

Advances in predicting omics profiles from imaging data

Alexa H. Beachum^{1,2}, Xue Xiao¹, Yuansheng Zhou¹, Qiwei Li³, Guanghua Xiao^{1,4,*}, Lin Xu^{1,5,*}

¹Quantitative Biomedical Research Center, Department of Health Data Science & Biostatistics, Peter O'Donnell Jr. School of Public Health, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390, United States

²Department of Statistics and Data Science, Southern Methodist University, 6425 Boaz Lane, Dallas, TX 75205, United States

³Department of Mathematical Sciences, The University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080, United States

⁴Department of Bioinformatics, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390, United States

⁵Department of Pediatrics, Division of Hematology/Oncology, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390, United States

*Corresponding authors. Lin Xu, Quantitative Biomedical Research Center, Department of Health Data Science & Biostatistics, Peter O'Donnell Jr. School of Public Health, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390, USA. Tel: (214)648-1403; Fax: (214) 648-1663; E-mail: Lin.Xu@UTSouthwestern.edu; Guanghua Xiao, Quantitative Biomedical Research Center, Department of Health Data Science & Biostatistics, Peter O'Donnell Jr. School of Public Health, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390, United States. Tel: (214) 648-4110; Fax: (214) 648-1663; E-mail: Guanghua.Xiao@UTSouthwestern.edu

Abstract

While traditional imaging techniques, such as histopathology, are often part of clinical workflows, molecular profiling remains more difficult to conduct and is less cost-effective. Thus, the prediction of molecular 'omics' data directly from imaging has emerged as an appealing alternative. While existing reviews have mentioned image-based prediction of biomarkers within specific disease contexts, this review provides a comprehensive overview of current methods that leverage imaging to predict (i) DNA-based aberrations, (ii) bulk transcriptomic profiles, (iii) single-cell transcriptomics, and (iv) spatial transcriptomics across disease contexts and imaging modalities. To address the complexity of these predictive tasks, we find that many studies employ cutting-edge deep learning strategies for image processing, feature extraction, feature aggregation, and downstream molecular prediction. In this review, we highlight the diverse applications of both deep learning-based and modern statistical frameworks designed for image-based omics prediction. The insights gleaned from these inferred molecular data have broad clinical relevance and will continue to improve our understanding of the relationships between molecular and visual features, paving the way for new diagnostic and therapeutic applications.

Keywords histology imaging, genomics, transcriptomics, deep learning

Introduction

In recent years, predictive models leveraging imaging and omics data have been developed to extract both molecular-level and clinical insights. While imaging data capture critical phenotypic characteristics, omics data deliver in-depth molecular insights. These two modalities offer complementary information; however, imaging is typically more cost-effective and accessible than omics profiling. As a result, the ability to accurately predict omics directly from imaging holds great promise as a practical and scalable surrogate for omics-based diagnosis and prognosis.

While many existing reviews summarize methods for image-omics integration and data fusion [1–5], these algorithms fundamentally differ from predictive frameworks designed to infer molecular information solely from images. Image-omics integration requires data from both modalities and typically combines features through concatenation, transformation, or model-based integration, often to enhance

performance on classification tasks such as identifying normal versus disease tissue. In contrast, models that predict omics profiles from imaging data alone aim to estimate underlying molecular states, producing more versatile outputs such as high-dimensional bulk transcriptomics data. It is important to note that image-omics fusion can still play a role within predictive pipelines, e.g. as an alignment mechanism during training, even when only imaging data are required at inference. These inferred transcriptomics profiles can then support a broad range of downstream analyses, including predicting responses to immune therapies [6] and visualizing spatial patterns of intratumor heterogeneity [7].

Meanwhile, several cancer-focused reviews have briefly discussed the prediction of known molecular markers from imaging data as part of surveys on artificial intelligence (AI)-based methods for cancer detection and therapy. However, these discussions are mostly limited to image-based prediction of DNA-based markers in specific cancer types, leaving other disease areas—such as cardiac defects,

Received: October 16, 2025. **Revised:** January 7, 2026. **Accepted:** January 31, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

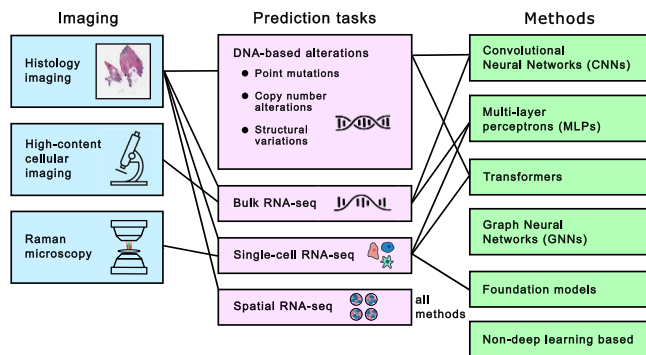


Figure 1 Schematic of end-to-end workflow for leveraging diverse imaging modalities to predict molecular and genomic features. (left) Input imaging data ranges from tissue-level histology to sub-cellular Raman microscopy. (center) These inputs are mapped to various prediction tasks, primarily focusing on transcriptomic profiles and DNA-based alterations. (right) A suite of computational methods, dominated by deep learning architectures, are employed to predict meaningful biological signals from the raw imagery. The connecting lines between imaging types to prediction tasks, and between prediction tasks to computational methods, represent the presence of published algorithms linking them. Notably, histology imaging is linked to all molecular prediction tasks, and spatial RNA-seq prediction is conducted using all listed computational methods.

neurological disorders, and various non-cancer conditions—largely unexplored. Further, recent advances using imaging data to predict omics have gone beyond biomarker prediction to infer gene expression profiles, particularly from single-cell and spatial transcriptomics datasets.

To fill in these gaps, we provide a comprehensive review of recent developments in predictive algorithms that leverage imaging data to infer molecular profiles across more than 20 disease types. We also cover a broad spectrum of omics modalities, including both DNA-based alterations [8–12] and gene expression profiles from bulk, single-cell, and spatial transcriptomics data [6, 13–16] (Fig. 1). To our knowledge, these directions have not yet been systematically reviewed.

In this review, we focus on commonly used imaging and omics data types that enable the prediction of molecular profiles from imaging information. The emergence of digital pathology has made it easier than ever to view, analyze, and share high-resolution whole-slide images (WSIs). One of the most widely used, publicly available resources for developing predictive models from WSIs is The Cancer Genome Atlas (TCGA), which provides an extensive collection of clinical annotations, molecular profiles, and whole-slide hematoxylin and eosin (H&E) stained images. Both traditional image processing techniques and modern computer vision methods have been critical in analyzing the complex patterns in H&E images. Notably, recent deep learning studies have found strong evidence linking morphological features in WSIs to specific molecular profiles [17, 18]. Although H&E images are the most prevalent form of imaging for omics prediction, other imaging modalities—such as high-content cellular imaging [19] and Raman microscopy [20]—are also discussed in this review.

Here, we highlight the methodological foundations of image-to-omics prediction methods as well as their broader applications. While many image-based predictive models are designed for direct clinical insights (e.g. survival, disease prognosis), our primary focus is on approaches that infer DNA-based aberrations, bulk transcriptomic profiles, as well as single-cell and spatial transcriptomics. These omics-level predictions not only reveal molecular mechanisms underlying

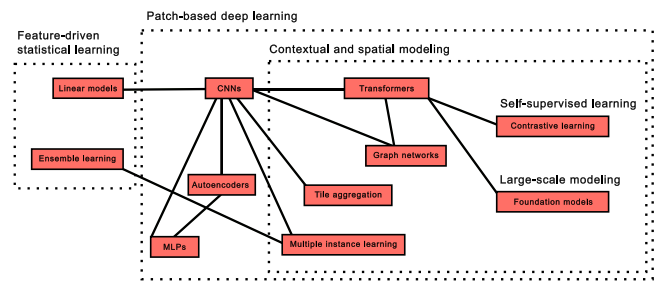


Figure 2 Taxonomy and interrelationships of methodological themes in image-based omics prediction. The hierarchical diagram organizes modeling approaches based on their conceptual lineage and data handling strategies, from feature-driven statistical learning to spatially aware deep learning. The connecting lines represent the presence of integrated frameworks that combine multiple approaches (e.g. CNN-based feature extraction combined with transformer-based spatial modeling).

image-based phenotypes but also inform downstream clinical applications. By consolidating these predictive methods, we aim to clarify the current landscape of image-to-omics prediction and outline opportunities for future development.

Methodological themes and computational challenges

As shown in Fig. 2, early image-omics studies relied on feature-driven statistical learning strategies based on curated regions of interest (ROIs), semantic features identified by radiologists, and broad textural features. However, modern image-to-omics prediction pipelines have increasingly leveraged deep learning to accelerate the modelling of complex image features. Still, these newer approaches carry their own challenges; high-resolution images, such as gigapixel WSIs, make it infeasible to train deep learning models directly on entire slides. As a result, WSIs are typically divided into smaller regions, known as tiles (or patches), to make computation more manageable. However, for methods that do not require tile-level supervision, aggregation of tile-level features or predictions becomes essential. A range of tile-level aggregation strategies have been proposed, ranging from simple averaging [6, 21] to more advanced weighted averages of clustered tiles, called ‘supertiles’ [14]. Other recent approaches [13, 22] avoid tile-level predictions altogether by leveraging the concept of multiple-instance learning (MIL), in which instances are organized into sets called ‘bags,’ enabling pooling- or attention-based aggregation mechanisms.

In parallel, neural network architectures designed for image-based omics prediction have evolved to address other recurring challenges in the field, including intra-tumor heterogeneity, modality alignment, and mismatched spatial resolution. While convolutional neural networks (CNNs) frequently play a central role in image feature extraction (e.g. PC-CHiP [9], MOMA [12], DeepPT [6]), CNNs are often biased towards detecting local patterns and may struggle to model long-range spatial dependencies relevant to heterogeneous tumor tissue. To better capture global spatial relationships, vision transformers, which use attention to integrate information across multiple spatial scales, have gained traction, particularly for spatial transcriptomics prediction [23–26] where molecular fidelity depends on balancing fine-grained local structures with broader tissue context.

Some of the newest predictive methods using H&E images have leveraged foundation models—large-scale neural networks trained on vast datasets for diverse applications. Visual-language models like CLIP [27] have been adapted for use in pathology, enabling image

captioning from histology slides, as demonstrated by CONCH [28] and PLIP [29]. Moving beyond captioning, pathology-specific foundation models have been applied to a wide range of computational pathology (CPath) tasks, including tissue classification, mutation prediction, disease sub-typing, tumor grading, and more [30–36]. Building on these histopathology foundation models, recent frameworks have incorporated contrastive learning [37–39] to improve alignment between molecular and image-derived representations before prediction. By generating shared image-omics embeddings, contrastive approaches aim to mitigate modality mismatch and compensate for the loss of molecular specificity caused by spatial aggregation.

Despite these advances, current evaluation standards for image-based omics prediction make it difficult to compare performance both within individual studies and across frameworks. While the detection of aberrant molecular states, such as mutation status, can be evaluated as simple binary outcomes, the prediction of bulk transcriptomic profiles is considerably more complex. Although training objectives vary for different bulk RNA-seq frameworks, the most commonly used metrics for comparing predicted and ground-truth expression levels are Mean Squared Error (MSE) and Pearson correlation coefficient (R). Pearson correlation is particularly prevalent for comparing predictive performance across datasets, architectural variants, and other published frameworks [13–16, 31, 33]. However, it can be problematic for gene expression data that routinely suffer from zeros and low-expression counts. Despite this concern, correlation thresholds (e.g. $R > 0.4$) and adjusted *P*-values are often applied to identify well-predicted genes worthy of downstream investigation. While such genes can be informative for within-study analyses, the number of well-predicted genes is relatively unstable and strongly influenced by dataset size [6, 14] and network architecture [7], limiting its utility for cross-study comparisons. As an alternative, robust concordance metrics, such as Lin's Concordance Correlation Coefficient (CCC), may offer a more reliable measure of agreement between ground truth and predicted gene expression.

Evaluating methods for predicting spatial transcriptomics introduces additional challenges due to strong spatial dependencies and concerns regarding reproducibility across tissues. Pearson correlation coefficient and MSE remain the standard quantitative metrics for assessing predictive performance, but they may not capture the interplay between spatial context and gene-level variability. In a benchmark of methods for spatial transcriptomics prediction, Wang et al. [40] discuss how aggregated correlation metrics may dilute biologically relevant signal, suggesting emphasis on spatially variable genes (SVGs) and highly variable genes (HVGs). From this perspective, visualization of predicted spatial gene expression using tissue-level heatmaps can be useful to highlight patterns of predictive performance, though such visualizations are inherently limited to individual genes. Because expression levels vary considerably across tissues, spatial concordance metrics could complement gene-level error metrics by focusing on agreement of spatial organization.

Predicting DNA-based alterations: point mutations, copy number alterations, and structural variations

In this section, we will discuss three major types of DNA-based alterations: point mutations as alterations involving one or a few base pairs, copy-number alterations as those typically affecting multiple genes while spanning less than a full chromosomal arm,

and structural variations as large-scale changes that often involve arm-level or entire chromosomal rearrangements. These categories of DNA-based alterations are primarily distinguished by the length and genomic scale. Previous studies suggest that point mutations, copy-number alterations, and structural variations are associated with distinct histological patterns and subtypes [41, 42]. Thus, the targeted prediction of DNA-based alterations is a common classification problem addressed by recent methods based on H&E images, some of which are summarized in Table 1.

Many modern methods start with established CNN architectures, such as Inception or ResNet, for feature extraction before downstream prediction of point mutations. For example, the modified Inception v3 architecture proposed in 2018 by Coudray et al. [8] was trained to predict the 10 most commonly mutated genes in lung adenocarcinoma (LUAD) and found that six genes—STK11, EGFR, FAT1, SETBP1, KRAS, and TP53—could be predicted from histopathology images. Similar models have been used to predict mutation status of a limited number of commonly mutated genes in hepatocellular carcinoma (HCC) [11], breast carcinoma [43], colorectal cancer [44], and endometrial carcinoma [45]. A recent review of AI-based approaches for predicting gene mutations from imaging further highlights emerging techniques and their applications in the study of common cancers [46].

Recently, H&E images have been used to predict tumor mutational burden (TMB), an important biomarker for predicting response to immunotherapy. The prediction of TMB is closely related to the identification of point mutations, as TMB is typically quantified by counting the number of somatic point mutations per megabase of DNA. It was recently hypothesized that these aggregated mutations result in global morphological changes captured in histopathology images [10]. As such, we have seen the emergence of machine learning methods designed to predict high versus low TMB from histopathology images [10, 47–49]. Image2TMB [10] uses the Inception v3 architecture, tile-level aggregation, and a random forest to predict the probability of high TMB for patients in the TCGA LUAD dataset. Compared to predictions derived from costly targeted gene panels, Image2TMB achieved high accuracy both overall and within stratified patient subgroups. In addition to predictive performance, Image2TMB and similar approaches have revealed clinically relevant histopathological features and spatial heterogeneity within tumor samples [47, 48].

Beyond point mutations and TMB, other DNA-based changes can be inferred from H&E images, providing additional insights. For instance, Pan-cancer Computational Histopathology (PC-CHiP) [9] successfully predicts copy number alterations and large-scale structural variations, including whole-genome duplications (WGDs), chromosome arm gains and losses, and focal amplifications and deletions. While PC-CHiP uses a CNN architecture and regularized linear models to accomplish these tasks, a more recent framework, Multi-omics Multi-cohort Assessment (MOMA) [12], applies vision transformers and multiple instance learning for genomic predictions. When evaluated on data from colorectal cancer patients, MOMA achieves superior performance in predicting copy number alterations when compared to PC-CHiP.

Collectively, these approaches demonstrate diverse capabilities for predicting both small- and large-scale genomic changes from histology images. Although these methods focus on a limited set of known molecular abnormalities rather than comprehensive omics profiles, the accurate prediction of commonly mutated, disease-relevant genes has the potential to guide clinical decision-making through early disease screening, risk assessment, and the detection of prognostic

Table 1 Summary of methods that use imaging to predict DNA-based alterations.

Algorithm or author name	Software	Methodology	Functions*
Coudray, 2018 [8]	https://github.com/ncoudray/DeepPATH	CNN	Predict point mutation status of commonly mutated genes in LUAD
Chen, 2020 [11]	https://github.com/drmaxchen-gbc/HCC-deep-learning	CNN	Predict point mutation status of commonly mutated genes in HCC
Bilal, 2021 [44]	Available upon request to authors	Three separate CNNs	Predict hypermutation, microsatellite instability (MSI), chromosomal instability, high versus low CpG island methylator phenotype (CIMP), and the mutation status of three tumor-related genes
Qu, 2021 [43]	https://github.com/huiqu18/GeneMutationFromHE	CNN for feature extraction and MLP predictor with self-attention	Predict point mutation status and copy number alterations
PC-CHiP (2020) [9]	https://github.com/gerstung-lab/PC-CHiP	Modified CNN for feature extraction and regularized linear models for transfer learning	Predict point mutations, copy number alterations, and structural variations
MOMA (2023) [12]	https://github.com/hms-dbmi/MOMA	CNN; vision transformers; attention-based multiple instance learning	Predict point mutation status of three clinically important genes, MSI, CIMP, and copy number alterations
Image2TMB (2020) [10]	https://github.com/msj3/Image2TMB	CNN; tile-level aggregation; random forest	Predict high versus low tumor mutational burden (TMB)
Niu, 2022 [47]	Not available	CNN	Predict high versus low TMB
Shimada, 2021 [48]	https://github.com/niigata-bioinfo/hypermutation-ai-code/	CNN	Predict high versus low TMB

*Unless otherwise stated, each method uses H&E stained WSIs to accomplish their function. The functions listed relate specifically to omics prediction and do not include clinical prediction tasks (tumor versus normal tissue classification, survival, etc.).

traits. Looking ahead, transfer learning—where learned histological features are used for other downstream tasks—offers a compelling path for expanding clinical insights. For example, PC-CHiP showed that histopathological features can provide complementary information to existing stage-based prognosis across multiple cancer sub-types. Overall, these approaches for image-based genomic prediction represent key computational advancements with the ability to support, rather than replace existing clinical workflows.

Predicting bulk gene expression data

Beyond simple classification tasks and prediction of molecular anomalies, recent models have demonstrated the ability to perform per-gene regression—such as using H&E image features to directly predict bulk gene expression levels for thousands of genes (Table 2). This is a more complex problem than categorical prediction, requiring modern deep learning architectures capable of efficient image processing and large-scale multi-output regression. Convolutional neural networks, multilayer perceptrons, transformers, and hybrid designs have all been applied to extract image features and estimate expression levels gene-by-gene (e.g. DeepPT [6], tRNAsformer [13], hist2RNA [21]). In many frameworks, the final step involves producing gene expression predictions using an MLP head [6, 7, 14], with the dimensionality of the output layer matching the number of input genes.

Inferred gene expression values can be used for a variety of downstream tasks, including virtual spatialization, which uses heatmaps to visualize the intensity of predicted gene expression over tissue images. These maps can reveal ROIs that correlate with tumor growth rate, histological stage, and cancer recurrence. Several methods use WSIs to generate both bulk RNA-seq data and spatial predictions, including HE2RNA [14], MOSBY [15], SEQUOIA [7], and EMO [50]. HE2RNA, a widely cited model, combines an MLP with transfer learning to

predict cancer-related gene signatures and produce virtual spatialization maps that distinguish cell subtypes. MOSBY maps RetCCL (contrastive self-supervised learning)-based image features to omics data by predicting levels of disease-related biomarkers, followed by tile-level inference and colocalization mapping for spatial biomarker discovery. SEQUOIA integrates a pre-trained histopathology foundation model, UNI [30], with a linearized transformer, boosting gene prediction accuracy and producing spatial maps linked to breast cancer recurrence. EMO takes a different approach, training a separate CNN for each gene to predict bulk gene expression and spatial variability. While EMO identifies a large number of genes significantly associated with observed RNA-seq data, it is very computationally intensive. As EMO and MOSBY lack publicly available code, SEQUOIA is the preferred choice among these methods, offering improved performance over MLP-based models such as HE2RNA.

While some histology-based predictive methods for bulk omics explore spatial relationships, others integrate omics regression with downstream clinical prediction. tRNAsformer [13], e.g. employs a dual-head architecture: one head for disease sub-typing and another for bulk RNA-seq prediction. Instead of parallel architecture, the ENLIGHT-DeepPT [6] framework is a two-step workflow for precision medicine applications. First, DeepPT [6] uses neural networks to predict slide-level bulk gene expression profiles, and then ENLIGHT uses the inferred expression to successfully predict patient responses to immune therapies. The accuracy of ENLIGHT’s unsupervised method was comparable to that of a supervised model predicting treatment responses directly from images—demonstrating the utility of using intermediate inferred gene expression in the absence of labeled clinical data. Taking a similar approach, Tavalara et al. [22] used predicted gene expression from WSIs as an intermediate step for classifying mice as susceptible versus supersusceptible to fulminant-like pulmonary tuberculosis. More recently, hist2RNA [21] applied

Table 2 Summary of methods that use imaging for bulk gene expression prediction.

Algorithm or author name	Software	Methodology	Function*
SEQUOIA (2024) [7]	https://github.com/gevaertlab/sequoia-pub	CNN or foundation model for histological feature extraction; MLP, transformer, or linearized transformer for prediction	Predict bulk RNA-seq gene expression profiles; resolve spatial gene expression at loco-regional levels
HE2RNA (2020) [14]	https://github.com/owkin/HE2RNA_code	Multi-layer perceptron (MLP); transfer learning	Predict bulk RNA-seq gene expression profiles; provide virtual spatialization of predicted expression; predict microsatellite instability (MSI) status
MOSBY (2024) [15]	Not available	CNN; contrastive self-supervised learning; multiple instance learning (MIL); MLP regression	Predict bulk profiles (gene expression, gene set expression, protein expression, and DNA-based features); provide virtual spatialization of profiles; produce colocalization maps
EMO (2021) [50]	Not available	Individual CNN models for each gene	Predict bulk RNA-seq gene expression profiles; predict intratumor spatial expression
DeepPT (2024) [6]	https://zenodo.org/records/11125591	CNN for feature extraction; autoencoder for feature compression; MLP regression for prediction	Predict bulk RNA-seq gene expression profiles
tRNAsformer (2023) [13]	https://zenodo.org/records/7613349	CNN for feature extraction; multiple instance learning; transformer encoder	Predict bulk RNA-seq gene expression profiles
Tavolara, 2021 [22]	https://github.com/cialab/image2gene	Gradient boosted trees for feature selection; attention-based multiple instance learning for prediction	Predict bulk RNA-seq gene expression profiles
hist2RNA (2023) [21]	https://github.com/raktim-mondol/hist2RNA	CNNs for feature extraction; patch-level feature aggregation; 1D convolutional layers	Predict bulk RNA-seq gene expression profiles
Image2Omics (2024) [19]	https://github.com/GSK-AI/image2omics	Modified CNN; multiple-instance learning	Predict bulk multi-omics (transcriptomics and proteomics) from high-content cellular imaging

*Unless otherwise stated, each method uses H&E stained WSIs to accomplish their function. The functions listed relate specifically to omics prediction and do not include clinical prediction tasks (tumor versus normal tissue classification, survival, etc.).

a gene expression-based classification model for identifying breast cancer subtypes. Collectively, these studies mark pivotal advances in image-based omics studies, producing comprehensive transcriptomic representations that not only capture molecular states but also enable clinically relevant predictions.

As research continues to investigate the links between omics and histopathology images, new approaches are beginning to explore the relationships between omics and cell imaging. Image2Omics [19], developed by the biopharmaceutical company GSK, adopts a multiple instance learning strategy for using high-content cellular imaging to predict bulk transcriptomics and proteomics. This novel deep learning framework employs a modified ResNet18 backbone for pre-training, followed by omics-specific fine-tuning. Notably, the study demonstrated that certain transcript and protein abundances could be predicted from cell imaging. This work broadens the scope of image-based omics prediction beyond tissue histology and opens new opportunities for linking cellular phenotypes with multi-omics data.

Overall, this collection of frameworks demonstrates the value and versatility of image-based inferred gene expression; the spatialization of predicted gene expression enables the study of tumor heterogeneity, while inferred transcriptomic profiles also serve as effective intermediaries between imaging data and downstream clinical applications. Methodologically, these approaches show clear

trade-offs between predictive performance, transferability, generalizability, and computational complexity. In terms of accuracy and transferability across clinical applications, SEQUOIA and DeepPT-ENLIGHT stand out over earlier methods, such as HE2RNA and tRNAsformer. With respect to generalizability, methods including SEQUOIA, DeepPT, and MOSBY used numerous TCGA cohorts and external validation datasets, whereas tRNAsformer and hist2rna were tested exclusively on kidney and breast cancer datasets, respectively. Although direct comparisons of training time and computational efficiency are lacking, EMO and MOSBY may have limited practical usage because of concerns about accessibility and computational feasibility and accessibility. Taken together, these advances highlight the promising role of image-derived omics representations as flexible and informative bridges between imaging data and molecular biology.

Predicting single-cell gene expression data

Predicting conventional dissociated single-cell transcriptomics (scRNA-seq) directly from imaging is a considerable challenge because the resolution of standard histology imaging does not align with the biological granularity of single-cell molecular profiles without spatial context. While H&E images can capture coarse cellular morphology, their resolution fundamentally underspecifies fine-scale

transcriptional states. Consequently, there are only a few approaches attempting direct image-to-single-cell omics prediction. SCHAF [51] uses H&E histology images to produce spatially-resolved single-cell gene expression profiles, offering two versions: Paired SCHAF, designed for datasets with matched spatial transcriptomics, and Unpaired SCHAF, for those without. Of the two methods, they found that leveraging available spatial transcriptomics (Paired SCHAF) produced more spatially accurate predictions. Regardless of the availability of spatial transcriptomics data, the inherent difficulty of aligning coarse image features with fine-grained single-cell data has prompted the aid of pre-trained models. For example, SCHAF's training integrates the H&E foundation model UNI [30], the scRNA-seq foundation model scGPT [52], and Tangram [53] for spatial alignment of scRNA-seq data. Using a similar strategy, DeepCell [54] modifies the architecture of the pre-trained DeepSpot [54] model to enhance its histology image-based prediction of spatial transcriptomics to single-cell level resolution.

Other methods predict gene expression at pseudo-single-cell resolution through the explicit integration of H&E images with spot-level spatial transcriptomics. For instance, Spotiphy [16] is a novel deconvolution-based approach that integrates three data types: scRNA-seq data, spot-level sequencing-based ST data, and histological images to infer cell-type proportions, impute single-cell RNA data, and generate whole-transcriptome images at pseudo-single-cell resolution. iStar [55], a predecessor of Spotiphy, uses hierarchical image feature extraction to combine histology images and spot-level gene expression to generate near-single-cell resolution gene expression, revealing fine-grained cellular details and patterns missed by existing approaches, such as XFuse [56]. However, in benchmarks comparing cell-type-level expression matrices for each spot, Spotiphy demonstrated superiority over iStar, achieving the highest correlation and cosine similarity and the lowest absolute error and mean square error against the ground truth.

Beyond histopathology imaging, Raman microscopy imaging has emerged as a new platform for image-to-omics prediction. Raman2RNA (R2R) [20] uses live-cell Raman microscopy images, with spectra captured at subcellular resolution, to predict single-cell gene expression profiles via anchor-based spatial smFISH or anchor-free adversarial autoencoders. Like SCHAF, R2R uses Tangram [53] for domain transfer from spatial to scRNA-seq profiles. In downstream trajectory analysis, the authors show that R2R-inferred scRNA-seq can track cell differentiation and gene expression dynamics.

Altogether, these approaches underscore both the promise and the difficulty of image-based single-cell omics prediction. Across histology- and Raman-based methods, the integration of spatial omics data has been critical to aligning coarse image features with fine-grained molecular profiles. However, the availability of associated spatial omics remains limited, restricting the scalability and generalizability of workflows like Spotiphy and iStar, which explicitly require spatial transcriptomics data. As a result, future progress will likely depend on innovative strategies that reduce reliance on paired spatial data, such as improved multimodal representation learning, transfer learning from foundation models, and more robust alignment techniques capable of operating under weak or indirect supervision.

Predicting spatial gene expression data

Within the past five years, numerous computational methods have emerged for the prediction of spatially resolved gene expression

from histological images (Table 3). Convolutional neural networks are among the most popular deep learning architectures for such tasks. For instance, ST-Net [57] builds on the pre-trained DenseNet-121 [58] model to predict the spatially localized gene expression of 250 genes using H&E-stained histopathology images. The authors found that ST-Net could accurately predict spatial variation in the expression of 102 genes and detect biomarkers of intra-tumor variation. A similar framework for spatial transcriptomics prediction, BrST-Net [59], used a combination of the CNN EfficientNet-b0 [60] and an auxiliary network, outperforming ST-Net by predicting more than twice the number of positively correlated genes.

Other CNN-based approaches for predicting spatial transcriptomics from H&E images go beyond the detection of positively correlated genes to capture spatial heterogeneity and enable downstream inference. For example, Levy-Jurgenson et al. [61] used Inception V3 [62] to train separate models for pre-selected traits, generating scaled tile-level expression scores used to produce molecular heat maps and heterogeneity maps. DeepSpaCE [63], adapted from VGG16 [64], predicts spot-level gene expression and supports practical applications such as super-resolution spatial mapping and imputation of gene expression between consecutive tissue sections; its super-resolution mapping of SPARC expression delineated a tumor invasion boundary more clearly than the original spot-level expression. Taking a different approach, BLEEP [39], adopts a contrastive learning strategy inspired by CLIP [27], jointly training image and expression encoders to generate bimodal latent space embeddings. This representation enables accurate imputation for query image patches, and the BLEEP approach outperforms existing methods such as ST-Net [57] and HisToGene [65].

Beyond CNNs, vision transformers have become the backbone of many predictive spatial omics models. For instance, iStar [55] employs a hierarchical vision transformer (HViT) to extract histology features before generating near-single-cell resolution gene expression. Notably, iStar also provides high-resolution annotation of tissue architectures via unsupervised segmentation and cell type prediction. Exemplar Guided Network (EGN) [23] combines an image reconstruction feature extractor, a ViT backbone, Exemplar Bridging (EB) blocks for refining ViT patch representations, and an attention-based gene prediction block. Using a modified vision transformer with dense image patch sampling, HisToGene [65] predicts super-resolution spatial gene expression and shows that predicted expression can be used to recover pathologist-annotated spatial domains. In contrast to vision transformer approaches, mclSTExp [37] applies an NLP-style transformer, treating spatial transcriptomics spots as 'words' and their sequence as 'sentences.' Its contrastive learning framework with parallel CNN and transformer networks captures visual and spatial gene expression features. When evaluated on three benchmark datasets, mclSTExp outperformed ST-Net [57], HisToGene [65], His2ST [24], THItGene [26], and BLEEP [39] for all considered genes as well as the top 50 highly expressed genes.

Graph neural networks have also become increasingly popular components of frameworks predicting spatial transcriptomics from WSIs. There is also evidence that integrating graph neural networks with CNNs improves histology image and omics fusion [66, 67]. Thus, recent predictive models have adopted similar hybrid graph-based models with CNNs or vision transformers. For instance, EGNv2 [68] combines a ResNet feature extractor with a graph convolutional network backbone, alongside a graph exemplar bridging block to guide the prediction of gene expression for slide image

Table 3 Summary of methods that use imaging to predict single-cell and spatial gene expression.

Algorithm or author name	Software	Methodology	Function*
SCHAF (2025) [51]	https://github.com/ccomiter1/SCHAF	Encoder-decoder architecture; histopathology foundation model; vision transformer; scRNA-seq foundation model; Tangram [53] to infer spatial location for scRNA-seq	Predict spatially resolved single-cell transcriptomic data
DeepSpot/DeepCell (2025) [54]	https://github.com/ratschlab/DeepSpot	Deep set neural network; pre-trained histopathology foundation models	Predict spatial gene expression profiles; aggregate expression into pseudo-bulk RNA profiles; predict single-cell spatial transcriptomics
Spotiphy (2025) [16]	https://github.com/jyyulab/Spotiphy	Probabilistic generative model; variational inference	Generate whole transcriptome images with pseudo-single-cell resolution; impute spot-level gene expression
iStar (2024) [55]	https://github.com/daviddaiweizhang/istar	Hierarchical vision transformer (HViT); self-supervised learning; MLP; score-based cell type inference	Infer super-resolution spatial gene expression; Annotate cell types at near-single-cell resolution
Raman2RNA (2024) [20]	https://github.com/kosekijkk/raman2rna	MLP to predict smFISH from Raman spectra; Tangram [53] to integrate predicted smFISH profile to scRNA-seq; adversarial autoencoders (AAEs) for anchor-free prediction	Predict single-cell RNA-seq profiles from Raman microscopy images; predict expression states of marker genes; track live cells and predict gene expression dynamics
ST-Net (2020) [57]	https://github.com/bryanhe/ST-Net	CNN	Predict spatial gene expression
BrST-Net (2023) [59]	https://github.com/Mamunur-20/BrSTNet	Evaluated 10 different architectures: CNNs and transformers with and without an additional auxiliary network	Predict spatial gene expression
Levy-Jurgenson, 2020 [61]	https://github.com/alonalj/PathoMCH	CNN	Spatially resolve bulk mRNA and miRNA expression; predict gene expression scores for molecular traits of interest; generate heterogeneity maps and compute heterogeneity index
DeepSpaCE (2022) [63]	https://github.com/tmonjo/DeepSpaCE	CNN; multi-task learning	Predict spatial gene expression; generate super-resolution spatial gene expression; impute ST data for blank tissue sections between consecutive sections
BLEEP (2023) [39]	https://github.com/bowang-lab/BLEEP	CNN image encoder; MLP expression encoder; contrastive learning; query-reference imputation; <i>k</i> -nearest reference expression profiles	Predict spatial gene expression
HisToGene (2021) [65]	https://github.com/maxpmax/HisToGene	Vision transformer	Infer super-resolution spatial gene expression
EGNv1 (2022) [23]	https://github.com/Yan98/EGN	Vision transformer blocks and exemplar bridging blocks; attention-based prediction	Predict gene expression for fine-grained areas (i.e. windows) of WSIs
mclSTExp (2024) [37]	https://github.com/ZhicengShi/mclSTExp	CNN image encoder; transformer expression encoder; contrastive learning; query <i>k</i> -nearest spots	Predict spatial gene expression
EGGN (EGNv2) (2024) [68]	https://github.com/Yan98/EGN	CNN feature extraction; Graph convolutional network (GCN); graph exemplar bridging block; attention-based prediction	Predict spatial gene expression
Hist2ST (2022) [24]	https://github.com/biomed-AI/Hist2ST	Transformer; Graph neural network (GNN); LSTM; ZINB distribution	Predict spatial gene expression
TCGN (2024) [25]	https://github.com/lugia-xiao/TCGN	CNN; transformer; GNN	Predict spatial gene expression; can be adapted for bulk RNA-seq prediction

(Continued)

Table 3 Continued

Algorithm or author name	Software	Methodology	Function*
THItToGene (2023) [26]	https://github.com/yrjia1015/THItToGene	Vision transformer; Graph attention network (GAT)	Predict spatial gene expression
HGGEP (2024) [70]	https://github.com/QSong-github/HGGEP	Gradient enhancement module; convolutional block attention module; vision transformer; hypergraph association module; LSTM	Predict spatial gene expression
BayesDeep (2023) [71]	https://github.com/Xijiang1997/BayesDeep	Bayesian hierarchical model	Predict single-cell resolution spatial gene expression
GeneCodeR (2022) [72]	https://github.com/AskExplain/GeneCodeR	Generative encoding via Generalized Canonical Procrustes	Transform histology images to gene expression vectors through functional mapping
STAGE (2024) [73]	https://github.com/zhanglabtools/STAGE	Supervised autoencoder	Generate high-density spatial gene expression; improve low-quality ST data; impute ST data for blank tissue sections between consecutive sections
NSL (2022) [74]	Not available	Neural stain deconvolution	Predict spatial gene expression
OmiCLIP/Loki PredEx (2025) [38]	https://github.com/GuangyuWangLab2021/Loki	Dual encoder framework: NLP-style transformer for text encoding (genes treated as sentences); ViT for image encoding; contrastive learning	Predict spatial gene expression

*Unless otherwise stated, each method uses H&E stained WSIs to accomplish their function. The functions listed relate specifically to omics prediction and do not include clinical prediction tasks (tumor versus normal tissue classification, survival, etc.).

windows. The Hist2ST [24] framework includes ConvMixer module [69], transformer, and GNN to learn features before following the ZINB distribution for gene expression prediction. Benchmark evaluations in predictive performance and spatial region detection showed that Hist2ST outperformed both ST-Net and HisToGene. Three additional hybrid architectures were introduced in 2024: TCGN [25], THItToGene [26], and HGGEP [70]. Each achieved superior predictive performance over ST-Net, HisToGene, and Hist2ST, reinforcing the notion that hybrid graph-based approaches can more effectively capture histology image and omics relationships.

Other methods using histology images for spatial transcriptomics prediction do not rely on CNN, transformer, or graph-based architectures and instead apply probabilistic modeling, matrix factorizations, or lightweight autoencoding frameworks. For example, BayesDeep [71] uses a Bayesian hierarchical model to reconstruct spatial transcriptomics at single-cell resolution by linking spot-level molecular profiles with morphological characteristics in histology images. GeneCodeR [72] takes a generative approach to map histology images to gene expression, drawing conceptual similarities to autoencoders, Procrustes analysis, Canonical Correlation Analysis, and Singular Value Decomposition (SVD). In a related strategy, STAGE [73] uses a custom supervised autoencoder framework to generate super-resolution spatial transcriptomics data and can be extended to generate spatial gene expression profiles of blank tissue sections between consecutive ST sections. Neural Stain Learning (NSL) [74], by contrast, predicts local spatial gene expression using a novel stain deconvolution layer to address the lack of explicit modeling of tissue staining. NSL has the key benefit of simplicity, requiring only 11 trainable weight parameters as opposed to the millions required in more complex CNN-based approaches.

Several recent models leverage foundation models and extend these capabilities to spatial transcriptomics prediction. DeepSpot [54] integrates pre-trained pathology foundation models [30, 31, 33] and tissue

context to predict spatial transcriptomics from H&E images. While DeepSpot relies on pre-trained models, OmiCLIP [38] represents a new image-omics foundation model that underlies the Loki platform, which supports multiple downstream tasks such as tissue alignment, image-transcriptomic retrieval, and spatial gene expression prediction. OmiCLIP uses contrastive learning to jointly train image and text encoders, where the text input consists of sentences representing the top-expressed genes in each spatial spot (e.g. ‘SNAP25 ENO2 CKB... VPS13D’). The resulting visual-transcriptomic embeddings are used by Loki PredEx, which predicts gene expression from image patches. Across diverse datasets, LokiPredEx consistently achieved competitive or superior performance in terms of MSE and PCC metrics when compared with Hist2ST, HisToGene, BLEEP, and mclSTExp—with comparisons to BLEEP and mclSTExp being particularly notable given their similar contrastive learning-based designs. Beyond predictive accuracy, OmiCLIP also offers practical advantages. It is more computationally efficient than models like Hist2ST and HisToGene, which require substantial hardware resources, and the availability of pretrained weights further enhances its accessibility and scalability within the field.

Given the growing number of tools for predicting spatial gene expression from histology images, a natural question is how to compare these methods to one another. A recent benchmark study of 11 methods compared the models’ predictive performance, generalizability, usability, translation potential, and efficiency on five spatial transcriptomics datasets [40]. While the authors did not identify a top performer across all categories, HisToGene, DeepSpaCE, and Hist2ST displayed noteworthy performance in model generalizability and usability.

In summary, the diverse landscape of methodologies for image-based spatial transcriptomics prediction reflects how the field has evolved to address challenges of spatial heterogeneity, data resolution, generalization, and computational efficiency. Heterogeneity within

and across tissue images has driven the adoption of vision transformers, graph-based models, and hybrid architectures, which often capture long-range spatial context more effectively than traditional CNNs. Contrastive learning strategies that jointly learn image and omics representations have had similar success by aligning multi-modal information in shared latent spaces. Foundation models such as OmiCLIP, trained on millions of histopathology images, extend these capabilities by improving generalizability across datasets and potential clinical applications. Nevertheless, the inherent difficulty of generating spatially resolved omics directly from images presents both a challenge and opportunity for continued development of advanced statistical approaches and neural representation learning.

Future perspectives and concluding remarks

While this review focuses on using imaging to predict omics data, tools for the reverse (i.e. using omics to predict imaging) are emerging as well. Zandavi et al. build on the concept of ‘omics imagification,’ which aims to transform high-dimensional molecular data into a 2D image [75]. Their proposed method, Fotomics, uses the fast Fourier transformation to generate image representations of single-cell RNA-seq data. A CNN trained on Fotomics-generated images showed superior classification accuracy of diverse cell types when compared to conventional scRNA-seq classifiers, as well as DeepInsight [76]—another method for imagification of non-image data. Going beyond abstract image representations of omics data, MorphNet [77] predicts biologically interpretable cell morphology images from gene expression. MorphNet requires training a Variational Autoencoder (VAE) and GAN on spatial transcriptomics data, and trained MorphNet can predict new cell morphology from scRNA-seq data. Despite these advances, predicting imaging from omics is still a relatively unexplored area of research.

Although the accessibility of large-scale datasets and recent advances in deep learning have made image-based omics prediction increasingly feasible, several open challenges remain. Variation in pre-processing pipelines, the heterogeneity of these large-scale cohorts, and inconsistent evaluation practices have hindered the development of standardized benchmarking procedures. In addition, the transferability and generalizability of models across imaging modalities, omics data types, and disease contexts remain largely unexplored. For example, comparing top-performing bulk RNA-seq prediction methods with wsi2rppa [78]—one of the only frameworks available for image-based prediction of protein expression levels—would provide valuable insight into cross-omics generalization. Beyond issues of transferability, the field would also benefit from greater emphasis on biological interpretability, including the integration of causal modeling approaches or explainable machine learning (XAI) methods, such as SHAP [79] or LIME [80].

Image-based omics prediction holds significant promise for clinical applications, including biomarker discovery, prognosis estimation, and personalized treatment, aligning closely with the goals of precision medicine. However, when molecular profiles are inferred rather than directly measured, their use in clinical decision-making requires careful ethical and regulatory consideration. Inferred omics may reflect uncertainty, bias, or limited generalizability i.e. not readily apparent in clinical settings, increasing the risk of misinterpretation. As a result, models producing inferred omics likely require rigorous validation, transparent communication of limitations, and restricted clinical use. With appropriate governance, these methods can enhance understanding of molecular and phenotypic

relationships while supporting, rather than replacing, established clinical assays.

Key Points

- With the growing availability of clinical imaging data, such as histopathology whole-slide images, there is an increasing interest in using imaging data to directly infer molecular omics information.
- While existing reviews have covered image-omics data fusion and image-based prediction of biomarkers in specific cancer types, this work is the first to comprehensively review predictive frameworks for direct image-to-omics inference.
- This review highlights current approaches for image-based molecular omics prediction, focusing on methods developed to predict DNA-based aberrations, bulk transcriptomic profiles, single-cell transcriptomics, and spatial transcriptomics.
- Deep learning-based strategies, including convolutional neural networks, multilayer perceptrons, and transformers, have proven particularly effective for leveraging imaging data to predict molecular omics data.

Acknowledgements

The resources from the Quantitative Biomedical Research Center (QBRC) and BioHPC at UT Southwestern Medical Center are gratefully acknowledged.

Author contributions

AB, GX, and LX conceived and designed the study. XX, YZ, and QL were involved in the writing. LX and GX acquired the funding. AB, GX, and LX wrote and revised the manuscript. All authors have read, revised, and approved the final manuscript.

Funding

This work was supported by the following funding: the Sam Day Foundation, Rally Foundation, Children’s Cancer Fund (Dallas), the Cancer Prevention and Research Institute of Texas (RP180319, RP200103, RP220032, RP170152 and RP180805), and the National Institutes of Health funds (R01DK127037, R01CA263079, R21CA259771, P30CA142543, UM1HG011996, and R01HL144969), and the support to the Data Science Shared Resource from Cancer Center Support Grant P30 CA142543 (to L.X.); the National Institutes of Health (1R01GM115473, 1R01GM140012, 5R01CA152301, P30CA142543, P50CA70907, R35GM136375); and the Cancer Prevention and Research Institute of Texas (RP180805, RP190107) (to G. X.).

Data availability

The datasets used in this study are publicly available, and the original publications for each dataset are cited in the manuscript.

Ethics declarations

All the authors declare that they have no competing interests.

References

- Antonelli L, Guarracino MR, Maddalena L. Integrating imaging and omics data: a review. *Biomed Signal Process Control* 2019;**52**: 264–80.
- Hériché J-K, Alexander S, Ellenberg J. Integrating imaging and omics: computational methods and challenges. *Annu Rev Biomed Data Sci* 2019;**2**:175–97.
- Huang W, Tan K, Zhang Z *et al*. A review of fusion methods for omics and imaging data. *IEEE/ACM Trans Comput Biol Bioinform* 2023;**20**:74–93.
- Schneider L, Laiouar-Pedari S, Kuntz S *et al*. Integration of deep learning-based image analysis and genomic data in cancer pathology: a systematic review. *Eur J Cancer* 2022;**160**:80–91. <https://doi.org/10.1016/j.ejca.2021.10.007>
- Watson ER, Taherian Fard A, Mar JC. Computational methods for single-cell imaging and omics data integration. *Front Mol Biosci* 2021;**8**:768106. <https://doi.org/10.3389/fmolb.2021.768106>
- Hoang DT, Dinstag G, Shulman ED *et al*. A deep-learning framework to predict cancer treatment response from histopathology images through imputed transcriptomics. *Nat Can* 2024;**5**:1305–17. <https://doi.org/10.1038/s43018-024-00793-2>
- Pizurica M, Zheng Y, Carrillo-Perez F *et al*. Digital profiling of gene expression from histology images with linearized attention. *Nat Commun* 2024;**15**:9886. <https://doi.org/10.1038/s41467-024-54182-5>
- Coudray N, Ocampo PS, Sakellaropoulos T *et al*. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018;**24**:1559–67. <https://doi.org/10.1038/s41591-018-0177-5>
- Fu Y, Jung AW, Torne RV *et al*. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Can* 2020;**1**:800–10. <https://doi.org/10.1038/s43018-020-0085-8>
- Jain MS, Massoud TF. Predicting tumour mutational burden from histopathological images using multiscale deep learning. *Nat Mach Intell* 2020;**2**:356–62.
- Chen M, Zhang B, Topatana W *et al*. Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. *NPJ Precis Oncol* 2020;**4**:14. <https://doi.org/10.1038/s41698-020-0120-3>
- Tsai PC, Lee TH, Kuo KC *et al*. Histopathology images predict multi-omics aberrations and prognoses in colorectal cancer patients. *Nat Commun* 2023;**14**:2102. <https://doi.org/10.1038/s41467-023-37179-4>
- Alsaafin A, Safarpour A, Sikaroudi M *et al*. Learning to predict RNA sequence expressions from whole slide images with applications for search and classification. *Commun Biol* 2023;**6**:304. <https://doi.org/10.1038/s42003-023-04583-x>
- Schmauch B, Romagnoni A, Pronier E *et al*. A deep learning model to predict RNA-seq expression of tumours from whole slide images. *Nat Commun* 2020;**11**:3877. <https://doi.org/10.1038/s41467-020-17678-4>
- Senbabaoglu Y, Prabhakar V, Khormali A *et al*. MOSBY enables multi-omic inference and spatial biomarker discovery from whole slide images. *Sci Rep* 2024;**14**:18271. <https://doi.org/10.1038/s41598-024-69198-6>
- Yang J, Zheng Z, Jiao Y *et al*. Spotiphy enables single-cell spatial whole transcriptomics across an entire section. *Nat Methods* 2025;**22**:724–36. <https://doi.org/10.1038/s41592-025-02622-5>
- Ash JT, Darnell G, Munro D *et al*. Joint analysis of expression levels and histological images identifies genes associated with tissue morphology. *Nat Commun* 2021;**12**:1609. <https://doi.org/10.1038/s41467-021-21727-x>
- Badea L, Stanescu E. Identifying transcriptomic correlates of histology using deep learning. *PLoS One* 2020;**15**:e0242858. <https://doi.org/10.1371/journal.pone.0242858>
- Mehrizi R, Mehrjou R, Alegro M *et al*. Multi-omics prediction from high-content cellular imaging with deep learning. *arXiv* 2023. <https://doi.org/10.48550/arXiv.2306.09391>
- Kobayashi-Kirschvink KJ, Comiter CS, Gaddam S *et al*. Prediction of single-cell RNA expression profiles in live cells by Raman microscopy with Raman2RNA. *Nat Biotechnol* 2024;**42**:1726–34. <https://doi.org/10.1038/s41587-023-02082-2>
- Mondol RK, Millar EKA, Graham PH *et al*. hist2RNA: an efficient deep learning architecture to predict gene expression from breast cancer histopathology images. *Cancers* 2023;**15**:2569. <https://doi.org/10.3390/cancers15092569>
- Tavolara TE, Niaz MKK, Gower AC *et al*. Deep learning predicts gene expression as an intermediate data modality to identify susceptibility patterns in mycobacterium tuberculosis infected diversity outbred mice. *EBioMedicine* 2021;**67**:103388. <https://doi.org/10.1016/j.ebiom.2021.103388>
- Yang Y, Hossain M, Stone E *et al*. Exemplar guided deep neural network for spatial transcriptomics analysis of gene expression prediction. *arXiv* 2022. <https://doi.org/10.48550/arXiv.2210.16721>
- Zeng Y, Wei Z, Yu W *et al*. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Brief Bioinform* 2022;**23**:bbac297. <https://doi.org/10.1093/bib/bbac297>
- Xiao X, Kong Y, Li R *et al*. Transformer with convolution and graph-node co-embedding: an accurate and interpretable vision backbone for predicting gene expressions from local histopathological image. *Med Image Anal* 2024;**91**:103040. <https://doi.org/10.1016/j.media.2023.103040>
- Jia Y, Liu J, Chen L *et al*. THltoGene: a deep learning method for predicting spatial transcriptomics from histological images. *Brief Bioinform* 2023;**25**:bbad464. <https://doi.org/10.1093/bib/bbad464>
- Radford A, Kim J, Hallacy C. Learning transferable visual models from natural language supervision. *Int Conf Mach Learn* 2021;**1**:8748–63.
- Lu MY, Chen B, Williamson DFK *et al*. A visual-language foundation model for computational pathology. *Nat Med* 2024;**30**:863–74. <https://doi.org/10.1038/s41591-024-02856-4>
- Huang Z, Bianchi F, Yuksekogun M *et al*. A visual-language foundation model for pathology image analysis using medical twitter. *Nat Med* 2023;**29**:2307–16. <https://doi.org/10.1038/s41591-023-02504-3>
- Chen RJ, Ding T, Lu MY *et al*. Towards a general-purpose foundation model for computational pathology. *Nat Med* 2024;**30**:850–62. <https://doi.org/10.1038/s41591-024-02857-3>
- Filiot A, Ghermi R, Olivier A *et al*. Scaling self-supervised learning for histopathology with masked image modeling. *medRxiv* 2023. <https://doi.org/10.1101/2023.07.21.23292757>
- Filiot A, Jacob P, Kain A *et al*. Phikon-v2, a large and public feature extractor for biomarker prediction. *arXiv* 2024. <https://doi.org/10.48550/arXiv.2409.09173>
- Saillard C, Jenatton R, Llinas-López F *et al*. H-optimus-0. [GitHub] 2024. <https://github.com/bioptimus/releases/tree/main/models/h-optimus/v0>
- Wang X, Zhao J, Marostica E *et al*. A pathology foundation model for cancer diagnosis and prognosis prediction. *Nature* 2024;**634**:970–8. <https://doi.org/10.1038/s41586-024-07894-z>
- Xu H, Usuyama N, Bagga J *et al*. A whole-slide foundation model for digital pathology from real-world data. *Nature* 2024;**630**:181–8.
- Zhang S, Xu Y, Usuyama N *et al*. A multimodal biomedical foundation model trained from fifteen million image–text pairs. *NEJM AI* 2025;**2**:1. <https://doi.org/10.1056/AIoa2400640>

37. Min W, Shi Z, Zhang J *et al*. Multimodal contrastive learning for spatial gene expression prediction using histology images. *Brief Bioinform* 2024;**25**:bbae551. <https://doi.org/10.1093/bib/bbae551>
38. Chen W, Zhang P, Tran TN *et al*. A visual-omics foundation model to bridge histopathology with spatial transcriptomics. *Nat Methods* 2025;**22**:1568–82. <https://doi.org/10.1038/s41592-025-02707-1>
39. Xie R, Pang K, Chung S *et al*. Spatially resolved gene expression Prediction from H&E Histology Images via Bi-modal contrastive learning. *Proc 37th IntConf Neur Inf Proces Sys (NeurIPS)*, Article 3095. New Orleans, LA, USA: Curran Associates Inc., 2023.
40. Wang C, Chan AS, Fu X *et al*. Benchmarking the translational potential of spatial gene expression prediction from histology. *Nat Commun* 2025;**16**:1544. <https://doi.org/10.1038/s41467-025-56618-y>
41. Calderaro J, Couchy G, Imbeaud S *et al*. Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. *J Hepatol* 2017;**67**:727–38. <https://doi.org/10.1016/j.jhep.2017.05.014>
42. Natrajan R, Sailem H, Mardakheh FK *et al*. Microenvironmental heterogeneity parallels breast cancer progression: a histology-genomic integration analysis. *PLoS Med* 2016;**13**:e1001961. <https://doi.org/10.1371/journal.pmed.1001961>
43. Qu H, Zhou M, Yan Z *et al*. Genetic mutation and biological pathway prediction based on whole slide images in breast carcinoma using deep learning. *NPJ Precis Oncol* 2021;**5**:87. <https://doi.org/10.1038/s41698-021-00225-9>
44. Bilal M, Raza SEA, Azam A *et al*. Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. *Lancet Digit Health* 2021;**3**:e763–72. [https://doi.org/10.1016/S2589-7500\(21\)00180-1](https://doi.org/10.1016/S2589-7500(21)00180-1)
45. Hong R, Liu W, DeLair D *et al*. Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models. *Cell Rep Med* 2021;**2**:100400. <https://doi.org/10.1016/j.xcrm.2021.100400>
46. Shao J, Ma J, Zhang Q *et al*. Predicting gene mutation status via artificial intelligence technologies based on multimodal integration (MMI) to advance precision oncology. *Semin Cancer Biol* 2023;**91**:1–15. <https://doi.org/10.1016/j.semcancer.2023.02.006>
47. Niu Y, Wang L, Zhang X *et al*. Predicting tumor mutational burden from lung adenocarcinoma histopathological images using deep learning. *Front Oncol* 2022;**12**:927426. <https://doi.org/10.3389/fonc.2022.927426>
48. Shimada Y, Okuda S, Watanabe Y *et al*. Histopathological characteristics and artificial intelligence for predicting tumor mutational burden-high colorectal cancer. *J Gastroenterol* 2021;**56**:547–59. <https://doi.org/10.1007/s00535-021-01789-w>
49. Wang L, Jiao Y, Qiao Y *et al*. A novel approach combined transfer learning and deep learning to predict TMB from histology image. *Pattern Recogn Lett* 2020;**135**:244–8.
50. Wang Y, Kartasalo K, Weitz P *et al*. Predicting molecular phenotypes from histopathology images: a transcriptome-wide expression-morphology analysis in breast cancer. *Cancer Res* 2021;**81**:5115–26. <https://doi.org/10.1158/0008-5472.CAN-21-0482>
51. Comiter C, Chen X, Vaishnav ED *et al*. Inference of single cell profiles from histology stains with the single cell omics from histology analysis framework (SCHAF). *bioRxiv* 2025. <https://doi.org/10.1101/2023.03.21.533680>
52. Cui H, Wang C, Maan H *et al*. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods* 2024;**21**:1470–80. <https://doi.org/10.1038/s41592-024-02201-0>
53. Biancalani T, Scalia G, Buffoni L *et al*. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nat Methods* 2021;**18**:1352–62. <https://doi.org/10.1038/s41592-021-01264-7>
54. Nonchev K, Dawo S, Silina K *et al*. DeepSpot: Leveraging spatial context for enhanced spatial transcriptomics Prediction from H&E Images. *medRxiv* 2025. <https://doi.org/10.1101/2025.02.09.25321567>
55. Zhang D *et al*. Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. *Nat Biotechnol* 2024;**42**:1372–7.
56. Bergenstrahle L *et al*. Super-resolved spatial transcriptomics by deep data fusion. *Nat Biotechnol* 2022;**40**:476–9. <https://doi.org/10.1038/s41587-021-01075-3>
57. He B, Bergensträhle L, Stenbeck L *et al*. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng* 2020;**4**:827–34. <https://doi.org/10.1038/s41551-020-0578-x>
58. Huang G, Liu Z, Maaten L *et al*. Densely connected convolutional networks. *Proc IEEE Conf Comput Vis Pattern Recognit* 2017;**1**:2261–9.
59. Rahaman MM, Millar EKA, Meijering E. Breast cancer histopathology image-based gene expression prediction using spatial transcriptomics data and deep learning. *Sci Rep* 2023;**13**:13604. <https://doi.org/10.1038/s41598-023-40219-0>
60. Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. *Proc 36th Int Conf Mach Learn* 2019;**1**:6105–14.
61. Levy-Jurgenson A, Tekpli X, Kristensen VN *et al*. Spatial transcriptomics inferred from pathology whole-slide images links tumor heterogeneity to survival in breast and lung cancer. *Sci Rep* 2020;**10**:18802. <https://doi.org/10.1038/s41598-020-75708-z>
62. Szegedy C, Vanhoucke V, Ioffe S *et al*. Rethinking the inception architecture for computer vision. *Proc IEEE Conf Comput Vis Pattern Recognit* 2016;**1**:2818–26.
63. Monjo T, Koido M, Nagasawa S *et al*. Efficient prediction of a spatial transcriptomics profile better characterizes breast cancer tissue sections without costly experimentation. *Sci Rep* 2022;**12**:4133. <https://doi.org/10.1038/s41598-022-07685-4>
64. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014. <https://doi.org/10.48550/arXiv.1409.1556>
65. Pang M, Su K, Li M. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. *bioRxiv* 2021. <https://doi.org/10.1101/2021.11.28.470212>
66. Ahmedt-Aristizabal D, Armin MA, Denman S *et al*. A survey on graph-based deep learning for computational histopathology. *Comput Med Imaging Graph* 2022;**95**:102027. <https://doi.org/10.1016/j.compmedimag.2021.102027>
67. Chen RJ, Lu MY, Wang J *et al*. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans Med Imaging* 2022;**41**:757–70. <https://doi.org/10.1109/TMI.2020.3021387>
68. Yang Y, Hossain M, Stone EA *et al*. Spatial transcriptomics analysis of gene expression prediction using exemplar guided graph neural network. *Pattern Recogn Lett* 2024;**145**:109966.
69. Trockman A, Kolter J. Patches are all you need. *arXiv* 2022. <https://doi.org/10.48550/arXiv.2201.09792>
70. Li B, Zhang Y, Wang Q *et al*. Gene expression prediction from histology images via hypergraph neural networks. *Brief Bioinform* 2024;**25**:bbae500. <https://doi.org/10.1093/bib/bbae500>

71. Jiang, X, Dong L, Wang S *et al*. Reconstructing spatial transcriptomics at the single-cell resolution with BayesDeep. *bioRxiv* 2023. <https://doi.org/10.1101/2023.12.07.570715>
72. Banh D, Huang A. Scalable parametric encoding of multiple modalities. *bioRxiv* 2021. <https://doi.org/10.1101/2021.07.09.451779>
73. Li S, Gai K, Dong K *et al*. High-density generation of spatial transcriptomics with STAGE. *Nucleic Acids Res* 2024;**52**:4843–56. <https://doi.org/10.1093/nar/gkae294>
74. Dawood M, Branson K, Rajpoot N *et al*. All you need is color: image based spatial gene expression prediction using neural stain learning. *arXiv* 2021. <https://doi.org/10.48550/arXiv.2108.10446>
75. Zandavi S, Liu D, Chung V *et al*. Fotomics: Fourier transform-based omics imagification for deep learning-based cell-identity mapping using single-cell omics profiles. *Artif Intell Rev* 2022;**56**:7263–78.
76. Sharma A, Vans E, Shigemizu D *et al*. DeepInsight: a methodology to transform a non-image data to an image for convolution neural network architecture. *Sci Rep* 2019;**9**:11399. <https://doi.org/10.1038/s41598-019-47765-6>
77. Lee H, Welch J. MorphNet predicts cell morphology from single-cell gene expression. *bioRxiv* 2022. <https://doi.org/10.1101/2022.10.21.513201>
78. Liu H, Xie X, Wang B. Deep learning infers clinically relevant protein levels and drug response in breast cancer from unannotated pathology images. *NPJ Breast Cancer* 2024;**10**:18. <https://doi.org/10.1038/s41523-024-00620-y>
79. Lundberg S, Lee S. A unified approach to interpreting model predictions. *Proc 31st Int Conf Neur Infor Process Sys*, pp. 4768–77. Long Beach, California, USA: Curran Associates Inc., 2017.
80. Ribeiro M, Singh S, Guestrin C. “Why should I trust you?”: Explaining the predictions of any classifier. *Proc 22nd ACM SIGKDD Int Conf Knowledge Discovery and Data Mining*, pp. 1135–44. San Francisco, California, USA: Association for Computing Machinery, 2016.