

# Targeted Cancer Therapy: From Scratch to Clinical Trials

Ojas Gupta

Lynbrook High School, 1280 Johnson Ave, San Jose, CA, 95129, Santa Clara Count, USA; ojasgupta1000@gmail.com

**ABSTRACT:** This paper attempts to flatten the learning curve for finding subtype-specific cancer drugs by outlining the different pathways to arrive at clinical trials starting from scratch. Targeted, or subtype-specific, cancer drugs work by targeting “cancer dependencies,” or the products of certain genes that are necessary for cancerous cell growth. Unfortunately, process are very prevalent in the industry. This paper will outline and compare current methods of identifying a cancer dependency and the subsequent drug molecule. Only a careful consideration of current limitations will improve our ability to correctly uncover cancer genetic dependencies and will facilitate the development of improved therapeutic agents.

**KEYWORDS:** Molecular Biology; Computational Biology; Genetics; Genomics; Computational Chemistry.

## ■ Introduction

Despite recent critical technological advancements in medicine, cancer remains the number one cause of death worldwide, with one new diagnosis every 30 seconds in the US alone.<sup>1,2</sup> Over 200 different types of cancer have been identified, and 1 in 3 people will develop one of them in their lifetime.<sup>1</sup> Despite various cancer-related guidelines for diagnosis, treatment, and follow-up, improving the long-term prognosis of certain cancer patients remains difficult. One reason for this bottleneck is because of the sluggish process involved in drug discovery. Drugs take 10-15 years to bring to market, and 97% of cancer drugs that enter clinical trials fail, one of the highest across all pharmaceutical fields.<sup>3</sup> Additionally, the drug discovery process is long, confusing, and filled with barriers. Thus, to maximize efficiency and improve outlooks, establishing and comparing current methods to find cancer drugs is imperative.

Cancer is the name given to a collection of related diseases that can start in nearly any organ or tissue when abnormal cells grow beyond their usual boundaries, invade adjoining parts of the body and/or spread to other organs. The latter process is called metastasis and is the major cause of death from cancer.<sup>1</sup> Tumors are often confused with cancer, however tumor is the word given to any uncontrolled cell growth, whereas cancer is a type of tumor that is malignant.<sup>1</sup> Cancer is caused by alterations, or mutations, in the genome. These changes in the DNA are induced by chemicals or radiation,<sup>4</sup> biological factors (virus and bacteria),<sup>4</sup> or spontaneously during DNA replication.<sup>5</sup> Mutations can result in genome instability, which can lead to aneuploidy, silencing of tumor suppressors, and/or alterations in protein functions, resulting in dysregulation of crucial biological pathways that lead to cancer.

Finding cancer dependencies is the norm for developing targeted cancer cell drugs. Like a person who is dependent on coffee to be productive, cancer cells are dependent on the products of certain genes in order to dominate their environment and grow.<sup>6</sup> Cancer cells will stop growing

and die when the activity of these gene products is blocked. Thus, a cancer dependency is a specific gene or protein that is required for cancer survival. Deleting that gene or inhibiting the encoded protein's activity, would result in repressed growth and/or death of the cancer cells.<sup>6</sup> As a result, researchers are constantly looking for cancer dependencies and developing drugs to block their activity.

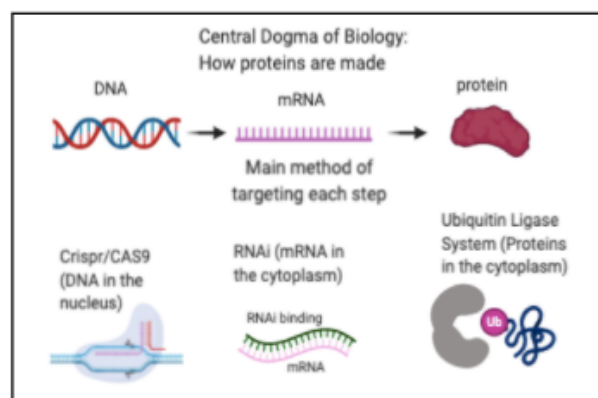
However, finding cancer dependencies is unfortunately much easier said than done. Most cancer dependencies result from alterations in numerous genes, making their identification nearly impossible due to the sheer number of possible genes and proteins in any given cell.<sup>7</sup> Additionally, techniques used to find dependencies are not perfect, and there have been many instances of revolutionary studies claiming to have found a cancer dependency only to be debunked later on (Table 1).<sup>6</sup> That is why cancer drugs have one of the highest failure rates of 97% and take over a decade to bring to market.<sup>3</sup> Thus, the goal of this paper is to review the available methods of finding a cancer dependency, and the methods of finding the subsequent cancer drug, as well as how to reduce the chances for failure or misidentification of a target.

**Table 1:** Examples of widely published yet misidentified cancer dependencies, adapted from Lin *et al.*<sup>6</sup>

Misidentified gene	Role/Cancers reported to be dependent
CD95 (or Fas)	Apoptosis & growth/ various cancers
HDAC6	Protein Maintenance/ various cancers
MELK	Cell Cycle/Triple-Negative Breast Cancer
MEN1	Tumor & growth suppressor/Glioma
MTH1 (or NUDT1)	Oxidation of DNA/ various cancers
PAK4	Signaling and cytoskeleton/ various cancers
PIM1	Glycolysis and growth/ various cancers
SMYD2	Signaling/ various cancers
STK33	Invasion and upregulation of transcription/ various cancers

## ■ Discussion

Cancer dependencies are almost always identified by using loss-of-function assays (LOF), where the effects of blocking the function of a gene is studied.<sup>7</sup> If the result is repressed growth and/or death of the cancer cells, then a promising cancer target has been identified. Following the central dogma of biology—transcription and translation—LOF assays can target 3 distinct molecules: DNA, RNA, or protein (Figure 1). By its definition, the inhibition of a cancer dependency will lead to repressed growth and/or cancer cell death. Thus, a successful cancer drug is one that can target a cancer dependency. For many of these methods, the DNA or RNA sequence needs to be known, which won't be detailed in this paper but has extensively been detailed in others.<sup>8</sup>



**Figure 1:** Summary of the main Loss-of-function techniques detailed in this paper, sorted by which molecule the method targets. By the central dogma of biology, these could target DNA, mRNA, or the protein, and are achieved using CRISPR/Cas9, RNAi, and PROTAC/ULS respectively.

### *DNA level targeting:*

Direct DNA alterations, or genome therapy, is primarily achieved through recently discovered CRISPR/Cas.<sup>9</sup> This system naturally evolved as the “immune system” of bacteria to remove foreign DNA injected by bacteriophages.<sup>9</sup> CRISPR stands for Clustered Regularly Interspaced Short Palindromic repeated DNA sequences. This term was coined because when sequencing bacterial genomes, scientists observed clustered repeat sections in DNA separated by non-repeating DNA sequences named spacers unlike typical tandem repeats in the genome.

At first, no conclusions or investigations were made into this strange coincidence. Eventually, the non-repeating spacer DNA sequences were identified as belonging to viruses and other mobile genetic elements. After extensive research, it was determined that small-hairpin RNAs known as CRISPR RNAs (crRNAs, also called guideRNA/gRNA or single guideRNA/sgRNA) are synthesized from bacteriophage DNA sequences stored in these spacers in bacterial DNA. This crRNA binds to a protein in the Cas enzyme family and stays in the cytoplasm just as antibodies would stay in human blood. When the same phage injects its DNA again, the crRNA will bind to it via complementarity. Then, the Cas9 enzyme will initiate a double strand break in the foreign DNA, preventing an infection. However, why the crRNA and Cas enzyme complex did not destroy the virus DNA stored in the

discovered that in order for Cas enzymes to be activated, it needs to bind to a protospacer-adjacent motifs (PAMs), which is a specific short sequence of nucleotides uniquely found on foreign DNA, but removed once integrated into the bacterial genome.<sup>9</sup> Individual Cas proteins have their own “activation” PAM sequences. The most common enzyme, Cas9, has a PAM site of NGG, or any base followed by two guanines.<sup>9</sup> Because of the simple PAM requirements and versatility, Cas9 has become a commonly used enzyme in genome manipulation.<sup>10</sup> There are two ways CRISPR can be introduced into the cell nucleus for gene knockout: transfection and transduction.<sup>10</sup> Transfection is when molecules enter the cell directly through their cellular membrane. In CRISPR transfection, there are three types of molecules that can be used: DNA, RNA, or in its final protein form (Ribonucleoprotein complex).<sup>10</sup> Electricity, liposomes, and microinjection are commonly used to introduce the molecule of choice into the cell.<sup>10</sup> Either way, the final protein complex of the crRNA and the Cas enzyme will be assembled in the cells. Transduction, on the other hand, utilizes a Lenti or Adeno Associated virus to carry the complex and insert it directly into the cell. From this point, the two methods work the same. The complex binds to the target sequence as directed by the crRNA and causes the double stranded break. From here, the cells will attempt to repair the double stranded break through one of two pathways: non-homologous end joining (NHEJ) or homology-directed repair (HDR). In NHEJ, error-prone DNA-Ligase attempts to fix the break, often adding random bases resulting in frameshift mutations. This generates a null, or total LOF.<sup>9</sup> HDR, on the other hand, can insert the missing gene from homologous or sister chromatids, or insert a completely new gene.<sup>11</sup> Additionally, scientists sometimes insert a gene like GFP or antibiotic resistance so they can label and observe cells that underwent the change. Usually, these selection markers are inserted along with the CRISPR complex. If transfection is used, then short single-stranded DNA (ssDNA) also called donor oligonucleotides or synthetic DNA, carry the genes.<sup>11</sup>

CRISPR is one of the most commonly used tools in LOF assays, and by far the most popular method of direct gene knockout.<sup>12</sup> There are many older and less-commonly used DNA editing tools that can also be used in knockout assays. One of them is the Artificial Restriction DNA Cutter (ARCUT) method. This method does not use enzymes and can knockout or insert genes of interest by using pseudo-complementary peptide nucleic acid (pcPNA). The pcPNA specifies the cleavage site and binds to each side of the DNA, unraveling it. Chemicals cerium (CE) and EDTA (mixture) carry out the actual splicing. Then, DNA ligase in the cell attaches genes of interest as provided. This technique is particularly useful in extreme conditions where enzymes cannot function properly; however, ARCUT is not popular due to longer turnaround time and need for site-specific pcPNA.<sup>12</sup> Another less common method of DNA knockout is Protein-Based Nuclease systems. Humans naturally contain Meganucleases, also called endonucleases, which are molecular DNA scissors. This technique has three steps: first is the recognition of a cleavage site, then Meganucleases splice out

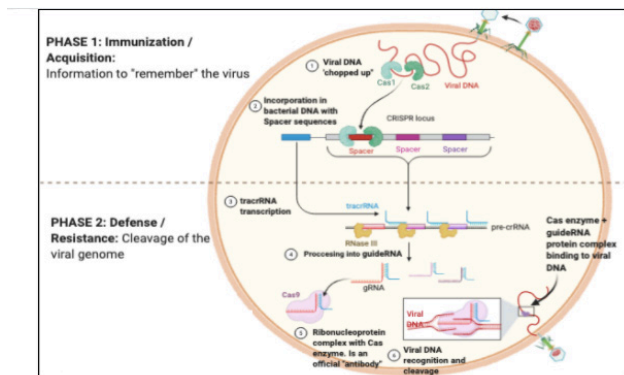
the region, and finally the desired DNA sequence is inserted by DNA ligase. The first recognition step is done by either Zinc-Finger-binding protein domains, or TALEN's. Zinc Fingers can be programmed to recognize combinations of codons, which makes TALEN's a little easier to work with as they can recognize combinations of individual base-pairs.<sup>12</sup> This technique is not preferred due to high cost and time involved in generating a Zinc-Finger or TALEN recognition sequence, even though it has shown to be less toxic than CRISPR since meganucleases naturally occur in the body whereas foreign Cas enzymes can trigger an immune response.<sup>12</sup>

Although CRISPR is much faster, cheaper, and more specific than all other DNA inhibition techniques, it is not perfect. The biggest issue is that crRNA tends to have relatively high mismatch tolerance. In other words, Cas9 commonly cleaves off-target sites, a problem shared by all LOF techniques to some degree.<sup>12</sup> Different Cas proteins and more specific gRNA's have combated this issue, but these efforts are far from perfect. However, it is worth mentioning that CRISPR is widely considered to be the most accurate and specific form of LOF assays.<sup>6,12</sup> Another issue is that the use of knockout techniques to remove one gene's expression might result in compensatory upregulation in a paralogous gene that achieves the same phenotype or role, thus falsely portraying the role of the targeted gene as null.<sup>13</sup> Additionally, the immunogenicity of Cas9 is of concern. Around 50% of humans contain antibodies to the two most common CRISPR nucleases. This leads to low knockout efficiency and possibly a serious immune storm *in vivo*.<sup>13</sup> Recent research has also shed light on CRISPR's

leading to apoptosis.<sup>13</sup> Finally, the CRISPR-induced mutation will not always completely ablate gene expression. Sometimes functional truncated proteins, or "zombie" proteins, can still be expressed in small amounts.<sup>14</sup> CRISPR/Cas9 has allowed us to manipulate gene expression and study gene function on a whole new scale, but it is still far from ideal (Figure 2).

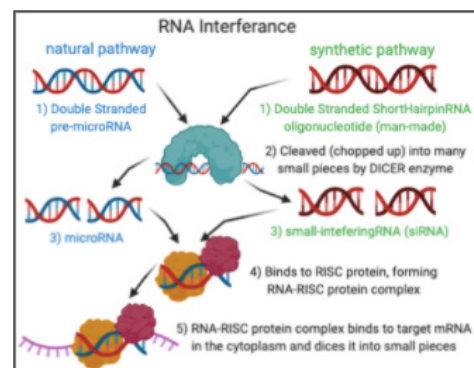
#### RNA level inhibition:

Unlike direct genome inhibition, perturbations at the RNA level do not completely knockout a gene: instead it induces significant knockdown, or downregulation, of a gene to the point where the importance of the gene can be determined.<sup>15</sup> It is important to note that RNAi is a natural mechanism used by cells to limit mRNA expression, and for use in LOF assays, a synthetic version that mimics the natural mechanisms is used (Figure 3). Natural RNAi works by introducing a double stranded microRNA (miRNA), that binds to the mRNA of the target gene in the cytoplasm.<sup>15</sup> Once bound, the natural miRNA can recruit cell machinery such as the DICER enzyme, forming a RISC complex, and cut the mRNA into small, useless chunks.<sup>15</sup> Alternatively, the RISC complex can bind to the target mRNA, simply preventing translation.<sup>15</sup>



**Figure 2:** Summarizes the 6 steps involved in natural CRISPR/Cas9 in bacteria as an "immune response for viral DNA infections. First, the intruding viral DNA is chopped up into many small sections. Then, it is incorporated into the bacterial chromosome with repeating "spacer" or dummy sequences between different viral segments. Next, in the Defense/Resistance half, tracrRNAs are sequenced directly from the viral sequences stored in the chromosome. These RNA's are modified and mature into guide RNAs by RNaseIII. Next, the mature gRNA binds to the Cas enzyme (Cas9 in this case) and forms the active and ready protein complex. In an analogy to the human body, this would be considered an active antibody in the blood. Finally, when the complex recognizes the foreign DNA, it will bind to it and prevent any damage from occurring.

role in disrupting the cell cycle and activating p53, which is a protein that senses DNA damage inside the cell. Since CRISPR/Cas9 induces double stranded breaks in the DNA, it can cause the p53 DNA-correction pathway to activate if the cell has the wild type, or natural (not mutated) version of p53,



**Figure 3:** RNAi is a natural mechanism used to limit gene expression. The clinical RNAi is a synthetic version which can mimic the natural mechanism and produce the same results. The natural pathway is shown blue, synthetic in green, and overlapping steps in black.

Short hairpin RNA (shRNA) and small interfering RNA (siRNA) are the two types of oligonucleotides, or man-made DNA, which can mimic the same role as natural occurring miRNA.<sup>16</sup> The method for application in humans is straightforward. Transduction via adeno or lenti viruses is most commonly used *in vitro*, as with CRISPR, but if assay is occurring in suspension, transfection is also used. Both methods get the amplified synthetic shRNA or siRNA, which needs to be complementary to the target mRNA sequence, into the cells of interest where the mechanism of binding to the RISC protein and subsequent steps occurs naturally.<sup>16</sup>

As mentioned before, because RNAi downregulates a gene's expression but does not remove it, RNAi harbors several advantages relative to other systems in the drug discovery process than any other LOF technique. This is due to the fact that it closely mimics the role of a drug, which also will downregulate the protein but not completely inhibit it.<sup>15</sup> Another advantage of RNAi over CRISPR is that it is more efficient as it is able to remove the function of homologs, as



RNAi targets mRNA instead of the actual genome, accounting for the issue of upregulation of homologs in CRISPR. If there are more mRNA then you would need more microRNA to knockdown the gene, hence its knockdown efficiency is reliant on gene prevalence.<sup>15</sup> However, RNAi's largest disadvantage is its prevalent off-target effects.<sup>16</sup> MELK, PIM1, PAK4, HDAC6, STK33 and other oncogenes were initially deemed essential to cancer cells by using RNAi (as listed in Table 1), but peer reviews revealed that the outcomes observed by most initial studies were actually due to RNAi degrading other mRNA strands that were not the initial target.<sup>6</sup>

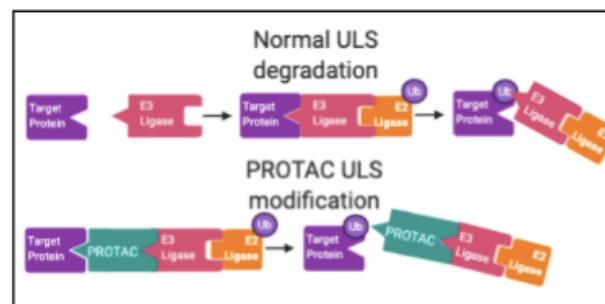
#### Protein level inhibition :

Finally, LOF assays can also inhibit function on the proteome level. Targeted proteolysis is desirable both in drug applications and in basic research. There are two main options for disrupting protein function: the inhibition of protein activity and the induction of protein degradation.<sup>16</sup> Small-molecule inhibitors (SMIs) are the most widely used tools for the first method of protein activity and are amenable to high-throughput screening, making their use and detection easier.<sup>16</sup> These are specially designed molecules that inhibit protein function by binding to a protein, thereby preventing it from doing its normal function. While these are fast acting, controllable, and efficient, SMI's have two main drawbacks preventing their widespread use. Each protein has a different structure, and the majority of proteins lack a SMI that can cause effective inhibition.<sup>16</sup> Additionally, SMI's have several off-target instances since many proteins in the body have similar receptors.<sup>16</sup>

Recently, the protein inhibition field has shifted towards the ubiquitin ligase system (ULS), a protein degradation system. Like RNAi, this mechanism was observed in organisms and then a synthetic version was created that can mimic the natural process and induce the same outcome (Figure 4). ULS is a tagging system that attaches ubiquitin to the protein of choice, which then signals degradation by a protease.<sup>17</sup> The process in the human body uses three steps, each with a different ubiquitin-activating enzyme.<sup>17</sup> The first step of the reaction is mediated by the E1 ubiquitin-activating enzyme that activates the ubiquitin monomer. Next, the activated ubiquitin is passed on to the E2 ubiquitin-conjugating enzyme, before ultimately being attached to the target protein through the action of an E3 ubiquitin ligase. The ubiquitin monomers can be further ubiquitinated to form chains, with distinct chain types linked to specific biological processes. If, for example, ubiquitin molecules added to the target protein form a lysine 48-linked chain, the ubiquitinated protein is marked for proteasomal degradation.<sup>18</sup>

The synthetic "copycat" version of ULS is PROTAC, or PROteolysis-TArgeting Chimeras. PROTACs are a more efficient way to cause protein degradation using the natural Ubiquitin recycling system. It consists of a ligand to the target protein bound to a ligand to the E3 Ubiquitin Ligase.<sup>19</sup> This binds the target protein directly to the E3 ligase, speeding up the ubiquitination process, triggering degradation in the same method.<sup>20</sup> In application, a ligand for the target protein

needs to be established and reproduced in order for PROTAC to work.<sup>20</sup> In summary, PROTAC mimics the natural ULS system mentioned before, by holding the protein close to the E3 enzyme by using ligands and more efficient binding



**Figure 4:** Similarly to RNAi, ULS is a natural pathway of limiting expression, but unlike RNAi, scientist have given a name to the synthetic version; PROTAC. The main difference is that PROTAC and natural ULS is that PROTAC can force controlled degradation, whereas the natural ULS pathway is only triggered by certain receptors. It is important to note that the function of the E1 ligase has not been included in this figure which normally brings the Ubiquitin protein from the ribosome and delivers it to the E2 ligase.

techniques, which forces ubiquitination and subsequent degradation by protease.

Currently, the PROTAC approach in LOF assays has four main limitations. The first three are: the lack of E3 ligases that can bind to the protein of interest, the lack of nanobodies or SSM to bind to the protein of interest, and the large size of PROTACs which occasionally prevent passage through the cellular membrane.<sup>18</sup> Luckily, the recent popularity has led to a huge increase in the enzyme and molecule libraries, as well as various adaptations with increased permeability.<sup>18</sup> In other words, due to the high efficiency and popularity, as well as the low toxicity, the first three problems are being rapidly solved. The fourth limitation, however, is more difficult to overcome. The Synthetic Small Molecules work by binding to the active site of a protein. However, 75% of the human proteome lacks active sites, as they don't actively perform any enzymatic function (ex: scaffolding or structural proteins & transcription factors).<sup>20</sup> However, there are still many proteins with active sites yet to be discovered and detailed, and humans have miles to go before we run out of "targetable" proteins.<sup>20</sup>

#### Comparison and Emerging Tools:

Currently, CRISPR and RNAi are the two most common methods for testing and identifying cancer dependencies and the only possible methods used for genome-wide screens.<sup>6</sup> As a head to head comparison, CRISPR generally outperforms RNAi in terms of single gene experiments regarding specificity and off target effects.<sup>21</sup> CRISPR also presents the unique opportunity to study cells with genetic changes, since the changes that are induced by CRISPR are inherited by the offspring.<sup>22</sup> RNAi, on the other hand, also possesses other benefits. It imitates the role of a natural drug better compared to CRISPR: decreasing but not eliminating the protein of interest. Additionally, unlike CRISPR, RNAi effects can be reversed and controlled, allowing researchers to test for the consequences of restoring gene expression.<sup>23</sup> Additionally, determining the DNA sequence that codes for a specific protein can be mis-

that are trimmed out before RNA translation. Thus, finding the DNA sequence for a set of amino acids isn't always straightforward, giving RNAi a slight advantage in convenience due to the relative simplicity in the direct mRNA-protein translation.<sup>23</sup> Finally, CRISPR sometimes initiates anti-tumorous cell machinery by activating the p53 DNA correction pathway, a problem that RNA does not have.<sup>13</sup>

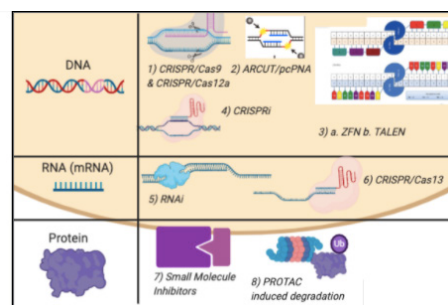
Although these techniques are far from perfect, recent alterations to the transcriptome editing toolbox have shown promise. The ubiquitin ligase system is one of these systems. By utilizing the body's natural recycling system, it has shown fewer off target effects and significant knockdown ability.<sup>18</sup> Additionally, three modifications to the CRISPR/Cas9 technology hold immense potential. CRISPR interference (CRISPRi) is one such modification. As its name suggests, it also relies on the CRISPR protein complex, which is a Cas enzyme bound to a guideRNA that is complementary to the target DNA sequence. Unlike CRISPR, CRISPRi implements an inactive or dead Cas enzyme (dCas) that lacks endonuclease activity. It acts as a repressor protein, binding to the target sequence and blocking transcription from RNA polymerase.<sup>24</sup> Thus, although CRISPRi does not directly edit DNA like CRISPR, it can efficiently block expression of targeted genes. CRISPRi also appears to produce a more consistent and robust knockdown with fewer off target effects and has advantage over CRISPR as it does not activate the p53 DNA correction pathway.<sup>21</sup> The application or method of delivery for CRISPRi is identical to that of CRISPR. Cas12a has also received significant attention recently as an alternative Cas enzyme utilized in normal CRISPR gene knockout. It has a significant advantage over traditional Cas9 because Cas12a contains DNAase and RNAase activity within the protein complex, which allows for "scanning" and processing of multiple guideRNAs simultaneously.<sup>25</sup> Although it operates at a slower pace, Cas12a offers higher efficiency if multiple genes need to be screened, which happens many times as multiple genes often code for proteins with redundant functions.<sup>25</sup> Other than multiple guideRNAs and a different Cas enzyme, the Cas12a gene-knockout mechanisms and application is identical to CRISPR. Finally, Cas13 is another recently discovered Cas enzyme that has attracted a lot of attention in LOF assays. Unlike other Cas enzymes, Cas 13 has the ability to recognize and bind to mRNA instead of DNA, inducing gene knockdown the same way as RNAi does.<sup>26</sup> Although there are many additional techniques used to find cancer dependencies, the most popular ones were discussed (Figure 5).

#### Drug Discovery:

Unfortunately, conventional drugs face the same problem as protein-level inhibition: many proteins are simply not able to be targeted. Traditional drugs work by binding to the active site of a protein, thus preventing its normal function. However, out of the millions of proteins in the body, a vast majority (75%) of the human proteome lacks these active sites and thus are undruggable by conventional medicine.<sup>20</sup> Of the 25% that we can exploit, a vast majority has still been untouched.

There are already a variety of methods and drugs that are commonly used to combat cancer. These include anti-growth

drugs and techniques such as chemotherapy and radiotherapy, which are generally used despite the type of cancer and can help limit the growth of most cancers.<sup>27</sup> However, the goal of LOF assays, as well as the goal of this paper, is to establish the steps for targeted-drug therapy for cancers that display a certain biomarker. Due to its specificity, these drugs tend to have twice as high success rates in approval and higher relative efficiency in treatment than drugs without specificity to certain cancer subtypes.<sup>28</sup> The goal of targeted drug development is to find and deliver a molecule that will bind to the target protein and prevent its function, without disrupting other proteins' activities. This paper only focuses on the first task of finding the right molecule. The path to finding a drug is long and complicated, with no "right way" or industry-accepted path. Once a cancer dependency has been found using LOF assays for the desired cancer subtype, then there are two general pathways for drug development: Structure Based Drug Design & Ligand



**Figure 5:** Summary of the eight main LOF methods detailed in this paper. These are sorted by which molecule they target. On the DNA level, we have the CRISPR complex first that searches for the target gene and induces a DSB. Both enzymes Cas9 and Cas12a can be used except Cas12a can hold/scan for multiple genes at once. Second, there is the ARCUT method which works by introducing pcDNA that unwinds and holds DNA where it is exposed to certain chemicals inducing DSBs. Third, there are the ZFN and TALEN methods, who both scan for the target gene and recruit naturally occurring meganuclease "DNA scissors" to induce DSB's. ZFN recognizes codons whereas TALEN recognizes individual bases. Fourth, the CRISPRi method which is a dCas protein that simply binds to the promoter, blocking transcription. On the mRNA level, there is the fifth method of RNAi which mimics the natural process to dice target mRNA into useless pieces. Sixth, the CRISPR/Cas13 complex binds to and dices mRNA, preventing translation. Finally on the protein level, there is the seventh method of SMI's which bind to the active site of a protein, preventing any useful action from occurring. The eighth and final system is PROTACs which imitate the body's natural protein recycling system and induce protein degradation.

Based Drug Design (Figure 6).

#### Structure Based Drug Design:

This pathway is undertaken when the structure of receptors (active sites) of the target protein is known.<sup>29</sup> As detailed in the protein inhibition section, the structure of the receptor is determined by X-ray crystallography, NMR, or through an online database. Protein structure can also be predicted by computational methods like threading and homology modeling. Threading predicts the shape based on established shapes of sequences of amino acids and domains.<sup>29</sup> Homology modeling predicts the shape based on a homologous protein whose shape has been identified.<sup>30</sup> Finally, the pseudo-receptor technique can also generate an accurate model of what the target protein's structure is. Once the target protein's receptors have been detailed and mapped, then the formation of a new drug can

occur. In the diagram, this is represented by the *de novo* drugs pathway as *de novo* means “from the beginning”. Some of the methods to create viable drugs are listed below.<sup>29</sup>



**Figure 6:** Although any possible method of identification is accepted, these are some of the most popular routes of finding a potential drug candidate. The methods are mainly grouped on whether they need to know the structure of the receptor of the target protein (cancer dependency) or not. The one exception to this is the Pseudoreceptor method. Although it is listed as a LBDD method in the text, this technique essentially utilizes the properties of a receptor protein, which means that it could be used to find the target structure as well as a drug.

1) **Fragment location methods:** To determine desirable locations of atoms or small fragments within the active site (the detailed receptor structure).

2) **Site point connection methods:** To split up the big hole into small gaps, or “site points”, and then place fragments in those small gaps until all such locations are occupied by suitable molecules.

3) **Fragment connection methods:**

Fragments are positioned and “linkers” or “scaffolds” are used to connect those fragments and hold them in a desirable orientation. This is usually used in conjunction with method 2.

4) **Sequential buildup methods:** Construct a ligand atom by atom, or fragment by fragment.

5) **Whole molecule methods:** Random established compounds are placed into the active site in various conformations in silico (computer model), assessing shape and/or electrostatic complementarity until the optimal molecule has been chosen.

In addition to *de novo* drug synthesis, Structure Based Virtual Screening (SBVS) has become a popular way to test for viable drugs.<sup>30</sup> First, a 3D scan of the target receptor is required. Then, using virtual compound libraries, computer-generated representations of known small molecules are “placed” in the 3D model of the receptor in a variety of positions, conformations and orientations.<sup>30</sup> Each such docking mode is called a ‘pose’, and are given a score based on its complementarity to the target in terms of shape and properties such as electrostatics. A good score for a given molecule indicates that it is potentially a good binder. This process is repeated for all molecules in the collection, which are subsequently rank-ordered by their scores, revealing the most viable drug options.<sup>30</sup> AutoDock is one of the many websites containing both the compound library and the random placement+scoring technology, which have played a crucial role in making SBVS accessible.<sup>31</sup> The one limiting factor of SBVS is that it only evaluates known molecular compounds instead of creating truly new compounds, which might

not always produce beneficial results or result in copyright battles.

### Ligand Based Drug Design:

Ligand Based Drug Design is an approach used in the absence of the target receptor structure, and instead it relies on knowledge of molecules that bind to the biological target of interest, or ligands. This approach can be broken down into four main strategies: Pseudoreceptors, QSAR, Scaffold Hopping, and Pharmacophore modeling.

The pseudoreceptor method not only functions as a valuable tool in the drug design process, but also provides an accurate idea of the structure of a target protein, which in some cases is enough to warrant Structure Based Drug Design approaches.<sup>29</sup> This method functions by studying known bioactive ligands, or proteins in the body that naturally bind to the target protein.<sup>29</sup> The drug is then modeled off of this natural substrate, and in some cases, the natural protein with slight editing actually becomes the drug.<sup>29</sup> Additionally, by studying the natural substrates of the target protein, much is learned about the target proteins requirements and structure, which aids in the drug design process.

Quantitative Structure Activity Relationship (QSAR) is one of the most widely used Ligand Based Drug Design approaches. At its core, QSAR is very simple. It is based on the assumption that structural or physiochemical properties yield similar activity.<sup>32</sup> Thus, by comparing the function of a molecule in other reactions, as well as known ligands or reactions involving the target protein, the reaction between a molecule and the target protein can be predicted. The end result is the same as the SBVS method. By performing this prediction on a large molecule library, and by assigning scores, the most promising “regions” and molecules are identified.<sup>33</sup> For 3D QSAR application, two main types of approaches are used: CoMFA and CoMSIA. Comparative Molecular Field Analysis (CoMFA) evaluates basic non-binding interactions between a ligand, or molecule, and the protein.<sup>34</sup> Comparative Molecular Similarity Indices Analysis (CoMSIA) measures more in-depth physico-chemical properties of the target protein, not molecule interactions.<sup>34</sup> Basic computation and cluster-mapping are used for prediction and scoring features.

The next method of LB Drug Design is scaffold hopping. This method is used when a ligand that binds to the protein of interest is already known. Scaffold hopping preserves the structure of the molecule but swaps out certain molecules and elements in favor of others.<sup>19</sup> Thus, it maximizes the efficiency of a molecule while making sure it still “fits”. Scaffold Hopping might be necessary for many reasons: improved pharmacokinetic properties, lower toxicity, better efficiency, and/or intellectual property rights.<sup>19</sup> It is important to note that this is more of a “finishing step” approach rather than starting from scratch. Research of a molecule (natural or synthesized) that already binds to the target is a prerequisite to this step.

The fourth main method of LB Drug Design is pharmacophore modeling. This is the oldest method of drug design, however recent online variations have revived popularity.<sup>29</sup> Pharmacophore modeling is a strategy focused on identifying



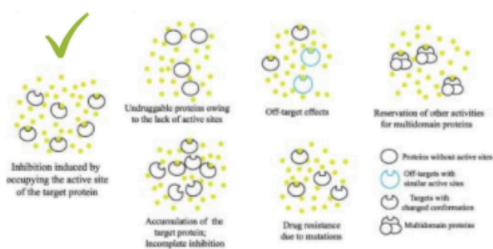
structural and biological features and putting together molecules that will meet the requirements. A popular used analogy is identifying and combining basic building blocks of molecules to make a compound that can perform all needed functions. For direct application, this happens in three simple steps:

- 1) finding the features required for a particular biological activity
- 2) determining the molecular shape or confirmation required
- 3) developing a superposition or order for the series of molecules.

This is usually now performed using online screening, with online 3D screening of databases to find the right molecules and order to make a compound, or drug. There are various commercially available websites, such as CATALYST, each with different filters promising better results.<sup>29</sup> Because a lot of the approaches are similar, some distinguishing is needed. This approach of pharmacophore modeling is similar to methods 2 and 4 in the de novo drug design process (Site point connection methods & Fragment connection methods), except those methods build up drugs mainly on structure whereas pharmacophore modeling builds up molecules mainly by function. Pharmacophore modeling is also similar to QSAR LBDD as they both build up drugs molecule by molecule, but pharmacophore modeling is focused more on starting from scratch and meeting the “final drug” attributes or features, whereas QSAR is focused more on scientific feasibility, chemical properties, and maximizing efficiency. However, there is a lot of gray area, and often the two are used in conjunction.

#### Validation:

Many of these approaches are often used together in the drug design process. Besides legal and financial problems, the top two reasons why 97% of drugs fail the development process



**Figure 7:** A summary of some problem drugs can run into. The one correct or positive outcome is shown with a green checkmark.

is due to lack of efficacy and off-target or toxic effects.<sup>35</sup> There are many causes for this, mainly due to other proteins with similar active sites that the molecule will unnecessarily effect (Figure 7). Thus, it is incredibly important to validate results using as many orthogonal techniques as possible to ensure that A) the drug target is actually a cancer dependency and B) the synthesized drug is able to effectively act only on the proposed target. This can be achieved in many ways, but the “golden-standard” for drug validation is a mutation assay.<sup>6</sup> This is done by inducing a mutation in the target protein that prevents the synthesized drug from binding. Thus, the drug should have no effect on growth of these mutant cells as it shouldn’t be able to bind to anything. If there is repressed growth or lack

of growth even from the mutant cells, an off-target effect has been identified and the drug design process needs to restart. However, inducing a mutation in the target protein isn’t always plausible, as not all proteins are easily mutated.<sup>6</sup> Additionally, the molecule might have other off-target or toxic effects when introduced in the body, which can’t be tested for in preclinical trials or in “golden-standard” mutation assays.<sup>6</sup> Currently, these other possibilities accounted for using mice or animal trials, although these don’t always represent the same factors in a human body. At this point, the drug is ready to go into clinical trials to test effects on a larger scale.<sup>36</sup>

#### Conclusion

This paper attempted to describe current methods for finding a cancer drug from scratch and flatten the learning curve for the overall drug discovery process. First, a definition and background of cancer was given. The three main causes of cancer, the body’s natural defense mechanisms for mutations, and the six hallmarks of cancer were given. Next, cancer dependencies were defined, which are needed for targeted-drugs therapy. Loss-of-function (LOF) assays were described as the primary experimental technique for establishing cancer dependencies, targeting either the DNA, RNA, or protein levels. Popular LOF techniques used to identify drug targets were discussed, with CRISPR/Cas9, RNAi, and PROTAC emerging as the most popular methods for every respective molecule. Additionally, a head-to-head comparison of CRISPR/Cas9 and RNAi was also listed, as well as an introduction to recent modifications that show enormous potential to the gene expression toolbox. Finally, the process of creating a drug targeting the protein arising from the cancer dependency was covered, split on whether the structure of the target protein was known or not. The validation process and the golden standard were also given.

The pathway to developing cancer drugs, however, still has a couple of major hurdles. One big hurdle is the fact that when creating a cancer drug molecule, there are possible off-target effects in the body which cannot be predicted in advance. This is only sufficiently observed in clinical trials, after a lot of time and money has already been invested. There are databases that are being created that try to contain all the active sites of proteins in the human body, and the compatibility or effects of a certain molecule can be simulated. Unfortunately, these databases are nowhere close to containing all the proteins or matching the complexity of reactions in the human body.

The future for fighting cancer is bright. Since the year 1991, there has been a 29% decrease in cancer deaths, with last year providing the single biggest yearly drop in deaths ever recorded. This has been fueled by many things. There has been a public awakening about carcinogens, smoking, and other environmental factors that increase cancer prevalence. There has been extensive research on obesity, diet, and some on the role of the gut microbiome in preventing or causing cancer. Increased public awareness and inevitable discoveries yielding more about cancer will further help decrease the number of cancer deaths and continue this trend in the future. The cancer research field will also undergo massive growth. Most of the

processes discussed in this paper were discovered at the end of the 20th century, apart from some of the DNA gene knockout techniques. Over the next decade, future technologies will no doubt render today's leading techniques obsolete as well. Out of those discussed in this paper, the field is shifting towards CRISPR/Cas12a, Cas13, and PROTAC systems in particular, although a variety of techniques both new and old are used in different labs. The biggest factor that needs improvement, as well as the benchmark for measuring the efficiency of new methods, is the prevalence of off-target effects: that is, when a target or molecule binds to something that wasn't intended, causing unintended consequences. Although cancer has been wreaking havoc on humanity for centuries, recent technologies have led to exponential amounts of research and development, and the end of cancer as we know it might be sooner than we expected.

### ■ Acknowledgements

Huge thanks to Ann Lin, Ph.D. candidate at Stanford for genetics and microbiology, who gave me this idea and introduced me to the scientific paper comprehension and writing process. She is both the “why” and the “how” of this paper. I truly could not have done it without her.

### ■ References

1. WHO. Cancer <https://www.who.int/westernpacific/health-topics/cancer> (accessed Jul 8, 2020).
2. Cancer Facts & the War on Cancer | SEER Training <https://training.seer.cancer.gov/disease/war/> (accessed Oct 2, 2020).
3. Lin, A.; Giuliano, C. J.; Palladino, A.; John, K. M.; Abramowicz, C.; Yuan, M. L.; Sausville, E. L.; Lukow, D. A.; Liu, L.; Chait, A. R.; Galluzzo, Z. C.; Tucker, C.; Sheltzer, J. M. Off-Target Toxicity Is a Common Mechanism of Action of Cancer Drugs Undergoing Clinical Trials. *Sci. Transl. Med.* 2019, 11 (509). <https://doi.org/10.1126/scitranslmed.aaw8412>.
4. Van Elsland, D.; Neefjes, J. Bacterial Infections and Cancer. *EMBO Rep.* 2018, 19 (11). <https://doi.org/10.15252/embr.201846632>.
5. Cooper, G. M. *Oncogenes*. Cell Mol. Approach 2nd Ed. 2000..
6. Lin, A.; Sheltzer, J. M. Discovering and Validating Cancer Genetic Dependencies: Approaches and Pitfalls. *Nat. Rev. Genet.* 2020, 1–12. <https://doi.org/10.1038/s41576-020-0247-7>.
7. Tsherniak, A.; Vazquez, F.; Montgomery, P. G.; Weir, B. A.; Kryukov, G.; Cowley, G. S.; Gill, S.; Harrington, W. F.; Pantel, S.; Krill-Burger, J. M.; Meyers, R. M.; Ali, L.; Goodale, A.; Lee, Y.; Jiang, G.; Hsiao, J.; Gerath, W. F. J.; Howell, S.; Merkel, E.; Ghandi, M.; Garraway, L. A.; Root, D. E.; Golub, T. R.; Boehm, J. S.; Hahn, W. C. Defining a Cancer Dependency Map. *Cell* 2017, 170 (3), 564–576.e16. <https://doi.org/10.1016/j.cell.2017.06.010>.
8. DNA Sequencing: History, Steps, Methods, Applications And Limitations. Genetic Education, 2019.
9. Adli, M. The CRISPR Tool Kit for Genome Editing and Beyond. *Nat. Commun.* 2018, 9. <https://doi.org/10.1038/s41467-018-04252-2>.
10. CRISPR Transfection Protocols Guide: How To Select The Best Method <https://www.synthego.com/guide/how-to-use-crispr/transfection-protocols> (accessed Aug 11, 2020).
11. Cortez, C. CRISPR 101: Homology Directed Repair <https://blog.addgene.org/crispr-101-homology-directed-repair> (accessed Jul 28, 2020).
12. Khan, S. H. Genome-Editing Technologies: Concept, Pros, and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application. *Mol. Ther. Nucleic Acids* 2019, 16, 326–334. <https://doi.org/10.1016/j.omtn.2019.02.027>.
13. You, L.; Tong, R.; Li, M.; Liu, Y.; Xue, J.; Lu, Y. Advancements and Obstacles of CRISPR-Cas9 Technology in Translational Research. *Mol. Ther. Methods Clin. Dev.* 2019, 13, 359–370. <https://doi.org/10.1016/j.omtm.2019.02.008>.
14. Meraldi, P. Bub1—the Zombie Protein That CRISPR Cannot Kill. *EMBO J.* 2019, 38 (7). <https://doi.org/10.15252/emboj.2019101912>.
15. Fire, A.; Xu, S.; Montgomery, M. K.; Kostas, S. A.; Driver, S. E.; Mello, C. C. Potent and Specific Genetic Interference by Double-Stranded RNA in *Caenorhabditis Elegans*. *Nature* 1998, 391 (6669), 806–811. <https://doi.org/10.1038/35888>.
16. Housden, B. E.; Muhar, M.; Gemberling, M.; Gersbach, C. A.; Stainier, D. Y. R.; Seydoux, G.; Mohr, S. E.; Zuber, J.; Perrimon, N. Loss-of-Function Genetic Tools for Animal Models: Cross-Species and Cross-Platform Differences. *Nat. Rev. Genet.* 2017, 18 (1), 24–40. <https://doi.org/10.1038/nrg.2016.118>.
17. Chamberlain, P. P.; Hamann, L. G. Development of Targeted Protein Degradation Therapeutics. *Nat. Chem. Biol.* 2019, 15 (10), 937–944. <https://doi.org/10.1038/s41589-019-0362-y>.
18. Advances in targeted degradation of endogenous proteins | SpringerLink <https://link.springer.com/article/10.1007/s00018-019-03112-6#Tab1> (accessed Aug 15, 2020).
19. Bergmann, R.; Linusson, A.; Zamora, I. SHOP: Scaffold HOPping by GRID-Based Similarity Searches. *J. Med. Chem.* 2007, 50 (11), 2708–2717. <https://doi.org/10.1021/jm061259g>.
20. An, S.; Fu, L. Small-Molecule PROTACs: An Emerging and Promising Approach for the Development of Targeted Therapy Drugs. *EBioMedicine* 2018, 36, 553–562. <https://doi.org/10.1016/j.ebiom.2018.09.005>.
21. Boettcher, M.; McManus, M. T. Choosing the Right Tool for the Job: RNAi, TALEN or CRISPR. *Mol. Cell* 2015, 58 (4), 575–585. <https://doi.org/10.1016/j.molcel.2015.04.028>.
22. Giuliano, C. J.; Lin, A.; Girish, V.; Sheltzer, J. M. Generating Single Cell-Derived Knockout Clones in Mammalian Cells with CRISPR/Cas9. *Curr. Protoc. Mol. Biol.* 2019, 128 (1), e100. <https://doi.org/10.1002/cpmb.100>.
23. Gupta, S.; Schoer, R.; Egan, J.; Hannon, G.; Mittal, V. Inducible, Reversible, and Stable RNA Interference in Mammalian Cells. *Proc. Natl. Acad. Sci. U. S. A.* 2004, 101, 1927–1932. <https://doi.org/10.1073/pnas.0306111101>.
24. Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* 2013, 152 (5), 1173–1183. <https://doi.org/10.1016/j.cell.2013.02.022>.
25. Gier, R. A.; Budinich, K. A.; Evitt, N. H.; Cao, Z.; Freilich, E. S.; Chen, Q.; Qi, J.; Lan, Y.; Kohli, R. M.; Shi, J. High-Performance CRISPR-Cas12a Genome Editing for Combinatorial Genetic Screening. *Nat. Commun.* 2020, 11 (1),



3455. <https://doi.org/10.1038/s41467-020-17209-1>.
26. Wessels, H.-H.; Méndez-Mancilla, A.; Guo, X.; Legut, M.; Daniloski, Z.; Sanjana, N. E. Massively Parallel Cas13 Screens Reveal Principles for Guide RNA Design. *Nat. Biotechnol.* 2020, 38 (6), 722–727. <https://doi.org/10.1038/s41587-020-0456-9>.
27. Cancer treatment – Mayo Clinic <https://www.mayoclinic.org/tests-procedures/cancer-treatment/about/pac-20393344> (accessed Aug 17, 2020).
28. Ch, W.; Kw, S.; Aw, L. Estimation of clinical trial success rates and related parameters <https://pubmed.ncbi.nlm.nih.gov/29394327/> (accessed Sep 15, 2020). <https://doi.org/10.1093/biostatistics/kxx069>.
29. Aparoy, P.; Kumar Reddy, K.; Reddanna, P. Structure and Ligand Based Drug Design Strategies in the Development of Novel 5-LOX Inhibitors. *Curr. Med. Chem.* 2012, 19 (22), 3763–3778. <https://doi.org/10.2174/092986712801661112>.
30. Kroemer, R. T. Structure-Based Drug Design: Docking and Scoring. *Curr. Protein Pept. Sci.* 2007, 8 (4), 312–328. <https://doi.org/10.2174/138920307781369382>.
31. AutoDock — AutoDock <http://autodock.scripps.edu/> (accessed Aug 23, 2020).
32. Acharya, C.; Coop, A.; Polli, J. E.; MacKerell, A. D. Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach. *Curr. Comput. Aided Drug Des.* 2011, 7 (1), 10–22. <https://doi.org/10.1093/biostatistics/kxx069>.
33. Grover, M.; Singh, B.; Bakshi, M.; Singh, S. Quantitative Structure–Property Relationships in Pharmaceutical Research – Part 1. *Pharm. Sci. Technol. Today* 2000, 3 (1), 28–35. [https://doi.org/10.1016/S1461-5347\(99\)00214-X](https://doi.org/10.1016/S1461-5347(99)00214-X).
34. Zhao, X.; Chen, M.; Huang, B.; Ji, H.; Yuan, M. Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) Studies on A1A-Adrenergic Receptor Antagonists Based on Pharmacophore Molecular Alignment. *Int. J. Mol. Sci.* 2011, 12 (10), 7022–7037. <https://doi.org/10.3390/ijms12107022>.
35. Fogel, D. B. Factors Associated with Clinical Trials That Fail and Opportunities for Improving the Likelihood of Success: A Review. *Contemp. Clin. Trials Commun.* 2018, 11, 156–164. <https://doi.org/10.1016/j.conctc.2018.08.001>.
36. Cancer Drug Development: How Therapies Move Through the Pipeline <https://www.cancertherapyadvisor.com/home/tools/fact-sheets/how-cancer-drug-development-pipeline-treatment-patient-fact-sheet/> (accessed Aug 18, 2020).

## ■ Author

Ojas Gupta is a current sophomore at Lynbrook High School in Santa Clara, California. His interest in science and genetics mainly through school competitions such as Science Olympiad led him into the wonderful world of bioinformatics and cancer research, where he was lucky enough to meet Ann Lin, a pioneer in the field. He hopes this paper can help educate the general population on the methodology of cancer research.