Marine Biology Research at Bahamas

Unique and exclusive partnership with the Gerace Research Center (GRC) in San Salvador, Bahamas to offer marine biology research opportunities for high school teachers and students.

- Terra has exclusive rights to offer the program to high school teachers and students around world.
- All trips entail extensive snorkeling in Bahamian reefs as well as other scientific and cultural activities.
- Terra will schedule the program with GRC and book the flights from US to the GRC site.
- Fees include travel within the US to Island, lodging, meals, and hotels for transfers, and courses.
- For more information, please visit terraed.org/bahamas.html

Terra is a N.Y. based 501.c.3 non-profit organization dedicated for improving K-16 education
<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Multiplexed Immunohistochemistry Reveals Cancer-Reactive Germinal Centers are Enriched with CD8+ and Th Cells</td>
<td>Aadi Bhattacharya</td>
</tr>
<tr>
<td>7</td>
<td>Aurora Kinase, EZH2, and BET Inhibitor Drug Synergy in Glioblastoma Multiforme</td>
<td>Ambica Sharma</td>
</tr>
<tr>
<td>14</td>
<td>Random Forest Identification of Pulsars</td>
<td>Ankhita Ravishankar Satyanur</td>
</tr>
<tr>
<td>23</td>
<td>A Search for Monochromatic Light from the Andromeda Galaxy</td>
<td>Ariana Paul</td>
</tr>
<tr>
<td>28</td>
<td>Oncolytic Virotherapy and The Immune System</td>
<td>Asya Yilmaz</td>
</tr>
<tr>
<td>33</td>
<td>Gendered Online Toxicity During COVID-19</td>
<td>Christine Lee</td>
</tr>
<tr>
<td>37</td>
<td>Development of Herb-based Moist Patch for Animal Healing Using Arginate Biofilm</td>
<td>Seungah Claire Cho</td>
</tr>
<tr>
<td>42</td>
<td>Analyzing the Views of Ethnic Minority Teenagers in Texas on Organ Donation and Organ Donation Policies</td>
<td>Emily Wang</td>
</tr>
<tr>
<td>51</td>
<td>A Novel Rapid Detection Method of <em>Escherichia coli</em> using PCR-based Nucleic Acid Amplification Using pH-sensitive Dyes</td>
<td>Eugene Shim</td>
</tr>
<tr>
<td>54</td>
<td>Caffeine Induces Adenosine A2A Receptor Triggering Brain Cell Damage in A172 Brain Cells</td>
<td>Hyunmin Lee</td>
</tr>
<tr>
<td>58</td>
<td>Time Evolution of the Mass of Primordial Black Holes in a Cosmic Microwave Background with a Time-Dependent Temperature</td>
<td>Jonathan Lee</td>
</tr>
<tr>
<td>63</td>
<td>Food Insecurity among US Households during the Covid-19 Pandemic: Prevalence and Inequality</td>
<td>Julia Yi</td>
</tr>
<tr>
<td>69</td>
<td>A Novel Association of EGFR Gene Alteration with Decreased Glioblastoma Patient Survival Rate</td>
<td>June Hyun</td>
</tr>
<tr>
<td>74</td>
<td>Certificate-Based Key Replacement in Single Use Hash-Based Signatures: A Post-Quantum Technique</td>
<td>Matthew D. Grupenhoff</td>
</tr>
<tr>
<td>80</td>
<td>Entrepreneurial Struggle: A Natural Language Processing Approach</td>
<td>Peter J. Yu</td>
</tr>
<tr>
<td>86</td>
<td>Association of Social Determinants of Health with Respiratory Disease: A County-by-County Analysis for Ohio</td>
<td>Rohan M. Desai</td>
</tr>
<tr>
<td>89</td>
<td>Automated Debugging System: Implementing Program Spectrum Analysis and Information Retrieval on Fault Localization</td>
<td>Yin-Siang Hao</td>
</tr>
<tr>
<td>98</td>
<td>Genomic Sequencing in Clinical Oncology</td>
<td>Aashna K. Soni</td>
</tr>
<tr>
<td>107</td>
<td>Informal Economy and Spending Shares</td>
<td>Tanush Sarubney</td>
</tr>
<tr>
<td>117</td>
<td>A Rapid One-step Field-based Detection Method for Nanoplastics in Water</td>
<td>Vidhatri L. Iyer</td>
</tr>
<tr>
<td>122</td>
<td>The Shortage of Access to Therapists in Cambridge, Massachusetts in the United States: A Pandemic Study</td>
<td>Ashton T. Yoon</td>
</tr>
<tr>
<td>126</td>
<td>Studies on Environmental Impact and Energy Efficiency of Artificial Intelligence and Machine Learning</td>
<td>Yavuz E. Damkaci</td>
</tr>
<tr>
<td>134</td>
<td>A Meta-analysis of Syndromic Autism Genes</td>
<td>Daria Lunina</td>
</tr>
</tbody>
</table>
Dr. Isaac Chan, UT Southwestern Medical Center Dallas, Texas.
Giuseppe Franchini, Dept. of Engineering and Applied Sciences.
Dr. Wolf-Dieter Steinmann, Thermal Process Techn. Stuttgart.
Dr. Noah Kalman, Miami Cancer Institute.
Dr. Michael Chuong, Radiation Oncologist Miami Cancer Inst.
Dr. Kavita Rawat, Zoologist at Delhi University.
Dr. Usha Sharma, Senior Program Scientist at NIAID.
Dr. Qiangmin Zhang, Allegheny Cancer Institute; Pittsburgh
Dr. Minyong Chen, Scientist III New England Biolabs.
Dr. Jiying Ning, Univ. of Pittsburgh; Pittsburgh/PA/USA.
Dr. David S. Millington, Duke University Medical Center.
Kavya Kadia, Udgam School For Children, Ahmedabad.
Byoung-young Choi, Korea Inst. of Geoscience and Mineral
Sang-Ryul Kim, Seongnam Federation for Envir. Movement.
Gavan Fitzsimons, Duke University Durham/NC/USA.
Donald Lehmann, Columbia University New York City.
Dr. Ravi Kumar Kadeppagari, Centre for Incubation,
Innovation, Research and Consultancy Karnataka, India.
Dr. S. Maheswari, Dep of Microbiology Indian Academy College
Harlan Downs-Tepper, Duke University.
Emir Yazici, Kirlareli University Turkey.
Jose Alemán, Fordham University.
Christopher Goscha, Université du Québec à Montréal.
VU Duc Liem; Trần Ngọc Dung, Ho Chi Minh Acad of Politics.
Ishara Bandara, School of Information Techn., Deakin Univ
Jiaxin Cai Xiamen, University of Technology.
Dmitrii Pugachev, INSEAD, France.
Vasudeva Sharma, Ex. Director, Child Rights Trust, India.
Dr. Donna D. Zhang, Musil Family Endowed Chair in Drug
Discovery Professor of Pharmacology and Toxicology.
R. Ken Coit, College of Pharmacy University of Arizona. Prof.
Ann Cowan, UConn Health, Farmington, Connecticut
Steven Girvin, Yale University.
James Whitfield, Dartmouth College.
Sungyoung Lee, YntoAb (CEO).
Euigil Jung, C. Life Lab, Namyangju-si, Gyeonggi-do Korea.
Sukhwa Kweon, Artinlabs(CEO), Seoul/ South Korea.
Song Sungho, Gyeonggi-do Agricultural Research & Extension
Services/ Hwaseong-si, Gyeonggi-do. South Korea
Dr. Julia Schaletzky, University of California.
Multiplexed Immunohistochemistry Reveals Cancer-Reactive Germinal Centers are Enriched with CD8+ and T\textsubscript{fh} Cells

Aadi Bhattacharya

Rye Country Day School, 3 Cedar Street, Rye, NY, 10580, USA; aadi.gbh@gmail.com

Mentors: Dr. Ramin Herati, Ms. Jen Doran

ABSTRACT: Lung cancer kills hundreds of thousands of people worldwide each year. Sites of immune cell activity called germinal centers (GCs) in lymph nodes are associated with improved survival outcomes in cancer patients. The mechanisms by which GCs influence patient survival are not well-understood but could hold the key to further improving survival. Our goal was to investigate the differences in the physical and cellular characteristics of healthy and cancer-reactive GCs to help elucidate the role of GCs in patients' response to cancer. To do so, GCs in a lung cancer–draining mediastinal lymph node were compared to GCs in a non-cancerous tonsil. We used multiplexed immunohistochemistry for 22 antibody targets to analyze the samples, thereby allowing simultaneous analysis of multiple biomarkers in a tissue sample. Although most cell types were present in similar proportions in the GCs of each tissue, there were more CD8+ T cells and follicular helper CD4+ T cells in the GCs of the cancer-draining node. A better understanding of the role of these cell types may lead to improved outcomes for lung cancer patients.

KEYWORDS: Biomedical and Health Sciences; Immunology; Cancer; Germinal Centers.

Introduction

Of all cancers, lung cancer is the leading cause of death in the world, killing more than 160,000 people each year in the US alone.\(^1\) Research into the causes and treatments of lung cancer shows that the immune system plays a vital role against the disease.\(^2\) In particular, tertiary lymphoid structures containing germinal centers (GCs) have been linked with improved survival outcomes in cancer patients.\(^3\) GCs are typically found in secondary lymphoid organs (SLOs), such as lymph nodes and tonsils.\(^4\) Still, their role in cancer is not well-understood. The cellular composition of tumor draining lymph nodes is also not widely studied, and little is known about the levels of specific T cell and B cell subgroups in comparison to healthy SLOs.\(^5\)

Although the presence of GCs is associated with improved survival likelihood,\(^3\) it is unknown which specific aspects of immunobiology are the most relevant. There are known relationships between T follicular helper (Tfh) CD4\(^+\) cells, B cells, and CD8\(^+\) T cells that are associated with increased survival in cancer patients.\(^6\)–\(^14\) It has been hypothesized that interactions of these adaptive immune cells in the GCs may contribute to this result. However, little is known about whether cancer-reactive and non-cancer-reactive GCs differ in biology, although a previous study comparing healthy tonsil and healthy lymph node GCs found few cellular differences between the two.\(^15\) Understanding the potential differences between these GCs could help clarify why some patients with cancer-associated GCs have higher survival chances than those without.

Tissue imaging methods that can detect several antibody targets in a sample can be used to study GCs. Conventional means of studying tissues, such as traditional immunofluorescence (IF), single–cell RNA sequencing (scRNA-seq), and flow cytometry, have yielded significant insights into biology but also have drawbacks. Traditional four-parameter IF can retain the spatial data of the tissue but can only resolve a small number of proteins in a tissue sample.\(^16\) In contrast, flow cytometry and scRNA-seq can resolve many more proteins but lose spatial information by dissociating tissue into single-cell suspensions.\(^16\)–\(^18\) The CO-Detection by IndEXing (CODEX) platform\(^19\)–\(^20\) aims to overcome these limitations by using reporters with oligonucleotides complementary to oligonucleotide-conjugated antibodies bound to the tissue. Reporters are imaged in groups followed by the removal of reporters and the addition of the next cycle's reporters. This process is repeated cyclically until all antibody targets are resolved. This approach gathers detailed cellular and biomolecular information in the context of spatial tissue organization and has opened the possibility of studying tissue, including GCs, in unprecedented detail.

We used the CODEX platform to compare cancer-reactive GCs from one thoracic lymph node of a non-small cell lung cancer patient to normal GCs in a tonsil from a different, cancer-free patient. We identified the major T and B cell subsets using high-dimensional clustering.

Methods

One non-small cell lung cancer mediastinal lymph node and one unmatched tonsil were formalin-fixed and paraffin-embedded (FFPE) prior to use in our study. Lymph node and tonsil were provided de-identified by the NYU Center for Biospecimen Research and Development, respectively, after being obtained under informed consent. Clinical information about the patients was not available.

The 5-micron thick tissue was mounted onto a poly-L-lysine-coated coverslip. Tissue staining was performed using the manufacturer's (Akoya Biosciences) instructions. The paraffin was melted at 64°C for twenty minutes, and the sample was incubated in 100% ethanol, then 90% ethanol, 70%
ethanol, 50% ethanol, and finally 30% ethanol solutions for five minutes each. The tissue sample was washed in distilled water twice for five minutes each at room temperature, and subsequently immersed in a 1M solution of sodium citrate (pH=6) at 11 psi for twenty minutes. The beaker was equilibrated to room temperature for ten minutes, after which the staining rack was immersed in distilled water to remove the citrate from the tissue. The tissue was placed into CODEX hydration buffer for four minutes and CODEX staining buffer for five minutes. A blocking buffer solution was also prepared following the manufacturer’s instructions. To create the antibody cocktail, 148.7 µL of this blocking solution was added to the dilutions of each of the 22 antibodies studied in this experiment, shown in Table 1, which were selected to identify specific subsets of B, T, epithelial, and cancer cells.²³ Some antibodies used were conjugated by Akoya, while most were manually conjugated. The conjugation process used blank cycles without reporters as controls to verify its success. Every antibody used in this experiment had been validated. CD19, CD20, and CD21 were used to identify B cells, while CD3e and CD45RO were used to identify T cells. CD4 and CD8 expression isolated the CD4+ and CD8+ T cell subgroups, and PD-1 and CD4 were used to isolate Tfh cells.

An empty pipette-tip box was filled to 20% capacity with distilled water to create a humidified chamber. The antibody cocktail was incubated on the tissue sample for three hours. After the staining, the tissue was incubated in CODEX staining buffer for four minutes and then placed in a 1.8% paraformaldehyde (PFA) solution at room temperature for ten minutes. After incubating in the PFA solution, the tissue was washed with PBS. The tissue section was then incubated in ice-cold methanol for ten minutes and washed in PBS. To make the final fixative solution, 20 µL of the CODEX fixative reagent was added to 1 mL of PBS. This solution was added to the tissue in a humidified chamber for twenty minutes.

A photobleaching solution was prepared by adding 25 mL of PBS, 4.5 mL of a 30% H2O2 solution, and 0.8 mL of a 1M NaOH solution. The tissue sample was placed in 4 ml of this solution in a six-well plate, sandwiched in between two bright LED panels at 25000 Lux for 45 minutes at room temperature. The tissue was then placed in a fresh 4 mL of the photobleaching solution and repeated. After the photobleaching was finished, the tissue was washed in PBS and refrigerated at 4°C until imaging.

The reporter plate was prepared by adding 2684 µL of nuclease-free water to 330 µL of a 10M CODEX buffer solution, 275 µL of an assay reagent, and 11 µL of a nuclear stain to create the reporter stock solution. For each cycle, every reporter that corresponded to each antibody in that cycle was added to the reporter stock solution at a 1:50 dilution into the corresponding wells of the reporter plate. After the reporter plate was prepared, a 1:2000 DAPI dilution was prepared, and the coverslip was placed into the stage insert. 700 µL of the diluted nuclear stain was added to the tissue sample to identify regions of interest for subsequent imaging.

The CODEX processor (version 1.8, Akoya Biosciences) was used with default parameters, employed background subtraction to remove autofluorescence, deconvolution to reduce out-of-focus light, extended depth of field to enable the collapse of a z-stack into the single best-focus image, and shading correction to adjust for optical shading. Cells were segmented using the default values of the radius, maximum cutoff, minimum cutoff, and size cutoff factor for the region-growing algorithm.²⁴ The processor, upon completion, produced an FCS file containing the expression levels of each marker for every cell, which was viewed in the CODEX Multiplex Analysis Viewer (MAV) (version 1.5.0.8) to assess the quality of the data.

### Results and Discussion

#### 1. Data Quality Assessments:

The data quality was measured qualitatively using both single-antibody expression and multiplexed images generated by MAV. Representative images of the expression for thirteen antibodies are shown (Figure 1).

![Figure 1: Representative images of expression for some antibodies for one germinal center from the cancer-reactive lymph node. DAPI is shown in white. Color denotes protein as indicated. The scale bar represents 720µm for each image.](image-url)

**Table 1:** List of antibody targets with the concentration and clone of each antibody.

<table>
<thead>
<tr>
<th>Target</th>
<th>Clone</th>
<th>Dilution</th>
<th>Manufacturer</th>
<th>Concentration</th>
<th>Clone</th>
<th>Dilution</th>
<th>Manufacturer</th>
<th>Concentration</th>
<th>Clone</th>
<th>Dilution</th>
<th>Manufacturer</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD45RO</td>
<td>CD3E-1</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
</tr>
<tr>
<td>PA688</td>
<td>CD122</td>
<td>1:150</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:150</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:150</td>
<td>Abcam</td>
<td>1.50</td>
</tr>
<tr>
<td>CD21</td>
<td>CD203</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
</tr>
<tr>
<td>PD-L1</td>
<td>CD86</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
</tr>
<tr>
<td>ND1</td>
<td>CD86</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
</tr>
<tr>
<td>CD4</td>
<td>CD4</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
</tr>
<tr>
<td>HLA-A</td>
<td>DRB1</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
</tr>
<tr>
<td>HLA-C</td>
<td>DRB1</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
</tr>
<tr>
<td>HLA-DR</td>
<td>DRB1</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
<td>Vacarıs</td>
<td>1:100</td>
<td>Abcam</td>
<td>1.50</td>
</tr>
</tbody>
</table>
We first asked whether proteins typically associated with T cells were indeed identifiable. To do this, the antibodies employed in downstream analysis were selected based on a combination of the signal quality and prior literature.²⁴ We observed expression of CD3e, CD4, CD8, CD45RO, and PD-1 in the GCs and T cell zones outside of the follicular mantle in accordance with prior studies.⁹ We then asked whether a similar association was found with B cells. Again, there was robust signal for proteins such as CD19, CD20, CD21, Ki-67, and HLA-DR inside the GC and in the follicular mantle, also in accordance with previous investigations.²⁴ Other antibodies, including Vimentin, Pan-cytokeratin, PD-L1, and Mac2/Gal3, are not expressed by cells in the GC and were excluded from further analysis. In addition, proteins such as Foxt3, ICOS, PAX5, T-bet, IRF4, CD11c, CD38, and CXCR3 did not show reliable staining in all regions of both tissue samples, so they were also not included in downstream analysis. These data demonstrate that CD3e, CD4, CD8, CD19, CD20, CD21, CD45RO, HLA-DR, Ki-67, and PD-1 could be used to identify relevant adaptive immune cells.

Next, the segmentation was evaluated for accuracy (Figure 2). Here, we asked whether the segmentation algorithm could distinguish cell boundaries such that only one nucleus was assigned to each region versus manual assessment. Across three distinct tissue regions, there were 114 errors out of 1438 total segmentation events, representing a 92% segmentation accuracy.

Figure 2: Segmentation overlay is shown for a cancer-reactive lymph node. DAPI is shown in white. Colored lines represent software-generated cell boundaries. The germinal center (scale bar is 360µm) (a) is also shown at higher magnification (scale bar is 80µm) (b).

To test whether known GC cell populations were identified, we manually drew boundaries around each GC in both samples based on the relative DAPI levels, CD21 expression, and Ki-67 expression in each follicle (Figure 3a). The X-shift clustering algorithm²⁴ was used to classify the GC cells into different subsets. This algorithm was performed first on B cells and separately for T cells. B cells clustered into two major clusters based on the expression of Ki67, whereas T cells clustered into three clusters representing T\textsubscript{fh}, CD4\textsuperscript{+}, non-T\textsubscript{fh} CD4\textsuperscript{+}, or CD8\textsuperscript{+} T cells.⁹ These results affirmed that the adaptive immune cells in the GC could be studied using this immunofluorescent method.

Next, to validate the results of the clustering algorithm, the relative expression levels of the antibodies used in analysis on each cell population identified were mapped (Figure 3b, 3c) and compared to the literature.²² We asked whether the expected proteins associated with B cells were indeed expressed on the GC-B cells. We found that CD19, CD20, CD21, and HLA-DR were expressed on these cells, with Ki-67 also highly expressed on the cells in the Ki-67+ cluster, as expected. Similarly, CD4\textsuperscript{+} T cells and Th\textsubscript{h} cells had higher expression of CD4 than other clusters, whereas CD8\textsuperscript{+} T cells had the highest CD8 expression. T\textsubscript{fh} cells were identified as the subset of CD4\textsuperscript{+} T cells expressing PD-1, as PD-1 is strongly expressed on T\textsubscript{fh}.⁹ All three T cell subgroups expressed more CD3e and CD45RO than the B cell clusters. These data demonstrated that the clustering algorithm identified cell subsets appropriately.

2. Comparing Morphological Characteristics of Germinal Centers:

We next asked whether the GCs in the cancer-draining lymph node were qualitatively different from GCs in the tonsil. GC boundaries were manually determined to facilitate the identification of cells within the GC in the healthy (Figure 4a) and cancer-reactive (Figure 4b) tissues.

Figure 4: Germinal centers were manually identified based on tissue morphology (blue). The white scale bar is 2.25mm in both images. a) GCs in the healthy tonsil tissue. b) GCs in the cancer-reactive lymph node.
samples. A greater proportion of all cells in the tissue section were located in the GCs in the tonsil than in the cancer-reactive lymph node (Figure 5b). Indeed, 8% of lymph node cells in the cancer-reactive lymph node were in GCs while 13% of tonsil cells were inside GCs. Finally, we assessed whether GCs were similarly sized. To do this, we identified the total number of cells in each GC and compared the averages in each tissue (Figure 5c). The tonsil GCs contained more cells than the lymph node GCs, as the average lymph node GC had 929 cells, compared to 2641 cells in the tonsil GCs. To determine statistical significance, we used a two-tailed hypothesis t-test, where the two sample sizes were the number of GCs analyzed in each tissue, and the comparison was performed on the mean number of cells per GC in both tissues. We obtained a p-value of 0.007, indicating that the observed differences in GC size are significant.

3. Comparing Cellular Characteristics of Germinal Centers:

To assess the cellular differences between the GC in different tissues, we compared the clusters generated by X-shift clustering between the cancer-reactive lymph node and the tonsil control. B cells constituted the majority of cells in both tissues, followed by CD4+ T cells, while CD8+ T cells were relatively rare (Figure 6a, b).

To quantitatively assess whether there were differences in GC composition, we determined the proportion of each cellular subset in the germinal center by dividing the total number of each cell type across all GCs by the total number of GC-resident cells (Figure 6c). We also showed a relative difference in the cancer-reactive lymph node GC cell populations from the non-cancer-reactive GC cell populations (Figure 6d). We observed 413% and 49.9% more CD8+ T cells and Tfh cells, respectively, in the cancer-reactive GCs relative to the non-cancer-reactive GCs. These data suggest that while the composition of germinal centers is mostly similar across tissues, CD8+ T cells and Tfh cells were significantly more abundant in the tumor-draining tissue. Figures 6e and 6f also show the differences in CD8+ T cells and Tfh cells in cancer-reactive GCs and non-cancer-reactive GCs. We again used two-tailed hypothesis t-tests to determine statistical significance, where the two sample sizes were the number of GCs analyzed in each tissue type, and the comparisons were performed on the averages of the proportions of CD8+ T cells and Tfh cells in each GC for the two tissue types. We obtained p-values of 0.008 for the comparison of CD8+ T cells and 0.025 for the comparison of Tfh cells, again indicating that the observed differences in the proportions of these two cell types are meaningful.

Figure 6: A cancer-reactive lymph node (a) and non-cancer-reactive tonsil (b) germinal center. Cell identities colored as shown. Blue arrows identify some CD8+ T cells. The scale bar represents 380μm. (c) GC cellular subset composition in the proportion of GC cells that were B, CD4+ T, CD8+ T, Tfh, Ki67+ B, and Ki67- B cells for both the tonsil and the lymph node. (d) The percent difference in proportions in subset composition in the cancer-reactive lymph node GCs compared to the tonsil GCs. (e) Percentage of CD8+ T cells in each GC. Error bars were constructed using an 80% confidence interval. (f) Percentage of Tfh cells in each GC. Error bars were constructed using an 80% confidence interval.

Conclusion

This research aimed to identify differences between cancer-reactive and non-cancer-reactive germinal centers using multiparameter immunofluorescent imaging of tissue because patients with cancer-reactive GCs have been shown to be associated with having higher likelihoods of survival. Using the CODEX platform, we compared GCs from a tumor-draining thoracic lymph node with GCs from a non-cancerous tonsil. We made several observations during this study. Using an optimized panel of antibodies targeting key proteins in the GC, we identified major cell types using manual and high-dimensional methods in each tissue. We looked at the levels of B cells, CD4+ T cells, CD8+ T cells, and Tfh cells, which are the most common lymphocytes that interact inside GCs. We found fewer GCs in the cancer-reactive tissue from these data compared to the non-cancer-reactive tissue. However, the GCs in the cancer-reactive GCs had more CD8+ T cells and Tfh cells. Together, these data highlight the importance of incorporating spatial biology into studying the immune response to cancer.

Our data observed qualitative and quantitative differences in GC biology between the two tissue types. The higher percentage of cells in GCs, number of GCs per unit area, and number of cells per GC in the non-cancer-reactive GCs were likely due to the noncancerous GCs being derived from tonsils. Tonsils are exposed to high levels of antigens derived from the mouth and upper airway, and tonsillar GCs have been studied as the “prototypical” GC in humans for decades. On the other hand, lymph nodes are likely less exposed to oral bacterial antigens and thus harbor less GC activity. Higher antigen exposure has been shown in previous literature to be correlated...
to an increase in GC formation,\textsuperscript{27} thus the positions of the two SLOs in the human body could make a difference in the sizes and quantities of GCs. However, other qualitative differences remain, as the cellular composition of healthy tonsil GCs and healthy lymph node GCs has been shown to be similar.\textsuperscript{20} We observed differences in the proportions of different cell types in noncancerous and cancer-reactive GCs, with elevated numbers of CD8$^+$ T cells and Tfh cells in the cancer-reactive GCs (Figure 6a-e). Thus, the differences observed here may reflect immune responses to the nearby tumor.

In previous studies, CD8$^+$ T cells are associated with the maintenance of cytolytic potential in the tumor and the lysis of tumor cells.\textsuperscript{28,29} Because of their well-described role in elimination of tumor cells, it is possible that they are a key factor in cancer-reactive GCs being associated with improved survival likelihoods. However, it is unclear when and where these cells are “licensed” to join the immune response. We observed a higher proportion of CD8$^+$ T cells in some, but not all, GCs in the cancer-draining lymph node relative to tonsillar GCs. One possibility is that only some of the GCs in the cancer-draining lymph node respond to the tumor, whereas the others may not be tumor-reactive. Because tumor-infiltrating CD8$^+$ T cells are associated with improved survival,\textsuperscript{7} determining whether the GCs in the tumor-draining lymph node with high CD8$^+$ T cell counts are indeed tumor-reactive would be of great interest. Indeed, the increased presence of CD8$^+$ T cells in cancer-reactive GCs could indicate an important role for these cells in the immune response to cancer.

Further study of additional proteins in these CD8$^+$ T cells would help clarify their phenotype and provide deeper insights into the specific roles that they may play in cancer-reactive GCs. For example, expression of Tim3 could imply reduced functional exhaustion, whereas positivity for CXCR5 could confirm the CD8$^+$ T cells’ ability to promote an antitumor response.\textsuperscript{27-30} Future experiments could ask whether CXCR5$^+$CD8$^+$ T cells or Tim3$^+$CD8$^+$ T cells, among other subsets, are increased in cancer-reactive GCs. It is important to understand exhaustion and activation levels in addition to just the number of CD8$^+$ T cells to understand whether these cells are actually having an impact on the tumor response. Such investigations into how these CD8$^+$ T cells interact with and influence tumors would help assess how to expand the improved outcomes to all patients.

Previous studies\textsuperscript{31} have also indicated that the CD8$^+$Tfh cell axis also has an important role in the cancer response. The accumulation of Tfh cells is an important contributor to antitumor responses that depend on cytotoxic CD8$^+$ T cells, and the absence of Tfh cells can lead to CD8$^+$ T cell dysfunction, lower cytokine production, and reduced cytotoxic capabilities.\textsuperscript{32} The increase in Tfh cells may be playing a role in supporting an antitumor response by the CD8$^+$ T cells. IL-21 produced by Tfh cells can enhance CD8$^+$ T cell function in the tumor, but little is known about whether this occurs outside the tumor as well. Future studies will be needed to assess whether this mechanism is occurring in cancer-reactive GCs.

The data from our studies help to connect the associations between GCs and patient response to cancer with the correlation between CD8$^+$ T cells and cancer outcomes. This can highlight opportunities for future study of the relevance of cancer-reactive GCs and the potential for their targeting to improve cancer outcomes.

\section*{Acknowledgments}
I would firstly like to thank Dr. Ramin Herati and his lab for their endless help with my research and data analysis. I would also like to thank my science research teacher, Ms. Jen Doran, for her consistent support, and the IJHSR team for providing the unique opportunity for me to present this work.

\section*{References}
\begin{enumerate}
\item Dela Cruz, C., Tanoue, L., Matthy, R. Lung Cancer: Epidemiology, Etiology, and Prevention; Clinics in Chest Medicine 32(4), 605–64: December 2011.
\item Mishra, A., Maruizza, R. Insights into the Structural Basis of Antibody Affinity Maturation from Next-Generation Sequencing; Frontiers in Immunology 9(117): February 2018.
\item Chandrasekaran, S., King, M. Microenvironment of Tumor Draining Lymph Nodes: Opportunities for Liposome-Based Targeted Therapy; International Journal of Molecular Sciences 15(11), 2020-2039: November 2014.
\item Stebbeg, M., Kumar, S., Silva-Cayetano, A., Fonseca, V., Linterman, M., Graca, L., Regulation of the Germinal Center Response; Frontiers in Immunology 9(2469): October 2018.
\item Crotty, S. T Follicular Helper Cells: A Decade of Discovery and Diseases; Immunity 50(5), 1132-1148: May 2019.
\item Choi, S., Morel, L. Immune Metabolism Regulation of the Germinal Center Response; Experimental and Molecular Medicine 52, 348-357: May 2020.
\item Vaccari, M., Franchini, G. T Cell Subsets in the Germinal Center: Lessons from the Macaque Model; Frontiers in Immunology 9(348): February 2018.
\end{enumerate}


27. Cirelli, K., Crotty, S. Germinal Center Enhancement by Extended Antigen Availability; Current Opinion in Immunology 47, 64–69: August 2017.


Author

Aadi Bhattacharya is a high school senior at Rye Country Day School in Rye, New York. He has grown interested in biological research in cancer immunology over the course of this research and hopes to study biomedical science/engineering in college.
Aurora Kinase, EZH2, and BET Inhibitor Drug Synergy in Glioblastoma Multiforme

Ambica Sharma

Washington-Liberty High School, 1301 N. Stafford St, Arlington, VA, 22201, USA; ambicasharma0122@gmail.com
Mentors: Thomas J. Walter, Nagi Ayad, Jason Brodowski

ABSTRACT: Glioblastoma (GBM) is the most common primary brain cancer with an incidence rate of 3.21 per 100,000 people. Even with standard-of-care treatment, the median survival time is 15 months. Therefore, new, more effective therapies are sought. In recent years, interest in combination therapies has grown because of the ineffectiveness of many monotherapies. In particular, combination therapies present the possibility of synergistic drug-drug interactions, which allow for promising clinical implementation. This investigation aims to create novel compound combinations of Aurora kinase, EZH2, and BET inhibitors to determine whether pharmacotherapies such as these can effectively treat GBM in vitro as an immunotherapy by interacting synergistically with one another. With five different combinations within two separate cell lines (LN229 and GBM22) of an Aurora kinase inhibitor drug (alisertib), EZH2 inhibitor drug (tazemetostat), and BET inhibitor drug (UM-002), the combination therapy of alisertib and UM-002 tested in vitro in LN229 cells displayed the most positive and synergistic results. Compared to the other combinations tested, the inhibition percentages were at least 34% greater for the alisertib and UM-002 combination, indicating promising drug-drug interactions between Aurora kinase and BET inhibitors which cause the GBM cell survival to decrease significantly. This illustrates the possibility of highly effective immunotherapy for this malignant disease with poor prognosis and survival.

KEYWORDS: Biology; Cancer Biology; Neuroscience; Glioblastoma Multiforme; Drug Synergy.

Introduction

Glioblastoma multiforme (GBM) is one of the most aggressive types of brain tumor.¹ It is categorized under the more common glialoma, a primary brain tumor classified according to its particular cell of origin; among these are astrocytic tumors, oligodendrogliomas, ependymomas, and mixed gliomas, which are all tumors related to the brain’s central nervous system.² Indeed, GBM is the most common primary malignant brain tumor as it accounts for approximately 60% of all brain tumors in adults.³,⁴ It is classified as a grade IV glioma and grade IV astrocytoma.¹ As the cancer invades adjacent tissue, distant organs are generally unaffected. GBM usually occurs in the cerebral hemispheres, especially in the frontal and temporal lobes.⁵ Although modern therapies continue to evolve for GBM (current standard treatment is surgery, radiation, and chemotherapy with temozolomide), due to its poor prognosis, patients are expected to have a median survival rate of 15 months post-diagnosis.⁴

GBM is highly resistant to several forms of therapeutic intervention, mainly due to its incredible complexity.¹ GBM is multiform grossly, microscopically, and genetically: the etiology of GBM consists of numerous regions of necrosis and local or systemic hemorrhage; microscopically, regions of pseudopalisading necrosis, pleomorphic cells and nuclei, and microvascular proliferation can be seen; genetically, GBM consists of several deletions, amplifications, and point mutations, which lead to the activation of abnormal cell production.⁶-⁸ Additionally, GBM shows intratumor genetic heterogeneity with subclones within the tumor cell population.⁹ It has been estimated that p53-deficient and cultured neoplastic cells that occur with GBM could have mutations in any gene at a rate of 1 in 1,000 cells.¹⁰ Assuming this is correct for GBM (in vivo), a tumor of 1 billion cells could harbor as many as 1 million cells with mutations in any given gene.¹¹

Nevertheless, current therapies for GBM include surgery (where a neurosurgeon works towards manually resecting the GBM), radiation therapy (use of high energy beams to kill cancer cells), chemotherapy (use of drugs), tumor treating fields therapy (use of an electric field to disrupt cancer cells from multiplying), targeted drug therapy (focus on abnormalities), and clinical trials (where patients can take part in studies leading to advanced treatment).² These standard treatments have been somewhat effective to the extent that GBM patients live around 2-4 more years. However, this regimen is not highly curative (not every tumor cell is killed), hence increasing the need for a significantly effective treatment strategy.¹²

Immunotherapy, a treatment in which the immune system is manipulated to attack tumor cells and minimizes adverse effects, has been on the rise as a possible advancement in the treatment of GBM.¹² Most immunotherapies are tested through clinical trials. By participating in these clinical trials, immunotherapies can lead to advances in treating GBM. Current clinical trials for GBM predominantly focus on peptide vaccines, adoptive T-cell therapy, oncolytic virotherapy, dendritic cell vaccines, and checkpoint inhibitors; as monotherapies prove ineffective; an alternative treatment strategy consists of combination therapies: a treatment modality that combines two or more therapeutic agents.¹⁵ Combination therapies present the
possibility of positive or even synergistic drug–drug interactions, which could conclude with clinical implementation.°

Aurora kinases, a family of serine and threonine kinases, regulate centriole and microtubule function and play an essential role in maintaining normal mitosis and regulated meiosis.° Overexpression or gene amplification of Aurora kinases leads to aneuploidy — the state of abnormal chromosome numbers that deviate from a multiple of haploid complement — thereby leading to cancer. In past decades, a series of Aurora kinase inhibitors (AKIs) developed have successfully repressed the progression and growth of many cancers both in vivo and in vitro, suggesting that Aurora kinases could be a novel therapeutic medium for the treatment of GBM.°² Alisertib, an AKI drug, has been extensively characterized using in vivo and in vitro preclinical models. It has been shown to display antiproliferative activity in various human tumor cell lines, including glioblastoma.°

The enhancer of the zeste homolog 2 (EZH2) gene provides instructions for making an enzyme modifying histone proteins. It has been implicated in oncogenesis as the catalytic methyltransferase within the PRC2 protein. This gene is overexpressed in cancers and is correlated with a lower survival rate of GBM. Pharmacological inhibition of EZH2 activity eventually leads to a reduction in tumorigenicity in GBM.°¹ Tazemetostat, an EZH2 inhibitor drug, has previously been shown to reduce viability in GBM cell lines. Furthermore, it is the first Food & Drug Administration-approved EZH2 inhibitor drug, suggesting it has undergone extensive trials to test its efficacy and safety for patients.°²

The bromodomain and extra-terminal (BET) proteins act as epigenetic readers with broad specificity on transcriptional activation, which includes the recruitment of positive transcription elongation factor and control of RNA polymerase II transcriptional activity. Bromodomain-containing protein 4 (BRD4) is a target in multiple cancers (in GBM, the inhibition or depletion of BRD4 reduces the expression of oncogenes).° Small molecule BET inhibitors reduce the growth of GBM and other brain tumors by competing with BET-histone interaction (reduced transcription of oncogenes essential for GBM cell interaction). There are multiple BET inhibitors, yet very few are brain penetrant: BET Inhibitor JQ1 is widely used in research and is brain penetrant, yet it is not clinically used because of its short half-life. BET Inhibitor; MK-8628 was used in a clinical study but was terminated when it was found to have no significant effect on GBM. BET Inhibitor; UM-002 is modified to increase potency (more than JQ1 or MK-8628) in reducing GBM cell proliferation in vitro.°²

To test each compound combination, synergy matrix screens were carried out. The standard plating and drugging procedure used for cell culture was followed. There are many ways to measure cell viability, including MTT Proliferation Assay. The MTT assay measured cellular metabolic activity as an indicator of cell proliferation, viability, and cytotoxicity. Instead of counting each individual cell, this colorimetric assay is based on the reduction of a yellow tetrazolium salt (expressed as 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide or MTT) to purple formazan crystals through metabolically active cells. Since the viable cells contain NAD(P)H-dependent oxidoreductase enzymes, the MTT is reduced to formazan. A solubilization solution is used to dissolve the insoluble formazan crystals, and the resulting colored solution is quantified by measuring absorbance at 500–600 nanometers with a cell plate reader. The darker or more purple the solution, the greater the number of viable, metabolically active cells.°³

Small molecule inhibitors each feature the ability to inhibit GBM proliferation individually, yet are of limited effectiveness against this highly aggressive brain tumor. This investigation aims to test novel compound combinations of these inhibitors to determine whether pharmacotherapies such as these, by interacting synergistically with one another, can more effectively treat GBM in vitro.

**Methods**

Cells were cultured in complete medium (45 mL Dulbecco’s Modified Eagle Medium, 5 mL Fetal Bovine Serum, and 1 mL Pen/Strep). Once cells were sufficiently confluent within a 15mL cell culture flask, the complete media was suctioned. Cells attached to the flask’s surface were separated with 5 mL of trypsin, which was added to the bottom of the flask for approximately 4 minutes and neutralized with 5 mL of complete media, ensuring the neutralized trypsin reached every part of the bottom of the flask. The cells were resuspended and counted with a hemocytometer and trypan blue to determine the volume of complete media needed to plate 5,000 cells per well. Cells and complete media were plated at 5000 cells/well in each of 64 wells, in an 8 x 8 pattern, for four 96-well plates and placed in a cell incubator.

After 24 hours, the complete media was removed from each of the wells, careful not to suction many cells. 200 µL of fresh complete media was added to all the wells. Cells were then treated with each drug in 7 1:4 serial dilutions, starting at 10 µM with a final Dimethyl Sulfoxide (DMSO) concentration of 0.25 – 0.5%, depending on the solubility of the solid drug in a DMSO and Phosphate Buffer Saline (PBS) stock solution. The positive control or vehicle was 25% DMSO. 10 µM Velcade was the negative control. Drugged cells were incubated for 72 hours.

10X stock solution of MTT was diluted with sterile 1X PBS to create the MTT working solution. Media was removed from cells. 100 µL of MTT working solution was added to each well. Cells were incubated for 1 hour. At the end of the incubation time, 100 µL of solubilization solution was added. The cells were kept at room temperature for another hour to allow the converted dye to dissolve completely. With a cell plate reader, the absorbance of the converted dye was measured at 570 nm and a reference wavelength of 650 nm. Each plate was measured without the cover to avoid condensation or smudge on the cover affecting the reading.

It was essential to take proper safety precautions during this experiment. All the experiments were conducted in a research laboratory with rules and safety precautions. Proper PPE equipment was also used throughout the investigation.
Results and Discussion

The optical density or transmission of light values of each well (produced as an Excel spreadsheet by the cell plate reader) were analyzed using the SynergyFinder software, which produced graphs depicting inhibition and synergy levels for each drug combination. A positive result indicates the inhibition of proliferation, and a negative result indicates the enhancement of proliferation. A synergy score can be expressed as the average excess response due to drug interactions. A synergistic effect occurs when the sum of the effect is more than the two individual chemical effects combined. An additive effect occurs when the sum of the effect equals the two individual chemical effects combined. Hence, a synergistic effect is ideal since more drug-drug interactions would be occurring that would allow a greater positive effect on cell viability.

**Alisertib and Tazemetostat in LN229:**

![Figure 1](image1.png)

**Figure 1:** Alisertib shows positive results on the percentage inhibition of LN229 cells.

![Figure 2](image2.png)

**Figure 2:** Tazemetostat shows negative results on the percentage inhibition of LN229 cells.

![Figure 3](image3.png)

**Figure 3:** Alisertib and tazemetostat combination shows slightly positive effects on LN229 cell.

![Figure 4](image4.png)

**Figure 4:** Alisertib and tazemetostat combination shows a slight positive effect on the percentage inhibition of LN229 cells.

![Figure 5](image5.png)

**Figure 5:** Alisertib and tazemetostat combination shows no significant synergy in LN229 cells.

The alisertib and tazemetostat combination in LN229 displayed moderate results. As seen in Figure 1, there was a greater percentage of inhibition at higher doses of alisertib. Figure 2 suggests that higher doses of tazemetostat induced a negative inhibition response. The combination of alisertib and tazemetostat moderately reduced cell viability, as seen in Figure 3 and increased inhibition by an average of 16%, as seen in Figure 4. Furthermore, based on the low mean and high p-value displayed in Figure 5, there was no synergistic effect between the alisertib and tazemetostat combination in the LN229 cells.

![Figure 6](image6.png)

**Figure 6:** UM-002 shows positive results on the percentage inhibition of LN229 cells.

![Figure 7](image7.png)

**Figure 7:** Tazemetostat shows negative results on the percentage inhibition of LN229 cells.

![Figure 8](image8.png)

**Figure 8:** UM-002 and tazemetostat combination shows slightly positive effects on LN229 cell viability.
The alisertib and UM-002 combination in LN229 displayed the most positive results. Figures 11 and 12 show that both alisertib and UM-002 demonstrated greater inhibition at higher doses. Figures 13 and 14 suggest that cell viability significantly decreased, and inhibition increased by 51% and 39% with the alisertib and UM-002 combination. In addition, Figure 15 indicates that, compared to the previous assays, there is some synergy between alisertib and UM-002 in LN229 cells given the positive mean value, but the p-value is not significant.

Figure 10: UM-002 and tazemetostat combination shows no significant synergy in LN229 cells.

The UM-002 and tazemetostat combination in LN229 displayed more positive results than the alisertib and tazemetostat combination in LN229. From Figure 6, as UM-002 doses increased, there was a greater percentage of cell inhibition. Figure 7 suggests the same dose response with negative inhibition of tazemetostat, as seen in Figure 2. The combination of UM-002 and tazemetostat decreased cell viability slightly more and increased inhibition by 12% than the previous combination results, as seen in Figures 8 and 9 compared to Figures 3 and 4. Despite positive results, Figure 10 suggests no significant synergy between UM-002 and tazemetostat in LN229 cells based on the low mean value and high p-value.

Figure 11: Alisertib shows positive results on the percentage inhibition of LN229 cells.

Figure 12: UM-002 shows positive results on the percentage inhibition of LN229 cells.

Figure 13: Alisertib and UM-002 combination shows positive effects on LN229 cell viability.

Figure 14: Alisertib and UM-002 combination shows a highly positive effect on the percentage inhibition of LN229 cells.

Figure 15: Alisertib and UM-002 combination shows no significant synergy in LN229 cells.

The alisertib and UM-002 combination in LN229 displayed the most positive results. Figures 11 and 12 show that both alisertib and UM-002 demonstrated greater inhibition at higher doses. Figures 13 and 14 suggest that cell viability significantly decreased, and inhibition increased by 51% and 39% with the alisertib and UM-002 combination. In addition, Figure 15 indicates that, compared to the previous assays, there is some synergy between alisertib and UM-002 in LN229 cells given the positive mean value, but the p-value is not significant.

Figure 16: Alisertib shows positive results on the percentage inhibition of GBM22 cells.
The alisertib and tazemetostat combination in GBM22 displayed results similar to that in LN229 cells. Figures 16 and 17 suggest that alisertib induced a positive inhibition and tazemetostat induced a slightly positive inhibition response, with only 5% inhibition. Figures 18 and 19 show that this combination leads to a slight decrease in cell viability and positive inhibition. Figure 20 indicates no significant synergistic effects of the alisertib and tazemetostat combination in GBM22 cells.

The alisertib and tazemetostat combination in GBM22 displayed results similar to that in LN229 cells. Figures 16 and 17 suggest that alisertib induced a positive inhibition and tazemetostat induced a slightly positive inhibition response, with only 5% inhibition. Figures 18 and 19 show that this combination leads to a slight decrease in cell viability and positive inhibition. Figure 20 indicates no significant synergistic effects of the alisertib and tazemetostat combination in GBM22 cells.

Figure 17: Tazemetostat shows slightly positive results on the percentage inhibition of GBM22 cells.

Figure 18: Alisertib and tazemetostat combination shows slightly positive effects on GBM22 cell viability.

Figure 19: Alisertib and tazemetostat combination shows a slight positive effect on the percentage inhibition of GBM22 cells.

Figure 20: Alisertib and tazemetostat combination shows no significant synergy in GBM22 cells.

Figure 21: UM-002 shows positive results on the percentage inhibition of GBM22 cells.

Figure 22: Tazemetostat shows slightly positive results on the percentage inhibition of GBM22 cells.

Figure 23: UM-002 and tazemetostat combination shows slightly positive effects on GBM22 cell viability.

Figure 24: UM-002 and tazemetostat combination positively affect the percentage inhibition of GBM22 cells.

Figure 25: UM-002 and tazemetostat combination shows no significant synergy in GBM22 cells.
The UM-002 and tazemetostat combination in the GBM22 cell line also yielded similar results as the same combination in LN229. Again, Figures 21 and 22 show that the UM-002 demonstrated positive inhibition while the tazemetostat displayed negative inhibition. The combination of these drugs also decreased cell viability and led to a moderate increase in percentage inhibition, as seen in Figures 23 and 24. However, there was no significant synergy between the UM-002 and tazemetostat combination in GBM22, as displayed in Figure 25.

**Conclusion**

Experimental results suggest that novel combinations of aurora kinase, EZH2 and BET inhibitors display promise, suggesting the possibility of devising more effective immunotherapy for GBM. The alternate hypothesis was partially supported with one of the five combinations: there was evidence of some synergy between alisertib and UM-002. The other combinations did not suggest any synergistic effects due to the high p-values. While conducting the experiments, there may have been a random or systematic error source. For example, since the cell plate reader was utterly computerized, it was impossible to determine an inaccurate reading unless the MTT color did not correspond to the reading. Hence, the experimental design made it challenging to decide on any type of statistical error due to the methods with which the data was analyzed. Therefore, all the figures were derived from a website with only the p-value and no other statistical test.

As previously described, alisertib and UM-002 individually decreased cell viability and increased the percentage inhibition in LN229 and GBM22, supporting previous findings relating to these inhibitor drugs. However, especially in the LN229 cells, tazemetostat displayed a negative percentage inhibition, which suggests that the cells proliferate in response to being drugged with tazemetostat. This seemed to cause the drug combinations, including tazemetostat, not to display the expected positive and synergistic results. This is indeed a puzzling finding as previous studies have found tazemetostat to decrease cell viability and displayed the most promising results overall. Compared to the other varieties tested, with an average of 68%, the GBM cell inhibition percentages were at least 34% greater for the alisertib and UM-002 combination (Figures 20–24), indicating that there are highly positive drug-drug interactions occurring, which cause the cancer cell survival to decrease significantly. Since this combination was tested in a cell line such as LN229, which is commonly used for *in vitro* GBM experimentation, the ability of this combination to penetrate through the cells is very encouraging for the development of a dual aurora kinase and BET inhibitor therapy for GBM. Ideally, if tested *in vivo*, small doses of alisertib and UM-002 would be administered, and a much more pronounced effect (significant decrease of cancer cells due to drug–drug interactions) would occur, thereby minimizing side effects of these drugs while simultaneously more effectively eliminating as many GBM cells.

With glioblastoma as the most common primary brain cancer, highly effective treatments must be found to treat this disease. With compound therapies on the rise, the alisertib and UM-002 drug combination effectively reduced cell viability and displayed the most promising results overall. Again, the random, systemic, and possible statistical error in this experiment could have been a factor in some of the results, which only displayed minimal synergistic effects in that the highest synergy score was 2 out of 20. Limiting any possible error could significantly aid in producing statistically significant results. Future directions for this experiment would focus on conducting additional assays to support and elaborate on the data found from this experiment. In addition to the LN229 and GBM22 cell lines, other primary GBM cell lines, such as GBM6, GBM10, and GBM39, could be tested since they have been shown to demonstrate positive effects.

Since tazemetostat displayed negative inhibition, other EZH2 inhibitors could be tested to determine whether EZH2 inhibitors are unsuitable for combination therapies. In addition to aurora kinase, BET, and EZH2 inhibitors, other inhibitors shown to reduce GBM cell proliferation, such as LIM kinase and HDAC inhibitors, could be tested.

Most importantly, among the combinations examined, the alisertib and UM-002 combination effectively reduced cell viability and displayed the most promising results overall. Compared to the other varieties tested, with an average of 68%, the GBM cell inhibition percentages were at least 34% greater for the alisertib and UM-002 combination (Figures 20–24), indicating that there are highly positive drug–drug interactions occurring, which cause the cancer cell survival to decrease significantly. Since this combination was tested in a cell line such as LN229, which is commonly used for *in vitro* GBM experimentation, the ability of this combination to penetrate through the cells is very encouraging for the development of a dual aurora kinase and BET inhibitor therapy for GBM. Ideally, if tested *in vivo*, small doses of alisertib and UM-002 would be administered, and a much more pronounced effect (significant decrease of cancer cells due to drug–drug interactions) would occur, thereby minimizing side effects of these drugs while simultaneously more effectively eliminating as many GBM cells.

With glioblastoma as the most common primary brain cancer, highly effective treatments must be found to treat this disease. With compound therapies on the rise, the alisertib and UM-002 drug combination and, possibly, other Aurora kinase and BET inhibitor drug combinations could be promising therapy with highly efficacious results on GBM cell death.

**Acknowledgments**

Ambica Sharma thanks Thomas J. Walter, M.D./Ph.D. for assistance in the laboratory, Nagi Ayad Ph.D. for allotting time in the laboratory, and Jason Brodowski for assistance with revisions to the manuscript prior to publication.

**References**

Ambica Sharma is a senior at Washington-Liberty High School. She is passionate about neuroscience and interdisciplinary STEM research, especially in cell biology. She is looking forward to majoring in molecular biology in college.
Random Forest Identification of Pulsars

Ankhita Ravishankar Sathanur
Eastlake High School, 400 228th Ave NE, Sammamish, WA 98074, USA; ankhitas@hotmail.com
Mentor: Kristen Surraro

ABSTRACT: This study explores the implementation of a random forest classifier to identify pulsar signals from a large sample of survey data. Pulsars are a unique type of rotating neutron star that emit pulses of radio emission in beams that sweep across Earth, allowing for the detection of their repetitive pulses. Traditionally, pulsar candidates have been identified through manual signal processing. As data volumes increase, automated methods, like artificial neural networks, have been proposed. In this study, the random forest classifier—an algorithm that takes the majority output of multiple decision trees—was used to separate pulsar signals from radio frequency interference (RFI) and other noise. 1,639 real pulsar examples and 16,259 samples of RFI/noise from the HTRU2 survey were used to create the model. Features of the data used include the mean, standard deviation, excess kurtosis, and skewness of the integrated pulse profile and DM–SNR curve. The model demonstrated a 95% accuracy in identifying pulsars. The excess kurtosis, skewness, and mean of the integrated profile were determined to be the most critical factors in differentiating between pulsars and interference. This tool could filter data from future surveys to reduce the number of candidates that need to be processed by humans.

KEYWORDS: Robotics and Intelligent Machines; Machine Learning; Random Forest; Physics and Astronomy; Pulsars.

Introduction

Pulsars form from neutron stars. A neutron star is a leftover core from when a star with a mass between 8 and 20 solar masses dies in an explosive supernova. This core is extremely massive, around 1.5 to 3 times the sun’s mass, and very small, typically only around 20 to 24 km in diameter (roughly the size of a small city block).¹ These features contribute to a pulsar’s high density, comparable to that of atomic nuclei, second only to black holes.

One of the most notable features of pulsars is their rapid spin rate. This spin is left over from the original rotation of the living star from which the pulsar formed. As a star collapses into a neutron star, its spin rate increases dramatically: as the size of the star shrinks significantly, the rotation rate must increase to conserve angular momentum.

Pulsars are also highly magnetic, with a magnetic field 100 million to 1 million billion times stronger than Earth’s magnetic field.² A neutron star with the right combination of extreme magnetism and rapid spin is described as a pulsar.

The light pulsars emit can be attributed to the rotating magnetic field. This creates an electric field where charged particles move and form an electric current. As these charged particles are accelerated to high speeds by the magnetosphere, the region above the surface of a pulsar, the pulsar radiates visible, radio, and x-ray light. Pulsars emit these beams of light from their north and south magnetic poles, which typically differ from their axis of rotation. This allows their beams of light to be viewed as pulses from Earth, as the coherent light (emitted like a laser rather than a lightbulb) sweeps across Earth. However, if the magnetic poles aligned with the rotation poles, the light would be viewed as a steady stream rather than as pulses, as the beams would not move as the star rotates.

Pulsars are valuable tools in studying the universe. For one, they are helpful in studying extreme states of matter due to their high density, which creates what astronomers call a “nuclear pasta” as atoms arrange in unusual shapes and patterns. For example, within neutron stars, atoms arrange themselves in patterns such as flat sheets, spirals, and small nuggets that are not seen anywhere else.²

While pulsars reduce speed as they lose energy to light, their rotation rate changes at a slow, almost imperceptible rate. They are, therefore, so dependable that astronomers can predict when a pulse will occur years in the future with an accuracy of 100 nanoseconds. This accurate, reliable spin rate makes pulsars helpful in finding exoplanets, as a nearby planet will often cause noticeable disturbances in a pulsar’s spin. In fact, the first planet discovered outside our solar system was discovered by observing the effects of pulsars orbiting a pulsar. The constant spin of pulsars as they move through space also makes them a useful tool in measuring cosmic distances as they move through space while blinking a known number of times per second. Finally, pulsars are useful in searching for gravitational waves, which also create disturbances in the regularly timed pulses.

Pulsar Detection:

Pulsars are extremely rare, and only around 2,000 have been detected to date. A typical method for detecting pulsars is all-sky surveys, where a telescope scans the entire sky and looks for light that flickers over time. The Parkes radio telescope in Australia has discovered the majority of pulsars. Still, other radio telescopes, such as the Arecibo telescope in Puerto Rico, the Green Bank telescope in West Virginia, the Molonglo telescope in Australia, and the Jodrell Bank telescope in England,
have also made contributions. In addition, the Fermi Gamma Ray Telescope has also detected gamma-ray-emitting pulsars.

Each pulsar has a unique spin rate and radio pulse profile, both of which are used in their detection. Traditionally, pulsar candidates have been identified through manual signal processing, where humans visually identify the emission spectrum from the data collected by telescopes. Not only is this process time-consuming and mentally demanding, but it also introduces human error into pulsar identification. However, as the technical capabilities of pulsar searches continue to grow and change - for example, through increasing bandwidth, sky coverage, sensitivity, and, most notably, frequency resolution - the number of candidates has risen. This makes manual processing no longer feasible. In response, some data filtering tools have been created to filter out most of the noise and interference before signals reach human eyes. However, even these filtering methods are no longer practical, as they return more viable candidates than can be manually processed. The lack of an efficient and accurate pulsar classification method has created this 'candidate selection problem'.

**Machine Learning:**

Machine learning is a branch of computer science and artificial intelligence that uses data and algorithms to imitate how humans learn. It allows computers to learn without being explicitly programmed in a way that continuously improves their accuracy.

The learning system of a machine learning algorithm can be split into a decision process, an error function, and a model optimization process. In the decision process, the model processes training data to create the model and then uses the patterns and correlations it identified in the training process to make a decision, for example, a prediction or classification, about new test data. An error function refers to the process of assessing the accuracy of the model. Finally, the model optimization process describes fine-tuning the algorithm to fit the training data better to attain a higher accuracy in the testing phase.

Machine learning can be subdivided into supervised and unsupervised machine learning. Supervised machine learning is a subset of machine learning where the machine uses the training examples with labels or targets. These labels help the algorithm to correlate features and ultimately identify patterns in the dataset. Supervised learning is instrumental in solving problems that fall under the categories of classification or regression. In classification problems, the machine predicts the most probable category, class, or label for new examples, while in regression problems, the machine predicts the value of a continuous response variable. Methods used in supervised learning include neural networks, linear regression, logistic regression, random forest, and support vector machines, and other methods.

Unsupervised learning refers to a model training process where the training data do not have attached labels. Instead, the machine searches for patterns in the data. Unsupervised learning is useful in solving problems that require clustering, as the machine can group data points with similar features.

**Decision Trees:**

A decision tree classifier is a supervised machine learning algorithm that uses a set of rules to make decisions, much like humans. The original collection of training data is entered at the start, called the root node. Decision trees use dataset features to create a list of yes/no questions. Each question is marked with a node. The questions continuously split the dataset into smaller subsets, where all the data points that correlate to the answer ‘yes’ branch into one group, and the remaining data points branch to create a second group. In this fashion, the data becomes organized into a tree structure. This splitting of nodes continues until there are no more rules to apply or no data points left. These final nodes are called leaf nodes. At this point, each leaf node must be assigned a class. If the data in a leaf node is fully isolated by class (all of the data points in the training set are in groups that only contain the same label), it is referred to as a pure leaf node, where the class is assigned as the common label. However, leaf nodes are often not 100% pure and are thus called mixed nodes. In this case, the algorithm assigns the most common class among the data points in the node as the common label.

Ideally, a decision tree will have the smallest number of splits possible while maintaining the highest accuracy. However, this approach could be more computationally infeasible, mainly as larger datasets are used and the time to build the tree grows. The next best approach in creating the best tree is through the greedy approach, which attempts to make the locally optimal decision at the current node rather than the best decision for the tree overall. In making the best split, decision trees aim to divide the dataset into the smallest subset possible, so ultimately, the goal is to minimize the loss function. Loss functions are mathematical equations, known as criteria, that calculate the information gained when a node is split. Criteria are used to decide which features are most efficient to split on. The three main loss functions used in decision tree algorithms are Gini, Entropy, and Log Loss. All three functions are measurements of error.

Decision trees are useful in solving both classification and regression tasks. They allow for easy interpretability, as decision trees are simple, easily visualized, and understood. They are also useful for their data robustness since they can process numerical, categorical, and Boolean data. Decision trees also readily provide information about the data and the relative importance of various features. The most important features are used to split nodes higher up in the tree by the need to create the most efficient splits possible.

**Random Forest Classification:**

Random forests are a collection of independent decision trees that act as an ensemble. When a test data point is run through a random forest, each tree will make its own class prediction based on its unique model. The class with the most votes (the class the majority of the trees predict) becomes the entire forest’s prediction. The underlying reasoning is that a group of relatively uncorrelated trees acting as a group will outperform any individual tree.

The low correlation between individual trees is key to the accuracy of the random forest. It ensures that an error in one
tree will not be matched in other trees and that the random forest is collectively protected from individual errors.

Two methods random forests use to ensure this diversity in trees are bagging and feature randomness.

Bootstrap Aggregation, also known as bagging, is the process where each tree uses a unique dataset sample to build its tree through random sampling with replacement. Rather than splitting the entire dataset into smaller chunks for each tree, random sampling with replacement allows each tree to be trained on a dataset equal to the size of the entire dataset. However, instead of feeding each tree the entire original dataset, each tree chooses a random sample of data points (equal to the total number of data points in the dataset) with replacement. Since the trees sample with replacement, duplicate data points can be sampled and used to train a single tree.

The second method used to ensure a low correlation between trees is feature randomness. When splitting a node in a regular decision tree, all the features of the dataset are considered. The split is made using the feature that creates the greatest separation between the two resulting branches. However, in random forest trees, each tree can only determine the best split from a randomly selected subset of the dataset features. This allows for even more variation among the trees in the forest.

Ultimately, random forests serve as highly accurate and efficient models for solving classification and regression problems and can generally produce better results than any single decision tree.

**Machine Learning in Pulsar Identification:**

Various machine learning methods have recently been proposed to address the growing need for automation in pulsar classification. However, the majority of these algorithms are based on artificial neural networks (ANNs).

ANNs are computing systems inspired by the biological neural networks used for decision-making in the human brain. They are instrumental when solving problems related to pattern recognition and classification, approximation, optimization, and data clustering. ANNs use an extensive collection of units or nodes that work as artificial neurons, and each act as individual simple processors that operate in parallel. Neural networks contain an input layer, one or more hidden layers, and an output layer. Each node is connected to every other neuron in the layers above and below it using a connection link, and each has a weight threshold. If the output of any individual node is above this threshold value, that node is activated and sends data to the next layer of the network. If the threshold is not met, no data is passed along to the following network layer.

ANNs have been frequently proposed as a solution to automating the classification of pulsars. For example, Bates et al. (2012) developed a neural network that could detect 85% of real pulsar candidates over two years of data from the mid-latitude portion HTRU survey. Candidate parameters used included the pulse period in milliseconds, the pulse width, the DM in cm$^{-3}$ parsecs, the signal-to-noise ratio of the detection, and a unique $\chi^2$ value calculated from fitting the pulse profile with a sine function. However, their study did not utilize a representative sample of the pulsar population during the training process. In addition, they used a single artificial neural network to detect different types of pulsars, likely contributing to the lower accuracy.

Eatough et al. (2010) also implemented a neural network to re-analyze Parkes Multibeam Pulsar Survey (PMPS) data. They were able to discover a previously unknown pulsar. This neural network was trained using a particular set of scores to identify credible pulsar candidates automatically. The tool recovered 92% of pulsars present in a test sample of approximately 2.5 million candidates. Shortcomings included the poor training of the ANNs on MSPs, unbalanced training sets, and abnormal candidate plots generated by search software, which made it unlikely that the tool would identify MSPs.

Morello et al. (2014) developed a third neural network, the Straightforward Pulsar Identification using Neural Networks (SPINN), to process HTRU survey data. The algorithm identified every known pulsar in the southern survey data with a false positive rate of only 0.64%. It also identified four new pulsars in re-processing the intermediate galactic latitude area of HTRU, three of which have spin periods shorter than five milliseconds. Pulse features used in developing SPINN include the S/N of the folded profile, which is a measure of signal significance, the ratio between period and DM, the intrinsic equivalent duty cycle of the pulse profile, which is the ratio of a pulsar's pulse width in seconds to its spin period, a measure of the validity of DM, the persistence of the signal through time, and a measure of the variability of the pulse shape during the observation.

This study aims to explore the implementation of another machine learning tool, the Random Forest Classifier, as a solution to the pulsar ‘candidate selection problem. In addition to ANNs and other previously proposed tools, the Random Forest Classifier can provide an efficient and accurate solution to identifying pulsar signals from a large sample of survey data.

**Methods**

**Dataset:**

In developing our tool, we used a sample of pulsar candidates from the South High Time Resolution Universe Survey (HTRU2). Data for the southern hemisphere portion of the HTRU survey was collected using the Parkes Multibeam Receiver. The dataset used came from the UCI Machine Learning Repository. The dataset contains 17,898 data points including 1,639 real pulsar examples and 16,259 samples of RFI/noise. The disparity between the number of real pulsar examples and the interference signals speaks to the rareness of true pulsar signals. Features of the data we used included the mean, standard deviation, excess kurtosis, and skewness of the integrated pulse profile and DM-SNR curve.

The integrated pulse profile is a superposition of hundreds of thousands of individual pulsar pulses. The integrated profile is unique to each pulsar and can be recreated anytime.

The Dispersion-Measure-Signal-to-Noise-Ratio curve, or the DM-SNR curve, accounts for the dispersion of pulses. Radio pulses arrive at different times across different radio frequencies due to the ionized interstellar medium radio signals travel through before they reach Earth. This delay across frequencies is referred to as dispersion. Astronomers fit the shape...
of the delay when creating the pulse profile to compensate for its effect; however, there remains uncertainty with the fit. This uncertainty is expressed through a DM-SNR curve.

The mean, standard deviation, excess kurtosis, and skewness of both the integrated pulse profile and the DM-SNR curve were used as features of the radio signals in training the machine learning model.

The mean refers to the average. Mathematically speaking, it is the sum of a collection of values divided by the number of values in the collection. In reference to the integrated profile, the mean refers to the average pulse energy associated with the profile. The mean of the DM-SNR curve is the average of the curve.

The standard deviation measures how much individual data points vary from the mean. For the pulse profile and DM-SNR curve, it measures how much individual pulses differ from the mean.

The kurtosis and skewness both refer to the shapes of the curves. Excess kurtosis describes how tailed distribution is relative to a normal distribution of data. As in Figure 1, a curve with fewer outliers would appear thin-tailed and have a negative kurtosis (less than 3), while a curve with many outliers would look fat-tailed and have a positive kurtosis (greater than 3). A positive skew indicates that the curve is left-modal and skewed to the left, and a negative skew suggests that the curve is right-modal and skewed to the right, as seen in Figure 2.

**Python Libraries and Packages:**

In creating the random forest classifier, we employed several python packages. Pandas, an open-source data analysis and manipulation tool, was used to format and visualize the data from a CSV file. The Scikit-Learn (sklearn) python machine-learning library was used in creating the bulk of the program. The method train_test_split was used from sklearn. model_selection to split the dataset into training and testing data, with 70% of the data used for training and 30% used for testing. The proportion of pulsars to noise was maintained during the split into testing and training data, and the method RandomForestClassifier was used from sklearn.ensemble to build the random forest. The metrics package from sklearn was used to calculate the accuracy of the random forest classifier, and the tree package was used to visualize individual estimators (a single decision tree). Finally, the matplotlib,pyplot library was used to create graphs and plots to visualize the impacts of various hyperparameters on the accuracy of the classifier. The matplotlib and seaborn libraries were also used to visualize relative feature importance.

**Dataset Balancing:**

The dataset used to build the random forest classifier is extremely imbalanced with roughly 90% noise data and only 10% real pulsar data. This has implications for the accuracy of the model, as generally, machine learning models tend to ignore and perform worse on the minority class. Again, this has consequences for the model as it identifies pulsars, the minority class. While testing dataset balancing methods, all hyperparameters of the random forest classifier were kept at their default values to ensure consistent results.

Initially, the model was built on the raw, unbalanced dataset, which contained 1,639 real pulsar examples and 16,259 samples of noise, to analyze the effects of data imbalance on the model. The resulting training and testing confusion matrices are shown in Figure 3.

The first method of dataset balancing tested, referred to as using a random subset of the noise data, involved cutting out noise data to the number of pulsar data before splitting it into training and testing data and building the model. This was achieved using the .sample() method through Pandas, which selected a random sample of 1,639 data points of noise data, equivalent to the number of pulsar data points included in the dataset. The resulting training and testing confusion matrices are shown in Figure 4.

The final dataset balancing method was the Synthetic Minority Oversampling Technique (SMOTE). This method addressed imbalance by oversampling the minority class. This is done by synthesizing new data points from existing examples. The resulting training and testing confusion matrices are shown in Figure 5.

**Figure 1:** Depiction of excess kurtosis.

**Figure 2:** Depiction of skew.

**Figure 3:** Confusion matrices resulting from the imbalanced dataset.

**Figure 4:** Confusion matrices resulting from a balanced dataset that was achieved by randomly sampling a subset of noise data equivalent to the pulsar data.

**Figure 5:** Confusion matrices resulting from the balanced dataset.
Detailed analysis was performed to determine each balancing approach’s strengths and weaknesses and compare the balanced approaches to the imbalanced method. In doing so, the blind approach was used as a baseline for comparing balancing method performance. It was derived as such: using the imbalanced dataset, if all cases are predicted as noise, the resulting accuracy is 16,259 noise cases / 17,898 total cases = 0.908426 in the full dataset. As shown in Figure 6, using this blind approach, a method that misses all pulsars, 100% of actual pulsar cases will be recognized incorrectly as noise, and 90.84% of all test cases predicted as noises will be real noise. Thus, 0% of actual pulsar cases will be recognized correctly as pulsars, as the method will not predict any pulsars.

In analyzing the balanced methods, using a random subset of the noise data, on average, provided better results than the imbalanced model. Using SMOTE, on average, provided comparable results to the imbalanced model.

Balancing using SMOTE and a random subset of the noise data provides comparable results when classifying noise data; SMOTE is only marginally more accurate. Using SMOTE, 97.97% of actual noise cases were recognized correctly as noise, comparable to 100% by the blind approach and 99.26% using an imbalanced dataset. Furthermore, using SMOTE, 90.83% of all test cases predicted as noise were real noise, comparable to 90.84% by the blind approach and 90.54% by the imbalanced method. However, SMOTE needed to be more accurate in classifying pulsar data than balancing using a random subset of the noise data. 89.43% of actual pulsar cases were recognized correctly as pulsars, which is still significantly higher than 0% by the blind method and 83.53% by the imbalanced method but is less accurate than 91.87% using a random subset of the data. Using SMOTE, 81.63% of all test cases predicted as pulsars were real pulsars, which is significantly less accurate than 92.04% by the imbalanced approach and 97.00% using a random subset of the noise data.

Balancing using a random subset of the noise data is more accurate in classifying pulsar data. Using a random subset, 91.87% of actual pulsar cases were recognized correctly as pulsars, higher than 83.53% by the imbalanced model and significantly higher than the 0% by the blind approach. In addition, 97.00% of all test cases predicted as pulsars were real pulsars, over 15% higher than the 81.63% using SMOTE.

However, it needed to be more accurate in classifying noise data. 97.15% of actual noise cases were recognized correctly as noise, which is notably less than 99.26% by the imbalanced approach and 100% by the blind approach. Also, 90.54% of all test cases predicted as noise were real noise, equal to 90.54% using the imbalanced method and comparable to 90.83% using SMOTE and 90.84% by the blind approach.

Cumulative averages show that using a subset of the noise data is roughly 4.175% more accurate overall than using SMOTE, 2.798% more accurate than the imbalanced method, and 30.527% more accurate than the blind approach.

The method train_test_split was used from sklearn.model_selection to split the dataset into training and testing data, with 70% of the data used for training and 30% used for testing. Next, the dataset was divided after balancing methods were applied. In the imbalanced method, the proportion between noise and pulsars was maintained, and 70% of each of the noise and pulsar cases was used for training, and 30% of the noise and pulsar cases were used for testing. In the approaches where the dataset was balanced, the entire dataset was split into 70% training data and 30% testing data. As a result, training and testing data were half pulsar data and half noise data.

**Hyperparameter Evaluation Metrics:**

Hyperparameters are adjustable parameters set before algorithm training, making it possible to control the model training process. Hyperparameters were tuned after balancing the data using a subset of the noise data, which was the most accurate dataset balancing method overall.

In assessing the accuracy of the model as different hyperparameters were tuned, the evaluation metric Classification Accuracy was selected. Classification accuracy is defined as the ratio of the number of correct classifications, where the class predicted by the model is the same as the class given in the input dataset, to the total number of predictions made by the model. This ratio is bound between 0 and 1. This evaluation metric was selected since the task was binary in nature (pulsar vs. non-pulsar), and all the predictions and prediction errors were therefore equally important. The results from this metric are illustrated as a plot.

The model’s accuracy relative to various hyperparameters was also visualized through a Receiver Operating Characteristic Curve (ROC curve). A ROC curve is a valuable evaluation metric for supervised binary classification problems. In creating a ROC curve, the true positive rate (TPR) is plotted against the false positive rate (FPR). It allows the tradeoff between sensitivity, the TPR, and specificity, the FPR to be visualized. As in Figure 7, the true positive rate is the ratio of "true" events correctly identified by the algorithm to the total number of "true" events in the testing dataset. The false positive rate is the ratio of "false" events that were incorrectly classified as "true" events to the total number of "false" events. In the best-case scenario, the true positive rate should be 1, while the false positive rate should be 0. For example, the true positive rate in pulsar classification refers to the number of pulsars correctly identified as pulsars. In contrast, the false positive rate refers to the number of non-pulsars wrongly identified as pulsars. ROC curves are especially useful in comparing the performances of
different supervised learning algorithms, in this case various random forest classifiers with different hyperparameters, by selecting the algorithm with the greatest area under the curve.

\[
TPR = \frac{TP}{TP + FN} \quad \text{FPR} = \frac{FP}{FP + TN}
\]

Figure 7: True positive rate and false positive rate equations. TP represents true positive, FN represents false negative, FP represents false positive, and TN represents true negative.

The greater the area under the ROC curve (AUC), the more accurate the classification model. Using integral calculus, the AUC is found by calculating the area under the curve from (0,0) to (1,1). AUC is a value from 0.0 to 1.0. An AUC of 0.0 means the model has classified 100% of the test data incorrectly, while an AUC of 1.0 means that 100% of the data was classified appropriately.

In using both metrics, the aim was to understand the model’s accuracy in identifying pulsars and to ensure that the false positive identification rate remained low. If the false positive value was to grow, and increasing numbers of false signals were incorrectly identified as real pulsar signals, valuable and expensive telescope time would be allotted to search in unnecessary places, thus defeating the purpose of this tool in reducing the time to identify pulsars accurately.

Hyperparameters:

Various hyperparameters were tested to increase the accuracy of the classifier. min_samples_leaf, min_samples_split, max_leaf_nodes, criterion, and n_estimators were the parameters tested. Each of these hyperparameters was tested individually to visualize the impact of increasing or decreasing their values. The remaining hyperparameters were kept at their default values during each test, with n_estimators = 100 across all tests. Several correlations were identified.

5.31 min_samples_leaf:

The min_samples_leaf hyperparameter determines the minimum number of samples required to be at a leaf node, the final nodes that make up the base of each decision tree. Values tested for min_samples_leaf were 1, 2, 5, 10, 50, and 100. As seen in Figure 8, at low values, from 1 to 10, the accuracies appeared to fluctuate but remained high in accuracy. However, the accuracy dropped at higher values. For example, setting min_samples_leaf to 5 gave the highest accuracy of 0.95024 and the highest AUC of 0.9482 (equal in accuracy to setting min_samples_leaf to 2), indicating that while the model is highly accurate, it is also maintaining the necessary low false positive rate. The observed results are likely because lower values of min_samples_leaf allow the tree to have more flexibility, create more splits, and have more individual leaf classes at the base of the tree, thus making it more accurate when tested.

The min_samples_split hyperparameter determines the minimum number of samples needed to split an internal tree node. The values 2, 5, 10, 50, and 100 were tested but demonstrated no clear correlation. As visualized in Figure 9, the AUC remained uniformly high between 0.94614 and 0.94817. Therefore, this parameter was not used in the final set of hyperparameters.

max_leaf_nodes:

Max_leaf_nodes is the hyperparameter that sets the limit on the splitting of nodes to reduce the depth of the tree, which reduces overfitting. A greater value of max_leaf_nodes will create a deeper tree with more splits. Values tested were 10, 20, 50, 100, 150, 200, 500, 750, and 1000. There was initially a positive correlation when plotting max_leaf_nodes versus accuracy. In Figure 10, as the number of leaf nodes grew, the accuracy also increased, with the greatest increase in accuracy between 10 and 20 leaf nodes. However, the accuracy dropped suddenly after 150 leaf nodes. The AUC at all tested values remained high (> 0.94), with the greatest AUC of 0.94817 when max_leaf_nodes was set to 150. As the max_leaf_nodes value grows, the tree can create more splits and become more constrained to the training data, allowing it to make more accurate decisions on the test data. However, suppose the value becomes too large. In that case, the tree can begin overfitting the training data to such precision that it cannot accurately address variances in the test data and thus begins making incorrect classifications, leading to the observed drop in accuracy.

Criterion:

Criteria, as mentioned prior, are mathematical equations used to determine the best split at a node and are used to maximize the information gained with each division. We tested the three main criteria, ‘Gini,’ ‘Entropy,’ and ‘Log Loss,’ to see which yielded the highest accuracy. Gini is calculated by subtracting the sum of the squared probabilities of each class from one. Log Loss is a formula for calculating information gain, mainly used to train binary classifiers. Entropy is a second criterion formula for calculating information gain. Lower entropy correlates to increased model accuracy.

The remaining hyperparameters were kept at their default values during each test, with n_estimators = 100 across all tests. Several correlations were identified.

Figure 8: Accuracy plot and ROC varying min_samples_leaf hyperparameter.

Figure 9: Accuracy plot and ROC varying min_samples_split hyperparameter.

Figure 10: Accuracy plot and ROC varying max_leaf_nodes hyperparameter.
11 showed that ‘Gini’ was the most accurate criterion for our model.

\[ n\text{-estimators} \]

\[ n\text{-estimators} \] refers to the number of estimators, or independent decision trees, present in the random forest. There is a pre-existing understanding that as the number of estimators increases, the accuracy will also increase as errors in individual trees are more frequently canceled out. This occurs only at the expense of run time, as increasing the number of estimators will take the processing unit more time to build the random forest.\(^{28}\) Since this correlation is well known, \( n\text{-estimators} \) was the last hyperparameter tuned after the other hyperparameters and their values were already optimized. The accuracy of the random forest was tested with 1, 2, 5, 10, 20, 50, 100, 200, 300, 400, and 500 estimators. As in Figure 12, plotting the resulting accuracies demonstrated that the accuracy appears to plateau after 50 estimators. The AUC remains high and fairly constant after 50 estimators, only slightly ranging from 0.9461 to 0.9502.

Results and Discussion

We looked at trends in the hyperparameter versus accuracy plots and ROCs to determine the parameters used in our final model. The final values were set as follows: \( n\text{-estimators} = 50, \) \( \text{min} \_\text{samples} \_\text{leaf} = 5, \) \( \text{max} \_\text{leaf} \_\text{nodes} = 150, \) criterion = ‘gini.’

In assessing the model’s accuracy using the evaluation metric Classification Accuracy as defined prior, our model demonstrated an accuracy of 0.9502. There are slight variations in the accuracy as the forest is rebuilt, as no two forests are identical.

The confusion matrices for the final model are shown in Figure 13. Analysis of the confusion matrices reveals that this model performs accurately on both noise and pulsar data. Using this model, 97.56% of actual noise cases were recognized correctly as noise, comparable to 100% by the blind approach. Using this algorithm, 90.84% of all test cases predicted as noise were real noise, equivalent to the accuracy of the blind approach. 91.67% of actual pulsar cases were recognized correctly as pulsars, higher than 0% by the blind method. Furthermore, 97.41% of all test cases predicted as pulsars were real pulsars. Cumulative averages show that this algorithm provides 30.757% more accuracy than the blind approach.

Decision tree 0 of the resulting random forest is depicted in Figure 14.

Feature Importance:

As visualized in Figure 15, in ranking feature importance in differentiating between pulsars and non-pulsars on a scale from 0 to 1, the excess kurtosis of the integrated profile was ranked as the most important feature with a relative feature importance of 0.249734. Next was the skewness of the integrated profile at 0.193486, the mean of the integrated profile at 0.176090, the mean of the DM-SNR curve at 0.105585, the standard deviation of the DM-SNR curve at 0.099761, the skewness of the DM-SNR curve at 0.074150, and the excess kurtosis of the DM-SNR curve at 0.065950. Finally, the least important feature in determining whether a signal is from a true pulsar or a form of interference was the standard deviation of the integrated profile at 0.035243.

Figure 15: The ranked relative importance of radio pulse signal features.

Figure 16 provides an insight into how an individual tree creates splits to create the most significant separation between the pulsar and non-pulsar classes. The root node of this tree divided based on feature 4, the mean of the DM-SNR curve,
indicating that the mean of the DM-SNR curve was the most important feature out of the random subset of features used to build decision tree 0 in differentiating the pulsars from the non-pulsars in the random data points used to create that tree. Feature 5, the standard deviation of the DM-SNR curve, and Feature 3, the skewness of the integrated profile, were the following two most important features, as they were then used to split the second layer of the tree.

Figure 15: The ranked relative importance of radio pulse signal features.

The model is very robust and can produce results with similar accuracy even with fewer features. When removing the standard deviation of the integrated profile from the model, the accuracy remained very stable at 0.9512. When removing both the standard deviation of the integrated profile and the excess kurtosis of the DM-SNR curve from the model, the accuracy remained at 0.9471. With the three least important features, the standard deviation of the integrated profile, the excess kurtosis of the DM-SNR curve, and the skewness of the DM-SNR curve removed, the accuracy of the model stayed constant at 0.9502. This broadens the applications of this machine-learning model in pulsar surveys. The algorithm can be used with very high accuracies even in cases where not all eight features' values are recorded or known.

Conclusion

Our model demonstrates the effectiveness of random forest machine learning classifiers to solve the pulsar candidate selection problem. Our model was able to identify pulsars to an accuracy of 95% based on the mean, standard deviation, excess kurtosis, and skewness of the integrated pulse profile and DM-SNR curve of radio pulse signals from the South High Time Resolution Survey.

The optimal hyperparameters used to achieve this accuracy were n_estimators = 50, min_samples_leaf = 5, max_leaf_nodes = 150, and criterion = 'gini'.

Our results demonstrate that the excess kurtosis, skewness, and mean of the integrated profile were the three most important factors in differentiating between real pulsar signals and interference or other noise.

Limitations:

While this model has demonstrated high accuracy in classifying real pulsar signals from radio data, it is yet to be trained to identify millisecond pulsars (MSPs) specifically. Thus, this tool is likely not useful in searching for MSPs.

Additionally, since the real labeled pulsars in the training data came from the signals emitted by typical pulsars, this model would not be able to successfully identify or flag novel and unusual phenomena that might otherwise be detected through manual signal processing by human eyes. Thus, human processing should still be involved in the pulsar identification process at some level, and automated methods such as the one proposed here should only act as a first pass over the data.

Future Work:

This tool can be applied to a broad spectrum of future surveys, for example those conducted by the future Low-Frequency Array (LOFAR), the Five Hundred Meter Aperture Spherical Telescope (FAST), and the Square Kilometer Array (SKA), all of which are expected to produce vast amounts of signal data. It can be used as a first pass over the recorded data to significantly reduce the number of possible pulsar candidates that require manual observation. In addition, the high accuracy of this classifier decreases the likelihood of missing promising pulsar candidates during the first pass over the data.

To continue to improve this model, it should be trained on larger labeled datasets in addition to the one provided by the HTRU2 survey.

To increase the likelihood of identifying MSP candidates, a separate machine learning classifier should be used that is trained on a dataset where MSPs are explicitly labeled. That will ensure that the unique characteristics of MSPs will be clearly identified by the machine learning tool.

Acknowledgments

I want to thank Kristen Surraro for her guidance on this project.

References

Author
Ankhita Sathanur is a senior at Eastlake High School. She plans to major in computer science and physics and pursue a career in computational astrophysics. She hopes to continue researching pulsars and their unique role in astronomy, and she plans to continue to implement computing solutions in the field of astrophysics.
A Search for Monochromatic Light from the Andromeda Galaxy

Ariana Paul

Valley Stream North High School, 750 Herman Ave, Franklin Square, New York, 11010, USA; arianapaul12@gmail.com
Mentors: Dr. Woo Rin Lee, Dr. Woo Rin Lee

ABSTRACT: The Andromeda Galaxy was surveyed for monochromatic sources by using an objective prism optical telescope. The survey consisted of capturing 150 images of the entire galaxy using exposure times of only 10 to 20 sec, giving optical spectra of every point within the galaxy. The goal was to detect any sources, including transients lasting less than 10 s, that emit a narrow range of wavelengths (<3 nm) inconsistent with known astrophysical emission lines. Two different cameras were employed, having pixel sizes of 11 and 3.7 microns, the latter offering superior cosmic ray discrimination from monochromatic light. The 150 images revealed no monochromatic emission, neither short-lived (under 20 s) nor lasting the 30-minute duration of the sequence of exposures. Injection and recovery of synthetic monochromatic sources verified that lasers could be detected and yielded a detection threshold of 1 photon per second. At the Andromeda Galaxy, that detection threshold corresponds to a laser having a power of roughly 100 Terawatts for a benchmark 10-meter laser launcher that is diffraction-limited. Any lasers more powerful and directed toward Earth would have been detected. None were found.

KEYWORDS: Physics and Astronomy Extraterrestrial Intelligence; galaxies, Individual. Andromeda; techniques, spectroscopic.

Introduction

Many scientists, philosophers, and religious thinkers have noted that finding intelligent life elsewhere in the universe would be an important discovery, perhaps motivating humanity to abandon notions that we are special creatures and to reassess our Earthly purpose and cosmic future.¹ There can be little doubt that extraterrestrial intelligent life exists somewhere in the vast entirety of the expanding universe. It contains at least hundreds of billions of galaxies, each containing billions of star systems with Earth-size planets.² In the last few decades, astrophysicists have begun the challenging and systematic search for evidence of life elsewhere. While these searches have yielded no detections, there are still numerous.

The search for intelligent life (SETI) is pursued with various approaches, most commonly a search for non-natural electromagnetic waves at various wavelengths, which is a fast and cheap way to send signals.¹ ³ The signals may come from civilizations, space stations, interstellar or intergalactic vehicles, or even exotic powerhouses near binary neutron stars or black holes. The electromagnetic waves become fainter as the distance squared, greatly decreasing the intensity over hundreds, thousands, or millions of light years. The communication may be powered by nuclear energy, photovoltaics, or as-yet-undiscovered energy sources, offering gigawatts of power, and they may be sent by civilizations perhaps long extinct. In contrast, communication continues to travel at light speed.

Telescopes searching for extraterrestrial communication may collect optical, ultraviolet, infrared, or radio waves, taking advantage of the transparency of the Earth's atmosphere at many wavelengths. Most of these waves pass through the interstellar medium with only small extinction. Once a telescope detects such waves, further analysis, and new observations can be done to validate its existence and interpret any messages. Still to this day, despite hundreds of warm Earth-size planets discovered³, evidence of life outside the Earth, including any that developed technology, has yet to be found.

SETI researchers have pursued the search for over 60 years with no success, motivating Philip Morrison and many others to explain "The Great Silence." In the 1980s Morrison noted that life on Earth is the consequence of a long, diverse sequence of chemical natural selection allowing specific chemical factors and environments to become reproductive systems we call "biology."¹ Any life that does emerge must make itself detectable over interstellar distances so that we can detect it, a challenge they may not be motivated to overcome.¹

The leading SETI "Breakthrough Listen" program concentrates on detecting radio waves of technological origin⁶, including from nearby stars and galaxies. Alternatively, one may search galaxies for laser beacons including from the nearby Andromeda Galaxy. Laser searches toward other galaxies may search galaxies for laser beacons including from the near by Andromeda Galaxy. Laser searches toward other galaxies would be sensitive only to powerful lasers used by civilizations that are interested in alerting us to their existence and sending high volumes of data.⁷ Laser beams are particularly useful for intergalactic communication because their small divergence concentrates the message toward the intended recipient.⁸ As optical lasers are rare in nature, if they exist at all, the detection of such beams would be extraordinary and worthy of follow-up study. ginger extract might support skin wound healing with strong antibacterial effects.
Searching for laser light from other galaxies represents an extreme domain of the plausible distributions of technological luminosity. Advanced civilizations are likely separated by millions of light years on average. The rare ones that flourish may be long-lived, allowing them to acquire the ability to send signals of high luminosity from their large energy availability. Accordingly, we describe here a search for laser beacons from the Andromeda Galaxy. Its structure, modest dust extinction, and tilted disk allow the entire disk and outer bulge to be observed for laser beacons originating within those regions. Images of Andromeda have been obtained through different filters, from UV to IR, but a search for monochromatic sources has not been done.

**Observational Methods:**

We searched for laser light from Andromeda by using an objective prism telescope having an entrance diameter of 0.28 meters and a prism over the aperture having a 7-degree wedge angle, as shown in Figure 1. We were fortunate to use the telescope built and maintained by the team at the Center for Space Laser Awareness. We accessed their telescope remotely using Zoom to command operations. A full description of the telescope, its prism, camera, and the resulting wavelength dispersion are already described. In brief, the 0.28-meter telescope (Celestron RASA-11) gathers light at each location in the sky within a 2 deg x 3 deg field of view. The prism over the front spreads the light into all optical wavelengths. Laser emission would be composed of a narrow range of wavelengths of optical light, either pulses or continuously on, that would appear as "dots" in the image. At the distance of Andromeda, 752 kiloparsecs, the plate scale (with the QHY600M) is 1.3 arcsec per pixel and a pixel spans a distance of 380 to 950 nm, with longer wavelengths to the left. These are spectra of stars in the foreground Milky Way Galaxy.

During the first observing run in Feb 2021, the data acquisition involved 100 images, each 10 sec in duration. During the second observing run in 2022, the data acquisition involved 50 images, each 20 sec in duration. The entire Andromeda Galaxy and its two dwarf spheroidal satellite galaxies are captured in each image. The spectra of foreground stars located in our Milky Way Galaxy are also captured in each image, with the faintest being visual magnitude 14. A typical 20-sec exposure of the Andromeda Galaxy is shown in Figure 2. The central nucleus and bulge regions of the galaxy are white due to the grayscale employed here, and the disk of the Andromeda Galaxy is the diagonal faint smear of light extending from the lower left to the upper right. The two dwarf spheroidal galaxy companions to Andromeda are also visible as spectra slightly puffed in width. Also visible are hundreds of stellar spectra, the horizontal streaks from left to right due to the prism dispersing the wavelengths, 380 to 950 nm, with longer wavelengths to the left. These are spectra of stars in the foreground Milky Way Galaxy.

This objective prism telescope system is sensitive to sources of a narrow range of wavelengths of optical light, either pulses or continuously on, that would appear as "dots" in the image. At the distance of Andromeda, 752 kiloparsecs, the plate scale (with the QHY600M) is 1.3 arcsec per pixel and a pixel spans a distance of 4.8 parsecs. Figure 3 shows the sum of the 50 images obtained with the KL400 camera, totaling an exposure time of 1000 sec. The three panels in Figure 3 display this summed image of Andromeda using three different grayscales. The different grayscales enable the identification of laser emission (dots) in regions that are faint outside the Andromeda Galaxy, medium-bright within the galactic disk, and very bright in the central region of Andromeda. These three grayscales permit the detection of laser point sources in the co-added image of all three domains of brightness. The co-added image reveals spectra of stars as faint as V magnitude 17.

**Analysis**

We inspected, by eye, each of the 150 images of the Andromeda Galaxy in search of laser light that would appear as a "dot" on the image having the same size as the point spread function. Visual analysis was preferred over an algorithm due to the mere 150 images, a small enough sample that visual...
inspection is possible, and offers the value of human pattern recognition and identification of unexpected noise in the images. With the KL400 camera and its 11-micron pixels, the full width at half maximum of the image shape of an unresolved laser source, i.e., the point spread function, is 2.5 pixels in the spatial direction and perhaps slightly wider in the wavelength direction. To examine the images, we found that the grayscale made a significant difference in the ability to filter out the stars and background galaxy light that constitute background noise. We adjusted the grayscale to provide a modest brightness of the background light from stars and the galaxy, allowing any laser light to be visible above that background.

Co-adding the images taken in Feb 2021 with the KL400 camera required technical help from the Space Laser Awareness team. We had positioned the telescope at the correct coordinates of the Andromeda Galaxy, RA= 0h 42m 44s and Dec = +41° 16 09 (equinox 2000), centering the fully visible galactic disk of Andromeda in the 2 x 2 deg field of view. But during the acquisition of 100 exposures, no guiding was done, allowing the telescope to drift by 30 arcseconds. The co-addition was done by the Space Laser Awareness, using a simple cross-correlation to reposition the images to be coincident before adding the 100 images together. The final grand image of those 100 exposures is shown in Figure 2. The images, including the sum, using all three grayscales were examined.

A visual inspection of each of the 100 images of Andromeda obtained with the KL400 camera and of the co-added image was conducted to search for "dots" that could be single-wavelength laser emission. The coadded image contained much less noise, as the fluctuations of photons arriving were averaged out. The photos were identified by the specific exposure number. We expect a laser "dot" to have a width of roughly 2.5 pixels due to the point spread function and perhaps occupy 3x3 pixels due to the resolution of the telescope. Among the 100 images obtained with the KL400 camera, neither the individual images nor the co-added image showed evidence of a "dot," i.e., no candidate laser light was found.

![Figure 2](ijhighschoolresearch.org)  
**Figure 2:** The three different grayscales are utilized to suppress the background of stars, allowing any dots of single-wavelength light to stand out. We examined the images displayed with different gray scales to search for any dots of monochromatic light.

For the 50 exposures of Andromeda taken with the QHY600M camera in Jan 2022, shown in Figure 3, the full width at half maximum of an unresolved laser should be 5.5 pixels wide, i.e., that of the point spread function of the telescope with the smaller pixels. We searched for any region in the Andromeda Galaxy that contained a dot standing above the background light of the galaxy and having that 5.5-pixel width. This visual approach allowed me to distinguish unresolved, monochromatic emissions from the dominant light sources that are primarily stars in the foreground Milky Way and smeared starlight across the Andromeda galaxy itself. Again, any laser light would appear as a "dot" roughly 5.5 pixels across containing wavelengths confined to within the spectral resolution of 3 nm in wavelength. The analysis involved examining both the 50 individual images and also the sum of those 50 images.

The final summed image of the Andromeda galaxy, observed with the QHY600M camera, was separated into 54 sub-images, allowing a magnified view of the sub-images to search for laser "dots." Figure 4 shows the sum of those 50 images, and the dashed lines show the 54 sub-images. A search concentrated on finding "dots" having a shape similar to the known point spread function of the optical system. We verified the point spread function by noting the widths in the spatial direction of the stellar spectra. The widths remained between 5 and 6 pixels; thus, the full width at half maximum of the point spread function was 5 to 6 pixels, corresponding to an angle of 6.5 to 7.8 arcsec on the sky. Careful examination of the 54 sub-images revealed no evidence of monochromatic emission.

To test the experimental search technique and to determine detection thresholds, the Space Laser Awareness team generated synthetic single-wavelength emission "dots" and added them to the co-added image shown in Figure 4. This constituted a blind test, as I was not told where the synthetic...
laser "dots" were placed nor how many. Figure 4 contains three such injected synthetic laser sources. The sub-image containing the synthetic laser "dot" is shown in Figure 5. Laser emission "dots" representing 10 photons/sec and 1 photon/sec are easily visible and detectable in Figure 5. We would have easily detected such laser emission. This represents the detection threshold.

![Figure 4](image1.png)

**Figure 4:** A coadded image of the Andromeda galaxy. 1000 sec total, with the QHY600M camera, displayed with a grayscale that reveals the medium bright regions in the bulge and inner disk of Andromeda. The hundreds of horizontal streaks are optical spectra of foreground stars in the Milky Way. Three synthetic sources of laser emission were placed along row 3750 at columns 3200, 3300, and 3400 having photon rates of 10 ph/s, 1 ph/s, and 0.1 ph/s. The sources have the shape of the point spread function with full width at half a maximum of 5.5 pixels. The first synthetic laser "dot" is barely visible here. The dashed grid shows the sub-image regions that were examined separately and in detail by eye (see Figure 5).

The synthetic point sources of single-wavelength emission injected into the images shown in Figures 2, 3, and 4 are two-dimensional Gaussian profiles with a full width at half the maximum of 5.5 pixels, i.e., the point spread function expected for lasers. The detection threshold of 1 photon/sec in a single-wavelength dot represents the limit of detection of laser emission in this project. Laser sources in the Andromeda galaxy that deliver more than 1 photon/sec to this 0.27-meter telescope would be detected.

![Figure 5](image2.png)

**Figure 5:** One sub-image within the coadded image of Andromeda galaxy (shown in Figure 4). This sub-image shows the region between columns 3000 and 4000 and rows 3000 to 4000 containing the galactic disk between 3 and 7 kiloparsecs from the nucleus of Andromeda. Three synthetic point sources of monochromatic emission are placed in the row in column 3200, 3300, and 3400 having injected 10 photons/sec, 1 photon/sec, and 0.1 photons/sec respectively. The first two synthetic laser emission "dots" are visible. The detection threshold is apparently 1 photon/sec. All 54 sub-images of the co-added image of Andromeda (1000 sec total) were searched for such monochromatic emission. None was found.

To summarize the analysis, the goal was to identify close clusters of pixels, and dots, that are approximately 3 pixels across for the KL400 camera and 5.5 pixels across for the QHY600M camera. Any laser “dot” must have some photons above the background consistent with the point spread function of the telescope. No such evidence of laser light was found, neither in individual images nor in the co-added images, with a detection threshold of 1 photon/sec.

### Results and Discussion

We obtained and examined 100 images of 10 sec each and 50 images of 20 sec each of the Andromeda Galaxy. No evidence of monochromatic emission was found, with a detection threshold of 1 photon per sec. No evidence of optical lasers was found. Laser emission would have appeared as a patch of pixels the size of the point spread function, having a peak of 1 photon per sec. None were found. By simple visual inspection and analysis of the objective prism images of the Andromeda Galaxy, we would have detected optical laser beams. For a benchmark laser transmitter of 10-meter size, similar in size to the largest telescope on Earth, a laser power of at least a terawatt would be required (see below) to enable minimal detection.

The purpose was to detect laser beams sent by technological life or their machines. The results showed that in two 30-minute time intervals, one in Feb 2021 and one in Jan 2022, the telescope did not pick up a laser beam. It remains possible that laser light does come from the Andromeda Galaxy that is either fainter than our detection threshold or was beaming while we were not observing.

With the KL400 camera, some dots did appear, but they were smaller than 2 pixels across. Those dots consisting of roughly 2x2 pixels were almost certainly caused by cosmic rays that hit the sensor of the camera. Cosmic rays are muons, protons, neutrons, electrons, or gamma rays that came from astrophysical sources, such as the Sun or supernova explosions. By bad luck, some of them hit our camera while we were observing that night. Their energy liberates electrons in the pixel, as if a photon hit it, appearing as a tiny dot smaller than the point spread function of a point of light seen through the telescope.

The intensity of any laser beam from the Andromeda galaxy is greatly weakened by its great distance, 2.5 million light years.
Lasers with a power of a megawatt are detectable from stars located 100 light years away.¹¹,¹² Thus, lasers in Andromeda must have a power of at least 100 terawatts (10¹⁴ W) to be detectable, as intensity decreases as the square of the distance. This power threshold is comparable to that of the future radio telescope searches for radio signals.⁴ This required power is also due to the relatively small 0.28-meter RASA telescope we used. Another limitation is that we captured only 150 exposures, taken within a total of 1 hour. This is a good sample for a first attempt.

The extinction of light due to dust has been measured carefully between the Earth and different regions of Andromeda.¹³ From half of its disk, more than 50% of the light arrives at Earth. From the other half, light extinction varies between 50% and 95% and is most severe for stars (or lasers) buried within dense molecular clouds. Thus, in the regions of dense dust, the laser light would have to be 2x to 20x stronger to permit detection.

A next attempt would involve taking more exposures, at least 1000, spanning a few hours. Also, it would be better to search for laser beams at a wide variety of times spread over a month or a year. We don’t know how often lasers in the Andromeda galaxy might be shining in our direction. If intelligent life occasionally cast a laser beam toward earth, we would have no way to know the exact time, or cadence, of their beaming. If any laser beam were detected, the next step would involve follow-up observations by other telescopes and by our telescope to verify the extraordinary discovery. Surely 150 exposures are not enough. Therefore, a larger sample size at different times is warranted. Another limitation is the position of the earth in its orbit around the Sun. The exposures were captured on one night in 2021 and one night in 2022. The Andromeda Galaxy is visible through the months June-February setting a time window when we could capture more exposures.

### Conclusion

No evidence of laser emission was found toward the Andromeda Galaxy in 150 exposures totaling 2000 s. The individual exposures of 10 or 20 s duration showed no laser emission. The sum of exposures totaling 2000 s also showed no evidence of laser emission, with the detection threshold of 1 photon per sec at the peak of the laser dot on the detector. Adopting a benchmark 10-meter aperture of a hypothetical laser launcher, the laser must have a power of roughly 100 terawatts (10¹⁴ W) to be detectable. More observations of Andromeda are warranted to detect non-astrophysical narrowband emission, using both higher spatial resolution and other wavelengths, notably x-ray, UV, IR, and radio. Other galaxies can be observed for narrowband emission using this technique, as the high end of the luminosity function of extraterrestrial technology remains unknown.

### Acknowledgments

This project was done as a high school “science research program” during the junior and senior years of A.P. I thank A.P.’s teacher, Michael Frazier, Science Research Coordinator, for his inspiration and guidance. We thank my mentor, Dr. Geoffrey Marcy, for his thoughtful and kind support. We also thank Valley Stream North High School on Long Island in New York for this wonderful opportunity to experience scientific research. We thank the team at the Center for Space Laser Awareness for use of the telescope and for outstanding technical help. We also thank the teams at Finger Lakes Instrumentation and QHYCCD who made the cameras we used, without which this research would not be possible.

### References


### Author

Ariana Samiah Paul is a student at North Carolina State University, Raleigh, NC, where she is pursuing a degree in Civil Engineering. She plans to continue to engage in research during her undergraduate career and beyond.
Oncolytic Virotherapy and The Immune System

Asya Yılmaz
Tarsus American College, Murathan Dist. 1202/8 St. No. 8 Evka-1, Buca, İzmir, 35400, Turkey; asayilmaz424@gmail.com
Mentor: Mr. Eric Trujillo

ABSTRACT: In order to awaken the innate and adaptive immune systems in response to malignant tumor growth, activation of effector cells like T- and B-cells through pro-inflammatory cytokine release prompted by apoptosis is vital in oncolytic virotherapy. As cancerous cells commonly exhibit abnormal behavior that escapes the radar of the immune system, the discovery of oncolytic virotherapy was essential to establishing an effective anti-tumor response. With certain combinations of immunotherapies, digging at the root of cancer, which is the uncontrolled division of abnormal cells, becomes possible, provoking a forcefully enacted treatment inside the host that uses both the immune system and external therapeutic agents in the fight against cancer. Such treatments as the combination of immune checkpoint inhibitors or macrophage-mediated responses and oncolytic virotherapy enable a higher chance of improvement in the prognosis.

KEYWORDS: Biomedical and Health Sciences; Immunotherapy; Oncolytic Virus; Inflammatory Response; Macrophage-mediated Response.

Introduction
From the beginning of the 20th century, oncolytic virotherapy attracted attention as naturally acquired virus infections seemed to cause the shrinkage of the tumors, possibly leading to complete disappearance. Chance encounters of tumors with viral infections starting from 1904 led researchers to conduct more comprehensive research on the possible usages of virotherapy on cancer.¹ In fact, the first commercialized oncolytic virus (OV), Oncorine (H101), was approved in 2005 by the Chinese State Food and Drug Administration (SFDA) for head and neck cancer.² Later, in 2015, the U.S. Food and Drug Administration (FDA) approved the T- VEC virus for melanoma.³

As OVs presented dual promise in cancer therapy, they were received with much interest. Apart from their oncolysis ability, a process in which OVs force the cancerous cells to burst due to continuous virus replication, the possible usage of combination therapies alongside oncolytic virotherapy made them a worthwhile investment in the fight against cancer. As the main target with oncolytic virotherapy is the positive response from the immune system, multiple ways to achieve this result have been orchestrated, mainly enhanced by their tendency to work well with other immunotherapies and their high immunogenicity. Therefore, the need for a brief but inclusive work to attract attention to this promising field was realized. All in all, this research treatise considers the body’s immune surveillance system and the possible methods to trigger the adaptive and innate immune systems into recognizing and eliminating tumors through using OVs alongside other complementary therapies.

Discussion
As early as 1909, Paul Ehrlich had laid the understanding of the importance of immune surveillance against tumors. He hypothesized that the host's defense prevented cells from evolving into tumors, but his hypothesis could not be proven due to the lack of equipment and understanding during his time.⁴ Experiments done with transplantation models later proved the theory of immune surveillance, as tumors were rejected in syngeneic hosts as opposed to normal tissues, hinting at the presence of tumor-specific antigens.⁴

OVs are indeed favorable in cancer treatment due to their various possible usages, from direct oncolysis to their effects in combination with other immunotherapies to elicit an immune response from the host. When cell lysis is initiated due to continuous virus replication inside the tumors, tumor-associated antigens (TAAs) are released, thereby triggering the immune system. Simultaneously, phagocytosis clearance of the apoptotic cells leads to triggering anti-inflammatory signaling pathways by apoptotic cell surface molecules via phagocyte receptors, also boosting the host autoimmunity.⁵

The combination of OVs with the dynamic characteristics of the host immune system remains a critical but, to some extent, unpredictable area. Findings on the issue of immune system effects on the injected OVs and the loss or gain of efficacy depend on various variables, from the genetic modification of the OVs to signaling pathway activation of the host. Interestingly, the advantage of OVs is apparently not reduced with the triggered anti-viral immune response as opposed to the results of prior research; instead, as long as there exists a balance between the triggered anti-viral response and the operation of OVs, for example, by deleting viral genes, activation of stronger anti-tumor response is possible.⁶

Targeting Tumor Vasculature with Oncolytic Viruses:
Tumor neovasculature is vital for the continuous feeding of the tumors; hence, targeting tumor vasculature by OVs in hopes of starving the tumor has gained attention in recent years. On the one hand, the collapse of the tumor vasculature was reported to lead to tumor hypoxia, nutrient restriction,
and pro-inflammatory cytokine release, showing promise in the field. On the other hand, the said vasculature is essential in transporting cancer therapeutics and various effector cells, leading researchers to aim for the normalization of tumor vasculature to maximize the anti-tumor activity of leukocytes and blood-borne therapeutic drugs. Although studies to finalize the debate of shutting down versus normalizing tumor vasculature is yet to produce a definitive answer, certain studies using OVs to disrupt the established tumor vasculature bear the potential of achieving a favorable prognosis. One such study by Breitbach et al. involved intravenous administration of the vesicular stomatitis virus (VSV), causing the tumor vasculature to be destroyed while the normal vasculature remained intact. Continuous VSV replication was followed by an inflammatory response, including neutrophil-dependent forming of micro-clots, enabling Breitbach and colleagues to observe blood clots in tumor vasculature in the following 24 hours upon intravenous administration. As such, it was proposed that OVs not only caused the lysis of the cell but also directly infected the tumor vasculature, demonstrating the factors VSV anti-tumor activity depended on, which were infection of tumor vasculature and intravascular coagulation, all the while underlining the importance of the route of delivery through which the VSV first came in contact with the endothelial cells, explaining its propensity to cause vascular collapse.

For further demonstration, Figure 1 (B) represents the formation of clots and initiation of vascular collapse after infection of the endothelium of the tumor vasculature by OVs and recruitment of neutrophils. In contrast, (A) represents oncolysis, and (C) represents the role of VEGF in the regulation of angiogenesis and tumor growth, yielding oncolytic effects.

Macroage-mediated Responses to Oncolytic Virotherapy:

Tumor-associated macrophages (TAMs) are an important part of the tumor microenvironment (TME). Depending on their activation, they may either have positive or negative effects on tumorigenesis. Due to their dual nature, they are promising partners with oncolytic virotherapy to act as an intermediary between the TME and cytotoxic lymphocytes of the innate immune system, such as natural killer (NK) cells; or T lymphocytes and B lymphocytes of the adaptive immune system, to eliminate tumors.

When macrophages elicit enhancing antitumor responses, adaptive and innate immune system cells will be alerted, ultimately causing the elimination of cancerous cells. On the other hand, since TAMs may cause the suppression of the immune system in TME, it is important to combine both therapies in a balance to restrain the immunosuppressive, tumor-supportive properties of macrophages. Macrophages are highly plastic cells, and depending on the environment, they can undergo drastic changes that can either affect tumorigenesis negatively or positively. Such duality has been separated as classically activated macrophages that elicit a positive response, also known as the M1 subtype, and alternatively activated macrophages that elicit a negative response, known as the M2 subtype. However, it has recently come to attention that such distinct separation was only sometimes possible due to the multiple activation states of the macrophages. Since OVs work as strong immunological stimuli,
they can be used to convert the phenotype functions of macrophages accordingly.¹⁶

The most abundant molecule in the TME is macrophages; therefore, they are an essential target in cancer therapy, mainly used in two ways. The first method is to reduce the TAM populace in the TME either through its depletion or the restriction of monocyte recruitment that gives rise to TAMs. The second is by re-modifying macrophages through converting M2-like TAMs into M1 phenotypes or giving rise to macrophage-mediated anticancer responses such as phagocytosis.¹⁴ As explained above, specific functions of macrophages can be altered by OVs as they are highly plastic, making OVs the ideal partner to therapies including macrophages, though their connection is still understudied.

**Combined Therapy of Immune Checkpoint Inhibitors and Oncolytic Viruses:**

Immuno-oncology (IO) is a specific cancer treatment area that targets the host immune system to fight against cancer. In particular, the oppression of the immunosuppression ability of cancer by antibodies (Ab) that target immune checkpoint molecules caused positive results in the field.¹⁷ Therefore, the first immune checkpoint inhibitor (ICI), ipilimumab, was approved by US FDA for the treatment of advanced melanoma in 2011.¹⁷

Since apoptosis caused by OVs is followed by the release of immune-stimulatory molecules such as PAMPS, TAAs, and DAMPS, the immunosuppression ability of tumors is canceled.¹⁷ In fact, there is strong evidence suggesting that the unresponsive, cold tumors may be converted into responsive hot tumors by OVs, thus showcasing their compatibility with IO drugs. Certain studies on their combination yielded the result of elevated pro-inflammatory cytokine release, prompting the activation of NK cells and T-cells inside of the tumors, changing the state of the TME into that of a hot tumor.¹⁷

However, in clinical trials, OVs were seen to lose the advantageous tumor specificity they possessed when administrated over a period of time, caused by the loss of essential signaling pathways due to mutation, especially in heterogeneous tumor populations.¹⁷ Therefore, first-generation OVs like Rigvir and ONXY-015 as monotherapies yielded ineffective results, further highlighting the importance of combination therapies.¹⁷ As IO drugs also stimulate the host immune system, their combination with OVs produces elevated results, possibly marking one of the most efficient combination therapies.

**Risks of Oncolytic Virotherapy:**

Although oncolytic virotherapy bears much potential, it is vital to be mindful of their adverse effects and to study the possible ways to decrease the incompatibility between the injected OV and the host, recognizing the importance of continuous experiments on the viruses. Below is Table 1, showing various types of OVs that are being clinically tested.

In OV usage, two main factors are important to consider. Firstly, potent anticancer immunotherapy may result in autoimmune diseases and very well implicate healthy cells, so their dosage and compatibility with the type of cancer should be arduously studied before they are approved. Secondly, they are highly immunogenic, meaning they are likely to elicit an immune response from the host.¹⁸ While the latter can be balanced and turned into a favorable end result through the combination of various immunotherapies, the former yields complications due to the difficulty of researching virus-cancer compatibility and the unpredictable patterns of adverse reactions the host may produce. Therefore, creating non-pathogenic viruses through modification of the viral genome in oncolytic virotherapy has aroused much interest due to their capacity to offer a safer prognosis by aiming to weaken virus pathogenicity, reducing side effects, increasing the target selectivity, and inserting exogenous therapeutic genes into the virus genome to increase its expression in the tumor, consequently improving treatment through modified OVs.¹⁹,²⁰

**Delivery Routes of OVs and Possible Adverse Effects:**

Two main routes of OV injection are direct intratumoral delivery and intravenous delivery, with either of them posing different advantages and disadvantages. Intratumoral delivery is the commonly used method of OV delivery, encompassing the direct injection of OVs into tumor masses. While the dose and spread of the OVs can be better controlled when delivered intratumorally, this delivery method is unsuitable for deep and organ-specific tumors. On the other hand, intravenous delivery encompasses the injection of the OV directly into the vein, enabling the virus to spread more efficiently, especially for organ-specific tumors. During a phase I clinical trial, it was also seen that intravenous delivery of oncolytic virus HSV G207 to children with progressive or recurrent malignant supratentorial brain tumors yielded positive results, confirming the ability of the OVs to breach the blood-brain barrier to reach brain tumor tissues.¹⁹,³⁴

However, despite the positive results observed in various trials, it is worth noting that intravenous delivery requires highly selective target tissues since increasing the concentration of the virus will not lead to better selectivity or wider spreading of the virus but instead will lead to toxicity, resulting in questionable biosafety of OV therapy.¹⁹ As some treatments require a
specific route of delivery due to the tumor environment, it is important to research OVs \textit{in vivo} and \textit{in vitro} on a deeper level to observe the cancer-specific behaviors of each treatment type.

**Decreasing the Risk of Oncolytic Viruses:**

According to Li \textit{et al.}, there are mainly three ways of improving OV safety: selecting non-pathogenic viruses, modifying the pathogenicity rate of the viruses, and recombination of various viruses.\textsuperscript{19} For the issue of selecting viruses that are non-pathogenic, the parvovirus is an example that is harmless to humans due to their natural hosts being rats. Since this kind of approach depends on the overexpression of cytokines from sites of tumor that can activate metabolic pathways, it has low selectivity when it comes to non-malignant tumors. The second approach involving pathogenicity modification is widely used in clinical trials and consists of the deletion, insertion, and binding of gene sequences of the viruses to decrease pathogenicity. Prior studies include the deletion of genes such as TK, VGF, hemagglutinin, and B18R in the oncolytic poxvirus, the deletion of the ICP-34.5 gene in HSV1716, and more, offering promising results.\textsuperscript{19,35}

The third method of conduct involves the recombination of types of OVs. As engineered by Abdullahi \textit{et al.}, the recombination of vesicular stomatitis virus (VSV) and Newcastle disease virus (NDV) into recombinant VSV-NDV (rVSV-NDV) virus by retaining the VSV backbone while replacing its glycoprotein with hemagglutinin-neuraminidase (HN) and the modified envelope proteins of the NVD eliminated the adverse reactions previously observed in the liver and the brain, prevented by the replacement of the glycoprotein.\textsuperscript{36}

Following continuous advancement in oncology and the discovery of other types of OVs will likely decrease the risk of virus therapy for cancer. Although the possibility of minor post-treatment adverse effects is not entirely gone, there is a high chance of improving the biosafety of oncolytic virotherapy with long-term observations of the cases and further research.

**Conclusion**

Oncolytic Virotherapy still has unlimited potential that is yet to be studied. Due to tumors’ immunosuppressive nature, OVs play a major role in cancer therapy and combinational therapy. Upon apoptosis, pro-inflammatory cytokines that activate T-cells and NK cells are activated, alerting the innate and adaptive immune systems. Considering their immunosuppression-evoking properties, OVs are ideal to go hand in hand with other therapies that struggle due to tumor expression. While it is important to consider the dichotomy of certain molecules that carry threats to the prognosis when incorporated into cancer therapy, certain clinical trials have offered worthwhile results in recent years, calling attention to the future of certain combined therapies. Apart from directly triggering the immune response upon injection, OVs showed positive outcomes when combined with other immunotherapies such as macrophage-targeted therapy and immune checkpoint inhibitors, highlighting that despite the direct effect of these viruses on alerting the immune system, they worked fine as mediators in the host bodies. Keeping the balance by selective therapies is important not only to eliminate tumors completely but also to avoid feeding them instead, notably calling attention to the different results yielded from M1 and M2 subtype TAMs due to their high plasticity or the differing outcome of acute and chronic inflammation. Therefore, minimizing the possibility of damage is a vital step through continuous in vitro and in vivo experiments by utilizing complementary therapies in a balance that will suppress tumor hallmarks such as immunosuppression, loss of apoptosis, and evasion of growth suppressors. Furthermore, seeing that all directed and specific treatments against the aforementioned properties of cancerous cells depended on awakening the hosts’ immune system, OVs are a strong candidate for better cancer therapy due to their highly immunogenic features. Undeniably, a better understanding of the properties of oncolytic viruses and their complementary therapies will lead to promising cancer therapies and even cancer prevention in the future.

**Acknowledgments**

The completion of this paper would not have been possible without the help of our biology teacher, Ms. Altan, our headmaster Mr. Trujillo, and our advisor Mr. Montalvo.

Special thanks to my lovely mother, who supported me greatly during my studies.

**References**


**Author**

Asya Yılmaz is a 10th-grade student studying at Tarsus American College. Her greatest aspiration is to become a Pharmacology Research Scientist after attending a Pharmacy College. She currently resides in İzmir, although she stays in Mersin for most of the year due to her high school studies.
Gendered Online Toxicity During COVID-19

Christine Lee
Riverdale Country School, 5250 Fieldston Rd, The Bronx, NY, 10471, USA; christine.lee.nyc8@icloud.com
Mentors: Killian McLoughlin, Rudy Nunez

ABSTRACT: The adverse effects of online sexism have become increasingly challenging to address due to their silencing effect, which is both normalized and subtle. Since behaviors specific to the COVID-19 Pandemic (e.g., wearing face masks) should not differ along gendered lines, any disparity in the genders’ outrage would suggest that gendered hostility and punishment are intrinsic to contemporary social life and are not confined to gender-coded behaviors with roots in our past. In other words, hostility is an expression of gender bias rather than a reaction to different behavior by gender. This research examines the differences in the volume and content of toxicity directed towards men and women using a sample of tweets posted to Twitter in English from December 2019 to December 2021 (N = 5,000,000). The LDA topic model results suggested that men received more tweets containing outrage than women. Still, women were targeted with more gendered hostility, including but not limited to gendered slurs and phrases that incite sexual violence. Social media platforms must take preventative measures to limit the spread of toxicity.

KEYWORDS: Behavioral and Social Sciences; Cognitive Psychology; Gender; COVID-19; Online Outrage.

Introduction

Social media use has been growing across the globe for more than a decade. Sharing opinions online has two main possible outcomes. The first outcome is the democratizing the public sphere, framing social media as a free, accessible forum that allows for public discourse open to diverse opinions without marginalizing any perspective.¹ The second outcome is the fragmentation of society — small, vocal, public section become the only voices that are heard online.² The latter may occur as the result of outrage and moralistic punishment, as humans tend to punish those who deviate from social norms.³ Research in social psychology suggests that punishment often targets a specific identity, such as the gender identity or race of the recipient.⁴ This punishment silences the voices of marginalized groups, thus replicating and even exaggerating existing power structures.⁵ A democracy, which relies on openness to diverse perspectives and opinions without the systematic exclusion of certain groups, is threatened, and thus, aggressive online harassment challenges unlimited free speech.⁶

There have been gender differences in compliance with COVID-19 safety recommendations. Men reported significantly lower intentions to wear face coverings in public, self-isolate when sick, avoid public gatherings, and wash hands more frequently.⁷ Individuals who decided not to comply with these recommended safety measures were publicly shamed and berated on social media; a phenomenon that the media has dubbed ‘pandemic shaming’ by influential news outlets such as the New York Times, The Guardian, The Atlantic, and The Washington Post. COVID-19 is a novel context in which to examine possible gendered behaviors. New behaviors required by the pandemic (e.g., wearing face masks) should not have been gendered when they emerged to ensure a fair society with equal treatment. If pandemic-specific behaviors evoked responses that differed along gendered lines, this would suggest that gendered hostility and punishment are intrinsic to contemporary social life and are not confined to gender-coded behaviors with roots in our past. To investigate the differences in online incivility experienced by men and women during the pandemic, I collected social media data from Twitter from December 2019 to December 2021 (N = 5,000,000) and measured the differences in the volume and content of toxicity directed towards men and women.

Past research examining Reddit’s r/AmITheAsshole, a thread in which users describe morally ambiguous scenarios in-depth and users vote on whether the Original Poster acted immorally, found that the Original Poster was more likely to be voted the aggressor when they were male.⁸ This research suggests that in certain online contexts men receive greater attention for their failure to abide by rules or live up to expectations. However, an exploration of aggressive online communication should not limit itself to the frequency or volume of harassment. The content of such communication must also be assessed. Toxic, outraged, or aggressive comments online may be constructive and designed to correct undesirable behaviors. Still, other comments can incite physical or sexual violence and deploy slurs and hate speech targeting individuals based on their group membership. Aggressive online communication styles may drive marginalized groups, such as oppressed gender identities and racial identities, into silence, measured by a lack of tweets from one demographic about a societal issue.⁹ Words such as “bitch” and “slut,” among others, target women.¹⁰ If any harassment is gendered at all, whether the frequency is different or not, there is a gender disparity that must be addressed. A possible consequence of this disparity is the silencing of marginalized groups as a response to gendered toxicity.

In this paper, I investigate potential gender disparities in both the amount and content of online toxicity targeting males and females for non-compliance with COVID-19 guidelines. I used a machine learning classifier to measure toxicity in a
dataset of tweets posted during the first two years of the pandemic, split by the gender of the target of the tweet. Using a topic model, I also explored whether the content of the tweets expressing toxicity differed across the target gender.

Methods
To test the differences in the volume and content of toxic posts targeting women versus men on Twitter during the pandemic, I collected a novel set of posts made to Twitter from December 2019 to December 2021 (N = 5,000,000). Tweets were collected using a strategy that allowed me to infer the gender of their targets. I also used an existing toxicity classifier (https://perspectiveapi.com) to estimate the probability that each tweet contained toxic language, defined as rude or disrespectful comments likely to lead the recipient to leave a conversation. Using this information, I could compare the counts of toxic tweets targeting men compared to women and test for differences. I also built two topic models, one for each group of tweets; those targeting men and those targeting women. This allowed me to conduct a preliminary investigation into the differences in content or type of language used to target women compared to men.

Dataset:
Twitter is a microblogging and social networking platform where users send out 280-character messages called tweets. Users may read tweets, post tweets, and follow other users to view their tweets. It is currently one of the leading social networking sites worldwide with 237.8 million monetizable daily active users as of 2020. I used the publicly accessible Twitter API to search for tweets that contained references to mask-wearing, hand-washing, or self-isolating alongside gendered pronouns (e.g., she/he). Each tweet had to tag another platform user to be included and posted immediately prior to and during a large portion of the COVID-19 pandemic, December 2019 - December 2021. This methodology allowed me to collect tweets that I could reasonably infer contained content about the unfolding pandemic and which addressed targets identified (at least in part) by their assumed gender. Due to restrictions on the number of tweets that can be collected at any one time from Twitter’s APIs, I developed a strategy to sample 5 million tweets equally distributed over the months of interest (i.e., December 2019 - December 2021) and gendered pronouns. My final sample contained exactly 5 million tweets split 50/50 across genders. An exclusive OR logic was used in the queries to Twitter’s API so that tweets containing both male and female pronouns were not collected.

Operationalizing Toxicity:
I made use of an existing, validated tool for the classification of toxic speech online. Perspective API (https://perspectiveapi.com) was developed by Google to assist content moderators in identifying various types of online text likely to be ‘toxic,’ where ‘toxic’ refers to rude and disrespectful language which appears designed to exclude the target from a conversation. Aside from its widespread use in industry and academia, Perspective API was useful for my purposes because it targets language that might be used to exclude users. Given the possibility that online toxicity was used during COVID-19 to exclude certain social groups from conversations about pandemic policy, this operationalization of toxicity can address whether women’s voices were sidelined in these debates. Perspective API estimates the probability that it contains severely toxic language for a given piece of online text. I classed as toxic any tweet in my dataset that the Perspective API assigned a probability of being toxic of .51 or higher.

Results and Discussion
Results:
I tested whether there were differences in the volume and content of toxicity targeting men and women on Twitter during the COVID-19 pandemic. Across all the tweets in my dataset, Perspective API classified 18.28% as severely toxic (Figure 1). When split by the gender of the target, a higher proportion of tweets targeting men were classified as toxic compared to those targeting women, 18.5% and 17.68%, respectively (Figure 2). While small, the results of a Fisher’s exact test revealed that this difference was significant, p < .001, such that men were targeted with a greater proportion of toxicity than women (Odds ratio = 0.95).

I also ran two GSDMM topic models on the portion of the data that was classed as severely toxic. I tested one model on the tweets that targeted men and the other on tweets that targeted women. This allowed me to explore the content-level differences between toxicity targeting men and women. I choose to use GSDMM models because they were explicitly designed to model short texts, such as tweets. Other common topic modeling techniques, such as LDA, perform less well on short texts. There are limited methods to present the results of GSDMM models visually. Below, I present for each model the 25 most frequently occurring words in the topic that contains the most tweets (Figure 3). As seen in the word clouds, the words most associated with men in the
toxic portion of the data invoke general stupidity and terms of abuse (i.e., ‘fucking’, ‘stupid,’ and ‘idiot’). However, some of the terms associated with women were specifically gendered slur terms (i.e., ‘slut’, ‘bitch,’ and ‘cunt’). The word ‘husband’ also appeared in this topic, signaling a woman’s relationship to a man. The internet trope of ‘Karen’ was also referenced. The specificity of the toxicity in the second model suggests content-level differences in tweets targeting men and women. Women appeared more likely to be targeted with gendered slurs and indexed to their husbands or to an internet meme of an entitled woman.

A)

B)

Figure 3: The 25 most frequently used words in the topic that contained the most tweets according to GSDMM models run on tweets that contained severely toxic language and that targeted (A) men and (B) women.

Overall, these results suggest that men received more tweets containing toxic language than women, but women were targeted with more gendered slurs.

Discussion
In this research, I hypothesized that 1) women would receive a higher volume of tweets containing toxicity, and 2) women would receive harsher and more offensive toxicity. The results surprisingly only supported the second hypothesis. However, the finding that women receive more gendered slurs than men while receiving fewer toxic tweets overall indicates that the gender disparity is stronger and more entrenched than the hypothesis may have predicted, urgently suggesting the necessity for political action.

While this research offers new insights and supports existing research, certain factors that limit my conclusions. For example, determining the gender identity of the users was limited to making inferences based on the pronouns she/her and he/him. Because they/them is both singular and plural, these data are too ambiguous to analyze and make conclusions about toxicity toward people who do not use she/her or he/him pronouns. Twitter does not provide exact demographic information to researchers. Additionally, this study only looked at toxicity pertaining to following safety guidelines and protocols for COVID-19, not all toxicity in the vast world of social media. This limits the generalizability of my findings to other domains.

Perhaps social media hate speech filters should add more warning signs, scanning for hate speech or language that incites violence before a tweet or allow users to filter those sentiments out of their feeds. While there exists a report feature for every post, the hateful speech is already online, creating a negative effect, making the change a reaction instead of a preventative measure.

Further research should focus on understanding why women receive more gendered and offensive posts online. Researchers should continue to suggest how to dismantle entrenched societal sexism and how society may work toward making outrage constructive instead of marginalizing.

Thus, although this research does not definitively provide a solution to women experiencing more gendered toxicity it exposes a problem to be addressed, one that likely affects all social media users.

Conclusion
While men receive a greater number of tweets containing outrage, women receive more gendered hostility in the content of the tweets, such as slurs that are offensive to women or phrases that incite sexual violence. This research demonstrates that sexism and punishment in society are still imposed by the community and are not yet obsolete, requiring action to be taken. To curb the negative effects, social media platforms should harness their influence to ensure that no voices are marginalized so that all may contribute to democratic public discourse.

Acknowledgments
I am deeply grateful to Killian McLoughlin, a Princeton Ph.D. student, for guiding me and offering resources for me to analyze my data. I am also appreciative of my teacher at Riverdale Country School, Rudy Nunez, for supporting and igniting my passion for pursuing data science to find out the truth about the world.

References


**Author**

Christine Lee is an eager data scientist to find impactful solutions to societal problems, specifically those related to the oppression of marginalized groups. Outside of research, Christine is interested in playwriting, a passion that complements data science because both require an understanding of influences that skew perspectives.
Development of Herb-based Moist Patch for Animal Wound Healing using Arginate Biofilm

Seungah Claire Cho
Marymount International School Rome, Via di Villa Lauchii, Rome, Italy; clairecho1015@gmail.com
Mentors: Dr. Woo Rin Lee, Dr. Woo Rin Lee

ABSTRACT: Inflammation is the biggest hindrance to wound healing, especially for animals. To solve this problem, we used widely known herb extracts that have significantly outstanding efficacy towards preventing wound healing inflammation: Hibiscus (H), Ginger (G), and Yarrow (Y) extracts. We directly extracted the herbs (H, G, and Y) and tested if these combined extractions prevented inflammation. Additionally, we produced alginate hydrogel mixed with the herb extracts to create wound-healing patches for animals. Next, we found that less than 0.5% of individual treatments of all three herb extracts showed low skin cell cytotoxicity. Next, we discovered that H and G + H + Y extracts decreased LPS-induced inflammation in mouse skin cells. Additionally, we used 2% alginate gel since the balance of tensile strength, flexibility, and stiffness was optimal. Finally, we created an alginate-based wound healing patch containing herbal extracts with a sticky bandage for animals. We concluded that our healing patches were appropriate for the animal because the patches were non-toxic and did not induce allergic reactions. Our novel wound healing patch may reduce animal suffering from wound inflammation. These novel patches will significantly progress toward new strategies for curing animal wound inflammation.

KEYWORDS: Biomedical and Health Sciences, Wound healing; Inflammation; Herb extracts; Alginate; Wound healing patch.

Introduction

Wounds are injuries that break the skin cell or other parts of the body tissues—identified as one of the most common damages for both animals and humans. Wound healing is a complex process supported by myriad cells to repair damaged tissue. Wound healing is restoring the normal anatomic continuity to an injured area of tissue. The level of contamination, blood supply, and the causation of the wound contribute to the development of the health conditions for infection.

Inflammation is the primary stage of wound healing, beginning immediately after the injury when the injured blood vessels leak transudate causing swelling. The main goal of inflammation is to infuse monocytes and lymphocytes, essential to wound healing, from white blood cells to the wound bed to prevent infections and control bleeding. Inflammation is classified into three phases: acute, subacute, and proliferative. Each phase tends to show the body's different clinical signs. For instance, during the acute phase, heat, redness, and swelling are representative signs. Alternatively, after the subacute inflammatory phase, tissue can repair and be strengthened during the remodeling phase.

During the inflammation process, bacterial infection may lead to the progression of a chronic wound. Furthermore, the increased inflammation in the skin and intrinsic causes of chronic wounds may lead to bacteria colonization worsening the chance of healing. Bacterial infection is the major factor that substantially influences the formation of a chronic wound injury.

Since ancient times, Hibiscus has been traditionally used for medical treatment in various ways. Hibiscus offers a long list of physical benefits for many different systems in the body, for instance, maintaining the healthy function of organs and supporting the reproductive system. Above all, Hibiscus tends to protrude in maintaining beneficial cholesterol concentration and healthy blood pressure. The wound-healing activity of Hibiscus was studied using three different models: excision, incision, and dead space wound. The Hibiscus extract has increased cellular proliferation and collagen synthesis at the wound site, regenerating skin tissue's total protein and collagen content. Predominantly, the extract-treated wounds were found to heal significantly faster, as indicated by enhanced rates of wound contraction.

Traditionally, ginger has been used for many different conditions, including treatment for colds, fevers, digestive abnormalities, and nausea. Ginger contains many active substances, such as triterpenoids, flavonoids and saponins, and 6-gingerol. Gingerol and shogaol are phenolic components that have anti-inflammatory effects. A previous study also showed that a combination of curcumin and ginger extract improves wound healing in rat skin. This study found that 6-gingerol is the most abundant in the ginger extract. Also, pretreatment of 6-gingerol increased the matrix metalloproteinase-1 (MMP-1) protein level, which is associated with skin vascular growth into collagen. Therefore, we hypothesized ginger extract might support skin wound healing with strong antibacterial effects.

Yarrow has been used for centuries for its outstanding potential health benefits, identified as a popular medicinal herb. People usually consume this herb by brewing it into tea. Based on recent research, it has been found that yarrow herb extracts
present anti-inflammatory and antioxidant properties, both of which aid wound healing. Specifically, yarrow extracts tend to increase the body cell's fibroblast concentration, the cell's main component responsible for regenerating injured tissues. Besides, yarrow tea reduces skin and liver inflammation, giving huge support in treating skin infections.

Since animal wounds are easily exposed to bacteria, which can enter wounds and release chemicals that prevent immune cells from eliminating bacteria. This process delays wound healing. Therefore, we hypothesized that creating a wound healing patch using an alginate biofilm mixed with the natural anti-inflammatory extracts from hibiscus, ginger, and yarrow may inhibit the bacteria-induced inflammatory response.

We successfully created alginate biofilm mixed with hibiscus, ginger, and yarrow extracts in this study. We also found that a 5% volume-to-volume (v/v) mixture of each hibiscus, ginger, and yarrow showed the most effective anti-inflammatory effect. This study may support the animals that suffer from severe inflammation caused by bacterial infection in their wound.

**Methods**

**Herb extraction**

Dried powder of ginger, hibiscus, and yarrow was purchased in the nearby supermarket. 10 g of each dried herb was boiled with 200 ml of distilled water for 2 hours. Then the extracted solution was filtered using filter paper (Whatman). The filtrate was stored at 4 °C.

**Herb extract treatment in mouse skin cells**

B16-F1 mouse skin cells were purchased from Korea Cell Line Bank. B16-F1 mouse skin cells were maintained in RPMI1640 (Gibco) supplemented with 10% fetal bovine serum (FBS) and 1% Penicillin-Streptomycin (PS). Trypsin-EDTA was used to detach the cells every three or four days to maintain the cells in healthy condition. The combined quantity of all substances, RPMI1640 and the herb extract, was set to 1000 µL. For instance, at 0.5% hibiscus concentration, 995 µL of all substances, RPMI1640 and the herb extract, was set to 1000 µL. EDTA was used to detach the cells every three or four days to maintain the cells in healthy condition. The combined quantity of all substances, RPMI1640 and the herb extract, was set to 1000 µL. For instance, at 0.5% hibiscus concentration, 995 µL of all substances, RPMI1640 and the herb extract, was set to 1000 µL.

**Cell confluency quantification using cell image**

The cell confluency was quantified using the EVOS M5000 imaging system (Thermofisher). Cell confluence measurement calculated the percentage confluence of cultured cells based on selected reference objects and background. The Confluence tool measures the percent area covered by cells in the image, representing the quality of live cells.

**Optimizing concentration of alginate gel**

100 mL of distilled water and 0.1 to 5 g of sodium alginate were mixed to generate a 0.1 - 5% sodium alginate solution. The properties of gel solidification, tensile strength, flexibility, transparency, and stiffness were analyzed.

**Results and Discussion**

We aim to test the cytotoxicity of ginger, hibiscus, and yarrow extracts in B16-F1 mouse skin cells. Since the previous experiment, concentrations over 5% treatment (all three herb extracts) caused the most cell death. Therefore, to determine the most appropriate concentration of herbs for inhibiting skin inflammation, we tested the concentrations within minimal gaps, having a maximum of 5% concentration. The cell confluency (vertical axis) refers to the proportion of the flask covered by adherent cells. Simply put, the higher the cell confluency, the more cells are alive, indicating cell viability (Figure 1). All three herbs tend to have alive cells in the 0.5% concentration, starting from around 75% of cell confluency (Figure 1). Herb extract shows outstanding cell viability with all six concentrations above 50% of cell confluency (Figure 1). On the other hand, hibiscus and yarrow tend to show higher cytotoxicity as the concentration increases, having almost annihilated cells at 5% concentration (Figure 1). We concluded that 0.5% of all three herb extracts show the most optimized concentration for the downstream experiments.

![Figure 1](image1.png)

**Figure 1:** Ginger, hibiscus, and yarrow herbal extracts decreased the cell confluency of B16-F1 mouse skin cells in culture plates. For cells that grow as a monolayer, cell confluency is defined as the percentage of the culture vessel surface area that appears covered by a layer of cells when observed by microscopy. The cell confluency represents cell viability.

Next, we tested whether herbal extracts have an anti-inflammatory effect on mouse skin cells. We quantified interleukin-2 (IL–2) expression, a bacteria-induced inflammation marker. Hence, increased IL–2 quantity means an increase in inflammation. We used lipopolysaccharides (LPS), a lipid and a polysaccharide produced from bacteria, to induce inflammation in mouse skin cells. In this study, with the presence of LPS, we treated 0.5% of herbal extracts in the mouse skin cells to find the effect of herbal extracts on inhibiting inflammation. We tested herb extract of ginger (G), hibiscus (H), and yarrow (Y) individual treatment and two or three combination treatments. As shown in Figure 2, configured the experiment with a total of nine treatments: no treatment, LPS only, G only, Y only, G + H, G + Y, H + Y, and G + H + Y. We incubated these nine treatments for...
Finally, to produce the most appropriate wound healing patch for inflammation, we concluded that alginate biofilm (nontoxic and transparent) is the best fit. After soaking the alginate biofilm into the herbal extraction solution, we successfully created the wound healing patch (Figure 3). To begin with, we used 2% concentrated alginate gel to produce alginate biofilm and dried it for 24 hours to remove all the water solution from the gel. At last, biofilm was successfully created and stored at 4°C. Each wound healing patch was then cut into a rectangular shape and soaked in an herbal extracted solution for 1 hour. After all, we produced the biofilm successfully (Figure 3). Since alginate is hydrophilic, the alginate wound healing patch contained the herbal extract solutions for more than 24 hours. The wound healing patch cannot be attached to the animal’s skin alone. Therefore, we attached the biofilm to the sticky bandage, making the final patch very flexible and easily attached to the animal’s skin (Figure 3). We created an alginate-based wound healing patch containing herbal extracts with a sticky bandage for animals.

To find out the best concentration of alginate film, we compared six different concentrations of alginate hydrogel. The agarose gel with 0.1 and 0.5 % did not solidify. Although the 0.5 % alginate gel solidified, it significantly lacked tensile strength and stiffness, causing it to rip easily (Table 1). The 1 % alginate gel was more durable, but it still lacked tensile strength and stiffness (Table 1). 4 % and 5 % were both highly durable but lacked transparency and flexibility (Table 1). Thus, we found that the 2 % alginate gel to be the most optimal because it had balanced characteristics between tensile strength, flexibility, and stiffness.

<table>
<thead>
<tr>
<th>Alginate (%)</th>
<th>Gel solidification</th>
<th>Tensile strength</th>
<th>Flexibility</th>
<th>Transparency</th>
<th>Stiffness</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>No</td>
<td>X</td>
<td>N/A</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>0.5</td>
<td>Yes</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>Yes</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>5</td>
<td>Yes</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

Finally, to produce the most appropriate wound healing patch for inflammation, we concluded that alginate biofilm (nontoxic and transparent) is the best fit. After soaking the alginate biofilm into the herbal extraction solution, we successfully created the wound healing patch (Figure 3). To begin with, we used 2% concentrated alginate gel to produce alginate biofilm and dried it for 24 hours to remove all the water solution from the gel. At last, biofilm was successfully created and stored at 4°C. Each wound healing patch was then cut into a rectangular shape and soaked in an herbal extracted solution for 1 hour. After all, we produced the biofilm successfully (Figure 3). Since alginate is hydrophilic, the alginate wound healing patch contained the herbal extract solutions for more than 24 hours. The wound healing patch cannot be attached to the animal’s skin alone. Therefore, we attached the biofilm to the sticky bandage, making the final patch very flexible and easily attached to the animal’s skin (Figure 3). We created an alginate-based wound healing patch containing herbal extracts with a sticky bandage for animals.

The position of the animal wounds varies depending on the type of injury; hence, the shape of the band is essential to cover the whole wound position. We created Alginate-based wound healing patches of different sizes and shapes. Cutting it into pieces using scissors allowed us to create different shapes easily. When cutting, we had to make sure the shape of the biofilm and bandage had equal in size. We produced circular, triangle, square, rectangle, plus shape, and other shape polygons (Figure 4). This advantage allows our band to shape into any formation depending on the animal’s wound position and shape. In case of wound inflammation, depending on the animal’s wound position, this patch provides the best-optimized shape based on the wound size and position (Figure 4).
Since we designed healing patches for animals, we had to make sure they were appropriate for animals by attaching the patches to them. We attached the healing patches to our pet’s body parts, such as the leg, head, and breast. We also attached these above their fur to ensure the patch’s adhesive strength. After securing the healing patches for about a day, we concluded that our healing patches were appropriate for the animals. The patches were non-toxic and did not occur any allergic reactions. Even though they were attached to the animal’s fur, we found that the patch was affixed for a long time (Figure 5). Since our patches could be produced in different formations, we were able to stick these patches to various parts of the animal’s body (Figure 4). Moreover, the animals did not tend to find these patches encumbered, allowing the patches to be long-lasting.

### Conclusion

Based on our first experiment, we found that 0.5% of all three herb extracts show the most optimized concentration for downstream experiments. (Figure 1) Also, we found that the herbal extract of the H-only sample and G + H + Y combination treatment decreased the LPS-induced inflammation in mouse skin cells after several tests. (Figure 2) After figuring out the best alternative concentration and type of wound healing patch for inflammation, we successfully created an alginate-based wound healing patch containing herbal extracts with a sticky bandage for animals. (Figure 3) By this, we could present our band’s advantage: the band could be formed in any shape. Depending on the animal’s injured position, we customized the patch to be the best-optimized shape based on the wound size and position in case of inflammation. Therefore, we could stick these patches to various parts of the animal’s body (Figure 4). Moreover, the animals did not tend to find these patches encumbered, allowing the patches to be long-lasting.

Considering that our experiments used tiny volumes of herbal extracts and LPS, a slight error in the injected concentration of extracts may cause some errors in our results. Likewise, while distributing the experimental cells in equal proportions, small contact with the atmosphere or cell culture dish may affect the viability of mouse skin cells. Additionally, to examine the stabilized results, this experiment required much patience and a long time. For example, to inspect the effect of combined herbal extracts on the LPS-induced cells by the IL-2 expression level, it required about a day for cells to be completely affected by the extracts.

Alginate film is a highly effective wound-healing patch that offers a range of benefits for patients and healthcare providers. One of the main advantages of using alginate film is its ability to retain moisture in the wound, which helps to promote healing and reduce the risk of infection. Alginate film creates a physical barrier that helps prevent bacterial contamination, making it an ideal option for patients with open or infected wounds. Another benefit of alginate film is its conformability, which is used on many wound types, including irregular or hard-to-reach wounds. In addition, alginate film is bio-compatible with human and animal tissue and easy to remove, causing minimal pain or discomfort to the patient. Furthermore, alginate is a natural material derived from seaweed and is biodegradable, making it an environmentally friendly option for wound healing and cost-effective.

Ecofriendly wound healing patches are important because they reduce the environmental impact of traditional wound care products, which often contain plastic and other non-biodegradable materials. We made eco-friendly patches from natural materials, which are biodegradable and compostable. Additionally, we develop these patches by using natural ingredients for wound healing. Using eco-friendly wound healing patches can help reduce waste and pollution and promote more sustainable and natural healing practices.

In conclusion, the final wound healing patches produced were non-toxic and did not occur any allergic reactions toward the animal specimens. We also conclude that our healing patches can be engaged with other strategies to develop various animal medical treatments further.

### Acknowledgments

This research was supported/partially supported by Dr. Woo Rin Lee. We thank our colleague, Dr. Woo Rin Lee, who provided insight and expertise that greatly assisted the research.

### References

domized Controlled Trials. *Nutrients* 2020, 12 (1).


**Author**

Claire Cho is currently a senior at Marymount International School Rome. She majors in Biology, Chemistry, and Advanced math at school. She is willing to extend her passion for biology by pursuing a career in veterinary medicine.
Analyzing the Views of Ethnic Minority Teenagers in Texas on Organ Donation and Organ Donation Policies

Emily Wang

Clear Creek High School, 2305 East Main Street, League City, TX, 77573, USA; emilywang1369@gmail.com
Mentors: Mr. Roberto Morales

ABSTRACT: The United States currently operates under an opt-in organ donation policy, in which individuals must sign up to become organ donors. Nevada Senate Bill 134 proposed a change from the current opt-in system to an opt-out system that would automatically register individuals as organ donors. It became more pressing than ever to examine the thoughts of the American population on possible organ donation policies. Thus, this research aimed to explore the views of Texas teenagers from minority ethnic backgrounds on organ donation and various organ donation policies. This research utilized a mixed methods approach by distributing a survey to Texas high school teenagers and conducting in-depth interviews with five teenagers from ethnic minorities. After an analysis of variance test and thematic analysis were conducted on the collected data, it was determined that there was a lack of knowledge about organ donation among teenagers. The extent to which family, media, and ethnic background factor into the views of teenagers on organ donation was also determined. Lastly, the research showed a general openness to an opt-out policy, indicating that it could be reasonably implemented in the next few decades.

KEYWORDS: Behavior and Social Sciences; Sociology and Social Psychology; Public Policy; Organ Donation; Ethnic Minorities.

Introduction

For decades, the United States has suffered from an organ donation shortage. Organ donation is surgically removing an organ from the donor and placing it into the recipient whose previous organ has failed.¹ The Health Resources and Services Administration² states that nearly 110,000 men, women, and children are on the national transplant waiting list.

The United States currently operates under an “opt-in” system, defined as a system in which individuals must actively register to donate their organs. Nevada Senate Bill 134 proposes that drivers “opt out” of organ donation instead.³ This means citizens would be automatically enrolled as organ donors at birth. If an individual or their family decides they do not want them to be an organ donor, they can be taken off the organ donor registry.

With a feasible transition to an opt-out policy, it was essential to determine the younger generation’s opinions about this issue. In addition to teenagers being future voters and possible organ donors, they are also heavily influenced by their family and cultural background. Organ donation was examined at length from several perspectives to provide a clearer understanding of the future of the field in the United States.

The Ethicality of Organ Donation:

The debate over the ethicality of organ donation has been ongoing for over forty years. At the time of its introduction, many thought transplantations were too dangerous. Unforeseeable effects on the organ donor raised the question of whether doctors should advise living patients to donate organs, even if it could negatively affect the patient’s health.⁴ Another concern at the time was the risk to the living donor in proportion to the recipient’s benefit. Essentially, there was a possibility that two lives could be ended trying to save one. Over time, the perception of death by both the public and the medical profession has shifted significantly. With technological advancements, like mechanical ventilation, organs can be increasingly salvaged for transplantation. According to Arthur Caplan⁵, this has resulted in a “shift in the legal definition of death toward the so-called ‘brain death’…. Brain-death statutes permitted organs to be harvested from those who had suffered an irreversible loss of brain function….”. A survey conducted in 2015 found that 71% of participants favored organ donation, even when transplantation was described as causing the donor’s death. This can explain why several countries have started to shift their organ donation policies to accommodate this newer view of organ donation.⁶

The Types of Organ Donation Policies:

The policy in the United States is an opt-in policy, which balances individual rights with virtues like sacrifice and altruism. The efficiency of this policy has been a point of debate for decades, as it is acknowledged that the supply of organs is struggling to keep up with the increasing demand. This system has also resulted in a complicated network of various procurement agencies operating differently. Thus, it can be supported that an opt-in system is not sufficient.⁵ As a result, an opt-out system that automatically registers individuals as organ donors has been implemented in several countries. In Belgium, which has an opt-out system, the family refusal rate is only 13% compared to the 49% rate in the U.S.² Researcher Romelie Rieu⁷ argues in his journal article that at least with an opt-out system, more people are benefitting from organ donation. Additionally,
Austria’s opt-out system has made participants feel that deciding to be an organ donor is less substantial. An opt-out policy implies that being an organ donor is the default. From this perspective, the question people ask themselves when registering or renewing their driver’s license is, “Do I want to be in the minority of people unwilling to help others in my community?”¹⁸ The impact of organ donation policies on the psychological understanding of the procedure is an essential component of the discourse around the hesitancy to become an organ donor in certain countries, such as the United States. A significant point of debate regarding a switch to an opt-out system is the system’s ethicality. In the United States, the American Society of Transplant Surgeons states that organ donation can only be considered ethical if the individual involved is aware of the decision they are making. It can be argued that if the entire population is not informed of what an opt-out system entails, this system could be unethical.¹⁹ However, in response, it could be argued that inaction can be considered consent, which can be acted on.¹⁰ The common use of the phrase “speak now or forever hold your peace” in modern weddings is an example of how failure to say “no,” in essence, constitutes a “yes.”

The other organ donation policy that has been proposed is an incentive system. These incentives can be non-monetary or monetary. The research found that more people donated blood when rewarded with gold, silver, and bronze medals, which supports the success of implementing a non-monetary policy. The other possible option is monetary compensation that can be given to the deceased’s family. However, researchers have acknowledged that this type of system would come with various risks. For instance, some families may be motivated by compensation, or the public may lose trust in the overall healthcare system.¹¹ In the past, five provinces in China tried a financial compensation policy in which the Red Cross Society of China paid for funeral expenses and the cost of the grave plot. However, it was concluded that this policy might need to be more efficient in the long term, as the incentive could become less appealing over time.¹² Of course, offering financial incentives to families of deceased individuals to encourage them to consent to organ donation is a controversial idea. There is a debate that this is unethical and undermines the altruistic appeal of the current system. Different types of incentives will have different degrees of acceptability to the public and healthcare professionals. Researchers sent questionnaires to transplant surgeons, transplant coordinators, and critical-care nurses to determine this. They found that all three professions believe alternative policies that offer non-monetary or indirect monetary incentives that benefit the donor to be morally appropriate. In contrast, policies that provide direct economic incentives to the donor family are ethically inappropriate.¹³ This suggests that medical professionals are also skeptical of a system involving financial incentives. Thus, an organ donation policy involving either non-monetary or monetary incentives would not be well received unless it directly benefits the donor.

**The Public Perception of Organ Donation:**

Since the start of organ donation, a gap between supply and demand has raised the question of who should receive what. According to professor and researcher Surjit Singh Dhooper¹⁴, there has always been a degree of bias among racial divisions in the distribution process. As of 1986, organs were distributed based on time, medical need, the relationship between donor and recipient, and quality of organ and tissue match. However, data has shown that African American candidates are less likely than White candidates to receive an organ transplant. This presents a concern because, according to Kierans and Cooper, “perception that the selection of organ recipients is fair and equitable is essential to the operation of the entire organ transplantation program.” This issue is due to the manifestation of racism in transplantation practices. Up until recently, kidney recipients and donors were matched by blood type. Group O is the most common blood type in the UK, but 38% of the Asian population has Group B blood type. This puts minorities at a disadvantage both when donating and waiting for organs. In the UK, there have even been media campaigns targeted at minority groups that urge them to help their communities by becoming organ donors.¹⁵ This indicates that media considerably influences a person’s willingness to become a registered organ donor. With media having an increasingly large presence in the lives of teenagers through social media and the 24-hour news cycle¹⁶, thoughts regarding organ donation may reflect the media being generated. There is also a possibility that reluctance among ethnic minority communities is due to medical mistrust and their cultural beliefs. For example, the deceased donor rate in China is significantly lower than in Western countries. Researchers have found this to be the result of traditional cultural beliefs in China that state a body should remain intact after death.¹² This suggests that cultural beliefs could have a prominent influence on opinions regarding organ donation. Since cultural and religious beliefs often overlap, there was a possibility that religion would influence thoughts on organ donation too. Range and Brazda¹⁷ tested this by gathering participants from students at a Catholic college in the southeastern United States. The participants were given questions that asked them to rate their agreement with a statement on a scale from 1 to 7. This showed that religious involvement did not correlate significantly with any view on organ donation. This study indicates that the effect of an individual’s ethnic background on their views toward organ donation should be further researched.

Organ donation continues to be a prevalent topic with several points of debate. With the proposal of a change to an opt-out system in Senate Bill 134, it is more pressing than ever to examine the thoughts of the American population on possible organ donation policies.³ Research regarding public opinion of guidelines has been done before in various countries, but the views of specifically teenagers from ethnic backgrounds on this issue have yet to be explored. Teenagers provide a unique perspective on the issue, as they are influenced by their own culture, family, experiences, and the media. Thus, this research aims to fill the gap present in the field today about the views of Texas teenagers from ethnic backgrounds on organ donation and various organ donation policies.

---

DOI: 10.36838/v5i4.8
### Methods

This research utilized a non-experimental design and an exploratory approach to make connections between the views of teenagers from ethnic minority populations and low organ donation rates. The study was also conducted through a mixed-method process. Quantitative data was collected using a survey, and qualitative data were collected through five interviews with teenagers from ethnic minorities. This method was chosen because it allowed for exploring trends across a sample group and personal experiences specific to an individual. Other research regarding organ donation, like that of Jasper et al.\(^\text{13}\), also employed a mixed methods approach.

The survey was created on Microsoft Forms, and the interviews were conducted on Microsoft Teams. This was the most accessible platform for the participants in this research to use. The digital survey first included questions regarding the participant’s age, location, and ethnic background. The rest of the survey questions followed the structure example below in Figure 1. The complete survey with all 15 questions can be found in Appendix A. All five participants signed an informed consent form and received parental consent before they completed their interview.

![Figure 1: Example of a question from the survey used.](https://www.example.com/figure1.png)

First, a survey was created with questions about organ donation and organ donation policies. Questions asked participants to rate their response to a statement on a Likert scale from 1 to 7. A Likert scale was chosen, as it is widely used and allows for degrees of opinions to be easily expressed. All questions also included an additional option outside the scale labeled “Not sure.” This prevented participants from choosing a random answer if they felt overwhelmed and did not know how to answer the question. Next, the survey was distributed to the sample group of approximately 100 Texas high school students between the ages of 13 and 18 from various ethnic backgrounds. Students signed an informed consent form and received parental consent before they began the survey. Their personal information was kept completely confidential by limiting the viewing of the survey responses to only the researcher. After the survey was distributed, participants that were both willing to be interviewed and were from an ethnic minority community participated in an interview to gain more detailed information on their responses.

### Results and Discussion

#### Survey Findings:

In total, 135 responses to the survey were collected over seven weeks. Of the 135 respondents, 122 answered each question thoroughly and provided usable contact information. Figure 2 displays the demographic characteristics of these 122 respondents.

<table>
<thead>
<tr>
<th>Age</th>
<th>Frequency</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>42</td>
<td>31.11%</td>
</tr>
<tr>
<td>15</td>
<td>30</td>
<td>22.22%</td>
</tr>
<tr>
<td>16</td>
<td>28</td>
<td>20.74%</td>
</tr>
<tr>
<td>17</td>
<td>19</td>
<td>14.07%</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>2.22%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region of Residence</th>
<th>Frequency</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Texas</td>
<td>17</td>
<td>12.59%</td>
</tr>
<tr>
<td>Southeast Texas</td>
<td>101</td>
<td>74.81%</td>
</tr>
<tr>
<td>West Texas</td>
<td>1</td>
<td>0.74%</td>
</tr>
<tr>
<td>Upper Rio Grande Texas</td>
<td>1</td>
<td>0.74%</td>
</tr>
<tr>
<td>Capital Texas</td>
<td>1</td>
<td>0.74%</td>
</tr>
<tr>
<td>Northwest Texas</td>
<td>1</td>
<td>0.74%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race</th>
<th>Frequency</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>68</td>
<td>50.37%</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>17</td>
<td>12.59%</td>
</tr>
<tr>
<td>Asian</td>
<td>23</td>
<td>17.04%</td>
</tr>
<tr>
<td>Black or African American</td>
<td>13</td>
<td>9.63%</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>1</td>
<td>0.74%</td>
</tr>
</tbody>
</table>

The fourth survey asked participants to rate their agreement to the statement “I am knowledgeable about organ donation” on a scale from one to seven, with one indicating they strongly disagree and seven indicating they strongly agree. The mean response of all teenage participants, including those that selected a 4 and excluding those that selected “Not sure,” was 3.82. The median of the data was a response of 4, and the mode was a response of 4. The responses of the participants can be seen in Figure 3 below.

![Figure 3: Participant agrees to the statement “I am knowledgeable about organ donation.”](https://www.example.com/figure3.png)

The fifth question of the survey asked participants to rate their agreement with the statement “My views on organ donation have been influenced by my ethnic background” on the same scale. The mean response of all teenage participants was 1.92. The data median was a response of 1, and the mode was a response of 1. The responses can be seen in Figure 4 below.
The twelfth survey question asked participants to rate their agreement to the statement "Under a monetary incentive system, I would be more likely to register as an organ donor." on the same scale. The mean response of all teenage participants was 3.92. The median of the data was a response of 4, and the mode was a response of 1. The responses of the participants can be seen in Figure 8 below.

The thirteenth question of the survey asked participants to rate their agreement to the statement "Under a non-monetary incentive system, I would be more likely to register as an organ donor." on the same scale. The mean response of all teenage participants, excluding those that selected "Not sure," was 3.20. The median of the data was a response of 3, and the mode was a response of 1. The responses of the participants can be seen in Figure 9 below.

The fifteenth and final question of the survey asked participants to rate their agreement to the statement "An incentive policy would be unethical." on the same scale. The mean response of all teenage participants was 2.56. The median of the data was a response of 3, and the mode was a response of 1. The responses of the participants can be seen in Figure 10 below.
Interview Findings:

Lastly, 24 participants said they wanted to participate in an interview. Five of these participants were from ethnic minority communities and were contacted to schedule a virtual interview. Only these participants were selected for interviews, as the research aimed to analyze the views of solely ethnic minority teenagers and not in comparison to their White counterparts.

Each interview transcript was then summarized and condensed to only the most prevalent or heavily discussed themes expressed by each participant in response to the questions asked. The main themes of each interview are listed in Figure 11 below. The de-identified transcripts are available upon request.

| 17-year-old Hispanic female | Never thought about donation
|                          | Family discouraged donation
|                          | Opt-out is not unethical
|                          | Incentive not much of a contributing factor
|                          | Donation should be more widely discussed
|                          | Received most influence from parents |
| 18-year-old Hispanic female | Already a registered donor
|                          | Feels a moral responsibility to help
|                          | Views differ from others in ethnicity
|                          | Influenced most by media
|                          | Monetary incentive policy would be best |
| 18-year-old Asian female  | Already a registered donor
|                          | Family does not talk about donation
|                          | Does not see donation in media
|                          | Comfortable under opt-out policy
|                          | Incentive would not contribute to decision
|                          | Change not necessary, but beneficial |
| 18-year-old Asian female  | Donation is scary topic in culture
|                          | Not knowledgeable on donation and policies
|                          | Opt-out policy requires additional education
|                          | Incentive does not influence decision
|                          | All policies are ethical
|                          | Policy change is necessary |

Figure 10: Participant agrees to the statement "An incentive policy would be unethical."

Analysis:

This research suggested that conclusions could be drawn regarding the views of teenagers from various ethnic minorities on organ donation and organ donation policies. One hundred thirty-five individuals in Texas took a survey asking them to rate their agreement to a statement on a Likert scale from 1 to 7.

An analysis of variance or ANOVA test with a Tukey Honestly Significant Difference post-hoc test with a 95% confidence level was used to determine whether there were statistically significant differences between the five ethnic minorities surveyed. This test was used for further quantitative data analysis because more than two groups were tested for differences.

The five groups used in the ANOVA test were as follows: White, Hispanic or Latino, Asian, Black or African American, and American Indian or Alaska Native. The data was first separated into these five groups. Then, for each analyzed survey question, selections of “Not sure” were excluded to accurately depict the views that the participants confidently submitted.

Six of the 15 questions asked in the initial survey were selected to be statistically analyzed. This was due to the amount of data collected and the value of the data in relation to the research question. Of the six questions selected, 2 produced results of significance.

The first question selected and of significance follows: "My ethnic background has influenced my views." This question was chosen because it further investigated how ethnicity could affect opinions on organ donation. The responses to this statement differed based on how participants felt their culture had impacted them. The number of participants in the ethnic group, their responses’ arithmetic mean, and their responses’ standard deviation were inputted into an ANOVA test calculator. The results of the test are shown below in Figures 12 and 13.

Source of Variation | Sum of Squares | d.f. | Variance | F    | p    |
---------------------|----------------|-----|----------|------|------|
Between Groups:      | 29.7396        | 4   | 7.4349   | 2.6392 | 0.0385 |
Within Groups:       | 287.6204       | 95  | 2.8171   |       |      |
Total:               | 297.3600       | 99  |          |      |      |

Figure 12: ANOVA table for responses to the statement "My views on organ donation have been influenced by my ethnic background."

<table>
<thead>
<tr>
<th></th>
<th>Diff</th>
<th>CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>White vs Hispanic or Latino</td>
<td>0.2143, 95%</td>
<td>-1.1804 to 1.6089</td>
<td>0.9929</td>
</tr>
<tr>
<td>White vs Asian</td>
<td>0.7815, 95%</td>
<td>-0.5110 to 2.0740</td>
<td>0.4502</td>
</tr>
<tr>
<td>White vs Black or African American</td>
<td>1.5952, 95%</td>
<td>0.1105 to 3.0800</td>
<td>0.0288</td>
</tr>
<tr>
<td>White vs American Indian or Alaska Native</td>
<td>-0.5714, 95%</td>
<td>-5.2804 to 4.1375</td>
<td>0.9972</td>
</tr>
<tr>
<td>Hispanic or Latino vs Asian</td>
<td>0.5672, 95%</td>
<td>-1.1173 to 2.2517</td>
<td>0.8820</td>
</tr>
<tr>
<td>Hispanic or Latino vs Black or African American</td>
<td>1.3810, 95%</td>
<td>-0.4552 to 3.2171</td>
<td>0.2323</td>
</tr>
<tr>
<td>Hispanic or Latino vs American Indian or Alaska Native</td>
<td>-0.7857, 95%</td>
<td>-5.6170 to 4.0455</td>
<td>0.9912</td>
</tr>
</tbody>
</table>

Figure 11: Main themes expressed by interview participants.
Since the p-value of the comparison between White and Black or African American participants was below a 0.05 value, there was a statistical difference in the thoughts between the two surveyed groups on the influence of their ethnic background on their views regarding organ donation. The large sample size difference between the two groups was accounted for when running the ANOVA test, but variables such as gender, household income level, and personal experiences were not. Due to these research limitations, it can only be shown that the ethnicity of the participants was a possible factor of this statistical significance rather than the sole precipitator. It is also important to note that the introspection of the participants limits the data. Due to the nature and subject of the research, there is inevitably some degree of bias and error.

The second statement selected and of significance follows: "Under a non-monetary incentive system, I would be more likely to register as an organ donor." This question was chosen because it focused on a non-monetary incentive system. The responses reflected how various participants viewed this system and how they believed they would be affected by it. The responses could be used to predict the effectiveness of a non-monetary incentive system within this demographic. The results of the ANOVA test can be seen below in Figures 14 and 15.

### Table 1: ANOVA table for responses to the statement "Under a non-monetary incentive system, I would be more likely to register as an organ donor."

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Variance</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups:</td>
<td>44.3712</td>
<td>4</td>
<td>11.0928</td>
<td>3.5786</td>
<td>0.0098</td>
</tr>
<tr>
<td>Within Groups:</td>
<td>247.9818</td>
<td>80</td>
<td>3.0998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td>292.3529</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since the p-value of the comparison between Hispanic or Latino and Black or African American participants was below a 0.05 value, there was a statistical difference in the thoughts between the two surveyed groups on the influence of a non-monetary incentive policy. The sample size difference between the two groups was accounted for when conducting the ANOVA test. However, it still cannot be proven that the ethnicity of the participants was the sole indicator. Again, other variables, such as gender, household income level, and personal experiences, could not be accounted for. There were also different interpretations of the question, as exposed through later interviews. Some participants may have believed the question to be about the personal benefit gained from monetary compensation.

In contrast, others may have thought it to be referring to the altruism of organ donation. In the future, a pre-survey should be done to gauge conflicting interpretations before the research is done. Once again, the data is limited by the introspection of the participants.

Of the 135 surveyed individuals, five participants were interviewed. The five interview transcripts were analyzed for themes to expose the general thoughts of teenagers from ethnic minority communities on organ donation.

The most prevalent theme across the multiple interviews was the influence of family on views of organ donation. Throughout all 5 interviews, the participants mentioned the term “family” 29 times and the terms “parents” or “grandparents” 7 times. When asked the question, “From where would you say you have received the most influence regarding your views on organ donation?”, 4 of the 5 responses were “My parents,” “My grandmother and my dad,” “My family, probably,” and “My home.” This indicates that the views of teenagers on organ donation have been primarily influenced by their families.

Another prevalent theme throughout the interviews was the role of media in developing views on organ donation. When asked, “In what way would you say media has or hasn’t influenced your views?” the participants’ views slightly differed. One participant stated, “I’ve never seen it like, being advertised before.” Other participants shared the same sentiment, saying that “...from just what I read, I haven’t had the topic of organ donation...” and “...I don’t know if I’ve seen things about organ donation...” In contrast, one participant expressed that media had a prominent role in how they think about organ donation. They stated, “…so many people like on social media or just in the media...need help...” This supports research done by Dhooper¹⁴, which found that organ donation heavily relies on its altruistic appeal to the public to succeed.
Additionally, when asked about how their ethnic background has influenced their views, the consensus was that they did not sense there was much of an influence at all. One participant stated, "...you're still going to have like a varying percentage of those who will be an organ donor and who won't be an organ donor...." Other responses include "...my family doesn't really talk about it much..." and "...I'm not sure what other South Asians besides my own family thinks...." The survey responses seemed to lean towards this view, as well. The mean response of all teenage participants, excluding those that selected “Not sure,” to the statement “My views on organ donation have been influenced by my ethnic background.” was 2.3. The results of this research align with a previous study done by Range and Brazda¹⁷, who found that religious involvement did not correlate significantly to any particular view on organ donation in college students. This suggests that while the views of older generations on organ donation may coincide with specific cultural beliefs, younger generations have a different perspective. One participant reflected this with their response that "...as I grew up, it (cultural beliefs) no longer influences me that much." Teenagers may receive most of their information about organ donation from their family and ethnic background, but their own views will continue to develop with time. One theme that was commonly mentioned across all interviews was a lack of knowledge about organ donation itself. When asked, “Do you meet the requirements to register as an organ donor?” several participants were unaware of the requirements. The same sentiment was reflected when asked to reflect on whether they felt that there was enough information available for them to form their own views on organ donation. Some of the participants’ answers included, “...it’s never seen in like the TV or the news or anywhere” and “...in school...I don’t hear about organ donation”. These responses highlight an overall lack of knowledge among the younger generation regarding organ donation. This was also seen in the quantitative results of the survey, in which the mean response of all teenage participants, excluding those that selected “Not sure” to the statement “I am knowledgeable about organ donation,” was 3.82. This highlights a need for more information about donations explicitly targeted at teenagers.

Regarding the various organ donation policies, there was a general openness to the idea and implementation of an out-out policy among the teenagers that stated they were already comfortable with being organ donors. One participant explained that they “feel like having an opt-out policy is better for people because...if you really don't want to, then you can go through the...hassle of...opting out.” This sentiment that an opt-out system would make the process of becoming an organ donor easier and, therefore, increase donation rates was reflected by another participant, who stated that “...by being registered at birth and then being able to opt-out at any moment, it gives a bigger opportunity to have organs donated.” However, it is essential to note that the other significant barrier to increasing organ procurement rates is the criteria for deceased organ donation. When asked about the ethicality of an opt-out system, 4 of the 5 interview participants believed that it would be ethical because there is still a choice to opt-out. One participant explained that “because it's something that you still have a choice of doing...it's really not unethical in my point of view....” The only participant that did not agree expressed reservations about the ethicality of an opt-out system when knowledge about it is not widespread enough. They stated, "...if our country has an opt-out policy and they (people not originally from the United States) don’t know that coming here, then that's not exactly very ethical...." This perspective echoes the views of Allard and Fortin⁹, who argued that a decision is not ethical if the individual is not aware of the decision in the first place. The two opposing views presented throughout the interviews demonstrate that even within the younger generation, the debate surrounding the ethicality of donation is still prevalent.

Lastly, the role of monetary and non-monetary incentives was discussed throughout the interviews. The five participants presented two views. The first was that incentives would motivate others to engage in living organ donation. One participant stated, “...it would help me be more inclined to do it because you’re still receiving some sort of benefit for yourself.” However, other participants said they “...would do it (register as an organ donor) just to help others rather than for the incentive.” These different views highlight a previously mentioned limitation in the research, as the various participants interpreted the question differently. The confusion over the interpretation of the questions could have been prevented by conducting a pre-survey.

**Conclusion**

Previous research regarding organ donation has explored debates surrounding the ethicality of assistance and alternative donation policies to the opt-in system. While research has examined influences like religion on organ donation views, a study focused on teenagers from different ethnic backgrounds was not yet done. This research aimed to fill that gap in knowledge through an exploratory and mixed-method approach, utilizing both a quantitative survey and several qualitative interviews. The results of this research present several new conclusions that can be used to guide future organ donation legislation.

First and foremost, it has become apparent through both survey results and interviews that teenagers need to be more knowledgeable about the organ donation process. This presents a vital concern, as knowledge was determined to be an essential factor in evaluating the ethicality of organ donation itself. More information must be provided to young people to guarantee an effective rate of organ donation in the United States and the ethicality of the act itself. This could be done through social media campaigns or educational resources distributed through high schools.

The research also presented several prominent factors that influence the views of teenagers on organ donation. While most participants stated that their family most influenced them, others were influenced by media about organ donation. The importance of family in developing views on organ donation suggests that to increase donation rates in the United States among the younger generation, older generations must first have a favorable view of the issue by increasing the trustworthi
ness of medicine and the medical system. The other influential factor of media can be further utilized to spread information to teenagers about organ donation and the steps they can take to register as organ donors. Lastly, participants were asked about the influence of their ethnic background on their views regarding organ donation. The consensus of those interviewed was that it was not a prominent influence. However, this could not be concluded due to several research limitations. First, there was a significant variation in the number of participants from each ethnic group, resulting in a sample group not consistent with the makeup of Texas. This means that the ANOVA tests were also less effective than possible and could be slightly problematic. Secondly, participants were not equally distributed across Texas, as most of them were located in its Southeast region. Lastly, the five interviewed participants could only represent Asian and Hispanic young women. More research with larger and more representative sample groups must be done to confirm the results of this research across several demographics.

The research concludes that teenagers who are already open to the idea of organ donation believe an opt-out policy to be both an ethical and effective way to increase donation rates in the United States. As for monetary or non-monetary incentive policies, several views were most prominent in this research. The first is that financial compensation would contribute to one’s decision to become an organ donor, resulting in an overall increase in donation rates. The second is that the idea of the compensation would not sway those with a positive or negative view of organ donation. This suggests that while an incentive policy would be effective in some regards, it would not be a motivating factor for all individuals. Additional research is needed to explore the benefits of implementing an incentive policy in the United States.

In conclusion, this research regarding the views of teenagers from different ethnic minority backgrounds found a general lack of knowledge surrounding organ donation. Additionally, it explored the extent to which family, media, and ethnic background factor into the views of teenagers on organ donation. Lastly, it analyzed the response of teenagers to the alternative donation policies of opt-out and incentive systems. The conclusions drawn from this research suggest that greater educational organ donation resources must be offered to young people to move forward in the current organ donation landscape.

Acknowledgments
Thank you to Mr. Roberto Morales for his support and guidance throughout the research process. I am also grateful to IJHSR for allowing me to share my research with others.

References

Author
Emily Wang is a Clear Creek High School senior in League City, Texas. She is passionate about public policy and plans to pursue a career in law.

Appendix A: Full Survey

For the following questions, please select the numerical value that best aligns with your emotional response to each statement within quotations, with 1 meaning you strongly disagree with the statement and 7 meaning you strongly agree with the statement. If you are not sure, select the answer choice labeled “Not sure.”

1. Believe organ donation to be a rewarding experience.
   1 2 3 4 5 6 7
   Not sure

2. I am willing to be a registered organ donor.
   1 2 3 4 5 6 7
   Not sure

3. I feel a responsibility to be a registered organ donor.
   1 2 3 4 5 6 7
   Not sure

4. I am knowledgeable about organ donation.
   1 2 3 4 5 6 7
   Not sure

DOI: 10.36833/v5i4.8
My views on organ donation have been influenced by my ethnic background.*

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not sure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

My views on organ donation align with those common in my ethnic background.*

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not sure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

My views on organ donation have been influenced by the media I consume.*

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not sure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The current United States organ donation policy can be categorized as an opt-in organ donation policy. This means that individuals must sign up to become organ donors. However, many countries around the world have adopted different policies to increase their organ donation rates. An opt-out policy has been implemented in several European countries, such as Belgium. This policy automatically enrolls individuals as organ donors at birth.

Under an opt-out policy, I would be comfortable being automatically enrolled as an organ donor at birth.*

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not sure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An opt-out policy would be more effective in increasing organ donation rates in the United States than an opt-in policy.*

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not sure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An opt-out policy would be unethical.*

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not sure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Another potential organ donation policy entails an incentive system. Essentially, the organ donor or the family of the organ donor would be offered a reward. One possible option is monetary compensation that can be given to the family of the deceased. However, other options do not necessarily have to be financial. For example, blood donors frequently receive medallions or are recognized for their service.

Under a monetary incentive system, I would be comfortable with my family receiving compensation on my behalf.*

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not sure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Under a monetary incentive system, I would be more likely to register as an organ donor.*

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not sure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Under a non-monetary incentive system, I would be more likely to register as an organ donor.*

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not sure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An overall incentive policy would be more effective in increasing organ donation rates in the United States than an opt-in policy.*

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not sure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An incentive policy would be unethical.*

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not sure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A Novel Rapid Detection Method of *Escherichia coli* using PCR-based Nucleic Acid Amplification Using pH-sensitive Dyes

Eugene Shim

Dulwich College Seoul, 6 Sinbanpo-ro 15-gil, Banpo 2(i)-dong, Seocho-gu, Seoul, South Korea; gosyber1@gmail.com

Mentors: Prof. Woo Rin Lee

**ABSTRACT:** Molecular diagnostic assays based on nucleic acid amplification are the primary choice for diagnosing *Escherichia coli*. However, it requires extended workflow time (more than 10 hours), sophisticated equipment, or both to detect and analyze the amplification product. Here, a novel method of amplification detection was developed that mediates the pH change resulting from amplification reactions performed with minimal buffering capacity. This novel method allows one to easily check the positive infection state of *Escherichia coli* by changing the color of the amplified DNA product, which mediates the low pH of the reaction buffer. Colony Polymerase Chain Reaction (PCR), a widely used method to determine the presence or absence of bacteria DNA, was performed to amplify the 16s rRNA sequence. In this paper, experiments were conducted several times to find the optimal conditions for the color change in the DNA-amplified reaction buffer. Firstly, 47 °C, 50 °C, and 53 °C annealing temperatures were tested, and all three conditions successfully amplified the target bands. Next, phenol red concentration was optimized with 0.35 mM condition among five different conditions tested: 0.14 mM, 0.35 mM, 0.71 mM, 1.07 mM. Also, various concentrations of HCl were added to the PCR reaction solution, and the addition of 0.6 µM HCl showed the optimal pH condition for color change detection. Using the optimal conditions, we successfully amplified the *Escherichia coli* DNA with the transparent color of the PCR reactions. This colorimetric detection of amplification demonstrates a generally applicable approach for visual detection of nucleic acid amplification, which enables molecular diagnostic assays to be analyzed rapidly in 2 hours without the need for expensive lab equipment (agarose gel electrophoresis, gel documentation, and real-time PCR thermocycler). This novel method is also possible to be applied in diagnosing coronavirus.

**KEYWORDS:** Cellular and Molecular Biology, pH, Color change; *E. coli*; PCR; DNA amplification.

**Introduction**

*Escherichia coli* (*E. coli*) is a widely known gram-negative bacillus being a crucial member of the normal intestinal microflora of mammals, including humans.¹ *E. coli* causes various human diseases, such as bloody diarrhea, seizures, and dysentery.² These are enteric diseases caused by at least six different pathotypes. Further pathotypes can possibly cause extra-intestinal infections such as urinary tract infections and meningitis.³ Therefore, rapid diagnosis of *E. coli* infection is crucial.

One method to identify *E. coli* is utilizing Polymerase chain reaction (PCR) with primers against the uidA gene encoding beta-D-glucuronidase from food samples.⁴ Overall, this method requires a colony of bacteria on an agar plate, which takes 24 h. Another method to identify *E. coli* is Enzyme-Linked Immunosorbent Assay (ELISA).⁵ This method detects the protein of *E. coli* and takes about 24 hours to pre-enrichment the culture samples. The two methods can be merged to produce even more robust results.⁶ PCR is an excellent method of *E. coli* detection compared to standard culture and ELISA methods being extremely sensitive as well as accurate and promising real-time quantitative.⁷

We developed a new PCR-based nucleic acid amplification in this study using pH-sensitive dyes. This method only requires performing the PCR reaction with the samples. When the *E. coli* is presented in the sample, then the PCR reaction buffer decreases the pH by amplifying *E. coli* DNA. This new method can save time because DNA amplification can be detected by the color change of the PCR reaction buffer (red to orange). Through this study, we optimized the PCR conditions, such as annealing temperature, initial pH, and the concentration of phenol red (pH-sensitive dye) that detects the *E. coli* DNA by changing the color of the PCR reaction buffer.

**Methods**

*Polymerase Chain Reaction (PCR)*

Firstly, a mixed solution 11 times larger than 20 µl, which is the amount needed per one tube for the PCR, was prepared. Ten solutions of 20 µl in total to run five different amounts of phenol red indicator for bacteria-containing and not containing conditions were prepared. The following PCR reaction condition was prepared: Water (15.5 µl), 10x buffer (2 µl), dNTP (1 µl), F primer (0.5 µl), R primer (0.5 µl), Enzyme (polymerase) (0.5 µl), *E. coli* bacteria (0.5 µl). For optimizing phenol red concentration, various volume of phenol red indicator (14.1 mM) was added to the PCR reaction buffer: 2µl, 1.5µl, 1µl, 0.5µl, and 0.2µl. Start the PCR program by denaturation at 95°C for 3 minutes, followed by a series of cycles of
denaturation at 95°C for 30 seconds, annealing at 50 °C for 30 seconds, and extension at 72°C for 40 seconds. The number of cycles was set to 35 cycles.

Optimizing pH condition for PCR

A mix solution which was 11 times larger than the original solution, was made. Then using pipettes, this solution was divided into ten solutions. Then to each solution, different amounts of concentrated hydrochloric acid were added to alter the solution’s acidity to easily detect color change after the PCR within the range of 0 µl to 1 µl. The following PCR reaction buffer was prepared: Phenol Red 0.5 µl, 10x buffer 2 µl, dNTP 1 µl, and water 17 µl. The solutions were mixed well, so there was no material lying on the bottom that was not getting thoroughly mixed.

Agarose Gel electrophoresis

First, the PCR samples with or without E. coli samples were prepared. In the gel tank, the agarose gel was ready to be poured into the TAE buffer in sinking the gel. Then to each well, the samples were inserted to visualize the band lengths. Then it was turned on for electricity, giving charge to the end so the DNA bands could move. It was left on for 20 minutes to move DNA bands from negative to positive charge flow. Smaller DNA travels faster; positioning on top and thicker DNA means it contains more DNA. Then agarose gel with DNA was placed under the UV light to check the results.

Results and Discussion

This experiment aims to find the optimal annealing temperature for PCR reaction. We tested three different annealing temperature conditions for PCR reaction: 47 °C, 50 °C, and 53 °C. Using agarose electrophoresis gel, we visualized the amplified DNA band in the blue light illuminator. All three temperatures amplified the E. coli DNA (Figure 1). The expected band size was about 900 bp, and the size of the DNA amplified product was confirmed with the DNA size marker (Figure 1). In conclusion, we used an annealing temperature of 50 °C for the PCR reaction.

![Figure 1: Optimization of annealing temperature for amplification of E. coli 16S rRNA gene. The agarose gel image of DNA amplified products using three different annealing temperature conditions: 47 °C, 50 °C, and 53 °C.](image)

![Figure 2: Optimization of phenol red concentration on PCR reaction buffer for inducing color change after DNA amplification. The image of the final PCR reaction buffer with different concentrations of phenol red after PCR reaction. Phenol red concentrations of 0.14 mM, 0.35 mM, 0.71 mM, 1.07 mM, and 1.41 mM were used in this experiment.](image)

![Figure 3: Ten HCl concentrations were added to each sample to find the range of color changes in the PCR reaction solution. The image portrays samples with the addition of HCl concentrations of 0.05 µM, 0.1 µM, 0.2 µM, 0.3 µM, 0.4 µM, 0.5 µM, 0.6 µM, 0.7 µM, 0.8 µM, and 0.9 µM.](image)

Phenol red is a widely used pH indicator because it is easy to use and can measure a wide range of pH values. It changes color depending on the pH of the solution, making it easy to determine the pH visually. It is also relatively inexpensive, stable, and non-toxic, making it a good choice for laboratory and industrial settings. Additionally, it can be used in liquid or solid form, making it versatile for different measurements. Since phenol red’s versatility, stability, and a wide range of pH measurements, we selected phenol red as a pH indicator.

Next, we aimed to optimize phenol red concentration showing the color change after proceeding with the PCR reaction under bacteria-containing conditions. Therefore, we hypothesized that comparing the negative and positive samples of bacteria would vividly portray apparent color changes in the PCR reaction buffer. Phenol red concentrations of 0.14 mM, 0.35 mM, 0.71 mM, 1.07 mM, and 1.41 mM were used in this experiment. We prepared five negative control samples without bacteria and five samples with bacteria. The color of the samples without the bacteria and with the bacteria showed no difference in all five different concentrations of phenol red (Figure 2). However, as the phenol red concentration increased, the color of the PCR reaction sample became darker red. The phenol red concentration is critical for transforming the color after the PCR reaction. Since the DNA amplification in PCR reaction decreases the pH of the samples slightly, a high concentration of phenol red may reduce the sensitivity of the color change. The concentration of 0.14 mM showed pale pink color, and 0.71 mM showed a dark red color (Figure 2). Therefore, we chose 0.35 mM phenol red concentration as the optimal concentration and fixed this condition in the following experiments. This result shows that we failed to induce the color change under a bacteria-containing sample.

![Table: Phenol Red Concentrations](image)
The intended purpose of this experiment was to deduce the range of pH conditions that allows clear visualization of the color changes. By adding different concentrations of HCl, a strong acid, the samples’ initial pH conditions were decreased, indicating a more precise color range. Ten various HCl concentrations, including 0 µM, 0.1 µM, 0.2 µM, 0.3 µM, 0.4 µM, 0.5 µM, 0.6 µM, 0.7 µM, 0.8 µM and 0.9 µM were utilized to evoke the range of color change. As the added HCl concentration increases, the samples become more acidic, displaying brighter shades (Figure 3). For example, HCl concentrations of 0 µM to 0.2 µM showed dark red, 0.3 µM to 0.5 µM showed orange-red, 0.6 µM showed orange, and 0.7 µM to 0.9 µM showed yellow shades of the sample. In conclusion, we deduced that 0.5 µM - 0.6 µM and 0.6 µM - 0.7 µM concentrations display to most visible color changes, successfully concluding the optimal concentrations needed for the PCR reactions.

Figure 4: The PCR reaction sample with 0.6 µM HCl showed an apparent color change from red to orange. A total of four samples with or without bacteria supplemented with 0.5 µM and 0.6 µM HCl were prepared, and PCR was conducted. Then the PCR reaction tube image and the gel electrophoresis displaying the amplified bands of E. coli 16s rRNA are presented. The two bands on each side represent the DNA size marker indicating different lengths of the DNA bands.

This experiment aimed to display clear amplified bands of E. coli 16s rRNA from gel electrophoresis and induce color changes from pH change. Gel electrophoresis was performed by loading four different conditioned PCR reaction samples with either 0.5 µM or 0.6 µM HCl and with or without E. coli. DNA size markers were loaded on each side of the gel representing DNA band lengths comparison. After approximately 30 minutes, the gel was put under blue light to display the bands clearly. As expected, the two samples without bacteria did not show amplified DNA bands, meaning the E. coli 16s rRNA was not detected (Figure 4). The two samples containing bacteria, with 0.5 µM and 0.6 µM HCl, indicated 900 bp of the amplified DNA (Figure 4). In the 0.5 µM HCl condition, there was not an apparent color change shown, however in the 0.6 µM HCl condition, a color change from pink to orange was displayed, meaning 0.6 µM HCl is the optimized condition for detecting color change due to low pH produced by DNA amplification. However, since phenol red and HCl were added to the PCR reaction buffer solution, DNA amplification was not efficiently conducted, showing the lower intensity of DNA bands compared to Figure 1. In conclusion, the PCR reaction successfully induced pH color changes showing whether it contains E. coli bacteria.

■ Conclusion
In conclusion, our hypothesized predictions were in line with the concluded results. The annealing temperatures of 47 °C, 50 °C, and 53 °C were utilized successfully, amplifying the target bands and displaying an optimal condition for color change in the DNA amplified reaction buffer. The 0.35 mM phenol red concentration was the optimized concentration allowing the apparent color change after the PCR reaction. The optimized pH condition of the HCl concentration added was 0.6 µM HCl. Using such optimal conditions, E. coli DNA was successfully amplified, enabling a practical approach for visual detection and the further potential of this method.

Acknowledgments
I want to thank Prof. Woo Rin Lee for his guidance and advice in conducting this research.

References

Author
Eugene Shim is a rising senior at Dulwich College Seoul. She is interested in studying biochemistry, especially nanotechnology and genetics. She hopes this research paper will provide her with wider perspectives as well as further sophisticated knowledge in her interested area in the near future for her career path.
Caffeine Induces Adenosine A2A Receptor Triggering Brain Cell Damage in A172 Brain Cells

Hyunmin Lee
The Hockaday School, 11600 Welch Rd, Dallas, TX 75229, USA; gosyber1@gmail.com
Mentors: Prof. Lee

ABSTRACT: The Adenosine A2A Receptor, ADORA2A, is a protein-coding gene that plays a crucial role in neuromodulatory functions. High expression of the A2A receptor in human brain cells is one of the known stress biomarkers in the human brain. Furthermore, a high dose of caffeine can cause brain diseases like stroke and dementia. However, the effect of a high dose of caffeine on ADORA2A gene expression in human brain cells has never been investigated. Therefore, we hypothesized that caffeine might induce stress through an increased intracellular H$_2$O$_2$ level, which overexpresses ADORA2A in brain cells. First, when A172 human brain cells were incubated with various concentrations of caffeine, the caffeine-treated A172 cells showed a significantly higher ADORA2A expression level than the no caffeine-treated control sample. Second, when A172 human brain cells were treated with 0 µM caffeine and 20 µM caffeine, the 20 µM caffeine sample had a 3-fold increase in its H$_2$O$_2$ levels, demonstrating that caffeine increases intracellular H$_2$O$_2$ levels in A172 cells. Third, 30 µM and 50 µM of H$_2$O$_2$ treatment significantly increased the ADORA2A expression level, indicating that a higher H$_2$O$_2$ concentration leads to a higher ADORA2A expression level. Overall, we found that high doses of caffeine significantly augmented the ADORA2A expression levels through increased intracellular H$_2$O$_2$ in human brain cells. The application of this novel molecular mechanism may contribute to the development of therapeutic drugs for caffeine-induced stress in human brain cells.

KEYWORDS: Biology; Neuroscience; Caffeine; Brain Damage; ADORA2A receptor.

Introduction
ADORA2A encodes Adenosine A2A Receptor, which plays a crucial role in neuromodulatory functions and facilitates neurotransmissions. When experiencing an increased workload or a significant amount of stress, the body uses a greater amount of intracellular ATP, increasing firing rate conditions and thus selectively activating A2A receptors. Furthermore, ATP also acts as a danger signal in the brain, prompting changes in neuronal circuits in which more extracellular ATP is produced. The high concentration of synaptic extracellular adenosine selectively activates Adenosine A2A Receptor. Then, it promotes synaptic plastic changes in neuronal circuits that contribute to neurotoxicity in the brain.

Pharmacological and genetic evidence demonstrates that when suffering from an acute brain injury or chronic brain disease, the overactivation of Adenosine A2A Receptors can trigger brain dysfunction. Therefore, Adenosine A2A Receptor antagonists showed a neuroprotective effect in different brain damage models, including Alzheimer’s disease, ischemic brain stroke, and Parkinson’s disease. While many people nowadays drink coffee to boost their alertness and enhance brain function, caffeine can actually cause brain damage. According to a study at UniSA’s Australian Centre for Precision Health, there was a significant association between reduced brain volume and the amount of coffee consumed. Furthermore, excessive drinking of coffee has been shown to put people at risk of various brain diseases, including stroke and dementia. In addition to these long-term damages, caffeine can also reduce cerebral blood flow and induce brain hypoperfusion, thus posing short-term and long-term risks to its consumers.

On the contrary, other studies suggest that drinking caffeine can protect the brain from the risk of brain diseases. In one study, positron emission technology (PET) scans demonstrated that regular consumption of coffee led to adenosine receptors being blocked by caffeine. If this blockage persists over time, it can lead to adaptations in the brain, which change the receptor’s availability. This chronic shift can lead to a decreased risk of brain diseases like dementia. Overall, caffeine can have both positive and negative effects on the brain depending on various factors, including the amount consumed. Thus, in this study, we aimed to focus on how high dosages of caffeine would impact the brain.

Previous research indicated that caffeine induces hydrogen peroxide in high dosages. Therefore, in this study, we hypothesized that caffeine would induce intracellular stress in human brain cells through an increased intracellular H$_2$O$_2$ level, which may increase ADORA2A expression. Since ADORA2A is a well-known brain cell stress maker, we focused on the ADORA2A expression level as an indicator of caffeine-induced stress in human brain cells.

Methods
Cell culture and maintenance:
A172 human brain cells were purchased from Korea Cell Line Bank. The cells were maintained in the RPMI1640 cell culture media (Welgene) supplemented with 5% pen strep and...
10% fetal bovine serum. The cells were kept healthy under a 37 °C cell incubator with 5% CO₂.

**RNA extraction:**
The total RNA was extracted from the A172 cells using AccuPrep® Universal RNA Extraction Kit (Bioneer), followed by the manufacturer’s provided instructions. 50 µL eluent was used for the downstream experiments.

**Agarose gel-based RT-PCR:**
After cDNA was synthesized using the extracted RNA by the TOPscript™ Reverse Transcriptase kit (Enzynomics), a 20 µL PCR reaction was performed to amplify ADORA2A and GAPDH, which is used for normalizing ADORA2A expression level. The following PCR reaction was used to synthesize both genes. For annealing temperature, 62 °C was used, and the extension time was set to be 20 sec. A total of 30 cycles were run using a Thermocycler machine (Bioneer).

**Intracellular H₂O₂ staining:**
The intracellular H₂O₂ staining solution (Invitrogen) was used to stain H₂O₂ in green fluorescence. DAPI staining solution was used to stain the nucleus in blue fluorescence. After the treatment, each sample was incubated with a one-hour staining solution inside the 5% CO₂ cell incubator. ImageJ was used to quantify the H₂O₂ level. The “Threshold” tool was used to identify and select the fluorescence signal in the image. This tool sets a threshold intensity level and will highlight all pixels in the image that exceed this intensity. "Analyze Particles" was used to measure the size and intensity of the fluorescence signal. This tool generated a table of measurements, including the area, mean intensity, and integrated density of each fluorescence signal. Lastly, the "Measure" tool was used to manually measure the fluorescence intensity at specific locations in the image.

**Statistical analysis:**
One-way ANOVA with Tukey’s post hoc test was used to calculate the p-value.

### Results and Discussion

![Image](image1.png)

**Figure 1:** Caffeine increased ADORA2A mRNA expression level in A172 human brain cells. (A) Agarose gel image representing the amplified band ADORA2A and GAPDH cDNA. (B) Bar graph illustrating the mean and standard deviation of normalized ADORA2A mRNA expression level. GAPDH expression level was used to normalize ADORA2A expression level (n = 2). One-way ANOVA with Tukey’s post hoc test was used to calculate the p-value.

Through this first experiment, we aimed to investigate the correlation between caffeine and ADORA2A mRNA expression levels in human brain cells. After incubating A172 brain cells for five days with cell culture media, we added caffeine solutions of different concentrations to five samples: 0, 20, 100, 500, 750, and 1000 µM. After performing RT-PCR on these solutions, we then amplified cDNA samples and loaded them in agarose gel, allowing us to analyze the respective band intensities. No discernible band appeared for the sample with 0 µM caffeine (Figure 1A). On the other hand, an evident band of amplified ADORA2A was shown for the samples treated with caffeine, demonstrating that caffeine treatment increased the ADORA2A expression level. Furthermore, all caffeine-treated cells showed a significant increase in their ADORA2A expression level compared to the control sample with 0 µM caffeine (Figure 1B). Thus, the experiment results demonstrated that caffeine increases ADORA2A mRNA expression levels in A172 human brain cells.

![Image](image2.png)

**Figure 2:** Caffeine increased the intracellular H₂O₂ in A172 human brain cells. (A) Fluorescence images indicating nucleus (blue) and intracellular H₂O₂ (green) in A172 cells (B) Bar graph representing mean and standard deviation of intracellular H₂O₂ level in A172 cells. (n = 2). One-way ANOVA with Tukey’s post hoc test was used to calculate the p-value.

Through this experiment, we sought to test whether caffeine affects intracellular H₂O₂. We tested two conditions: 0 µM caffeine and 20 µM caffeine, incubating them for 48 hours with A172 brain cells. After the experiment, we checked the intracellular H₂O₂ levels through the green fluorescence representing them. The samples treated with 20 µM caffeine showed stronger fluorescent signals when compared to the 0 µM caffeine treated samples (Figure 2A). Moreover, quantifying the green fluorescence levels revealed that the H₂O₂ levels had a 3-fold increase in the 20 µM caffeine-treated sample (Figure 2B). This result indicates that caffeine increased intracellular H₂O₂ levels in A172 cells.

![Image](image3.png)

**Figure 3:** H₂O₂ increased ADORA2A mRNA expression level in A172 human brain cells. (A) Agarose gel image representing the amplified band ADORA2A and GAPDH cDNA. (B) Bar graph illustrating the mean and standard deviation of normalized ADORA2A mRNA expression level. GAPDH expression level was used to normalize ADORA2A expression level (n = 2). One-way ANOVA with Tukey’s post hoc test was used to calculate the p-value.
Based on the results of the previous experiments, which demonstrated a positive correlation of caffeine with H$_2$O$_2$ and ADORA2A, we aimed to investigate caffeine’s effect on ADORA2A expression levels. Therefore, we hypothesized that caffeine might increase the ADORA2A expression level. After incubating A172 brain cells with 0, 5, 10, 30, and 50 µM H$_2$O$_2$, we used agarose gel and its band intensities to quantify the level of ADORA2A in each sample. While the A172 brain cells with H$_2$O$_2$ 0, 5, and 10 µM did not significantly affect the ADORA2A band intensity, the cells incubated with H$_2$O$_2$ concentrations of 30 and 50 µM significantly increased the band intensity of ADORA2A (Figure 3A). Furthermore, the cells with H$_2$O$_2$ concentrations of 50 µM had a normalized ADORA2A expression level approximately nine times greater than that of the 0 µM H$_2$O$_2$ cells, demonstrating the positive correlation between H$_2$O$_2$ concentration and ADORA2A expression level (Figure 3B). Thus, the results of this experiment indicate that a higher level of intracellular H$_2$O$_2$ increases ADORA2A expression levels in A172 human brain cells.

Figure 4: Discovery of a novel pathway for how caffeine overdose damages the human brain. When caffeine overdose induces intracellular H$_2$O$_2$ stress, ADORA2A expression is increased, and finally, brain stress damage is induced.

Based on the first experiment, caffeine caused an increase in ADORA2A mRNA expression levels in A172 human brain cells. Furthermore, caffeine also increased the intracellular H$_2$O$_2$ levels in the cells. Results from a subsequent experiment indicated that higher levels of intracellular H$_2$O$_2$ levels increase ADORA2A expression levels in A172 human brain cells. Overall, this study demonstrated a positive correlation between caffeine, intracellular H$_2$O$_2$ levels, and ADORA2A expression levels in human brain cells (Figure 4).

Conclusion
Though our study demonstrates the effects of caffeine on A172 human brain cells, we must validate this by testing different human brain cells and ensuring that our conclusion is not uniquely applicable to A172 brain cells. Conducting additional tests to uncover the detailed molecular mechanism of caffeine inducing such changes will also allow us to further our finding that caffeine increases intracellular H$_2$O$_2$ levels. Additionally, based on our finding that H$_2$O$_2$ levels increased the ADORA2A expression level, we expect that future studies will allow us to find the H$_2$O$_2$-dependent transcription factors that may have directly induced the change in ADORA2A expression levels.

With nine in ten Americans reporting consumption of caffeine and three in four having caffeine at least once a day, it is evident that many Americans consume caffeine regularly. However, one in four consumes it three or more times a day, perhaps going beyond healthy consumption and raising concerns about caffeine safety and overdose. Despite our familiarity with caffeine, there is not enough awareness of these concerns, and studies on the effects of caffeine on brain cells are minimal. Our study found that a caffeine overdose can induce ADORA2A, a brain stress marker, indicating that a caffeine overdose may damage the human brain. In addition to this finding, we also discovered details on how caffeine may damage the brain cells; caffeine increases ADORA2A expression through increased intracellular H$_2$O$_2$ levels. Since this is the first study of discovering a concrete, detailed mechanism regarding how caffeine can damage the brain, it expands opportunities for further study of caffeine in the future. Our discovery of this mechanism involving H$_2$O$_2$ will facilitate efforts to develop novel therapeutic drugs, allowing not only the awareness of how caffeine overdose tangibly impacts our brain but also enabling the minimization of future risks.

Acknowledgments
I am extremely grateful to Prof. Lee from the University of Suwon for his support and assistance in performing experiments.

References


**Author**

Hyunmin Lee is a junior at The Hockaday School in Dallas, Texas. She is interested in studying neuroscience and its implications for understanding human psychology.
Time Evolution of the Mass of Primordial Black Holes in a Cosmic Microwave Background with a Time-Dependent Temperature

Jonathan Lee
St. Margaret’s Episcopal School, 31641 La Novia Ave, San Juan Capistrano, California, 92675, USA; jonathanlee@college124@gmail.com
Mentors: Rifath Khan

ABSTRACT: Primordial black holes have been long theorized to exist, with their formation predicted to have been a short time after the Big Bang. In this paper, we discuss the different properties of primordial black holes, such as the possibility of primordial black holes being the seeds for supermassive black holes that are present in the center of galaxies and primordial black holes being a candidate for dark matter. Then, we show how we derived an equation for the evolution of the mass of a primordial black hole in a cosmic microwave background with a time-dependent temperature. Also, we discuss techniques that we used to solve the equation we derived and how we calculated the mass range of a primordial black hole in the early universe. The results of these calculations show the possible masses of primordial black holes that could account for all of the dark matter in the universe. If primordial black holes could explain the dark matter problem, then it would help extend our understanding of the composition of the universe and the formation of the early universe.

KEYWORDS: Physics and Astronomy; Astronomy and Cosmology; Black Holes; Primordial Black Holes; Thermodynamics.

Introduction

The existence of primordial black holes has been a topic of interest for scientists since 1967 when Yakov Borisovich Zel’dovich and Igor Dmitriyevich Novikov first suggested the existence of these black holes and also when Stephen Hawking studied these black holes in 1971.¹ These primordial black holes are predicted to have formed near the beginning of the universe soon after the Big Bang. The formation of primordial black holes cannot be due to stellar collapse like the black holes known to exist today because primordial black holes would have formed long before the existence of the first stars. Instead, it is theorized to be a result of irregularities in the uniformity of the cosmic microwave background. If primordial black holes do, in fact, exist, their existence means that the universe was not distributed equally in terms of mass, as there must have been some areas in the early universe with a high enough density to cause gravitational collapse and the formation of a black hole. The fact that primordial black holes could not have formed from stellar collapse also means that the initial mass of a primordial black hole is not bound by the same limits of initial mass that a black hole formed from stellar collapse is. This makes primordial black holes a strong topic of research among scientists, as it would better explain our understanding of the early universe and the distribution of mass in the early universe; it may explain the formation of supermassive black holes and the properties of primordial black holes also make it a possible candidate for dark matter.

In this paper, I will be exploring the possible existence of primordial black holes, how they could have survived to the present day, and the possibility of primordial black holes being the dark matter that is theorized to exist in our universe. The methods section of the paper will be a basic review of black holes and the properties and formation of a black hole. I will also discuss black hole thermodynamics, Hawking radiation, and black holes as blackbodies. There will then be a review of the cosmic microwave background and its temperature over time. This will explain the temperature of the cosmic microwave background in the early universe. The final part of the methods section will be an overview of primordial black holes, supermassive black holes, the possible formation of supermassive black holes, and primordial black holes as a candidate for dark matter. In the results and discussion section, I will explain the methods used to find the balance between the evaporation rate and absorption rate of primordial black holes and how the initial mass range of primordial black holes can be found using those rates. The methods to find this initial mass range involve creating an equation that finds the mass of a black hole as a function of time by taking into account the temperature of the cosmic microwave background over time, the absorption rate of a black hole from the cosmic microwave background, and the evaporation rate of a black hole due to Hawking radiation. I will then cover how I went about numerically solving the differential equation for the mass of the black hole as a function of time. This section will also explain why the equation is hard to solve and where the use of approximations was needed. It will also discuss the results on the initial mass range of primordial black holes found using the program. Finally, the conclusion section will summarize our findings and then compare our solution to other research that has been done on primordial black holes. Then, I will discuss the similarities and differences between our results and other research findings and compare our methods.
Methods

Review of Black Holes:

According to Einstein's theory of General Relativity, a black hole is a region in spacetime where the force of gravity is too strong to allow anything to escape, including all matter and electromagnetic radiation. This is why black holes are completely invisible, as there is no light being reflected. The only way to detect black holes is by observing their gravitational effects on their surroundings. At the center of a black hole, there is a point where spacetime becomes infinitely curved, called a singularity. As the singularity is a singular point in spacetime, the volume of the singularity is zero, which means that the singularity also is a point of infinite density. Any object that falls into a black hole will fall towards this singularity without the possibility of escape. Black holes also contain an event horizon, which determines the boundary of a black hole. Once an object crosses the event horizon, it will not be able to escape and will inevitably fall toward the singularity. This also applies to any information from an event that occurs within a black hole, so any information that happens within the event horizon cannot be communicated to anyone outside of the black hole. The radius of a black hole, known as the Schwarzschild radius, can be determined as the distance from the singularity to the event horizon, which can be found with the equation

$$ r = \frac{2GM}{c^2}, \quad (1) $$

with $r$ being the radius of the black hole, $G$ being the gravitational constant, $M$ being the mass of a black hole, and $c$ being the speed of light.

The only way astrophysical black holes can form is through stellar collapse. This occurs at the end of a star's lifetime when the star's mass gets drawn inward towards the center of gravity. While in stars not large enough to create a black hole, the force of gravity is not strong enough to overcome degeneracy pressure resulting in the formation of a white dwarf or a neutron star. However, in stars large enough to create a black hole, in which their remnants are larger than the Tolman–Oppenheimer–Volkoff limit, the force of gravity overwhelms the degeneracy pressure, resulting in the formation of a black hole.

Black Hole Thermodynamics:

The mass of a black hole over time depends on the thermodynamics of a black hole. The way in which black holes gain mass is through gaining energy from the cosmic microwave background, which converts to mass with Albert Einstein's famous $E = Mc^2$ equation. As stated in the previous section, the absorption rate of a black hole can be found using the Stefan-Boltzmann law, but this law is also effective in determining the emission rate of a black hole. This is because the Stefan-Boltzmann law determines the amount of energy radiated from a black body using its temperature. Black holes are the perfect black body as they absorb all radiation that comes into contact with them. Using the Stefan-Boltzmann law, the equation for the absorption rate of a black hole can be given as

$$ P_{\text{absorption}} = 4\pi r^2 \sigma T_{\text{CMB}}^4, \quad (2) $$

where $P_{\text{absorption}}$ is the power absorbed by a black hole or the total amount of energy absorbed by a black hole in one second, $r$ is the radius of the black hole, $\sigma$ is the Stefan-Boltzmann constant, and $T_{\text{CMB}}$ is the temperature of the cosmic microwave background.

Black holes emit radiation through Hawking radiation, which Stephen Hawking first theorized in 1974. Hawking radiation happens when particles created from quantum fluctuations form near the event horizon of a black hole. While the particle and its antiparticle usually annihilate each other out of existence, when the particle pair forms near the event horizon of a black hole, sometimes one particle falls into the black hole. In contrast, the other is emitted out into the universe. However, the particle that falls into the black hole contains negative energy, which means the black hole will lose energy and mass from that particle, causing the black hole to evaporate slightly. To determine the amount of energy lost by a black hole due to Hawking radiation, the equation for the Stefan-Boltzmann law can be used again with the temperature of the black hole itself being used in the equation. The equation for the temperature of a black hole can be derived using the equation for the first law of thermodynamics,

$$ dE = dQ - dW, \quad (3) $$

with $dE$ being the change in energy of the black hole, $dQ$ being the change in the heat of the black hole, and $dW$ being the rate of work being done by the black hole. Since a black hole does no work on its surroundings, $dW$ can be assumed to be zero. This equation is combined with the equation for the second law of thermodynamics,

$$ dS = \frac{dQ}{T_{\text{BH}}}, \quad (4) $$

with $dS$ being the change in entropy of the black hole and $T_{\text{BH}}$ being the temperature of a black hole, to form the equation

$$ T_{\text{BH}} = \frac{dE}{dS}. \quad (5) $$

Then, Einstein's $E = Mc^2$ equation can be differentiated to get

$$ dE = dMc^2, \quad (6) $$

which can be inserted into the temperature equation found above to get

$$ T_{\text{BH}} = \frac{dM}{dS}c^2 = \frac{dS^{-1}}{dM}c^2. \quad (7) $$

$dS/dM$ can then be found by using the Bekenstein–Hawking formula for the entropy of a black hole,

$$ S = \frac{h \pi c^3 A}{4 \hbar G}, \quad (8) $$

and the formula for the area of the black hole,

$$ A = \frac{16\pi GM^2}{c^4}, \quad (9) $$

to find the entropy of a black hole in terms of mass,

$$ S = \frac{4\pi \hbar G M^2}{c^4}. \quad (10) $$
which can be differentiated to
\[ \frac{dS}{dM} = \frac{8\pi k_B G M}{\hbar c}. \] (11)

When this equation for \(dS/dM\) is inserted back into the equation found for the temperature of a black hole, the result is the equation
\[ T_{BH} = \frac{\hbar c}{8\pi G M k_B}. \] (12)

Using the Stefan-Boltzmann law, the equation for the emission rate of a black hole can be given as
\[ P_{\text{emission}} = 4\pi r^2 c T_{BH}^4, \] (13)
and \( T_{BH} \) can be inserted into the equation resulting in the equation
\[ P_{\text{emission}} = 4\pi r^2 c \left( \frac{\hbar c}{8\pi G M k_B} \right)^4. \] (14)

**Overview of the Cosmic Microwave Background and its Temperature Over Time:**

The cosmic microwave background is the radiation left over from the Big Bang and its temperature can be used to determine the universe’s temperature on a large scale. This is important because the cosmic microwave background’s temperature determines a black hole’s absorption rate with the Stefan-Boltzmann law. While the temperature of the cosmic microwave background is currently around 2.73 Kelvin, it has been decreasing ever since the Big Bang. It will eventually approach absolute zero at some point in the far future.³ The reason for this decrease in the temperature of the cosmic microwave background is that the universe’s expansion increases the wavelengths of radiation in the universe, causing a reduction in the radiation’s energy, meaning the radiation’s temperature is cooling down. Using the Friedmann equations, the temperature of the cosmic microwave background has decreased at different rates during different epochs. During the radiation-dominated era, the cosmic microwave background temperature was proportional to \( t^{-1/2} \). In contrast, during the matter-dominated era, the temperature of the cosmic microwave background was proportional to \( t^{-2/3} \).³ Currently, the universe is in the dark energy-dominated era where the temperature of the cosmic microwave background is proportional to \( e^{-t \cdot (Hubble \ constant \times t)} \).³ Using these rates, the temperature of the cosmic microwave background can be determined at different points in time, which can then be used to determine the absorption rate of a black hole given its mass and point of time in the universe.

**Primordial Black Hole Formation:**

Primordial black holes could have been formed in the early universe during the radiation-dominated era, where the radiation density was higher than the density of matter in the universe. While the radiation-dominated era lasted until the universe was around 47,000 years old, the formation of primordial black holes is thought to have happened less than a second after the Big Bang.³ This means that the formation of primordial black holes happened before the formation of the first stars in the universe, showing that the formation of primordial black holes could not have been due to stellar collapse. Therefore, primordial black holes would have formed from irregularities in the density of the early universe that were at a high enough density to cause gravitational collapse. Since primordial black holes could not have been formed by stellar collapse, the lower bound of the initial mass range of primordial black holes is not restricted by the Tolman-Oppenheimer-Volkoff limit. Black holes formed from stellar collapse have a lower bound on initial mass with the Tolman-Oppenheimer-Volkoff limit of around 1.5 to 3.0 solar masses.⁷ Any star remnant with a mass below that limit will collapse into a neutron star or a white dwarf rather than a black hole. However, primordial black holes can form with an initial mass below that limit, meaning that if a black hole was found with a mass far below the Tolman–Oppenheimer–Volkoff limit, there is a strong possibility of the black hole being a primordial black hole, since there has not been enough time since the early universe for a black hole formed from stellar collapse to evaporate to that small of a mass from Hawking radiation. Even with this expanded mass range, there is still a lower bound on the initial mass of a primordial black hole that has survived to the present day, around \(10^{11} \) kg. This is because any primordial black hole with an initial mass lower than \(10^{11} \) kg would have completely evaporated before the present day due to Hawking radiation.

**Supermassive Black Holes:**

Supermassive black holes are black holes that are defined to be between 0.1 million and 10 billion solar masses.⁸ Most large galaxies in our universe are believed to contain supermassive black holes at their centers, but their origin is unknown. One possibility is that primordial black holes could be seeds for supermassive black holes. The possibility of supermassive black holes forming from primordial black holes is more likely than the possibility of supermassive black holes forming from black holes that were formed from stellar collapse and grew from accretion. This is because primordial black holes would have had more time to grow by accretion, as they are predicted to have existed long before the existence of the first stars.⁹

**Primordial Black Holes as Dark Matter:**

The possibility of primordial black holes being a candidate for dark matter has been explored in many different studies, even though there has not been any evidence showing that this is the case. This is due to the properties of primordial black holes that show that it is entirely plausible that primordial black holes account for some, if not all, of the dark matter in the universe. One of the main properties that make primordial black holes a good candidate for dark matter is that primordial black holes are non-baryonic matter. Baryonic matter is matter made up of baryons, particles that are made up of an odd number of quarks, like protons and neutrons, making all atoms baryonic matter. Because primordial black holes formed during the radiation-dominated era, before Big Bang nucleosynthesis formed the baryonic matter in the universe, primordial black holes must be considered non-baryonic matter.¹⁰ Most of the universe’s dark matter is considered non-baryonic because it has not been directly observed and barely interacts with baryonic matter, making primordial black holes an entirely plausible candidate for dark matter. Primordial black holes are
that would explain observations in galaxy halos that suggest the presence of dark matter. However, many candidates for massive compact halo objects are composed of baryonic matter, making primordial black holes an interesting candidate as they are non-baryonic. The possibility of primordial black holes being all of dark matter was explored in a 2019 study, which concluded that there is still a possibility of primordial black holes in the mass range of $3.5\times10^{17}$ solar masses to $4\times10^{18}$ solar masses or $7.0\times10^{33}$ kg to $8\times10^{38}$ kg making up all of the dark matter in the universe. Because of this possibility, this is the mass range we will be using for the current mass of primordial black holes in my calculations of finding the mass of primordial black holes as a function of time.

**Results and Discussion**

**Evolution Equation of the Mass of a Black Hole:**

To find the evolution equation for the mass of a primordial black hole, the energy that the primordial black hole gains and loses over time must be found as that energy is directly converted to mass using Einstein’s $E = Mc^2$ equation, which can be converted to

$$M = \frac{E}{c^2}.$$  

(15)

Since the evolution equation is finding the mass of a primordial black hole as a function of time, the equation can be differentiated in terms of time to

$$\frac{dM}{dt} = \frac{dE}{dt} \cdot \frac{1}{c^2}.$$  

(16)

Power ($P$) is the derivative of energy with respect to time, so $P$ can be substituted to $dE/dt$ making the equation able to be written as

$$\frac{dM}{dt} = \frac{P}{c^2}.$$  

(17)

The power of a black hole is equal to the power that the black hole is absorbing minus the power that the black hole is emitting, which can be written as

$$P = P_{\text{Absorption}} - P_{\text{Emission}}.$$  

(18)

and using the equation for $P_{\text{Absorption}}$ and the equation for $P_{\text{Emission}}$, the power equation can be written as

$$P = 4\pi r^3 \sigma T_{\text{CMB}}^4 - 4\pi r^3 \sigma \left( \frac{\hbar c^2}{8\pi GM_{\text{BH}}} \right)^{\frac{4}{3}}.$$  

(19)

$P$ can then be inserted back into the equation, finding the change in mass of a black hole as a function of time, resulting in the equation

$$\frac{dM}{dt} = \frac{4\pi r^3 \sigma T_{\text{CMB}}^4 - 4\pi r^3 \sigma \left( \frac{\hbar c^2}{8\pi GM_{\text{BH}}} \right)^{\frac{4}{3}}}{c^2},$$  

(20)

or

$$\frac{dM}{dt} = \frac{16\pi G^2 \sigma}{c^2} M^2 T_{\text{CMB}}^4 - \frac{25\pi \sigma T_{\text{CMB}}^4}{M^2}.$$  

(21)

The integrated equation ended up becoming

$$M_f = \int_{M_i}^{M_f} \left[ \frac{2\pi}{3} \frac{H_{\text{CMB}}}{c} \log \left( \frac{1 + \frac{H_{\text{CMB}}}{c}}{1 + \frac{H_{\text{CMB}}}{c}} \right) \right] M_f^3 = \Delta t,$$  

(25)

where $M_f$ is the mass of the primordial black hole at the present day, $M_i$ is the mass of the primordial black hole at a given point in time, and $\Delta t$ is the time between the present day and the given point in time.

With the present-day mass range of primordial black holes already known from the methods section, the mass range during each era can be calculated using the integrated equation. However, programs such as Mathematica and MATLAB could not provide results due to the equation’s complexity, so we devised a way to solve the equation by splitting the equation

$$x_t = \frac{A^t M T_{\text{CMB}}^4}{B^t},$$  

(26)

into different parts. First, we set

$$x_t = A^t M T_{\text{CMB}}^4 = B^t T_{\text{CMB}}^4.$$  

(27)

Solving the Evolution Equation:

The equation found in the previous section must be integrated to find the mass of a primordial black hole when it formed in the early universe. To make the integration of this equation easier, the constants found in the equation can be combined to make the equation

$$A = \frac{16\pi G^2 \sigma}{c^2} \approx 1.74 \times 10^{-74},$$  

(22)

and

$$B = \frac{\sigma T_{\text{CMB}}^4}{25\pi GM_{\text{BH}}^2} \approx 3.92 \times 10^{19},$$  

(23)

to make the equation

$$\frac{dM}{dt} = AM^2 T_{\text{CMB}}^4 - B M^3.$$  

(24)

A variable in the equation is $T_{\text{CMB}}$, which decreases over time. However, since a function to determine $T_{\text{CMB}}$ over time is difficult to incorporate into the equation, we turned to splitting up the time axis into five different eras and treating $T_{\text{CMB}}$ as a constant within these different eras, with era 1 being the period where the age of the universe was 10-12 seconds to 10-6 seconds, era 2 being the period from 10-6 seconds to 3 minutes, era 3 being the period from 3 minutes to 300,000 years, era 4 being the period from 300,000 years to 10^9 years, and era 5 being the 10^9 years to present day, which is estimated to be around 10^10 years. Since the temperature of the cosmic microwave background is decreasing at a logarithmic rate, it can be assumed that the temperature during an era will mostly be a number near the temperature at the most recent point in an era. Therefore, the temperature of the cosmic microwave background at the most recent point in these eras can be used as the temperature for the entire era, which is given as $10^{13}$ K for era 1, $10^{10}$ K for era 2, 3,000 K for era 3, 10 K for era 4, and 3 K for era 5. The integration of this equation is still difficult, which is why the program Mathematica was used to integrate this equation.
resulting in the simplified version of the equation being

\[
x_i = \frac{A^2 M_i T_{\text{CMB}}}{B^1},
\]

and

\[
\left(2 \arctan(x_i)+\log\left(\frac{1-x_i}{1+x_i}\right)\right) - \left(2 \arctan(x_e)+\log\left(\frac{1-x_e}{1+x_e}\right)\right) = \Delta t(A^2 B^1 T_{\text{CMB}}^3).
\]

Then, \(x_i\) was found using MATLAB and then using Desmos to graph the function, \(2 \arctan(x_i)+\log\left(\frac{1-x_i}{1+x_i}\right)\), and a line for \(x = x_i\) a numerical value for that portion of the equation using the intersect point of the two lines. Next, a numerical value was found for \(\Delta t(A^2 B^1 T_{\text{CMB}}^3)\), using the number of seconds in the era and the temperature during the era, and then added onto the numerical value for \(2 \arctan(x_i)+\log\left(\frac{1-x_i}{1+x_i}\right)\) to find \(y_i\). Using Desmos, a line, \(y = y_i\) was then graphed along with the function, \(2 \arctan(x_i)+\log\left(\frac{1-x_i}{1+x_i}\right)\), to find \(x_i\), which is the \(x\) value at the intersection point. Finally, \(M_i\) is found using MATLAB using the equation for \(x_i\), finding the mass of the primordial black hole at the beginning of the era. This process is then repeated for each era and done for both the upper and lower bound of the mass range. The results were then compiled into two different tables with columns showing the different eras, the age of the universe at the beginning of each era, the temperature of the cosmic microwave background at that point in time, and the lower bound for the mass of a primordial black hole in Table 1 and the upper bound in Table 2.

### Table 1

<table>
<thead>
<tr>
<th>Era</th>
<th>Age of Universe</th>
<th>Temperature of CMB</th>
<th>Mass of Primordial Black Hole</th>
</tr>
</thead>
<tbody>
<tr>
<td>End of Era 5</td>
<td>10^10 years</td>
<td>3 K</td>
<td>(7.0000000000000000000010^3) (\text{kg})</td>
</tr>
<tr>
<td>Beginning of Era 5</td>
<td>10^9 years</td>
<td>10 K</td>
<td>(7.0000171463604040010^3) (\text{kg})</td>
</tr>
<tr>
<td>Beginning of Era 4</td>
<td>300,000 years</td>
<td>3000 K</td>
<td>(7.0000598014906910000010^3) (\text{kg})</td>
</tr>
<tr>
<td>Beginning of Era 3</td>
<td>3 minutes</td>
<td>10^10 K</td>
<td>(7.0001714636040560000010^3) (\text{kg})</td>
</tr>
<tr>
<td>Beginning of Era 2</td>
<td>10^9 seconds</td>
<td>10^10 K</td>
<td>(6.9999354879242937x10^3) (\text{kg})</td>
</tr>
<tr>
<td>Beginning of Era 1</td>
<td>10^12 seconds</td>
<td>10^10 K</td>
<td>(6.9999352879232600x10^3) (\text{kg})</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Era</th>
<th>Age of Universe</th>
<th>Temperature of CMB</th>
<th>Mass of Primordial Black Hole</th>
</tr>
</thead>
<tbody>
<tr>
<td>End of Era 5</td>
<td>10^10 years</td>
<td>3 K</td>
<td>(8.0000000000000000000010^4) (\text{kg})</td>
</tr>
<tr>
<td>Beginning of Era 5</td>
<td>10^9 years</td>
<td>10 K</td>
<td>(8.00013623518096798641x10^4) (\text{kg})</td>
</tr>
<tr>
<td>Beginning of Era 4</td>
<td>300,000 years</td>
<td>3000 K</td>
<td>(8.00013623518068879147x10^4) (\text{kg})</td>
</tr>
<tr>
<td>Beginning of Era 3</td>
<td>3 minutes</td>
<td>10^10 K</td>
<td>(8.00013623518068934411x10^4) (\text{kg})</td>
</tr>
<tr>
<td>Beginning of Era 2</td>
<td>10^9 seconds</td>
<td>10^10 K</td>
<td>(8.00013623518068344115x10^4) (\text{kg})</td>
</tr>
<tr>
<td>Beginning of Era 1</td>
<td>10^12 seconds</td>
<td>10^10 K</td>
<td>(8.00013623518066917869x10^4) (\text{kg})</td>
</tr>
</tbody>
</table>

**Acknowledgments**

I would like to thank my mentor Rifath Khan, a Ph.D. candidate for Applied Mathematics and Theoretical Physics at the University of Cambridge, for being a supervisor in developing this research paper.

**References**

3. Willis, J. The cosmic microwave background - UVIC.
11. Djorgovski, S. G. WIMPs and MACHOS.
Food Insecurity among US Households during the Covid-19 Pandemic: Prevalence and Inequality

Julia Yi

East Brunswick High School, 380 Cranbury Rd, East Brunswick, NJ 08816, USA; juliayithebest@gmail.com
Mentors: Mrs. Valerie Alimonti, Professor Robson

ABSTRACT: Using the Household Pulse Survey conducted by the U.S. Census Bureau from April 2020 to November 2021, this study investigates whether the Covid-19 pandemic escalated the prevalence and inequality of food insecurity in the United States. It finds that the pandemic statistically increased the prevalence of food insecurity – 21.86% and 6.01% of US households were either low or very low food secure, double the 2019 level. The top reason was the lack of affordability resulting from income loss and inflated food prices. Inequality of food insecurity was pronounced -- people of color, low-income households, households with income loss, and households with children were impacted most during the pandemic. Given that these household types had a disproportionately higher prevalence of food insecurity before the pandemic, the pandemic elevated the inequality of food insecurity even more. Among food-insecure households, 18.66% of households received free meals, and 29.66% received SNAP benefits. Food pantries or food banks provided most of the free groceries and meals. The government can use the findings, nonprofit organizations and community programs to identify and target households who need help to stay food secure – minority households, households that lost their income, households with children, and low-income households.

KEYWORDS: Behavioral and Social Sciences, Food Insecurity; Covid-19 Pandemic; Prevalence; Inequality; US Households.

Introduction

Since December 2019, the world has been facing a big challenge, the Covid-19 pandemic. Unlike other outbreaks in living memory, such as SARS, MERS, or Ebola, COVID-19 was a global disease with high transmissivity, quickly affecting poor, rich, rural, and industrialized countries. The World Health Organization (WHO) declared the Covid-19 outbreak a Public Health Emergency of International Concern (PHEIC) on January 30, 2020, and a pandemic on March 11, 2020.¹ ² As of Dec. 3, 2021, the Coronavirus Resource Center at Johns Hopkins University reported 264,834,612 cases and a death toll of 5,242,078 globally; the United States had seen 4,898,223 infections and 787,678 deaths. The New York Times reported that the highest peak for the number of infections, hospitalization, and death was in January 2021, and the second most prominent peak was in September 2021 in the United States due to the Delta variant.³ The socioeconomic impacts of this pandemic were comparable only to the crash of the Great Depression.

The frightening fast rate of spread led to a harsh response – a complete shutdown in major cities and expanded to many states in March-April 2020.³ Shelter at home orders were issued by local governments, enforcing curfews and social distancing. Nonessential businesses were shut down, while even essential ones, such as restaurants that provided food, were only allowed to ship takeout during the lockdown. Grade schools and colleges started online learning. By the end of 2020 and early 2021, all the countries began a race to vaccinate. Cases and deaths began to fall. However, the variants were still a threat. For example, by July 2021, the Delta variant was found in 65 countries, including the United States, which resulted in the second highest peak for infection cases, hospitalization, and death in the United States.³ In December 2021, the world had another spike infection cases due to the new variant, Omicron.

The Covid-19 pandemic brought economic, social, health, and emotional shocks to people regardless of their backgrounds and socioeconomic status. Food insecurity has been one of the most critical problems.⁴ Figure 1 the trend of food insecurity in US households from 1995 to 2020, distinguishing low food security and very low food security according to the USDA definitions. Low food-secure households have enough food to avoid substantially disrupting their eating patterns or reducing food intake by using various coping strategies, such as eating less varied diets, participating in Federal food assistance programs, or getting food from community food pantries. Among very low food-secure households, the standard eating patterns of one or more household members are disrupted, and food intake is reduced at times during the year because they have insufficient money or other resources for food. Figure 1 shows that before the pandemic, the United States had the lowest food insecurity prevalence – approximately 10.54% of households were low food secure, and 4.11% were very low food insecure in 2019.
The World Bank Report states that “COVID-19 impacts led to severe and widespread increases in global food insecurity, affecting vulnerable households in almost every country, with impacts expected to continue into 2022 and possibly beyond.” Many US households faced a significant challenge to staying food secure for different reasons. Due to staying at home for long periods, some people developed “fear from infection or fear from the virus.” People feared contracting Covid-19 and did not want to go out or were affected by lockdowns. Another reason could be that some households had sick family members and/or school children who did not go to school because of the pandemic, which also made people unable to get out of the house shopping for groceries. The pandemic also disrupted the U.S. food supply chain, and people could not get enough food or sufficient varieties at grocery stores. Food prices increased significantly, making households unable to purchase groceries and meals, especially low-income households and those whose family members either lost jobs or employment income. One of the most important reasons was that people lost jobs, had pay cuts, or had their businesses either temporarily or permanently closed. The income shock not only caused households with low food security before the pandemic to be trapped in severe food insecurity but also dragged households who were food secure before the pandemic into either low or very low food insecure status.

The objectives of this study are to investigate: 1) whether the pandemic increased food insecurity among US households, 2) the inequality of food security by socioeconomic profiles, and 3) channels from which food insecure households get help.

## Literature Review

Before the pandemic, about 10.5% of U.S. households were classified as low food secure, and 4.1% were very low food secure in 2019 (see Figure 1). Coleman-Jensen et al. (2020) show that food insecurity among US households in 2019 has the following patterns. First, the following households had a statistically higher rate of food insecurity than the national average: households with incomes below 185% of the poverty line (27.6%), households with children headed by a single woman (28.7%), and households with children (13.6%). Second, among all food-insecure households in 2019, 36.6% were households with children, 19.7% were female-headed households with children, and 13.5% were married-couple households with children. Third, food insecurity exhibits significant racial inequality – non-Hispanic black and Hispanic households were more likely to be food insecure (19.1% and 15.6%) than non-Hispanic white households (7.9%).

The pandemic caused economic and social shocks to US households. The pandemic elevated the prevalence of food insecurity, from 18% in late spring to 35% in summer of 2020. During the pandemic, especially during lockdowns, households were recommended to stock up on food and limit trips to the grocery store. Some households may lack financial resources and physical space to store food. The pandemic brought price spikes for almost all food categories, which caused negative impacts on food security. Lack of affordability was the single most significant barrier to staying food secure.

### Methods

This study uses data from the Household Pulse Survey (HPS) conducted by the U.S. Census Bureau at the outset of the pandemic. HPS is a 20-minute online survey focusing on how the Covid-19 Pandemic affected households throughout the United States economically, socially, and mentally. The survey’s first week lasted from April 23, 2020, to May 5, 2020. The first phase of the HPS consists of 12 weekly surveys until the week of July 16, 2020. Starting from the second phase, the U.S. Census Bureau changed the HPS to every two weeks. This study’s most recent survey was conducted from September 29, 2021, to October 11, 2021, week 39. According to the US Census Bureau, the sample represents the U.S. population aged 18 years and above.

The US Household Pulse Survey participants were asked whether their household had experienced any food insufficiency in the last seven days. The choice options are 1) enough of the types of food wanted; 2) enough food, but not always the types wanted; 3) sometimes not enough to eat, and 4) often not enough to eat. Following the definition of food security by the USDA, I assume that the participants who chose the first two options were food secure. More specifically, I classify households being low food secure if they decided “sometimes not enough to eat”; and very low food secure if they chose “often not enough to eat.”

The participants were asked to identify their race and ethnic background among non-Hispanic white, Hispanic, non-Hispanic black, Asian, and others. They were also asked to report whether they have children under age 18 and whether any household members lost their employment income. The participants were also asked to choose their household income level among the following intervals: less than $25,000, $25,000-$34,999, $35,000-$49,999, $50,000-$74,999, $75,000-$99,999, $100,000-$149,999, $150,000-$199,999.

---

Figure 1: The trend of food insecurity among the US households (1995-2020).


Last Accessed on December 4th, 2021.
and $200,000 and above. According to the 2020 Census, the federal poverty line for an average household with four members was $26,200, and the median household income was $67,521. I created four income categories: less than $25,000 for those living in poverty; $25,000-$75,000 for those living out of poverty but below the median household income; $75,000-$100,000; and $100,000 and above.

I used graphical and statistical analyses to examine the proposed research questions on food insecurity among US households during the pandemic. Statistical student t-tests were conducted to assess the inequality of food insecurity between different types of households.

**Results and Discussion**

The empirical analyses investigate the prevalence, reasons, and inequality of food security and channels through which food-insecure US households get their free meals and groceries to improve their food security during the pandemic.

**Overall Food Insecurity and its Reasons among US Households:**

Table 1 shows that the prevalence was 21.86% for low food security and 6.01% for very low food security during the pandemic, higher than the 2019 level (10.54% and 4.11%). Approximately 27.86% of US households faced either low or very low food security, more than double the level in 2019 (14.65%).

<table>
<thead>
<tr>
<th>Food Insecurity Measures</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low food security</td>
<td>16.29%</td>
<td>25.65%</td>
<td>21.96%</td>
<td>1.50%</td>
</tr>
<tr>
<td>(sometimes not enough to eat)</td>
<td>(Week 4)</td>
<td>(Week 33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very low food security</td>
<td>4.08%</td>
<td>8.27%</td>
<td>6.01%</td>
<td>1.17%</td>
</tr>
<tr>
<td>(often not enough to eat)</td>
<td>(Week 11)</td>
<td>(Week 30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall food insecurity</td>
<td>23.45%</td>
<td>32.70%</td>
<td>27.46%</td>
<td>2.54%</td>
</tr>
<tr>
<td>(low and very low food security)</td>
<td>(Week 9)</td>
<td>(Week 31)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1:** Food insecurity and reasons for US households during the pandemic (%).

Table 1 summarizes the self-reported reasons for food insecurity: some could not afford to stay food secure (60.41%); others were afraid to go out to get groceries and meals (13.59%); 11.44% could not get out of their houses to get groceries and/or meals. From Week 1 to Week 27, the participants were provided a choice that stores did not offer the type of food they would like to buy – 11.16% cited this as the main reason. Figure 2(a) shows a significantly high prevalence of low food security and very low food security during the pandemic. Table 1 shows that the peak time for low food security and very low food security were in Week 33 (June 23, 2021 – July 5, 2021) and Week 38 (September 15, 2021 – September 27, 2021), respectively. The widespread of the Delta variant likely attributed to the peak. Figure 2(b) shows that an increasing proportion of the participants attribute the lack of affordability to their food insecurity situation.

(a) Low and very low food security during the pandemic

(b) Reasons for food insecurity during the pandemic

**Figure 2:** Food insecurity and its reasons among US Households during the pandemic.

The ERS Food Price Outlook tracks changes in the Consumer Price Index for food at home and food away from home. Retail food prices were higher in September 2021 than in September 2020 for all food categories. The highest price increases were for beef and veal (17.6%) and pork (12.7%). The disruptions caused by the pandemic to the food supply chain are attributed to the rising food prices. A concerted effort between the public and private sectors including food industries and major food companies, should be promoted to mitigate the impacts of the pandemic disruptions and improve efficiency.

**Household Food Insecurity between Difference Households:**

Inequality of food insecurity was pronounced in the US before the pandemic and further exacerbated it. I examined the inequality of food security by race and income level between households with and without children and between households with and without income loss.

**Race-specific Inequality of Food Insecurity**

Figure 3 shows a clear pattern of food insecurity by race – non-Hispanic black households experienced the most food insecurity, followed by Hispanic and Latino households, non-Hispanic white households, and Asian households had the least exposure to food insecurity during the pandemic. Table 2 shows that the proportion of low food-secure households was 29.50% for non-Hispanic black, 26.24% for Hispanics and Latinos, 18.42% for non-Hispanic white, and 14.22% for Asians.

**Figure 3:** Food insecurity of US households by race during the pandemic by races.

<table>
<thead>
<tr>
<th>Table 2: Food insecurity during the pandemic among different households (%)</th>
<th>White</th>
<th>Hispanic, Latino</th>
<th>Non-Hispanic Black</th>
<th>Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Food Security: Sometimes not enough to eat</td>
<td>18.42</td>
<td>2.84</td>
<td>32.99</td>
<td>3.78</td>
</tr>
<tr>
<td>Average</td>
<td>18.42</td>
<td>2.84</td>
<td>32.99</td>
<td>3.78</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>5.92</td>
<td>6.75</td>
<td>3.16</td>
<td>3.32</td>
</tr>
<tr>
<td>Very Low Food Security: Often not enough to eat</td>
<td>5.52</td>
<td>1.56</td>
<td>32.99</td>
<td>3.78</td>
</tr>
<tr>
<td>Average</td>
<td>5.52</td>
<td>1.56</td>
<td>32.99</td>
<td>3.78</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>5.92</td>
<td>6.75</td>
<td>3.16</td>
<td>3.32</td>
</tr>
<tr>
<td>Low or Very Low Food Security: Often not enough to eat</td>
<td>5.52</td>
<td>1.56</td>
<td>32.99</td>
<td>3.78</td>
</tr>
<tr>
<td>Average</td>
<td>5.52</td>
<td>1.56</td>
<td>32.99</td>
<td>3.78</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>5.92</td>
<td>6.75</td>
<td>3.16</td>
<td>3.32</td>
</tr>
</tbody>
</table>
I formulated the null and alternative hypotheses below:

Null Hypothesis: $X_{\text{Hispanic/Latino}} = X_{\text{Asian}}$
Alternative Hypothesis: $X_{\text{Hispanic/Latino}} > X_{\text{Asian}}$

Based on the corresponding $p$-value for Student $t$-tests in Table 3, we reject all null hypotheses at the 1% statistical significance level and accept the alternative hypotheses. I, therefore, conclude that race-specific inequality of food insecurity was pronounced: non-Hispanic black households most likely experienced both low security and very low security, followed by Hispanic and Latino households, non-Hispanic white households, and Asian households the least likely to experience food insecurity. The race-specific inequality of food insecurity found in this study is consistent with the literature.

Table 3: Comparison of food insecurity between different households (%).

<table>
<thead>
<tr>
<th>Household Income</th>
<th>Low Food Security</th>
<th>Very low food security</th>
<th>Low and very low food security</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Difference (p-value)</td>
<td>Difference (p-value)</td>
<td>Difference (p-value)</td>
</tr>
<tr>
<td>Non-Hispanic Black vs. Hispanic/Latino</td>
<td>$2.12***$</td>
<td>$0.00$</td>
<td>$4.16***$</td>
</tr>
<tr>
<td>Hispanic/Latino vs. Non-Hispanic White</td>
<td>$1.53***$</td>
<td>$0.00$</td>
<td>$3.29***$</td>
</tr>
<tr>
<td>Non-Hispanic White vs. Asian</td>
<td>$0.12$</td>
<td>$0.00$</td>
<td>$0.00$</td>
</tr>
</tbody>
</table>

Table 3 shows the prevalence of low food insecurity was 26.62% for households with employment income loss and 16.36% for households without employment income loss. The pattern of the majority of very low food security was the same. Still, in a much smaller magnitude: 8.01% for households lost employment income and 3.93% for households without any employment income loss.

I formulated the null and alternative hypotheses below:

Null Hypothesis: $X_{\text{With Employment Income Loss}} = X_{\text{Without Employment Income Loss}}$
Alternative Hypothesis: $X_{\text{With Employment Income Loss}} > X_{\text{Without Employment Income Loss}}$

Based on the corresponding $p$-values of several Student $t$-tests shown in Table 3, I reject all the null hypotheses at the 1% statistical significance level likely to experience low food security than households without children during the pandemic.

Inequality of Food Insecurity by Household Income Levels:

Figure 6 demonstrates that the lower the household income, the more likely they experience food insecurity.
that the prevalence of low food security was 32.35% for low-income households (less than $25,000), 21.44% for households with incomes between $25,000 and $75,000, 12.25% for households with incomes between $75,000 and $100,000, and 9.25% for households with incomes of $100,000 or greater. The patterns were the same for the prevalence of very low food security: 11.45% for households with incomes less than $25,000, 4.65% for households with incomes between $25,000 and $75,000, 2.45% for households with incomes between $75,000 and $100,000, and 2.51% for households with incomes of $100,000 or greater. Almost half of the households living in poverty experienced either low or very low food security (43.80%), followed by those with income between $250,000 and $750,000 (26.10%). Fewer households with income above $750,000 experienced food insecurity (14.69%) for those between $25,000 and $75,000, 10.76% for the highest income households -- $100,000 or greater.

Food banks and pantries played an essential role among different channels in providing emergency food. More than a quarter (25.07%) of households received free meals or groceries from food banks and pantries. Food banks and pantries were forced to develop and implement emergency plans to accommodate social-distancing guidelines, lack of staff, and the sudden increase in demand. Many food banks and pantries furloughed volunteers and implemented sanitation measures to protect workers and ensure food safety. They prepacked emergency food boxes rather than allowing clients to select their food and ask their clients to remain outside rather than enter the building. With a reduced labor force and volunteers, food banks and pantries had to accommodate an increased demand than before the pandemic, including those newly food-insecure individuals unfamiliar with navigating assistance.

As schools closed most of the time during the pandemic, school districts developed different strategies to support emergency feeding programs for their students and families without clear federal mandates to continue feeding students and guidelines for effectively executing feeding programs. About a quarter (24.37%) of the respondents received free meals or groceries from school and other programs aimed at children, who served meals and snacks to students outside the cafeteria and used strategies such as grab-n-go’s, drive-thru’,s, and food deliveries to meet the needs of children. School districts should work closely with all the parties, including the local government, industries, parents, and volunteer groups, to continue such support to help students and families in need.

Table 4 also shows that about 20% of the respondents received free meals or groceries from family, friends, and neighbors; followed by religious organizations such as churches (14.43%), shelters or soup kitchens (2.89%), and home-delivered meal services (3.36%). Home-delivered meal services such as Meals on Wheels are essential to people bound at home, including the elderly living in their own homes or care facilities. My research finds that only 3.36% of free meals and groceries came from home-delivery meal services. One solution to mobilize home-delivery meal services is to provide free or reduced-price meals or groceries. An alternative solution is to create more innovative supporting environments. For example, we can create

| Table 4: Free meals, SNAP, and provider or sources of free groceries and meals (%) |
|---------------------------------|-----------------|-----------------|-----------------|
| Source of free groceries and/or meals | Min. | Max. | Mean | Standard Deviation |
| Food pantry or food bank | 20.73% | 28.41% | 25.07% | 2.07% |
| School programs aimed at children | 18.35% | 32.64% | 24.37% | 3.79% |
| Family, friends, or neighbors | 14.84% | 24.74% | 19.26% | 2.47% |
| Religious organization (e.g., Church) | 11.29% | 18.47% | 14.43% | 2.59% |
| Shelter or soup kitchen | 1.14% | 6.93% | 3.97% | 1.30% |
| Home-delivered meal services (e.g., Meals on Wheels) | 0.70% | 6.57% | 3.98% | 1.51% |
| Other community program | 7.99% | 13.42% | 10.65% | 1.43% |

DOI: 10.36838/v5i4.12
community fridges or food baskets, where people could put in groceries they do not need, and others could come to grab what they need.

- **Conclusions and Policy Implications**

US households experienced significant challenges in staying food secure during the pandemic due to lockdowns and isolations, disruptions to the food supply chain, and increased food prices in the United States and worldwide. The Household Pulse Survey conducted by the U.S Census Bureau, a 20-minute online survey, has collected information on how the pandemic affected US households economically and socially. I used the 39 weeks of data from the Household Pulse surveys from April 23, 2020, to September 29, 2021, to examine the food insecurity of US households during the pandemic. I draw the following important conclusions based on graphical and statistical analyses of this nationally representative dataset.

First, the pandemic significantly increased the prevalence of food insecurity among US households: 21.86% and 6.01% of households were either low food secure or very low food secure, double the pre-pandemic level in 2019. The primary underlying reason was the lack of affordability resulting from income loss and inflated food prices.

Second, the inequality of food insecurity was pronounced and increased. People of color, low-income households, households with income loss, and households with children were impacted most by food insecurity. Given that these household types had a disproportionately higher prevalence of food insecurity before the pandemic, the impact of the pandemic elevated the inequality of food insecurity even more.

Third, about 18.66% of households received free meals, and 29.66% received SNAP benefits. Among all the channels, the food pantry or food bank provided most free groceries and meals, followed by school programs aiming to help children stay food secure. Almost all the organizations providing emergency food to needed families and communities were affected by the pandemic as they faced challenges in having sufficient staff and volunteers, food supplies, and the logistics of providing free meals and/or food. The efficacy of these organizations in improving the food security of households and communities is challenging.

The study provides the following policy implications. First, this study helps identify households that need help most to stay food secure – minority households, households that lost their income, households with children, and low-income households. The government, non-profit organizations, and community programs may want to target these types of households to reduce food insecurity and its inequality. Second, SNAP benefits were one of the critical resources for low-income households to fight against hunger and improve food security. The pandemic also calls for changes in the SNAP on the amount, duration, and allowed shopping methods of SNAP benefits.

- **Acknowledgments**

I thank my teacher, Mrs. Valerie Alimonti, for her encouragement and support and, most importantly, her belief in me. I also thank Professor Robson at Rutgers University for the Zoom discussion and his encouragement.

- **References**


- **Author**

Julia is a tenth grader at East Brunswick High School in New Jersey, United States. The possible majors she has entertained are psychology, education, and arts. She has been participating in many volunteer events to build a better community by using her talents in arts and dance.
A Novel Association of \textit{EGFR} Gene Alteration with Decreased Glioblastoma Patient Survival Rate

June Hyun

Harrow School, 5 High St, Harrow, Middlesex, HA1 3HP, United Kingdom; junehyun2024@gmail.com
Mentor: Professor Woo Rin Lee

\textbf{ABSTRACT:} Glioblastoma multiforme (GBM) is the most aggressive and prevalent adult central nervous system (CNS) tumor. Epidermal growth factor receptor (EGFR) alterations are prognostic and diagnostic markers of GBM. Our study aims to investigate the association of \textit{EGFR} alteration with changes in GBM patients’ overall survivability. cBioPortal genomic data analysis was used to retrieve genomic data and clinical attributes from 1,122 GBM patients. We performed a comparative study on \textit{EGFR}-altered and non-altered patient groups, mutation and amplification frequency analysis, \textit{EGFR} mutation analysis, and clinical attribute analysis. The mutation and amplification frequency analysis showed that only the \textit{EGFR} gene had higher amplification and mutation counts in the deceased group, indicating a notable association between \textit{EGFR} alteration and decreased survivability. Mutation analysis identified the mechanism by which \textit{EGFR} mutation altered signal transduction and cell proliferation, leading to uncontrolled cell division. A clinical attribute analysis on the relationship between \textit{TERT} expression, \textit{EGFR} alteration, and chromosome 7 gain/chromosome 10 loss (Chr 7 gain/Chr 10 loss) suggested the viability of targeting \textit{EGFR} signaling for GBM therapy. Genomic data analysis revealed a strong association between \textit{EGFR} gene alteration and decreased GBM patient survival rate. Thus, potential \textit{EGFR} molecular alteration is a field of therapeutic value.

\textbf{KEYWORDS:} Biomedical Engineering, Cancer Biology; Genetics; \textit{EGFR}; Patient Survival; Glioblastoma.

\section*{Introduction}

Glioblastoma multiforme (GBM) is a grade IV astrocytoma that accounts for 47.7\% of all central nervous system (CNS) tumor diagnoses; over 12,000 Americans receive a GBM diagnosis annually.\textsuperscript{1} 90\% of GBMs occur \textit{de novo} in the brain, while the remaining 10\% develop from low-grade astrocytomas. Malignant cells transfer to adjacent brain cells, contributing to the tumor’s rapid growth. Web-like vein synthesis and the sheer size of the mass often cause peritumoral edema, intracranial hypertension, and tumor-induced epilepsy. Thus, the median survival time is 15 months, and the 5-year survival is less than 5\%.\textsuperscript{2} Although surgical removal of over 85\% of the tumor does prolong survival, radio-chemo-therapy-resistant GBM relapse is common. \textit{IDH1} mutation, MGMT methylation status, \textit{TERT} promoter mutation, and \textit{EGFR} mutation are all prognostic and diagnostic markers of GBM.\textsuperscript{3} Therefore, our research aims to describe the relationship between \textit{EGFR} mutation and overall patient survivability to address the therapeutic value of the molecular alteration.

The accumulation of genetic and epigenetic alterations causes cancer development.\textsuperscript{4} DNA mutations, rearrangements, deletion, and amplification are well-known abnormalities in cancer cells. Recent advancements in sequencing technology allow the study of the human genome in nucleotide resolution. Identifying the genetic differences that may cause cancer is possible by sequencing DNA or RNA. By sequencing DNA or RNA, identifying the genetic differences that may cause cancer is possible. This approach also quantifies the activity of genes encoded in human DNA to understand the proteins involved in cancer cell characteristics, such as uncontrolled cell growth.\textsuperscript{5}

When cancer-causing genetic alterations are identified, we can better understand the molecular basis of cancer development. Therefore, clinical data that describes each cancer patient and their genomic data are essential.\textsuperscript{6} Recently, cBioPortal, the open-source cancer genomic database with patient clinical data, was shared with researchers worldwide. As cBioPortal provides a significant portion of worldwide cancer genomic data, it opens new opportunities for discovering gene alterations that cause cancer development.\textsuperscript{7}

Epidermal growth factor receptor (EGFR) is a transmembrane glycoprotein receptor tyrosine kinase (RTK) responsible for cell proliferation, differentiation, and survival.\textsuperscript{8} Seven ligands, EGF, TGF\textalpha, AREG, EREG, BTC, HBEGF, and EPGN, bind to the extracellular domain. The ligand interaction with the receptor homodimerizes or heterodimerizes EGFR with a protein of the ErbB (RTK) family.\textsuperscript{9} The dimerization enables the intracellular c-terminus of the EGFR to autophosphorylate, triggering many different canonical and non-canonical intracellular cascades. Point mutation, amplification, and up-regulation of the \textit{EGFR} gene are hallmarks of tumorigenesis and are often used in diagnosing non-small cell lung cancer (NSCLC) and anal cancer.\textsuperscript{10} \textit{EGFR}-positive adenocarcinoma of the lung has a median overall survival of 36.9 months, while \textit{EGFR}-negative adenocarcinoma has a median overall survival of 55.4 months. Loss of regulation of \textit{EGFR} is present in 50\% of GBM patients. Therefore, investigating the relationship between \textit{EGFR} gene up-regulation and patient survival is vital to identify a new \textit{EGFR}-targeting treatment pathway in GBM.\textsuperscript{11}
Since gene copy number alteration and mutation may cause the loss of regulation of EGFR, we investigated the gene using the genomic data of 1,122 GBM patients in this study. The low survivability of GBM and common relapses necessitate investigations to discover novel genetic associations with GBM. However, research has yet to thoroughly investigate the relationship between patients’ survival rate and EGFR gene mutation in GBM patients. Therefore, we hypothesized that genetic alteration of EGFR may affect the GBM patient survival rate.

Methods

**cBioPortal genomic data analysis:**

We used the cBioPortal database was used to retrieve genomic data and clinical attributes from 1,122 GBM patients and performed two comparative survival analyses. The first analysis divided the patients into two groups - EGFR-amplified and EGFR-non-amplified patients. The second analysis divided the patients into EGFR-mutated and EGFR-non-mutated groups. Then, we analyzed the overall patients’ survival with statistical analysis by calculating the Log-rank p-value.

**Analyzing EGFR amplification and mutation frequency:**

We performed gene amplification and mutation enrichment analysis using cBioPortal. After collecting the genomic and survival data of 1,122 GBM patients, we divided the patients into living and deceased groups. First, enrichment analysis ranked the most amplified and mutated genes in 1,122 GBM patients. Then, we performed statistical analysis to calculate the p-value using Wilcoxon Test.

**EGFR mutation analysis:**

The cBioPortal mutation analysis function plotted the illustration of EGFR mutation numbers and positions. The x-axis represented the position of the amino acid sequence of EGFR and the position of the mutations within the protein domain; the y-axis represented the number of mutations found in GBM patients.

**GBM patient’s clinical attribute analysis:**

cBioPortal is a web-based tool that allows users to analyze and visualize the clinical attributes of cancer patients. Using cBioPortal, we analyzed the clinical attributes of 1,122 GBM patients within two groups: patients with or without EGFR alterations. In addition, cBioPortal generates a visual representation of the data, such as a bar chart or scatter plot.

Results

Since gene copy number alteration and mutation may cause the loss of regulation of EGFR, we investigated the gene using the genomic data of 1,122 GBM patients in this study. The low survivability of GBM and common relapses necessitate investigations to discover novel genetic associations with GBM. However, research has yet to thoroughly investigate the relationship between patients’ survival rate and EGFR gene mutation in GBM patients. Therefore, we hypothesized that genetic alteration of EGFR may affect the GBM patient survival rate.

We analyzed EGFR amplification and EGFR mutation against the patient’s overall survival months to investigate the genetic association between the EGFR gene and GBM survivability. We used the cBioPortal database to retrieve genomic data from 1,122 GBM patients and performed two comparative survival analyses. The first analysis divided the patient group into EGFR-amplified and EGFR-non-amplified patients; the second analysis divided the patient group into EGFR-mutated and EGFR-non-mutated patients. The overall survival analysis result indicates that EGFR-amplified patients have a significantly decreased survival rate than EGFR-non-amplified patients (Figure 1A). The median overall survival in months is 14.30 for EGFR-amplified patients; the median overall survival in months is 36.80 for EGFR-non-amplified patients (Figure 1A). The difference in the median overall survival, 22.50 months, indicates EGFR amplification’s role in poor prognosis in GBM patients (Figure 1A). The second patient’s comprehensive survival analysis shows that EGFR mutation decreases the survival rate compared to EGFR non-mutated patients (Figure 1B). EGFR-mutated patients’ median overall survival in months is 13.80, while EGFR-non-mutated patients’ median overall survival in months is 44.40 (Figure 1B). The even more significant difference between the median overall survival, 30.60 months, suggests that EGFR mutation influences GBM patient prognosis negatively (Figure 1B). EGFR amplification and EGFR mutation are significant factors that affect GBM patient survival rates. However, since EGFR mutation has a more considerable difference between median overall survival in months than EGFR amplification, the EGFR gene mutation might be a more significant factor in overall survival.

**Figure 1:** EGFR amplified and mutated GBM patients show significantly lower survival rates than EGFR non-amplified and non-mutated GBM patients. (A) A graph indicating the probability of overall survival (y-axis) a-

![Figure 2](image-url)
and TSPAN31 (Figure 2). All ten genes showed significantly higher amplification frequency in the deceased group (Figure 2). The EGFR gene showed the highest amplification frequency in the deceased group and the largest difference between amplification frequency in living and deceased groups (Figure 2). Since the deceased group had a higher amplification frequency than the living group in all ten genes, the entire genome of the deceased group may have low stability. Furthermore, the high amplification frequency of EGFR in the deceased group suggests that the amplification of EGFR may contribute to a lower survival rate of GBM patients.

Mutation occurs throughout the genome of a GBM patient; thus, we analyzed the mutation frequency of the whole genome of living and deceased patients. We found the top ten genes with the most difference in mutation frequency between living and deceased groups. Only four genes, IDH1, ATRX, EGFR, and CIC, showed a significant difference in mutation frequency (Figure 3). Of the four genes, IDH1, ATRX, and CIC showed a higher mutation frequency in the living group, while EGFR displayed a higher mutation frequency in the deceased group (Figure 3). The significantly higher mutation frequency in the living group for the IDH1, ATRX, and CIC genes suggests that a high mutation frequency in the living group may not be associated with a lower patient survival rate. Since EGFR is the only gene with a significant mutation frequency in the deceased group, a high mutation frequency of EGFR may lower the survival rate of GBM (Figure 3).

The EGFR gene is significantly mutated in deceased patients. Thus, we analyzed the position of EGFR mutation in the EGFR amino acid sequence to find important mutation positions associated with lower survival in GBM patients. We found five separate domains in the EGFR amino acid sequence: Receptor L domain (57-167 aa), Furin-like cysteine-rich region (185-338 aa), Receptor L domain (361-480 aa), Growth factor receptor domain IV (505-636 aa), and Protein tyrosine kinase (713-965 aa) (Figure 4). The number of EGFR mutations was highest in the Furin-like cysteine-rich region (Figure 4).

The growth factor receptor domain IV had the second highest number of EGFR mutations out of the domains (Figure 4). Previous studies confirm that EGFR mutations at alanine 289 (EGFR A289D/T/V) of the Furin-like cysteine-rich region are associated with a significant reduction in the overall survival of GBM patients. Therefore, the high number and frequency of EGFR mutations in the Furin-like cysteine-rich region and the growth factor receptor domain IV suggest that mutations in the two domains may be significantly associated with lower survival rates in GBM patients.

**Discussion**

This paper is the first study to investigate the association of EGFR gene amplification and mutation with overall survivability in GBM patients. Figure 1 shows the comparison of survival months of EGFR-amplified, non-amplified, mutated,
and non-mutated patients. The graph shows an apparent reduction in survival for EGFR-amplified and mutated patients, highlighting a significant association between amplification, mutation, and the deceased patient group. Furthermore, EGFR-mutated and non-mutated groups have a larger difference in median overall survival than EGFR-amplified and non-amplified groups, suggesting that EGFR mutation, rather than amplification, may more significantly affect patient survivability in GBM. The strong association between EGFR gene status and patient survival suggests that the gene status may be used as a new biomarker factor to consider when predicting GBM patient prognosis, improving the prediction accuracy, and providing patients and relatives with a confident disclosure.

Previous studies show similar EGFR gene relationships with patient survivability in NSCLC. The likelihood of EGFR amplification increases in advanced clinical stages of lung cancer. Furthermore, EGFR amplification increases with EGFR tyrosine kinase inhibitor (TKI) resistance. As TKI is a common drug used to treat NSCLC, TKI resistance is associated with poor outcomes in NSCLC patients. Therefore, EGFR amplification is significantly associated with low survivability of patients, which agrees with our study's result from Figure 1.

Past studies investigating EGFR in GBM corroborate the result. EGFR amplification observed in this paper may lead to EGFR overexpression in approximately 50-60% of GBM tumors. Furthermore, EGFR mutations lie primarily in the extracellular domain. Most mutations in Figure 4 lie between amino acids 0-600 aa in the extracellular domain showing consistency with the previous study.

We analyzed the gene amplification and mutation frequency in the whole genome of living and deceased GBM patients to further investigate the association of EGFR amplification and mutation with deceased patient survivability. Many genes, including EGFR, displayed significantly higher amplification frequency in deceased patient groups than in living patient groups, suggesting that genetic instability of the whole genome contributes to the lower survivability of patients with EGFR amplification observed in previous analyses. On the other hand, mutation frequency analysis identified four genes (IDH1, ATRX, EGFR, and CIC) with significant differences between living and deceased group mutation frequency. IDH1, ATRX, and CIC displayed higher mutation frequency in living groups, suggesting that mutations in these genes do not lower the survivability of GBM patients. However, EGFR had a higher mutation frequency in the deceased group. Therefore, the analysis found that EGFR is the only gene in the GBM patient’s whole genome with a higher amplification and mutation frequency which may lower survivability, highlighting the importance of further research on the role of EGFR in GBM. This study analyzed the position of mutations in the EGFR amino acid sequence to identify mutations that significantly affect survivability. We categorized five distinct domains in the sequence, of which three (furin-like cysteine-rich region, growth factor receptor domain IV, and receptor L domain) contained many EGFR mutations. Furin-like domain facilitates receptor aggregation for signal transduction by RTKs. A mutation in the furin-like domain may enhance the internal signaling, promoting uncontrolled cell growth. Furthermore, previous studies indicate that mutation in the furin-like domain correlates with less effective targeted therapy and potential drug resistance, decreasing survivability in GBM patients.

Growth factor receptor domain IV facilitates the dimerization of EGFR when EGF is received, which regulates phosphorylation and subsequent signal transduction for cell growth. Ligand stimulation in mutated growth factor receptor domain IV results in the formation of a tyrosine-phosphorylated, disulfide-bonded dimer without EGF. Therefore, the mutation causes EGFR to trigger cell growth without EGF binding, leading to unregulated cell growth in GBM.

The analysis of mutation positions in the EGFR amino acid sequence identified that most mutations in the receptor L domain occurred in the L1 subdomain, located near the N-terminus and responsible for the binding of ligands. The L1 subdomain causes a conformational change in EGFR when a ligand binds to the receptor, influencing the intracellular domain to trigger downstream signaling pathways that promote cell growth and division. Therefore, a mutation in L1 can disrupt ligand binding to EGFR or enhance the signaling pathways, promoting cancer growth. Furthermore, previous studies suggest a link between L1 subdomain mutation and cancer drug resistance. The analysis highlights the multitude of EGFR mutations that affect signal transduction triggering cell growth and division, suggesting that EGFR may be an area of interest for therapeutic research for GBM drugs.

We investigated EGFR alteration association with TERT expression and Chr 7 gain/Chr 10 loss to compare the clinical attributes of GBM patients with and without EGFR alteration. The previous study indicated that EGFR signaling induces TERT expression by transcription factors such as c-Myc, Sp1, and Ets-2. Since TERT is often increased in cancer cells that activate the infinite cell division, targeting EGFR signaling is considered an effective therapeutic strategy for lung cancer. 21 Chr 7 gain/Chr 10 loss is also often associated with EGFR gene mutation in lung cancer. Previous studies indicate that these mutations may activate EGFR signal pathways, contributing to cancer development and growth. Our study shows a similar relationship between EGFR, TERT, and Chr 7 gain/Chr 10 loss exists in GBM, demonstrating the importance of further investigation into the targeting of EGFR alteration for novel therapeutic pathways.

**Conclusion**

In this study, we found that EGFR amplification and mutation significantly reduce GBM patient survival rate and that mutation might significantly influence overall survivability. EGFR amplification had the highest frequency in the deceased group out of the top ten genes, with the largest disparity in frequency between living and deceased groups, suggesting that EGFR amplification may affect survivability. Out of four significant mutations, only the EGFR mutation had a significant mutation frequency in the deceased group; thus, a higher EGFR mutation frequency may lower the survival rate. When we further analyzed the EGFR mutation position, the high mutation frequency was detected in the furin-like cysteine-
rich region. Growth factor receptor domain IV suggests a significant association between lower survivability and the two domains. We also analyzed the clinical attributes associated with EGFR-altered patient groups. We found that high levels of TERT expression and Chr 7 gain/Chr 10 loss in EGFR-altered groups show that alteration in EGFR may lower survivability in GBM patients and increase genomic instability.

This study analyzed patient genomic data associated with decreased survival rates from one published database. Therefore, we must conduct in vitro experiments to further analyze the effect of EGFR amplification and mutation on cancer cell function, such as the speed of cancer cell division, migration, and invasion. Also, we should perform an in vivo mouse experiment or xenograft to verify the results, indicating that EGFR amplification and mutation decrease the patient’s survival rate. In addition, many other genes, such as SEC61G and LANCL2, have high amplification frequencies in the deceased group. Thus, we must analyze other amplified genes in the deceased group indicated in Figure 2 to discover any interactions between genes that lead to decreased survival rates in GBM patients.

Acknowledgments

I want to thank Professor Woo Rin Lee at the University of Suwon for acting as a mentor and remotely guiding my research. Furthermore, I would like to show appreciation for Dr. Luca Passamonti at Biogen, who inspired my passion for neurology and degenerative diseases.

References


Author

June Hyun is a junior at Harrow School in the United Kingdom. As the Scientific Society President, he is involved in various school volunteering and research activities. His interest in neurology, neurodegenerative diseases, and robotic neurorehabilitation drive his aspiration to become a physician-researcher.
Certificate-Based Key Replacement in Single Use Hash-Based Signatures: A Post-Quantum Technique

Matthew D. Grupenhoff
Henry M. Gunn High School, Palo Alto, California, 94306, USA; mg27713@munge.com
Mentor: Dr. Dong, Polygence

ABSTRACT: The recent advancement in quantum computing poses a concern over the security of current public key cryptography systems. Since then, hash-based signature schemes have been created to replace vulnerable schemes, although they have limitations such as requiring a large key and/or signature size or only working a limited number of times. This paper discusses a solution to these two problems through key replacement to be used an arbitrary number of times with a reasonable key size. The experiments were done by extending the Lamport signature scheme with this key replacement, leading to a signature scheme that is slower than DSA and ECDSA, but faster than RSA for sufficiently large security levels. As a result, while applying this technique does not lead to very fast performance, it still has the advantage of post-quantum security, and it is not excessively slow.

KEYWORDS: Systems Software; Cybersecurity; Digital Signatures; Post-Quantum; Hash-Based.

Introduction

The current internet is secure because of cryptographic systems such as digital signatures. However, the recent advancement in quantum computing poses a concern over the security of current public key cryptography systems. This is the result of an algorithm from 1994 known as Shor's algorithm, which is capable of solving the discrete logarithm problem as well as factoring large numbers, both in polynomial time.¹ As a result, the public key cryptography used every day, such as transmitting private information online, will eventually be rendered completely insecure.

Today's internet uses a number of concepts from cryptography to remain secure. A key concept in cryptography is encryption, which is the transformation of secret information into a form that can only be transformed back if a small secret referred to as a key is known; this reverse process is known as decryption. To distinguish between encrypted data from different parties, a key is also required to encrypt the data, to begin with. This key is important because it ensures that even if an attacker knows how the security works, they still need additional information that cannot be efficiently discovered.²

Some algorithms, known as symmetric algorithms, use the same key for both encryption and decryption, which is generally easy to implement; they rely on the fact that certain functions, together with their inverses, can be trivially computed if a piece of secret information is known. In cryptography, this secret information is stored as the key. While symmetric cryptography is generally very secure, there are times when it cannot be used, such as allowing anyone to verify the authenticity of a message while restricting who can create authentic messages.

The security of a cryptographic algorithm is defined by its security level n, which is the number such that computing unintended information, such as a private key from a public key, requires a maximum of $2^n$ comparisons;⁴ therefore, it requires $2^{n-1}$ comparisons on average. Security levels are helpful for comparing the performance of new algorithms with existing algorithms, and deciding what parameters to use when switching algorithms.

One limitation of symmetric algorithms is that they cannot be used to prevent parties from decrypting data while allowing them to encrypt it, or vice versa. By using an algorithm that uses a separate key for each direction, this problem is solved; such algorithms are known as asymmetric algorithms. One of the two keys, known as the public key, is typically shared to allow anybody to either encrypt or decrypt data, depending on what the encryption is being used for, while the other key, known as the private key, is not shared with the public, and is typically only used by the party that generated it; the private key is required to use a function known as a trapdoor function, which is used for the operation that is not intended to be used publicly. These keys must be generated through a method such that the private key cannot be efficiently computed from the public key; if it is possible to derive the private key from the public key, then the security is effectively reduced to that of a symmetric algorithm.

One particular algorithm, RSA (Rivest, Shamir, and Adleman), has been used for decades due to it receiving no critical attacks. Its key generation algorithm starts by generating 2 random prime numbers, known as p and q, and multiplying them, releasing the product as part of the public key. This is significant because there is no efficient algorithm for finding factors of a large number on a classical computer but being able to efficiently find factors would mean p and q are recoverable. Therefore, the private key can be computed. This defeats the security of RSA because, with the private key, an attacker can take unauthorized actions such as forging secure messages.

It is also useful at times to be able to verify a piece of data without storing it in its entirety, either because it is inefficient...
or insecure to store the entire piece of data. For this, a one-way function is also used to transform a variable-length input into a fixed-length output with an extremely low probability of two distinct inputs with the same output being discovered. These one-way functions are known as hash functions, and their outputs, known as hashes, are often stored next to critical data to ensure it was not accidentally modified or used for cases such as password storage. In both of these cases, the data can be verified by taking the hash a second time and comparing the result, but there is an extremely low probability of the hash being reversed.

Sometimes, verifying a large piece of data at once is not an option. In these cases, the data can be verified in smaller parts by hashing each part, then hashing the list of hashes and storing that master hash securely. As this list grows, multiple master hashes can be evenly divided with a higher master hash, and this tree can be extended as deep as necessary. A tree like this is known as a Merkle tree, or a hash tree.

One significant application of asymmetric cryptography is the concept of digital signatures, which can be verified by anybody but only generated by the person with a private key. Given an asymmetric encryption scheme, such as RSA, and a hashing algorithm, signatures are commonly implemented by hashing the content and encrypting the hash with the private key; the signature can then be verified by decrypting it with the public key and comparing the result with the hash of the content.

Quantum computers are capable of operating on quantum states of 0 and 1, as opposed to how everyday computers, known as classical computers, operate solely on 0 and 1. This allows them to try every possible input to a one-way function at the same time. However, due to limitations imposed by quantum mechanics, a special circuit is needed at the end to extract any useful results.

Shor’s algorithm is an algorithm for quantum computers capable of finding factors of a number n in $O((\log n)^3 \log(\log(\log n)))$ time, which is far more efficient than the $O((\log n)^{O(\log \log n)})$ time required on a classical computer. As a result, p and q can be recovered from a public key, and the corresponding private key can be calculated, breaking the security. Most other asymmetric encryption algorithms, such as ECDSA (Elliptic Curve Digital Signature Algorithm), were also broken, since they ultimately rely on problems that depend on the discrete logarithm problem, a problem that Shor’s algorithm is capable of solving.⁵,⁶

Fortunately, hash functions and symmetric encryption algorithms are believed to be secure from quantum computing. The only quantum attack that can be used against these algorithms is Grover’s algorithm, which finds the position of a known element in a set of size $n$ in $O(\sqrt{n})$ time. This is not very threatening, since hashes are generally very large.⁷ For example, SHA-256⁸ creates hashes of size 256 bits, meaning the time needed to crack it with Grover’s algorithm is somewhere in the ballpark of $2^{128}$ iterations, which keeps it safe.

As a result of quantum computing, many other schemes for asymmetric encryption have been invented. These are currently secure because their best-known general attack is through Grover’s algorithm, which still requires a large amount of time due to the square root of a large power of 2 still being a large power of 2. Areas of focus for post-quantum cryptography include hash-based cryptography, code-based cryptography, multivariate cryptography, and lattice-based cryptography.⁹,¹⁰ However, all of these areas have barriers that prevent them from completely replacing existing schemes, such as lattice-based cryptography only being effective against worst-case attack complexity, not average-case.¹¹

Hash-based signatures are of particular interest because their security directly relies on the security of the hash function used, meaning when a given security level is deemed insecure, the hash function itself can simply be replaced to increase the security level. Lamport created a one-time hash-based signature scheme, meaning each key can only be used to sign one hash before losing its security, that signs data by mapping each bit to one of two entries in the private key, and keeping a hash of every entry in the public key.¹² Winternitz created another one-time signature scheme with a reduced risk of the key being insecure where the public key is computed by hashing the private key 256 times. Signing a message is done by hashing it, then hashing each element of the private key $p$ a total of 256–$b_i$ times, where $b_i$ is the corresponding byte of the input hash, and storing the result in the respective part of the signature. The signature can then be verified by hashing each hash in the signature $b_i$ times and comparing the result against the public key.¹³ Compared to Lamport signatures, Winternitz signatures are smaller due to only needing to store a hash for every byte of the input instead of every bit. Finally, there is also a few-time signature scheme called Hash of Random Subsets that is similar to Lamport signatures but is constructed in a way that not all security is lost from reusing the key.¹⁴

In addition to these one-time and few-time signature schemes, there are hash-based signature schemes that allow many uses through the switching of keys after each signature. One significant hash-based signature scheme is the Extended Merkle Signature Scheme (XMSS), a scheme where a hash tree is maintained, ending in keys for a one-time or few-time algorithm such as Lamport or Winternitz. Each time a message needs to be signed, the first unused key in the tree is used to sign the message. This allows a single large tree to be used to sign a fixed number of messages, although it still requires the signer to maintain state to remember which key to use next. This need to maintain state was removed in SPHINCS, a scheme similar to XMSS that uses a few-time signature scheme where instead of using the first unused key for the signature, the key index is computed directly from the input data.¹⁵ While this algorithm does have a small risk of reusing a key in the tree, a sufficiently large tree does not have this issue. However, because any path in the tree needs to be shared with anybody who needs to verify the signature, the tree depth also needs to be small enough to be able to efficiently share it.

## Methods

The obvious issue with every hash-based signature scheme is that either there is a limit to the number of times a given key can be used, or the key and/or signature is extremely large.
When one party needs to communicate a sequence of data to another party, this can be solved with a certificate-like approach, where each key signs a new public key together with the main data. This allows for a variably long series of secure messages.

To implement this, a hash-based signature scheme \( S \) must be split into a hash function \( H \) and a scheme-specific transformation \( T \), such that given any message \( m \), \( S = T(H(m)) \). For example, in RSA, \( T \) is the decryption function. In the case of Lamport, \( T \) is the process of taking each bit and locating it in the private key to obtain a signature element, while in Winternitz, \( T \) is the process of hashing each unit of data in the public key a given number of times.

The most straightforward method of replacing keys using certificates would then involve prepending the hashed data with a new public key and applying a scheme-specific transformation to the entire sequence as the signature, storing the new public key as the signer's state. The signature could then be verified by decrypting the signature, verifying the data hash, and storing the new public key.\(^*\) This scheme is shown in Figure 1. Note that unlike a traditional stateful signature scheme, this scheme requires that the verifier stores the most recent public key as their own state.

Suppose there is a scheme with a scheme-specific transformation of \( T \) and an inverse of \( T_{\text{inv}} \), a key-generation algorithm \( \text{KeyGen} \), and a public key derivation algorithm \( \text{PublicKey} \), and there is a hash function \( H \) that outputs \( b_h \) bits. Then, signing a message \( m \) with private key \( sk \) of size \( N \) uses the following pseudocode:

\[
\text{secure data} = \text{pk new} \ || \ H(m) \\
\text{signature} = T(sk, \text{secure data})
\]

To verify this signature, the following pseudocode can be followed:

\[
\text{secure data} = T_{\text{inv}}(\text{signature}) \\
\text{pk new} = \text{secure data}[N:b_h] \\
\text{assert} \ H(m) == \text{secure data}[b_h:]
\]

While this scheme looks ideal, the issue is that to sign a message with \( n \) bits, \( 2nb_b \) bits are needed in a Lamport private key if each bit corresponds to a \( b_b \)-bit value in the private key; \( b_b \) will be referred to as the key range. This means \( 2Nb_b \) bits are needed in the public key where \( b_b \) is the number of bits in the hash, which is greater than the \( n \) bits that this key is able to sign. Although this could be solved by reducing the key size over time, that brings back the issue of there being a limited number of signatures.

Instead, this can be solved by applying the scheme-specific transformation the hash of the key followed by the hash of the data and prepending the transformed result with the new public key, storing the result as the signature. The signature can then be verified by decrypting the transformed part and comparing the hashes for both the new public key and the data. This removes size issues due to the required signature size for a key now being \( 2b_b \), so any key size above that threshold can be used. This modified scheme is shown in Figure 2.

\[ \text{(sk new, pk new)} = \text{KeyGen}(1^N) \\
\text{secure data} = \text{pk new} \ || \ H(m) \\
\text{signature} = T(sk, \text{secure data}) \]

To verify this signature, the following pseudocode can be followed:

\[
\text{secure data} = T_{\text{inv}}(\text{signature}) \\
\text{pk new} = \text{secure data}[N:b_h] \\
\text{assert} \ H(m) == \text{secure data}[b_h:]
\]

Figure 1: Intuitive layout of key replacement signatures

When using a linear list of keys for this algorithm, the range of numbers allowed in the key was tested for \( 2^{16} \) and \( 2^{256} \). The case of \( 2^{16} \) computed very fast, but the \( 2^{256} \) case took too much memory as a result of the code responsible for preventing random numbers from repeating themselves taking too much memory. Fortunately, for a sufficiently large range, it is extremely unlikely for a number to repeat, so after a key range threshold of \( 2^{17} \), this anti-repetition code was disabled.

In this situation, the time complexity is \( O(b_h^2 + n + b_i) \) (proved in the appendix) where \( n \) is the private key size, \( b_h \) is the number of bits in the hash, and \( b_i \) is the input size. This is similar to the Lamport time complexity of \( O(n + b_i) \), and for sufficiently large key sizes, it becomes identical.

In this new scheme, signing a message \( m \) with private key \( sk \) of size \( N \) uses the following pseudocode:

\[
\text{(sk new, pk new)} = \text{KeyGen}(1^N) \\
\text{secure data} = H(pk new) \ || \ H(m) \\
\text{encrypted secure data} = T(\text{signature}) \\
\text{signature} = \text{pk new} \ || \ \text{encrypted secure data}
\]

To verify this signature, the following pseudocode can be followed:

\[
\text{encrypted secure data} = \text{signature}[N:b_h] \\
\text{secure data} = T_{\text{inv}}(\text{encrypted secure data}) \\
\text{assert} \ H(pk new) == \text{secure data}[1:b_h] \\
\text{assert} \ H(m) == \text{secure data}[b_h:]
\]

\[ \text{Figure 2: Ideal layout of key replacement signatures} \]

\[ \text{† A similar scheme was described by Jonathan Katz,}^3 \text{ but unlike that scheme certificate-based key replacement is intended for a series of messages between two parties.} \]
Results and Discussion

The data collected while testing Lamport with key replacement consisted of the security level, running time, and the running time for existing algorithms at the same security level. The times are measured in seconds and were collected for both signing and verifying.

All data were collected on a 3.2 GHz Intel Core i7 processor with 32 GB 2667 MHz DDR4. The Lamport algorithm was self-implemented, while pycryptodome 3.11.0 was used for RSA, DSA (Digital Signature Algorithm), and ECDSA. The software code to reproduce these results can be found at †.

First, Lamport signing with key replacement using SHA-512 (key size of 2048) was compared with RSA, DSA, and ECDSA, with random 32-byte messages. Common RSA key sizes were used with their corresponding security levels, and Lamport key ranges were chosen so that the security level is approximately equal to the corresponding RSA security level, using the formula $S = \log_2(b_h + 2) + b_k - 1$ for $b_h$ bits in the hash and $b_k$ bits per key entry, proved in the appendix. In this case, $S = \log_2(514) + b_k - 1$ = $b_k + 8$, which means the key entry size for a given security level in this experiment is $S$ = $b_k$ - 8. The data is shown in Table 1.

From Table 1, it is clear that using certificate-based key replacement with Lamport signatures takes roughly 50 times as long as DSA and ECDSA, and a varying amount with RSA. The comparison with RSA is significant because it shows that although Lamport key replacement takes longer, it is ultimately faster than what will happen with RSA, so if RSA can be made to work efficiently at a security level of 256, Lamport signatures with key replacement can as well.

The verification data in Table 2 was taken at the same time as the above signing data, using the signatures generated from the signing phase. One important change in RSA to obtain verification data that grows similarly to signing data was changing e from the typical 65537 to $2^{256}$ - 1. As a result, it should be noted that in practice, RSA verification is much faster than the times indicated in Table 2.

Table 1: Signing times for 4 signature schemes at various security levels

<table>
<thead>
<tr>
<th>Security level</th>
<th>Lamport with Key Replacement</th>
<th>RSA</th>
<th>DSA</th>
<th>ECDSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>0.11562</td>
<td>0.0081</td>
<td>0.0027</td>
<td>N/A</td>
</tr>
<tr>
<td>112</td>
<td>0.17746</td>
<td>0.00488</td>
<td>0.0053</td>
<td>N/A</td>
</tr>
<tr>
<td>128</td>
<td>0.29688</td>
<td>0.01487</td>
<td>0.0079</td>
<td>0.0082</td>
</tr>
<tr>
<td>192</td>
<td>0.35206</td>
<td>0.16313</td>
<td>N/A</td>
<td>0.00131</td>
</tr>
<tr>
<td>256</td>
<td>0.53875</td>
<td>0.09614</td>
<td>N/A</td>
<td>0.00236</td>
</tr>
</tbody>
</table>

Table 2: Verification times for 4 signature schemes at various security levels

<table>
<thead>
<tr>
<th>Security level</th>
<th>Lamport with Key Replacement</th>
<th>RSA</th>
<th>DSA</th>
<th>ECDSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>0.03889</td>
<td>0.00331</td>
<td>0.0019</td>
<td>N/A</td>
</tr>
<tr>
<td>112</td>
<td>0.04754</td>
<td>0.00205</td>
<td>0.0063</td>
<td>N/A</td>
</tr>
<tr>
<td>128</td>
<td>0.08333</td>
<td>0.00578</td>
<td>0.00122</td>
<td>0.00159</td>
</tr>
<tr>
<td>192</td>
<td>0.09591</td>
<td>0.07413</td>
<td>N/A</td>
<td>0.00370</td>
</tr>
<tr>
<td>256</td>
<td>0.13701</td>
<td>0.39958</td>
<td>N/A</td>
<td>0.00485</td>
</tr>
</tbody>
</table>

This data shows that certificate-based key replacement with the Lamport scheme is also slower than existing algorithms, being slower than DSA and ECDSA roughly by a factor of 30. Similar to signing, RSA is unique among the schemes because Lamport with key replacement becomes faster than the RSA with a larger e at a security level of 256, and minimum-security levels are bound to increase in the future.

In most cases, Lamport with key replacement is much slower than discrete logarithm-based schemes for both signing and verification. However, this slowness would be acceptable even if the scheme were not quantum-safe, and the quantum safety makes it optimal compared to the other schemes.

Compared with XMSS and SPHINCS, certificate-based key replacement has the immediate advantage of having a smaller key size. The key size is double the size needed for the underlying algorithm, whereas XMSS and SPHINCS have signature sizes that are high multiples of the base algorithm's size. Additionally, this key replacement scheme can be used infinitely, while XMSS can only be used a set number of times, and SPHINCS can only be used a finite number of times unless the tree is extremely large so that few enough input hashes map to each public key.

One disadvantage of certificate-based key replacement with single use signature schemes such as Lamport signatures is the fact that state must be updated every time data is signed, and the size of the updated state is large. This is important, since in a scenario where many parties need to verify a set of signatures, they need to verify the entire list of keys as they would with a certificate path, or there would be a risk of one of the keys being created by an attacker, and used to sign the rest of the chain of keys.

However, this can be solved by using key replacement on SPHINCS instead of a one-time scheme and updating the key every N signatures using a similar certificate pattern, as shown in Figure 3. When doing this, there is a trade-off between tree size and state when a given security level is desirable. This can be seen intuitively by the fact that the most primitive version of SPHINCS would be directly passing to the onetime signature scheme, where a key update must be done after every signature, while with the largest SPHINCS key where every input is mapped uniquely, a key update is never required. Assuming the input is infinitely large, the optimal relationship between key size and signatures per state update.

![Figure 3: Key replacement on a many-time signature scheme](image-url)
where given any unit of size (e.g., bits), $L$ is the total size of the tree, $L_k$ is the signature size, $L_k$ is the individual key size, and $C$ is a security parameter. The proof for this is given in the appendix.

Certificate-based key replacement is safe from both general forgery and existential forgery. It provides non-repudiation, given that the underlying algorithm used, such as Lamport, is also safe from these attacks. Because the signing of the data itself is almost identical to the underlying algorithm, there is no issue there; as a result, it suffices to show that the key replacement is secure. First, existential forgery is not a problem with the key replacement because the key is hashed, so the hash function must be reversed to get any meaningful public key, and even if this was done, more hashes would need to be reversed to get the private key or sign a message; because quantum computers cannot efficiently reverse hashes, this remains true for quantum computers. Certificate-based key replacement is also safe from general forgery because the only new channel for forgery is to use a different replaced key to legitimately sign future messages. Still, because the new key is effectively signed by the underlying algorithm, this remains safe provided that the underlying algorithm is safe. Finally, non-repudiation is provided because claiming that any data was not legitimately signed falls back on the security of the underlying algorithm, either because the key used to sign the message was broken or an earlier key was broken that led to the most recent key not being generated by the intended signer.

## Conclusion

A persistent stream of hash-based signatures with a small signature size can be achieved by replacing keys in a certificate-like manner. When applied to the Lamport signature scheme, this showed reasonable performance with the signing time growing slower than that of RSA signatures. It has the advantage of post-quantum security that is not present in traditional signature schemes. Certificate-based key replacement also has the advantage of having a smaller key size than SPHINCS. It can be used when there is a need for repeated public key signing during communication between two parties. The next step to research is the application of key replacement to many-time signature schemes such as SPHINCS, which may have useful applications such as man-in-the-middle resistant key exchange.

## Acknowledgments

I would like to thank Dr. Dong for helping me learn relevant cryptographic concepts and guiding my research. Dr. Dong was an excellent mentor for me in this project, and this paper would not exist without him. I would also like to thank Polycryptodome for helping structure my research and for providing an excellent platform that fosters research.

## References

7. Fluhrer, S. Reassessing Grover’s Algorithm. 9.

## Author

Matthew Grupenhoff is a senior at Henry M. Gunn High School in Palo Alto, CA. He intends to major in math or computer science, with a focus on cryptography and machine learning.

## Appendix

**Proof of Lamport with Certificate Key Replacement time complexity:**

Assuming the same hash function is used for both the data hash and the hashing of key entries, and this function outputs $bh$ bits, a total of $2h$ public key entries are required, which results in $4bh$ hashes being required to compute the public key. For signing and verification, only half of these are computed, resulting in $2bh$ key hashes required. In all cases, the number of hashes that need to be computed is $O(bh)$. 

**References**

Because a hash with an N bit input requires O(N) time,18 if each key entry is b₂ bits, then the computation of each hash requires O(b₂) time, which means key generation, as well as the part of signing and verification outside the root hashing require O(b₂) time.

When the main hashes are included, a total of b₃ bits must be hashed to authenticate the new key, and b₁ bits must be hashed to sign the input. Because of this, the main hashing component of this signature scheme grows with O(b₃) time. Adding these time complexities, the resulting time complexity is O(b₃+b₂+L). Since the total number of bits in the key is 4b₃, b₃, this is equivalent to O(n+b₃), which is identical to the standard Lamport time complexity. However, the new public key, which has a size of b₃², has not been accounted for in this equation. When that is added, the time complexity is O(b₃²+n+b₃).

Proof of Lamport with Certificate-based Key Replacement security level:

The main way to crack a certificate-based key replacement signature is to obtain the relevant part of a private key at any iteration. To do this, 2h hashes must have preimages computed for a complete private key, where h is the total number of bits that can be handled by the private key (to sign both the next key and the data itself). However, if the goal is only to replace the current signature or to create a different chain of keys, then only n/2 entries are relevant. Of these n/2 entries, the needed preimage for n/4 of the entries, on average, will already be known after a signature is given, leaving one hash per entry for n/4 entries needing a preimage computed. Therefore, let N be the number of hashes whose preimages are needed, which is n/2 if a signature is currently not known, or n/4 if a signature is known.

Assuming each private key entry contains b₁ bits and the hash contains b₃ bits, let b be the minimum of the two; each private key entry can be any of 2b values. The optimal attack would consist of comparing each hash obtained with all N hashes to be reversed before checking another hash, skipping public key entries that have already been calculated; because the average number of attempts required per entry is then 2b/N, the total number of attempts is n²b²(1+3b)+n+4b=2b²((n+1)+b+1)+log₂(N+1)+b+1+4, which becomes log₂(2n+2)+b. Assuming the scenario where this algorithm would be weakest, the security level is log₂(2n+2)+b-log₂(2n+2)+b-log₂4. This simplifies to log₂(2n+2)+b-log₂4.

In cases where the same hashing algorithm is used for both the key/signed data and the individual private key entries, and the key range is greater than or equal to the number of possible hashes (meaning b=n), n²b=2b due to the nature of the signature being equal to the number of bits. As a result, the security level becomes log₂(2b+2)+b-log₂(2b+2)+b-log₂2-2=log₂(b+2)+b-log₂2-2=log₂(b+2)+b-log₂2-2=log₂(b+2)+b-log₂2-2, which simplifies to log₂(b²+2)+b-log₂2-2.

Proof of C formula:

In many-time key replacement, D and N can be varied to adjust the variables. 1/N is the S value here, while L is the total size of a tree in any unit, which is L₈(2D⁻¹)·L₉(2D⁻²). As a result, L=2⁻D(L₈·L₉)·L₁₀⁻¹·L₁₀⁻², so D=2⁻D(L₈·L₉)·L₄⁻¹·L₄⁻². Since there are 2⁻D⁻¹ unique keys in the tree, after n unique keys are used, the probability of key n being a copy of another key is n⁻112⁻¹, and the probability of key n being unique is 1-n⁻112⁻¹=n⁻112⁻¹-n⁻12⁻¹.

As a result, the probability that all keys are unique after n keys are
Entrepreneurial Struggle: A Natural Language Processing Approach

Peter J. Yu
Louisiana School for Math, Science, and the Arts, 715 University Parkway, Natchitoches, Louisiana, 71457, USA; propeter35@gmail.com

ABSTRACT: Although entrepreneurial struggle is an important topic to explore in entrepreneurship research, it has received only limited attention from researchers. Some studies have begun to focus on the relationship between entrepreneurs’ struggle and their mental health. However, there is little understanding of the issues that lead to entrepreneurial struggle. This research explores different areas in which entrepreneurs struggle in their venture-building process. Using a corpus of 10,150 semi-anonymous Reddit posts, this study identifies the areas of entrepreneurial struggle in the venture-building process. A total of 691 posts that contained struggle-related words were collected from the dataset. After applying Natural Language Processing (NLP) techniques, four areas of entrepreneurial struggle were found: product concept and business model; resources; market entry strategy and entry timing; and customer care, service, and communication. The findings of this study indicate that entrepreneurs should be prepared to deal with challenges in these areas in their entrepreneurial endeavors.

KEYWORDS: Behavioral and Social Sciences, Entrepreneur; Entrepreneurship; Entrepreneurial Struggle; Natural Language Processing; Machine Learning.

Introduction

While entrepreneurial activities are the engine of economic development, ninety percent of new ventures fail. The high rate of new venture failure indicates that entrepreneurs need help in the venture-building process. In the entrepreneur's process of building new ventures, struggling with difficult tasks or situations is an important issue, but it has been understudied in contemporary research on entrepreneurship. Recent studies have started to explore entrepreneurial struggle, and more notably, the impact of struggle on mental health. However, despite the obvious importance of studying entrepreneurial struggle, the term has yet to be defined in the literature. In this paper, entrepreneurial struggle is conceptualized as the difficulty entrepreneurs encounter during the venture-building process.

It is critical to investigate struggle-eliciting events as such information would help inform entrepreneurs about potential setbacks they may encounter when trying to start a new business. If not dealt with successfully, those setbacks could lead to entrepreneurial failure, significantly influencing investors’ support for the entrepreneur’s subsequent ventures. Furthermore, venture failure has an impact on the confidence of entrepreneurs themselves. Confidence is a critical force that motivates the entrepreneur’s venture-building effort and increases positive emotions while decreasing uncertainty. With no confidence, entrepreneurs who have struggled in the venture-building process may hesitate to try again.

Although entrepreneurial struggle may be a beneficial learning experience for entrepreneurs, it can be an emotionally negative and traumatic experience. Seventy-two percent of entrepreneurs reported mental health concerns including depression, ADHD, illegal substance abuse, and bipolar disorder. During the venture-building process, entrepreneurs often work in an unpredictable environment that requires many tasks for which they are frequently unprepared. In addition, mental health for entrepreneurs is critically important. To improve our understanding of mental health in entrepreneurship, it is crucial to identify and study struggle-eliciting events in the venture-building process.

As entrepreneurial struggle can lead to business failure and mental health issues, it is a highly relevant topic in entrepreneurship scholarship. However, the limited research on the aspects of entrepreneurial struggle has been mostly post hoc and has yet to explore the struggle-causing events as they happen. Furthermore, the types and nature of entrepreneurial struggle have not been studied in depth. The current study investigates the areas in which entrepreneurs struggle and aims to achieve a better understanding of entrepreneurial struggle across a wide range of business contexts.

Literature Review

Researchers have realized the importance of studying entrepreneurial struggle. For example, past research has examined the relationship between mental health and entrepreneurial failure. Often, mental health issues are a result of struggling through the process of starting a new business. Researchers explored the roles of confidence and emotion in the venture-building process. Failed ventures, which are associated with struggle-related events, can affect entrepreneurs’ emotional and cognitive resilience and confidence. They propose five outcomes that are more likely to exist among confident entrepreneurs than those less confident: positive emotions, emotional resilience, social support from team members, financial resilience, and subsequent ventures. In addition, greater failures in venture building can damage entrepreneurs’ reputations, which can critically impact social and financial capital.

Other researchers investigated the role of depression in entrepreneurial exit. Depression impacts an entrepreneur’s personality, which plays a role in the intention to start a new venture.
business and business success. Furthermore, depression has an impact on self-efficacy. Entrepreneurs with low self-efficacy may perceive complex tasks as threats and negatively evaluate themselves. Low self-efficacy increases the likelihood of exiting the venture-building process. On the other hand, entrepreneurs with high self-efficacy will commit more and persevere in challenging tasks. Out of the 12,293 cases the researchers investigated, there are 2,496 instances of entrepreneurial exit. The researchers found that depression positively correlates with exiting from self-employment, and self-efficacy mediates 32% of the relationship between depression and exit.

Recently, several studies have attempted to use machine learning and Natural Language Processing (NLP) techniques to investigate entrepreneurial mental health using social media. For instance, one study used machine learning on a corpus of 27,906 semi-anonymous posts on Reddit to study entrepreneurial disappointment. A supervised machine learning classification task was performed, and the Logit Boost algorithm was selected to detect disappointment-related posts with an 88% accuracy. The study established five attributions of entrepreneurial disappointment: self-related, norms-related, others-related, entrepreneurship-process-related, and venture-performance-related. It found that internal, global, and stable causes resulted in a higher frequency of depression compared with external, specific, and temporary causes. Although this study provided valuable insight regarding the association between disappointment and entrepreneurship, the factors that lead to entrepreneurs’ disappointment were not investigated.

Using NLP techniques to analyze social media posts on platforms such as Reddit has also been carried out by other researchers. These studies demonstrated the value of this particular research methodology. For example, researchers from a recent study collected a corpus of 22,808 posts on Reddit over a period of 3 months to study anxiety disorders. To classify anxiety-related posts, they generated features of the posts by applying Natural Language Processing techniques, more specifically N-gram language modeling, vector embeddings, topic analysis, and emotional norms. They were able to classify posts with 98% accuracy. Latent Dirichlet Allocation (LDA) topic modeling, an unsupervised feature generation technique, was used to find correlations between specific topics and anxiety. This method could be very effective in identifying struggling-related topics that entrepreneurs run into during the venture-building process.

Another study employed machine learning and NLP techniques to detect the presence of depression in Reddit posts. Specifically, it used a MultiLayer Perception (MLP) classifier, a combination of Linguistic Inquiry and Word Count (LIWC), LDA, and bigram techniques to achieve a 91% accuracy in detecting depression in Reddit posts. The researchers discovered that depression-related posts contained words connected to anxiety, sadness, and other negative emotional states. Although their research successfully detected depression in Reddit posts, it did not investigate the reasons behind the negative emotional states.

In short, social media has become a popular resource for detecting and predicting mental issues and other topics. Studies have used artificial intelligence and computational linguistics to utilize social media datasets. This paper takes inspiration from them to identify the types and nature of entrepreneurial struggle.

### Methods

#### Data Collection:

Reddit is one of the largest social networks for news aggregation, content rating, and discussion. Currently, Reddit has over 50 million daily active users and 100,000 communities. The platform allows users to have a relatively large body of text compared with other social networks. In addition, unlike other social media sites, Reddit offers anonymity to users. This enables candid discussions of the problematic issues, and hardships entrepreneurs deal with.

The “r/startups” subreddit comprises more than one million members who discuss starting new ventures. A total of 10,150 posts were extracted from the “r/startups” subreddit group from January 1, 2020, to October 1, 2022, using the Pushshift API. To search for posts related to entrepreneurial struggle, a word list containing struggle-related words was used to filter through the corpus. The word list is given below:

- anxiety, worry, fret, concern, agitation, apprehension, nervousness, unease, stress, pressure, trouble, struggle

It is important to note that the algorithm used in the study included all the verb tenses where appropriate, which are not shown above. Many of the posts confirmed the importance of user anonymity in this study. For example, one anonymous entrepreneur stated: “I have been dealing with many doubts and am unmotivated. I am thinking about leaving my own startup. I need support from someone in the startup world, and everyone I confront doesn’t fully understand the problem.” Posts like this were viewed and discussed by many Reddit users.

Of the original corpus, 691 posts contained terms related to entrepreneurial struggle. A series of NLP tools were performed to pre-process the dataset. First, URLs, punctuation, and stop words such as “the,” “a,” and “I” were removed from the dataset through the Natural Language Toolkit library in Python because they do not contribute to our ability to understand the substantive issues under study. Next, lemmatization was performed on each post to remove inflectional endings or to return the base forms of words. Finally, lemmatization was used instead of stemming because the context of the word is considered.

#### Topic Modeling:

Topic modeling is an unsupervised statistical modeling technique that identifies topics that describes a set of documents. In this research, Latent Dirichlet Allocation (LDA), a generative topic modeling technique, is used to classify text in the posts into specific topics. To use the LDA topic modeling approach, the study has to specify the number of topics in the model. An LDA model was trained on the dataset to generate latent unlabelled topics characterized by a selected distribution of the top 10 individual words.

Following the data pre-processing, Term Frequency–Inverse Document Frequency (TF-IDF) vectorization was used to transform text into numerical data and highlight the importan-
ce of words in the corpus. TF-IDF is a statistical measure calculated by multiplying the term frequency and inverse document frequency. The term frequency can be found by the raw count of a word that appears in a document. Term Frequency (TF) is captured in equation 1, where \( t \) represents a term and \( d \) is a document.

\[
TF(t, d) = \frac{\text{number of times } t \text{ appears in } d}{\text{total number of terms in } d}
\]

Inverse Document Frequency (IDF) is a statistical measure of how common or rare a word is across a document corpus. The IDF can be calculated by taking the logarithm of the total number of documents and the number of documents that contain the term. To avoid division by zero, one is added to the number of documents that contain the term. In equation 2, IDF is captured, where \( N \) is the total number of documents and \( df \) is the number of documents that contain the term \( t \).

\[
IDF(t) = \log \left( \frac{N}{1 + df} \right)
\]

TF and IDF are then multiplied to find the TF-IDF, represented in equation 3.

\[
TF-IDF(t, d) = TF(t, d) \times IDF(t)
\]

The proposed number of topics in the LDA model was adjusted several times and some of them had unbalanced word distributions and needed to display clear themes. For example, when the proposed number of topics was 6, posts in one topic had keywords that were not closely related to one another, such as market timing and product conceptualization. After experimenting with models that have 3–7 topics, the 4-topic solution was deemed the best as all topics had relatively consistent themes.

### Results and Discussion

#### Latent Dirichlet Allocation Topic Modeling Results and Discussion:

The results of the LDA model when the proposed number of topics was four are shown in Table 1. Most of the top topic words for Topic A are related to finding a product concept or business model. In other words, posts assigned to this topic indicate entrepreneurs needed help understanding the market, the needs of customers, and the product or service they should offer. For Topic B, the top topic words primarily hinted at a struggle for resources. These words show that entrepreneurs in this category are looking for solutions regarding funding, the time needed for the business, human resources, etc. Topic C has top words related to the struggle of deciding the best market entry strategy and entry timing (e.g., plan, time, and start). The “time” topic word for Topic C differs from the one in Topic B because it relates to the optimal timing of marketing actions such as when to enter the market. These entrepreneurs seem to need help coming up with a plan to enter the market with the right approach for the target market. Finally, the top words for Topic D are associated with customer care, service, and communication.

Of the 691 posts containing struggle-related words, 321, 251, 102, and 17 were assigned to Topic A, B, C, and D, respectively. The distribution of each struggle category can be seen in Figure 1.

#### Sample Posts:

To help illustrate how entrepreneurs discuss their struggles in the four topics, this section provides some sample posts from the four topics. As discussed earlier, Topic A is about entrepreneurial struggle regarding the business model or product idea. This is evident in the post below:

“Sometimes life can be hard or too much, and you need someone who isn’t a part of your life to talk to. Therapists, friends, and family want to fix you. But, sometimes, finding someone who wants to listen to or talk with you is hard. I really want to start a business that helps people and gives them someone to talk to in these times. It helps those with social anxiety to have a friend. I want to offer some people free trials in the beginning because I want to know what they would want that would make it better for them. So, if anyone is interested or you know anyone that would be interested let me know. To be clear, I am not a therapist or a mental health professional and am just wanting to start a business for people to be able to pay for convenient friendship.”

In the post above, the would-be entrepreneur had a vague idea of a potential product for their intended market but needed to know how attractive the product would be. Thus, they reached out to the audience on Reddit to discuss. This person apparently has a strong desire to understand the target market and tries to provide a service that could meet the market needs. This pattern could be observed in many posts in the Topic A category.

Posts in the Topic B category reported resource struggles (e.g., money, time, and human resources). These entrepreneurs are putting in the effort to assemble the necessary resources to build their businesses but seem to need more of them. For example, in the sample post below, one entrepreneur had difficulty in preparing for pitching the business to potential investors:

“Hi everyone! We’re a toy startup preparing for our first investment round. Honestly, my team and I are nervous about it,
and I wanted to ask for advice from people with actual pitching experience. How did you get in touch with funds? Did you have a warm intro? What’s the most important in your opinion?”

Market entry strategy and timing were discussed by entrepreneurs who requested help with posts in Topic C. These entrepreneurs need to figure out the optimal market strategy, including market timing. For example, the person that posted the request below was trying to figure out the appropriate price for a new app, with consideration for the affordability of the product in different geographical locations.

“I have recently launched an app and have struggled with finding the right price to charge the users. I know that people from the USA and Europe can pay higher prices, but I feared that it was too expensive for the rest of the world. I looked into the ability to charge different prices based on users’ country and could not find a product like that other than for WooCommerce. My website is built from the ground up so that is not an option. My question to ask is whether you guys already do this. Do you know of a web app that does it? Would you use something like this?”

The final topic area, Topic D, includes posts regarding providing the best customer care, service, and communication. One entrepreneur, who had the goal of targeting senior customers, struggled with communicating with the targeted group:

“Hi, all. About a month ago, I launched a product for seniors and retired people to get 100 users in 100 days. Are there any suggestions on how I could get those 100 users? I have two personas: people aged 65+ with specific habits and their 35–45-year-old children, who can also sign up for the service for their parents. To test product-market fit, I’m trying to test a few channels: google ads (10% CTR), Facebook groups (very positive comments, dozens of clicks, but no traction), forums, and emails. I managed to scrape hundreds of emails from several communities and launched a cold email campaign. There was a 60% open rate and many clicks, but no results. I also initiated some partnerships with websites with a senior audience in order to get referral traffic: crosswords, sudoku, knitting, and gardening. A few answered back and one of them accepted. Long story short, I’m struggling to find potential clients for the service. Facebook ads are an option of course, which I will test, but my budget is not unlimited. Any ideas that could help? Thank you so much for your support!”

Discussion

While previous research has investigated some negative aspects of the venture-building process such as entrepreneurial anxiety, there is little research on the types and nature of entrepreneurial struggle. This research suggests that entrepreneurs need help in four primary areas: product concept and business model; resources; market entry strategy and entry timing; and customer care, service, and communication. Most of the struggles are in the marketing area. This points to the difficulty entrepreneurs often have in connecting what they contemplate offering to the market and the market itself. Past research has not sufficiently investigated entrepreneurs’ struggles in the marketing area and the constraints of limited resources. Topics such as product model, resources, market strategy, and customer service are explored for the first time in the current study. It could be argued that marketing and assembling resources to make the new venture a reality are the two main factors hindering entrepreneurs’ efforts to become successful.

Based on these findings, entrepreneurship education programs should focus on the four struggle areas identified in this study to prepare entrepreneurs better before starting ventures and during the venture-building process. Teaching entrepreneurs in areas such as marketing and fundraising is most needed. This will help entrepreneurs acquire knowledge and skills critical to starting companies and improve the likelihood of new business success. On the other hand, new business owners and aspiring entrepreneurs should understand that they must develop skills in the four areas identified.

Although previous studies have, to some extent, researched the affective aspect of entrepreneurial struggle⁸, the term “entrepreneurial struggle” until now has been used without an explicit definition. In this paper, we define entrepreneurial struggle as the difficulty entrepreneurs encounter during the venture-building process. This definition allows scholars to distinguish between entrepreneurial struggle and events that may not be related to the entrepreneurial process, such as personal or family issues. Furthermore, more clarity is established for future entrepreneurship research by defining entrepreneurial struggle this way.

Intense mental illness symptoms can arise because of struggle-related events entrepreneurs face.⁵ The resulting emotional response after entrepreneurs face challenges may lead to depression and unproductive behavioral changes. Research has found that entrepreneurs may stop socializing and isolate themselves when coping with negative emotional responses.⁵ Conversely, entrepreneurs who expressed disappointment because of loneliness had the desire to maintain meaningful relationships. The behavioral response in entrepreneurs, when faced with struggle-related events, is not beneficial for their mental health.⁵ This further points to the importance of helping entrepreneurs deal with the struggle areas identified in this research as doing so will have a direct positive impact on the mental health of entrepreneurs. For example, if given the opportunity, an entrepreneur who has never had experience in communicating with customers could improve the skills required in this area and have a better experience in the entrepreneurial journey. Therefore, this enhancement in skills would lead to an improvement in the entrepreneur’s mental health.

Limitations and Future Research:

The findings and limitations of this study have important implications for future research in entrepreneurship. Although several types of entrepreneurial struggle were identified in this study and their relations with the affective experience of entrepreneurs pointed out, it still needs to be made clear whether entrepreneurs with mental health issues encounter more struggle during the venture-building process. In other words, it remains to be investigated whether mental health issues and entrepreneurial struggles actually influence each other. Furthermore, while the results of this study show that entrepreneurs primarily struggle in the four areas, the extent of
the effect of the struggle-related events on new venture success should have been researched. Further research is required to investigate this issue as some areas of struggle may lead to new business failure more than others. While Reddit provides anonymous posting and allows for researching topics associated with a stigma or shame for those posting messages, it does not let researchers examine the entrepreneurs’ previous venture experiences, industries, and other backgrounds. These factors may directly impact the extent to which entrepreneurs experience potential struggles and how they deal with struggles. When posts were extracted from Reddit, only the body text of posts was collected. Any information that could reveal the post’s author (i.e., username) was not collected. The identities and descriptions of the entrepreneurs were kept unknown throughout the study.

Furthermore, the Reddit user base is different from the general population. According to Pew Research Center, the Reddit user base in the United States is majority male (67% of users), white (70% of users), and young (64% of users are under 29 years of age). Additionally, Reddit users have higher levels of digital literacy and web access. Therefore, this study may have yet to capture entrepreneurs with different characteristics.

Also, the mood and circumstances under which the entrepreneurs made the posts could influence the contents of the posts themselves. For example, it is possible that some entrepreneurs posted during a night of poor sleep or a cold. Future studies should make efforts to identify these conditions and their associated effects. Additionally, when a team of entrepreneurs starts new ventures, more than one individual entrepreneur can encounter difficulty building a new venture. The interactions and relationships among a team of entrepreneurs may influence decisions when building new ventures. These decisions may be crucial when acquiring resources and meeting public expectations. Exploring struggle among a team of entrepreneurs is a valuable extension of entrepreneurial struggle research.

Finally, future research can examine the effects of entrepreneurial struggle on the personal lives of entrepreneurs. Although the struggle areas identified in this study impact the venture-building process, more investigation is needed to examine the relationship between struggle and entrepreneurs’ life and career paths. In addition, further research is required to explore how entrepreneurs constructively grow and build resilience following entrepreneurial struggle and their inclination to continue being an entrepreneur.

**Conclusion**

In this study, 10,150 semi-anonymous posts were extracted from Reddit to examine the struggles of entrepreneurs. After implementing NLP techniques, four primary struggle areas that entrepreneurs encountered were identified: product concept and business model; resources; market entry strategy and entry timing; and customer care, service, and communication. As entrepreneurial struggle is connected with entrepreneurial failure and the mental health of entrepreneurs, there are numerous directions for future research in this area. It is also essential to use the findings of this research to create educational and training materials to help entrepreneurs better prepare for the inevitable trials and tribulations during their venture-building endeavors. Organizations, government agencies, and educational institutions that provide entrepreneurship education and assistance programs should especially pay attention to the four struggle areas identified in this paper in their programs to help entrepreneurs cope with the challenges they encounter.

**Acknowledgments**

I want to acknowledge the mentorship of Professor Brad Burkman of the Louisiana School for Math, Science, and the Arts. Furthermore, I am grateful to Dr. Lidan Fan of the University of Kansas, Dr. Long Pham of Texas A&M University, and Dr. Javier Flores of Emporia State University for their comments that helped improve this paper. In addition, I would like to thank my instructors at the Louisiana School of Math, Science, and the Arts: Dr. Zheng Chen for motivating me to grow my knowledge in STEM; Professors Randy Key and David Andersen for inspiring my interest in mathematics; and Professor Sanjeetha Peters for encouraging me to pursue computer science. Finally, I am grateful to my parents for their support.

**References**


DOI: 10.36838/v5i4.15


**Author**

Peter Yu is a sophomore at the Louisiana School for Math, Science, and the Arts in Natchitoches, Louisiana. He participates in competitive swimming and plays the piano. In the future, he plans to pursue computer science in college.
Association of Social Determinants of Health with Respiratory Disease: A County-by-County Analysis for Ohio

Rohan M. Desai
Hawken School, 12465 County Line Rd, Chesterland, OH 44026, USA; desairohan2005@gmail.com
Mentors: Raoul R. Wadhwa, Jacob G. Scott

ABSTRACT: A recent research focus on social determinants of health has revealed their importance in the healthcare setting. Even prior to the COVID-19 pandemic, about 20% of United States zip codes experienced economic distress, marked by high poverty rates and unemployment. These areas are particularly vulnerable to adverse clinical outcomes. Here, we studied clinical outcomes of respiratory diseases in the state of Ohio, specifically COVID-19 and lung cancer, and how they were affected by social determinants of health, specifically median county-level income. Three publicly available datasets were merged, and associations between pairs of variables were evaluated using Spearman’s rank correlation coefficient and Student’s t-test. Our analysis found that clinical outcomes of both respiratory diseases were significantly correlated with median county-level income (p < 0.001). This implies that although variation between disease outcomes exists across Ohio, social determinants of health play a common, underlying role. As a result, public health intervention across Ohio is necessary to support health equity. Similar studies should be undertaken for other states and territories within the United States to determine the need for public health intervention across the country.

KEYWORDS: Biomedical and Health Sciences; Other; Cancer; COVID-19; Social Determinants of Health.

Introduction

Despite significant improvements in diagnosis and management of sub-acute and chronic diseases, there continue to be challenges of unexplained residual risk accounting for disparate outcomes. In recent times, community-level socioeconomic distress has been identified as a potential driver of outcomes following many common diseases, including common respiratory illnesses like chronic obstructive pulmonary disease and lung cancer.¹³ Unfortunately, a fifth of United States zip codes currently demonstrate social and economic distress, including high poverty rates, joblessness, and economic recession.⁴ Measures of community distress such as low education levels, low median household income, high rates of poverty, and air quality are important determinants of its residents’ overall health and well-being.⁵⁻⁶ Consequently, patients with respiratory diseases living in socioeconomically distressed neighborhoods have a higher risk of adverse clinical outcomes.

Since 2019, the world has been dealing with the novel coronavirus disease 2019 (COVID-19), which predominantly manifests with acute respiratory symptoms. Interestingly, the aforementioned social determinants of health (SDOH) might play a key role in determining outcomes, even considering a relatively novel and acute disease like COVID-19. The state of Ohio represents a bell-weather state in the USA with a blend of socioeconomic strata, political ideologies, and urban versus rural areas. Whereas most published research uses national-level data that potentially obfuscate associations due to the presence of unrelated variables, such as varying culture and diet, using state-level data permits a more reasonable assumption of homogeneity with regard to cultural variables. We hypothesized that, in the state of Ohio, there was a significant association between household income, an important surrogate for SDOH, and acute (COVID-19 infection) or sub-acute (lung cancer) respiratory illnesses. Furthermore, we predicted that higher household income would be associated with lower COVID-19 fatality and lower incidence of lung cancer.

Methods

Median county-level household income was used as a proxy for social determinants of health. The severity of COVID-19 was measured by case fatality, defined as the percentage of confirmed cases of COVID-19 with a fatal outcome. Lung cancer incidence was defined as the number of newly diagnosed cases per 100,000 members of the population. The proxy variables were chosen based on reasonable relation to the outcome we intended to measure⁷⁻¹⁰ and based on the availability of datasets. Association between these variables was measured with Spearman’s rank correlation coefficient, which measures the association between two quantitative variables based solely on order instead of magnitude. The statistical significance of the pairwise correlations was evaluated with Student’s t-test, which permits statistical testing of a predetermined hypothesis. The false-positive threshold was set to 5%, or 0.05; a two-tailed test was used in accordance with the principle of conservatism. Although relying on clinically relevant data, the study only used publicly available and aggregated datasets. As such, it is exempt from IRB approval under category 4: secondary research of publicly available data without identifiable information.

The current study utilized 3 publicly available datasets: (1) county-level median income, which was collected from the United States Census Bureau; (2) county-level lung cancer incidence, which is publicly available on the National Cancer Institute website; and (3) county-level COVID-19 case fatality.
ratios, downloaded from Johns Hopkins University's publicly available repository. All datasets contained complete information for Ohio's counties (n = 88) and were merged by county name with no name conflicts or missing values. Data manipulation, visualization, and the statistical inference were conducted in the R programming language. The R code used to conduct the statistical analysis for this manuscript is available via a publicly accessible GitHub code repository at the following link: https://www.github.com/desairohan2005/covid-lung-sdoh.

Figure 1: Distribution of median income, COVID-19 case fatality ratio, and lung cancer incidence in counties of Ohio. Higher (lighter color) median county-level income (left) visually corresponds to lower (darker color) county-level COVID-19 case fatality ratio (center) and lower (darker color) county-level lung cancer incidence (right).

Results and Discussion

Results:

Median county-level income across Ohio was $52,723 (IQR: 47,274-59,167). As a result, the median county-level lung cancer incidence was 68.4 (IQR: 62.0-80.6) per 100,000 population, and the median COVID-19 case fatality ratio was 1.98% (IQR: 1.65-2.36). The distribution of the above parameters on a county-by-county basis is shown in Figure 1. A relationship between median income (Figure 1, left) and both of the SDOH variables (Figure 1, center and right) is visually apparent. Specifically, counties with a higher median income (lighter color) also have decreased COVID-19 case fatality (darker color) and decreased lung cancer incidence (darker color). Data for 7 counties with the highest and lowest median income each are shown in Table 1. There also appears to be a clear pattern within the numerical data in this table. That is, the 7 counties with the highest median incomes tend to have lower values for COVID-19 case fatality and lung cancer incidence in comparison with the 7 counties that have the lowest median incomes in the state of Ohio.

The patterns deduced from the figure and table were confirmed quantitatively. The two SDOH variables, lung cancer incidence and COVID-19 case fatality were not significantly correlated (r = -0.20, p = 0.1). In contrast, median income was significantly correlated with both lung cancer incidence (r = -0.42, p < 0.001) and COVID-19 case fatality (r = -0.49, p < 0.001). This implies that 24% of the variation in lung cancer incidence and 18% of the variation in COVID-19 case fatality ratio in counties across Ohio was explained by median county income. Scatterplots of all three pairwise associations are included as supplemental figures at https://www.github.com/desairohan2005/covid-lung-sdoh.

Significance of Results:

We have demonstrated that as the median income of an Ohio county increases, the likelihood of dying due to COVID-19 and being diagnosed with lung cancer decreases, suggesting that SDOH plays a crucial role in clinical outcomes of 2 distinct respiratory diseases, one acute and another subacute/chronic. Interestingly, COVID-19 case fatality and lung cancer incidence were not significantly associated, thereby reducing the likelihood of confounding. Previous studies have explained how management, in addition to incidence, of lung cancer is affected by SDOH.²³ We validated these studies at the state-level and demonstrated the need for public health intervention.

We also encourage incorporating SDOH consideration into clinical care. This can be done qualitatively and quantitatively within healthcare settings. For instance, qualitative interventions would include emphasizing health counseling by primary care physicians. Although primary care physicians serve large patient populations, they can focus their efforts based on the specific SDOH vulnerabilities present in the area they serve. On the other hand, quantitative interventions could include the incorporation of SDOH variables into clinical decision-support tools. In particular, risk calculators have gained traction within clinical care. Yet, despite the clear effect of SDOH on clinical outcomes, no risk calculators currently use SDOH variables in predicting outcomes. Our study clearly indicates that this is a limitation that future risk calculators should overcome.

Table 1: Counties in Ohio with the highest and lowest median incomes (seven of each). Corresponding values for lung cancer incidence

<table>
<thead>
<tr>
<th>Rank</th>
<th>County</th>
<th>Median Income ($1000s)</th>
<th>Lung cancer incidence (per 100k)</th>
<th>Risk of lung cancer relative to median</th>
<th>COVID-19 case fatality (%)</th>
<th>Risk of fatality relative to median</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Delaware</td>
<td>104.3</td>
<td>51.4</td>
<td>0.75</td>
<td>0.71</td>
<td>0.36</td>
</tr>
<tr>
<td>2</td>
<td>Warren</td>
<td>87.1</td>
<td>62.9</td>
<td>0.92</td>
<td>1.24</td>
<td>0.63</td>
</tr>
<tr>
<td>3</td>
<td>Union</td>
<td>82.8</td>
<td>66.2</td>
<td>0.97</td>
<td>0.85</td>
<td>0.43</td>
</tr>
<tr>
<td>4</td>
<td>Seagea</td>
<td>80.0</td>
<td>48.0</td>
<td>0.70</td>
<td>2.20</td>
<td>1.11</td>
</tr>
<tr>
<td>5</td>
<td>Medina</td>
<td>74.2</td>
<td>56.9</td>
<td>0.83</td>
<td>1.73</td>
<td>0.87</td>
</tr>
<tr>
<td>6</td>
<td>Greene</td>
<td>67.1</td>
<td>56.5</td>
<td>0.83</td>
<td>1.62</td>
<td>0.82</td>
</tr>
<tr>
<td>7</td>
<td>Clermont</td>
<td>67.0</td>
<td>75.9</td>
<td>1.11</td>
<td>1.24</td>
<td>0.63</td>
</tr>
<tr>
<td>82</td>
<td>Highland</td>
<td>43.3</td>
<td>73.5</td>
<td>1.07</td>
<td>1.74</td>
<td>0.88</td>
</tr>
<tr>
<td>83</td>
<td>Pike</td>
<td>42.8</td>
<td>99.0</td>
<td>1.45</td>
<td>1.42</td>
<td>0.72</td>
</tr>
<tr>
<td>84</td>
<td>Morgan</td>
<td>42.3</td>
<td>80.5</td>
<td>1.18</td>
<td>2.20</td>
<td>1.11</td>
</tr>
<tr>
<td>85</td>
<td>Gallowa</td>
<td>42.1</td>
<td>83.7</td>
<td>1.22</td>
<td>2.06</td>
<td>1.04</td>
</tr>
<tr>
<td>86</td>
<td>Athens</td>
<td>40.9</td>
<td>74.4</td>
<td>1.09</td>
<td>1.13</td>
<td>0.57</td>
</tr>
<tr>
<td>87</td>
<td>Scioto</td>
<td>39.7</td>
<td>86.7</td>
<td>1.30</td>
<td>1.59</td>
<td>0.80</td>
</tr>
<tr>
<td>88</td>
<td>Adams</td>
<td>39.1</td>
<td>93.6</td>
<td>1.37</td>
<td>2.25</td>
<td>1.14</td>
</tr>
</tbody>
</table>

Strengths and Limitations:

Data reliability is a strength of this study. All the data sources were established organizations that make their data publicly available. This not only increases data reliability by opening it to public critique but also increases the reproducibility of the study. Thus, researchers and reviewers of this study can themselves replicate the results to validate quantitative outputs. Notably, as time passes, researchers can also replicate this study in the future to determine whether the effects of SDOH over time are changing. This feature will be particularly useful after implementing public health interventions to determine whether they have had a valuable impact on healthcare outcomes.
A limitation of this study is the county-level aggregation. Many SDOH studies have been conducted at the national level, which limits their application to individual states in a large, heterogeneous country like the United States. Although our study limits this issue by focusing on county-level data within a single state, counties themselves still possess a measure of heterogeneity that is lost by aggregation. However, this aggregation is a necessary limitation, as having less aggregation would risk confidentiality. Additionally, publicly available data is generally limited to county-level or zip code-level aggregation, both of which contain group sizes of similar magnitude.

Future Directions:

We are particularly interested in two extensions of our study. First, this cross-sectional analysis of Ohio should be repeated individually for other states and territories within the United States. Given the level of funding and autonomy provided to state governments, individual studies are warranted as public health interventions could be more effective if customized to target the SDOH vulnerabilities in each state. Second, this cross-sectional analysis could be extended to a longitudinal one. At the same time, we evaluate the association between SDOH and clinical outcomes at a specific point in time; these associations represent complex relationships that will likely evolve over time due to a multitude of reasons, such as legislative changes or public health programs. Thus, comparing our results to the associations a few years in the future will provide valuable insight into understanding whether health inequity is growing or decreasing over time within the state of Ohio.

Conclusion

We have shown that social determinants of health are significantly associated with respiratory disease-related health outcomes within the state of Ohio. For example, patients diagnosed with COVID-19 are more likely to die because of it when they reside in a lower-income county. Additionally, people living in lower-income counties are more likely to be diagnosed with lung cancer. Although the numerical results might vary across the United States, it is likely that our results generalize across the United States. Therefore, public health interventions are necessary to manage the potential detriment to health caused by social determinants and address healthcare inequity.

Acknowledgments

The authors thank the Hawken School, Cleveland Clinic, and Case Western Reserve University for their support. This was a no-cost, unfunded study.

References


Authors

Rohan M. Desai is a high school student in Ohio interested in battling healthcare inequity in the United States. He will attend Vanderbilt University, where he will major in Medicine, Health, and Society.

Raoul R. Wadhwa, MD, recently completed medical school in Ohio. He is interested in primary care medicine.
Automated Debugging System: Implementing Program Spectrum Analysis and Information Retrieval on Fault Localization

Yin-Siang Hao
Taipei Wego Private High School, No.50, Zhuhai Rd., Beitou Dist., Taipei City, 11254, Taiwan (R.O.C.); samyhao0507@gmail.com
Mentors: Yi-Ting Liao, Sun-Yuan Hsiehs

ABSTRACT: Debugging is often the most time-consuming phase during program development, lengthening the development time and lowering efficiency. Even though there are currently existing tools that help raise debugging efficiency, their functions are largely limited, as they only present a more comprehensive analysis behind the compiling and running processes to save complicated steps in debugging; however, developers are still required to test and verify the reasonability of every line of code. Unlike the current manual debuggers, this research aims to build a system that automatically detects bugs in the programs through program spectrum analysis and information retrieval. In the section of program spectrum analysis, the system will statistically analyze every code block’s probability of containing bugs in the input source code file according to the provided test cases, later forming an initial suspiciousness ranking based on previous calculations. Afterward, the system retrieves historical files that resemble the current source code and uses their suspiciousness rankings to modify the initial suspiciousness ranking, generating the final suspiciousness ranking as the system’s output. This research integrates the two techniques and optimizes the formula of forming the suspiciousness values in program spectrum analysis and the comparison mechanisms in information retrieval, reaching performances of higher comprehensiveness and precision.

a. The word “bugs” here refers to logic errors that cause wrong answers or runtime errors, not syntax errors that lead to compile errors.

KEYWORDS: Systems Software; Algorithms; Program Analysis, Natural Language Processing, and information retrieval.

Introduction

Background:
In recent years, computer science and programming have become popular fields due to their infinite potential for innovations as well as their significant contributions to the planet. However, while developing projects ranging from single-file codes to cross-platform software, most programmers are struggling to debug – a process to find the errors that occurred in the codes and resolve them – and it increases inefficiency and time consumption during the process.

Computer programming can be divided into four phases: identifying problems, finding solutions, coding them, and debugging.¹ Relative to the first three phases, debugging often makes the least number of changes, yet it usually requires the longest time length. According to the CVP survey (Figure 1), debugging has accounted for 50% (312 billion US Dollars) of the global software developer wages, equivalent to the wages for designing and building programs.²

Currently, multiple tools can assist developers to find bugs, such as debuggers and reversible debugging software. By setting breakpoints and limitations in the debugger mode, users can stop at lines that they think are suspicious, track the execution routes, and monitor the changes of variables and memory allocations line by line.³ On the other hand, reversible debugging software records all the memory access, computations, modifications to variables, as well as calls to the operating sys
tems. By moving forward and backward among lines, the users can inspect the reasonability of the current program states and identify the errors in the codes.⁴

Motivation:
Although there are techniques that help developers to detect bugs more effectively, they still require developers to evaluate the correctness of each program state manually. Currently, although there is research discussing the validity of fault localization via spectrum analysis and information retrieval, the accuracy and effectiveness are not as significant as expected, and no tools are developed to automatically indicate potential bugs for users.

Figure 1: Impact of debugging on time spent developing code and its cost in terms of wages per annum.⁴
Purpose:
Due to the lack of automatic debugging tools, this research aims to develop an automated debugging system that automatically locates potential bugs in given programs. The following are three primary objectives of this research:
1) Save the effort of the developers on assessing and locating bugs
2) Develop algorithms that can implement the functions of this system
3) Enhance the accuracy of the automated debugging outcome

System Architecture:
In the program spectrum analysis section, the buggy source code file is first inputted and processed. A new file is created by integrating the original codes and a line-tracing mechanism. Next, it is compiled and executed with input and output files provided by the users, and the system generates a coverage matrix consisting of m block-hit spectra, each recording whether the code blocks are traversed in the execution or not. The error log records the result of every execution. By comparing the similarity between the coverage matrix of each code block and error log, the system measures the suspiciousness values, or the probability of containing bugs, of the code block – the higher they resemble each other, the more possible that the code block is the reason that leads to failures and contains bugs. Finally, an initial ranked list of code blocks is generated based on the suspiciousness values.

In the information retrieval section, the input source code is vectorized into a term vector through the process of TF-IDF. It is then compared with the TF-IDF vector of every dataset, a collection of similar source code files, in the historical file database. Of all datasets, one dataset with the highest relevance score is chosen. In the dataset, all historical source code files are vectorized into term vectors through TF-IDF and compared with the term vectors of the input source code file, where a relevance score is generated. The final buggy code block ranking of each historical source code file and then alters the initial ranked list of the input source code file to the extent proportional to the relevance score between the historical and input files. Finally, a final ranking of the current code file is generated and outputted, indicating the suspiciousness values of each code block. (Figure 2)

Related Works
Since this research is based on a variety of algorithms and theories, the essential concept of each should be briefly explained to avoid further ambiguity. Multiple papers in support of this research are cited in this section for higher thoroughness and authority.

Program Spectra:
The program spectrum represents different perspectives toward a program and focuses on different features during program executions.⁵ The two types of spectra are hit and count, of which the former only records true or false, and the latter records the number of times the spectrum is executed. Branch spectra only record the steps regarding conditional statements, such as “if”, “for”, and “while.” Complete-path spectra track the complete routes of execution, including conditional branches, loops, and statements. Different from the complete-path spectra, path spectra only record partial paths based on an acyclic control flow graph, exclusive of any loops. Different from the normal control flow graph, the acyclic eliminates the back edges that form loops, becoming loop-free. Data-dependence spectra record definition-use pairs, each of which has the form (d, u), respectively meaning the definition statement of a variable, statements using the variable, and variable name. Output spectra save the output of the execution. Similar to complete-path spectra, execution-trace spectra record the entire route; yet, the main difference between them is that execution trace includes real codes, whereas complete-path spectra only contain line numbers.⁶ Block spectra form program blocks that compound statements.⁷ For example, statements under if or else are included in the same block because they are always run together under an execution. Table 1 illustrates the spectra, including its profiled code lines, and execution records based on hit and count, for example, the program Number of n Digits of Figure 3.

![Figure 2: The algorithmic process of the implemented automated debugging system](image)

![Figure 3: Example code Number of n Digits and its control flow graph](image)

Table 1: Spectra for Number of n Digits of Figure 3
Spectrum-Based Fault Localization:
Spectrum-based fault localization, or SBFL, evaluates the suspiciousness of every program block. This technique measures the frequency at which each block is executed during failed executions, and this number of frequencies is seen as the suspiciousness value of this program block. There have been different program spectra proposed for this technique, and the most commonly used is block-hit, due to the high availability of its result and the low cost of collecting them. In a process of spectrum-based fault localization, provided test cases are used for the execution of the source code program. During every execution, program blocks are marked with dots if they are executed, as shown in Figure 4. This forms the coverage matrix, which has a row number equal to the number of test cases and a column number equal to the number of program blocks, as shown in Figure 5. After the execution, the system will get a list of outcomes, each representing the success or failure of execution, known as the error vector. In Figure 4, the error vector is in the last row labeled “Execution results.” The process can be shown in another form with only matrices, presented in Figure 5. M spectra indicate the number of runs of the program, and N stands for the number of code blocks. The M runs generate M results, each recorded either with 0 for successful (no errors) or 1 for failed (with errors), recorded in the error vector.

Figure 4: An example of spectrum-based fault localization 5

<table>
<thead>
<tr>
<th>ID</th>
<th>Program block</th>
<th>T</th>
<th>T</th>
<th>E</th>
<th>E</th>
<th>N</th>
<th>S</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 5: The Coverage Matrix and Error Vector 7

The purpose of calculating the suspiciousness value of every code block is to find to what extent a block reflects the error vector in the M runs. The more closely they resemble each other, the more probable of the block being the bug. This deduction is based on the fact that if the block is involved in executions that turn out to be the failed ones, it may be the factor that leads to the program’s failure, thus being where the bug is located. Measures for the similarity between the vector of \( x_{ij} \) to \( x_{Mj} \) and the error vector are called similarity coefficients. There are four kinds of similarity coefficients, each calculated through a different formula, namely Jaccard, Tarantula, AMPLE, and Ochiai.

TF-IDF:
TF-IDF, which stands for “term frequency-inverse document frequency,” evaluates the importance of a word in a document based on its occurring frequency in a document and the corpus. As seen in the mathematical definition below, it is the product of term frequency and inverse document frequency.

\[
tf(t,d, D) = \frac{\text{frequency of } t \text{ in } d}{\text{total number of words in } d}
\]

Inverse document frequency indicates the universality of a word in the corpus. Mathematically, it is calculated by dividing the total number of documents in the corpus by the number of documents with the term t included. The lower this number is, the more common t is, and vice versa.

Cosine Similarity:
Cosine Similarity is an approach that calculates the similarity between two documents. The documents are presented in the form of vectors, with each value representing the term frequency of a word. Then, the cosine formula of vectors is applied to the measurement of the distance between two vectors.

\[
sim(a, b) = \frac{a \cdot b}{\|a\| \cdot \|b\|}
\]

In the formula, the similarity is calculated by dividing the product of vectors a and b by the product of the length of the two vectors. The length of a vector is measured using the Euclidean norm. It is defined as the square root of the sum of the square of every vector component.
Information Retrieval-Based Fault Localization:

Information Retrieval-Based Fault Localization, or IRFL, aims to find out a ranked list of program elements based on their probability to be bugs. Throughout the process, it uses bug reports, documents that contain specific information about the failure of a program, to generate textual similarities with each program element, such as "for," "if," and "while," and rank them using these relevance scores. The technique that most system uses to calculate relevance scores is a combination of TF-IDF and cosine similarity. In the TF-IDF section, the compared documents (bug report and program element files) are changed into a vector of numbers, each representing the importance of every word. Then, the cosine similarity formula will be applied to calculate the distance (in this case, the similarity) between vectors.

Two major relevance functions carry out document comparisons, respectively direct and indirect relevance functions. In the direct relevance function, the relevance score between the current bug report and each program element file is calculated, creating an initial ranking of program element files. In the indirect relevance function, the current bug report is first compared with every history bug report related to the current case, with its relevance score calculated. Then, the system finds the program elements fixed in every history bug report, and multiplies the relevance score between the history bug report and fixed elements to the previous relevance scores. This final score turns out to be the indirect relevance score between the current bug report and those fixed elements.¹⁴

Combining the results of the two relevance functions, the system will generate an ultimate program element ranking that indicates their ranked suspiciousness to be bugs, as shown in Figure 6.

Methods

This research mainly focuses on the debugging of the C++ programming language and uses C++ as the language for the implementation of the automated debugging system.

Facilities:

The hardware used in this research includes a laptop (CPU: 2 GHz Quad-Core Intel Core i5 with Intel Iris Plus Graphics, Memory: 16 GB, SSD: 512GB) for researching, developing the automated debugging system, conducting and analyzing experiments, as well as a notebook for recording the experimental data. The development environment for the automated debugging system is Visual Studio Code, and the programming language for development is C and C++

Program spectrum analysis:

In this section, the system aims to generate a suspiciousness sequence that indicates the suspiciousness of code blocks to contain bugs. The most important components in the measurement of suspiciousness values are the execution paths, which constitute the coverage matrices, and the execution result, which constitutes the error vectors. The process can be divided into three main steps: Modifying the source code file to make it fit the following operations, executing the source code file, and analyzing the data collected in the previous executions.

Since the initial source code files don’t have the functions of collecting execution path and result, a new file is created by the system, including the initial codes along with additional mechanisms, as shown in Figure 7.

Figure 7: The contrast between an input source code file (left) and its modified source code file (right)

__LINE__ is added to the end of every available line. It is a preprocessor macro that provides the line number of the current statement.¹⁵ The variable bracket is declared to record the level of curly braces the current statement is wrapped in. A line is available to add __LINE__ when bracket > 0, or the current statement is wrapped in at least one curly braces pair, and when the statement is not ended with a closing curly brace. During every execution, the value of __LINE__ is recorded in route.txt, which will contain a complete execution path when the execution finishes.

In the program, the modified source code file is executed when pairs of input and expected output files are provided. To compile and execute with the code file, the function system() is needed. It invokes the command-line interface to execute commands given as the function's parameters.¹⁶ After every execution, the output file is checked by comparing it with the expected output file provided by the users, after which the result is generated (Figure 8).

Figure 8: Compilation and Execution of source codes with input and output files
When all executions end, the system generates a set of route files and an error log, indicating whether an execution fails. Using these data, the system calculates the numbers of successful and failed execution that each line of code is involved in and the total number of successful and failed executions. An optimized Tarantula Formula is used as the program spectrum formula to identify every line of code's suspiciousness value:

\[ S_{ui} = \left( \frac{\log \frac{n_{i,FB}}{n_{i,F} + n_{i,B}}} {\log \frac{n_{i,S} + \sum_{j=1}^{u} n_{j,F} + n_{j,B}} {n_{i,F} + n_{i,B} + n_{i,S}}} \right) \times \left( \frac{\log \frac{n_{i,B}}{n_{i,F} + n_{i,B}}} {\log \frac{n_{i,S} + \sum_{j=1}^{u} n_{j,F} + n_{j,B}} {n_{i,F} + n_{i,B} + n_{i,S}}} \right) \]

, \( n_{i,F} = \{ |j|; \text{block}_{j} \cap \text{error}_{i} = 1 \}\) \( n_{i,B} = \{ |j|; \text{block}_{j} \cap \text{error}_{i} = 0 \}\) \( n_{i,S} = |\text{block}_{i}| \) \( n_{i,S} = |\text{block}_{i}| \)

Where \( u \) equals the number of times the latter parameter of max function is the maximum. Besides the original Tarantula formula, a new formula calculating the probability of another circumstance is generated and compared with the original one. It considers the possibility that bugs come from “not passing correct code blocks,” which differs from the original formula examining bugs from executions “passing certain buggy code blocks.”

However, it is risky to expect all code blocks to be passed to get successful outcomes since some are open to restricted conditions and designed not to pass in these failed test cases. Thus, \( 1/(u+1) \) is applied as a coefficient to rationalize the probabilities.

After the suspiciousness value calculations of every line, adjacent lines with the same suspiciousness values are combined into blocks of code \( \text{codeBlocks}[i] \) and sorted so that their suspiciousness values, \( \text{codeBlocks}[i].\text{sus} \), are arranged in descending order, becoming a ranked list of code blocks.

**Information Retrieval:**

The goal of applying information retrieval is to utilize previous debugging experiences to help optimize the accuracy of the current suspiciousness value of each code block. The historical file database provides information that contributes to the optimizations during the process.

The database contains folders indicating highly relevant sets of historical debugging analysis. In every historical debugging analysis, all information about the code blocks of the historical source code file is recorded, including every code block's starting, ending line, suspiciousness value, and description.

Throughout this section, the TF-IDF vectorization function is used to help evaluate the relevance among documents and suspiciousness sequences. The input text is first preprocessed through a series of text preprocessing methods. Non-alphabetical and non-numerical characters are removed, all characters are converted to lowercase, and all words in the text are tokenized into string vector \( \text{vocabList} \). Finally, the tokens are stemmed, or to remove their inflections to simpler forms of words, using OleanderStemmingLibrary.\(^{17}\)

Two maps record the token's term frequency (TF) and inverse term frequency (IDF), respectively. The former counts the occurrences of every term in \( \text{vocabList} \) and divides them by the total number of tokens in the text. The latter uses Code Description Corpus (Figure 9) to measure the document frequency. The corpus is read and outputted as token list \( \text{allVocabList}[i] \) referencing document i in the corpus. During the calculation of every term's inverse document frequency, the system iterates over \( \text{allVocabList}[i] \) and identifies whether the document contains the term through \textit{binary_search}.

*Figure 9: Code Description Corpus*

First, the system vectorizes the general description of the current source code file provided by the users into the TF-IDF vector, calculates the relevance score between the vector and all folders' description vectors using the cosine similarity formula, and finds the most relevant folder among the database, which is the primary reference for information retrieval in the following processes.

The system uses cosine similarity to measure the relevance between the current source code file and historical debugging analysis to determine how the analysis can affect the current suspiciousness ranking. The relevance score is a multiplication of three sub-relevance scores, namely \textit{fileRelevanceScore}, \textit{bugRelevanceScore}, and code block suspicion relevance, \( \text{cbMatchSus}[i] \) of the \( i \)th code block.

Based on the provided code block descriptions of the current and historical source code file, the system pairs \( \text{codeBlocks}[i] \) with the most relevant code block of the historical file. \( \text{cbMatch}[i] \), where \( i \) is the code block's index of the current source code file, indicates the index of \( \text{codeBlocks}[i] \)'s corresponding code block of the historical source code file. For every \( \text{codeBlocks}[i] \), the system calculates the cosine similarity between code block description \( \text{codeBlocksDesc}[i] \) and \( \text{cmpCodeBlocksDesc}[i] \), and finds the historical code block, \( \text{cmpCodeBlocks}[\text{cbMatch}[i]] \), whose description has the maximum cosine similarity with \( \text{codeBlocksDesc}[i] \).

With the matches of code blocks, the system vectorizes the string form of the combination of all code block descriptions \( \text{codeBlocksDescStr} \) and the corresponding historical code block descriptions \( \text{cmpCodeBlocksDescStr} \), and generate the relevance score \( \text{fileRelevanceScore} \) using cosine similarity.

While \( \text{fileRelevanceScore} \) represents the similarity of content between the current and historical source code file, \( \text{bugRelevanceScore} \) indicates the consistency of bug conditions between the two files. It is the cosine similarity between the ranked suspiciousness sequence and the corresponding suspiciousness sequence of the historical source code file.

Besides \( \text{fileRelevanceScore} \) and \( \text{bugRelevanceScore} \), \( \text{cbMatchSus}[i] \) indicates the consistency of bug conditions between the \( i \)th pair of code blocks. It is calculated through restricted growth formula to extremize the values at both ends (0% and 100%):
\[ cbMatchSus[i] = e^{-k*\Delta sus} \]

Where \( \Delta sus \) is the difference of suspiciousness value between the pair of code blocks. Logically, a difference of 0.5 in suspiciousness value indicates a 50% suspiciousness relevance of the two code blocks since the case is placed under an ambiguous circumstance where the possibility of each being completely irrelevant equals that of being completely relevant. Therefore, by substituting \( \Delta sus \) with 0.5 and \( cbMatchSus[i] \) with 50%, 
\[ -k*ln\ 0.5/50\% = -1.38629436112 \]

Therefore, the formula presented is:
\[ cbMatchSus[i] = e^{-1.38629436112*\Delta sus} \]

Multiplying the three relevance scores gets the final relevance score for updating the current suspiciousness ranking.

After retrieving the historical file’s result, \( buggyCodeBlocks[i] \) records the index of the final fixed buggy code blocks in the historical file. The system refers back to the corresponding code block of the current source code file using \( cbMatch \), and adds the relevance score to \( updateWeight \), which measures the final weight of updating the suspiciousness sequence:
\[
updateWeight[i] = \frac{\Sigma_{j \neq i} (relevance\ score_{file, i})^2}{\Sigma_{j \neq i} relevance\ score_{file, j}}
\]

Where \( file_{ij} \) is the file with one of its fixed buggy code blocks index equal to \( cbMatch[i] \). The final suspiciousness value \( finalSus[i] \) is updated according to \( updateWeight[i] \):
\[ finalSus[i] = codeBlocks[i].sus + (1 - codeBlocks[i].sus) \times updateWeight[i] \]

After the process, the code block information of current source code files is written to a new debugging analysis and submitted to the database.

### Results and Discussion

In order to test the system’s accuracy of correctly indicating bugs in the source codes, multiple pairs of buggy and fixed source code files written in c++ are used to examine the consistency of the output suspiciousness rankings and the buggy lines fixed in the correct source code file.

#### Data and Preprocessing:

In order to exhibit the validity of the information retrieval section, the testing focuses on one coding problem "Wanna Go Back Home" from AtCoder Grant Contest 003.\(^{18}\) Throughout the testing, 15 pairs of buggy and fixed source code files are randomly chosen from the submission page of the problem (Figure 10). Each pair of buggy and correct source code files are written by the same user in AtCoder. The buggy source codes are labeled “WA (Wrong Answer)” in the online judge, and the correct one with “AC (Accepted)” (Figure 10).

Figure 10: One of the pairs of buggy and correct source code files for Wanna Go Back Home

Source code and code block descriptions are generated for every pair of buggy and fixed source code files. Source code descriptions are the paraphrased form of the problem statement and are unique from other source code descriptions, while code block descriptions present the content of code blocks in the form of plain words, as close to the codes as possible.

Complete input and output test case folders are downloaded from its official folder atcoder_testcases on Dropbox.\(^{19}\) All of the test cases are involved in the execution of the buggy source code files.

Preprocessing:
1. Removal of Comments
   In order to condense the code length and unnecessary runtime, additional comments are removed.
2. Addition of curly braces for single-line loop or conditional statements
   Since the line “route ≪ to_String(__LINE__) ≪ “” is added to the end of a line, single-line loop or conditional statements disable the mechanism to detect whether the statements within are passed. Besides, if additional lines are added after an ‘if’ that is followed by an else if or else, the syntax will be incorrect and lead to compile errors. Therefore, curly braces are added to the statements’ end to wrap the content in curly braces.

#### Result

Among the 20 pairs of buggy and correct source code files, 48 code blocks are buggy and fixed, each receiving a final suspiciousness value and rank.

The suspiciousness values of the buggy code blocks are concentrated between 70% to 89%, illustrating that the system has highlighted most of the buggy code blocks as highly suspicious, as shown in Figure 12.

Also, 70.83% of the buggy code blocks are ranked above the 70th percentile (Figure 13), indicating that these code blocks are also considered as being most likely to contain bugs compared to other code blocks in the source code.

Overall, the suspiciousness values of the buggy code blocks are updated by an average of 22.08% from program spectrum analysis to the information retrieval section (Figure 14). The significant improvement of the suspiciousness values demonstrates the effectiveness of information retrieval on relevant
historical files and increases its accuracy in identifying buggy code blocks.

Besides the rank prominence, the calculated suspiciousness values of buggy code blocks also have outstanding suspiciousness values compared to other correct code blocks. The overall average of the deviation score of the buggy code block’s suspiciousness values is 0.8, and 75% of them have deviation scores above 0.8 (Figure 15), signifying that the suspiciousness values of the buggy code blocks are not only ranked top but distant to that of other code blocks in the suspiciousness rankings (nearly one standard deviation away from the average).

**Table 2:** Mean (± Standard Deviation) of the suspiciousness value, rank, and absolute accuracy of the fixed code blocks

<table>
<thead>
<tr>
<th>Average Suspiciousness Value of Buggy Code Blocks</th>
<th>Average Ranking of Buggy Code Blocks</th>
<th>Average Percentile Rank of Buggy Code Blocks</th>
<th>Absolute Accuracy of Buggy Code Blocks ranked No. 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>70.94% ± 13.32%</td>
<td>No. 2.63 ± 2.10 / 11.08</td>
<td>76.69% ± 14.69%</td>
<td>43.75%</td>
</tr>
</tbody>
</table>

**Discussion**

After testing with the source code files, multiple conclusions are drawn from the statistical results.

**Problems:**

1. **Weakened effectiveness under all-WA situations**
   
   Based on the program spectrum formula:
   
   \[
   \left( \frac{n_{i,F,0} + n_{i,F,1}}{n_{i,F,0} + n_{i,F,1}} \right) \times \left( \frac{n_{i,S,0} + n_{i,S,1}}{n_{i,S,0} + n_{i,S,1}} \right)
   \]
   
   when there are only failed cases, \(n_{i,S,0}\) and \(n_{i,S,1}\) both equals 0, making the suspiciousness value 100%, which represents definite bugginess of the code block and is therefore inaccurate.

2. **High Time Complexity**

   The time complexity of this system is \(O(n_f \times n_c^2 \times n_{term} \times n_{doc} \times \log \log n_{token})\) where \(n_f\) is the number of historical files in the most relevant folder, \(n_c\) is the number of code blocks in a source code file, \(n_{term}\) is the number of terms in an input text file, \(n_{doc}\) is the number of documents in the corpus, and \(n_{token}\) is the number of tokens in each document of the corpus. This system becomes time-consuming when \(n_c\) is large.

3. **Inaccurate Code block Matchings**

   Occasionally, the current source code blocks are matched with irrelevant historical code blocks, making the result inaccurate.

4. **Noise Problem**

   Since the system uses relevance scores as the updating weights, files and code blocks with little relevance still affect the final suspiciousness ranking, making it less accurate.

**Solution:**

To solve the problem that occurred under all-WA conditions, the design of the program spectrum formula is changed by adding special cases to make the suspiciousness value be 0.5 under the condition that \(n_{i,S,1}/n_{i,S,0} + n_{i,S,1}\) equals 0 and \(n_{i,F,1}/n_{i,F,0} + n_{i,F,1}\) doesn’t. The database then updates the suspiciousness sequence according to the retrieval of the historical debugging analysis, therefore enhancing the accuracy of the result. An example is shown in Figure 16, where initially the suspiciousness values of all the potentially buggy code blocks are 100%, yet through the optimization mentioned above, the more reasonable result of the program spectrum section provides space for the information retrieval section to update the values according to relevant historical files.

**Figure 12:** Suspiciousness Value Distribution of Buggy Code Blocks

**Figure 14:** Initial and Final Suspiciousness Value Distribution of Buggy Code Blocks

**Figure 15:** Distribution of Buggy Code Block’s Suspiciousness Value Deviation Scores

**Figure 16:** Buggy Source Code of AGC003_12cpp (Left), Comparison of its suspiciousness values before and after the optimization is implemented (Right)
Solution Proposal:

1. After calculating the IDF value of terms, the system can save the results onto the database and directly retrieves them when there are identical input terms afterward. This approach is especially helpful for common terms.

2. To solve the inaccuracy of code block matching and noise problem, changing the code block matching algorithm from textual similarity to control flow graph comparison is more accurate regarding the similarity of the functions of the code block pairs.

3. Increase the weight of relevance score in the update weight formula (i.e., change the numerator from $\sum \text{[relevance score]}$ to $\sum \text{[relevance score]}$). Simultaneously, broaden the number of historical files and folders to provide ample and more likely valuable references. Once the system can retrieve sufficient relevant files, the irrelevant ones will have little impact on the final suspiciousness rankings.

## Conclusion

### Summary of Findings:

By implementing program spectrum analysis and information retrieval in the system, it is demonstrated that the system can detect and rank most of the buggy code blocks as highly suspicious, reflected in the final code block ranking. To further improve the system's accuracy and efficiency in localizing bugs, calculations should be saved and used when the next similar request is made. Also, the code block matching algorithm can be changed to control flow graph comparison to keep the matchings consistent with the relevance of code blocks' content. At the same time, the database should constantly expand to handle a wider range of source code files.

### Future Prospects and Applications:

Currently, this system only supports single-file code projects written in c++. In the future, it is expected to expand the types of supported programming languages and give more users access to the system.

The system can also be integrated into code editors and IDE (Integrated Development Environment) and combined with debuggers. By utilizing the large quantities of source codes developed on the platforms and the comprehensive functions of debuggers, the system can expand the database and implement higher quality fault localization on the input source codes.

By saving the developers' development time, this system can ultimately increase the production of software programs and even fasten technological growth.

### Acknowledgments

I want to acknowledge and give my warmest thanks to my two advisors, Teacher Yi-Ting Liao and Dr. Sun-Yuan Hsieh. Without their guidance on the content, overall structure, and format of my research, I wouldn't be able to achieve the current quality.

### References


The source code of the automated debugging system developed in this research is publicized on Github. Link: https://github.com/Sammyhao/Automated-Debugging-System.

Authors
Sam is a senior at Wego High School who loves programming and is always eager to explore the newest findings in computer and information systems. Besides doing research, he is currently developing web applications and leading the informatics club at school. He intends to extend her interest in coding by majoring in computer science in college.
Genomic Sequencing in Clinical Oncology

Aashna K. Soni
BASIS Independent Fremont, 39706 Mission Blvd, Fremont, CA 94539, USA; soniaashnakaur@gmail.com
Mentor: Gauravjit Singh

ABSTRACT: Next-generation sequencing (NGS) technologies can simultaneously sequence millions of small DNA or RNA fragments. They mark a paradigm shift over the preceding first-generation sequencing methodologies, notably Sanger sequencing. With its low cost and high throughput, NGS holds value for many different clinical applications. This review aims to discuss the application of NGS in a clinical oncology setting—namely, how sequencing is used to detect an individual's unique mutations and inform targeted therapy options. This review will focus specifically on how Foundation Medicine, a leading company in cancer biomarker detection, conducts its testing and interprets the results obtained from sequencing. Additionally, this review will explore the underlying mechanism of NGS by illustrating the Illumina sequencing workflow, one of the most commonly used NGS workflows today that is also used in Foundation Medicine’s products. Finally, the review will look ahead to newer innovations and improvements that will enhance the ability of genomic sequencing to deliver tangible benefits to cancer patients.

KEYWORDS: Translational Medical Sciences; Disease Treatment and Therapies; Genomic Sequencing; Illumina Sequencing; Genomic sequencing in cancer applications.

Introduction

Cancer is a disease in which cells grow uncontrollably.¹ For many years, conventional chemotherapy was the standard treatment for most cancers. Chemotherapy works by targeting cells at different stages of the cell cycle.² Since cancer cells divide much more rapidly than normal cells due to the specific mutations they carry, chemotherapy is often highly successful at killing cancerous cells. However, because chemotherapy cannot distinguish between healthy and cancerous cells, normal cells are also damaged in the process, which can lead to serious side effects.³ For example, because hair follicle cells divide rapidly, chemotherapy often damages these cells, resulting in significant hair loss.⁴

As a result, a new class of treatments, collectively known as targeted therapies, has emerged to target an individual's unique mutations by looking for characteristic biomarkers through a tumor's DNA sequence. Whereas chemotherapy kills cancer cells in the body, targeted therapy prevents cancer cells from replicating by targeting these specific biomarkers. Mechanisms by which targeted therapies work include stimulating the immune system to kill cancer cells, interrupting growth-stimulating signals from reaching cancer cells, interfering with signals that trigger angiogenesis, or causing cancer cells to undergo apoptosis.⁴ These drugs significantly improve over standard chemotherapy, demonstrating the benefit of using an individual tumor’s DNA sequence to target cancer cells.

Most targeted therapies fall into two categories: monoclonal antibodies and small-molecule drugs. Monoclonal antibodies are designed to function as "substitute antibodies,"⁵ binding to antigens on cancer cells. They can serve many functions, including flagging cancer cells for destruction by the immune system, blocking proteins that cancer cells need to divide and proliferate, or inhibiting immune checkpoints.⁵ An example of a monoclonal antibody is trastuzumab, which is used to treat human epidermal growth factor receptor 2 (HER2) positive breast cancer. This molecule attaches to HER2 on cancer cells, which prevents HER2 from sending signals that allow cancer cells to continue growing.⁶ On the other hand, small-molecule drugs have small molecular weights and can enter cancer cell membranes to target intracellular molecules.⁷ An example of a small-molecule drug is sorafenib, a tyrosine kinase inhibitor.⁸

Many targeted therapies today rely on NGS to determine a tumor’s DNA sequence. For example, Foundation Medicine conducts NGS for solid tumors to suggest FDA-approved therapies. Though NGS has shown great success in oncology applications, some downsides exist. For example, the short read lengths generated from a DNA sample make it difficult to identify larger structural variations. As a result, an emerging area of interest is single-molecule sequencing, which allows for longer read lengths to be produced. Thus, single-molecule sequencing represents an area in which to explore new types of complex mutations involved in cancer, such as large insertions, deletions, inversions, duplications, and translocations. Using short-read technologies, structural variations are determined indirectly by aligning multiple reads to a reference genome.⁹ In some cases, certain types of mutations, such as copy number variations, may not even be detected from short-read data.⁹ Long read coverage, on the other hand, can allow for entire structural variations to be covered in a single read in many instances, increasing sensitivity to these types of mutations.⁹ Additionally, long-read sequencing performs particularly well in highly repetitive regions of DNA and hard-to-access regions of the genome, such as telomeres and centromeres.⁹
Most reviews focus on precision medicine in cancer focus on a specific aspect of the process, such as sequencing techniques, the development of targeted therapies, different types of biopsies in the clinic, or the genetic complexity and functional relevance of different mutation types. This review is different in that it aims to combine all of these areas to coherently describe the process of genomic sequencing and targeted therapies in a clinical context, as well as to point to newer techniques that can improve the current state of cancer care.

Results and Discussion

The First Generation of DNA Sequencing: Sanger Sequencing

Frederick Sanger’s chain-termination sequencing technique, developed in 1977, marked a significant achievement in the field of sequencing technology. An illustration of the process of Sanger sequencing is shown in Figure 1. The technique uses modified nucleotides called dideoxynucleotides triphosphates (ddNTPs), whose sugar molecules “lack a hydroxyl group on the 3’ carbon,” preventing extension of the DNA chain. The given DNA sample is first denatured to separate the two strands. ddNTPs, which are labeled with a different dye color for each nitrogenous base, and dNTPs, are put together into the reaction, though the concentration of ddNTPs is less, along with DNA polymerase and the desired DNA sample to be sequenced. A primer is annealed to the DNA template strand, allowing DNA polymerase to extend the chain of nucleotides. Whenever a ddNTP is incorporated instead of a dNTP, the chain is terminated due to the lack of a hydroxyl group on the 3’ carbon of the nucleotide. After many cycles, millions, or even billions, of DNA fragments of different lengths are generated, guaranteeing that a ddNTP would have been incorporated at each location in the template DNA. These fragments undergo capillary electrophoresis, which separates the fragments based on size, allowing the terminating nucleotides on the fragments to be read in order from smallest to largest fragment. The results are shown on a chromatogram, in which a peak represents a nucleotide. By detecting a ddNTP and its characteristic dye from the end of each DNA fragment, the instrument can determine the sequence of the DNA sample nucleotide by nucleotide.

Figure 1: Sanger sequencing workflow.

The Second Generation of DNA Sequencing:

The second generation of DNA sequencing represents a significant shift over preceding first-generation sequencing technologies in that it produces short reads in a massively parallel fashion. Sanger sequencing was able to produce high-quality reads with low error rates. However, this process is costly and inefficient. On the other hand, in NGS, multiple DNA fragments are sequenced simultaneously and later assembled to generate the complete DNA sequence, making it much faster and cheaper than Sanger sequencing.

There are many different types of NGS. Examples include Roche/454 sequencing, which utilizes a pyrosequencing approach, and Ion torrent sequencing, which works by detecting hydrogen ions released as DNA polymerase incorporates nucleotides. However, this review will focus specifically on the Illumina NGS approach, as it is one of the most widely used workflows today and is utilized by many companies, including Foundation Medicine. The general process of Illumina sequencing consists of the following steps: library preparation, in which adapters are ligated to DNA fragments; cluster generation, in which bridge amplification is used to produce multiple copies of the DNA fragment; sequencing by synthesis, in which fluorescent signals emitted from the flow cell after each nucleotide addition are detected to determine the DNA sequence; and data analysis, in which the millions of reads generated are aligned to a reference genome and variants are determined. The first three steps of the process are shown in Figure 2. First, the given DNA sample is randomly fragmented, and adapters are added to the ends of each fragment. These adapters allow the fragments to hybridize to the flow cell in the next step. Then, through bridge amplification, in which double-stranded DNA “bridges” are formed as DNA fragments bend over and bind to vacant flow cell oligos, DNA polymerase synthesizes new strands, generating many clusters of DNA fragments. In the sequencing stage, the fluorescently-labeled bases are called through the incorporation of complementary nucleotides. As the flow cell is imaged, the “emission wavelength and intensity,” which are produced from each cluster, determine the base call. Then, the reversible terminator on the nucleotide is cleaved, allowing for the incorporation of the next base. This process is repeated many times to generate a read of the desired read length. This procedure is known as sequencing by synthesis, as the template DNA strand is read through the incorporation of complementary nucleotides by DNA polymerase. Finally, the millions of reads generated are aligned to the reference genome to synthesize the full sequence, and variants are identified.

Figure 2: Illumina sequencing workflow.
NGS in Oncology:

NGS technology has shown tremendous value in clinical cancer applications due to its ability to identify targeted therapies for patients. As the quality, efficiency, and cost of NGS have improved, more and more patients are being treated with targeted therapy and immunotherapy. The following sections describe the process of biomarker testing in a clinical context, from the biopsy to the interpretation of results.

Biopsy:

The process of cancer care begins with a biopsy, in which cells or tissues are removed from a specific location in the patient’s body to be analyzed by the pathologist.¹⁶ A biopsy is necessary to confirm a cancer diagnosis, and many different types of biopsies can be administered. For this review, the two main types of biopsies that will be discussed are tissue biopsies and liquid biopsies.

Tissue biopsies are the standard for biopsies in cancer patients owing to the different types of data that can be analyzed from a tissue specimen.¹⁷ Tissue biopsies allow for a histologic analysis of the specimen and the detection and analysis of important “non-DNA-based alterations.”¹⁷ Frequently, the tissue is preserved in formalin-fixed paraffin-embedded (FFPE) format. This preparation is ideal for immunohistochemistry, as “cell structures and proteins are well preserved.”¹⁸ The drawback of using FFPE tissue is that the DNA and RNA in these samples are often degraded and chemically damaged during the fixation process due to chemical crosslinking between nucleic acids and proteins.¹⁹ Nevertheless, many companies, such as Foundation Medicine, use FFPE-preserved tissue in their genomic analysis process. On the other hand, fresh frozen tissue biopsies are much better for analyzing DNA or RNA, but they are difficult to store for long periods.¹⁸

The second main type of biopsy is a liquid biopsy. Liquid biopsies are performed when a tissue biopsy is infeasible or when a biopsy might pose a significant danger to the patient.¹⁷ In a liquid biopsy, cell-free DNA (cfDNA), or small DNA fragments located in “the noncellular component of the blood,”¹⁷ is isolated from bodily fluids to analyze the amount and sequence of circulating tumor DNA (ctDNA) present. Many different types of bodily fluids can be used for liquid biopsies, such as urine, cerebrospinal fluid, saliva, and blood.¹⁷ Blood is the most commonly used fluid in a liquid biopsy.¹⁷ The benefit of a liquid biopsy is that it is a minimally-invasive procedure, as compared to a tissue biopsy, which could be painful and dangerous.¹⁷ Furthermore, studies have shown that analysis of cfDNA is better able to capture the “molecular heterogeneity associated with resistance,”¹⁷ since there are often many different tumor subpopulations within a single individual, and each tumor subclone carries unique mutations. The unique mutations carried by different tumor populations in a metastatic cancer can be better captured with a blood sample that contains ctDNA released throughout the body.¹⁷ However, liquid biopsies may not always yield optimal results for patients if the concentration of ctDNA in the blood is too low, as current technologies are unable to detect ctDNA below a certain concentration.¹⁷

Analysis of the Biopsy:

After a tissue or blood sample is collected, it is sent to a pathologist, who performs different tests to provide a final diagnosis and compiles information about the gross and microscopic description of the specimen in a pathology report.²⁰ After the pathology report is issued, the medical oncologist may send the biopsy for NGS to drive a targeted treatment strategy. The results of the sequencing procedure are compiled in a biomarker testing report. Foundation Medicine’s biomarker testing approach is described below.

Foundation Medicine Genomic Sequencing:

Foundation Medicine performs a type of NGS called comprehensive genomic profiling (CGP) on solid tumors. Foundation Medicine’s tissue-based and blood-based tests analyze 324 genes. The combined DNA and RNA test (FoundationOne® Heme) analyzes 406 DNA genes and 265 RNA genes to better understand gene fusions and rearrangements.²¹,²² In addition to looking for mutations in a wide range of genes, Foundation Medicine analyzes the clinically significant genomic signatures of microsatellite instability (MSI) and tumor mutational burden (TMB). The specific sequencing technique used by Foundation Medicine is hybridization-based capture technology,²³-²⁵ which sequences a subset of genomic regions that are isolated through “hybridization to target-specific biotinylated probes.”²⁶ Hybridization-based capture technology allows a large number of genes to be sequenced at a high mutation resolution to detect and understand new and known variants.²⁶

Foundation Medicine Genomic Signatures Analyzed:

Foundation Medicine analyzes two main genomic signatures in DNA samples: MSI and TMB.²¹ A microsatellite is a “short segment of DNA, usually, one to six or more base pairs in length, repeated multiple times in succession at a particular genomic location.”²⁷ MSI occurs when there are differences in the number of microsatellite repeats, making the region of nucleotides unstable.²⁸ MSI occurs due to deficient DNA mismatch repair (dMMR). Mismatch repair is responsible for correcting such errors during nucleotide incorporation. MSI could indicate that an individual has Lynch syndrome, which could put them at high risk for colorectal cancer.²⁹,³⁰ TMB measures the number of somatic mutations “per coding area of a tumor genome.”³¹ It has been shown that high TMB (TMB-H) generally occurs in cancers linked to high mutagen exposure, an example being the connection between smoking and the presence of non-small cell lung cancer (NSCLC).³¹

MSI and TMB are important biomarkers that predict a patient’s response to immunotherapy treatment. One specific type of immunotherapy used for MSI-H or TMB-H cancers is immune checkpoint inhibitors (ICI). Immune checkpoints are pathways that ensure that the immune system does not attack healthy cells in the body.³² However, cancer cells utilize these pathways to their advantage to render immune cells inactive.³² A common mechanism through which immune checkpoints function is through “ligand-receptor interactions.”³² In this process, a ligand on the cancer cell binds to a receptor on the T-cell, inactivating the T-cell and preventing it from attacking
the cancer cell. The process of ICI takes advantage of these interactions. In ICI treatment, monoclonal antibodies are delivered to target “negative regulators of T-cell function,” or proteins that are part of the ligand-receptor interactions. Examples of such negative regulators include PD-L1 (present on tumor cells), PD-1 (present on T-cells), and CTLA4 (present on T-cells). The monoclonal antibody binds to one of these proteins, preventing it from binding to the corresponding protein on either a cancerous cell or a T-cell. This allows T-cells to remain active and find and destroy cancer cells. Figure 3 illustrates the mechanism of action of immune checkpoint inhibitors, specifically targeting PD-1 or PD-L1.

**Figure 3: ICI mechanism of action.**

**Driver Genes: Background:**

Driver genes, or genes that, when mutated, contribute to cancer development, are grouped into two main categories: proto-oncogenes and tumor suppressor genes. Proto-oncogenes code for proteins that trigger cell division. When mutated, proto-oncogenes become oncogenes, which code for proteins that excessively stimulate cell division. One mechanism through which a proto-oncogene can be converted to an oncogene is when mutations within a proto-oncogene result in an overactive protein that stimulates cell division being produced in the normal amount. Other mechanisms include multiple copies of the proto-oncogene being created and mutations within a control region, such as promoter, both of which lead to an abnormally large quantity of the normal growth-stimulating protein. Examples of clinically important proto-oncogenes are **BRAF, KRAS,** and **MYC.**

On the other hand, tumor suppressor genes produce proteins that inhibit cell division. Mutations to tumor suppressor genes destroy safeguards designed to inhibit cell division, resulting in cell proliferation. Examples of tumor suppressor genes include **TP53, PTEN,** and **CDKN2A.** Furthermore, a specific group of tumor suppressor genes functions in DNA repair. An example of a repair mechanism that, when malfunctioning, contributes to cancer development is mismatch repair (MMR), which is primarily involved in correcting base-base errors and small indels. Mutations to DNA repair genes increase the likelihood of cells developing other driver mutations, which further contribute to cancer development.

**Foundation Medicine Genomic Alteration Classes Analyzed:**

Genomic alterations that occur in driver genes play an important role in cancer development. The main genomic alteration classes analyzed by Foundation Medicine are base substitutions, insertions and deletions (indels), copy number alterations (CNAs), and select genomic rearrangements.

Single base substitutions (SBSs) occur when one nucleotide in the DNA sequence is replaced with another. Base substitutions can be either missense, which results in an incorrect amino acid, nonsense, which results in a premature protein, or silent, which does not affect the amino acid produced. According to a review published in 2014, the number of SBSs present in an individual can vary greatly, both among different cancer types and within cancer types. Cancers, like lung cancer, whose presence is linked to “chronic mutagen exposure,” were found to carry a large number of SBSs. SBS mutations that are quite prominent in cancer genomes are “C→T transitions at CG:CG sites and substitutions at C:G base-pairs in the context of YC:GR dinucleotides,” where “Y” represents a pyrimidine and “R” represents a purine, and a colon separates dinucleotides on opposite strands. For example, SBS signatures found in lung cancer include C→A:T and T:A→G:C transversions, whose putative causes are tobacco smoke and arsenic exposure, respectively.

Indels occur when nucleotides are added to or deleted from the DNA sequence. Indel mutations are usually confined to less than 1 kb of nucleotides. Indels lead to a shift in the reading frame if the number of bases added or deleted is not a multiple of three, and this generally results in a nonfunctional protein. These types of indels are called non-3n indels. On the other hand, 3n indels, in which the number of bases added or deleted is a multiple of three, will cause amino acid changes but will not have as drastic effects as non-3n indels. According to a study published in 2010, since indels can significantly affect gene function, such as through the downstream effects of frameshift mutations, indels can serve as a type of driver mutation in cancer development. This was demonstrated by the abundance of indels in the COSMIC database analyzed in the 2010 study. Additionally, this study found that frameshift mutations, or non-3n indels, are much more prevalent in tumor suppressor genes than in oncogenes, whereas 3n indels are much more common in oncogenes, as shown by the fact that there were 7.9 times more 3n indels in oncogenes than in tumor suppressor genes when looking at genes with greater than or equal to 100 mutations. Overall, the study found that those “one- or two-bp indels and non-3n indels are dominant in both genome and coding sequences,” this can vary considerably based on the specific niche location of indels in the genome.

CNAs are a type of structural change in genomic material in which regions of the DNA sequence are deleted or amplified. CNAs can affect areas ranging from just a few kilobases up to entire chromosomes. Somatically acquired copy number alterations (SCNAs), or those acquired during an individual’s life and not passed on to future generations, play a pivotal role in tumorigenesis. SCNAs allow tumor cells to increase the expression of certain genes and decrease the expression of...
others, thus likely conferring a certain advantage to these cells.⁴³ The “first comprehensive pan-cancer analysis of SCNAS,"⁴⁴ published in 2013, revealed that large amplifications or deletions are commonly present in ovarian carcinomas, and the number of SCNAS varies drastically based on the tumor type.⁴⁴

Finally, rearrangements are large-scale changes to chromosome structure that include duplications, deletions, insertions, inversions, and translocations.⁴⁵ One important result of many genomic rearrangements is gene fusion. Gene fusions are “hybrid genes formed when two previously independent genes become juxtaposed."⁴⁶ Fusions can result from both structural and non-structural rearrangements, and they are a common source of driver mutations in many cancers.⁴⁶

The mechanisms through which fusion genes contribute to tumorigenesis include causing one of the fusion partners (such as a tumor suppressor gene) to lose its function, coding for an abnormal protein that has “oncogenic functionality,"⁴⁶b or deregulating one of the fusion partners (such as a proto-oncogene).⁴⁶ The frequency of gene fusions is highly variable among different cancers.⁴⁶ One important example of gene fusion occurs in lung cancer. In patients with ROS1-positive lung cancer, a gene fusion occurs between the ROS1 gene and a portion of another gene, the most common one being the CD74 gene.⁴⁷ This fusion results in an activation of the ROS1 gene that leads to cell proliferation.⁴⁷

Interpretation of Data:

Information found through NGS can predict which types of treatment have the highest likelihood of working for a given patient based on the driver mutations identified. For example, common driver mutations in lung cancers occur in the genes EGFR and KRAS.⁴⁸ EGFR is a gene that codes for the EGFR receptor protein involved in triggering cell proliferation and survival upon a ligand binding to it.⁴⁹ KRAS codes for the K-Ras protein which is involved in the RAS/MAPK signaling pathway, and it is involved in transmitting signals from the cell’s environment to the nucleus to trigger cell division or differentiation.⁵⁰ Based on a patient’s unique mutation, the Foundation Medicine test will recommend a targeted therapy. For example, currently, patients with the KRASG12C mutation will still undergo first-line systemic therapy.⁵¹ However, if the patient does not respond well to this standard treatment, they may be prescribed the Kras inhibitor, sotorasib, an FDA-approved small molecule.⁵² Sotorasib works by binding to KRAS G12C and inhibiting “KRAS oncogenic signaling.”⁵²

Genomic signatures can also guide treatment. MSI-H or TMB-H tumors would likely respond well to immunotherapy treatment, such as pembrolizumab, an immune checkpoint inhibitor. Pembrolizumab is an FDA-approved anti-PD-1 drug to treat patients with MSI-H or TMB-H cancers under specific circumstances (such as when prior treatment has not been successful), where TMB-H is defined as greater than or equal to 10 mutations/megabase.⁵³

If no FDA-approved therapies are available for the patient, they may be encouraged to enroll in a clinical trial to access new drugs still in development. Clinical trial enrollment has the added benefit of helping researchers gather data on the safety and efficacy of new drugs, promoting the development of new treatment options for a broad group of patients.

Despite the significant amounts of valuable data generated by NGS, medical oncologists still look at many other factors when considering which drugs to prescribe to a particular patient, including cancer type, stage, age, and overall health. For example, an aggressive drug would likely not be used to treat an elderly patient with late-stage cancer, as the side effects could be unmanageable. Nevertheless, data collected from NGS are pivotal in identifying targeted therapies that are more likely to work than conventional chemotherapy.

Advantages of Foundation Medicine’s Approach:

Because Foundation Medicine accommodates both tissue and blood samples through its FoundationOne® CDx and FoundationOne® Liquid CDx, even patients with tumors in difficult-to-access locations can benefit from the large-scale genomic analysis. Additionally, the massively parallel nature of the test allows for simultaneously testing hundreds of genes to detect many different types of biomarkers.

Proven Benefits of NGS: Non–Small Cell Lung Cancer:

A patient suspected of having lung cancer will be sent for a biopsy. The type of biopsy can vary based on the location of the tumor and the patient’s specific conditions. Examples of biopsy procedures include bronchoscopy, mediastinoscopy, and transthoracic needle aspiration.⁵¹ Tissue biopsies are the standard for biopsies.⁵¹ However, liquid biopsies can be performed in cases when an invasive tissue biopsy would be unsafe for a patient based on their medical conditions.⁵¹

Let us assume that a patient is diagnosed with NSCLC. A patient with advanced or metastatic disease will have their histologic subtype established.⁵³ Based on these results, the patient will be diagnosed with an NSCLC subtype, such as adenocarcinoma or squamous cell carcinoma.⁵³ If the patient’s histologic subtype is determined to be adenocarcinoma, then a tumor sample will be sent for biomarker testing to look for mutations in genes such as EGFR and ALK, as well as genes such as KRAS, ROS1, and BRAF.⁶⁰ In addition to testing for these genes, the patient’s PD-L1 status will also be tested to determine the potential for immunotherapy treatment.⁶⁰ If the patient has an EGFR exon 19 deletion or L858R mutation and the “EGFR mutation [is] discovered prior to first-line systemic therapy,”⁶⁰ the preferred drug for the patient will be osimertinib,⁵¹ a third-generation EGFR tyrosine kinase inhibitor (EGFR-TKI).⁵⁵

Protein tyrosine kinases are proteins that play an important role in cell-signaling pathways that are involved in controlling “cellular differentiation and proliferation.”⁵⁶ Osimertinib works by inhibiting mutant EGFR, the protein product of activating mutations in EGFR.⁵⁵ Osimertinib differs from earlier-generation EGFR-TKIs in that in addition to targeting exon 19 deletions and exon 21 L858R mutations, it also targets the EGFR T790M mutation. This mutation often causes tumors to develop resistance to earlier EGFR-directed therapies.⁵⁷ Osimertinib has an overall response rate (ORR) of 80% based on the FLAURA clinical trial, performing significantly better than standard EGFR-TKIs.⁵⁸ However, tumors can still acquire resistance to osimertinib treatment through mechanisms such
as “loss of the T790M mutation during osimertinib therapy.”\textsuperscript{55}

If this change is accompanied by other resistance mechanisms, such as MET amplification and KRAS mutation, even first-generation EGFR-TKIs may be ineffective.\textsuperscript{55} Based on the symptoms (or lack thereof) that the patient experiences, they may progress to a new therapy, such as surgery or stereotactic ablative radiotherapy (SABR), or may even continue osimertinib.\textsuperscript{51} Overall side effects of osimertinib include diarrhea, constipation, rash, itching, and vomiting.\textsuperscript{59} Currently, other drugs are being studied in clinical trials to target tumors that have developed specific resistance mechanisms to osimertinib.\textsuperscript{58}

**Future of Genomic Sequencing in Cancer:**

Though NGS marks a significant improvement over conventional chemotherapy approaches for treating patients, it still has certain disadvantages. For example, NGS generates very short read lengths, which makes it difficult to assemble the reads to the reference genome and accurately detect structural variations (SVs).\textsuperscript{60} Additionally, the amplification-based procedures involved often obscure important mutations that are present at a lower frequency.\textsuperscript{60} Single-molecule sequencing (SMS) has the potential to address these gaps to make the process of genomic sequencing in cancer more fruitful. In the following section, the benefits and pitfalls of SMS are described.

**Single-molecule Sequencing:**

SMS, or third-generation sequencing (TGS), is a newer sequencing approach, currently spearheaded by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), that allows for much longer read lengths to be generated compared to NGS workflows.\textsuperscript{60}

Figure 4 compares NGS with two types of TGS: PacBio SMRT sequencing and ONT sequencing. The main difference between the two types of sequencing methods is the size of the read lengths. While the insert, or DNA fragment sandwiched between two adapters, is less than 300 bp in NGS,\textsuperscript{60} PacBio can sequence tens of kilobases of DNA.\textsuperscript{61} And, Nanopore sequencing has an even longer read length capacity, sequencing as low as 20 bases all the way up to millions of bases.\textsuperscript{62} In PacBio SMRT sequencing, a circular DNA template is generated (as opposed to linear templates in NGS) and then sequenced as DNA polymerase incorporates nucleotides.\textsuperscript{60} The DNA polymerase sequencing by synthesis method is a similarity between PacBio SMRT sequencing and NGS. In ONT sequencing, DNA fragments are passed through nanopores situated in an electrically resistant membrane in which an electrical current is run.\textsuperscript{63} As the DNA fragment passes through the nanopore, electrical current disruptions occur, which are measured to determine the base call.\textsuperscript{63}

The advantage of SMS approaches is that they can sequence long fragments of DNA directly, which differs from current NGS approaches that rely on amplification techniques such as bridge amplification.\textsuperscript{60} Additionally, they are substantially better at identifying SVs compared to conventional NGS approaches due to the long reads produced.\textsuperscript{60} Currently, there is difficulty in identifying SVs using NGS because of the short read lengths.\textsuperscript{64} In fact, current short-read approaches detect within the range of 10% to 70% of SVs and have false positive rates ranging up to 89%.\textsuperscript{64} Short-read approaches are also not designed to capture the full complexity of SVs.\textsuperscript{64} On the other hand, SMS, by utilizing longer reads, can better identify larger, more complex SVs in the given DNA sample. Overall, SMS can also detect mutations that occur at low frequencies in the DNA sample.\textsuperscript{60}

SVs, including large insertions, deletions, inversions, duplications, and translocations, represent an important class of mutations in cancer.\textsuperscript{65} For example, some serious small cell lung cancer cases that are high-grade are driven by somatically acquired SVs.\textsuperscript{66} There are many mechanisms through which SVs can contribute to tumorigenesis. They can alter the copy number of genes, obstruct the crucial function of tumor-suppressor genes, form fusion genes, or pair one gene’s coding sequence with another’s regulatory machinery.\textsuperscript{65} In fact, new types of complex SVs are being discovered,\textsuperscript{67} and these variants could help develop new anti-cancer drugs.\textsuperscript{68}

Sakamoto et al. found a unique class of mutations, cancerous local copy-number lesions (CLCLs), using SMS.\textsuperscript{66} CLCLs are composed of “complex combinations of copy-number changes (duplications), inversions, and deletions,”\textsuperscript{67} and they were found within important cancer genes “such as the \textit{STK11}, \textit{NF1}, \textit{SMARCA4}, and \textit{PTEN} genes.”\textsuperscript{67} These types of mutations were difficult to identify and characterize solely based on short-read sequencing,\textsuperscript{67} suggesting the need for long-read sequencing approaches to identify novel mutation types that could have clinical relevance as drug targets. Additionally, this study confirmed that long-read sequencing approaches alone can still identify small-scale mutations, such as point mutations, which are changes in single nucleotide pairs.\textsuperscript{67}

Currently, the main disadvantage of SMS is the high error rate.\textsuperscript{66} It has been shown that for PacBio SMRT sequencing and ONT sequencing, most errors can be attributed to “false insertions or deletions.”\textsuperscript{60} In the future, it is possible that NGS will be combined with SMS for a “hybrid sequencing strategy.”\textsuperscript{66} This approach would allow for better quantification of fusion genes and mRNA transcript isoforms and reduce the error rate.\textsuperscript{60} Furthermore, to successfully incorporate long-read sequencing into the clinic, the format of the tissue biopsy collected will need to change.\textsuperscript{68} Currently, the downside of using FFPE samples is that the nucleic acids are not well preserved and experience chemical damage and fragmentation.\textsuperscript{68} Due to this limitation, fresh frozen tissue, which does a considerably better job at preserving nucleic acid structure, is needed for long-read sequencing.\textsuperscript{68}
Therefore, with certain modifications to the process of genomic sequencing in a clinical setting, a wider variety of SVs can be determined, which could both serve as an important way to detect new mutation targets as well as tailor treatment to an individual’s specific mutations.

**Conclusion**

Clinical oncology is continuously evolving as new technologies are developed and improved upon. The progress made in this field within the last few decades is commendable, and an increasing number of people are receiving personalized cancer treatment driven by genomic sequencing. Current areas of interest include incorporating newer techniques, such as SMS, to improve the detection of SVs through their ability to generate long reads. Thus, new technologies continue to expose the underlying genomic complexity of cancer that continues to humble experts in the field. However, no technology is perfect: each has its advantages and disadvantages. SMS suffers from higher error rates than NGS technologies. The future of clinical oncology likely lies in a combination of SMS and NGS, along with the development of bioinformatics pipelines that process long-read data, to leverage the unique benefits of each approach to deliver different kinds of actionable insights.

**Acknowledgments**

I want to thank radiation oncologist, Dr. Gauravjit Singh for guiding me throughout the process of writing this literature review and providing feedback on my work.

**References**

22. *FoundationOne® Heme | Foundation Medicine*. FoundationOne Heme Overview. FoundationMedicine®, 2021. https://assets.ctfassets.net/w98cd481qy0p/4mnCQaQuXydmH2lI5ujuZ/45a31ee91f38e090c0099e8a1c9a3/Foundatio nOne_Heme_Overview.pdf (accessed 07-02-22).
23. *FoundationOne® Liquid CDx | Foundation Medicine*. FoundationOne Liquid CDx Technical Specifications (Gene List). FoundationMedicine®, 2021. https://assets.ctfassets.net/w98cd481qy0p/wVEm7V7ICYR67S5C1VbU7/f055e0476 183a6acd4eaeb583a3a00/F1LCDx_Technical_Specs_07 2021.pdf (accessed 07-02-22).


**Authors**

Aashna Soni is a high school student at BASIS Independent Fremont, Upper School. She is interested in biology, particularly how tools like machine learning can accelerate progress in areas such as cancer research. She hopes to pursue a career in biological research. In her free time, Aashna likes to code and write short stories.
Informal Economy and Spending Shares

Tanush Sawhney
Sanskriti School, Dr. S. Radhakrishnan Marg, Chanakyapuri, New Delhi, Delhi - 110021, India; tanushsawhney@gmail.com
Mentor: Prof. Ceyhun Elgin

ABSTRACT: This paper conducts an empirical analysis to understand the relationship between the size of the informal economy and the composition of GDP in the form of its spending shares. I use comparable cross-country panel data from Penn World Tables and regression to model the relationship. The results suggest that the spending shares of consumption, investment, government consumption, exports, and imports have different yet interrelated relationships with informality (measured as a share of GDP). Investment, exports, and imports result in significant negative covariances, and consumption in significant positive covariance. The paper proceeds to analyze the possible hidden explanatory variables underlying these relationships, endeavoring to understand how different expenditure compositions of GDP across time and political geography can be applied to understand informality.

KEYWORDS: Behavioral and social sciences. Shadow economy; informal sector; consumption; investment; government spending.

Introduction

This paper studies the relationship between the degree of informality with the constituents of the expenditure side of GDP. Commensurate cross-country panel data from the Penn World Tables and comprehensive estimates of informality based on a deterministic DGE model from Elgin's The Informal Economy are used with regression to model the relationship.¹² The spending shares of GDP have distinct yet connected relationships with the informal economy (as a share of GDP). The results indicate that while the consumption share has a positive covariance, investment, exports, and imports have a significant inverse relationship with informality. Possible channels of action resulting in these relationships are analyzed by discussing likely hidden explanatory variables.

Studying informality is one necessity for inclusive growth and development for large proportions of the world. An ILO report estimated that more than 60% of the world's population made their living in the informal sector in 2018,³ with a likely increase due to the countercyclical safety net-like features of the informal sector as a result of COVID; it only gains importance. Expanding our understanding can aid in effective policymaking and interventions, and this paper hopes to contribute to developing the frontier of this understanding by the humblest bit more.

With Amartya Sen's seminal conception of development,⁴ in "the enhancement of freedoms that allow people to lead lives that they have reason to live," inclusive growth expands opportunity. It provides this safety net, especially for those that are most vulnerable. The informal sector is an essential contributor to this, cushioning the economy during downturns and employing assets and firms that could not operate in the formal sector due to obstacles to involvement, discrimination, and lack of resources, both knowledge and skill-based, and financial.

The literature on informality has developed considerably in the recent era, with various contending models emerging. Alternative definitions are a key feature of the discourse, with the alternate problem componentization playing a key role in developing policies. We establish the context for analysis through a non-exhaustive list of key publications.

When characterizing definitions of informality, we must look at its ontology. In his pioneering work, anthropologist Keith Hart put forth the informal economy as a term.⁵ Through the development of the literature, there are many indicators for the determination of the boundaries of the informal sector ranging from the size of firms, registration status, the legality of form, entitlements, and property rights, as well as terms of employment and social protections.

In "Informality and Development, "La Porta and Shleifer support the dual view of informality consistent with Lewis,"⁶ seeing the formal and informal as segregated in serving different customers by producing other products and using different inputs. This is supported by demand constraints modeling due to the differences in consumer income. This is further added to supply-side constraint modeling through a lack of managerial human capital, which tends to produce a much more significant effect than the human capital stock of workers. However, Ulyssea, through the own work of the author and the literature, shows the coexistence of formal/informal firms in industries contradicting the dualistic model with the lack of a missing middle in firm productivity distributions.⁸ The author discusses that endogenous variables, conditional on skill and self-selection, can largely explain the wage gap between formal and informal workers. Loayza defines informality as the set of firms, workers, and activities outside legal frameworks with indications of noncompliance and precariousness of agent situations.⁹

When looking at the determinants of informality, through empirical analysis of Korea, Chile, and Peru, La Porta and Shleifer suggest that the growth of output largely drives the increase of formalization in the process of economic growth through formal firms and stagnation of informal firms.⁶ La-
bor force growth has an associated negative covariance to formalization. Loayza marks the causes of the informal sector as inefficiencies in the production and delivery of government services, with excessive regulation causing non-compliance.⁹

E. Dabla-Norris et al. point to the inverse relationship between the quality of the legal system and the degree of informality in the economy.¹⁰ The quality of the legal system is shown to reduce the elasticity of informality to regulatory burdens. Preliminary findings link the importance of private contracting institutions to reducing informality compared to constraints on the legal executive. A model is developed to understand decision-making for agents in choosing between running firms in the formal or informal sector or seeking employment in the same. Financial development, as a whole, is found to be statistically insignificant in being correlated to informality.

Friedman et al. show that overregulation is a key indicator of a larger informal economy due to the excess burden imposed on firms in addition to taxation.¹¹ Tax rates themselves, controlled for GDP per capita, do not seem to affect the size of the informal economy, possibly because while higher tax rates would encourage informal production, the better legal and supportive environment they may provide with increased public resources may incentivize official production. Corruption and weaker institutions were shown to be correlated positively to the size of the informal economy, with a perceived lack of delivery of public goods by the government and excess burdens encouraging the entrepreneur to shift production to the informal economy, further reducing revenue and hurting delivery of public goods in a vicious cycle.

In discussing interventions to increase formalization, La Porta and Shleifer suggest that an addition to the magnitude of the costs of informality through taxation and regulation of informal firms can have net adverse effects on the economy.⁶ This explanation is based on a characterization of informal firms as lacking human capital with inherent low productivity, making growth and survival unsustainable for them in a competitive formal sector. Such policies would, hence, more likely drive them out of business, leading to the net effects of greater poverty.

Meanwhile, Ulyssea divides the policy interventions to deal with informality as increasing the costs of informality or reducing the costs of formality (or increasing its benefits).³ The author analyzes that reducing entry costs, applicable to De Soto’s view,¹² has the least effect on formalization. A greater effect is observed for lowering costs of staying formal, not only for entry, and intensifying government crackdown on informality. The author differentiates between the extensive margin of firm compliance and the intensive margin of the degree of social protection of employment. The paper observes that interventions on the extensive margin are more effective in creating productivity and output effects than the intensive margin, which could produce adverse effects.

Loayza suggests a homage to Danish flexicurity with labor market flexibility and social protection combined with tax rationalization and an efficient regulatory and judicial environment.⁹ The paper recommends using both punishments on informality and emphasizing the benefits of formality in specific comprehensive plans adjusted per country based on political considerations of economic and social costs.

While spending shares have been studied with regard to economic growth, this paper is unique in exploring the linkages of informality with the composition of GDP in the form of the spending shares of the interaction of its actors. This paper helps expand the understanding of the determinants of informality through their action through the channel of structural differences in the composition of GDP. It can lead to the development of the composition of GDP as a key proxy for identifying specific actions of the determinants.

The rest of the paper is organized as follows: In the next section, I will discuss the empirical methods used in the paper, before which I present the data used, discussing its sources and descriptive statistics. Then in section III, I apply these methods and show the results I obtained. Finally, I provide some concluding remarks and discussion in the last section.

### Methods

#### Data and Conceptual Framework:

The data about spending shares as a proportion of GDP is obtained from the latest edition of Penn World Tables 10.0 prepared by the Groningen Growth and Development Centre.¹ Data used is the spending shares as a proportion of CGDPO, calculated at the output side at the current purchasing power parity, enabling us to understand relative productive capacities across nations at different periods of time by holding prices constant, giving it centrality in the analysis. Finally, the author thanks Elgin for the informal sector (as a percent of the GDP) series, calculated based on a deterministic dynamic general equilibrium model.² After data adjustment to remove outlier values and ensure the availability of complete data for all observations, we have 7791 observations across 154 countries. These range from the years 1950 to 2017, depending on availability.

The underlying data for spending shares are most popularly known for estimating Gross Domestic Product through the expenditure method, wherein the summation of spending by different economic actors takes place. However, they can shed light on other key parts of the economy. In terms of GDP being classified as Y, as a monetary measure of output produced within the territorial boundary of a nation within a given year, spending shares are defined using the following system of equations:

\[ Y = C + I + G + X - M \]

that is, Gross Domestic Product = consumption expenditures + investment expenditures + government expenditures + exports – imports.

Hence, we define the spending shares of GDP in the following manner. These are real shares of GDP to control for differences in relative price levels between the sectors where expenditure is measured. Such a method finds basis in earlier work studying the relationship of constituents of GDP with other factors such as productivity.¹³
This is also known as the share of household finance consumption expenditure (HFCE) or private final consumption expenditure (PFCE). It refers to the expenditure undertaken by private actors to satisfy needs or wants. All included goods and services are considered consumed in the given year and not stored, regardless of status as durable goods like cars and electronics. It is important to note that this includes transactions covered in whole or partly by unilateral transfers from the state. Therefore, it consists of the amounts spent by the household on activities such as healthcare, education, and housing, after subsidization by the government. These transfer payments are not taken to be part of the Government's final consumption expenditure due to actual spending of goods and services happening through the households. Households will include in their consumption expenditure remuneration received in kind from their employers. The value of goods produced for self-consumption, such as the imputed value of rent from owner-occupied housing, will be included in the same.

It has largely remained stable around the 70% mark for the United States over the past decade, pre-pandemic, and averages around the 60% mark for India in the same period. The mean size for consumption share is 0.624 in the data, with a standard deviation of 0.159. The values ranged from 0.066 to 0.999, with a median value of 0.626. Negative values and values above one were removed from the data due to the likelihood of such values being outliers and possible inaccuracy.

**I/Y: Investment Share:**

In terms of GDP calculations, investment is componentized into gross fixed capital formation, inventory investment, and the net acquisition of valuables by enterprises and households (NAV). Drawing from Nurkse, its meaning lies in society's choice to not apply the whole of its production capacity to present desires of consumption but instead devote a considerable portion to its future consumption through the expansion and maintenance of its productive capacity in the form of capital goods such as plants and equipment, transportation and communication infrastructure and other instruments that can increase the productivity of effort.

It is an essential disaggregate of GDP due to the role played by capital accumulation in economic theory as the basis for growth and proved through empirical evidence. The composition and amount of this capital stock and changes in productive activity are often analyzed through gross fixed capital formation. Therefore, with the centrality of capital formation in growth and raising productivity, a parameter on which low performance plays a role in definitions of informality becomes important in our discussion.

Inventory investment corresponds to the change in stock lying with a producer, usually kept to meet unanticipated manufacturing or sales variations. The term also conforms to those currently part of the aforementioned processes. For example, a result of stock produced within a given year that remained unsold is included in the domestic production for that year, and discounts for previously produced inventory may have been sold in the given year. Hence, a net change in the stock is used to calculate GDP.

Gross fixed capital formation is further divided. It is divided into three components. The first is fixed capital investments by firms or the net change in the value of a producer's fixed assets, i.e., those tangible or intangible assets used repeatedly in the production process for more than one accounting year.

An essential point to note is that this differs from ordinary intermediate expenditure on the maintenance of the current capital stock but will include improvements to fixed assets expected to extend their length of usage and production capacity. This will also include commercial and industrial buildings and other constructions of the like by the firm. Public Investment in similar fixed assets, and notably in the fashioning of public infrastructure, including the establishment of roads, railways, schools, hospitals, etc. Finally, expenditure on residential dwellings is part of this. Since the 1993 UN System of National Accounts, valuables classified as precious stones and metals not to be used in the process of production, works of art and antiques, and other valuables have been included as a part of the capital account at the actual or estimated values amounts payable on the transferring of ownership, inclusive of any associated transaction costs. Three approaches are generally used for estimating the form of commodity flow, expenditure, and financing approach.

Investment shares are centered around a median of 0.208 with a mean of 0.215 and a standard deviation of 0.105. The data ranged from 0.004 in Bulgaria in 1996 to 0.95 in Nigeria in 1997. Negative values and values above one were removed from the data due to the likelihood of such values being outliers and possible inaccuracy.

**G/Y: Government Spending Share:**

The government spending share of GDP includes expenditure by the government on purchasing goods and services as part of general government expenditure. Spending by the government on producing nonmarket final foods and services and social transfers provided in kind will be included in the same. It does not include transfer payments made to consumers, including social security, which would be included in consumption, or to firms with subsidies for capital investment as that is included as part of the investment. Contribution to Gross Fixed Capital Formation will be included as part of the investment, and its addition will lead to the share of government output, a different figure.

It can be componentized into various expenses, including, first, a group of costs reflecting collective consumption that provides utility for society as a whole or a significant proportion of it and are hence known as public goods and services. The second would focus on those focused around an individual household, which are the social transfers in kind mentioned earlier and market goods and services provided to households. An important note here is that since government-provided goods and services often do not possess a market price, they are valued at the sum of the costs of their production, which may include compensation of government employees, intermediate consumption, and depreciation. This division between individual and borderline consumption may not be well-
defined, such as expenses of Ministries of Health at a national level being collected, during the funding of a particular hospital or individual. However, this differentiation is irrelevant to our analysis and can be dismissed.

The size of the government’s final consumption expenditure in GDP ranged from 0.005 in Nigeria in 2003 to 0.816 in Guinea-Bissau in 1980. The average size was 0.183, with a standard deviation of 0.0913. The value at the 50th percentile was 0.165.

**XY and M/Y: Exports and Imports Shares:**

Exports include the sale of goods and services, included in the boundary of production of GDP, from residents to non-residents. Barter transactions or goods sent abroad as part of gifts or grants will also be included. Imports follow the reversed criteria, with goods not included in this boundary of production bought from non-residents by residents. Since imports enter the calculation of GDP as a negative term, we use negative import share as the relevant variable of analysis in this section.

The definition of residents here hinges on the center of economic interest of the individual residing in a particular economic territory, corresponding to the political boundaries of a nation-state, its territorial waters, and enclaves in the world. The physical passage of goods within a country’s borders may not be necessary for this classification, except for food consumed in ships or planes and goods produced in the international territory by residents and then sold to non-residents. Similarly, goods sent abroad for minor processing and transportation equipment are not classified as exports when they pass international boundaries. The shares are defined as the value of exports over the value of GDP, and similarly for imports.

These shares are dependent on not only the country but the world situation. Relative prices between foreign and domestic products influenced by changes in real exchange rates due to interest rates, political factors, changes in product demand, etc., will affect these proportions yearly. Recessions or booms within international markets will also affect these interactions with the “Rest of the World” for each nation and have multiple effects on the domestic economy within a considerably co-dependent world.

The average size of exports as a ratio to GDP was 0.223, with a standard deviation of 0.247. The dataset is skewed to the right, but with a significant number of observations, it should be fine for our analysis. The median value is 0.4, ranging from 0.00002 in Georgia in 1994 to 2.822 in Singapore in 1995.

The negative of the import share has a mean of -0.257 with a standard deviation of 0.263. The minimum value, or the largest absolute value of import share, was [-3.392] in Singapore in 1995, with the smallest absolute value of [-0.0001] in Georgia in 1994, corresponding with the export counterparts. The median value is -0.177.

**Informal Sector Share:**

A discussion on definitions must also describe the various interchangeable terms that tend to be used, such as the shadow economy, underground economy, and black or murky economy, indicating in itself the importance of the definition in the direction of the debate and research. The margin definition from Ulyssea provides a vital building point in the form of the distinction in the form of the extensive margin of informality, defined with regard to the payments of entry fees and legal registration by firms. However, it is important to note that mere registration does not indicate the following distinct practices from other “informal” firms. The intensive margin focuses on whether firms that are formal in the first sense hire workers without a contract. The precariousness and vulnerability inherent in informal employment are captured here, but this may be subject to different criteria across geographies and cultures. Illegality in labor practices, tax avoidance, unaccounted-for production, and lack of differentiation from household finances may be other characteristics.

The share of the informal sector is defined as the value of the output of the informal sector upon the value of gross domestic output. Our dataset’s mean informal sector size (unadjusted for weightage by GDP) was 0.346, with a standard deviation of 0.145. The median value was 0.337. Values ranged from 0.079 in Switzerland in 2017 to 1.126 in Thailand in 1960.

Elgin notices that the share of the informal sector as a percentage of world GDP has decreased due to lower informal shares in higher-income economies, though it does so at a decreasing rate.² The relationship between per capita income and informality may be somewhat non-linear, but Elgin and Birinci find a U-shaped relationship between informality and economic growth.

### Table 1

Below provides descriptive summary statistics of all variables used in the empirical analysis.

<table>
<thead>
<tr>
<th></th>
<th>GY</th>
<th>IV</th>
<th>GY</th>
<th>M</th>
<th>NY</th>
<th>EY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.624</td>
<td>0.216</td>
<td>0.183</td>
<td>0.233</td>
<td>0.237</td>
<td>0.546</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.159</td>
<td>0.108</td>
<td>0.091</td>
<td>0.247</td>
<td>0.363</td>
<td>0.145</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.000</td>
<td>0.004</td>
<td>0.005</td>
<td>0.000</td>
<td>0.002</td>
<td>0.079</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.999</td>
<td>0.555</td>
<td>0.816</td>
<td>2.822</td>
<td>0.990</td>
<td>1.126</td>
</tr>
<tr>
<td>Median</td>
<td>0.026</td>
<td>0.208</td>
<td>0.168</td>
<td>0.140</td>
<td>0.177</td>
<td>0.337</td>
</tr>
</tbody>
</table>

### Analytical Methods:

Our empirical analysis methodology will be based on a correlation analysis of the scrubbed data. A linear relationship between expenditure components of Gross Domestic Product and Informal Sector Size is calculated in terms of the Pearson Product-Moment Correlation to understand its intensity and direction. Cross-country and longitudinal regression methodologies allow for the accounting of structural differences, political, economic, social, etc., to a considerable extent. It is reported in the following section. The correlation methodology is based on calculating a line of best fit between the two variables, and its measure r indicates how close data points are to this line. We then form an ordinary least squares linear regression-based model with the constituents of GDP as our dependent variable to attempt to predict levels of informality with regard to constituents of GDP.

Before this analysis, we confirm that our data is appropriate. Our variables are continuous and measured at the ratio level. We assume a linear relationship exists between them and test this by applying correlation. For the same, as described in the Data section, they have been scrubbed for outliers. Normality is assumed as n is a large value for this t series. It is important
to note that inputs about causation cannot be drawn as a result of correlation analysis due to the possible presence of confounding variables. As a rule, the fluctuations in variable x are caused by fluctuations in multiple variables.

I will calculate and report correlation coefficients between informal sector size and spending shares of GDP. As well known, a correlation coefficient always takes a value between -1 and 1. A negative correlation implies that the two series are moving in the opposite direction, whereas a positive correlation suggests that they move in the same direction. However, not all correlation values are statistically significant. Generally, as well known, the rule of thumb is that a positive correlation should be larger than 0.10, and a negative correlation should be smaller than -0.10 to be statistically significant. I will also supplement the correlation analysis with scatter plot diagrams, utilizing trendlines to understand the effects of the correlation.

**Results and Discussion**

Table 2: Correlations of Spending Shares with Informal Sector Size.

<table>
<thead>
<tr>
<th></th>
<th>C to IS</th>
<th>I to IS</th>
<th>G to IS</th>
<th>X to IS</th>
<th>M to IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.299</td>
<td>-0.346</td>
<td>0.0312</td>
<td>-0.376</td>
<td>0.339</td>
</tr>
</tbody>
</table>

Table 2 presents correlations of the informal sector with all the individual spending shares. Now having shown these correlations below, we will plot each relationship on a separate scatter plot diagram and discuss the nature of the relationship between the two variables. The author explores possible explanatory variables behind the relationships shown as a result of the analysis.

**Consumption Share and Informality:**

![Figure 1: Consumption share and informal sector.](image)

Consumption as a share of GDP sees a statistically significant correlation with the informal sector. Table 2 indicates a moderate positive association with an approximate value of 0.3. The value of R2 is 0.105 for the regression model, indicating that the change in the consumption share explains 10.5% of the change in the informal economy. The model given by the line of best fit shows an approximate relationship of an increase of 0.0027 in the size of the informal share for a given increase of 0.01 in the share of consumption as a proportion of GDP. An initial theory was that high and rising corruption had been linked to income inequality and poverty increases. Many in the literature on informality have also expounded on the effect of this corruption as an effective additional regulatory burden, incentivizing many to switch to the informal sector. This common variable of corruption can possibly explain the relationship between informality and consumption.

The 2012 Economic Report of the President postulated that sluggish growth in consumer spending might reflect the sharp divergence of the income distribution in the United States with increased inequality. There are two related reasons in economic theory for the same. The first, composed of the Keynesian view of consumption, predicts a reduction in consumption with increased inequality due to the lower marginal propensities to consume the more affluent households, earning greater proportions of the country’s income.

The second, as in Inequality and Aggregate Demand, relates to a more permanent change. The volatility in incomes associated with low-income households, especially in developing countries, leads to permanent raises in aggregate savings. Many of these households are likely to be involved in the informal economy due to their low-income, low-productivity nature. They may have a proportion of this volatility derived from such involvement, especially in developing countries. This would indicate that lower spending on consumption could be correlated with higher informality, which doesn't seem to be the case. Recent literature suggests that an empirical relationship between consumption and inequality may not exist as idealized in theory, perhaps explaining why the initial hypothesis is inaccurate.

To attempt to explain the positive relationship, we may rely on the equations of national income; we know that injections and leakages from the flow, assuming the effect of the balance of payments is held constant, give us

\[
Y = C + S + T
\]

Disposable income here is given by \( Di = C + S - T \)

Where Y is national income which is spent in the following components as C is consumption, S is saving, T is Taxes.

With the effect of constant government expenditure, taxes, and constant income, a greater consumption share implies a smaller share of savings. Saving finances the supply of loanable funds with a constant balance of payments. Greater shares for consumption will mean less money available for investment, raising interest rates and disincentivizing the investment expenditure, reducing its share in GDP. This will lead to a greater consumption share implying a smaller investment share, and hence greater informality, possibly explaining the positive relationship.

Understanding the effect of the Balance of Payments, as a result of decreased savings, some amount of investment could be met through net capital inflows, which could increase real exchange rates. This would cause a reduction in exports. While there would be an increase in imports relative to exports, absolute values may decrease due to the lesser availability of foreign exchange. As explained later, this reduction in the shares of exports and possible reduction in the import share, which are negatively related, point to greater consumption’s positive relationship through these channels.
credit that can aid expansion and small firm size, relative low burden in a difficult situation with regards to securing formal titation that may be characteristic of firms avoiding regulator noncompliance and evasion of state rules, put enterprises in the isolation of activity that defined informality,²³ along with the low productivity growth outcomes. The precariousness and has emphasized the role of credit constraints miring firms in constraints offers an additional viewpoint. Research on informality hinging on the definition of informality hinging on the escape from the regulatory burden and as a means of evasion and non-compliance would indicate that informal firms face a constant risk of the confiscation of their assets by state authorities, discouraging investment in the fixed capital stock of considerable value due to uncertainty of its future availability for use. Regulatory burdens, compounded by firms’ lack of compliance and preference for evasion and avoidance of costs of navigation of the system, make it likely that securing permissions that may become necessary for expansion will be difficult. Difficulties in enforcing property rights may add to the uncertainty, with possible questions over asset ownership and acquiring other resources that may be necessary to fuel expansion and drive investment. The strong negative association further aids this argument in the quality of the legal sector and, as a result, the enforcement of property rights and informality.¹⁰ Thus, greater investment could happen due to lower regulatory burdens and better property rights, which are positively associated with informality.

Second, the relationship between informality and credit constraints offers an additional viewpoint. Research on informality has emphasized the role of credit constraints miring firms in low productivity growth outcomes. The precariousness and isolation of activity that defined informality,²³ along with the noncompliance and evasion of state rules, put enterprises in the informal sector, with the absence of registration and documentation that may be characteristic of firms avoiding regulator burden in a difficult situation with regards to securing formal credit that can aid expansion and small firm size, relative low-er asset wealth may make terms of credit unfavorable. This would lead firms to forgo investment or access normal sources charging high-interest rates, reducing their quantity. The subsistence functioning of parts of the sector would not allow the accumulation of enough capital for investment. Lower levels of entrepreneurial capital in the sector would add to the credit constraints discussed in making investment unlikely.

Third, we focus on the role of productivity in informal activity. Much of the work on informality has emphasized that extensive crackdowns on the sector would see a fall in employment and output. This has been characterized by the inherent low productivity nature of the sector due to the relative lack of capital and selection, some of it self-selection, of lower productivity inputs and processes.² Due to these reasons, survival in the formal sector with the costs of compliance with taxes would make firms uncompetitive and unable to survive. By increasing the capital stock available to these firms, investment would increase productivity and allow firms to create conditions that would enable them to move to the formal sector in a way that allows them to partake of enough benefits to balance costs. Additionally, catering firms to different markets is a point to consider. Many theorists point out that the informal and formal sectors seem to cater to different sets of consumers altogether, with the informal sector producing more low-quality, lower-cost goods for lower-income consumers. Drawing on the fundamentals of Smith in the Wealth of Nations,²⁴ the investment would require larger production and a larger market to ensure its feasibility. Due to the nature of the market, the normal sector caters to would make it infeasible.

Considering that these low-income markets could contain unskilled laborers’ characteristics of low productivity used in the definitions discussed above is important. The informal sector tends to engage in unskilled labor-intensive activity in a small-scale, household-financed way, consistent with many studies.²⁵ The availability of unskilled labor makes a case for low-cost conditions that make it preferable to capital as a factor of production due to the substitution effect, discouraging investment in capital stock. These low wages could add to the cycle of a differentiated low-income market. This is consistent with models that view unskilled labor and capital as substitutes in the informal sector, while skilled labor may act as a complement.²⁶

**Government Share and Informality:**

![Figure 3: Government Spending Share and Informal Sector.](image-url)
The share of final government consumption in GDP shows a statistically insignificant relationship with informality, with a value of r below 0.1 and near zero, as from Table 2. The scatter plot of the informal sector and government spending share is shown in Figure 3. The linear regression model is inapplicable here. This can be explained by the action of opposing factors in the relationship between government spending on informality.

Government spending through the provision of public goods, including infrastructure, which could make formality’s benefits more apparent, and health and education, supporting increasing formalization, aiding the reduction of informality, would indicate a positive covariance. However, its ability to finance this expenditure could be severely impaired by the presence of a large informal economy from which resources cannot be collected. The perceived lack of delivery of public goods in particular periods, especially during political and economic unrest, can incentivize the shifting of production to the informal economy, which in turn can make the delivery of these public goods more difficult, supporting a negative covariance. Hence, while higher tax rates to finance government spending as a share of GDP could incentive informality through the imposition of a fiscal burden, the better legal or regulatory environment offered could counteract and incentivize official production. This is consistent with work around the regulatory burden in the field.⁹,²⁷

Wagner’s law points to an absolute and relative increase in spending by the government with the growth of per capita income, with an essential long-run elasticity of public spending above 1. Studies have found empirical evidence for the same.²⁸ Informality, however, has a strong negative linear covariance with per capita GDP.⁶ Thus, increases in per capita income, a decrease in informality, and an increase in relative nominal government expenditure are associated, pointing to a negative correlation between informality and government share of GDP.

However, increases in per capita income may lead to relative price level increases in government expenditure through the Balassa-Samuelson effect, reducing the impact on the share of real output. Government expenditure is estimated mainly through the production of the non-tradable provided by it to the public, which would exhibit greater price levels with regard to tradable goods, with growth coming from productivity. Stagnation in the share of real output could mean no real change in the provision of public goods, explaining why informality is not linked to the government’s share in output.

Much public corruption, however, finds itself linked to government involvement. Increasing the spending share of the government could indicate more opportunities for rent-seeking by the government bureaucracy and greater imposition of a regulatory burden. This would again incentive informality, with the costs of corruption effectively adding on, as discussed in the section on consumption.

**Exports Share and Informality:**

The data shows an empirical relationship between exports as a share of GDP and informal share in Figure 4. A statistically significant negative association of correlation coefficient value of -0.376 is present, as shown in Table 2. In the regression model, a value of R² of 0.141 indicates that the variation explains 14.1% of the variation in the share of the informal sector in exports as a share of GDP. The line of best fit offers a model with an increase in exports of 0.01 as a share of GDP associated with a decrease in the share of the informality of 0.0022 basis points.

It stands to reason that large amounts of exports as a percentage of GDP could mean a large proportion of the population engaged in such activities. The exporting firms will require industrial production bases, possible only through investment. If exports are to remain or become competitive, they may need to keep up with upgrades in technology requiring investment. Hence, reasons applicable for investment being negatively associated with informality could extend to exports through its indirect effect. Additionally, exports tend to need financial support in the exchange of currency, insurance contracts, and credit until payments are collected, making it difficult for informal firms to access due to the financial frictions they tend to facts.

Greater trade through exports and imports can result and indicate greater integration in the world economy. This leaves countries more dependent on others and more subject to global standards involving labor laws and protections that may require crackdowns on the informal sector, leading to its reduction. This greater integration, acting as greater knowledge transfer and Foreign Direct Investment inflows, can lead to rapid technological development, as experienced in Japan and South Korea in the second half of the 20th century, where American investment fueled growth. A resultant decrease in the size of the informal sector with greater productivity labor and individuals is probable, with firms and individuals able to better take advantage of and ensure delivery of formal sector benefits such as public goods and finding it more challenging to hide production from the government eye due to its absolute increase. This is seen in Figure 5.1; whereas real GDP per capita increases over the years, informality decreases. In Figure 5.2, on plotting real GDP per capita logarithmically on the horizontal axis, we can observe an almost constant percentage rate of change in informality.
Development literature suggests that higher percentages of exports to GDP indicate higher growth rates. This context may hold relevance in our discussions of informality, considering established relations between growth, productivity, and informality. The Dutch disease possibility may offer some outliers in this analysis.

While the discussion around informality and economic growth points to a non-linear relationship, its componentization can help form a possible theory. Elgin and Birinci (2016) find support for a negative correlation of informality with economic growth in low-income economies.¹¹ The literature has suggested that formal firms can better take advantage of higher investment credit opportunities, engage in asset collateralization efficiently, and increase access to public goods like infrastructure.¹² Their resultant higher productivities may allow them to become competitive exporters, aiding the share of such in GDP.

Economic growth is positively correlated with increases in the export share of GDP.³⁰ The typical relationship with exports is positively associated with economic growth. However, economic growth is negatively associated with informality, which may explain some parts of the relationship between exports and informality. Elgin and Birinci also found that in high-income economies, which may tend to have larger shares of exports in GDP,¹⁹ informality may cause increases in growth. This can explain the panning out of the effect at higher levels in our analysis.

Exploring productivity further, as alluded to above, while variations among economic sectors exist, studies have found support for higher productivity in exporting firms compared to those focused on the domestic market.³¹ In Taiwan, Chen, and Tang found contribution to productivity improvements through exports due to channels including economies of scale.³² As economies become more export-oriented, with exports consuming a larger share of the proportion of GDP, firms that contribute to more significant proportions of this export-oriented GDP are likely to be formalized, considering the productivity differentiations inherent in them from the informal economy and in the access to capital, investment, and quality of inputs available to them. There will also be additional productivity derived from entrepreneurial capital that is more likely to be employed in the formal sector.¹⁰

Imports Share and Informality:

The action of the negative imports share corresponds to an associated increase in informality, underlying a statistically significant negative relationship between imports and informality, as seen in Figure 6. Table 2 gives us a correlation coefficient value of 0.339 with the negative of the import share. The value of R² of 0.115 indicates that 11.5% of the variation is explained by the variation in the share of imports. The line of best fit offers a model wherein an increase of 1/100 in the imports share corresponds to an associated decrease of 0.0019 in the informal share of GDP.

Similar to exports, imports as a share of GDP have shown positive covariance with economic growth.³⁰ In a similar manner to as suggested following this discussion in the export result in analysis, imports are suggested to be negatively correlated to the share of the informal sector. For the growth of imports as a share of GDP, either the current account will have to grow, likely fueled by merchandise trade, making the reasons for exports valid, or through transfers such as remittances which are limited in terms of growth ceilings comparatively. Surpluses on the capital account could fuel imports. Still, incentivizing inflows would require good governance, favorable laws for firm performance, and economic growth, all factors associated with reductions in informality. Similarly, the international aid part of the current account could also be contingent on the implementation and upholding of specific

---

**Figure 5.1:** Real GDP per capita sourced from the World Bank national accounts data (NY.GDP.PCAP.CD) has been plotted with the informal sector share from Elgin (2021) for Japan and the Republic of Korea in 1960-2000.

**Figure 5.2:** Real GDP per capita sourced from the World Bank national accounts data (NY.GDP.PCAP.CD) has been plotted against the informal sector share in Japan (on a logarithmic scale).

**Figure 6:** Imports Share and Informat Sector.

The action of the negative imports share corresponds to an associated increase in informality, underlying a statistically significant negative relationship between imports and informality, as seen in Figure 6. Table 2 gives us a correlation coefficient value of 0.339 with the negative of the import share. The value of R² of 0.115 indicates that 11.5% of the variation is IS/Y is explained by the variation in the share of imports. The line of best fit offers a model wherein an increase of 1/100 in the imports share corresponds to an associated decrease of 0.0019 in the informal share of GDP.
standards, with greater integration in the world economy as a whole, similar to the case of exports, making countries more subject to regulations such as labor laws, leading to crackdowns on informality, and requiring certain practices, such as perhaps reduction in corruption, again being associated with lesser informality. Their relevance helps explain the relationship between imports and informal shares.

**Informality (logarithmic) and Spending Shares:**

Table 3: Informality (logarithmic) and Spending Shares.

<table>
<thead>
<tr>
<th>Correlation to log (S/Y)</th>
<th>CY to ln(S/Y)</th>
<th>I to ln(S/Y)</th>
<th>QV to ln(S/Y)</th>
<th>X/Y to ln(S/Y)</th>
<th>MY to ln(S/Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.325</td>
<td>-0.372</td>
<td>0.068</td>
<td>-0.401</td>
<td>0.358</td>
</tr>
</tbody>
</table>

Table 3 shows, for all covariances, stronger relationships when modeling using the logarithmic of the share of informality in the same direction of the association. The data indicates a panning out of the effect of the associated increase in informality at higher levels of spending shares relative to the mean values for the spending shares, with stronger effects observed at changes in lower levels of spending shares. Our analysis of probable causes remains valid and is characterized by this behavior consistent with the explanation that at higher levels, associated effects of share of spending are already near maximum and have a diminishing marginal impact.

**Conclusion**

The paper indicates a statistically significant positive relationship between the share of the informal sector and consumption, with a possible reason explored, including the resultant reduction of investment as a result of greater consumption. The investment share shows a statistically significant negative relationship, with property rights and regulatory burden, credit constraints, and the impact of productivity growth playing possible roles. Analysis shows government final consumption spending as a share of GDP to be unrelated, likely due to counteracting effects with a possible explanation of inverse relationship through the action of Wagner’s law and the association of income levels with informality tempered by the Balassa-Samuelson effect. Greater trade, both in the form of exports and imports, shows a significant negative relationship with the share of informality, expectedly supported by greater integration and economic growth and expansion of a higher productivity base for exports. Many of the spending shares likely have interlinked effects through their impacts on each other.

There is scope for a full-fledged econometric analysis in a multivariate model, with control for the effect of constituents on GDP on each other, to develop an understanding of the exact nature of the relationship between the constituents of GDP and informality. Modeling the impact of change in different shares in a dynamic equilibrium model could offer light on “ideal” structural compositions of GDP with regard to the informal economy. It will also be necessary to explore the composition of informality in terms of productivity, growth, and quality of employment and how that may be affected regarding the spending shares.

The study’s relevance is key in policymaking, perhaps holding answers to critical questions regarding expanding the tax base and providing social support to all vulnerable economic actors. Developing an understanding of the effects of the shares of GDP through the channels proposed in this paper, in terms of their extent and applicability, can help open up new ranges of policy options while illuminating the impact of structural differences across countries and, as the result of adopted policies in affecting the informal sector. The middling relationship with consumption relative to other spending shares for informality suggests it may not be the most effective channel, especially as our analysis points to multiple other explanatory variables, such as corruption and inequality. This understanding, however, could provide a greater impetus to dealing with the same through the methods of formalization. For economic growth, the paper helps indicate that informality is associated with a lower investment share due to disincentives caused by or as pre-conditions for it. Thus, crowding in large-scale investment across enterprise strata to unlock productivity gains would likely see a highly-associated decrease in informality. In the context of our analysis, the self-selection approach would suggest that investment in human capital, along with a reduction in regulatory burden and credit constraints, would be key to this process. The non-existent relationship with government spending can allude to how rather than the quantity of spending, the quality of norms associated with doing business and the preconditions responsible for it are more effective in dealing with informality.

Greater integration into the world economy associated with opening up markets to greater competition, incentivizing capital usage, and loosening burdensome restrictions would likely aid in reducing informality. A greater focus on export-oriented industries instead of domestic ones in the economy would incentivize businesses to innovate and compete. It would do the same through an additional channel, inviting a greater demand for public goods. The meeting of this demand could help disincentivize informality among other industries. The diminishing marginal impact observed in Table 3 indicates that a combination of such efforts will be necessary, rather than single silver bullets, which can achieve desired outcomes.

**Acknowledgments**

I want to thank my mentor, Professor Ceyhun Elgin, for his constant support and guidance in producing this research article and for the privilege of working with him, someone on the cutting edge of informality. He has been crucial in facilitating the expansion of my understanding of this consequential area of research.

**References**


DOI: 10.36838/v514.19


18. National accounts of OECD Countries. DOI: 10.1787/2221433x.


**Author**

Tanush Sawhney is a student in Grade 12 at Sanskriti School, New Delhi. The Best in Economics Award winner at the International Economics Olympiad, he is fascinated by public, development, and behavioral economics and their ability to facilitate impactful policy interventions in uplifting the most vulnerable.
A Rapid One-step Field-based Detection Method for Nanoplastics in Water

Vidhatri L. Iyer

University High School of Indiana, 2825 W. 116th Street, Carmel, Indiana-46032; vid2008mp@gmail.com
Mentor: Mr. Bob Roudebush

ABSTRACT: The widespread use of plastics has raised concerns about their persistence in aquatic and terrestrial environments. Tiny plastic particles, microplastics, and nanoplastics generated due to degradation lead to real environmental problems. Rapid identification of potential hotspots of nanoplastics contamination has been challenging due to a lack of field-based detection methods. This study aimed to develop a single-step portable method to detect nanoplastics in water samples. Using a custom-designed fluorometer, an optimized Nile Red-based method has been successfully extended to detect nanoplastics within 10 minutes from wastewater samples with a limit of detection of 35 µg/mL. This method can be broadly used to monitor nanoplastic load during wastewater treatment and in different surface water streams which drain into urban watersheds.

KEYWORDS: Earth and Environmental Sciences; Water Science; Nanoplastics; Detection Method; Field-technique.

**Introduction**

Plastics are widely used because of their low manufacturing cost, durability, and versatile use in many consumer goods. According to a recent study in 2020, the annual global production of plastic has increased to 360 metric tons per year.² The vast amount of plastic waste has caused environmental concern due to the contamination of soil and water bodies such as rivers and oceans. Moreover, plastics can adsorb many toxic aromatic hydrocarbons, heavy metals, and pharmaceutical and personal care products and serve as reservoirs for these toxic agents.¹Microplastics and nanoplastics can pose severe health issues such as inflammation and physical damage to lung epithelial cells.³ Quantification and tracking of plastic pollution have been further complicated due to the trapping of debris in complex matrices such as wastewater samples.¹,² Macroplastics are easy to detect by visual inspection. Still with additional expensive laboratory detection techniques, degraded plastic products such as microplastics and nanoplastics are often easier to quantify in field samples.

Small plastic fragments are generated due to the action of chemical and environmental agents such as, soil fungi. Commonly used household products, packaging materials, clothing and bath scrubs release many plastic fragments into the daily wastewater reservoirs.¹,² Plastic fragments > 100 nm to < 5mm are characterized as microplastics and fragments < 100 nm are called as nanoplastics. Microplastics and nanoplastics in river streams and oceans can harm aquatic life and human health.³ Microplastics but not nanoplastics in water can be detected under the microscope, but the method could be more reliable due to differences in size, transparency, and fiber types.⁴ More sophisticated techniques, such as Fourier Transform infrared or Raman Spectroscopy techniques, are useful for reliably detecting the chemical composition of the polymer types.⁴ However, these detection methods are expensive and often time-consuming for routine field use.

Microplastics are hydrophobic and can be visualized by staining with lipophilic fluorescent dyes such as Nile Red.⁵-⁸ The interaction of plastics with these fluorophores is facilitated in the presence of organic solvents such as chloroform, acetone, or methanol. The staining protocol is usually conducted with different microplastics using a filter paper method and quantifying the particles by visual or semi-automated counting techniques. The Nile Red method was proven effective in staining several plastic types, including polyethylene, polypropylene, and polystyrene.⁵-⁸ However, these staining protocols often involve several steps of sample preparation, washing, and data capture.

Two recent studies have described a plate-based method to detect nanoplastics using polystyrene beads to generate a standard curve.⁹-¹⁰ Plate-based readers are expensive and unsuitable for routine field-based testing. Moreover, the non-specific binding of nanoplastic beads to these clear plastic plates has not been ruled out. In the study by Gagne et al., known amounts of 100 or 50 nm diameter nanoplastics were added to tissue extracts prepared from freshwater adult mussels (Elliptio complanata), which shifted the emission peak to 623 nm from the normal Nile Red peak of 660 nm.⁹ The authors also tested the effects of detergents such as Triton X-100 or Tween-20 as a proxy for lipid rich-environments. The fluorescent intensity increased with increasing concentration of the detergents, consistent with the solvatochromic nature of Nile Red.⁹ A recent study that used similar experimental conditions to the Gagne et al. paper showed that the proportion of methanol over water in the microwells was not reproducible using a different motor rotor dye, DCVJ to detect nanoplastics.¹⁰ Methanol concentrations of 20% were optimal to detect a 620 nm emission peak with DCVJ. The study by Moraz and Breider, however, did not
compare the performance of Nile Red dye using these optimal conditions.¹⁰ As suggested earlier, these studies used 96-well clear plastic plates, which did not wholly rule out non-specific binding of Nile Red or DCVJ to the plates in the absence of nanoplastics.⁹,¹⁰

The goal of the present study was to develop a cost-effective, simple, portable, and preferably a one-step nanplastic detection method suitable for field applications. A custom-designed hand-held portable fluorometer with Ex/Em of 450/620 nm was used with commercial 50 nm polystyrene nanoplastic beads to optimize various assay parameters such as linearity of signal, effects of shaking, and assay incubation times. After optimization of the assay conditions, field samples from influent and effluent wastewater samples were filtered using a 0.45 µm syringe filtered before staining with Nile Red dye in the presence of 20% methanol. This optimized protocol from this study has broader applications to detect hotspots of nanoplastics contamination in urban and rural water streams.

### Methods

#### Instrumentation:

An easily portable hand-held single-tube fluorometer was custom-built by Amiscience Corporation (Fremont, CA). The instrument was set up to read a single wavelength with excitation/emission wavelengths of 450/620 nm. The instrument had a touch-screen LCD display and operated in either a 5V DC power adaptor or 4 AA batteries (Figure 1). Signals were read as relative fluorescent units (RFU) and converted to relevant plastics concentrations using a standard curve with 50 nm polystyrene beads.

![Figure 1: Hand-held fluorometer and mini-glass tubes.](ijhighschoolresearch.org)

#### Results and Discussion

**Instrument Reproducibility:**

Instrument reproducibility was studied by mixing 160 µL of Milli-Q water with 40 µL of either methanol or Nile Red working solution (final Nile Red concentration is 8 µM) and incubated for 10 minutes at room temperature. Results were generated from 7 independent trials. As shown in Table 1, the background signals from methanol and Nile Red solution were at 1000 and 10,000 RFUs, respectively, in all seven trials pointing to a high level of reproducibility in instrument performance.

<table>
<thead>
<tr>
<th>Trial No.</th>
<th>Methanol</th>
<th>Nile Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000</td>
<td>10000</td>
</tr>
<tr>
<td>2</td>
<td>1000</td>
<td>10000</td>
</tr>
<tr>
<td>3</td>
<td>1000</td>
<td>10000</td>
</tr>
<tr>
<td>4</td>
<td>1000</td>
<td>10000</td>
</tr>
<tr>
<td>5</td>
<td>1000</td>
<td>10000</td>
</tr>
<tr>
<td>6</td>
<td>1000</td>
<td>10000</td>
</tr>
<tr>
<td>7</td>
<td>1000</td>
<td>10000</td>
</tr>
</tbody>
</table>

**Standard Curve Generation Using the Bead Titration method:**

Calibration curves were generated using serial dilutions of polystyrene beads in water at 400, 200, 100, 50, 25, 12.5, 6.25, 3.125, 1.562 and 0 µg/mL. In addition, a mixture containing 160 µL of beads with 40 µL of 40 µM Nile Red solution was incubated for 10 minutes at room temperature. Samples were read at 450 nm. As shown in Figure 2, a linear relationship was obtained with increasing concentration of the beads with a limit of detection (LOD) of 35 µg/mL using the formula 3.3µ ± S, where µ is the standard deviation of the response and S is the slope of the calibration curve. Bead concentrations above...
Based on the optimization of different parameters, the protocol utilized to test nanoplastics in influent and effluent wastewater is shown in Figure 4.

**Effect of different incubation times on calibration curve performance:**

The impact of different incubation times on the slope of the linear regressions was tested. The assay mixture with increasing bead concentrations was incubated for 10, 20, 30, or 60 min at room temperature, and RFUs were read using the hand-held fluorometer. As shown in Table 2, the slope of the linear regression obtained at 620 nm is remarkably consistent across different incubation times. Therefore, an incubation time of 10 min for subsequent experiments was chosen for assay optimization.

**Table 2:** Effect of incubation times on slope of linear regression at 620 nm.

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>Nile Red</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>971.88</td>
<td>0.995</td>
</tr>
<tr>
<td>20</td>
<td>933.66</td>
<td>0.991</td>
</tr>
<tr>
<td>30</td>
<td>964.9</td>
<td>0.990</td>
</tr>
<tr>
<td>60</td>
<td>1003.6</td>
<td>0.990</td>
</tr>
</tbody>
</table>

**Effect of shaking on calibration curve performance:**

The effect of optimal mixing of the final solution without (Figure 3A) or with (Figure 3B) shaking was tested using an orbital shaker. Serial dilutions of bead samples were mixed with Nile Red dye and left on an orbital shaker at 200 rpm. Mixed samples without shaking were left on the benchtop for 10 min as control. As shown in Figure 3, the slope of linear regression obtained by both methods was very similar.

**A. No Shake (10 min)**

**B. Shake (200 rpm, 10 min)**

**Figure 3:** Effect of shaking on the calibration curve.

Based on the optimization of different parameters, the protocol utilized to test nanoplastics in influent and effluent wastewater is shown in Figure 4.

**Influent and Effluent parameters in a wastewater treatment facility:**

Sewage water is treated daily at the TriCo Regional Sewer Facility (Zionsville, IN). The plant collects about 4 million gallons of wastewater from the residents in the western half of Carmel, IN, and sections of Indianapolis, IN. Several water parameters are collected daily from influent and effluent water samples to monitor quality when the treated water is allowed back into the Indiana White River. The facility tested influent and effluent water samples collected on five independent days for different parameters and data was provided for this study. As shown in Tables 3A and B, the effluent samples showed a significant reduction in total suspended solids, total ammonia, and phosphorus content.

**Table 3:** Influent and Effluent water quality

<table>
<thead>
<tr>
<th>Day</th>
<th>Suspended Solids - mg/L</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Influent</td>
</tr>
<tr>
<td>1</td>
<td>214</td>
</tr>
<tr>
<td>2</td>
<td>244</td>
</tr>
<tr>
<td>3</td>
<td>280</td>
</tr>
<tr>
<td>4</td>
<td>220</td>
</tr>
<tr>
<td>5</td>
<td>164</td>
</tr>
</tbody>
</table>
Nanoplastics levels in influent and effluent wastewater samples:
Nanoplastics levels were measured using the one-step fluorometric technique. First, wastewater samples were filtered using a 0.45 μm syringe filter to remove all suspended debris. As described in the materials and methods section, 160 μL filtered wastewater was mixed with 40 μL of Nile Red working solution and incubated for 10 minutes at room temperature before reading at 450 nm. As shown in Table 4, nanoplastics were detected from influent samples from all five collection days. On the other hand, a significant reduction of nanoplastics was observed in effluent water samples from 4 out of 5 days of collection likely due to the trapping of nanoplastics in the sewage sludge waste. In all cases, the levels were near or lower than the limit of detection of the assay. Interestingly, the levels of nanoplastics were similar in the influent and effluent water samples on day 2. These data clearly point to the utility of the one-step nanoplastic detection technique to monitor plastic load in treated water samples.

Table 4: Nanoplastics in influent and effluent wastewater samples.

<table>
<thead>
<tr>
<th>Day</th>
<th>pH (I)</th>
<th>pH (E)</th>
<th>Phosphorus (mg/L) (I)</th>
<th>Phosphorus (mg/L) (E)</th>
<th>Ammonia (mg/L) (I)</th>
<th>Ammonia (mg/L) (E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.20</td>
<td>7.53</td>
<td>4.38</td>
<td>0.816</td>
<td>25.7</td>
<td>0.058</td>
</tr>
<tr>
<td>2</td>
<td>7.58</td>
<td>7.55</td>
<td>3.94</td>
<td>1.385</td>
<td>24.8</td>
<td>1.085</td>
</tr>
<tr>
<td>3</td>
<td>7.30</td>
<td>7.63</td>
<td>4.14</td>
<td>0.951</td>
<td>24.2</td>
<td>0.500</td>
</tr>
<tr>
<td>4</td>
<td>7.32</td>
<td>7.64</td>
<td>4.15</td>
<td>0.633</td>
<td>24.6</td>
<td>0.046</td>
</tr>
<tr>
<td>5</td>
<td>7.29</td>
<td>7.57</td>
<td>6.28</td>
<td>0.953</td>
<td>23.2</td>
<td>0.036</td>
</tr>
</tbody>
</table>

Table 5: Nanoplastics numbers in effluent wastewater samples. Standard 50 nm polystyrene particles were available at 25 mg/ml at a concentration of 3.64 x 10^4 particles/ml. Nanoplastics concentration at μg/ml was converted to particles per mL using this information from the package insert.

A comparison of the nanoplastics content of influent samples to the number of particles revealed a strikingly high nanoplastics load in all five collection days (Table 5).


Author

Vidhatri Iyer is a freshman at high school and is very interested in environmental research. She has conducted prior research to monitor water quality in a wastewater treatment facility and is passionate about developing field-based techniques to detect plastic contaminants in water samples. In addition, she is interested in a medical career and would like to be a physician-scientist to pursue her passion for medicine and research.
The Shortage of Access to Therapists in Cambridge, Massachusetts in the United States: A Pandemic Study

Ashton T. Yoon
Chadwick School, 26800 S Academy Dr, Palos Verdes Peninsula, CA, 90274, USA; yoon.ashton.25@gmail.com
Mentors: Sophie Kim

ABSTRACT: The COVID-19 pandemic and lockdown have affected the health and safety of students of all ages across the globe, causing insecurity, confusion, emotional isolation, distress, and other psychiatric conditions. Specifically, young people between the ages of 5 to 19 are more vulnerable to mental health issues, and COVID-19 has become a source of many young people's anxiety. This study aims to investigate the availability of mental health professionals, especially during the post-pandemic time when telehealth is readily available, but there is a lack of therapists. To message therapists, we used the website psychologytoday.com. Our results showed that only 6 out of 636 therapists responded to emails from the Cambridge database. Of those 6, only 0.31% of the 636 therapists replied with an open slot for a new client. 0.63% accounted for those with no availability. In conclusion, our findings showed a massive mental health professional shortage. This study emphasizes the need to address the current therapist shortage issue in the United States.

KEYWORDS: Behavioral and Social Sciences, Psychology; Therapist; Accessibility; Shortage.

Introduction

After the COVID-19 lockdown, an increasing number of people have been searching for and needing mental health help via therapy. The COVID-19 pandemic affects the health and safety of students of all ages, causing insecurity, confusion, emotional isolation, distress, and other psychiatric conditions. Social distancing and school lockdowns forced children to experience school through the internet. This further escalated the mental health problems in children: young people between the ages of 5 to 19 are more vulnerable to mental health issues, and COVID-19 has become the source of the anxiety many young people face.¹ Studies show that access to healthcare services is vital for all stages of life, and accessibility of these services allows for more prompt diagnosis of health problems. This, as well as faster detection of mental health problems, allows for conditions to be treated more proactively.² However, despite this need for therapy spiking, there is still limited access to therapy; demand greatly outweighs the supply, with students and adolescents not having access to professional care.³,⁴

A sample of 50 thousand adults reported that 95.6% have at least one barrier preventing them from healthcare access, including mental health services. Extreme shortages of mental health prescribers (psychiatrists) or non-prescribers (i.e., other licensed mental health professionals like psychologists, social workers, counselors, therapists, etc.) have been found in 77% of U.S. counties, and 96% have had an unmet need for prescribers. One-third of Americans with mental health issues receive treatment from mental health professionals, meaning that two-thirds cannot access professional help.⁴,⁵

This study aims to analyze the accessibility of mental health professionals, especially during the post-pandemic time when telehealth is readily available, but there is a lack of therapists. We define in our study therapists as psychologists, social workers, and other therapists. We excluded other professionals who may be helpful but are unconventional providers and often need to be included in licensed therapy databases, like personal aides, energy workers, reflexologists, hypnotherapists, etc. We hypothesize that after the COVID-19 pandemic, people's access to therapists has been limited due to high demand and not enough therapists.

Methods

The zipcodes 02138 and 02139 were used from Cambridge to find therapists. To message therapists, the website psychologytoday.com was used.⁷ Using their email messaging system, all of the therapists contacted in Cambridge were sent the same message (below). The 02138 and 02139 zip codes covered the MIT school zones. This was to examine the availability of therapists for students in these universities. The availability/options of therapy are essential for university students, especially those from Harvard and MIT, so this was a critical area to check. All therapists in each zip code were sent the same message (below). The 02138 and 02139 zip codes covered the author's location, which includes the Harvard University and MIT school zones. This was to examine the availability of therapists for students in these universities. The availability/options of therapy are essential for university students, especially those from Harvard and MIT, so this was a critical area to check. All therapists in each zip code were sent the same message (below).

The therapists who were not taking new clients had their names still written down in the spreadsheet, but it was indicated that they needed to be messaged. Clinics and counseling centers were also messaged, and no types of therapists were excluded. The goal for all of the messages was to find the therapist's soonest appointment availability. In the case of vague answers, therapists were explicitly asked for the date and time of their earliest availability. To record the data, the date of re-
quest and the date of response from the therapist was recorded. To record the data, the date of request and response from the therapist was written in a google spreadsheet. If provided, the date of the earliest consultation was written in the spreadsheet. Those who said that they had a waitlist in their psychologytoday.com profile were still emailed for their waitlist. Only some responded, and those who did were asked about the waitlist times.

**Results and Discussion**

In the Cambridge Round 1 dataset (Table 1), 865 therapists were viewed on the PsychologyToday website, and 636 emails were sent to them (Table 1 and Table 3). Out of the 636 emails, only six replied to the email, and of those 6, 4 therapists replied with no availability, making up 0.63% of the total therapists. The other two therapists replied with availability, making up 0.31% of 636 therapists. Forty-one therapists indicated on psychologytoday.com that they were accepting waitlists, making up 4.74%. However, 0 therapists responded with any waitlist. Two-hundred twenty-nine therapists were already not taking new patients, so they were not emailed, and 630 therapists still need to respond to their email, taking up the extreme majority of 99.05% (Table 1).

In the Cambridge Round 2 dataset (Table 2), 826 therapists were viewed on the PsychologyToday website, and 603 emails were sent out to them (Table 2 and Table 4). Out of the 603 emails, only eight replied to the email, and of those 8, 4 therapists replied with no availability, making up 0.66% of the total therapists. The other four therapists replied with availability, making up 0.66% of 603 therapists. Twenty-eight therapists indicated on psychologytoday.com that they were accepting waitlists, making up 3.39%. However, only one therapist responded with any waitlist, making up 0.17% of therapists contacted. Two-hundred twenty-three therapists were already not taking new patients, so they were not emailed, and 595 therapists still need to respond to their email, taking up the extreme majority of 98.67% (Table 2).

**Table 1:** Data Collection Round 1: Therapists Contacted in February-March 2022 in Cambridge: 02138, 02139

<table>
<thead>
<tr>
<th>Types of Responses</th>
<th># of Messages</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes responses</td>
<td>2</td>
<td>0.31%</td>
</tr>
<tr>
<td>No responses</td>
<td>4</td>
<td>0.63%</td>
</tr>
<tr>
<td>Waitlist responses</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>Total responses</td>
<td>6</td>
<td>0.94%</td>
</tr>
<tr>
<td>Total messages Sent</td>
<td>630</td>
<td>99.05%</td>
</tr>
</tbody>
</table>

**Table 2:** Data Collection Round 2: Therapists Contacted in July 2022 in Cambridge: 02138, 02139

<table>
<thead>
<tr>
<th>Types of Responses</th>
<th># of Messages</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes responses</td>
<td>4</td>
<td>0.66%</td>
</tr>
<tr>
<td>No responses</td>
<td>4</td>
<td>0.68%</td>
</tr>
<tr>
<td>Waitlist responses</td>
<td>1</td>
<td>0.17%</td>
</tr>
<tr>
<td>Total responses</td>
<td>8</td>
<td>1.33%</td>
</tr>
<tr>
<td>Total messages Sent</td>
<td>603</td>
<td>98.67%</td>
</tr>
</tbody>
</table>

**Table 3:** Comparison of data (Feb-Mar 2022) vs. therapist profile info on PsychologyToday

<table>
<thead>
<tr>
<th>From Data Collection (Feb-Mar 2022)</th>
<th>From PsychologyToday Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses Received</td>
<td># of Messages</td>
</tr>
<tr>
<td>Yes responses</td>
<td>1</td>
</tr>
<tr>
<td>No responses</td>
<td>3</td>
</tr>
<tr>
<td>Waitlist responses</td>
<td>0</td>
</tr>
<tr>
<td>Total Messages Sent (Yes + Waitlist)</td>
<td>636</td>
</tr>
<tr>
<td>Total Messages Not Sent (labeled &quot;No, not taking patients&quot;)</td>
<td>229</td>
</tr>
<tr>
<td>Total Therapists listed in 02138 + 02139</td>
<td>665</td>
</tr>
</tbody>
</table>

**Discussion**

Our results show the lack of accessibility that people have when searching for therapists. With only six out of 636 therapists messaged responding at all to the email, the data shows an extreme lack of responses from therapists and the number of available therapists.

This is highly detrimental and can cause undue harm to those searching for help who may have undiagnosed mental disorders by making them feel more isolated, like they did something wrong in their search, despite the effort it took to search for therapists. This may cause those looking for therapists to give up on their search. Not even receiving a response from a therapist could lead to dangerous results, and it may lead some people to make decisions that would be detrimental to themselves or others.

People searching for therapists must contact many therapists to get a response, then decide whether or not that therapist can accommodate their needs. This increases the time between a client’s need for help and the client receiving support, forcing
people to spend more time with undiagnosed and untreated conditions. When accounting for children ages 5-19 needing mental health therapy, especially after and during the pandemic, these results demonstrate that children and adolescents must wait an incredibly long amount of time before receiving help, if at all.¹²

The first round of data was obtained in February and March of 2022 from Cambridge, an area with a large number of adolescents and college students, primarily due to the presence of Harvard University, MIT, Boston Latin High School, and Cambridge Rindge High School. Students experiencing stress caused by school face restricted access to mental health help due to the minimal number of therapists willing to respond to requests and may or may not be available after that. Despite the large number of therapists available to contact via PsychologyToday, 99.1% of the therapists in the 02138 and 02139 zip codes have yet to respond to requests for an initial consultation or meeting.

The second round of data in Cambridge was obtained in July 2022. This was to determine whether the presence of college and high school students, who comprise a large proportion of the Cambridge population, may be associated with the number of therapists available in the 2 Cambridge zip codes tested in February and March 2022: 02138 and 02139. High school and college students who may have been getting therapy during the first round of data collection, which occurred during the school year, were likely out of the city due to summer programs and internships that take them to cities across the country. Therefore, the hypothesis for the second round of data collection was that there would be more availability of therapists during the summer when high school and college students are not in town.

The same methodology was followed to determine therapist response rates and availability. The data collected in July 2022 showed therapist response rates to be 1.33% and availability rates to be 0.66%. This demonstrates that, unfortunately, response rates were not much higher during the summertime, invalidating the hypothesis that the lack of available therapists was due to large numbers of students taking up much of the therapists’ time during the school year. These results reveal the unfortunate fact that there was no marked difference in therapist response rate or availability based on reaching out to them via PsychologyToday, a significant resource for those seeking therapy.

The data also showed that the information on therapist profiles on PsychologyToday was not current. For example, in the first round of data collection in February-March 2022, the number of therapists that claimed to be taking patients on a waitlist was 41 (6.45%); in the second round of data collection in July 2022, the number of therapists that claimed to be taking patients on a waitlist was 28 (4.64%). In addition, in February-March 2022, the number of therapists who responded and confirmed they had a waitlist was 0 (0.00%); in July 2022, the number of therapists who responded and confirmed they had a waitlist was 1 (0.17%).

It is difficult to compare the exact differences between real-time therapist availability / up-to-date information compared to their PsychologyToday profiles because of the extreme lack of responses. We only messaged therapists who labeled their profiles “currently accepting new patients.” However, of the few responses received, 4 (0.66%; July 2022) and 1 (0.16%; February-March 2022) were not accepting patients, which contradicted their status at the time of data collection on their PsychologyToday therapist profile.

The limitations of this study were the method of communication, number of websites used, amount of zip codes and states, and the lack of data pre-covid-19 pandemic. Since the analysis was only performed on zip codes in Cambridge, it is challenging to apply our conclusion to the entire United States. In addition, our method of communication was limited to only email, allowing for gaps between responses and a lack of thorough interaction with therapists if they could only call.

A possible solution to the problem of the therapist shortage is eliminating geographical limitations and utilizing a new technology. The number of people looking for remote therapy has been increasing recently. In addition, other studies by numerous researchers have demonstrated that telehealth is effective and adaptable.¹⁵,¹⁶ Many online platforms, such as Betterhelp and Talkspace, two companies based in the US, are widely used by patients globally. Such online platforms deal with more than 2.9 million messages, video sessions, phone calls, and live chats. This shows how people are increasingly adapting to online therapy to avoid unwanted delays. In addition, online therapies can be more convenient, cover a wider range of therapies, and make quick availability and more affordable.

Numerous free healthcare applications, such as MindShift, PTSD Coach, BellyBio, Relax with Andrew Johnson Lite, and eMoods Classic, are available to treat anxiety, post-traumatic stress disorder (PTSD), depression, and bipolar disorder. Chatbots can also help fulfill the deficiency of therapists and other mental health professionals. The usage of artificial intelligence (AI), and machine learning (ML), a form of AI, is gaining more interest among therapists to treat patients.¹⁷ Lyssn, a health clinic, uses algorithms to examine conversations with patients and therapists to generate clinical quality metrics, which range from general counseling to fidelity or evidence-based counseling. AI technology enables therapists to access family history, client behavior, and their response to previous treatment. This information allows therapists to make more accurate diagnoses, leading to more insightful and effective treatment decisions and frees time to accept new patients.

After the end of the COVID-19 pandemic, the healthcare sector has been finding modern ways to fulfill the increasing demand for healthcare professionals for people to adopt quickly and safely. The government should be interested in aiding new policies and providing funds to mental health organizations to overcome the limitation of people’s affordability. Programs, such as Enhanced Primary Care (EPC), cover medical allowance for up to 5 therapy sessions with a vast range of allied health therapists. This program so far covers only chronic health issues but is a fundamental start. With government involvement and aid, programs like these can extend their range to cover more patients.
Conclusion
Our study concluded that the most dominant barrier to healthcare professionals’ access is the lack of accessibility that people have when searching for therapists. People searching for therapists must contact many therapists to get a response, then decide whether or not that therapist can accommodate their needs. This study emphasizes the need to address the current therapist shortage issue in the United States. More efforts and resources should be allocated to provide greater therapist access and thus aid the mental health of millions of patients.

References
12. Cunningham, P. J., Beyond Parity: Primary Care Physicians’ Perspectives On Access To Mental Health Care: More PCPs have trouble obtaining mental health services for their patients than have problems getting other specialty services (2009). Health Affairs, 28 (Suppl1), w490-w501.

Authors
Ashton Yoon is a sophomore at Chadwick School in Palos Verdes, California. He plans on pursuing a medical career.
Studies on Environmental Impact and Energy Efficiency of Artificial Intelligence and Machine Learning

Yavuz E. Damkaci
Jamesville-Dewitt High School, Fayetteville, NY 13066, USA.; yavuzdamkaci@gmail.com

ABSTRACT: The consideration of environmental impact and energy efficiency when developing artificial intelligence (AI) and machine learning (ML) models has received some traction recently as ML models get bigger. It has been shown to increase the energy efficiency of models without significantly affecting accuracy. We reviewed five recent studies that focus on investigating and refining the energy efficiency of ML systems through numerous models and methods with either software or hardware variations. The first article proposed using the total number of Floating-Point of Operations (FPO) to better measure efficiency without depending on hardware, location, and AI approaches used in the model. The second article evaluated and compared two deep learning frameworks, Pytorch and Tensorflow, and how each has its strengths and weaknesses regarding its efficiency in the training phase versus the inference phase. The third article analyzed energy-efficient architectures for ML, which includes utilizing more approximation to increase efficiency while having a minimal effect on accuracy significantly. The fourth article focused on holistic architecture solutions for creating efficient run times and energy consumption of neural networks. The last article developed a Java tool to help with increasing energy efficiency on edges with minimal decrease in accuracy. Lastly, we touched on foundation models and their energy efficiency considerations.

KEYWORDS: Artificial Intelligence; Machine Learning; Energy Efficiency; Edge.

Introduction

Artificial Intelligence (AI), especially Machine Learning (ML), has significantly increased in the last five years with the introduction of several common applications with widespread use by the general public. With the recent growth, research has been primarily focused on the capabilities and accuracy performance of ML while neglecting its green aspects, such as energy consumption and its efficiency, and increased environmental impact.¹ As an example, for the increased environmental impact of machine learning, an ordinary car produces about 126,000 pounds of carbon dioxide over the course of its lifespan, and a common GPU training neural network model produces about 625,000 pounds of carbon dioxide.²

The popularity of Leaderboards³,⁴ can be seen as evidence of the strong focus on capabilities and accuracy performance using AI while omitting any mention of cost or efficiency. Even though there are obvious advantages to increasing model accuracy, concentrating solely on this statistic ignores the effects on the economy, environment, and society.

Furthermore, an important class of machine learning applications is mainly defined as “edge computing” devices - in autonomous cars, at the mobile edge, etc., where overall power consumption and energy efficiency are of prime importance due to dependence on rechargeable battery technology. Therefore, system designs must support accelerated computation with improved energy efficiency and minimal compromise on computational accuracy or hardware costs. Currently, only 10% of the enterprise data is processed at the edge, and it is expected that it will rise to 75% by 2025.⁵

Internet of Things (IoT) devices are expected to generate over ten quintillions of data daily by 2025, with Connected and Autonomous Vehicles (CAVs) generating 8TB of data per hour.⁶ Edge computing is expected to handle these constraints by handling data locally,⁷ but energy-efficient software is crucial for avoiding hardware overheating.⁸ Achieving energy efficiency will increase the usage time of the battery. Edge devices like mobile phones and electric vehicles rely on batteries, and their battery life significantly impacts user experience. Improving energy efficiency by 20% can increase an electric car’s continuous driving distance from 500 to 600 kilometers. This enhances edge availability and extends the radius of life.⁹

The application of ML has improved several applications (e.g., in the medical, financial, and transportation sectors) that, for example, use speech and image recognition, machine translation, and natural language processing (NLP),¹⁰,¹¹ Clone detection,¹² de-obfuscation,¹³ language migration,¹⁴ code summarization,¹⁵ auto-correction,¹⁶ auto-completion,¹⁷ code generation,¹⁸ and program comprehension.¹⁹ these advancements would not have been conceivable without the significant ML advancement. ML models consist of two processing phases: training and inference. Training is a phase to learn the pattern using selected training data. Inference is the phase to put a predicted outcome using a trained ML model.

Discussion

This review is concentrated on five selected recent studies⁹,²⁰-²³ focusing on improving and analyzing the energy efficiency of machine learning systems through various methods and models with either hardware or software modifications.
Energy Efficiency and Environmental Impact: Green vs. Red AI

As recently reported by Schwartz et al.,20 DL research increased the computational costs of state-of-the-art AI research as big as 3000,000x between 2012 and 2018, as shown in Figure 1. This surge also increased Green Software Engineering research, which aims to decrease software environmental footprints and support Green AI.²⁴

In the literature,²⁰ two terms have been used to describe the current trends: Green AI seeks sustainable and environmentally friendly computing solutions by seeking to treat efficiency as a primary evaluation criterion alongside accuracy, while Red AI seeks to improve accuracy (or related measures) using massive computational power while disregarding the cost—essentially “buying” more robust results.

To achieve a shift from Red AI to Green AI, efficiency must be a more common evaluation criterion for AI products alongside accuracy and related measures. Despite the clear benefits of improving model accuracy, focusing on a single accuracy metric needs to pay attention to the economic, environmental, and social costs of reaching the reported results.

Schwartz et al.²⁰ propose making efficiency a more common evaluation criterion for AI research alongside accuracy and related measures to lead the practice in the field by proposing to reduce Red AI practices while increasing the use of Green AI practices.

Schwartz et al.²⁰ empirical analysis provides evidence that relatively little attention was paid to computational efficiency by the AI research community in comparison to accuracy. They show that Red AI is on the rise despite the well-known diminishing returns of increased cost. As shown in Figure 2, taken from Mahajan et al.,²⁵ accuracy was increased linearly and reached a plateau in some cases. In contrast, training examples increased exponentially, resulting in higher training costs.

To move away from Red AI to Green AI, Schwartz et al.²⁰ identify “key factors that contribute to Red AI and advocate the introduction of a simple, easy-to-compute efficiency metric that could help make some AI research greener, more inclusive, and perhaps more cognitively plausible.”

To demonstrate the prevalence of Red AI, Schwartz et al.²⁰ randomly sampled 60 papers from top AI conferences (ACL, NeurIPS, and CVPR). Their results are summarized in Figure 3, and as shown, a large majority of the papers target accuracy (90% of ACL papers, 80% of NeurIPS papers, and 75% of CVPR papers), while only a small portion (10% of AC, and 20% of CVPR) argue for a new efficiency result. Authors²⁰ state that the AI community focuses on performance measures, such as accuracy, at the expense of efficiency measures, such as speed or model size.

The proposed²⁰ equation illustrates three major quantities related to the cost of generating a result while ignoring data augmentation. It can be applied for calculating either training or inference costs. According to this equation, models using large parameters will incur high costs. As an example,²⁰ for such large models which have high costs for processing each example, which leads to large training costs, Google’s BERT-large contains roughly 350 million parameters, Grover contains 1.5 billion parameters, NVIDIA’s Megatron-LM contains over 8 billion parameters, Google’s T5-11B contains 11 billion parameters, and openAI’s openGPT-3.4 contains 175 billion parameters. In comparison, open GPT-2 contained only 1.5 billion parameters. This shows the current trend of making

Figure 1: The amount of computing used to train deep learning models between 2012 and 2018, showing the 300,000 times increase which is correlated to the cost and environmental impact.²⁰

Figure 2: Linear increase of object detection top-1 accuracy in comparison to the exponential growth of training sample size (Instagram).²⁰

Figure 3: Number of papers, based on 60 randomly selected, presented at top AI conferences targeting accuracy, efficiency, both, or other. ACL: Association for Computational Linguistics, CVPR: Computer Vision and Pattern Recognition, NeurIPS: Neural Information Processing Systems.²⁰

Authors²⁰ proposed to include the computational price tag of developing, training, and running models as a key Green AI practice. The proposed equation is Cost (R) a E · D · H where the cost of an AI (R)esult grows linearly with the cost of processing a single (E)xample, the size of the training (D)ataset and the number of (H)yperparameter experiments.

The proposed²⁰ equation illustrates three major quantities related to the cost of generating a result while ignoring data augmentation. It can be applied for calculating either training or inference costs. According to this equation, models using large parameters will incur high costs. As an example,²⁰ for such large models which have high costs for processing each example, which leads to large training costs, Google’s BERT-large contains roughly 350 million parameters, Grover contains 1.5 billion parameters, NVIDIA’s Megatron-LM contains over 8 billion parameters, Google’s T5-11B contains 11 billion parameters, and openAI’s openGPT-3.4 contains 175 billion parameters. In comparison, open GPT-2 contained only 1.5 billion parameters. This shows the current trend of making
models larger, which increases faster than the model performance.

Several measures have been reported to assess the efficiency of different models, such as carbon dioxide emissions, electricity usage, elapsed real time, and the number of parameters. According to Schwartz et al., each of these suffers from a deficiency; carbon dioxide emissions highly depend on local electrical structures and location dependent; electricity usage is hardware dependent and does not allow fair comparison between different models and machines, elapsed in real-time is highly hardware dependent, and also influenced by other programs running on the device.

Schwartz et al., suggest using the total number of Floating-Point of Operations (FPO) to better measure efficiency without depending on the model’s hardware, location, and AI approach. In addition, FPO directly computes the amount of work done and is tied to the amount of energy consumed. FPO is often correlated with the running time of the model and the amount of work done at each time step.

The authors also discussed the limitations of FPO. FPO does not capture the difference in work based on model implementation, as two different implementations of the same model could result in very different amounts of processing work. Also, FPO ignores the memory used by the model, which has additional energy and monetary costs.

The authors demonstrated the use of FPO compared to various models with different parameters and their top-1 accuracy levels. As shown in Figure 4a (top), the increase in FPO and the number of parameters of ResNet to SENet increases around 90-95%, and top-1 accuracy increases only 3.9%. Similarly, as shown in Figure 4b (bottom), when the number of layers in a model is increased from 50 to 152 (300% increase for different versions of ResNet for single object recognition), the number of parameters and FPO increased significantly. In comparison, top-1 accuracy increased only 2.4%.

The energy efficiency of DL frameworks:

Georgiou et al.’s paper on Green AI looked into two DL frameworks, in particular, PyTorch and TensorFlow, to see whether they have different costs. The authors performed an in-depth empirical analysis to investigate and compare the energy consumption and run-time performance of the frame-works by using six large models for DL from different AI domains, as listed in Table 1.

<table>
<thead>
<tr>
<th>Category</th>
<th>Model (Date Set)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommender Systems (RS)</td>
<td>NCF (v2021)</td>
<td>It filters and provides feedback based on the NCF specifications</td>
</tr>
<tr>
<td>Natural Language Processing (NLP)</td>
<td>GANMT (2019-2020)</td>
<td>It enables capturing longer-term dependency and resolves the concept fragmentation problem</td>
</tr>
<tr>
<td>Computer Vision (CV)</td>
<td>ResNet-50 (Coco, 2014)</td>
<td>It is a convolution-based neural network for the model of object instance segmentation</td>
</tr>
</tbody>
</table>

The authors studied the most energy and run-time efficient and the most accurate among the two tested DL frameworks. The results of their empirical study showed that there is a significant cost difference between the two DL frameworks in 94% of the cases tested. TensorFlow achieves better energy and run-time performance in the training phase for 100% of the cases, and PyTorch shows better energy and run-time performance in the inference phase for 66% of the cases.

The authors’ energy and run-time performance test results, both in the training and inference phases, were shown in Tables 2 and 3, respectively.

Authors found that the cheapest DL framework changes, either in the training or inference phase, depending on the model’s domain, such as recommender, vision, or NLP. The statistical tests (see Table 2) confirm that the results in the training phase are statistically different, with a large effect size in 24 out of 24 cases (100%), with TensorFlow significantly outperforming PyTorch in 16 out of 24 cases (76%). The statistical test (Table 3) in the inference phase favors PyTorch in 16 out of these 21 cases (76%). Overall, the authors concluded that the cheapest framework for the training phase is TensorFlow, while PyTorch is the least costly for the inference phase.

Table 2: Training Phase: Mean values for the energy consumption (in Joules) and run-time performance (in seconds) of training. The Wilcoxon statistical significance test results (p-value) and effect size (A₁₂) are also reported. PKG is the core and non-core components of the processor. RAM is the main memory, and GPU is the graphics card.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformer</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
</tr>
<tr>
<td>ResNet-152</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
</tr>
</tbody>
</table>

Table 3: Inference Phase: Mean values for the energy consumption (in Joules) and run-time performance (in seconds) of training. The Wilcoxon statistical significance test results (p-value) and effect size (A₁₂) are also reported. PKG is the core and non-core components of the processor. RAM is the main memory, and GPU is the graphics card.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformer</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
<td>14,779</td>
</tr>
<tr>
<td>ResNet-152</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
<td>216,391</td>
</tr>
</tbody>
</table>
The authors used the combined results of training and inference from Tables 2 and 3. They graphed it as shown in Figure 5 because obtaining the final accuracy involves both the training and inference steps. Authors concluded that better energy consumption and run-time performance—in most cases—yield better accuracy results as well. This outcome suggests that Green AI efforts should be increased in AI research and application for better environmental outcomes while obtaining a similar accuracy performance.

**Figure 5:** Comparison of energy consumption, run-time performance (Rt), and accuracy (Acc) between TensorFlow and PyTorch. The size of a circle represents the run-time performance (i.e., the bigger the process, the lower the run-time), and color represents the value of the accuracy metric (i.e., the darker the circle, the higher the accuracy); the interpretation depends on the metric of the model. Accuracy has been computed with Hit Rate (NFC), Perplexity (Transformer-XL), BLEU (GNMT), Top-5 error rate (ResNet-50), Avg. Precision (Mask R-CNN) and Precision (SSD).

The authors concluded that DL developers should choose the most appropriate framework for the model at hand while optimizing accuracy, run-time performance, and energy consumption. They also found that the training phase is more expensive than the inference one, thus resulting in a higher carbon footprint impact. Therefore, both researchers and developers should take appropriate steps to reduce its environmental impact.

Several research studies introduced practices on how to use traditional ML efficiently to reduce energy consumption. According to McIntosh et al., empirical studies showed that J48, SMO, and MLP are the algorithms using less energy in training ML models for Android devices, delivering more energy efficiency, better accuracy, and a correlation to algorithms’ complexity. Wang et al. achieved energy savings ranging from 60% to 90%, with an accuracy loss of 1.2% to 2% by dropping unnecessary computations from Convolutional Neural Network (CNN) models running on FPGAs. In addition, several other studies on energy estimation or measurements for ML studies exist.

**Energy efficient architectures for machine learning:**

Shafique and his coworkers study focuses on creating adaptive and energy-efficient machine learning architectures. They showed that building an energy-efficient architecture using approximate computing and multi-objective evolutionary algorithms may result in high accuracy performances (92-94%) while being energy-efficient systems. There are several existing ways of increasing energy efficiency in ML, such as using FPGAs, ASICs, and other specialized compute fabrics like Application-Specific Instruction Set Processors (ASIPs) and Coarse-Grained Reconfigurable Processors (CGRAs) for CNN acceleration. Shafique et al. study would like to further the energy consumption by “systematic development of energy-efficient accelerators for Machine Learning, specifically in the context of Deep Neural Networks (DNNs), where a high degree of adaptivity can be realized by employing quality-energy-configurable approximate modules inside the accelerator datapath.”

The authors developed a systematic methodology for realizing energy-efficient accelerators, specifically for CNN and DNN-based AI systems, as shown in Figure 6.

**Figure 6:** Shafique et al. developed the methodology for designing energy-efficient and adaptive neural network accelerator-based architectures for Machine Learning and AI systems.

In the literature, there are several examples of energy-efficient and high-performance hardware accelerators using approximate arithmetic modules, which sometimes result in loss of accuracy and, therefore, require selective approximate arithmetic configuration.

In a study where the use of low-power imprecise multipliers using the UCI Machine Learning repository is studied, it has been shown that the loss of accuracy could be overcome by retraining the network by an energy savings of up to 62.49% over accurate face recognition. In a similar study, constrained re-training of the neural network is done to get rid of undesired combinations due to approximation, which resulted in only an accuracy loss of 2.83% for 8-bit and ~0.25% for 12-bit implementations for handwritten digit recognition on the MNIST dataset, as shown in Figure 7.

Shafique et al. showed that a careful analysis of data distribution across various data-paths of neural networks can aid in selecting an approximate circuit that provides minimum loss to classification accuracy. As shown in Figure 8, the authors demonstrated that due to the fixed architectural design of DNN, each datapath has a particular data distribution. The highest classification accuracy was achieved when datapaths comprising FM5 and FM6 were approximated compared...
to the case when FM1 and FM3 were approximated, which showed the lowest accuracy. The authors\textsuperscript{22} showed that high accuracy can be obtained via approximation computing, which results in energy-efficient architectures.

**Figure 7:** Classification accuracy of Neural Networks (NN) with approximate multipliers for MNIST and SVHN datasets and 8/12 bit precision.\textsuperscript{22}

**Figure 8:** Classification accuracy of Neural Networks (NN) with approximate multipliers for MNIST and SVHN datasets and 8/12 bit precision.\textsuperscript{22}

**Energy efficiency of DNN implementations:**

DNNs have achieved high accuracy in AI tasks, but their computation produces high energy costs.\textsuperscript{33} According to Ganguly \textit{et al.},\textsuperscript{23} scaling hardware architectures for higher performance in deep learning can worsen energy costs, increasing dataset sizes and network layers. Neuromorphic workloads, which result from a paradigm shift, change processing from compute-centric to data-centric, requiring emerging memory technologies like 3D-stacked DRAMs,\textsuperscript{34} NVMs,\textsuperscript{35} and near-data processing (NDP).\textsuperscript{36} However, determining the best architecture for specific end-user scenarios and devices is challenging due to different tradeoffs.

Ganguly and coworkers\textsuperscript{23} state that the energy efficiency of deep neural networks (DNN) implementations is challenging to evaluate using standard models, benchmarks, frameworks, and tools. Using a large number of GPUs in parallel works, but power consumption is a drawback. Accelerators and ASICs can operate in low-power envelopes but are inflexible. FPGAs are flexible but large, hard to program, and not cost-effective. Runtimes and energy efficiency of CNNs, used in services like image recognition, are critical for users and cloud service providers. Quantifying, measuring, modeling, and predicting energy consumption is crucial for system designers to determine whether to keep DNN computations locally or offload them to a cloud or data center.

DL frameworks like Caffe,\textsuperscript{37} Torch,\textsuperscript{38} and Tensorflow\textsuperscript{39} provide tools for benchmarking applications’ performance. Still, there is no support for runtime power/energy measurements across all system components, including CPU, GPU, memory subsystem, fabrics/interconnects, FPGAs, and accelerator/ASIC. Ganguly \textit{et al.}\textsuperscript{23} ask the following important questions regarding energy efficiency from a system architecture point of view: How much energy is consumed for an inference made by a neural network, and can we predict the energy consumption of a neural network on a specific implementation for a given set of parameters/constraints? The authors\textsuperscript{23} suggest that based on the answers to these questions, a holistic approach is needed to evaluate the energy consumption of neural networks for specific usage scenarios and target devices.

Recently proposed Paleo’s model\textsuperscript{40} for predicting CNN runtime is limited to real-world performance. NeuralPower\textsuperscript{41} developed a predictive framework for the power, runtime, and energy of CNNs without running them on a target platform. The framework provides a breakdown of runtime and power across different components, identifying bottlenecks and analyzing energy precision ratios. NeuralPower models the power and runtime of key layers in CNNs and predicts network performance. However, according to Ganguly and his co-workers,\textsuperscript{23} NeuralPower may help evaluate energy-efficient choices for CNN implementations, allowing devices to weigh battery usage constraints and available resources on the cloud if extended beyond GPUs.

**Energy Efficiency on Edges:**

In their work, Kumar and coworkers\textsuperscript{9} studied the JEPO tool as an Eclipse IDE plugin that offers energy-efficient Java suggestions for developers, which can be used in Edge applications. It analyzes Java files and matches them to a pool of suggestions. JEPO1 is an Eclipse plugin that helps developers write energy-efficient machine learning or refactor existing code in real time. It analyzes code and checks for specific patterns, generating suggestions and determining energy-hungry methods in Java projects using the Javassist library.

Kumar \textit{et al.}\textsuperscript{9} created Table 4 regarding these patterns, which relate to several Java programming language elements, are displayed along with some suggestions.

**Table 4:** Java components and energy efficiency suggestions\textsuperscript{9}

<table>
<thead>
<tr>
<th>Java Components</th>
<th>Suggestion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primitive data types</td>
<td>int is the most energy-efficient primitive data type. Replace if possible.</td>
</tr>
<tr>
<td>Scientific notation</td>
<td>Scientific notation results in lower energy consumption of decimal numbers.</td>
</tr>
<tr>
<td>Wrapper classes</td>
<td>Integer Wrapper class object is the most energy efficient. Replace if possible.</td>
</tr>
<tr>
<td>Static keyword</td>
<td>static keyword consumes up to 17,700% more energy. Avoid if possible.</td>
</tr>
<tr>
<td>Arithmetic operators</td>
<td>Modulus arithmetic operator consumes up to 1,620% more energy than other arithmetic operators.</td>
</tr>
<tr>
<td>Ternary operator</td>
<td>Ternary operator consumes up to 37% more energy than if-then-else statement.</td>
</tr>
<tr>
<td>Short circuit operator</td>
<td>Put most common case first for lower energy consumption.</td>
</tr>
<tr>
<td>String concatenation operator</td>
<td>StringBuilder append method consumes much lower energy than String concatenation operator.</td>
</tr>
<tr>
<td>String comparison</td>
<td>String compareTo method consumes up to 33% more energy than the String equals method.</td>
</tr>
<tr>
<td>Arrays copy</td>
<td>System.array.copy() is the most energy-efficient way to copy Arrays.</td>
</tr>
<tr>
<td>Array traversal</td>
<td>Two-dimensional Array column traversal result in up to 793% more energy.</td>
</tr>
</tbody>
</table>
According to Kumar and co-workers,⁹ JEPO also helps measure energy consumption at method granularity, injecting energy and time measurements at the code level and end of each method. According to their initial evaluation showed a 14.46% improvement in package energy consumption, 14.19% in core energy consumption, and 12.93% in execution time, with only 0.48% drop in classifier accuracy. Based on these results, for data centers, supercomputers, and autonomous vehicles, where vast amounts of data are processed quickly, the energy consumption of software can be significantly reduced with the aid of JEPO. Table 5 is the culmination (adopted from Kumar et al.⁹) of hardware and software packages developed specifically for IoT and Edge devices to increase efficiency.

Table 5: Hardware and software packages developed specifically for IoT and Edge devices with energy efficiency in mind.

<table>
<thead>
<tr>
<th>General Type</th>
<th>Product</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware: Accelerator Architecture</td>
<td>Edge TPUs</td>
<td>embedded version of TPU for edge computing</td>
</tr>
<tr>
<td>Hardware: Full-Stack Optimization</td>
<td>MNMNet</td>
<td>modular and efficient library for deep learning to support CNN and long short-term memory networks (LSTM)</td>
</tr>
<tr>
<td>Software: Machine Learning Packages</td>
<td>QuinPack</td>
<td>optimized library, by Facebook.</td>
</tr>
<tr>
<td>Software: Operating System Packages</td>
<td>TensorFlow Lite</td>
<td>operates on quantized 8-bit tensors to achieve low energy and high-performance</td>
</tr>
<tr>
<td>Software: Operating System Packages</td>
<td>Paddle Lite</td>
<td>easy to perform inference on edge and IoT devices and is compatible with PaddlePack and other pre-trained models.</td>
</tr>
</tbody>
</table>

### Foundation models:

The term foundation models was first used in 2021 by researchers at Sandford University.⁴² Unlike current NLP models, which require extensive training on labeled data which requires extensive time and energy, foundational models require little to no fine-tuning. They are trained on a large amount of unlabeled data that will be employed for various downstream tasks.⁴³ Foundation models can serve as the basis for several applications of the AI model, as their name implies. The model can transmit knowledge it has learned about one circumstance to another by using self-supervised learning and transfer learning. Foundation models require a vast corpus of training data; however, different models can be built on it for various tasks and applications once developed. The energy and time efficiency provided by the foundational models comes from being a foundation for multiple tasks and application without further training. Since they do not rely on labeled data, it also has time savings at the training phase of the model. Foundation models are a newly defined but fast-growing field of AI, with nearly seventy-five models identified⁴⁴ and 200 foundation model startups have emerged, collectively raising $3.5 billion as of October 2022.⁴⁵ Foundation models were recently defined as “pervasively influencing society” since unprecedented adoptions of applications such as ChatGPT and Bert.⁴⁶

### Conclusion

Real-time processing requirements of machine learning and local battery constraints on Edge make energy efficiency more challenging. Even though both hardware and software are important to increase efficiency, software is a critical bottleneck for IoT and Edge devices, as it can compromise performance and energy efficiency. Extensive research focuses on optimizing software energy consumption through programming languages, programming languages, optimized architectures, and compiling options.

As stated in our introduction, energy efficiency and utilizing the full performance of hardware with reduced heating on the end user devices will extend the battery life, provide a better user experience, and, most importantly, reduce the environmental impact. Since the number of devices is increasing rapidly as their use is also expanding in various scenarios, any improved energy efficiency on the devices using machine learning will substantially impact the environment.

### Acknowledgments

The author would like to thank Drs. Bekir Turkkan and Tevfik Koser for their help and feedback in writing this review article. He also would like to thank his high school AP computer science teacher, Jay Lang, for getting him interested in computer science and software engineering and for his excellent guidance since ninth grade.

### References


8. B. Luo, S. Tan, Z. Yu, and W. Shi, “‘EdgeBox: Live edge video anal

DOI: 10.36838/v5i4.22


TalwalkarLab/paleo


**Authors**

Yavuz Emre Damkaci is a rising senior high school student at Jamesville-Dewitt High School in Upstate New York. He is interested in software engineering and interned at a software start-up to develop an algorithm for their product. He is also been passionate about the environmental, solving related issues, and aiding with its protection.
A Meta-analysis of Syndromic Autism Genes

Daria Lunina
European Gymnasium, Moscow, 107113, Russian Federation.; luninadasha@gmail.com
Mentor: Kif Liakath-Ali

ABSTRACT: Autism is a neurodevelopmental disorder characterized by severe impairment of various behavioral functions. Mutations in several genes are known to be associated with autism. In this study, a meta-analysis was performed on 130 genes implicated in syndromic autism by the SFARI database. Results show that several genes are associated with known human pathological conditions, such as delayed speech and language development. Further analysis reveals that many genes are associated with functions related to head development, Rett syndrome (psychoneurological syndrome), and neuron projection morphogenesis. Protein network analysis revealed closely associated phenotypic terms with head development and neuron projection morphogenesis functional groups, suggesting that mutations in these genes significantly affect neuronal formation and, in turn, result in autistic characteristics. This study sheds light on the general role of autism-related genes, their interactions, and how mutations in them may disrupt normal neuronal function.

KEYWORDS: Behavioural and Social Sciences; Neuroscience; ASD; Autism Spectrum Disorder; Genetics; Meta-analysis; Causes of ASD.

Introduction

Autism, or Autism Spectrum Disorder (ASD), is a neurodevelopmental disorder in the category of pervasive developmental disorders characterized by severe and pervasive impairment in reciprocal socialization, qualitative impairment in communication, and repetitive or unusual behavior.¹ The word “autism” was first used in 1911 by German psychiatrist Eugen Bleuler to describe schizophrenia symptoms, such as replacing unsatisfying reality with hallucinations and vivid visions.² However, in the 1960s, the meaning radically changed as British child psychologists started using it instead to reflect a lack of imaginative thinking.³ ASD is a relatively widespread condition, with the Centers for Disease Control stating that 1 in 54 children are affected by autism in the USA.⁴ Moreover, estimated rates of autism have increased in the period from the 1960s to the 1980s from 5 to 72 cases per 10 000 children in Europe and the USA.⁵,⁶ The underlying neurobiological pathology of ASD remains unclear. Research findings in this sphere suggest that children with confirmed autism have macrocephaly at 2–3 years old. Brain growth is rapid at 12 months, and the frontal, temporal lobes, and limbic structures enlarge during the first two years of life, correlating with symptom development.⁷ Additionally, fMRI studies have found that people with autism have abnormal hypoactivation in the fusiform face area, which can be connected with problems in human perception compared with objects.⁸ ASD symptoms usually appear around 18 months, and treatment is most effective when applied as early as possible.⁹ Neurochemical studies have identified genetic differences in serotonin transport which are supported by empirical data on ASD-related symptoms.

One of the main problems with identifying the cause of this disorder is that it is unknown whether environmental factors or the genetic heritage is most influential (although the genetic cause is considered most probable), but also the sheer number of genes associated with Autism. As estimated by SFARI (Simons Foundation Autism Research Initiative), a comprehensive database, there are 1003 genes implicated in ASD.¹⁰ Data from this resource shows that 130 of these genes are considered syndromic, meaning they were found to be associated with ASD symptoms and linked to additional characteristics not required for an ASD diagnosis. Analysis of these syndromic genes will further our understanding of the genetic basis of ASD. The study presented here aims to conduct a meta-analysis of these 130 syndromic genes using tools provided by open resources, including Metascape¹¹ (a gene annotation and analysis resource), and reveal their ontology and biological role in brain development. Results from this study will contribute towards a better understanding of this disorder and will provide a computational analysis of syndromic genes, which will be useful in ASD-related research.

Methods

To perform a meta-analysis of syndromic genes implicated in ASD, a curated list of genes was downloaded from the SFARI Gene database, which classifies genes based on four scores: S (Syndromic), 1 (High Confidence), 2 (Strong Candidate), and 3 (Suggestive Evidence). The syndromic category includes genes with mutations reported to be associated with a substantial degree of increased risk and consistently linked to additional characteristics not required for an ASD diagnosis. It was reasoned that performing a meta-analysis on the Syndromic category of genes would be essential to understand how mutations in these genes may manifest ASD. Therefore, a total of 130 genes in the “Syndromic” category were entered into the Metascape under express and custom analysis (Supplementary Table 1). The genes were analyzed for gene summary, development of disorders, disease and gene association, gene placement...
on the chromosome, molecular function, RNA tissue category, and gene expression in different tissues.

The characteristics used for the custom analysis are reflected in Supplementary Table 2. Metascape has provided CAME analysis (Id conversion - convert input gene identifiers into Entrez gene IDs of a target species, annotation - extract annotation columns for the gene list, including gene descriptions, functions, and protein classes, membership - flag genes which fall under GO biological process terms that contain "invasion" as a keyword, function enrichment analysis of the gene list - identifying pathways (or complexes, published hit lists, etc.) that have statistically significant p-values and protein network analysis) on the bases of Homo sapiens genes. This information was further reflected in graphs and tables, taken from express analysis, which was discussed in this research’s “Results” section.

Results

The gene summary provided by Metascape reflected that all genes are protein encoding. Additionally, it was found that all genes are expressed, at least in some tissues. In particular, about 12.3 percent of genes are expressed in brain tissue (16 out of 130 genes) (Supplementary Table 2). Analysis of syndromic genes for known human phenotypes showed pathological conditions associated with ASD (Figure 1). The most frequent phenotypes associated with these genes are delayed speech and language development, autistic behavior, downward slant of the palpebral fissure, strabismus, and neurodevelopmental disorder. Of them, two (downward slant of palpebral fissure-behavior and strabismus) are the causes of eye development pathologies. As stated by DisGeNET, a database of diseases associated with genes concerning NCBI, autistic behavior is defined by: “Persistent deficits in social interaction and communication and interaction as well as a markedly restricted repertoire of activity and interest as well as repetitive patterns of behavior” while a neurodevelopmental disorder is described as: “behavioral and cognitive disorder with onset during the developmental period that involves impaired or aberrant development of intellectual, motor, or social functions.” Thus, it is evident that mutations in these genes are associated both with behavioral abnormalities and eye development pathologies. The lowest p-value negative logarithm is associated with a depressed nasal bridge and is -23.

The syndromic gene list was analyzed for enriched terms based on their functional roles. The highest p-values of the functions linked with the syndromic genes are found within the groups the head development, Rett syndrome (psychoneurological syndrome), neuron projection morphogenesis, and covalent chromatin modification (Figure 2). The most significantly enriched genes were found to be associated with head development (22 genes), neuron projection morphogenesis (19 genes), and covalent chromatid modification (16 genes), as indicated in the Metascape express analysis shown in Table 1.

![Figure 1: DisGeNET summary. Bar graph shows -\log_{10}(P) values of syndromic SFARI genes implicated in known human pathological conditions. Note that top gene enrichment is with delayed speech and language development, which is a clinical phenotype in ASD.](image1.png)

![Figure 2: Bar graph of enriched terms across input gene lists, colored by p-values. The bar graph shows that most syndromic genes are associated with head development, Rett syndrome (psychoneurological syndrome), neuron projection morphogenesis, and covalent chromatid modification.](image2.png)

Next, the syndromic genes were analyzed for the network of enriched terms, the relationship between genes, and the functions of the proteins they code for. Gene function could be defined as the encoding of particular molecules and proteins. In contrast, protein function depends on the processes in which they are involved, such as the structural integrity of cells or control over cell division. The nodes represent the similarity within a gene’s function, and links between nodes represent similarities between functions of proteins, which are encoded by ASD-related genes (Figure 3). One of the first things that can be seen is the abundance of links between nodes of the same color (cluster ID).

An example is the gene nodes, which influence sensory organ morphogenesis (brown cluster ID). They are strongly linked with each other and, due to the small number of links with genes of different cluster IDs, are separated from the main body of the diagram. In other words, genes of a similar coding function tend to code for proteins with a similar function.

Slightly above the center of the diagram, the group of small nodes that mostly have different gene coding functions, judging from the cluster IDs, appear to have many links between each other. For example, the chromatin remodeling node (magenta cluster ID) is surrounded by many links, which form a uniform circle around it. That means that small groups of genes, which code for proteins of different groups, make up the proteins with similar functions. It is also evident that these links appear to be thick. For example, DNA repair (turquoise cluster ID), which has few associated genes (small node), has many thick links between other nodes (different cluster ID). The thickness of the link represents that the function of one group of genes is strongly related to the function of the other group. This could mean that there are small groups of genes with different coding functions, which are firmly related to one another in two ways: similarity in the functions of proteins, which they code for, and the procedure of executing the gene’s

DOI: 10.36838/v5i4.23
coding function. It is also evident that cluster IDs of the smaller nodes represent processes that are not directly connected to ASD symptoms, such as chromatin remodeling mentioned before.

This data suggests two possible explanations for the genetic causes of ASD. The first option is that the interactions between small separate groups of genes, which code for proteins not directly linked to the ASD symptoms, result in the biological processes, which lead to the ASD-related phenotype, such as enlargements of the frontal, temporal lobes, and limbic structures. The second is that interactions between closely related genes of the same function, for example, head development genes, could result in the development of ASD. However, because even closely related groups of nodes have some links with other nodes, it could be stated that a more complex system of interactions between all the clusters results in the formation of ASD.

As shown in Figure 3b, which depicts the network of enriched terms from this analysis, the larger the cluster of genes with closely related functions, the lower the associated p-value. This relationship reflects a higher probability of the stated genes expressing associated proteins. Head development and neuron projection morphogenesis functional groups have the lowest p-values compared to other cluster IDs. Such a conclusion supports the idea that ASD causation might lie in the interactions between closely related genes of the same cluster-ID (function), as they are more likely to code for the associated proteins. In particular, those genes might be in head development and neuron projection functional morphogenesis groups.

For each given gene list, protein-protein interaction enrichment analysis has been carried out with the following databases: STRING,¹⁴ BioGrid,¹⁵ OmniPath, and InWeb_IM.¹⁷ The resultant network contains the subset of proteins that form physical interactions with at least one other member in the list. The Molecular Complex Detection (MCODE) algorithm has been applied to identify densely connected network components. The protein-protein interaction network (Figure 4) reflects a close relationship between four groups of genes encoding for similar proteins: Cohesin-SA1 complex, membrane transport-associated proteins, chromatin-associated genes, and modification-related genes. The fact that the proteins expressed by some ASD syndromic genes are closely related in their functional communication suggests possible mechanisms of ASD causation. However, it should also be stated that proteins depicted in the physical protein-protein interaction network, which do not hold high p-values for the assigned functions, are not included in large groups of nodes, or are interconnected with many nodes of different cluster IDs in the Network of enriched terms. This condition differentiates illustrated findings from the Figure 3a data. Some of the connections within the group of proteins with the same functions depicted in Figure 3 have been deleted. Apart from that, it is evident that many genes stated in Table 4, such as the histone modification gene, are not represented in Figure 3a legend. The possible explanation for these differences lies in the emphasis on the physical interactions between proteins and the fact that only genes with higher p-values were chosen to be depicted, as pointed out in the description of Figure 4. For instance, this condition limits the list of genes that could be included in the interactions between their proteins. Due to that fact, genes with more specific functions had to be chosen, while in Figure 3a, they might have belonged to more general function cluster IDs.

![Figure 3: a) Network of enriched terms colored by cluster ID, where nodes that share the same cluster-ID are typically close to each other. The links between the nodes reflect the similarities between the function of the proteins expressed by the genes. The thickness of links represents the genes closely related to their function within the expressed phenotype, reflected in a color; (b) Network of enriched terms (nodes): colored by p-value, where terms containing more genes tend to have a more significant p-value.](image-url)
The analysis results reflect that most syndromic genes analyzed from the SFARI database, indeed, most frequently, are associated with ASD-related symptoms. The data in Figure 1 showing the phenotypes associated with p-values reflects a similar idea, as, for example, head development may be linked with ASD symptoms. It is characterized by enlargements of the frontal, temporal lobes, and limbic structures, which occur at the age of 2, simultaneously developing other symptoms. Also, there are two trends found within the network of enriched terms. ASD is formed by genes’ interactions in many small functional groups not directly linked with ASD symptoms. In a few big functional groups, directly connected with ASD symptoms, such as head development. The latter option also corresponds with smaller p-values, indicating that these genes are more likely to be associated with their function. Protein–protein interactions reflect the enrichment term analysis graph trend, with a few differences. For instance, several links between the gene’s functional groups are deleted, compared to Figure 3. Although possible reasons for this and other differences were stated in the “Results” section, there is a lack of certainty.

Further investigations into detailed interactions between proteins of genes may provide more solid explanations for these findings. The genes represented in Figure 4 and Table 1 are responsible for the phenotype, which is not directly linked with ASD symptoms. It should also be stated that the p-values of genes depicted in Table 1 broadly vary even in one indicated group, even though the genes were chosen about their p-values. For instance, in group MCODE_1 covalent chromatin modification gene (GO:0016569) has a Log10(P) of -8.1, while the histone modification gene (GO:0016570) has a Log10(P) of -6.6. It means that the probability that particular gene codes for the stated protein, which in turn interacts with other proteins in the group, is uneven. This identifies the possibility that a different gene, not included in the analysis, may take part in the interaction. The idea also compromises the validity of the gene-to-gene communications reflected both in Figure 3a and Figure 4, showing that the relationship between proteins may not correspond to the assigned genes.

The findings stated above reflect links between genes and ASD-related phenotypes. The links could be explored in further research to understand better genes that may cause ASD. Additionally, relationships between groups of genes and links between groups may shed light on ASD formation, which could result from interactions between these genes rather than an expression of one particular gene.

One study examining the causative role of SFARI genes in ASD reported confounding findings about the database.¹⁸ It combined information about gene expression and clinical data across 80 samples and built a gene co-expression network analyzed at different granularity. The study found no significant link between SFARI’s gene expression and differential expression patterns comparing the group with diagnosed ASD and the control group. Also, no strong correlations between the expression of SFARI genes and ASD diagnosis were found through the gene co-expression network.

### Table 1: Color codes and descriptions for Figure 4. Three best-scoring terms by p-value as the functional description of the corresponding components, according to network plots within Figure 3.

<table>
<thead>
<tr>
<th>Color</th>
<th>MCODE</th>
<th>GO</th>
<th>Description</th>
<th>Log10(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>MCODE_1</td>
<td>GO:0016569</td>
<td>covalent chromatin modification</td>
<td>-8.1</td>
</tr>
<tr>
<td>Red</td>
<td>MCODE_1</td>
<td>GO:0016570</td>
<td>histone modification</td>
<td>-6.6</td>
</tr>
<tr>
<td>Red</td>
<td>MCODE_1</td>
<td>PID: P38 Alpha Beta Pathway</td>
<td>-6.1</td>
<td></td>
</tr>
<tr>
<td>Blue</td>
<td>MCODE_2</td>
<td>CORUM:63</td>
<td>Cohesin-SA1 complex</td>
<td>-15.2</td>
</tr>
<tr>
<td>Blue</td>
<td>MCODE_2</td>
<td>CORUM:164</td>
<td>Cohesin-SA1 complex</td>
<td>-15.2</td>
</tr>
<tr>
<td>Blue</td>
<td>MCODE_2</td>
<td>CORUM:165</td>
<td>Cohesin-SA1 complex</td>
<td>-15.2</td>
</tr>
<tr>
<td>Green</td>
<td>MCODE_3</td>
<td>R-HSA:4839726</td>
<td>Chromatin organization</td>
<td>-6.0</td>
</tr>
<tr>
<td>Green</td>
<td>MCODE_3</td>
<td>R-HSA:3247509</td>
<td>Chromatin modifying enzymes</td>
<td>-6.0</td>
</tr>
<tr>
<td>Green</td>
<td>MCODE_3</td>
<td>GO:0003009</td>
<td>morphogenesis of an epithelium</td>
<td>-4.2</td>
</tr>
<tr>
<td>Purple</td>
<td>MCODE_4</td>
<td>R-HSA:8856828</td>
<td>Clathrin-mediated endocytosis</td>
<td>-6.9</td>
</tr>
<tr>
<td>Purple</td>
<td>MCODE_4</td>
<td>R-HSA:1999971</td>
<td>Membrane Trafficking</td>
<td>-4.9</td>
</tr>
<tr>
<td>Purple</td>
<td>MCODE_4</td>
<td>R-HSA:5653656</td>
<td>Vesicle-mediated transport</td>
<td>-4.9</td>
</tr>
</tbody>
</table>
These findings suggest that genes analyzed in this study might not fully represent the ASD symptoms and causes.

However, the correlation between levels of expression and SFARI’s gene scores was statistically significant. It is valuable for the justification of the genes chosen for the study. SFARI genes were also connected with other neurodevelopmental disorders, such as Schizophrenia, Bipolar Disorder, Depressive Disorder, and others.

The study also suggested a new approach to using such databases against biases. Analysis of co-expression networks, including data from the whole network, can provide more meaningful information about genes linked to ASD diagnosis and even suggest new candidates. Incorporating this approach in further studies would be paramount for deeper investigation of ASD-related genes.

Another study investigated the variations of genes linked to ASD. To be more precise, they examined rare CNVs (copy-number variation) and proteins disrupting SNVs (single-nucleotide variants). The study used Infinium PsychArray as a tool for the analysis. It was characterized as an array of SNP (single-nucleotide polymorphism). Using it, researchers performed an integrated analysis of CNVs and SNVs collection of the homogeneous cohort of 127 Italian ASD families genotypes. Researchers found a higher number of rare CNVs, most frequently deletions, in ASD individuals compared to the unaffected group. They also observed a multitude of rare CNVs in the SFARI gene collection used in the study. It was indicated that ASD affected group had an increased transmission of rare SNV’s variants from heterozygous parents to probands. These findings support the idea of multi-genetic nature of ASD, which also corresponds with the conclusions of this paper.

Researchers also stated the significant role of the following genes in ASD development: VPS13B, WWOX, CNTNAP2, RBFOX1, MACROD2, APBA2, PARK2, GPHN, and RNF113A. From this list, only one gene (CNTNAP2) was included in this paper’s analysis. Hence, further investigation of the abovementioned genes would also provide more insight into ASD gene research.

As this research examined only human genes, it would be insightful to investigate genes of other species, such as mices (Mus musculus) or zebrafish (Danio rerio). Analyzing ASD genes in other organisms may provide a new understanding of their interactions and a fuller comprehension of ASD causation. Furthermore, as only syndromic genes from the SFARI database were analyzed, further research may examine other score sections from the same database or genes from other databases. Such examination would enrich the conclusions derived from the analysis described above and provide a better understanding of ASD.

It should also be stated that some of the genes used in this research have been marked in Allen Brain Atlas (actb abi1 vamp2 syt1 syn1 rp6ka3 prkdh phf8 pax6 ntrk2 ntng2 ntng1 nr2f1 mtor hcn1 gabra3 egf1a2 cux2 csp290 slc1a2 ctnnap2 cnk2a). Analysis of the ASD-related genes in this database and others would provide an in-depth comprehension of ASD causes.

**Acknowledgments**

I want to thank Dr. Kif Liakath-Ali for his supervision and mentorship in this project. Without his help and support, this research would not be possible. I would also like to thank the reviewers for their helpful suggestions.

**References**


**Author**

Daria Lunina, the author of this research, is a senior in the IB program in the European gymnasium. She is interested in neuroscience and wants to investigate behavioral disorders further in her career. She also wants to major in neuroscience for her bachelor's degree.

### Supplementary materials

#### Table 1: List of SFARI human syndromic genes used for the following research.

<table>
<thead>
<tr>
<th>stat us</th>
<th>gene-sy mbol</th>
<th>gene-name</th>
<th>ens embel-id</th>
<th>chromos ome</th>
<th>genetic - categor y</th>
<th>gene e - type</th>
<th>syndro mic</th>
<th>num ber-of-report s</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>ACTB</td>
<td>actin beta</td>
<td>ENSG000001 75624</td>
<td>7</td>
<td>Rare Single Gene Mutation</td>
<td>Synodic</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>ACTB9B</td>
<td>actin like 6B</td>
<td>ENSG000001 77809</td>
<td>7</td>
<td>Rare Single Gene Mutation</td>
<td>Synodic</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>ACY1</td>
<td>aminoacylase 1</td>
<td>ENSG000002 43989</td>
<td>3</td>
<td>Rare Single Gene Mutation</td>
<td>Synodic</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>Ah1</td>
<td>Abelson helper integration site 1</td>
<td>ENSG000001 35541</td>
<td>6</td>
<td>Rare Single Gene Mutation</td>
<td>Synodic, Genetic</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>9</td>
<td>ALDH1A3</td>
<td>aldehyde dehydrogenase family member A3</td>
<td>ENSG000001 94204</td>
<td>15</td>
<td>Rare Single Gene Mutation</td>
<td>Synodic</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>ANKS1B</td>
<td>ankyrin repeat and sterile alpha motif domain containing 1S</td>
<td>ENSG000001 85846</td>
<td>12</td>
<td>Rare Single Gene Mutation</td>
<td>Synodic</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>AP1S2</td>
<td>adaptor related protein complex 1 sigma 2 subunit</td>
<td>ENSG000001 82827</td>
<td>X</td>
<td>Rare Single Gene Mutation</td>
<td>Synodic</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>ARID2</td>
<td>AT-rich interaction domain 2</td>
<td>ENSG000001 88079</td>
<td>12</td>
<td>Rare Single Gene Mutation</td>
<td>Synodic</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>ATP1A1</td>
<td>ATPase transporting alpha 1</td>
<td>ENSG000001 63399</td>
<td>1</td>
<td>Rare Single Gene Mutation</td>
<td>Synodic</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>ATP1A3</td>
<td>ATPase transporting alpha 3</td>
<td>ENSG000001 56409</td>
<td>10</td>
<td>Rare Single Gene Mutation</td>
<td>Synodic, Function al</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Gene/Description</td>
<td>Chromosome</td>
<td>Expression Level</td>
<td>Molecular Function</td>
<td>References</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCO2L1, BCL6 corepressor like 1</td>
<td>ENSG00000080015</td>
<td>X</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG6, ALG6, alpha-1,3-glucosyltransferase</td>
<td>ENSG00000080030</td>
<td>1</td>
<td>Syndromic</td>
<td>1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRWD3, brevomedullin and WD repeat domain containing 3</td>
<td>ENSG00000080028</td>
<td>X</td>
<td>Syndromic</td>
<td>1 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C12orf57, chromosome 12 open reading frame 57</td>
<td>ENSG0000008011760</td>
<td>12</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAGNAI1A, Calcium channel voltage-dependent, PIG type, alpha 1A subunit</td>
<td>ENSG0000008041837</td>
<td>19</td>
<td>Rare Single Gene Mutation, Genetic Association</td>
<td>1 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAMK2A, calcium/calmodulin dependent protein kinase II alpha</td>
<td>ENSG0000008006850</td>
<td>5</td>
<td>Rare Single Gene Mutation, Syndrome, Genetic Association, Function</td>
<td>3 1 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAMK2B, calcium/calmodulin dependent protein kinase II beta</td>
<td>ENSG00000080480414</td>
<td>7</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCNK, cyclin K</td>
<td>ENSG0000008006851</td>
<td>14</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDK13, cyclin dependent kinase 13</td>
<td>ENSG00000080065883</td>
<td>7</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDK8, cyclin dependent kinase 9</td>
<td>ENSG00000080029684</td>
<td>13</td>
<td>Syndromic</td>
<td>1 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEP70A, Centrosomal protein 200kDa</td>
<td>ENSG0000008008707</td>
<td>12</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>3 1 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDK9, cyclin dependent kinase 19</td>
<td>ENSG00000080051111</td>
<td>6</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>3 1 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHD1, chromodomain helicase DNA binding protein 1</td>
<td>ENSG00000080031922</td>
<td>5</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>3 1 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHK2, Chk1-like kinase</td>
<td>ENSG00000080020822</td>
<td>22</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLCN4, chloride voltage-gated channel 4</td>
<td>ENSG00000080073464</td>
<td>X</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>3 1 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNKSR2, connexin enhancer of kinase suppressor of Ras 2</td>
<td>ENSG00000080069970</td>
<td>X</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>2 1 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNTAI2, contactin associated protein-like 2</td>
<td>ENSG00000080074469</td>
<td>7</td>
<td>Rare Single Gene Mutation, Syndrome, Genetic Association</td>
<td>2 1 69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTNNNA2, catenin alpha 2</td>
<td>ENSG00000080064632</td>
<td>2</td>
<td>Syndromic</td>
<td>1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUX2, cut-like homeobox 2</td>
<td>ENSG000000800171248</td>
<td>12</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>3 1 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYFPP2A1, cytoskeleton PALS family 27 subfamily A member 1</td>
<td>ENSG00000080068920</td>
<td>2</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEPDC5, DEP domain-containing 5</td>
<td>ENSG000000800610310</td>
<td>22</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DHX30, DEAH-box helicase 30</td>
<td>ENSG000000800310153</td>
<td>3</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLL1, delta-like canonical Notch ligand 1</td>
<td>ENSG00000080048719</td>
<td>6</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMD, dystrophin (muscular dystrophy, Duchenne and Becker types)</td>
<td>ENSG00000080068947</td>
<td>X</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOLK, dolichol kinase</td>
<td>ENSG00000080072035</td>
<td>9</td>
<td>Syndromic</td>
<td>1 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EEF1A2, Eukaryotic translation elongation factor alpha 2</td>
<td>ENSG000000800612010</td>
<td>20</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBXO11, F-box protein 11</td>
<td>ENSG00000080038081</td>
<td>2</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>3 1 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRMIPD4, FERM and PDZ domain containing 4</td>
<td>ENSG00000080069933</td>
<td>X</td>
<td>Rare Single Gene Mutation, Syndrome</td>
<td>1 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OAEBR2, gamma-amino- butyric acid type B receptor subunit 2</td>
<td>ENSG00000080056928</td>
<td>9</td>
<td>Rare Single Gene Mutation, Syndrome, Function</td>
<td>3 1 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene ID</td>
<td>Name</td>
<td>Description</td>
<td>Ensembl ID</td>
<td>Reference</td>
<td>Mutation Type</td>
<td>Phenotype</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>-------------</td>
<td>------------</td>
<td>-----------</td>
<td>--------------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GABRA3</td>
<td>Gamma-aminobutyric acid (GABA) A receptor, alpha 3</td>
<td>ENSG000000011677</td>
<td>Rare Single Gene Mutation</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GATM</td>
<td>Glycine amidotransferase (L-arginine-glycine amidotransferase)</td>
<td>ENSG0000000171765</td>
<td>Synthetase</td>
<td>15</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBRSL1</td>
<td>fibroblast-like</td>
<td>ENSG0000000127767</td>
<td>Synthetase</td>
<td>12</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGFR3</td>
<td>fibroblast growth factor receptor 3</td>
<td>ENSG000000029652</td>
<td>Synthetase</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GALNT2</td>
<td>polypeptide N-acetylglucosaminyl transferase 3</td>
<td>ENSG000000043641</td>
<td>Synthetase</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCFC1</td>
<td>host cell factor C1</td>
<td>ENSG000000072634</td>
<td>Rare Single Gene Mutation</td>
<td>1</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCN1</td>
<td>hyperpolarization activated cyclic nucleotide-gated potassium channel 1</td>
<td>ENSG000000064088</td>
<td>Rare Single Gene Mutation, Genetic Association</td>
<td>5</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDAC4</td>
<td>histone deacetylase 4</td>
<td>ENSG000000066024</td>
<td>Rare Single Gene Mutation, Genetic Association</td>
<td>2</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDAC8</td>
<td>histone deacetylase 8</td>
<td>ENSG000000047099</td>
<td>Rare Single Gene Mutation</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEPACAM</td>
<td>hepatic and glial cell adhesion molecule</td>
<td>ENSG000000066478</td>
<td>Rare Single Gene Mutation</td>
<td>11</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HERC2</td>
<td>HECT and RLD domain containing E3 ubiquitin protein ligase 2</td>
<td>ENSG000000028731</td>
<td>Rare Single Gene Mutation</td>
<td>15</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOXA1</td>
<td>homeobox A1</td>
<td>ENSG000000050991</td>
<td>Rare Single Gene Mutation</td>
<td>7</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HWE1</td>
<td>HECT, UBA and WWP domain containing 1, E3 ubiquitin protein ligase</td>
<td>ENSG000000060758</td>
<td>Rare Single Gene Mutation</td>
<td>1</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTS1</td>
<td>integrator complex subunit 1</td>
<td>ENSG000000064880</td>
<td>Rare Single Gene Mutation</td>
<td>7</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KAT6A</td>
<td>K(lysine) acetyltransferase 6A</td>
<td>ENSG000000083168</td>
<td>Rare Single Gene Mutation</td>
<td>8</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIF5C</td>
<td>kinesin family member 5C</td>
<td>ENSG000000066780</td>
<td>Rare Single Gene Mutation</td>
<td>2</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KPTN</td>
<td>kappin, actin binding protein</td>
<td>ENSG000000018182</td>
<td>Rare Single Gene Mutation</td>
<td>19</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUNPK</td>
<td>lunapark, ER junction formation factor mediator complex subunit 12L</td>
<td>ENSG000000044693</td>
<td>Rare Single Gene Mutation, Genetic Association</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBL3</td>
<td>MBL complex subunit 3</td>
<td>ENSG000000005302</td>
<td>Rare Single Gene Mutation</td>
<td>2</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTOC</td>
<td>mechanistic target of rapamycin (serine/threonine kinase)</td>
<td>ENSG000000098793</td>
<td>Rare Single Gene Mutation</td>
<td>1</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF1B</td>
<td>nuclear factor I B</td>
<td>ENSG000000047582</td>
<td>Rare Single Gene Mutation</td>
<td>9</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFIX</td>
<td>nuclear factor IX (CCAT-binding transcription factor)</td>
<td>ENSG000000048841</td>
<td>Rare Single Gene Mutation</td>
<td>10</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOVA2</td>
<td>NOVA alternative splicing regulator 2</td>
<td>ENSG000000047827</td>
<td>Rare Single Gene Mutation</td>
<td>5</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NR2F1</td>
<td>nuclear receptor subfamily 2, group F member 1</td>
<td>ENSG000000075745</td>
<td>Rare Single Gene Mutation</td>
<td>19</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTNG1</td>
<td>netrin G1</td>
<td>ENSG000000062631</td>
<td>Rare Single Gene Mutation</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTNG2</td>
<td>netrin G2</td>
<td>ENSG000000096358</td>
<td>Rare Single Gene Mutation, Genetic Association</td>
<td>9</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTRK2</td>
<td>neurotrophic receptor tyrosine kinase 2</td>
<td>ENSG000000048653</td>
<td>Rare Single Gene Mutation</td>
<td>9</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OCRL</td>
<td>osteoclast/renal tyrosine kinase of the low</td>
<td>ENSG000000022126</td>
<td>Rare Single Gene Mutation</td>
<td>14</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACS2</td>
<td>phospholipid scramblase 2</td>
<td>ENSG000000079864</td>
<td>Rare Single Gene Mutation</td>
<td>14</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCCG</td>
<td>propargyl-CoA carboxylase alpha subunit</td>
<td>ENSG00000175198</td>
<td>13</td>
<td>Rare Single</td>
<td>Syndrome, Genetic Association</td>
<td>1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>PCCB</td>
<td>propargyl-CoA carboxylase beta subunit</td>
<td>ENSG00000140904</td>
<td>3</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>1</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>NSD2</td>
<td>nuclear receptor binding SET domain protein 2</td>
<td>ENSG00000109685</td>
<td>4</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>PHF8</td>
<td>PHD finger protein B</td>
<td>ENSG00000172943</td>
<td>X</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>PKN2</td>
<td>protein phosphatase Mg2+/Mn2+ dependent 1D</td>
<td>ENSG00000105647</td>
<td>19</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>POU3F3</td>
<td>POU class 3 homeobox 3</td>
<td>ENSG00000108891</td>
<td>2</td>
<td>Syndrome</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>PPM1D</td>
<td>protein phosphatase Mg2+/Mn2+ dependent</td>
<td>ENSG00000170838</td>
<td>17</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>3</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>PPP2CA</td>
<td>protein phosphatase 2 catalytic subunit alpha</td>
<td>ENSG00000113075</td>
<td>5</td>
<td>Syndrome</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>PRKD1</td>
<td>Protein kinase D1</td>
<td>ENSG00000108309</td>
<td>14</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>PRODH</td>
<td>Proline dehydrogenase (oxidase) 1</td>
<td>ENSG00000100033</td>
<td>22</td>
<td>Syndrome, Genetic Association</td>
<td></td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>POLR3A</td>
<td>RNA polymerase II subunit A</td>
<td>ENSG00000148906</td>
<td>10</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>3</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>RAC1</td>
<td>Rac family small GTPase 1</td>
<td>ENSG000001036238</td>
<td>7</td>
<td>Syndrome, Genetic Association</td>
<td></td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>RAD21</td>
<td>RAD21 cohesin complex component</td>
<td>ENSG000001064754</td>
<td>8</td>
<td>Syndrome</td>
<td></td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>RALA</td>
<td>Ras like protooncogene A</td>
<td>ENSG000001068451</td>
<td>7</td>
<td>Syndrome</td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>RHEB</td>
<td>Ras homolog mTORC1 binding</td>
<td>ENSG00000100615</td>
<td>7</td>
<td>Syndrome</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>RILM</td>
<td>Ring finger protein 1/CDK10 domain interacting</td>
<td>ENSG000001031263</td>
<td>X</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>RNF135</td>
<td>Ring finger protein 135</td>
<td>ENSG00000181481</td>
<td>17</td>
<td>Rare Single Gene Mutation</td>
<td></td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ROR2</td>
<td>RAR-related orphan receptor A</td>
<td>ENSG00000109667</td>
<td>15</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome, Genetic Association</td>
<td>1</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>RFX4</td>
<td>regulatory factor X4</td>
<td>ENSG00000111793</td>
<td>12</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>RFX7</td>
<td>regulatory factor X7</td>
<td>ENSG000001081827</td>
<td>15</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>RLM2</td>
<td>regulating synaptic membrane exocytosis 2</td>
<td>ENSG00000107406</td>
<td>8</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome, Genetic Association</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>RPS6K3</td>
<td>Ribosomal protein S6 kinase, mTORC1</td>
<td>ENSG00000177119</td>
<td>X</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>3</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>RSRC1</td>
<td>arginine and serine rich coiled-coil 1</td>
<td>ENSG00000174891</td>
<td>3</td>
<td>Syndrome, Genetic Association</td>
<td></td>
<td>1</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>SATB2</td>
<td>SATB homeobox 2</td>
<td>ENSG00000119804</td>
<td>2</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome, Genetic Association</td>
<td>3</td>
<td>1</td>
<td>33</td>
</tr>
<tr>
<td>SETD1B</td>
<td>SET domain containing 1B</td>
<td>ENSG00000139718</td>
<td>12</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>3</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>SGSH</td>
<td>N-sulfoglucosamine sulfohydrolase</td>
<td>ENSG00000181523</td>
<td>17</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>SIK1</td>
<td>Sirtuin 1</td>
<td>ENSG000001042718</td>
<td>21</td>
<td>Rare Single Gene Mutation</td>
<td>Syndrome</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
<td>Mutation Type</td>
<td>Syn</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------</td>
<td>------</td>
<td>----</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLC1A2</td>
<td>Solute carrier family 1, member 2 (glial high affinity glutamate transporter), member 2</td>
<td>Rare Single Gene Mutation, Genetic, Association, Functional</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLC45A1</td>
<td>Solute carrier family 45, member 1</td>
<td>Rare Single Gene Mutation</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMARC A2</td>
<td>SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily a, member 2</td>
<td>Rare Single Gene Mutation, Syn</td>
<td>1</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMC1A</td>
<td>Structural maintenance of chromosomes 1A</td>
<td>Rare Single Gene Mutation, Genetic</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMC3</td>
<td>Structural maintenance of chromosomes 3</td>
<td>Rare Single Gene Mutation, Syn</td>
<td>3</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNX14</td>
<td>Sorting nexin 14</td>
<td>Rare Single Gene Mutation, Genetic</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAG1</td>
<td>Stromal antigen 1</td>
<td>Rare Single Gene Mutation, Genetic</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOX6</td>
<td>SRY-box transcription factor 6</td>
<td>Syndromic</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SYNE1</td>
<td>Spectrin repeat containing, nuclear envelope 1</td>
<td>Rare Single Gene Mutation, Genetic, Association</td>
<td>3</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SYT1</td>
<td>Synaptotagmin 1</td>
<td>Rare Single Gene Mutation, Genetic, Association</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAF1</td>
<td>TATA-box binding protein associated factor 1</td>
<td>Rare Single Gene Mutation, Syn</td>
<td>1</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBC1D23</td>
<td>TBC1 domain family member 23</td>
<td>Syndromic</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBX1</td>
<td>T-box 1</td>
<td>Syndromic, Functional</td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUPP1</td>
<td>SPT5 homolog, facilitates chromatin remodeling subunit</td>
<td>Rare Single Gene Mutation</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#</td>
<td>Gene</td>
<td>Description</td>
<td>Gene ID</td>
<td>Function</td>
<td>Mutation</td>
<td>Associated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>----------</td>
<td>----------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ZSWIM6</td>
<td>Zinc finger SWIM-type containing 6</td>
<td>ENSG00000130440</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>YWHAQG</td>
<td>Tryptophan monoxygenase/hydroxylase S-adenosylhomocysteine S-monoxygenase activation protein gamma</td>
<td>ENSG00000170027</td>
<td></td>
<td>Rare</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ZMIZ1</td>
<td>Zinc finger MIZ-type containing 1</td>
<td>ENSG00000108175</td>
<td></td>
<td>Rare</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ZMYM2</td>
<td>Zinc finger MIZ-type containing 2</td>
<td>ENSG00000121741</td>
<td></td>
<td>Rare</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Characteristics used for custom analysis


