IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Spatial Transcriptomics and Adaptive Multi-Modal Ensemble Encoding: Early Metastasis Profiling and Targeting

Nidhi Yadalam

Jesuit High School, 9000 SW Beaverton Hillsdale Hwy, Portland, Oregon, 97229, USA; nidhiyadalam@gmail.com
Mentor: Lara Shamieh

ABSTRACT: Lung cancer remains the leading cause of cancer-related mortality, primarily due to its high metastatic potential and the difficulty of early metastasis detection. Current clinical approaches rely on low-sensitivity assessments and can also cause complications. To address these challenges, this project introduces a computational framework that predicts metastatic potential directly from the primary tumor site, thereby eliminating the need for secondary site biopsies and enabling earlier intervention. This framework leverages spatial transcriptomics, a cutting-edge technology that maps tumor tissue, capturing cellular interactions that are missed by traditional methods. By analyzing the spatial and molecular features of primary tumors, this system identifies high-risk tumor regions and key metastatic drivers, integrating neural networks, autoencoders, and unsupervised clustering. In-silico validation aligned the received outputs with known signatures from open datasets. Building on these findings, a deep learning-guided lipid nanoparticle optimization pipeline was developed to design drug carriers for siRNA, aiming to silence the inhibition of genetic pathways. Further supervised analysis and validation revealed the heightened properties of the enhanced carriers. This study bridges AI-driven metastasis prediction and treatment with precision nanomedicine for metastatic cancer at its earliest stages by shifting from late-stage intervention to proactive, gene-targeted suppression.

KEYWORDS: Computational Biology and Bioinformatics, Computational Pharmacology, Multimodal Ensemble Learning, Spatial Transcriptomics, Metastasis Prediction and Treatment.

Introduction

Bronchogenic Carcinoma, or Lung Cancer (LC), is deemed to be the leading cause of death by cancer in the United States, with 234,580 new cases of LC and 125,070 deaths just in 2024. LC accounts for 20% of all cancer-related deaths, accounting for more than breast, colon, and prostate cancers combined.[1] LC is a disease caused by the uncontrollable division of mutated cells, which can cause abnormalities, including tumor masses.[2] A common cause involves repeated exposure to carcinogens, which can lead to an abnormal growth pattern of the epithelial cells lining the lung pathways, resulting in lung cancer if left untreated.[3] There are two main categories of LC: Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC).[4] SCLC is a cancer that grows and spreads at an exceptionally high rate, with a centrally localized lung mass or bulky thoracic lymph node involvement; the main subtypes of NSCLC are adenocarcinoma, squamous cell carcinoma, and large cell carcinoma.[5]

Lung Cancer Metastasis (LCM) refers to the dynamic process by which the cancer cells escape the primary tumor, migrate, and colonize distant organs. The process involves several complex mechanisms and commonly affects specific organs, contributing to the lethality of most forms of lung cancer. The spread of lung cancer cells is influenced by various factors, including interactions within the tumor microenvironment (TME) and the process of epithelial-mesenchymal transition (EMT). EMT enables cancer cells to acquire migratory and invasive properties.[6] Lung cancer can metastasize to almost any part of the body, but the most common locations for metastasis include the liver, bones, brain, and adrenal glands (Figure 1a).[7] The process begins with angiogenesis, where new blood vessels form to supply nutrients and oxygen to the tumor. This supports rapid tumor growth and provides a pathway for cancer cells to escape the primary site.[8] Through EMT, lung cancer cells lose their epithelial characteristics, such as tight cell-cell adhesion and polarity, and gain mesenchymal features such as increased motility and invasiveness. This transition enables cancer cells to invade the surrounding tissue, penetrate the extracellular matrix, and enter the bloodstream or lymphatic system, a process known as intravasation. Once in circulation, the cancer cells survive various stresses (such as the immune surveillance system) by developing resistance mechanisms.[9] To establish secondary tumors, the cells undergo a mesenchymal-epithelial transition (MET), reverting to an epithelial phenotype that supports proliferation and colonization (Figure 1b).[10]

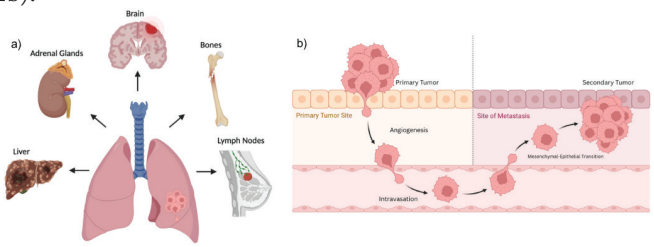

Figure 1: Biological context for metastasis. a) Anatomical illustration of primary lung cancer and its most frequent metastatic destinations. b) Schematic representation of the general metastatic cascade: angiogenesis provides vascular access, followed by intravasation into the bloodstream, and eventual colonization of distant tissues through mesenchymal-epithelial transition.

DOI: 10.36838/IJHSR85.1

The primary challenge in managing metastatic lung cancer is the difficulty in early detection and precise identification of metastatic spread. Unlike localized tumors that can be surgically removed or targeted with localized radiation, metastases often involve multiple distant sites, requiring systemic treatment approaches that are less effective in advanced stages.[11] Current diagnostic techniques rely on imaging scans (CT, PET, MRI) and biopsies from suspected secondary sites.[12] However, these methods have several invasive limitations – posing risks of complications, infections, and additional stress for patients.[13] Furthermore, metastatic tumors are often heterogeneous, meaning that biopsies from one part of the metastatic site may not fully capture the molecular characteristics of the entire metastatic population.[14] Many metastases also remain undetectable until they reach a clinically significant size, by which point treatment options become limited. Some micrometastases are too small to be detected through conventional imaging, leading to missed diagnoses and delays in targeted treatment.[15]

Recent technological advancements have revolutionized the ability to analyze gene expression while preserving spatial characteristics, providing valuable insights into the interactions within the TME, unlike traditional bulk RNA-sequencing methods, which average gene expression across an entire tissue. Spatially resolved gene expression, also known as spatial transcriptomics (ST), is a technique that enables scientists to study gene expression within the physical context of intact tissue.[16] This approach provides valuable insights into how cells interact with one another and their environment, which is critical for understanding cell functions and tissue dynamics.[17] By analyzing intact tissue, spatial transcriptomics retains critical spatial information, offering a more holistic view of cell biology.[18]

However, despite its potential, the vast, high-dimensional datasets generated by spatial transcriptomics remain difficult to analyze and interpret. Machine Learning (ML) has emerged as a transformative approach to addressing technical challenges and analyzing patterns.[19] Utilizing deep learning allows for an understanding of the complex ST organization of tumor cells in their microenvironment.

RNA-based therapeutics, particularly small interfering RNA (siRNA), offer a highly specific approach to cancer treatment by silencing genes that drive metastasis at the translational level.[20] Unlike traditional treatments, siRNA can target specific oncogenes or metastatic regulators, shutting down pathways responsible for tumor invasion, angiogenesis, and immune suppression.[21] This precision therapy has the potential to revolutionize cancer treatment by directly inhibiting metastatic progression rather than broadly targeting cell division. However, siRNA delivery presents a major challenge. Naked siRNA is highly unstable, rapidly degrading in the bloodstream and failing to reach tumor sites effectively.[22] Additionally, siRNA molecules must overcome multiple biological barriers, including immune clearance, poor cellular uptake, and endosomal degradation, before they can exert their therapeutic effects.[23] To address these challenges, Lipid Nanoparticles (LNPs) have emerged as a promising delivery platform for RNA-based therapies, offering enhanced stability, targeted delivery, and efficient cellular uptake (Figure 15).[24]

LNPs are engineered drug delivery systems designed to encapsulate and protect siRNA, improving its stability and ensuring efficient transport to tumor cells (Figure 2a). LNPs are widely used in gene therapy, immunotherapy, and vaccine development, proving their effectiveness in delivering nucleic acids with high efficiency. The core of an LNP typically contains ionizable cationic lipids that bind to and condense the negatively charged RNA, surrounded by structural lipids such as cholesterol and phospholipids that enhance membrane stability. A polyethylene glycol (PEG)-lipid is incorporated on the outer surface to improve circulation time and reduce immune clearance.[25] Upon reaching the target tissue, the acidic environment of endosomes activates the ionizable lipids, promoting endosomal escape and releasing the siRNA into the cytoplasm, where it silences gene expression by degrading target mRNA (Figure 2b).[26] Compared to viral delivery systems, LNPs offer greater biocompatibility, lower immunogenicity, and scalable production—making them ideal for siRNA-based cancer treatment.

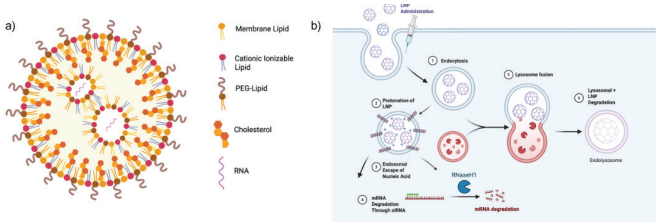

Figure 2: RNA-loaded lipid nanoparticles and delivery mechanism. a) Schematic of an LNP showing its major components: ionizable lipids, cholesterol, membrane lipids, PEG-lipids, and encapsulated RNA. b) Mechanism of siRNA delivery via LNPs: cellular uptake, endosomal escape triggered by pH-dependent activation of ionizable lipids, and gene silencing through RNA-induced silencing complex (RISC) activation. Modified from ref 36.

Despite their potential, optimizing LNPs for siRNA delivery remains a complex challenge. The ideal LNP formulation must strike a balance between stability, size, charge, and encapsulation efficiency, while ensuring precise targeting of metastatic cells. Current research focuses on improving LNP composition, surface modifications, and ligand-based targeting to maximize therapeutic efficacy.[27] By leveraging AI-driven optimization techniques, LNP formulations can be tailored for metastatic inhibition, marking a significant step toward next-generation cancer therapeutics.

Recent years have seen rapid progress in integrating spatial transcriptomics with advanced computational methods – particularly deep learning and multimodal learning – to dissect tumor and microenvironment architecture in cancer types. For example, a 2024 study by Zuani *et al.* demonstrated that combining single-cell and spatial transcriptomics in NSCLC enables a high-resolution "map" of tumor signaling modules that drive cellular reprograming of macrophages.[28] Furthermore, Wang *et al.* utilized a combination of scRNA-seq and ST in human multiple primary lung cancer lesions to map a novel epithelial sub-population and reveal inter-lesion spatial heterogeneity.[29] On the computational front, deep-learning frameworks have begun to integrate histopathological images

with ST maps. For example, Zhao *et al*. introduced a model named GIST, which fuses histology and spatial transcriptome data to predict cell-type distributions and latent spatial signatures.[30] Another model, TransST, uses transfer-learning and spatial factor modelling to improve deconvolution of ST datasets and reveal biologically meaningful areas of interest in tumor samplings.[31]

In parallel, deep, generative learning is rapidly accelerating the optimization of LNPs and ionizable lipids for nucleic acid delivery, critical for siRNA and mRNA therapies. The AGILE platform, developed by Xu *et al.*, uses a combination of deep learning neural networks and combinatorial chemistry to determine the best possible ionizable lipids for specific cell types.[32] Similarly, Wang *et al*. applied AI-driven virtual screening with LightGBM to predict lipid pKa and delivery efficiency, yielding novel candidates outperforming traditional lipid candidates such as DLin-MC3-DMA.[33] Comprehensive reviews, such as Dorsey *et al*., highlight challenges and opportunities in ML-guided formulation, including data sparsity, feature engineering, and multi-objective optimization of particle size, charge, and targeting.[34]

This study aims to utilize spatial transcriptomics and machine learning to analyze the molecular and spatial characteristics of primary lung tumors. This method will enable the early identification of tumors that are most likely to metastasize by analyzing the gene expression patterns associated with metastasis, the spatial organization of tumor cells, and the TME interactions. Using the encodings learned from the metastasis profiling, the pipeline also aims to provide an optimized nucleic acid-based therapy to inhibit metastasis. This will allow clinicians to intervene earlier and more precisely, tailoring treatments to suppress metastatic progression before it becomes widespread.

Methods

Data Preprocessing:

The spatial transcriptomics datasets used in this study were obtained from the Visium Spatial Gene Expression platform, provided by 10x Genomics.[35] The Visium platform enables high-resolution spatial transcriptomic profiling, capturing gene expression data alongside spatial coordinates within tissue sections. The datasets include spatial coordinates for tissue spots, corresponding gene expression profiles, and cluster annotations derived from prior analyses.

To prepare the data for downstream analysis, spatial coordinates were normalized to maintain consistency across samples. High-variance genes were selected from the gene expression data, as they are more likely to capture biologically significant patterns. Modified from the SpatialPCA reduction methodology proposed by Shang and Zhou, dimensionality reduction was then applied using Principal Component Analysis (PCA).[36] This retains the top 50 principal components to reduce noise and computational complexity while preserving key features of the data. Finally, the spatial coordinates, transcriptomic data, and cluster information were merged into a unified dataset, creating a comprehensive foundation for subsequent analysis, including clustering and spatial modeling. This preprocessing pipeline ensures that the high-dimensional data is optimized for accurate and robust exploration of spatial gene expression patterns.

Graph Neural Network for Feature Learning:

The spatial and transcriptomic data are processed using a graph neural network (GNN), which is designed to effectively capture the relationships between spatially adjacent regions and their gene expression profiles. The GNN operates on a graph structure where each spatial spot or cell is treated as a node, and edges represent spatial proximity. The input to the GNN consists of a graph where nodes are characterized by the PCA-reduced gene expression data and their spatial coordinates. The graph edges are constructed based on a k-nearest neighbor (k-NN) algorithm applied to the spatial coordinates, ensuring that each node is connected to its most relevant neighbors.[37] Edge weights are computed using a Gaussian kernel on spatial distances, allowing closer neighbors to have a stronger influence while reducing noise from distant nodes.

The GNN is composed of multiple graph convolutional layers designed to propagate and aggregate information across the graph.[38] Each layer updates the features of a node by aggregating the features of its neighbors, weighted by the edge connections. This equation is expressed mathematically as:

$$h_i^{(l+1)} = \sigma\left(W^l \sum_{j \in N(i)}^{n} \frac{h_j^{(l)}}{|N(i)|} + b^l \right)$$

Where:

$h_i^{(l)}$ is the feature vector of the node i at layer l.

W^l is the trainable weight ma dtrix at layer l.

$N(i)$ represents the neighbors at the node i.

b^l is the bias term at the layer l.

σ is a non-linear activation function, such as the rectified linear unit (ReLU).

The aggregation function combines information from neighboring nodes, ensuring that each node's updated representation incorporates both its local features and its spatial context. After passing through the GNN, each node is represented by a high-dimensional embedding that captures both the spatial relationships of the tissue and the transcriptomic variation. These refined features are then passed into the autoencoder for dimensionality reduction and further representation learning in the latent space. By propagating and aggregating information across spatially connected nodes, the GNN enables the model to capture the subtle spatial dependencies and gene expression patterns crucial for tasks such as clustering and domain identification.

Latent Space Representation:

The autoencoder plays a pivotal role in learning meaningful latent space representations by processing data through its input, bottleneck, and output layers.[39] The input to the autoencoder is derived from the encodings of the GNN, which combines PCA-reduced gene expression data and the corresponding spatial embeddings. The latent space, or bottleneck,

DOI: 10.36838/IJHSR85.1

is the most critical component of the autoencoder architecture. This layer reduces the input to a compact 10-dimensional representation, forcing the network to focus on the most essential features of the data. The latent space effectively captures a distilled version of the spatial transcriptomic data, encoding both the spatial relationships and the most significant patterns in gene expression. This compact embedding serves as a versatile representation, enabling tasks like clustering to uncover meaningful spatial domains or regions with shared biological properties. The process of distillation in the latent space also ensures that noise and irrelevant features are filtered out, leaving only the core biological signals for downstream analysis. The output of the autoencoder seeks to reconstruct the input as closely as possible. Starting from the 10-dimensional latent space, the decoder progressively transforms the reduced representation back into the original input dimensionality. The accuracy of the reconstruction is measured using the mean squared error (MSE) loss function, which quantifies the difference between the original input and the reconstructed output. Minimizing this loss ensures that the autoencoder has constructively captured the most critical features in the latent space while preserving the integrity of the data during reconstruction. By training the autoencoder on this task, the latent space representation becomes highly expressive, capturing both the spatial and transcriptomic relationships present in the input.

Clustering:

Clustering is a crucial step in analyzing spatial transcriptomic data, as it enables the identification of biologically meaningful patterns and spatial domains within the tissue. In this study, k-means clustering was applied to the learned 10-dimensional latent representations generated by the autoencoder to group data points into distinct clusters based on their feature similarity. The algorithm begins with the initialization step, where cluster centroids are randomly placed in the latent space, as shown in previous studies.[40] Each centroid represents the potential center of a cluster.[41] In the assignment step, each data point, represented as a vector in the 10-dimensional latent space, is assigned to the nearest cluster centroid based on the Euclidean distance metric:

$$d\left(x_i,\ c_k\right) = \sqrt{\sum_{j=1}^{10}\left(x_{ij} - c_{kj}\right)^2}$$

Where:

$x_i = (x_{i1}, x_{i2}, \cdots, x_{i10})$ represents a datapoint in the latent space.

$c_i = (c_{k1}, c_{k2}, \cdots, c_{k10})$ represents a datapoint in the latent space.

$d\ (x_i, c_k)$ is the Euclidean distance between the datapoint x_i and the centroid c_k.

Each data point is assigned to the cluster k^* that minimizes the Euclidean distance:

$$k^* = \underset{k}{\operatorname{argmin}}\, d\left(x_i,\ c_k\right)$$

This ensures that data points with similar gene expression and spatial characteristics, as encoded in the latent space, are grouped. The recalculation of the centroid iteratively occurs until a solution is identified where the clusters are well-defined and separated.

By clustering in this reduced-dimensional space, k-means identifies distinct groups of spatially and transcriptionally similar regions, corresponding to biologically or pathologically distinct spatial domains. To facilitate interpretation of the clustering outcomes, spatial coordinates from the original data are used to create scatter plots where each data point is plotted according to its physical location within the tissue. The clusters assigned by k-means are visualized by coloring each point according to its cluster label. Once clusters are identified, they are visualized in two dimensions. Spatial coordinates are then color-coded by cluster labels, allowing the mapping of clusters back onto the tissue to reveal their spatial distribution and biological significance.

To validate the biological relevance of the identified clusters, Gene Set Enrichment Analysis (GSEA) is performed. This analysis determines whether specific gene sets, such as those associated with metastatic pathways, are enriched within each cluster. For instance, gene sets related to epithelial-mesenchymal transition (EMT), cellular invasion, migration, proliferation, or angiogenesis can reveal clusters likely representing metastatic regions. By linking clusters with well-documented metastatic pathways, GSEA provides a critical layer of validation, enabling researchers to hypothesize about the role of each cluster in metastatic progression.

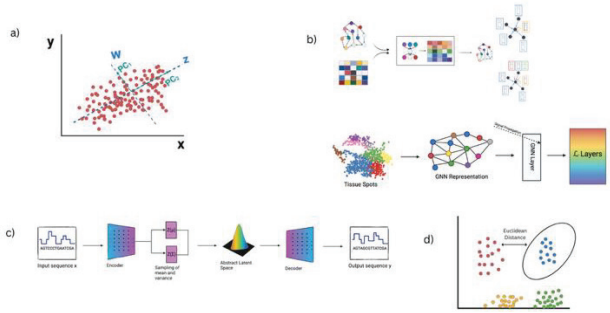

Figure 3: Pictorial diagram of the ensemble multi-modal machine learning pipeline. a) Principal Component Analysis (PCA) reduces dimensionality of high-throughput transcriptomic data. b) Graph Neural Networks (GNNs) capture spatial and transcriptional relationships between tissue regions. c) A variational autoencoder learns abstract latent features from gene expression profiles. d) K-means clustering is applied in latent space to group biologically relevant tissue domains based on Euclidean distance.

Metastasis Risk Score (Supervised Post-Analysis):

After identifying clusters potentially associated with metastatic behavior, a Metastasis Risk Score is calculated for each spatial location within the tissue. The computation of the Metastasis Risk Score considers three primary components: gene expression signatures, spatial context, and cluster membership.

The first component, gene expression signatures, involves evaluating the expression levels of known genes associated with metastasis. Regions with elevated expression of these genes are considered to have a higher likelihood of exhibiting metastatic behavior. The second component, spatial context, assesses the proximity of a spatial location to the primary tumor and evaluates the spatial patterns of gene expression. For instance, locations closer to the primary tumor or showing spatial char-

acteristics consistent with metastatic progression may receive higher scores. Lastly, cluster membership is incorporated by assigning higher weights to clusters identified as enriched for metastatic pathways through GSEA, linking the unsupervised clustering results to the risk-scoring process.

To refine the Metastasis Risk Score, supervised machine learning models such as Support Vector Machines (SVMs) and Random Forests (RFs) are employed. These models take input features derived from the latent space embeddings, expression levels of metastasis-associated genes, spatial proximity to high-risk clusters, and pathway enrichment scores. By training on labeled datasets where metastatic status is known, these models learn to predict the likelihood of metastasis at each spatial location. Incorporating known metastatic markers as features further enhances the accuracy of the risk predictions while maintaining biological interpretability.

Unsupervised Therapy Design:

The Variational Autoencoder (VAE)-Bayesian inference method is implemented in this study because it allows for a continuous, molecule-based algorithm that can derive features from its own latent space, inspired by the methodology from Ochiai *et al.*[42] This method is modified from the chemical LNP optimization methodology proposed by Nidhi Yadalam in 2024 for cystic fibrosis therapeutics.[43] As the input is a valid Simplified Molecular-Input Line-Entry System (SMILES) entry of a combinatorially formulated LNP (composed of the four lipids as described previously), the VAE traverses through its encoder/decoder network, recognizing the principal components of the entry.

The primary chemical formulation system is manned by principles of iterative, combinatorial chemistry to initiate different syntheses of LNPs. After manually retrieving various cationic ionizable lipids, cholesterol, phospholipids, and PEG-lipids, the client class iterates through the database, identifying the SMILES input for each compound. The canonical smiles were manually inputted at the beginning for easy access. The code then generates combinatorial libraries of molecules by enumerating possible combinations of R-groups on the scaffold of cationic ionizable lipids. The composition finalizes with a large database of LNPs.

The encoder portion for the VAE is made based on Relational Graph Convolutional Networks (R-CGN). The main inputs for the encoder are the adjacency and feature matrices. After the relational convolutions, the dimensionality of the graph is then further reduced from 2D to 1D, allowing the molecule to be easily represented for random selection later. However, the 2D dimensionality is retained, allowing it to represent the latent space. In the latent chemical space, the features, or properties, of the compounds were reduced to lower dimensionality and then optimized using Gaussian properties and the loss function. The custom loss used in the VAE function consists of two terms: a reconstruction loss and a KL divergence loss. The reconstruction loss term measures how well the model reconstructs the input data, while the KL divergence term encourages the learned latent space to resemble a predefined prior distribution.

The decoder reconstructs the primary input SMILES from the latent space. After defining the latent space input, densely connected layers are applied inside the latent space to learn a nonlinear mapping from the latent space representation to the adjacency matrix and feature matrix. Therefore, the generated outputs capture meaningful graph structures and node features while mitigating the risk of overfitting. The decoder's dense layers are then mapped to a continuous adjacency tensor and reshaped to match the specified adjacency shape, generating a representation of the graph's adjacency matrix in the SMILES format.

Results and Discussion

Overview of Identified Clusters:

Unsupervised clustering methods, including k-means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN), were applied to the latent space generated by the autoencoder. These methods identified distinct spatial clusters representing biologically meaningful groupings based on gene expression patterns. Scatterplots of spatial coordinates, colored by cluster labels, demonstrated clear separation between tissue regions, with clusters exhibiting high variability in gene expression predominantly located at the tumor's invasive front or distant metastatic sites. Figure 4a and Figure 4b show the groupings of spatial coordinates and clusters, as well as the transcriptome and heatmap of genes for the Human Lung Cancer FFPE tissue tumor block.

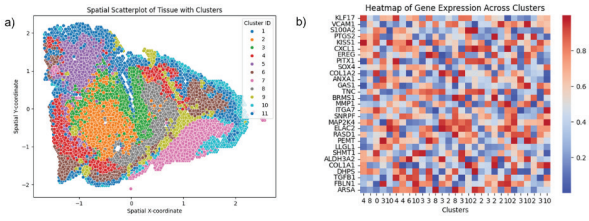

Figure 4: Spatial heterogeneity of gene expression. a) Spatial scatterplot showing transcriptionally distinct clusters across tissue regions, each color-coded to indicate unique spatial domains. b) Heatmap displaying gene expression variation across clusters, highlighting differential expression of metastasis-related genes.

Metastatic Pathways Identified:

Pathway enrichment analysis of metastatic clusters revealed significant associations with key metastasis-related pathways for each specific dataset. Clusters demonstrated strong enrichment for EMT, as evidenced by the upregulation of genes such as SNAI1, ZEB1, and TWIST1, indicating enhanced cell mobility and invasiveness. This is demonstrated in publications that support the EMT signaling pathway, as mediated by the most famous EMT-related cytokine, transforming growth factor β (TGF-β).[44] Additionally, angiogenesis pathways exhibited elevated expression of genes such as VEGFA, which has been shown to bind with receptors such as FT1 and KDR — supporting vascular remodeling and tumor expansion.[45] Immune evasion pathways were also enriched, marked by upregulation of MHC class II genes in the absence of co-stimulatory signals, enabling the tumor to suppress immune responses through regulatory T cell induction.[46,47] ECM modeling was observed through the overexpression of matrix

DOI: 10.36838/IJHSR85.1

metalloproteinases (e.g., MMP1, MMP9), which correlated with tissue invasion and metastasis.[48] All the different metastatic capabilities were discovered through the various tests of the individual datasets and compared with existing literature, with certain biochemical integrations shown (Figure 5). Dynamic comparisons with metastasis-related gene sets from TCGA and GEO datasets confirmed the robustness of these clustering results. A real-time correlation with the published cancer datasets was also run to ensure the practicality of the clustering results and the pathway likelihood (Figure 6a).

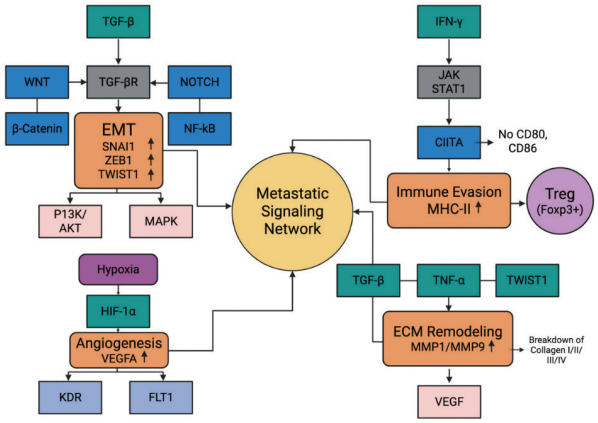

Figure 5: Integrated signaling pathways in lung cancer metastasis. Key pathways contributing to metastasis, including EMT, immune evasion, angiogenesis, and ECM remodeling, are shown converging on a central metastatic signaling network. Arrows indicate activation or upregulation of downstream effectors.

Gene Set Enrichment Analysis:

Gene Set Enrichment Analysis (GSEA) further validated the enrichment of metastatic pathways within the identified clusters. EMT and invasion pathways displayed normalized enrichment scores (NES) greater than 3, with false discovery rates (FDR) of less than 0.05. A high NES indicates strong and statistically significant enrichment of these pathways within the metastatic clusters, suggesting that genes involved in EMT and invasion are highly activated compared to background gene expression levels. The low FDR ensures that these findings are not due to random chance, increasing confidence in the biological relevance of these pathways to metastasis. This provides robust evidence that EMT and invasion processes play a critical role in driving metastatic progression within the identified clusters, reinforcing the validity of the clustering and enrichment analysis. Additionally, immune suppression and angiogenesis pathways exhibited consistent enrichment across multiple metastatic clusters. Enrichment maps visualizing these findings revealed cluster-specific pathway activation, underscoring the heterogeneity in metastatic mechanisms across different tissue regions (Figure 6a and 6b).

Cross-Validation with External Datasets:

External validation was conducted using publicly available transcriptomic datasets from The Cancer Genome Atlas (TCGA) from the Genomic Data Commons (GDC) Portal[49] and Gene Expression Omnibus (GEO)[50] to assess the robustness of the identified metastatic clusters. To ensure consistency, metastatic signatures were extracted from independent patient-derived tumor samples in these datasets and compared against the gene expression profiles from the identified clusters. The enriched pathways are known to be statistically and biologically significant, with the False Discovery Rate (FDR) values of less than 0.05 (Figure 6c).

Validation was performed by applying the trained clustering model to these external datasets and computing classification performance metrics. The metastatic clusters demonstrated high predictive accuracy, with an average precision of 94.2% (TCGA) and 91.5% (GEO), recall values of 92.8% (TCGA) and 90.2% (GEO), and F1 scores of 93.5% (TCGA) and 90.8% (GEO) (Figure 6d). Additionally, a receiver operating characteristic (ROC) analysis was performed, revealing high area under the curve (AUC) scores of 0.9 (TCGA) and 0.89 (GEO). The receiver operating characteristic (ROC) curve (Figure 6e) illustrates that metastatic classification shows clear separation between true positive and false positive rates.

Metastasis Risk Scoring:

A Metastasis Risk Score was computed for each spatial region based on gene expression, spatial context, and cluster membership. High-risk regions, defined as those scoring in the top 20th percentile, exhibited a mean risk score of 0.87 ± 0.05, while low-risk regions scored significantly lower at 0.34 ± 0.07 ($p < 0.05$) (Figure 6f). The 20th percentile was chosen to capture the upper quantile of spatial regions exhibiting extreme metastatic gene-expression patterns while retaining sufficient statistical power for group comparisons. Cross-validation of the risk scores against clinical metastatic samples was performed using an independent dataset of histologically confirmed metastatic regions. The predictive accuracy of the risk scoring system was evaluated using sensitivity and specificity metrics, achieving a sensitivity of 92.8% and a specificity of 89.4%. A Spearman correlation analysis ($\rho = 0.81$) further validated the strong association between predicted high-risk regions and actual metastatic outcomes, reinforcing the robustness of the risk assessment framework. These results highlight the effectiveness of the risk-scoring system in identifying regions with high metastatic potential.

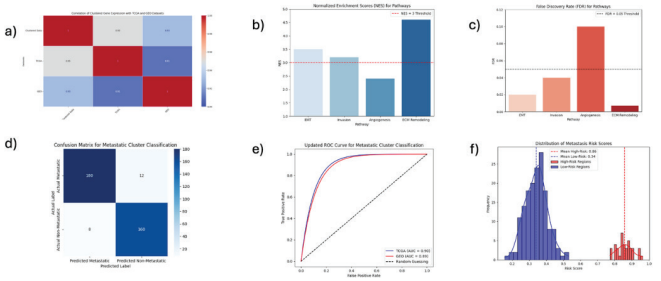

Figure 6: Validation of biological relevance and predictive strength of identified metastatic clusters. a) Clustered gene expression profiles show high correlation ($r \geq 0.91$) with TCGA and GEO datasets. b) Enrichment analysis reveals strong activation of metastasis-related pathways, all with NES ≥ 3. c) All enriched pathways show statistically significant FDR values (FDR < 0.05). d) The confusion matrix shows accurate classification of metastatic vs. non-metastatic clusters with minimal error. e) ROC curves for TCGA and GEO confirm strong model performance (AUC = 0.90 and 0.89). f) Metastasis risk scores show clear separation between high- and low-risk tissue regions.

Spatial Analysis of Metastatic Clusters:

Mapping of identified clusters onto spatial tissue coordinates demonstrated that metastatic clusters were predominantly located at the tumor's invasive front or in distant regions. Quantitative spatial analysis revealed that for tumor sections ranging from 400 to 1200 μm, metastatic clusters were positioned at an average distance of 600-950 μm from the tumor core, whereas low-risk clusters were placed within 250–450 μm of the tumor boundary. The spatial mapping (Figure 7b) illustrates that high-risk metastatic clusters (red) tend to be localized at the tumor periphery and distant regions. In contrast, low-risk clusters (blue) are concentrated near the tumor core. Most metastatic clusters exhibit a skewed distribution toward increased distances from the tumor core, reinforcing the notion that tumor cells preferentially disseminate away from the primary tumor mass. For the cluster distances to the vascular regions, the distances averaged at ~550 μm (Figure 7a), and for the cluster distances to the tumor core, the distances averaged at ~650 μm (Figure 7c).

Spatial autocorrelation analysis confirmed that metastatic clusters were non-randomly distributed (Moran's I = 0.67, $p < 0.03$), indicating strong spatial dependence of metastatic behavior. These findings suggest that metastatic clusters do not arise randomly within the tissue but instead follow predictable patterns of invasion towards vascularized and low-density extracellular matrix regions.

These findings were consistent with biological expectations of metastasis, particularly migration toward vascularized or low-density regions of the extracellular matrix. The statistical correlation between high-risk clusters and proximity to vasculature provided additional spatial validation supporting the identified clusters.

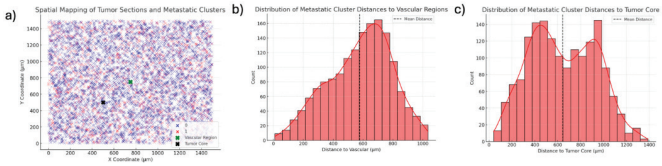

Figure 7: Spatial patterns associated with metastatic risk. a) Map of metastatic clusters overlaid on tissue coordinates, with high-risk (red) and low-risk (blue) areas, and annotated vascular regions and tumor core. b) Metastatic clusters are located closer to vascular regions on average (~550 μm), suggesting vascular-directed migration. c) Metastatic clusters show wider spatial spread from the tumor core (~650 μm mean distance), indicating dispersal beyond the primary tumor mass.

Unsupervised Therapy Analysis:

Following the development of the Variational Autoencoder (VAE) model and the exploration of its latent space to identify optimal molecular representations through decoding, an essential post-processing step ensures that only chemically valid SMILES strings are retained. Using RDKit modules, the model filters out invalid molecules by checking for structural feasibility, including bond integrity and atom valency. This step prevents the inclusion of synthetically impossible or unstable molecules. Beyond chemical validity, the model further evaluates the drug-likeness and pharmacological suitability of the generated lipid nanoparticles (LNPs) using two key metrics: Quantitative Estimate of Drug-likeness (QED) and Lipinski's Rule of Five. The QED score quantitatively assesses how "drug-like" a molecule is based on a composite of physicochemical properties. At the same time, Lipinski's rules provide guidelines (e.g., molecular weight < 500 Da, ≤5 hydrogen bond donors, ≤10 hydrogen bond acceptors, and logP ≤5) to gauge its oral bioavailability.

To further refine the selection, a supervised classification model is used to predict whether the newly generated LNP candidates are more optimized. This classifier integrates QED, Lipinski features, and additional molecular descriptors as inputs, enabling high-confidence predictions about the therapeutic viability of the generated molecules. The supervised classification model implemented a Random Forest Regressor, trained with built-in QED and Lipinski features for biomolecules and aggregates. The model received a 93% accuracy. This multilayered filtering process ensures that only chemically sound, biologically relevant, and pharmaceutically promising LNPs proceed to experimental validation.

Significance and Limitations:

This study advances the field by integrating spatial transcriptomics, graph neural networks, and variational autoencoding to create a unified pipeline capable of predicting metastatic risk and generating optimized siRNA-loaded lipid nanoparticles. Previous spatial transcriptomics studies mainly characterized tumor microenvironments,[16,18] but few have connected spatial gene expression with actionable therapeutic design. By leveraging spatially-aware ML, this work identifies high-risk metastatic regions directly from the primary tumor—addressing a major clinical gap, as metastasis confirmation typically requires invasive secondary-site biopsies.[13] The model also demonstrates strong performance when validated against TCGA and GEO datasets, supporting its biological robustness. Furthermore, coupling metastasis profiling with AI-driven nanoparticle optimization introduces a new computational framework for targeted RNA therapeutics.

However, several limitations must be acknowledged. The first includes the reliance on Visium's spot-level resolution, which restricts single-cell interpretability. The absence of wet-lab validation prevents full confirmation of metastatic behavior or LNP performance. The Metastasis Risk Score uses an arbitrary percentile threshold that may not generalize across tumor morphologies, and QED/Lipinski filters do not fully capture biological delivery constraints. Finally, proteomic, metabolic, and other biophysical factors could enhance metastasis qualities as well as transcriptomic signatures. Future work integrating multi-omic datasets, higher-resolution spatial platforms, and experimental validation will strengthen the translational potential of this framework.

■ Conclusion

Metastasis remains the greatest barrier to effective cancer treatment, responsible for over 90% of cancer-related deaths due to its late detection, lack of targeted therapies, and rapid resistance to conventional treatments. This study presents

DOI: 10.36838/IJHSR85.1

a first-of-its-kind computational framework that integrates AI-driven metastasis prediction with RNA-based precision therapy, addressing critical gaps in early detection and targeted intervention.

By leveraging spatial transcriptomics and machine learning, this work introduces an AI-driven metastasis prediction model that can identify high-risk tumor regions before they spread, offering a non-invasive alternative to traditional biopsy-based diagnostics. Beyond detection, this study pioneers the development of an AI-optimized lipid nanoparticle framework, designed to stabilize and deliver siRNA therapeutics to metastatic cells with high precision. In silico validation confirms that the optimized therapies exhibit heightened potential for efficacious delivery at precisely targeted locations, as historically verified.

This research bridges the gap between computational oncology and precision nanomedicine, showing that AI can not only predict metastasis but also guide the design of next-generation therapeutics to stop it at the molecular level. By replacing invasive biopsies with AI-driven spatial analysis and broad-spectrum chemotherapy with gene-targeted siRNA delivery, this study presents a transformative approach to treating metastatic cancer. Future advancements will focus on *in vitro/vivo* validation, expanded clinical applications, and further refinement of AI, paving the way for the real-world implementation of non-invasive, patient-specific metastasis therapies.

This work attempts to establish a new standard in cancer therapeutics, demonstrating that AI and RNA-based medicine can collaborate to detect, target, and treat metastasis with unprecedented precision, thereby pushing the boundaries of what is possible in oncology.

■ Acknowledgments

The results here are in whole or part based upon data generated by the TCGA and GEO Research Networks: https://www.cancer.gov/ccg/research/genome-sequencing/tcga and https://www.ncbi.nlm.nih.gov/geo/. Figures made with Canva, Python, and BioRender.

Thank you to Dr. Lara Shamieh for her mentorship and guidance throughout this research project.

Continuation from prior work done with unsupervised therapy optimization for lipid nanoparticles.

■ References

1. American Cancer Society. Lung Cancer Statistics | How Common is Lung Cancer? www.cancer.org. https://www.cancer.org/cancer/types/lung-cancer/about/key-statistics.html.
2. Cleveland Clinic. Lung Cancer. Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/4375-lung-cancer.
3. Siddiqui, F.; Siddiqui, A. H.; Vaqar, S. Lung Cancer. Nih.gov. https://www.ncbi.nlm.nih.gov/books/NBK482357/.
4. Dela Cruz, C. S.; Tanoue, L. T.; Matthay, R. A. Lung Cancer: Epidemiology, Etiology, and Prevention. Clinics in Chest Medicine 2011, 32 (4), 605–644. https://doi.org/10.1016/j.ccm.2011.09.001.
5. American Cancer Society. What Is Lung Cancer? www.cancer.org. https://www.cancer.org/cancer/types/lung-cancer/about/what-is.html.
6. Xie, S.; Wu, Z.; Qi, Y.; Wu, B.; Zhu, X. The Metastasizing Mechanisms of Lung Cancer: Recent Advances and Therapeutic Challenges. Biomedicine & Pharmacotherapy 2021, 138, 111450. https://doi.org/10.1016/j.biopha.2021.111450.
7. Popper, H. H. Progression and Metastasis of Lung Cancer. Cancer and Metastasis Reviews 2016, 35 (1), 75–91. https://doi.org/10.1007/s10555-016-9618-0.
8. Ngaha, T. Y. S.; Zhilenkova, A. V.; Essogmo, F. E.; Uchendu, I. K.; Abah, M. O.; Fossa, L. T.; Sangadzhieva, Z. D.; D. Sanikovich, V.; S. Rusanov, A.; N. Pirogova, Y.; Boroda, A.; Rozhkov, A.; Kemfang Ngowa, J. D.; N. Bagmet, L.; I. Sekacheva, M. Angiogenesis in Lung Cancer: Understanding the Roles of Growth Factors. Cancers 2023, 15 (18), 4648. https://doi.org/10.3390/cancers15184648.
9. Ribatti, D.; Tamma, R.; Annese, T. Epithelial-Mesenchymal Transition in Cancer: A Historical Overview. Translational Oncology 2020, 13 (6), 100773. https://doi.org/10.1016/j.tranon.2020.100773.
10. Kalluri, R.; Weinberg, R. A. The Basics of Epithelial-Mesenchymal Transition. Journal of Clinical Investigation 2009, 119 (6), 1420–1428. https://doi.org/10.1172/JCI39104.
11. Gerstberger, S.; Jiang, Q.; Ganesh, K. Metastasis. Cell 2023, 186 (8), 1564–1579. https://doi.org/10.1016/j.cell.2023.03.003.
12. Łukaszewski, B.; Nazar, J.; Goch, M.; Łukaszewska, M.; Stępiński, A.; Jurczyk, M. U. Diagnostic Methods for Detection of Bone Metastases. Contemporary Oncology 2017, 21 (2), 98–103. https://doi.org/10.5114/wo.2017.68617.
13. Hiroyasu Kameyama; Priya Dondapati; Simmons, R.; Leslie, M.; Langenheim, J. F.; Sun, Y.; Yi, M.; Rottschaefer, A.; Pathak, R.; Shreya Nuguri; Fung, K.-M.; Shirng-Wern Tsaih; Inna Chervoneva; Rui, H.; Tanaka, T. Needle Biopsy Accelerates Pro-Metastatic Changes and Systemic Dissemination in Breast Cancer: Implications for Mortality by Surgery Delay. Cell Reports Medicine 2023, 4 (12), 101330–101330. https://doi.org/10.1016/j.xcrm.2023.101330.
14. Gilson, P.; Merlin, J.-L.; Harlé, A. Deciphering Tumour Heterogeneity: From Tissue to Liquid Biopsy. Cancers 2022, 14 (6), 1384. https://doi.org/10.3390/cancers14061384.
15. Mao, X.; Mei, R.; Yu, S.; Shou, L.; Zhang, W.; Li, K.; Qiu, Z.; Xie, T.; Sui, X. Emerging Technologies for the Detection of Cancer Micrometastasis. Technology in Cancer Research & Treatment 2022, 21, 153303382211003-153303382211003. https://doi.org/10.1177/15330338221100355.
16. Marx, V. Method of the Year: Spatially Resolved Transcriptomics. Nature Methods 2021, 18 (1), 9–14. https://doi.org/10.1038/s41592-020-01033-y.
17. Du, J.; Yang, Y.; An, Z.; Zhang, M.; Fu, X.-H.; Huang, Z.; Yuan, Y.; Hou, J. Advances in Spatial Transcriptomics and Related Data Analysis Strategies. Journal of Translational Medicine 2023, 21 (1). https://doi.org/10.1186/s12967-023-04150-2.
18. Williams, C. G.; Lee, H. J.; Asatsuma, T.; Vento-Tormo, R.; Haque, A. An Introduction to Spatial Transcriptomics for Biomedical Research. Genome Medicine 2022, 14 (1). https://doi.org/10.1186/s13073-022-01075-1.
19. Nature Materials. Ascent of Machine Learning in Medicine. Nature Materials 2019, 18 (5), 407–407. https://doi.org/10.1038/s41563-019-0`0-1.
20. Li, S.-D.; Chono, S.; Huang, L. Efficient Oncogene Silencing and Metastasis Inhibition via Systemic Delivery of SiRNA. Molecular Therapy 2008, 16 (5), 942–946. https://doi.org/10.1038/mt.2008.51.
21. Ngamcherdtrakul, W.; Yantasee, W. SiRNA Therapeutics for Breast Cancer: Recent Efforts in Targeting Metastasis, Drug Resistance, and Immune Evasion. Translational research : the journal

of laboratory and clinical medicine 2019, 214, 105–120. https://doi.org/10.1016/j.trsl.2019.08.005.
22. Tatiparti, K.; Sau, S.; Kashaw, S.; Iyer, A. SiRNA Delivery Strategies: A Comprehensive Review of Recent Developments. Nanomaterials 2017, 7 (4), 77. https://doi.org/10.3390/nano7040077.
23. Zaidi, A.; Fatima, F.; Zaidi, A.; Zhou, D.; Deng, W.; Liu, S. Engineering SiRNA Therapeutics: Challenges and Strategies. Journal of Nanobiotechnology 2023, 21 (1). https://doi.org/10.1186/s12951-023-02147-z.
24. Delivery, a Landscape of Research Diversity and Advancement. ACS Nano 2021, 15 (11). https://doi.org/10.1021/acsnano.1c04996.
25. Hou, X.; Zaks, T.; Langer, R.; Dong, Y. Lipid Nanoparticles for MRNA Delivery. Nature Reviews Materials 2021, 6 (6), 1078–1094. https://doi.org/10.1038/s41578-021-00358-0.
26. Kalita, T.; Dezfouli, S. A.; Pandey, L. M.; Uludag, H. SiRNA Functionalized Lipid Nanoparticles (LNPs) in Management of Diseases. Pharmaceutics 2022, 14 (11), 2520. https://doi.org/10.3390/pharmaceutics14112520.
27. Wang, J.; Ding, Y.; Chong, K.; Cui, M.; Cao, Z.; Tang, C.; Tian, Z.; Hu, Y.; Zhao, Y.; Jiang, S. Recent Advances in Lipid Nanoparticles and Their Safety Concerns for MRNA Delivery. Vaccines 2024, 12 (10), 1148. https://doi.org/10.3390/vaccines12101148.
28. De Zuani, M.; Xue, H.; Park, J. S.; Dentro, S. C.; Seferbekova, Z.; Tessier, J.; Curras-Alonso, S.; Hadjipanayis, A.; Athanasiadis, E. I.; Gerstung, M.; Bayraktar, O.; Cvejic, A. Single-Cell and Spatial Transcriptomics Analysis of Non-Small Cell Lung Cancer. *Nature Communications* **2024**, *15* (1), 4388. https://doi.org/10.1038/s41467-024-48700-8.
29. Wang, Y.; Chen, D.; Liu, Y.; Shi, D.; Duan, C.; Li, J.; Shi, X.; Zhang, Y.; Yu, Z.; Sun, N.; Wang, W.; Ma, Y.; Xu, X.; Otkur, W.; Liu, X.; Xia, T.; Qi, H.; Piao, H.; Liu, H.-X. Multidirectional Characterization of Cellular Composition and Spatial Architecture in Human Multiple Primary Lung Cancers. *Cell Death & Disease* **2023**, *14* (7), 1–16. https://doi.org/10.1038/s41419-023-05992-w.
30. Ge, Y.; Leng, J.; Tang, Z.; Wang, K.; Kaicheng U; Zhang, S. M.; Han, S.; Zhang, Y.; Xiang, J.; Yang, S.; Liu, X.; Song, Y.; Wang, X.; Li, Y.; Zhao, J. Deep Learning-Enabled Integration of Histology and Transcriptomics for Analyzing Single-Cell Spatial Profiles. *Research* **2024**, *8*. https://doi.org/10.34133/research.0568.
31. Liu, S. S.; Wang, S.; Chen, Y.; Rustgi, Anil K; Yuan, M.; Hu, J. *TransST: Transfer Learning Embedded Spatial Factor Modeling of Spatial Transcriptomics Data*. arXiv.org. https://arxiv.org/abs/2504.12353 (accessed 2025-11-13).
32. Xu, Y.; Ma, S.; Cui, H.; Chen, J.; Xu, S.; Gong, F.; Golubovic, A.; Zhou, M.; Wang, K. C.; Varley, A.; Xing, R.; Wang, B.; Li, B. AGILE Platform: A Deep Learning Powered Approach to Accelerate LNP Development for MRNA Delivery. *Nature Communications* **2024**, *15* (1). https://doi.org/10.1038/s41467-024-50619-z.
33. Wang, W.; Chen, K.; Jiang, T.; Wu, Y.; Wu, Z.; Ying, H.; Yu, H.; Lu, J.; Lin, J.; Ouyang, D. Artificial Intelligence-Driven Rational Design of Ionizable Lipids for MRNA Delivery. *Nature Communications* **2024**, *15* (1). https://doi.org/10.1038/s41467-024-55072-6.
34. Dorsey, P. J.; Lau, C. L.; Chang, T.-C.; Doerschuk, P. C.; D'Addio, S. M. Review of Machine Learning for Lipid Nanoparticle Formulation and Process Development. *Journal of Pharmaceutical Sciences* **2024**, *113* (12), 3413–3433. https://doi.org/10.1016/j.xphs.2024.09.015.
35. 10x Genomics. Home Page. 10x Genomics. https://www.10xgenomics.com/.
36. Shang, L.; Zhou, X. Spatially Aware Dimension Reduction for Spatial Transcriptomics. Nature Communications 2022, 13 (1). https://doi.org/10.1038/s41467-022-34879-1.
37. Li, L.; Yang, W.; Bai, S.; Ma, Z. KNN-GNN: A Powerful Graph Neural Network Enhanced by Aggregating K-Nearest Neighbors in Common Subspace. Expert Systems with Applications 2024, 253, 124217–124217. https://doi.org/10.1016/j.eswa.2024.124217.
38. Anand, R. Math Behind Graph Neural Networks. Rishabh Anand. https://rish-16.github.io/posts/gnn-math/.
39. Kamal Berahmand; Fatemeh Daneshfar; Elaheh Sadat Salehi; Li, Y.; Xu, Y. Autoencoders and Their Applications in Machine Learning: A Survey. Artificial Intelligence Review 2024, 57 (2). https://doi.org/10.1007/s10462-023-10662-6.
40. Walker, T. Discovering Latent Clusters with K-means - Codezillas - Medium. Medium. https://medium.com/codezillas/discovering-latent-clusters-with-k-means-f1670a1320e8 (accessed 2025-06-11).
41. Heath, A. P.; Ferretti, V.; Agrawal, S.; An, M.; Angelakos, J. C.; Arya, R.; Bajari, R.; Bilal Baqar; Justin; Burt, J.; Catton, A.; Chan, B. F.; Chu, F.; Cullion, K.; Davidsen, T.; Do, P.-M.; Dompierre, C.; Ferguson, M. L.; Fitzsimons, M. S.; Ford, M. The NCI Genomic Data Commons. Nature Genetics 2021, 53 (3), 257–262. https://doi.org/10.1038/s41588-021-00791-5.
42. Ochiai, T.; Inukai, T.; Akiyama, M.; Furui, K.; Ohue, M.; Matsumori, N.; Inuki, S.; Uesugi, M.; Sunazuka, T.; Kikuchi, K.; Kakeya, H.; Sakakibara, Y. Variational Autoencoder-Based Chemical Latent Space for Large Molecular Structures with 3D Complexity. Communications Chemistry 2023, 6 (1), 1–14. https://doi.org/10.1038/s42004-023-01054-6.
43. Yadalam, N (2024). Lipophilicity-Based Genetic Delivery Formulation of Cystic Fibrosis Therapeutics via LNP-VACCO: Lipid Nanoparticle Variational Autoencoder-Guided Combinatorial-Chemistry Optimization (No. 13145). EasyChair.
44. Gonzalez, D. M.; Medici, D. Signaling Mechanisms of the Epithelial-Mesenchymal Transition. Science Signaling 2014, 7 (344), re8–re8. https://doi.org/10.1126/scisignal.2005189.
45. Shibuya, M. Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies. Genes & Cancer 2011, 2 (12), 1097–1105. https://doi.org/10.1177/1947601911423031.
46. Lei, P.-J.; Pereira, E. R.; Andersson, P.; Zohreh Amoozgar; Willem, J.; O'Melia, M. J.; Zhou, H.; Chatterjee, S.; Ho, W. W.; Posada, J. M.; Kumar, A. S.; Morita, S.; Menzel, L.; Chung, C.; Ilgin Ergin; Jones, D.; Huang, P.; Semir Beyaz; Padera, T. P. Cancer Cell Plasticity and MHC-II–Mediated Immune Tolerance Promote Breast Cancer Metastasis to Lymph Nodes. The Journal of Experimental Medicine 2023, 220 (9). https://doi.org/10.1084/jem.20221847.
47. Thibodeau, J.; Bourgeois-Daigneault, M.-C.; Lapointe, R. Targeting the MHC Class II Antigen Presentation Pathway in Cancer Immunotherapy. Oncoimmunology 2012, 1 (6), 908–916. https://doi.org/10.4161/onci.21205.
48. Niina Reunanen; VeliMatti Kähäri. Matrix Metalloproteinases in Cancer Cell Invasion. Nih.gov. https://www.ncbi.nlm.nih.gov/books/NBK6598/.
49. Heath, A. P.; Ferretti, V.; Agrawal, S.; An, M.; Angelakos, J. C.; Arya, R.; Bajari, R.; Bilal Baqar; Justin; Burt, J.; Catton, A.; Chan, B. F.; Chu, F.; Cullion, K.; Davidsen, T.; Do, P.-M.; Dompierre, C.; Ferguson, M. L.; Fitzsimons, M. S.; Ford, M. The NCI Genomic Data Commons. Nature Genetics 2021, 53 (3), 257–262. https://doi.org/10.1038/s41588-021-00791-5.
50. Barrett, T.; Wilhite, S. E.; Ledoux, P.; Evangelista, C.; Kim, I. F.; Tomashevsky, M.; Marshall, K. A.; Phillippy, K. H.; Sherman, P. M.; Holko, M.; Yefanov, A.; Lee, H.; Zhang, N.; Robertson, C. L.; Serova, N.; Davis, S.; Soboleva, A. NCBI GEO: Archive for Functional Genomics Data Sets—Update. Nucleic Acids Research 2012, 41 (D1), D991–D995. https://doi.org/10.1093/nar/gks1193.

DOI: 10.36838/IJHSR85.1

Author

Nidhi Yadalam is a senior at Jesuit High School in Portland with a deep interest in computational biology and the intersection of AI and medicine. Her research using machine learning in the biological field has been recognized at many international-level fairs and conferences, including ISEF. Outside the lab, she enjoys leading educational and charity-based outreach. Nidhi is also an accomplished Indian classical vocalist and teacher. She volunteers regularly at her local temple and memory care center, combining science, service, and creativity in everything she does.