HEST-1k: A Dataset for Spatial Transcriptomics and Histology Image Analysis

Guillaume Jaume[1,2,*]   **Paul Doucet**[1,3,*]   **Andrew H. Song**[1,2]   **Ming Y. Lu**[1,2,4]
Cristina Almagro-Pérez[1,2]   **Sophia J. Wagner**[1,6,7]   **Anurag J. Vaidya**[1,2,5]
Richard J. Chen[1,2]   **Drew F.K. Williamson**[8]   **Ahrong Kim**[1,9]   **Faisal Mahmood**[1,2]

[1]Mass General Brigham, Boston, USA   [2]Harvard Medical School, Boston, USA
[3]ETH Zurich, Switzerland   [4]EECS MIT, Cambridge, USA
[5]HST MIT, Cambridge, USA   [6]TUM, Munich, Germany
[7]Helmholtz Munich, Munich, Germany [8]Emory School of Medicine, Atlanta, USA
[9]Pusan National University, South Korea
gjaume@bwh.harvard.edu   faisalmahmood@bwh.harvard.edu

Abstract

Spatial transcriptomics enables interrogating the molecular composition of tissue with ever-increasing resolution and sensitivity. However, costs, rapidly evolving technology, and lack of standards have constrained computational methods in ST to narrow tasks and small cohorts. In addition, the underlying tissue morphology, as reflected by H&E-stained whole slide images (WSIs), encodes rich information often overlooked in ST studies. Here, we introduce HEST-1k, a collection of 1,229 spatial transcriptomic profiles, each linked to a WSI and extensive metadata. HEST-1k was assembled from 153 public and internal cohorts encompassing 26 organs, two species (*Homo Sapiens* and *Mus Musculus*), and 367 cancer samples from 25 cancer types. HEST-1k processing enabled the identification of 2.1 million expression–morphology pairs and over 76 million nuclei. To support its development, we additionally introduce the HEST-Library, a Python package designed to perform a range of actions with HEST samples. We test HEST-1k and Library on three use cases: (1) benchmarking foundation models for pathology (HEST-Benchmark), (2) biomarker exploration, and (3) multimodal representation learning. HEST-1k, HEST-Library, and HEST-Benchmark can be freely accessed at `https://github.com/mahmoodlab/hest`.

1   Introduction

Advances in molecular profiling enable spatially-resolved gene expression analysis with increasingly large gene panels, enhanced spatial resolution, and greater sensitivity [9, 122]. From the early days of bulk RNA sequencing constrained by its coarse resolution and limited gene panels, spatially-resolved technologies have progressed to achieve whole-transcriptome sequencing at sub-cellular resolution [81]. In cancer research, spatial transcriptomics (ST) holds particular promise for characterizing the tumor microenvironment, a key element in understanding disease progression and treatment response [133, 97, 140, 151]. With the large amount of transcriptomics data generated by a single ST sample (e.g., >10 million transcripts are detected in a typical 10x Genomics Xenium assay), computational methods are often used to uncover promising biomarkers, such as employing clustering methods for cell phenotyping [122].

However, high costs and rapidly evolving technology have constrained computational methods to narrow tasks and data cohorts of only a few patients [103, 63, 135]. Consequently, we observe a lack

---

[*] Equal contribution

38th Conference on Neural Information Processing Systems (NeurIPS 2024) Track on Datasets and Benchmarks.
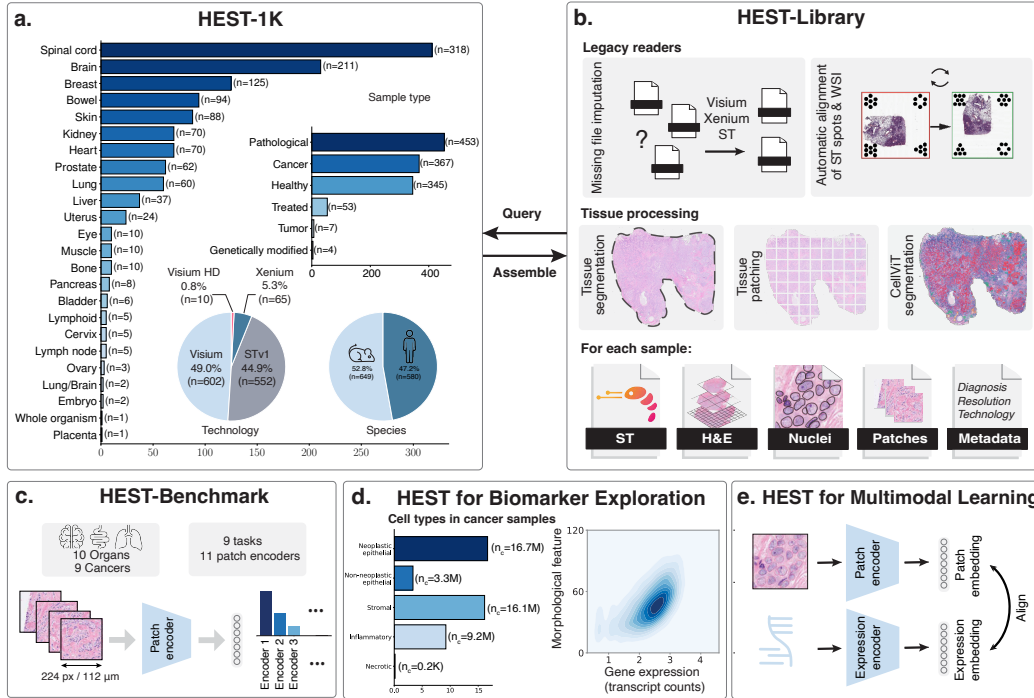
Figure 1: **The HEST environment. a.** Overview of HEST-1k, a dataset of n=1,229 paired spatial transcriptomics, H&E-stained whole-slide images and metadata. "Pathological" cases refer to non-tumor/non-cancer samples; "Tumor" refers to non-cancer samples. **b.** Overview of HEST-Library functionalities. **c., d., e.** Applications of HEST-1k include benchmarking foundation models for histology (**c.**), biomarker exploration (**d.**) and multimodal representation learning (**e.**).

of standardized resources and unified formats for handling ST, which limits the development of deep learning models on a large scale [109]. In addition, the underlying tissue morphology, traditionally visualized in hematoxylin and eosin (H&E)-stained tissue sections (whole-slide images, WSIs), is often overlooked in ST studies, despite encoding valuable information. In particular, pairs of ST and WSI enable analyzing expression changes in their morphological context, which may facilitate the identification of morphological biomarkers (e.g., changes in nuclear shape) that correspond to gene regulation patterns. Alternatively, pairs of ST and WSI can enable multimodal tissue representation learning for joint modeling of the morphomolecular signature of tissue at a scale and resolution beyond bulk RNA sequencing [28]. Finally, the development of "foundation models" for encoding histopathology images [147, 42, 29, 59, 88] has increased the need for new, diverse, and challenging benchmarks beyond diagnostic tasks. Using ST, new tasks can be defined to predict gene expression changes from histology.

Here, we introduce HEST-1k, a collection of paired ST and H&E-stained WSIs curated from public and internal cohorts (Figure 1.a). HEST-1k comprises 1,229 samples from 153 cohorts encompassing 26 organs, two species (*Homo Sapiens* and *Mus Musculus*), and 367 cancer samples from 25 different subtypes. Processing all samples in HEST-1k resulted in 2.1 million expression–morphology pairs and 76 million detected nuclei. With new cohorts frequently made public, we also introduce the HEST-Library, a Python package for interacting with HEST-1k data and assembling new samples as they become available (Figure 1.b). We highlight the potential of HEST-1k through three use cases: (i) benchmarking foundation models for histology using the HEST-Benchmark, a set of nine tasks (eight human cancer types and nine organs) for gene expression prediction from histology and evaluated on eleven state-of-the-art models (Figure 1.c), (ii) a proof-of-concept demonstrating the use of HEST-1k for biomarker characterization (Figure 1.d), and (iii) a proof-of-concept for expression-guided fine-tuning of foundation models for histology (Figure 1.e).

2

2 Related work

Libraries for ST analysis. Libraries to process, visualize, and analyze ST have been built around two core pipelines: *Scanpy* [149] (and the ANNDATA format) in Python and *Seurat* [24] in R. *Scanpy* has served as the foundation for several subsequent developments such as *Squidpy* [115] for spatial data exploration at cellular-, gene-, and morphological-level, *SpatialData* [95] for multi-technology integration and deep learning interfacing, *STlearn* [119] for cell-cell interactions and spatiotemporal trajectory analyses, and SOPA [21] for designing multistep pipelines. In R, *Seurat* [24] has been consolidated with packages such as *BayeSpace* [159] for clustering and spot super-resolution, and *Giotto suites* [26] for preprocessing, data integration and visualization of multiple ST technologies. 10x Genomics also includes proprietary software analytics through the Xenium and Visium Explorer pipelines for multimodal visualization, nuclear segmentation, and cell deconvolution. However, none of these pipelines were designed to handle the diversity of legacy data, where datasets can suffer from missing or incorrect data, such as alignment mismatches, incorrect pixel resolution, inconsistent image file formats, etc.

Molecular profile prediction from H&E. Molecular profiling from histology images has been explored both at (1) *slide-level* to predict bulk molecular status/changes from a WSI and at (2) *patch-level* to predict local molecular status/changes from regions-of-interest. (1) Slide-level profiling has been explored to predict the gene mutations [125, 143, 72, 38, 45, 145, 86], microsatellite instability [143, 71], and gene expression changes [127, 39, 55, 3], among others. The motivation is two-fold: First, patient screening to substitute or complement costly clinical molecular assays, and second, to identify morphological correlates of molecular alterations for discovering novel biomarkers. Such studies can be conducted on large patient cohorts as they mainly rely on data generated by the routine clinical workflow (e.g., using TCGA cohorts with >11,000 cases from 33 cancer types). (2) With ST, several works have explored predicting expression changes from regions-of-interest [55, 152, 116, 160, 121, 108, 107, 32, 144]. Due to limited cohort sizes (typically one to ten patients), transfer learning has become the norm using pretrained models based on ConvNets [55] or Vision Transformers [152, 116]. Due to the inherent noise found in transcriptomic measurements, several methods have been developed for integrating context that can account for global and local information from surrounding ST spots [58, 32, 160, 144]. While recent technologies offer near-single-cell resolution (such as Visium HD and Xenium), legacy assays operate at a more coarse resolution, which can be upsampled using super-resolution techniques [159, 20, 157]. The potential clinical and research implications of such methods are still being explored, with HEST-1k potentially catalyzing their large-scale development.

Foundation models in pathology. A fundamental task in computational pathology is to extract *general-purpose* embeddings of image patches (typically 256×256 to 512×512-pixel regions) that can then be used for downstream tasks, such as diagnosis or prognosis prediction. To achieve this, self-supervised learning (SSL) has been extensively applied [75, 42, 142, 147, 68, 33, 146, 11, 29, 88, 65, 64], such as based on the DINOv2 framework [112]. General-purpose patch encoders are trained on increasingly large and diverse patient cohorts (e.g., UNI [29] uses a ViT-Large trained on 100k WSIs, Virchow [142] uses a ViT-Huge trained on more than 1.5M WSIs). Recently, vision-language encoders designed for pathology have also been proposed [46, 92, 59, 88, 87] and rely on large-scale paired data scraped from social media, textbooks, or publications. As the number of such models rapidly increases, new, diverse, and challenging benchmarks are needed to replace or complement well-established tasks where performance has saturated. HEST-Benchmark aims to address this by offering a set of nine patch-level tasks for gene expression prediction from histology.

Patch-level benchmarks in histopathology. Early task and dataset contributions in computational pathology revolved around classifying small regions of interest. Over the years, a variety of benchmarks have been established: In prostate cancer, Gleason grading at pixel- and patch-level has been widely explored, with public resources such as AGGC [60], DiagSet [77], and SICAPv2 [129]. In colorectal cancer, datasets have been proposed for tissue classification, such as HunCRC [118], UniToPatho [15], MHIST [148], and CRC-100k [73]. In breast cancer, morphological subtyping has been vastly explored (e.g., for atypical ductal hyperplasia detection), such as BACH [8], BRACS [22], and BreakHis [131], and for lymph node metastasis detection with Patch CAMELYON (pCAM) [137], respectively. However, the performance on many of these datasets has saturated; for instance, Gleason scoring reaches similar or better performance than pathologists [23], which limits objective comparisons of new methods and hinders well-informed model selection for developing better features.

3

Instead, HEST-Benchmark provides a collection of diverse and challenging tasks that enable assessing the predictive capabilities of foundation models for histology.

3 HEST-1k Dataset

We present HEST-1k, a dataset of paired ST, H&E-stained WSIs, and metadata (Figure 1.a). To this end, we extracted all publicly available cohorts that provide ST with H&E-stained whole-slide images. Specifically, we harvested data from 10x Genomics public datasets (TENX)[1], Mendeley (MEND)[2], Spatial-Research (SPA)[3], Zenodo (ZEN)[4], the National Center for Biotechnology Information (NCBI)[5], GitHub[6], the Human Cell Atlas[7], BioStudies[8], HTAN[9], and internal data cohorts. A summary of all sources is provided in Appendix Table A1 with specifics in Appendix Table A2, A4, A6,A7,A8,A9, and A10.

3.1 Metadata

As spatial transcriptomics experiments were not intended for large-scale computational research, they are provided in various formats (e.g., images can be in JPG or TIFF format, with or without cross-modal alignment files) and resolutions. We unified all data with comprehensive metadata with *generic-*, *histology-*, and *expression*-related descriptors for all samples. **Generic:** We provide the reference to the original publication, download link, year of publication, license, and sample species. Each sample is then categorized as either healthy, cancer, tumor (non-cancer), treated (which refers to a post-compound administration), genetically modified (mostly knock-out mouse samples), or pathological (i.e., non-tumorous with extra specification). All cancer samples were unified using the OncoTree code, a taxonomy of cancer types provided by the Memorial Sloan Kettering Cancer Center[10]. Finally, we provide the organ using the highest level of the OncoTree taxonomy as a reference. **Expression:** We report the number of genes and spots per sample, the spot resolution and spacing, the total number of reads, and the mean number of reads per spot. We additionally provide the transcriptomic technology (ST, Visum, Visium HD, or Xenium). **Histology:** We provide the image resolution (in μm/pixel) and magnification as $10\times$ (1.15 to 0.8 μm/px), $20\times$ (0.8 to 0.4 μm/px) and $40\times$ (0.4 to 0.1 μm/px). All images with a pixel size higher than 1.15 μm/px were discarded to ensure an acceptable image quality. In addition, we provide the image size at the highest resolution and the tissue preparation protocol (frozen or formalin-fixed paraffin-embedded, FFPE).

3.2 Histology

All tissue sections were normalized and transformed into a generic TIFF object, a pyramidal image that can easily be integrated into computational frameworks using OPENSLIDE or viewers such as QUPATH [14]. In addition, we provide a contour object that delineates all the tissue regions identified in the image. We developed a robust tissue *vs.* background detection method where we fine-tuned a DeepLabV3 [27] model with an ImageNet-pretrained ResNet50 backbone on a set of annotated segmentation regions (including pen marks, fiducials, multiple stains, artifacts, etc.). From the tissue segmentation, we extracted 224×224-pixel patches at $20\times$ magnification around each spot. For Xenium samples, we generated "pseudo-Visium" spots by pooling transcripts on 55×55-μm patches without spacing. This yielded 2.1 million valid patches for which a corresponding expression profile was derived. Such patching can readily be used for various downstream tasks, such as employed in the **HEST-Benchmark** or multimodal fine-tuning of foundation models for histology (Section 5 and 7).

3.3 Nuclear segmentation and classification

In addition to patching, we include nuclear segmentation that delineates each nucleus identified in all slides from HEST-1k. We used CellViT [61], a state-of-the-art nuclear segmentation model that was trained on the PanNuke dataset [47, 48]. CellViT enables joint instance segmentation and classification of each nucleus into five classes: neoplastic epithelial, non-neoplastic epithelial, inflammatory, stromal, and necrotic. On average, we identified 62.1k nuclei per slide, for a total

---

[1] https://www.10xgenomics.com/datasets   [2] https://data.mendeley.com/   [3] https://www.spatialresearch.org/
[4] https://zenodo.org/   [5] https://www.ncbi.nlm.nih.gov/gds/   [6] https://github.com/
[7] https://data.humancellatlas.org/   [8] https://www.ebi.ac.uk/biostudies/   [9] https://humantumoratlas.org/
[10] https://oncotree.mskcc.org

4

of 76.4 million nuclei identified across all samples. Among those, 17.6 million are classified as neoplastic, 21.5 million as stromal, 4.9 million as normal epithelial, 15.4 million as inflammatory, and 76 thousand as necrotic. The resulting nuclear segmentation and classification can easily be visualized using QUPATH (using geojson) or loaded as Python/R objects (using JSON). For all Xenium samples, we additionally provide the nuclear and cell segmentation derived from the DAPI staining finely aligned with the H&E slide.

3.4 Gene expression

All expression data were unified in a ANNDATA object that can be loaded with *scanpy*. ANNDATA encodes the gene names (as *var*) and number of spots (as *obs*). Each entry represents the raw transcript counts of a gene in a given spot. No additional normalization was conducted, and we let users explore various normalization strategies based on needs, e.g., using total count normalization, log-normalization, etc. In addition, we include metadata to specify the number of genes, the gene panel, and the tissue site. For all Xenium samples, we also provide the list of all measured transcripts with their exact 2D position in the tissue (aligned with the H&E slide).

To use the expression in tandem with the WSI, an alignment file describing the mapping between the image and the spots is needed. However, relying on publicly available alignment information brings three challenges: (1) most datasets report alignment with respect to a low-resolution version of the image, (2) they are not standardized, and (3) alignment quality can be low. To address these limitations, we re-aligned all samples under the same unified format between the WSI and the corresponding expression profile. For all Visium samples, we developed an automatic alignment pipeline based on fiducial detection (see Section 4) and embedded the alignment in the *scanpy* object. For all Xenium samples, we used the publicly available VALIS [50] pipeline for fine-grained image registration to align the DAPI image (aligned with the transcripts by design) and the H&E slide.

4   HEST-Library

The HEST-Library is built around *scanpy* and ANNDATA. At its core, the HEST-Library enables (1) assembling and querying HEST-1k, (2) visualizing and mitigating batch effects, and (3) running the HEST-Benchmark (Section 5). We describe its core functionalities, particularly for unifying legacy data.

Conversion to generic TIFF. We integrate functions to convert a WSI from common formats found in public ST datasets (e.g., OME.TIF, JPG, BigTIFF, etc.) to a pyramidal generic TIFF format. Pyramidal formats offer seamless integration with OPENSLIDE (commonly used in computational pathology pipelines) and QUPATH (open-access software for WSI visualization and annotation).

Automatic alignment in Visium. Spot alignment is crucial to ensure an accurate match between the ST spots and the WSI. While software such as LoupeBrowser enables manual alignment using fiducials (i.e., reference markers placed at the corners of the capture area), the process remains time-consuming when processing large batches of samples. Instead, we implemented an automatic fiducial detection algorithm based on YOLOv8 [123] for processing Visium samples (Appendix Figure 5). Specifically, we manually annotated 119 fiducial regions that we further augmented using tissue and fiducial mixing. We then fine-tuned YOLOv8 pretrained on the COCO dataset. In early versions that do not provide corner fiducials (e.g., STv1), we realigned using the provided spot position files. In Xenium, we use VALIS [50] to register the DAPI staining (aligned with the transcripts) with the H&E image.

Automatic detection of image resolution. From the alignment and the spot resolution, we can infer the exact pixel size. To this end, we compute the distance in pixel between two neighboring spots and leverage the known inter-spot distance in μm to estimate the pixel width in μm/px. For Xenium samples, we use the H&E alignment file provided as part of the assay, which provides an affine transformation from the DAPI-stained image (with known pixel size) to the H&E image. We then compared the self-reported image resolution and our re-estimations to manually inspect and correct discrepancies.

Conversion to ANNDATA. ST data is provided in multiple formats, such as CSV, MEX, TXT, h5, etc. We provide functions to unify a large set of existing formats into a ANNDATA object that stores the

5

raw transcript counts as a matrix of genes by the number of spots, in addition to metadata about the samples (e.g., the (x,y) coordinates of each spot, the pixel resolution, etc.).

Tissue segmentation and patching. We provide a tissue segmentation pipeline optimized for Visium/Xenium images. The segmentation can then automatically tessellate the tissue into fixed-size image patches at a predefined resolution (expressed in μm/px) around each spot.

Automatic HEST-1k download. To facilitate downloading part or all of the HEST-1k dataset (over >1TB), we implemented an easy download option where the user can specify entries of the metadata, for instance, to query all human invasive breast cancer cases.

Batch effect visualization and mitigation. We provide functions to help visualize batch effects using dimensionality reduction techniques with user-prompted stratification (e.g., tissue site, institution, disease, etc.). In addition, we provide a wrapper of well-established batch effect mitigation strategies (namely ComBat [158], Harmony [76] and matching mutual nearest neighbors [54]), which can be applied to a list of HEST samples.

5 HEST-Benchmark

From HEST-1k, we curated the HEST-Benchmark, a set of nine tasks for gene expression prediction from histology in human cancer samples. The goal is two-fold: (i) benchmarking foundation models for histology under a diverse and challenging benchmark and (ii) understanding the predictive capabilities of state-of-the-art models in predicting expression from morphology. Compared to existing tasks (e.g., Camelyon16 [18]), the HEST-Benchmark brings increased morphological diversity and more complex challenges, particularly with the inherent difficulty of expression prediction.

5.1 Task definition

We define nine tasks with data from eight human cancers and nine organs (eight primary and one metastatic dataset), which include **invasive ductal carcinoma** (breast cancer, IDC, Task 1), **prostate adenocarcinoma** (prostate cancer, PRAD, Task 2), **pancreatic adenocarcinoma** (pancreatic cancer, PAAD, Task 3), **skin cutaneous melanoma** (skin cancer, SKCM, Task 4), **colonic adenocarcinoma** (colon cancer, COAD, Task 5), **rectal adenocarcinoma** (rectum cancer, READ, Task 6), **clear cell renal cell carcinoma** (kidney cancer, ccRCC, Task 7), **lung adenocarcinoma** (lung cancer, LUAD, Task 8), and **axillary lymph nodes in IDC** (metastatic, LYMPH-IDC, Task 9). Additional information is provided in Appendix Table A11.

For each task, we predict the expression of the top 50 genes with the highest normalized variance across all samples from 112×112 μm H&E regions (equivalent to 224×224-pixel patches at $20\times$). To avoid train/test patient-level data leakage, we use patient-stratified splits, resulting in a k-fold cross-validation, where k is the number of patients. In ccRCC, we use $k/2$-fold cross-validation due to the large number of patients.

5.2 Evaluating foundation model for pathology

We use the HEST-Benchmark to evaluate 11 foundation models for pathology. Namely, **ResNet50 (IN) [90]** (ImageNet pretrained), **CTransPath [146]** (adapted MoCov3 pretrained on TCGA and PAIP), **Remedis [11]** (SimCLR [30] pretrained on TCGA), **Phikon [42]** (iBOT pretrained on TCGA), **UNI [29]** (DINOv2 ViT-Large pretrained on internal hospital data and GTEx), **CONCH [88]** (visual-language model using CoCa pretrained on captions from publications and educational resources), **GigaPath** [154] (DINOv2 ViT-giant pretrained on proprietary data), **Virchow** [142] (DINOv2 ViT-Huge pretrained on proprietary data), **Virchow 2** [162] (DINOv2 ViT-Huge pretrained on proprietary data), **H-Optimus-0** (DINOv2 ViT-giant pretrained on proprietary data), and **UNIv1.5** (DINOv2 ViT-giant pretrained on public and proprietary data). Additional information is provided in Table A12 and Appendix C.3.

We learn a regression model to map model-specific patch embeddings (512 to 2,048 dimensions) to the log1p-normalized expression of the top 50 highly variable genes. All tasks are evaluated using the Pearson correlation between the predicted and measured gene expression. We report mean and standard deviation across all folds (or patients). All experiments were run on a single NVIDIA 3090 GPU. We report performance using three downstream regression models: (i) PCA-reduced

6

Table 1: **Comparison of 11 patch encoders evaluated on the HEST-Benchmark.** Reported results are based on PCA with 256 factors followed by a ridge regression. Model performance measured with Pearson correlation. Standard deviation is reported across all folds (i.e., patients). Best is **bold**, second best is underlined.

	IDC	PRAD	PAAD	SKCM	COAD	READ	ccRCC	LUAD	LYMPH IDC	Average
ResNet50 (IN)	0.4741	0.3075	0.3889	0.4822	0.2528	0.0812	0.2231	0.4917	0.2322	0.326
	± 0.047	± 0.0309	± 0.0754	± 0.1141	± 0.0372	± 0.0517	± 0.0554	± 0.0119	± 0.0491	
CTransPath	0.511	0.3427	0.4378	0.5106	0.2285	0.11	0.2279	0.4985	0.2353	0.3447
	± 0.0531	± 0.0458	± 0.0664	± 0.0827	± 0.0557	± 0.0764	± 0.0475	± 0.0414	± 0.0477	
Phikon	0.5327	0.342	0.4432	0.5355	0.2585	0.1517	0.2423	0.5468	0.2373	0.3656
	± 0.0914	± 0.0767	± 0.0684	± 0.0549	± 0.0056	± 0.0822	± 0.0263	± 0.0045	± 0.0457	
CONCH	0.5363	0.3548	0.4475	0.5791	0.2533	0.1674	0.2179	0.5312	0.2507	0.3709
	± 0.0842	± 0.0099	± 0.0729	± 0.0542	± 0.0075	± 0.0476	± 0.0353	± 0.0107	± 0.042	
REMEDIS	0.529	0.3471	0.4644	0.5818	0.2856	0.1145	0.2647	0.5336	0.2473	0.3742
	± 0.069	± 0.0074	± 0.0722	± 0.0421	± 0.02	± 0.0987	± 0.0539		± 0.0585	
GigaPath	0.5508	<u>0.3708</u>	0.4768	0.5538	<u>0.301</u>	0.186	0.2391	0.5399	0.2493	0.3853
	± 0.0726	± 0.021	± 0.0489	± 0.0586	± 0.0145	± 0.0704	± 0.0364	± 0.0369	± 0.0522	
UNI	0.5702	0.314	0.4764	0.6254	0.263	0.1762	0.2427	0.5511	0.2565	0.3862
	± 0.0833	± 0.0715	± 0.0687	± 0.0338	± 0.0311	± 0.0565	± 0.0368	± 0.0198	± 0.0436	
Virchow	0.5702	0.3309	0.4875	0.6088	**0.311**	0.2019	0.2637	0.5459	0.2594	0.3977
	± 0.0939	± 0.0081	± 0.0412	± 0.0733	± 0.0083	± 0.0467	± 0.039	± 0.0262	± 0.043	
Virchow2	0.5922	0.3465	0.4661	0.6174	0.2578	0.2084	**0.2788**	**0.5605**	0.2582	0.3984
	± 0.0814	± 0.029	± 0.0676	± 0.0174	± 0.0189	± 0.0502	± 0.0516	± 0.0172	± 0.0296	
UNIv1.5	**0.5989**	0.3645	<u>0.4902</u>	<u>0.6401</u>	0.2925	<u>0.2240</u>	0.2522	<u>0.5586</u>	**0.2597**	<u>0.4090</u>
	± 0.0842	± 0.0308	± 0.0502	± 0.041	± 0.0142	± 0.0378	± 0.04	± 0.026	± 0.0424	
H-Optimus-0	<u>0.5982</u>	**0.385**	**0.4932**	**0.6432**	0.2991	**0.2292**	<u>0.2654</u>	0.5582	<u>0.2595</u>	**0.4146**
	± 0.0843	± 0.0008	± 0.0443	± 0.0668	± 0.0007	± 0.041	± 0.0309	± 0.0324	± 0.04	

embeddings (with n=256 factors) followed by Ridge regression trained with adaptive regularization as shown in Table 1, (ii) Ridge regression model as shown in Appendix Table A13, and (iii) an XGBoost regression model with 100 estimators and a maximum depth of 3 as shown in Appendix Table A14. Our main results are reported using PCA+Ridge (i) and XGBoost (iii). Directly applying Ridge regression may unfairly penalize models with larger embedding dimensions. To guarantee a fairer and more objective comparison, we chose to utilize PCA reduction.

5.3 Scaling laws in HEST-Benchmark

Overall, H-Optimus-0 brings the best average Pearson correlation in both PCA+Ridge and XGBoost evaluation, outperforming the second-best model, UNIv1.5, by 0.56% and 0.69%, respectively. ResNet50 (IN), the only model that was not pretrained on histology images, leads to the lowest performance in both PCA+Ridge and XGBoost. Legacy domain-specific models, such as CTransPath, are outperformed by all recent models, including UNIv1.5, UNI, GigaPath, Virchow, and H-Optimus-0. The disparity between the top and bottom domain-specific models is notable, showing an absolute improvement of 7.0% for PCA+Ridge and 4.8% for XGBoost. When inspecting individual performance, we observe large differences across tasks from 0.6432 Pearson correlation in SKCM to 0.2292 in READ for H-Optimus-0 evaluated using PCA+Ridge.

Model scaling law. By inspecting the number of trainable parameters within the vision encoder for each model, we can describe how model size influences performance (measured with average Pearson correlation across all tasks, Figure 2.a). Performance increases with model size following a logarithmic scaling law (Pearson correlation of R=0.81, P-value<0.01). Models considered "parameter-efficient" are represented on top of the log-transformed linear regression line (e.g., CONCH, UNIv1.5, and H-Optimus-0). This observation suggests a trade-off between downstream performance and model size. Despite the observation of a model scaling law, significant variations in performance among models of identical size persist, such as between H-Optimus-0 and GigaPath, both of which are ViT-giant models with over one billion parameters.

Data scaling law. We further explored how the number of training samples used for pretraining each model (i.e., the number of image patches) affects performance. We observe that increasing the number of patches moderately correlates with the average performance (Pearson correlation of R=0.48, P-value=0.13). This correlation is weaker than model size, which we hypothesize is due to this analysis overlooking both the absolute number of WSIs used for pretraining (image patches are not independently and identically distributed per WSI) and disparities of morphological variety among WSIs (e.g., staining variation, disease diversity, artifacts, etc.).

7
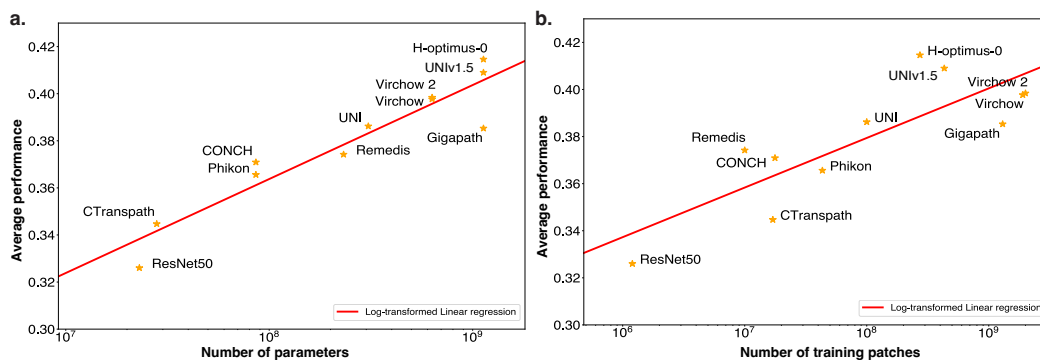
Figure 2: **Scaling laws in HEST-Benchmark. a.** Model scaling law comparing the number of training parameters in the vision encoder (log-scale) and the average performance on the HEST-Benchmark. Pearson correlation between parameters and performance of R=0.81 (P-value < 0.01). **b.** Data scaling law comparing the number of image patches used for pretraining (log-scale) and the average performance on the HEST-Benchmark. Pearson correlation between number of patches and performance of R=0.48 (P-value=0.13).

Overall, HEST-Benchmark brings new insights into the performance of foundation models for pathology. We observe that (1) Scaling the model size strongly correlates with average performance, but the gains grow logarithmically with the number of trainable parameters. (2) Scaling the number of training patches weakly correlates with a higher performance (also on a logarithmic scale). (3) Performance remains low for some tasks (e.g., READ and ccRCC), which suggests that (i) the morphology might not be as reflective of gene expression for some cancer types or (ii) some cohorts have more noise than others (e.g., due to batch effects, low sensitivity, dropout events, or spillover between adjacent spots).

6   HEST for biomarker exploration

HEST-1k also enables the analysis of interactions and correlations between tissue morphology (as seen in H&E) and local gene expression (as provided in ST). Here, we showcase the capabilities of HEST-1k (1) by studying morphological correlates of expression changes in invasive breast cancer and (2) by visualizing tumor heterogeneity both on the morphological and molecular sides. Specifically, we focus on invasive ductal carcinoma (IDC) samples imaged with Xenium. Using CellViT nuclear segmentation and classification, we identified neoplastic nuclei (exemplified in two samples: Figure 3.a with n=168,033 nuclei and Appendix Figure 6.a with n=342,018 nuclei). We then overlay the WSI with the expression of specific genes, such as *GATA3*, a known prognostic gene in breast cancer [102](Figure 3.b). This qualitatively shows that high *GATA3* expression is associated with cancerous regions and reveals heterogeneity within invasive regions (e.g., the right-most region shows higher expression of *GATA3* than the rest of the tumor, Figure 3.b). Using the nuclear segmentation, we can compute human-interpretable features related to nuclear size (area, perimeter, major axis length, minor axis length, and equivalent diameter), topology and shape (roundness, ellipticity, eccentricity, extent, and roughness), and cell distribution (cell density and crowdedness). A heatmap of the nuclear area of neoplastic cells also indicates morphological heterogeneity among neoplastic regions (Figure 3.c,d). Regions with a high nuclear area and elevated GATA3 expression notably overlap, suggesting that this tumor exhibits molecular heterogeneity, which to some degree is morphologically expressed.

To investigate this hypothesis, we measured the Pearson correlation between the expression of *GATA3* and nuclear area in neoplastic cells (Figure 3.e). We observe a moderate correlation (R=0.47, P-value$< 10^{-4}$), which is also observed in other genes and morphological features (Figure 3.f). Overall, out of the 12 human-interpretable features we analyzed, we found the highest association with gene expression for size-related features, while features involving topology, shape, and cell distribution had a lower correlation (R<0.2). A similar analysis in another IDC sample (Appendix Figure 6.b,c) further asserted these observations. In particular, we found the highest associations between nuclear size and expression for the genes *FLNB* (R=0.45, P-value$< 10^{-4}$) and *TPD52* (R=0.47, P-value$< 10^{-4}$), both

8
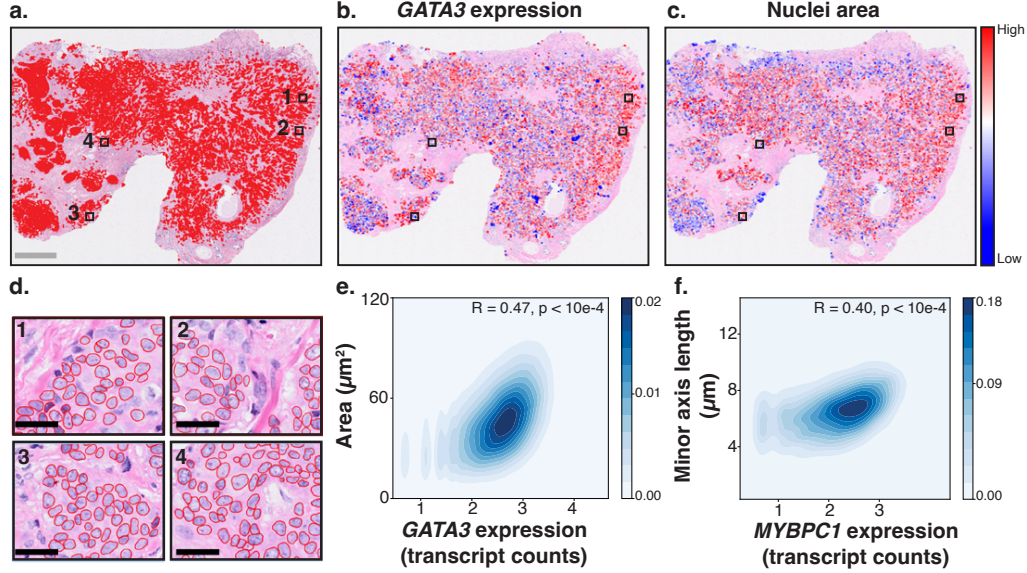
Figure 3: **HEST for biomarker exploration: Analysis of an invasive ductal carcinoma sample imaged with Xenium. a.** IDC Xenium sample with neoplastic nuclei overlaid in red ($n_c = 168,033$ detected nuclei). Gray scale bar represents 2 mm. **b.** Heatmap of Xenium expression of gene *GATA3*. Blue and red values indicate above and below the mean (in white), respectively. **c.** Heatmap of neoplastic nuclear area. **d.** Four randomly selected regions with CellViT segmentation of the neoplastic nuclei. Black scale bar represents 30 μm. **e., f.** Correlation between nuclear area and *GATA3*, and minor axis length and *MYBPC1*.

involved in breast tumor growth and proliferation [13, 124], and *FOXA1* (R=0.47, P-value$< 10^{-4}$), a known prognostic factor associated with better survival [12, 150].

Such analysis highlights how HEST-1k can be used to identify fine-grained morphological correlates of expression. Similar approaches can be used to characterize morphological and molecular tumor heterogeneity at a larger scale.

7   HEST for multimodal representation learning

Access to spatially-resolved expression–morphology pairs unlocks new directions for multimodal representation learning. Several problem statements can be explored, such as cross-modal alignment and retrieval, multimodal fusion, etc. Here, we fine-tune CONCH [88] (ViT-Base model) on five Xenium invasive breast cancer cases (four ductal and one lobular case) using multimodal contrastive alignment. We hypothesize that the resulting breast cancer-specific patch encoder, termed CONCH-FT, can better encode the underlying molecular landscape associated with disease-specific morphologies. To validate the hypothesis, CONCH-FT is benchmarked on an independent breast cancer cohort for molecular subtyping against its non-finetuned version.

Specifically, for each Xenium sample, we extract 112×112-μm image patches centered around each spot at 20× magnification (0.5μm/px). This yields 47,051 pairs of 224×224-pixel patches and corresponding expression profile (n=238 common genes in the panel, log1p normalized). We then embed the data using modality-specific encoders: the image patches using a pretrained CONCH model and the expression data using a 3-layer MLP (normalized expression data are encoded as tabular data). The modality-specific embeddings are then aligned using a contrastive objective, i.e., InfoNCE loss [111] by fine-tuning the image encoder and training the expression encoder from scratch. To mitigate over-fitting, we use the following training recipe: (1) Finetune only the last 3 layers of CONCH, (2) employ a layer-wise learning decay factor of 0.7, and (3) employ patch-level image augmentation. Additional details are provided in Appendix.

We evaluate the resulting CONCH-FT model to predict ER, PR, and HER2 expression status (binary) from WSIs in the BCNB dataset [153] (n=1,058 WSIs). To generate a slide representation for a

9

Table 2: **CONCH fine-tuning on invasive breast cancer.** Logistic regression evaluation for ER/PR/HER2 status on BCNB (binary task, n=1,058 WSIs). A WSI is represented by the average of the patch embeddings within each WSI. We report the mean ± standard deviation computed over all folds (or patients) for ROC-AUC (AUC) and balanced accuracy (Bal.acc.). Best is **bold**.

	Rank	ER		PR		HER2	
		AUC	Bal.acc.	AUC	Bal.acc.	AUC	Bal.acc.
CONCH	144.66	0.881	0.745	0.810	0.698	0.715	**0.624**
CONCH-FT	**146.47**	**0.884**	**0.752**	**0.818**	**0.714**	**0.724**	0.615

WSI, we take the average of the patch embedding in the WSI (mean pooling), which is subsequently mapped to the expression status using logistic regression (Table 2). The simple mean pooling approach to embedding the slide without additional fine-tuning on the downstream tasks highlights the expressivity of the learned latent space. We observe that CONCH-FT outperforms CONCH on most metrics, demonstrating that pan-tissue histology patch encoders can be further fine-tuned to obtain better tissue-specific patch encoders. This is further validated by the larger rank induced by the patch embedding space [49] for CONCH-FT, suggesting better expressivity of the patch embeddings. While these results are based on only five paired WSIs, we anticipate additional benefits when training with larger disease-specific cohorts.

8 Discussion

Summary. We assembled HEST-1k, a dataset comprising paired spatial transcriptomics, H&E-stained whole-slide images, and comprehensive metadata built from public and internal cohorts. HEST-1k includes 1,229 samples, encompassing 2.1 million spots and over 76 million cells. The scale and comprehensiveness of HEST-1k, supported by the HEST-Library, enable exploring directions such as biomarker exploration and multimodal representation learning. Additionally, motivated by the need for new, diverse, and challenging patch-level benchmarks, we curated the HEST-Benchmark, a set of nine tasks covering eight cancer types and nine organs for gene expression prediction from histology. The HEST-Benchmark revealed data and model scaling laws across 11 foundation models of different dimensions and pretraining scale [88].

Limitations. Our study includes a few limitations. First, research data, such as those generated in spatial transcriptomic, are inherently noisy. While we tried to minimize "label" noise (e.g., by re-estimating image magnification and alignment, and unifying cancer samples using oncotree code taxonomy), staining and compression artifacts, varying acquisition protocols, among others, can negatively impact the quality of HEST-1k. Second, batch effects (on both the imaging and transcriptomic sides) can be significant across samples, datasets, and technologies. While this study does not explore batch effects quantification or mitigation, we provide a set of helpers in HEST-Library to let users explore this direction. Lastly, although the HEST-Library was designed for versatility, it cannot cover all existing formats and should rather be viewed as a blueprint for processing ST data in a consistent and unified manner.

Future work. Spatial transcriptomics is rapidly evolving, with new datasets frequently published. As they become available, we will keep updating HEST-1k with new resources. This study merely starts to uncover the potential of HEST-1k for advancing translational research and biomarker exploration, and we plan to explore these capabilities further. Additionally, the prospects for multimodal representation learning with HEST-1k are promising and are expected to grow with the addition of more data.

10

Acknowledgements

We thank Dr. Maxime Meylan for his insights and guidance on accessing data published in [103]. We thank Rushin Gindra for his support in inspecting HEST-1k data, reporting issues, and providing references. HEST is supported by the Brigham and Women's Hospital (BWH) President's Fund, Mass General Hospital (MGH) Pathology, and the National Institute of Health (NIH) National Institute of General Medical Sciences (NIGMS) through R35GM138216. S.J.W. is supported by the Helmholtz Association under the joint research school "Munich School for Data Science - MUDS" and the Add-on Fellowship of the Joachim Herz Foundation.

Checklist

1. Do the main claims made in the abstract, and introduction accurately reflect the paper's contributions and scope? [Yes] Each claim: HEST-1k, HEST-Library, HEST-Benchmark, HEST for biomarker exploration, and multimodal fine-tuning are supported by dedicated sections in the main text, in addition to supplementary information provided in the appendix. In addition, all the code to reproduce results is made available.

2. Did you describe the limitations of your work? [Yes] We discuss limitations in the **Discussion**.

3. Did you discuss any potential negative societal impacts of your work? [Yes] We discuss potential negative societal impacts in the section **Ethical considerations, intended usage, and license**.

4. Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

5. Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] In the abstract, we provide a link to access the HEST page on GitHub. HEST-Library includes a link to download all data and to run the HEST-Benchmark. Finally, we provide all metadata associated with HEST-1k in a CSV as part of the supplementary material.

6. Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] When relevant, we provide training details, such as in the **HEST-Benchmark**.

7. Did you report error bars? [Yes] HEST-Benchmark results include standard deviation computed from cross-validation across all patients.

8. Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]

9. If your work uses existing assets, did you cite the creators? [Yes] All public resources used in this study are cited in Appendix Table A2, A4, A6, A7, A8, A9 and A10.

10. Did you mention the license of the assets? [Yes] Metadata associated with HEST-1k includes the license under which data were originally published. We ensured that the reported license allowed the distribution and creation of derivatives of the data.

11. Did you include any new assets either in the supplemental material or as a URL? [Yes] As part of HEST-1k, we include internal datasets (see Appendix Table A9).

12. Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No] We used public resources for which the license was allowing redistributing the work. Users are welcome to inspect the individual IRBs of each publicly available resource.

13. Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] We manually ensured that none of the published and distributed data includes personally identifiable information or offensive content, such as personal health information.

11

Appendix

A Ethical considerations, intended usage and license   **12**

B Background   **12**

    B.1   Computational pathology 13

    B.2   Spatial transcriptomics (ST) 13

C HEST   **13**

    C.1   HEST-1k 13

    C.2   HEST-Library 13

    C.3   HEST-Benchmark 14

D HEST for multimodal representation learning   **16**

E HEST for discovery   **17**

F Datasheet for HEST-1k   **17**

    F.1   Motivation for dataset creation 17

    F.2   Dataset composition 18

    F.3   Data collection process 18

    F.4   Data preprocessing 19

    F.5   Dataset distribution 19

    F.6   Legal and ethical considerations 19

G Author statement   **20**

A   Ethical considerations, intended usage and license

All resources provided as part of this study are strictly for research purposes and must not be utilized to support any diagnostic procedures. Users are hereby notified that the nuclear segmentation and classification components are derived from a publicly available model. Consequently, this model should not be regarded as the definitive standard, and users should exercise particular caution when utilizing this part of the dataset. Despite our efforts to exclude sensitive information, such as patient names, addresses, and social security numbers, users are expressly prohibited from attempting to reverse engineer the data to extract confidential patient information. In the presumption that users will adhere to the aforementioned restrictions, we have not identified any potential adverse social impacts that could arise from the use of HEST-1k.

The dataset is hosted on the HuggingFace Dataset webpage. All instructions are provided in the main README of HEST-Library. From there, users can choose to download HEST-1k in its entirety or a subset (e.g., only breast cancer samples). The HEST-1k, HEST-Benchmark, and HEST library are released under the Attribution-NonCommercial-ShareAlike 4.0 International license (CC BY-NC-SA 4.0 Deed)[11].

B   Background

This study connects two fields: (1) computational pathology, which primarily uses routinely acquired clinical data to determine outcomes such as disease diagnosis from H&E-stained digitized tissue

---

[11] https://creativecommons.org/licenses/by-nc-sa/4.0/

12

sections, and (2) spatial transcriptomics, which so far has been confined to biological research and aims to identify new biomarkers predictive of disease progression or response to treatment, among others.

B.1   Computational pathology

Research in computational pathology [130] has primarily focused on classifying digitized WSIs into clinical outcomes. Unlike natural image classification tasks such as ImageNet, a WSI may reach sizes of up to 150,000 × 150,000 pixels at 20× magnification (0.5μm/pixel). The challenge of managing the large size of WSIs has been one the central themes of the field, primarily through the adoption of multiple instance learning (MIL) for weakly-supervised classification [62]. MIL employs a two-step process: (1) Initially, the tissue is segmented from the background and then tessellated into patches, usually 256 × 256 pixels, akin to an ImageNet sample, and each patch is compressed into a patch embedding using a pretrained patch encoder. (2) Subsequently, these patch embeddings are aggregated using a learnable neural network, such as an attention-based network, a graph neural network, or a Transformer, to produce a slide embedding [62, 128]. This slide embedding is then used to classify specific targets of interest, such as cancer histological subtyping, morphological subtyping, mutation prediction, or survival analysis.

Such frameworks have been shown to achieve better or similar performance than humans for Gleason grading in prostate cancer[23], metastasis detection in lymph nodes[18], determining the origin of a cancer of unknown primary[89], predicting heart transplant rejection[83], among others.

B.2   Spatial transcriptomics (ST)

ST enables the measurement of gene activity and the mapping of its corresponding location in the tissue. In this study, we collected samples from two ST paradigms: sequencing-based (ST, Visium, Visium HD) and imaging-based (Xenium).

Visium (HD) / Spatial Transcriptomics: Visium-HD and its predecessors Visium and Spatial Transcriptomics (STv1) refer to a family of sequencing-based products for spatially resolving large transcript panels, whose main difference lies in the resolution and spacing between the expression measurement, called a spot. These spots capture mRNA from tissue sections placed on the chip, and the location-specific barcodes contained in each of the spots bind to the RNA to retain spatial information. The RNA molecules are then washed off the slides and processed by a sequencing instrument. Using a sequencing-based method allows the reuse of existing sequencing instruments developed in the fields of single-cell and bulk transcriptomics, hence benefiting from existing technological advancements and allowing whole transcriptome analysis. A fundamental drawback of current sequencing-based methods is the inherent RNA resolution limitation imposed by the size of the spots (e.g., 55μm in Visium).

Xenium: Xenium is an imaging-based spatial profiling technology that offers in situ RNA capturing on tissue sections by imaging fluorescent RNA markers derived from padlock probes and rolling circle amplification chemistry. This approach provides the exact 2D location of each measured transcript. As of 2024, Xenium cannot perform whole transcriptome measurements and is limited to gene panels of up to 5,000 genes.

C   HEST

C.1   HEST-1k

We provide a comprehensive description of all publicly available and internal cohorts integrated into HEST-1k.

C.2   HEST-Library

The HEST-Library helps transform unstructured spatial transcriptomics and histology data into a unified format. An overview of the HEST-Library is provided in Figure 4. An example of fiducial detection is presented in Figure 5.

13

Table A1: **HEST-1k data overview.** All samples include a license that allows sharing and redistributing. National Center for Biotechnology Information: NCBI.

Resource	Number of datasets	Number of samples	Size (GB, raw)
10x Genomics	87	112	275
Mendeley	9	118	181
Spatial-Research	4	139	18
Zenodo	4	21	18
NCBI	43	696	298
Internal	3	28	60
Miscellaneous	4	114	147
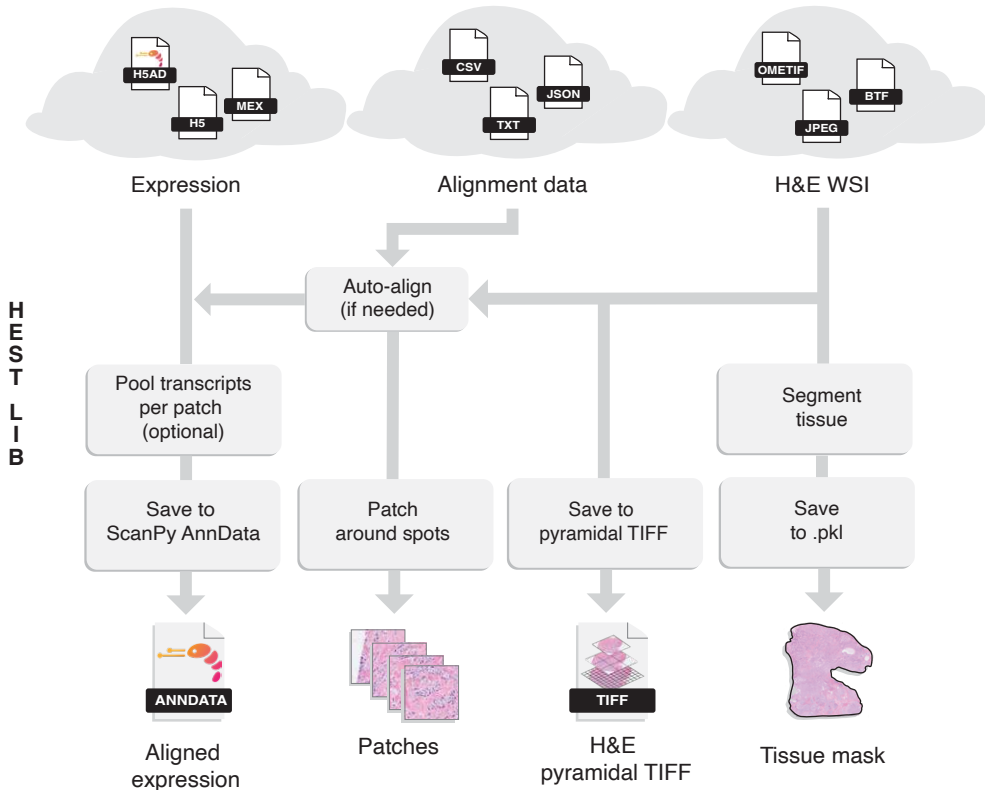
Figure 4: **Overview of HEST-Library functionalities.** HEST was designed to transform legacy data scrapped in multiple public repositories, such as NCBI, into unified HEST objects that can easily be integrated into computational pipelines.

C.3 HEST-Benchmark

Gene selection, XGBoost Forest, and Ridge regression models: We learn a regression model that maps the patch embeddings of each encoder to its corresponding gene expression profile. The XGboost model uses 100 estimators, a 0.1 learning rate, a max depth of 3, 0.8 subsampling, gamma of 0.0, regression alpha of 0.0, and regression lambda of 1.0. Additional information can be found in the XGBoost API[12]. The Ridge regression uses a fixed $L2$ regularization coefficient λ set to $100/MC$, where M is the embedding dimension and $C = 50$ is the number of targets trained with the Regularized Least-Squares Routine solver (sklearn implementation). Both regression models are trained to predict a panel constituted of the 50 most variable genes of each task. Specifically, for each

---

[12] `https://xgboost.readthedocs.io/en/stable/python/python_api.html`

14

Visum 6.5x6.5mm/11x11mm   **Spots aligned with the H&E**
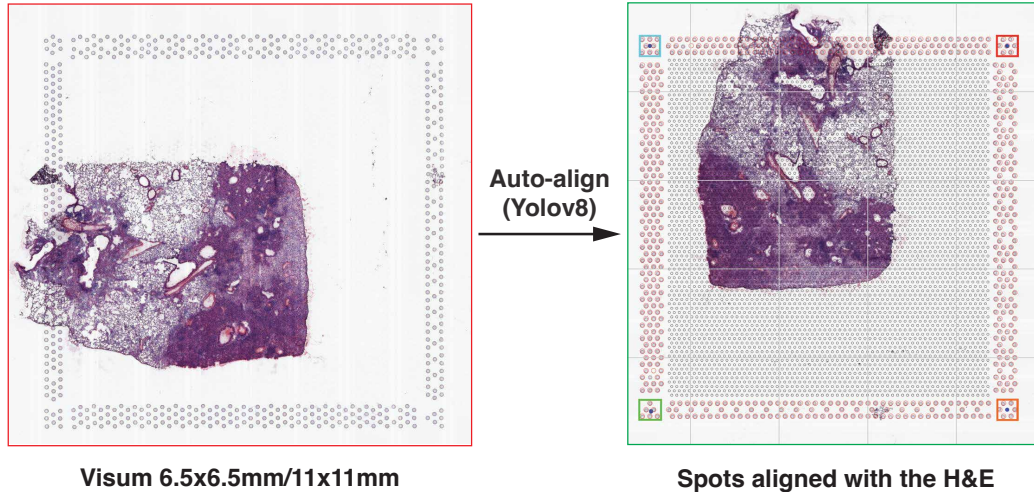
Figure 5: **Fiducial detection and automatic alignment in Visium.** Corner fiducials on 6.5×6.5mm and 11mm×11mm Visium slides are automatically detected with a finetuned Yolov8 model. The spot coordinates are then derived if at least 3 of the 4 corner fiducials are detected. This process enables automatically estimating the pixel resolution.

task, we select the 50 most variable genes across all spots and samples after excluding the genes that have non-zero counts in less than 10% of the spots.

Benchmark task description: We provide complementary information on each task introduced as part of the HEST-Benchmark.

Task 1: Prediction of expression in invasive ductal carcinoma (breast cancer, IDC). We used all publicly available Xenium samples available on 10x Genomics ("FFPE Human Breast using the Entire Sample Area", 2 patients) and two samples published in [63] (TENX95, TENX99, NCBI783, NCBI785). All samples are FFPE sections imaged with the Xenium pipeline v1.

Task 2: Prediction of expression in prostate adenocarcinoma (prostate cancer, PRAD). We used all 23 Visium samples (fresh frozen sections) from 2 patients published in [40] (MEND139 to MEND162). Both patients were diagnosed with prostatic acinar adenocarcinoma with a (4+3) Gleason score (ISUP group 4).

Task 3: Prediction of expression in pancreatic adenocarcinoma (pancreatic cancer, PAAD). We used 3 samples from 3 different patients from 10x Genomics ("FFPE Human Pancreas with Xenium Multimodal Cell Segmentation" and "Pancreatic Cancer with Xenium Human Multi-Tissue and Cancer Panel"). All samples are FFPE sections processed with Xenium pipeline v1 (TENX116, TENX126, TENX140).

Task 4: Expression prediction in skin cutaneous melanoma (skin cancer, SKCM). We used 2 samples from 2 different patients from 10x Genomics website ("Human Skin Data with Xenium Human Multi-Tissue and Cancer Panel"). All samples are FFPE sections processed with Xenium pipeline v1 (TENX115, TENX117).

Task 5: Prediction of expression in colon adenocarcinoma (colon cancer, COAD). We used 4 COAD samples from 2 different patients available on 10x Genomics (TENX111, TENX147, TENX148, TENX149). All samples are fresh frozen sections processed with Visium.

Task 6: Prediction of expression in rectal adenocarcinoma (rectum cancer, READ). We used 4 READ samples from 2 different patients published in [135]. All samples are fresh frozen sections processed with Visium (ZEN36, ZEN40, ZEN48, ZEN49).

Task 7: Prediction of expression in clear cell renal cell carcinoma (kidney cancer, ccRCC). We used the 24 ccRCC samples of 24 different patients published in [103]. All samples are fresh frozen sections processed with Visium (INT1 to INT24).

15

Task 8: Prediction of expression in lung adenocarcinoma (lung cancer, LUAD). We used 2 LUAD samples from 2 different patients from 10x genomics ("Preview Data: FFPE Human Lung Cancer with Xenium Multimodal Cell Segmentation"). All samples are fresh frozen sections processed with Xenium pipeline v1 (TENX118, TENX141).

Task 9: Prediction of expression in axillary lymph nodes in IDC patients. We used 4 axillary lymph node samples from 2 IDC patients published in [84]. All samples are fresh frozen sections processed with Visium (NCBI681, NCBI682, NCBI683, NCBI684).

We provide a brief description of each patch encoder assessed with the HEST-Benchmark.

ResNet50 (IN) [90]: This model uses a ResNet50 backbone [56] trained on ImageNet [35] (1.2 million natural images). Following prior work [90], the patch embeddings are extracted by taking the representation at the penultimate layer before final classification.

CTransPath [146]: This model uses a "Tiny" Swin Transformer backbone [85] with a window size of 14 (Swin-T/14, 28 million parameters) pretrained on TCGA and PAIP datasets (17 million images) using MoCoV3 [31].

Remedis [11]: This model uses a ResNet-152×2 (232 million parameters) initialized with the "Big Transfer"-medium protocol [74] on ImageNet-22K and pretrained with SimCLR [30] on TCGA.

Phikon [42]: This model uses a Vision Transformer-Base (ViT-B, 86 million parameters) [37] trained on TCGA data using iBOT [161].

UNI [29]: This model uses a ViT-Large (ViT-L, 307 million parameters) [37] trained on 100 million histology images (over 100,000 slides) from proprietary and public data using DINOv2 [112].

CONCH [88]: This model uses a ViT-B (86 million parameters) trained on a smaller version of UNI using iBOT, and then fine-tuned on 1.17 million histology image–caption pairs extracted from online educational and research resources using CoCa [156].

GigaPath [154]: This model uses a ViT-giant (1.13 billion parameters) trained on 1.3 billion image patches from 171,189 WSIs at $20\times$ magnification using DINOv2.

Virchow [142]: This model uses a ViT-Huge (632M parameters) trained on 2 billion image patches and 1.5M WSIs at $20\times$ magnification using DINOv2.

Virchow 2 [162]: This model uses a ViT-Huge (632M parameters) trained on 1.9B patches and 3.1M WSIs using DINOv2

H-Optimus-0: This model uses a ViT-giant (1.13B parameters) trained 273 million image patches from 500,000 WSIs at $20\times$ magnification using DINOv2.

UNIv1.5: This model uses a ViT-giant (1.13B parameters) trained on 432 million image patches from 350,000 WSIs using DINOv2.

D   HEST for multimodal representation learning

We provide additional information regarding CONCH fine-tuning using multimodal alignment. CONCH-FT model, a ViT-Base model initialized with CONCH weights, was fine-tuned for 50 epochs using a cosine learning rate scheduler, with a base learning rate of 10^{-4} for the image encoder and 10^{-3} for the expression encoder. Only the last 3 layers of the model were fine-tuned, with a layer-wise learning decay rate of 0.7. For training with the infoNCE loss, a contrastive temperature of 10^{-2} and batch size of 1,024 pairs of patch and transcriptomics were used. A combination of random horizontal/vertical flip and color jittering was employed for image patch augmentation.

The rank of the embedding space (also referred to as smooth rank measure [49]) measures the quality of the embeddings produced from encoders trained in unsupervised or self-supervised manners. Given the patch embedding matrix $H \in \mathcal{R}^{N \times d}$ and $d < N$, where N is the number of patches and d is the feature dimension, we compute the rank as the entropy of the d L1-normalized singular values of H.

16

E HEST for discovery

Cells were segmented and classified using CellViT [61]. To find the gene expression profile of each neoplastic cell, we matched each cell to its corresponding cell index in Xenium by assigning the index for which the distance between the cell centroids was the smallest. After matching all neoplastic cells, only those cells for which the assignment was unique were kept. After this filtering step, an average of 91% of the cells per sample were kept while 9% of the cells were discarded.
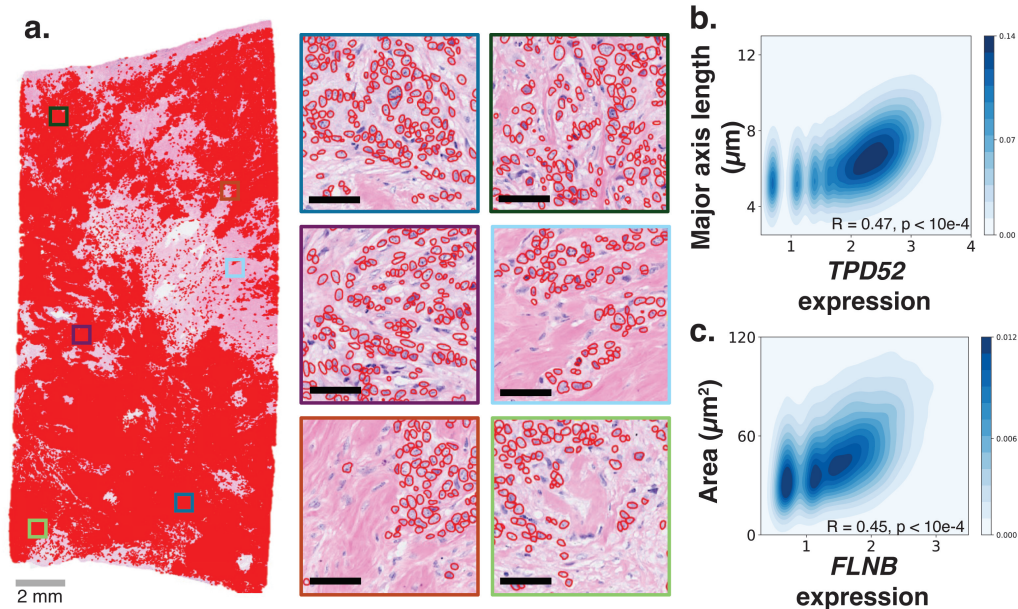
Figure 6: **HEST for biomarker discovery: Analysis of an invasive ductal carcinoma Xenium sample. a.** IDC Xenium sample with neoplastic nuclei overlaid in red (n_c=342,018 detected nuclei). Six randomly selected regions with CellViT segmentation of the neoplastic nuclei. Black scale bar represents 30 μm. **b.** Pearson correlation between the major axis length of neoplastic nuclei and the log1p-normalized expression of *TPD52*. **c.** Analogous analysis between nuclear area and *FLNB* expression.

F Datasheet for HEST-1k

We provide a DataSheet for HEST-1k that summarizes the contributions, analyses, and intended usages presented in the study.

F.1 Motivation for dataset creation

- **Why was the dataset created?** HEST-1k was designed with three key applications: (1) multimodal representation learning of histology and transcriptomics, (2) biomarker exploration and characterization, and (3) benchmarking foundation models for pathology. Despite many publicly available resources, no existing unified and user-friendly formatting was available to bring ST into the world of deep learning.

- **What (other) tasks could the dataset be used for? Are there obvious tasks for which it should not be used?** Users are welcome to introduce new, creative ways to use the dataset. However, users are not allowed to try to retrieve patient information from the existing data. A dedicated section is provided to discuss ethical considerations and intended usage.

- **Has the dataset been used for any tasks already? If so, where are the results so others can compare (e.g., links to published papers)?** The metadata attached to HEST-1k reports all samples that were made public as part of a publication (peer-reviewed or not).

17

- **Who funded the creation of the dataset?** HEST is supported by the Brigham and Women's Hospital (BWH) President's Fund, Mass General Hospital (MGH) Pathology, and the National Institute of Health (NIH) National Institute of General Medical Sciences (NIGMS) through R35GM138216.

F.2 Dataset composition

- **What are the instances?** The modalities used in this study are histopathology whole-slide images, gene expression data, and derivatives of these two modalities, such as nuclear segmentation and classification maps.

- **Are relationships between instances made explicit in the data** Each whole-slide image maps to a unique gene expression profile in an unequivocal way.

- **What data does each instance consist of?** Imaging data consists of Generic TIFF objects stored in a pyramidal format, and gene expression data consists of *scanpy* objects. Derivatives are stored in JSON files, parquet files, and Hierarchical Data Format (HDF) files.

- **Is there a label/target associated with instances? If the instances are related to people, are subpopulations identified (e.g., by age, gender, etc.), and what is their distribution?** Each sample pair (slide and expression profile) is associated with comprehensive metadata. All metadata information is thoroughly described in the main paper. Age and gender are only reported in a subset of cases.

- **Is everything included or does the data rely on external resources? (e.g., websites, tweets, datasets) If external resources, a) are there guarantees that they will exist, and remain constant, over time; b) is there an official archival version. Are there licenses, fees or rights associated with any of the data?** We provide all data as part of the HEST-1k release. In addition, a link to the original data is provided in the metadata. Each sample is associated with a license as provided by the original publication, where we ensured that the reported license allowed for distributing and creating derivatives of the data.

- **Are there recommended data splits or evaluation measures?** HEST-1k comes with the HEST-Benchmark, a series of tasks for gene expression prediction from histology images. All patient-stratified splits are specified in the attached comma-separated values (CSV) files.

- **What experiments were initially run on this dataset? Have a summary of those results and, if available, provide the link to a paper with more information here.** All experiments run with HEST-1k are described in this study. The reader can refer to the main text for a thorough description of all experiments (see **HEST-Benchmark**, **HEST for biomarker exploration**, **HEST for multimodal representation learning**).

F.3 Data collection process

- **How was the data collected?** The data were manually inspected and curated by the authors of the present study.

- **Who was involved in the data collection process?** All authors of the present study were involved in the data collection, inspection, and curation. The reader can refer to the original publication to understand how the data were originally acquired.

- **Over what time frame was the data collected? Does the collection time frame match the creation time frame?** The original data comprise publications from 2016 to 2024. As the dataset grows, more recent publications might be included in HEST-1k.

- **Does the dataset contain all possible instances? Or is it, for instance, a sample (not necessarily random) from a larger set of instances?** All pairs of gene expression data and whole-slide images of the underlying studies were included and are unique.

- **Is there information missing from the dataset and why? (this does not include intentionally dropped instances; it might include, e.g., redacted text, and withheld documents) Is this data missing because it was unavailable?** Original publications may include some missing information, such as the alignment file between the slide and the expression profile. We developed computational tools to minimize missing information and reach near-complete metadata.

18

- **Are there any known errors, sources of noise, or redundancies in the data?** All whole-slide images have been manually inspected. The quality from one sample to another varies significantly, for instance, due to poor staining, compression artifact, lower resolution, etc. Gene expression data are inherently noisy. Users can decide to apply post-hoc normalization methods to reduce noise, e.g., stain normalization on the imaging side or batch effect mitigation on the transcriptomics side.

F.4 Data preprocessing

- **What preprocessing/cleaning was done?** All whole-slide images were converted into pyramidal TIFF objects with re-estimated pixel resolution. All alignment files have been manually inspected and included if missing. All gene expression data have been transformed into *scanpy* objects following the same process.

- **Was the "raw" data saved in addition to the preprocessed/cleaned data? (e.g., to support unanticipated future uses)** Raw data are downloaded but not publicly shared. In the case of public samples, users can re-download them using the metadata provided as part of the dataset release.

- **Is the preprocessing software available?** Yes, the source code to preprocess HEST-1k is made publicly available as part of the HEST library.

F.5 Dataset distribution

- **How is the dataset distributed?** HEST-1k is distributed using HuggingFace Datasets.

- **When will the dataset be released/first distributed?** The dataset is public and can be accessed through the HuggingFace Datasets interface.

- **What license (if any) is it distributed under? Are there any copyrights on the data?** The dataset is distributed under the Attribution-NonCommercial-ShareAlike 4.0 International license (CC BY-NC-SA 4.0 Deed).

- **Are there any fees or access/export restrictions?** No access/export restrictions unless they violate the terms of the above-mentioned license (CC BY-NC-SA 4.0 Deed).

- **Who is supporting/hosting/maintaining the dataset? How does one contact the owner/curator/manager of the dataset?** The dataset is maintained by the authors of the publication.

- **Will the dataset be updated? How often and by whom? How will updates/revisions be documented and communicated (e.g., mailing list, GitHub)? Is there an erratum?** The dataset might evolve as additional samples become publicly available. Dataset versioning will be put in place.

- **If the dataset becomes obsolete how will this be communicated?** The GitHub README will be updated.

- **Is there a repository to link to any/all papers/systems that use this dataset?** There is no repository to link papers that use HEST-1k. Users are required to cite HEST-1k if they use it in their own research.

- **If others want to extend/augment/build on this dataset, is there a mechanism for them to do so? If so, is there a process for tracking/assessing the quality of those contributions. What is the process for communicating/distributing these contributions to users?** Users are welcome to contact us if they would like to provide additional data that meets our standards. We do not have a dedicated system to communicate these contributions. Newly added data will be tracked in the versioning.

F.6 Legal and ethical considerations

- **If the dataset relates to people (e.g., their attributes) or was generated by people, were they informed about the data collection? (e.g., datasets that collect writing, photos, interactions, transactions, etc.)** HEST-1k does not include patient information (such as name, address, etc.).

19

- **If it relates to other ethically protected subjects, have appropriate obligations been met? (e.g., medical data might include information collected from animals)** For animal samples (*Mus musculus* tissue), we refer to the original publication for an in-depth analysis.

- **If it relates to people, were there any ethical review applications/reviews/approvals? (e.g. Institutional Review Board applications)** For human tissue, we refer to the original publication for an in-depth analysis. Internal cohorts were ethically reviewed and collected as part of dedicated IRBs.

- **If it relates to people, could this dataset expose people to harm or legal action? (e.g., financial, social or otherwise) What was done to mitigate or reduce the potential for harm?** No, patients cannot be linked to the corresponding histology and gene expression profile.

- **If it relates to people, does it unfairly advantage or disadvantage a particular social group? In what ways? How was this mitigated?** Most datasets do not include specific demographics. When reported, we include this information in the metadata associated with each sample. To our knowledge, the representation of HEST-1k does not unfairly advantage or disadvantage a particular social group.

- **Does the dataset contain information that might be considered sensitive or confidential? (e.g., personally identifying information)** No.

- **Does the dataset contain information that might be considered inappropriate or offensive?** No.

G   Author statement

The authors of this paper bear all responsibility in case of violation of rights associated with HEST-1k, HEST-Library, and HEST-Benchmark.

20

Table A2: **Datasets gathered from 10x Genomics portal.** n: number of samples in the cohort.

Collection name	Organ	Technology	n	Num. genes
Adult Mouse Brain (FFPE)	Brain	Visium	1	19,465
Adult Mouse Brain Coronal Section (Fresh Frozen) 1	Brain	Visium	1	32,285
Adult Mouse Brain Coronal Section (Fresh Frozen) 2	Brain	Visium	1	32,285
Adult Mouse Kidney (FFPE)	Kidney	Visium	1	19,465
Adult Mouse Olfactory Bulb	Brain	Visium	1	32,285
Characterization of immune cell populations in the tumor microenvironment of colorectal cancer using high definition spatial profiling [110]	Bowel	Mixed	8	18,085
FFPE Human Breast using the Entire Sample Area	Breast	Xenium	2	541
FFPE Human Breast with Custom Add-on Panel	Breast	Xenium	2	541
FFPE Human Breast with Pre-designed Panel	Breast	Xenium	2	541
FFPE Human Pancreas with Xenium Multimodal Cell Segmentation	Pancreas	Xenium	1	541
FFPE Human Prostate Adenocarcinoma with 5K Human Pan Tissue and Pathways Panel	Prostate	Xenium	1	10,006
FFPE Human Skin Primary Dermal Melanoma with 5K Human Pan Tissue and Pathways Panel	Skin	Xenium	1	10,017
Fresh Frozen Mouse Colon with Xenium Multimodal Cell Segmentation	Bowel	Xenium	1	541
Fresh Frozen Mouse Brain Hemisphere with 5K Mouse Pan Tissue and Pathways Panel	Brain	Xenium	1	13,780
Fresh Frozen Visium on CytAssist: Human Breast Cancer, Probe-Based Whole Transcriptome Profiling	Breast	Visium	1	18,085
Fresh Frozen Visium on CytAssist: Mouse Brain, Probe-Based Whole Transcriptome Profiling	Brain	Visium	1	19,465
Human Bone and Bone Marrow Data with Custom Add-on Panel	Bone	Xenium	3	541
Human Brain Cancer, 11 mm Capture Area (FFPE)	Brain	Visium	1	18,085
Human Breast Cancer (Block A Section 1)	Breast	Visium	1	33,538
Human Breast Cancer (Block A Section 2)	Breast	Visium	1	33,538
Human Breast Cancer: Ductal Carcinoma In Situ, Invasive Carcinoma (FFPE)	Breast	Visium	1	17,943
Human Breast Cancer: Targeted, Immunology Panel	Breast	Visium	1	1,056
Human Breast Cancer: Visium Fresh Frozen, Whole Transcriptome	Breast	Visium	1	36,601
Human Breast Cancer: Whole Transcriptome Analysis	Breast	Visium	1	32,285
Human Cerebellum: Targeted, Neuroscience Panel	Brain	Visium	1	1,186
Human Cerebellum: Whole Transcriptome Analysis	Brain	Visium	1	1,186
Human Cervical Cancer (FFPE)	Cervix	Visium	1	17,943
Human Colon Preview Data (Xenium Human Colon Gene Expression Panel)	Bowel	Xenium	2	541
Human Colorectal Cancer, 11 mm Capture Area (FFPE)	Bowel	Visium	1	18,085
Human Colorectal Cancer: Targeted, Gene Signature Panel	Bowel	Visium	1	1,142
Human Colorectal Cancer: Whole Transcriptome Analysis	Bowel	Visium	1	36,601
Human Glioblastoma: Targeted, Pan-Cancer Panel	Brain	Visium	1	1,253
Human Glioblastoma: Whole Transcriptome Analysis	Brain	Visium	1	36,601
Human Heart	Heart	Visium	1	36,601
Human Heart Data with Xenium Human Multi-Tissue and Cancer Panel	Heart	Xenium	1	541
Human Intestine Cancer (FPPE)	Bowel	Visium	1	17,943
Human Kidney Preview Data (Xenium Human Multi-Tissue and Cancer Panel)	Kidney	Xenium	2	541
Human Kidney, 11 mm Capture Area (FFPE)	Kidney	Visium	1	18,085
Human Liver Data with Xenium Human Multi-Tissue and Cancer Panel	Liver	Xenium	2	541
Human Lung Cancer (FFPE)	Lung	Visium	1	18,085
Human Lung Cancer, 11 mm Capture Area (FFPE)	Lung	Visium	1	18,085
Human Lymph Node	Lymph node	Visium	1	36,601

21

Table A3: **Datasets gathered from 10x Genomics portal. Continuation.**

Collection name	Organ	Technology	n	Num. genes
Human Ovarian Cancer (FFPE)	Ovary	Visium	1	17,943
Human Ovarian Cancer, 11 mm Capture Area (FFPE)	Ovary	Visium	1	18,085
Human Prostate Cancer, Acinar Cell Carcinoma (FFPE)	Prostate	Visium	1	17,943
Human Prostate Cancer, Adenocarcinoma with Invasive Carcinoma (FFPE)	Prostate	Visium	1	17,943
Human Skin Data with Xenium Human Multi-Tissue and Cancer Panel	Skin	Xenium	2	541
Human Skin Preview Data (Xenium Human Skin Gene Expression Panel with Custom Add-On)	Skin	Xenium	1	541
Human Skin Preview Data (Xenium Human Skin Gene Expression Panel)	Skin	Xenium	1	541
Human Tonsil Data with Xenium Human Multi-Tissue and Cancer Panel	Lymph node	Xenium	2	541
Mouse Bone Data with Custom Add-on Panel	Bone	Xenium	3	541
Mouse Brain Coronal Section 1 (FFPE)	Brain	Visium	1	19,465
Mouse Brain Coronal Section 2 (FFPE)	Brain	Visium	1	19,465
Mouse Brain Section (Coronal)	Brain	Visium	1	31,053
Mouse Brain Serial Section 1 (Sagittal-Anterior)	Brain	Visium	1	31,053
Mouse Brain Serial Section 1 (Sagittal-Posterior)	Brain	Visium	1	31,053
Mouse Brain Serial Section 2 (Sagittal-Anterior)	Brain	Visium	1	31,053
Mouse Brain Serial Section 2 (Sagittal-Posterior)	Brain	Visium	1	31,053
Mouse Embryo, 11 mm Capture Area (FFPE)	Embryo	Visium	1	19,465
Mouse Kidney Section (Coronal)	Kidney	Visium	1	31,053
Mouse Tissue Microarray in 3x3 Layout with 1 mm Edge to Edge Spacing (FFPE)	Lung/Brain	Visium	1	19,465
Mouse Tissue Microarray in 3x3 Layout with 2 mm Edge to Edge Spacing (FFPE)	Lung/Brain	Visium	1	19,465
Mouse Tissue Microarray in 5x5 Layout with 1 mm Edge to Edge Spacing (FFPE)	Kidney/Brain	Visium	1	19,465
Normal Human Prostate (FFPE)	Prostate	Visium	1	17,943
Pancreatic Cancer with Xenium Human Multi-Tissue and Cancer Panel	Pancreas	Xenium	1	538
Preservation Method Comparison on CytAssist: FFPE Mouse Brain (Sagittal), 11 mm Capture Area	Brain	Visium	1	19,465
Preservation Method Comparison on CytAssist: Fixed Frozen Mouse Brain (Sagittal), 11 mm Capture Area	Brain	Visium	1	19,465
Preservation Method Comparison on CytAssist: Fresh Frozen Mouse Brain (Sagittal), 11 mm Capture Area	Brain	Visium	1	19,465
Preservation Method Comparison on Visium CytAssist: FFPE Mouse Brain (Sagittal), 11 mm Capture Area	Brain	Visium	1	19,465
Preservation Method Comparison on Visium CytAssist: Fixed Frozen Mouse Brain (Sagittal), 11 mm Capture Area	Brain	Visium	1	19,465
Preservation Method Comparison on Visium CytAssist: Fresh Frozen Mouse Brain (Sagittal), 11 mm Capture Area	Brain	Visium	1	19,465
Preview Data: FFPE Human Lung Cancer with Xenium Multimodal Cell Segmentation	Lung	Xenium	1	541
Preview Data: FFPE Human Lymph Node with 5K Pan Tissue and Pathways Panel	Lymphoid	Xenium	1	11,094
Visium CytAssist Gene Expression Libraries of Post-Xenium Human Colon Cancer (FFPE)	Bowel	Visium	4	18,085
Visium CytAssist Gene Expression Libraries of Post-Xenium Mouse Brain (FF)	Brain	Visium	4	19,465
Visium CytAssist, Mouse Embryo, 11 mm Capture Area (FFPE)	Embryo	Visium	1	19,465
Visium HD Spatial Gene Expression Library, Mouse Kidney (FFPE)	Kidney	Visium HD	1	19,059
Visium HD Spatial Gene Expression Library, Mouse Embryo (FFPE)	Embryo	Visium HD	1	19,059
Visium HD Spatial Gene Expression Library, Human Pancreas (FFPE)	Pancreas	Visium HD	1	18,085
Whole Mouse Pup Preview Data (Xenium Mouse Tissue Atlassing Panel)	Whole organism	Xenium	1	541

22

Table A4: **Datasets gathered from NCBI.**

Publication	Organ	Technology	n	Num. genes
10X Visium Spatial transcriptomics of murine colon at d14 (mucosa healing) in B cell sufficient-deficient mice [117]	Bowel	Visium	2	31,053
10X Visium Spatial transcriptomics of murine colon in steady state and during recovery after DSS colitis [117]	Bowel	Visium	2	31,053
A Spatial Transcriptomic atlas of the human kidney papilla identifies significant immune injury in patients with stone disease [25]	Kidney	Visium	7	36,601
A cellular hierarchy in melanoma uncouples growth and metastasis [69]	Skin	Visium	3	31,053
A new epithelial cell subpopulation predicts response to surgery, chemotherapy, and immunotherapy in bladder cancer [2, 52]	Bladder	Visium	4	33,538
A novel model of binge ethanol exposure reveals enhanced neurodegeneration with advanced age [7]	Brain	Visium	4	32,285
A single-cell transcriptomic analysis of endometriosis [43]	Uterus	Visium	2	36,601
Distinct mesenchymal cell states mediate prostate cancer progression [114]	Prostate	Visium	2	32,589
Epithelial Plasticity and Innate Immune Activation Promote Lung Tissue Remodeling following Respiratory Viral Infection [19]	Lung	Visium	1	32,285
Image-based spatial transcriptomics identifies molecular niche dysregulation associated with distal lung remodeling in pulmonary fibrosis [136]	Lung	Xenium	20	17,145
Gene expression within a human choroidal neovascular membrane using spatial transcriptomics [141]	Eye	Visium	5	36,601
Genome-wide Spatial Expression Profiling in Formalin-fixed Tissues [53]	Kidney	Visium	14	33,538
High-resolution mapping of the tumor microenvironment using integrated single-cell, spatial, and in situ analysis [63]	Breast	Xenium	4	541
Identification of TREM1+CD163+ myeloid cells as a deleterious immune subset in HCC [51]	Liver	Visium	2	36,601
Integration of spatial and single cell transcriptomics localizes epithelial-immune cross-talk in kidney injury [41]	Kidney	Visium	4	33,538
Molecular Atlas of the Adult Mouse Brain [113]	Brain	Spatial Transcriptomics	75	23,371
Regional differential gene expression analyses of brains from four 24w-old Nf1+- mice	Brain	Visium	4	32,285
SARS-CoV-2 Niches in Human Placenta Revealed by Spatial Transcriptomics [16]	Uterus	Visium	16	36,612
Schwann Cells Shape Tumor Cells and Cancer-Associated Fibroblasts in the Pancreatic Ductal Adenocarcinoma Microenvironment [155]	Pancreas	Visium	4	20,615
Single Cell and Spatial Analysis of Human Squamous Cell Carcinoma [66]	Skin	Spatial Transcriptomics	12	17,138
Single-cell profiling of primary and paired metastatic lymph node tumors in breast cancer patients [84]	Lymph node	Visium	4	33,931

23

Table A5: **Datasets gathered from NCBI. Continuation.**

Publication	Organ	Technology	n	Num. genes
Single-cell and spatial transcriptomics characterization of the immunological landscape in the healthy and PSC human liver [5]	Liver	Visium	4	36,601
Single-nucleus Ribonucleic Acid-sequencing and Spatial Transcriptomics Reveal the Cardioprotection of Shexiang Baoxin Pill (MUSKARDIA) in Mice with Myocardial Ischemia-Reperfusion Injury [82]	Heart	Visium	2	32,285
Spatial Multimodal Analysis: MALDI-MSI and Spatial Transcriptomics within the same tissue section [138]	Brain	Visium	19	32,285
Spatial RNA sequencing of regenerating mouse hindlimb muscle [100]	Muscle	Visium	3	33,217
Spatial Total RNA-Sequencing of regenerating mouse hindlimb muscle and Type 1-Lang reovirus-infected mouse heart [101]	Muscle	Visium	7	55,414
Spatial localization with Spatial Transcriptomics for an atlas of healthy and injured cell states and niches in the human kidney [79]	Kidney	Visium	23	33,538
Spatial sequencing of Foreign body granuloma [78]	None	Visium	1	15,524
Spatial transcriptomics landscape of non-communicable inflammatory skin diseases [126]	Skin	Visium	59	20,613
Spatial transcriptomics of adenoid cystic carcinoma of the lacrimal gland [106]	Eye	Visium	1	17,943
Spatial transcriptomics of the mouse brain across three age groups	Brain	Visium	6	32,285
Spatial transcriptomics reveal unnresolved wound repair as potential driver of PFA Ependymoma progression [44]	Brain	Visium	14	36,601
Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis [94]	Spinal cord	Spatial Transcriptomics	302	12,572
The neurons that restore walking after paralysis [70]	Spinal cord	Visium	16	22,127
Visium spatial transcriptomics analysis of lacrimal gland during chronic inflammation progression [98]	Eye	Visium	4	32,285
YAP Drives Assembly of a Spatially Colocalized Cellular Triad Required for Heart Renewal [80]	Heart	Visium	2	32,285
Zika virus co-opts miRNA networks to persist in placental microenvironments detected by spatial transcriptomics [17]	Placenta	Visium	8	32,298
Mouse model Heptablastoma spatial transcriptomics [120]	Liver	Visium	10	31,053

Table A6: **Datasets gathered on Spatial-Research.**

Publication	Organ	Technology	n	Num. genes
A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart [10]	Heart	Spatial Transcriptomics	19	39,739
Integrating spatial gene expression and breast tumour morphology via deep learning [55]	Breast	Spatial Transcriptomics	68	16,744
Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions [4]	Breast	Spatial Transcriptomics	36	15,045
Visualization and analysis of gene expression in tissue sections by spatial transcriptomics [134]	Brain	Spatial Transcriptomics	16	16,573

24

Table A7: **Datasets gathered on Mendeley.**

Publication	Organ	Technology	n	Num. genes
Ex-ST [6]	Brain	Visium	5	31,053
Genome-wide spatial expression profiling in formalin-fixed tissues [139]	Brain	Visium	15	31,053
Human ileum, Visium [104]	Bowel	Visium	4	33,538
Human squamous cell carcinoma [1]	Skin	Visium	4	33,538
Prostate needle biopsies pre- and post-ADT: Count matrices, histological-, and Androgen receptor immunohistochemistry images [96]	Prostate	Spatial Transcriptomics	24	26,437
Spatially resolved clonal copy number alterations in benign and malignant tissue [40]	Prostate	Visium	23	33,538
Spatially resolved transcriptomic profiling of degraded [105]	Bowel	Visium	35	17,943
spatialRNAseq heart raw suppdata	Heart	Visium	4	54,848
spatialRNAseq ileum raw suppdata	Bowel	Visium	4	54,848

Table A8: **Datasets gathered on Github and the Human Cell Atlas data explorer.**

Publication	Organ	Technology	n	Num. genes
A spatially resolved atlas of the human lung characterizes a gland-associated immune niche[93]	Lung	Visium	20	17,922
Molecular cartography uncovers evolutionary and microenvironmental dynamics in sporadic colorectal tumors[57]	Colon	Visium	41	19,366
Spatially resolved multiomics of human cardiac niches[67]	Heart	Visium	41	33,538
Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex[99]	Brain	Visium	12	33,538

Table A9: **Internal datasets.**

Publication	Organ	Technology	n	Num. genes
Prostate ST Internal	Prostate	Visium	4	17,943
Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer [103]	Lymph node	Visium	24	17,943

Table A10: **Datasets gathered on Zenodo.**

Publication	Organ	Technology	n	Num. genes
Charting the Heterogeneity of Colorectal Cancer Consensus Molecular Subtypes using Spatial Transcriptomics: datasets [135]	Bowel	Visium	14	36,601
Demo 10x Visium dataset for STQ [34]	Skin	Visium	1	68,886
Nextflow Pipeline for Visium and H&E Data from Patient-Derived Xenograft Samples [36]	Skin	Visium	4	68886
Spotiphy: generative modeling in single-cell spatial whole transcriptomics [132]	Brain	Visium	2	32,285

25

Table A11: **Overview of the HEST-Benchmark.** Each task involves predicting the expression levels of the 50 most variable genes from 112×112 μm H&E-stained image patches centered on each spatial transcriptomics spot. The tasks are formulated as multivariate regression problems. The Oncotree code describes the cancer type diagnosed in samples, e.g., PAAD denotes pancreatic adenocarcinoma. Additional information is provided in the Appendix.

Task ID	Oncotree	Organ	Number of Patients	Number of Samples	Technology
Task 1 [63]	IDC	Breast	4	4	Xenium
Task 2	PRAD	Prostate	2	23	Visium
Task 3	PAAD	Pancreas	3	3	Xenium
Task 4	SKCM	Skin	2	2	Xenium
Task 5 [135]	COAD	Colon	2	4	Xenium
Task 6 [135]	READ	Rectum	2	4	Visium
Task 7 [103]	ccRCC	Kidney	24	24	Visium
Task 8	LUAD	Lung	2	2	Xenium
Task 9 [84]	IDC	Axillary lymph nodes	4	4	Visium

Table A12: **State-of-the-art foundation models for histology evaluated on HEST-Benchmark.** B: Base, L: Large, H: Huge, G: Giant. *: number of patches during pretraining, **: number of patches during fine-tuning.

Name	Number of slides	Number of patches	Magnification	Model	Training recipe	Number of parameters
ResNet50 (IN) [91]	N/A	1.2M	N/A	ResNet-50	Supervised	23M
CTransPath [147]	32k	17M	10×	Swin-T	MoCov3	28M
Remedis [11]	10k	10M	20×	ResNet-152	iBOT	232M
Phikon [42]	6k	43.3M	20×	ViT-B	iBOT	86M
UNI [29]	100k	100M	20×	ViT-L	DINOv2	307M
CONCH [88]	N/A	16M* + 1.17M**	Many	ViT-B	CoCa	86M
GigaPath [154]	171k	1.3B	20×	ViT-g	DINOv2	1.13B
Virchow [142]	1.5M	2B	20×	ViT-H	DINOv2	632M
Virchow 2 [162]	3.1M	1.9B	Many	ViT-H	DINOv2	632M
H-Optimus-0	500K	273M	20x	ViT-g	DINOv2	1.13B
UNIv1.5	350K	432M	20x	ViT-g	DINOv2	1.13B

Table A13: **HEST-Benchmark evaluated using Ridge regression**. Model performance measured with Pearson correlation. Best is **bold**, second best is <u>underlined</u>.

	IDC	PRAD	PAAD	SKCM	COAD	READ	ccRCC	LUAD	LYMPH IDC	Average
REMEDIS	0.4936	0.2632	0.2881	0.4117	0.151	0.0776	0.2201	0.3114	0.1694	0.2651
	± 0.0725	± 0.0821	± 0.0544	± 0.0384	± 0.0147	± 0.0684	± 0.0418	± 0.0432	± 0.0365	
GigaPath	0.532	0.3035	0.3172	0.2231	0.163	0.1236	0.2172	0.3144	0.1925	0.2652
	± 0.0812	± 0.0279	± 0.0165	± 0.0071	± 0.041	± 0.0479	± 0.0871	± 0.0304		
UNIv1.5	0.5657	<u>0.3065</u>	0.3004	0.258	0.1982	0.1077	0.2023	0.3084	0.1998	0.2719
	± 0.0866	± 0.02	± 0.0199	± 0.0514	± 0.0262	± 0.0222	± 0.0595	± 0.0747	± 0.0129	
H-Optimus-0	**0.5789**	0.2561	0.3367	0.2778	0.1605	0.1228	0.2342	0.3143	0.1976	0.2754
	± 0.0899	± 0.0003	± 0.0428	± 0.0048	± 0.0522	± 0.0309	± 0.0373	± 0.083	± 0.0253	
Virchow2	<u>0.5666</u>	0.2972	0.2718	0.303	0.1814	0.1208	0.2257	0.3017	0.2172	0.2762
	± 0.0848	± 0.037	± 0.0387	± 0.0184	± 0.0326	± 0.0526	± 0.0433	± 0.1199	± 0.019	
Virchow	0.5583	0.2744	0.3361	0.3389	0.1825	0.0955	0.2375	0.2897	0.2081	0.2801
	± 0.0876	± 0.0019	± 0.037	± 0.0063	± 0.0369	± 0.0527	± 0.0371	± 0.0785	± 0.0234	
ResNet50	0.4453	0.2753	0.3432	<u>0.413</u>	<u>0.2009</u>	0.0669	0.2103	0.4001	0.203	0.2842
	± 0.0377	± 0.0622	± 0.0654	± 0.0814	± 0.061	± 0.0646	± 0.0548	± 0.0637	± 0.0536	
CTransPath	0.4996	0.2895	<u>0.3826</u>	0.4038	0.1751	0.0909	0.2139	0.4026	0.2089	0.2963
	± 0.0594	± 0.0724	± 0.066	± 0.065	± 0.0423	± 0.0808	± 0.0438	± 0.07	± 0.0367	
Phikon	0.5259	0.2493	0.3594	0.3684	0.1697	0.1136	**0.253**	<u>0.4224</u>	0.2151	0.2974
	± 0.0791	± 0.1264	± 0.0707	± 0.1061	± 0.0562	± 0.0749	± 0.0483	± 0.0579	± 0.0416	
UNI	0.563	0.257	0.3768	0.3433	0.1839	<u>0.1239</u>	<u>0.2395</u>	0.3714	<u>0.2236</u>	<u>0.2981</u>
	± 0.0771	± 0.0819	± 0.0555	± 0.0556	± 0.0509	± 0.0434	± 0.0557	± 0.1098	± 0.0289	
CONCH	0.528	**0.3604**	**0.4224**	**0.5079**	**0.2467**	**0.1443**	0.2356	**0.4957**	**0.2462**	**0.3541**
	± 0.0794	± 0.0135	± 0.0773	± 0.0281	± 0.0045	± 0.0455	± 0.0387	± 0.0203	± 0.0349	

26

Table A14: **HEST-Benchmark evaluated using XGBoost regression.** Model performance measured with Pearson correlation. Best is **bold**, second best is <u>underlined</u>.

	IDC	PRAD	PAAD	SKCM	COAD	READ	ccRCC	LUAD	LYMPH IDC	Average
ResNet50 (IN)	0.4646	0.3433	0.4017	0.4707	0.2892	0.0586	0.181	0.4967	0.2284	0.326
	± 0.0353	± 0.0168	± 0.0648	± 0.0834	± 0.0115	± 0.069	± 0.0502	± 0.01	± 0.0511	
CTransPath	0.4738	0.3514	0.4257	0.5304	0.2921	0.0996	0.2026	0.5225	0.234	0.348
	± 0.0394	± 0.0032	± 0.0701	± 0.073	± 0.0018	± 0.0766	± 0.0387	± 0.0063	± 0.0613	
Phikon	0.4704	**0.3943**	0.3988	0.5323	0.277	0.1451	0.213	0.542	0.2443	0.3575
	± 0.0672	± 0.0123	± 0.0598	± 0.0607	± 0.0098	± 0.0851	± 0.0362	± 0.017	± 0.0632	
GigaPath	0.5222	0.3749	0.4415	0.5297	0.2876	0.1609	0.2207	0.5506	0.2464	0.3705
	± 0.0641	± 0.0103	± 0.058	± 0.0376	± 0.0039	± 0.0777	± 0.0402	± 0.0108	± 0.0526	
CONCH	0.5175	0.3784	0.4428	0.5766	0.3215	0.1431	0.1738	0.5581	0.2554	0.3742
	± 0.0602	± 0.0124	± 0.0665	± 0.0519	± 0.0062	± 0.0665	± 0.0544	± 0.0081	± 0.0605	
REMEDIS	0.5116	0.3526	**0.4621**	0.5885	0.319	0.1129	0.2303	0.562	0.2521	0.3768
	± 0.0594	± 0.0073	± 0.0555	± 0.0253	± 0.0101	± 0.0846	± 0.0393	± 0.0057	± 0.0601	
Virchow2	0.5378	0.3772	0.4237	0.5565	0.281	**0.1779**	**0.2428**	0.5641	<u>0.2582</u>	0.3799
	± 0.0685	± 0.007	± 0.0525	± 0.0152	± 0.0162	± 0.077	± 0.0361	± 0.0069	± 0.0504	
UNI	0.538	0.3513	<u>0.451</u>	<u>0.6089</u>	0.2921	0.1679	0.235	0.5357	0.2456	0.3806
	± 0.0603	± 0.0162	± 0.0587	± 0.0294	± 0.0191	± 0.0641	± 0.0381	± 0.0057	± 0.05	
Virchow	0.5309	0.3447	0.4448	0.6089	0.3275	0.1419	0.2307	<u>0.5643</u>	**0.2617**	0.3839
	± 0.0764	± 0.0117	± 0.0501	± 0.0165	± 0.0254	± 0.0669	± 0.0336	± 0.0091	± 0.0537	
UNIv1.5	<u>0.555</u>	0.3654	0.434	0.6025	**0.336**	<u>0.1742</u>	0.2166	0.5634	0.2515	<u>0.3887</u>
	± 0.0763	± 0.0098	± 0.0568	± 0.0385	± 0.0179	± 0.0568	± 0.0337	± 0.0054	± 0.0434	
H-Optimus-0	**0.5564**	<u>0.3829</u>	0.4445	**0.6502**	0.2922	0.1731	<u>0.2402</u>	**0.5654**	0.2555	**0.3956**
	± 0.0777	± 0.0049	± 0.0563	± 0.0326	± 0.0063	± 0.0777	± 0.0348	± 0.0084	± 0.0522	

27

References

[1] Xesús Abalo et al. *Human squamous cell carcinoma, Visium*. Version V1. 2021. DOI: 10. 17632/2bh5fchcv6.1. URL: https://data.mendeley.com/datasets/2bh5fchcv6/1.

[2] H. A. Abdel-Hafiz et al. "Y chromosome loss in cancer drives growth by evasion of adaptive immunity". In: *Nature* 619.7970 (2023), pp. 624–631. DOI: 10.1038/s41586-023-06083-w.

[3] Areej Alsaafin et al. "Learning to predict RNA sequence expressions from whole slide images with applications for search and classification". In: *Communications Biology* 6.1 (2023), p. 304.

[4] Alma Andersson et al. "Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions". In: *Nat. Commun.* 12.6012 (Oct. 2021), pp. 1–14. ISSN: 2041-1723. DOI: 10.1038/s41467-021-26271-2.

[5] Tallulah S. Andrews et al. "Single-cell, single-nucleus, and spatial transcriptomics characterization of the immunological landscape in the healthy and PSC human liver". In: *J. Hepatol.* 80.5 (May 2024), pp. 730–743. ISSN: 1600-0641. DOI: 10.1016/j.jhep.2023.12.023. eprint: 38199298.

[6] Zaneta Andrusivova and Yuhang Fan. *Ex-ST*. Version V1. 2023. DOI: 10.17632/nrbsxrk9mp.1. URL: https://data.mendeley.com/datasets/nrbsxrk9mp/1.

[7] Paige E. Anton et al. "Binge ethanol exposure in advanced age elevates neuroinflammation and early indicators of neurodegeneration and cognitive impairment in female mice". In: *Brain Behav. Immun.* 116 (Feb. 2024), pp. 303–316. ISSN: 0889-1591. DOI: 10.1016/j.bbi.2023.12.034.

[8] Guilherme Aresta et al. "Bach: Grand challenge on breast cancer histology images". In: *Medical image analysis* 56 (2019), pp. 122–139.

[9] Michaela Asp, Joseph Bergenstråhle, and Joakim Lundeberg. "Spatially Resolved Transcriptomes—Next Generation Tools for Tissue Exploration". In: *BioEssays* 42.10 (Oct. 2020), p. 1900221. ISSN: 0265-9247. DOI: 10.1002/bies.201900221.

[10] Michaela Asp et al. "A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart". In: *Cell* 179.7 (Dec. 2019), pp. 1647–166019. ISSN: 1097-4172. DOI: 10.1016/j.cell.2019.11.025. eprint: 31835037.

[11] Shekoofeh Azizi et al. "Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging". In: *Nature Biomedical Engineering* (2023), pp. 1–24.

[12] Sunil Badve et al. "FOXA1 expression in breast cancer—correlation with luminal subtype A and survival". In: *Clinical cancer research* 13.15 (2007), pp. 4415–4421.

[13] S Bandaru et al. "Targeting filamin B induces tumor growth and metastasis via enhanced activity of matrix metalloproteinase-9 and secretion of VEGF-A". In: *Oncogenesis* 3.9 (2014), e119–e119.

[14] Peter Bankhead et al. "QuPath: Open source software for digital pathology image analysis". In: *Scientific Reports* 7 (Dec. 2017). DOI: 10.1038/s41598-017-17204-5.

[15] Carlo Alberto Barbano et al. "UniToPatho, a labeled histopathological dataset for colorectal polyps classification and adenoma dysplasia grading". In: *arXiv* (Jan. 2021). DOI: 10.1109/ICIP42928.2021.9506198. eprint: 2101.09991.

[16] Enrico R. Barrozo et al. "SARS-CoV-2 niches in human placenta revealed by spatial transcriptomics". In: *Med* 4.9 (Sept. 2023), 612–634.e4. ISSN: 2666-6340. DOI: 10.1016/j.medj.2023.06.003.

[17] Enrico R. Barrozo et al. "Zika virus co-opts microRNA networks to persist in placental niches detected by spatial transcriptomics". In: *Am. J. Obstet. Gynecol.* 230.2 (Feb. 2024), pp. 2511–25117. ISSN: 1097-6868. DOI: 10.1016/j.ajog.2023.08.012. eprint: 37598997.

[18] Babak Ehteshami Bejnordi et al. "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer". In: *JAMA* 318.22 (2017), pp. 2199–2210.

[19] Andrew K. Beppu et al. "Epithelial plasticity and innate immune activation promote lung tissue remodeling following respiratory viral infection". In: *Nat. Commun.* 14.5814 (Sept. 2023), pp. 1–16. ISSN: 2041-1723. DOI: 10.1038/s41467-023-41387-3.

28

[20] Ludvig Bergenstråhle et al. "Super-resolved spatial transcriptomics by deep data fusion". In: *Nat. Biotechnol.* 40 (Apr. 2022), pp. 476–479. ISSN: 1546-1696. DOI: 10.1038/s41587-021-01075-3.

[21] Quentin Blampey et al. "Sopa: a technology-invariant pipeline for analyses of image-based spatial omics". In: *Nature Communications* 15 (2024). Publisher: Nature Publishing Group, p. 4981. DOI: 10.1038/s41467-024-48981-z. URL: https://www.nature.com/articles/s41467-024-48981-z.

[22] Nadia Brancati et al. "BRACS: A Dataset for BReAst Carcinoma Subtyping in H&E Histology Images". In: *arXiv preprint arXiv:2111.04740* (2021).

[23] Wouter Bulten et al. "Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge". In: *Nature Medicine* 28.1 (2022), pp. 154–163.

[24] Andrew Butler et al. "Integrating single-cell transcriptomic data across different conditions, technologies, and species". In: *Nat. Biotechnol.* 36 (May 2018), pp. 411–420. ISSN: 1546-1696. DOI: 10.1038/nbt.4096.

[25] V. H. Canela et al. "A spatially anchored transcriptomic atlas of the human kidney papilla identifies significant immune injury in patients with stone disease". In: *Nature Communications* 14.1 (2023), p. 4140. DOI: 10.1038/s41467-023-41340-8.

[26] Jiaji George Chen et al. "Giotto Suite: a multi-scale and technology-agnostic spatial multi-omics analysis ecosystem". In: *bioRxiv* (Nov. 2023), p. 2023.11.26.568752. eprint: 2023.11.26.568752. URL: https://doi.org/10.1101/2023.11.26.568752.

[27] Liang-Chieh Chen et al. "Rethinking Atrous Convolution for Semantic Image Segmentation". In: *arXiv* (June 2017). DOI: 10.48550/arXiv.1706.05587. eprint: 1706.05587.

[28] Richard J Chen et al. "Pan-cancer integrative histology-genomic analysis via multimodal deep learning". In: *Cancer Cell* 40.8 (2022), pp. 865–878.

[29] Richard J. Chen et al. "Towards a General-Purpose Foundation Model for Computational Pathology". In: *Nature Medicine* (2024).

[30] Ting Chen et al. "A simple framework for contrastive learning of visual representations". In: *International conference on machine learning*. PMLR. 2020, pp. 1597–1607.

[31] Xinlei Chen, Saining Xie, and Kaiming He. "An Empirical Study of Training Self-Supervised Vision Transformers". In: *arXiv* (Apr. 2021). DOI: 10.48550/arXiv.2104.02057. eprint: 2104.02057.

[32] Youngmin Chung et al. "Accurate Spatial Gene Expression Prediction by integrating Multi-resolution features". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, pp. 11591–11600.

[33] Ozan Ciga, Tony Xu, and Anne Louise Martel. "Self supervised contrastive learning for digital histopathology". In: *Machine Learning with Applications* 7 (2022).

[34] *Demo 10x Visium dataset for STQ*. [Online; accessed 16. May 2024]. Feb. 2024. DOI: 10.5281/zenodo.10654467.

[35] Jia Deng et al. "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.

[36] Sergii Domanskyi et al. "Nextflow Pipeline for Visium and H&E Data from Patient-Derived Xenograft Samples". In: *bioRxiv* (2023). DOI: 10.1101/2023.07.27.550727. eprint: https://www.biorxiv.org/content/early/2023/07/30/2023.07.27.550727.full.pdf. URL: https://www.biorxiv.org/content/early/2023/07/30/2023.07.27.550727.

[37] Alexey Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *International Conference on Learning Representations*. 2021. URL: https://openreview.net/forum?id=YicbFdNTTy.

[38] Amelie Echle et al. "Deep learning in cancer pathology: a new generation of clinical biomarkers". In: *British journal of cancer* 124.4 (2021), pp. 686–696.

[39] Omar SM El Nahhas et al. "Regression-based Deep-Learning predicts molecular biomarkers from pathology slides". In: *Nature communications* 15.1 (2024), p. 1253.

[40] Andrew Erickson et al. "Spatially resolved clonal copy number alterations in benign and malignant tissue". In: *Nature* 608 (Aug. 2022), pp. 360–367. ISSN: 1476-4687. DOI: 10.1038/s41586-022-05023-2.

29

[41] Ricardo Melo Ferreira et al. "Integration of spatial and single-cell transcriptomics localizes epithelial cell–immune cross-talk in kidney injury". In: *JCI Insight* 6.12 (June 2021). ISSN: 0021-9738. DOI: 10.1172/jci.insight.147703.

[42] Alexandre Filiot et al. "Scaling Self-Supervised Learning for Histopathology with Masked Image Modeling". In: *medRxiv* (July 2023). DOI: 10.1101/2023.07.21.23292757.

[43] Marcos A. S. Fonseca et al. "Single-cell transcriptomic analysis of endometriosis". In: *Nat. Genet.* 55 (Feb. 2023), pp. 255–267. ISSN: 1546-1718. DOI: 10.1038/s41588-022-01254-1.

[44] R. Fu et al. "Spatial transcriptomic analysis delineates epithelial and mesenchymal subpopulations and transition stages in childhood ependymoma". In: *Neuro-Oncology* 25.4 (2023), pp. 786–798. DOI: 10.1093/neuonc/noad070.

[45] Yu Fu et al. "Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis". In: *Nature Cancer* 1.8 (2020), pp. 800–810.

[46] Jevgenij Gamper and Nasir Rajpoot. "Multiple Instance Captioning: Learning Representations from Histopathology Textbooks and Articles". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

[47] Jevgenij Gamper et al. "PanNuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification". In: *European Congress on Digital Pathology*. Springer. 2019, pp. 11–19.

[48] Jevgenij Gamper et al. "PanNuke Dataset Extension, Insights and Baselines". In: *arXiv preprint arXiv:2003.10778* (2020).

[49] Quentin Garrido et al. "Rankme: Assessing the downstream performance of pretrained self-supervised representations by their rank". In: *International Conference on Machine Learning*. PMLR. 2023, pp. 10929–10974.

[50] Chandler D. Gatenbee et al. "Virtual alignment of pathology image series for multi-gigapixel whole slide images". In: *Nat. Commun.* 14.4502 (July 2023), pp. 1–14. ISSN: 2041-1723. DOI: 10.1038/s41467-023-40218-9.

[51] Julie Giraud et al. "TREM1+CD163+ regulatory myeloid cells expand in steatohepatitis-HCC and associate with poor prognosis and therapeutic resistance to immune checkpoint blockade". In: *bioRxiv* (Nov. 2022), p. 2022.11.09.515839. eprint: 2022.11.09.515839. URL: https://doi.org/10.1101/2022.11.09.515839.

[52] K. H. 3rd Gouin et al. "An N-Cadherin 2 expressing epithelial cell subpopulation predicts response to surgery, chemotherapy and immunotherapy in bladder cancer". In: *Nature Communications* 12.1 (2021), p. 4906. DOI: 10.1038/s41467-021-25205-2.

[53] Eva Gracia Villacampa et al. "Genome-wide spatial expression profiling in formalin-fixed tissues". In: *Cell Genomics* 1.3 (Dec. 2021), p. 100065. ISSN: 2666-979X. DOI: 10.1016/j.xgen.2021.100065.

[54] Laleh Haghverdi et al. "Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors". In: *Nat. Biotechnol.* 36 (May 2018), pp. 421–427. ISSN: 1546-1696. DOI: 10.1038/nbt.4091.

[55] Bryan He et al. "Integrating spatial gene expression and breast tumour morphology via deep learning". In: *Nature biomedical engineering* 4.8 (2020), pp. 827–834.

[56] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[57] Cody N. Heiser et al. "Molecular cartography uncovers evolutionary and microenvironmental dynamics in sporadic colorectal tumors". In: *Cell* 186.25 (Dec. 2023). Publisher: Elsevier, 5620–5637.e16. ISSN: 0092-8674. DOI: 10.1016/j.cell.2023.11.006. URL: https://doi.org/10.1016/j.cell.2023.11.006 (visited on 10/25/2024).

[58] Jian Hu et al. "SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network". In: *Nature methods* 18.11 (2021), pp. 1342–1351.

[59] Zhi Huang et al. "A visual–language foundation model for pathology image analysis using medical Twitter". In: *Nature Medicine* 29 (Aug. 2023), pp. 1–10. DOI: 10.1038/s41591-023-02504-3.

30

[60] Xinmi Huo et al. *Comprehensive AI Model Development for Gleason Grading: From Scanning, Cloud-Based Annotation to Pathologist-AI Interaction.* [Online; accessed 7. May 2024]. July 2022. DOI: 10.2139/ssrn.4172090.

[61] Fabian Hörst et al. "CellViT: Vision Transformers for precise cell segmentation and classification". In: *Medical Image Analysis* 94 (2024), p. 103143. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2024.103143. URL: https://www.sciencedirect.com/science/article/pii/S1361841524000689.

[62] Maximilian Ilse, Jakub Tomczak, and Max Welling. "Attention-based Deep Multiple Instance Learning". In: *Proceedings of the 35th International Conference on Machine Learning.* 2018, pp. 2132–2141.

[63] Amanda Janesick et al. "High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis". In: *Nat. Commun.* 14.8353 (Dec. 2023), pp. 1–15. ISSN: 2041-1723. DOI: 10.1038/s41467-023-43458-x.

[64] Guillaume Jaume et al. "Multistain Pretraining for Slide Representation Learning in Pathology". In: *Proceedings of the European Conference on Computer Vision (ECCV).* Springer. 2024.

[65] Guillaume Jaume et al. "Transcriptomics-guided Slide Representation Learning in Computational Pathology". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 2024.

[66] Andrew L. Ji et al. "Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma". In: *Cell* 182.2 (July 2020), pp. 497–51422. ISSN: 1097-4172. DOI: 10.1016/j.cell.2020.05.039. eprint: 32579974.

[67] Kazumasa Kanemaru et al. "Spatially resolved multiomics of human cardiac niches". In: *Nature* 619.7971 (July 2023), pp. 801–810. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06311-1. URL: https://doi.org/10.1038/s41586-023-06311-1.

[68] Mingu Kang et al. "Benchmarking Self-Supervised Learning on Diverse Pathology Datasets". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2023, pp. 3344–3354.

[69] Panagiotis Karras et al. "A cellular hierarchy in melanoma uncouples growth and metastasis". In: *Nature* 610 (Oct. 2022), pp. 190–198. ISSN: 1476-4687. DOI: 10.1038/s41586-022-05242-7.

[70] Claudia Kathe et al. "The neurons that restore walking after paralysis". In: *Nature* 611 (Nov. 2022), pp. 540–547. ISSN: 1476-4687. DOI: 10.1038/s41586-022-05385-7.

[71] Jakob Nikolas Kather et al. "Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer". In: *Nature medicine* 25.7 (2019), pp. 1054–1056.

[72] Jakob Nikolas Kather et al. "Pan-cancer image-based detection of clinically actionable genetic alterations". In: *Nature Cancer* 1.8 (2020), pp. 789–799.

[73] Jakob Nikolas Kather et al. "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study". In: *PLoS Med.* 16.1 (Jan. 2019). DOI: 10.1371/journal.pmed.1002730.

[74] Alexander Kolesnikov et al. "Big Transfer (BiT): General Visual Representation Learning". In: *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V.* Heidelberg, Germany: Springer-Verlag, 2020, pp. 491–507. ISBN: 978-3-030-58557-0. DOI: 10.1007/978-3-030-58558-7_29.

[75] Navid Alemi Koohbanani et al. "Self-Path: Self-supervision for Classification of Pathology Images with Limited Annotations". In: *IEEE Transactions on Medical Imaging* (2021).

[76] Ilya Korsunsky et al. "Fast, sensitive and accurate integration of single-cell data with Harmony". In: *Nat. Methods* 16 (Dec. 2019), pp. 1289–1296. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0619-0.

[77] Michał Koziarski et al. "DiagSet: a dataset for prostate cancer histopathological image classification". In: *Sci. Rep.* 14.6780 (Mar. 2024), pp. 1–14. ISSN: 2045-2322. DOI: 10.1038/s41598-024-52183-4.

[78] Thomas Krausgruber et al. "Single-cell and spatial transcriptomics reveal aberrant lymphoid developmental programs driving granuloma formation". In: *Immunity* 56.2 (Feb. 2023), 289–306.e7. ISSN: 1074-7613. DOI: 10.1016/j.immuni.2023.01.014.

31

[79] Blue B. Lake et al. "An atlas of healthy and injured cell states and niches in the human kidney". In: *Nature* 619 (July 2023), pp. 585–594. ISSN: 1476-4687. DOI: 10.1038/s41586-023-05769-3.

[80] Rich Gang Li et al. "YAP induces a neonatal-like pro-renewal niche in the adult heart". In: *Nature cardiovascular research* 3.3 (Mar. 2024), p. 283. DOI: 10.1038/s44161-024-00428-w.

[81] Xinmin Li and Cun-Yu Wang. "From bulk, single-cell to spatial RNA sequencing". In: *Int. J. Oral Sci.* 13.36 (Nov. 2021), pp. 1–6. ISSN: 2049-3169. DOI: 10.1038/s41368-021-00146-0.

[82] Wenyong Lin et al. "Single-nucleus ribonucleic acid-sequencing and spatial transcriptomics reveal the cardioprotection of Shexiang Baoxin Pill (SBP) in mice with myocardial ischemia-reperfusion injury". In: *Front. Pharmacol.* 14 (May 2023), p. 1173649. ISSN: 1663-9812. DOI: 10.3389/fphar.2023.1173649.

[83] Jana Lipkova et al. "Deep learning-enabled assessment of cardiac allograft rejection from endomyocardial biopsies". In: *Nature medicine* 28.3 (2022), pp. 575–582.

[84] Tong Liu et al. "Single cell profiling of primary and paired metastatic lymph node tumors in breast cancer patients". In: *Nat. Commun.* 13.6823 (Nov. 2022), pp. 1–17. ISSN: 2041-1723. DOI: 10.1038/s41467-022-34581-2.

[85] Ze Liu et al. "Swin transformer: Hierarchical vision transformer using shifted windows". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 10012–10022.

[86] Chiara Maria Lavinia Loeffler et al. "Artificial Intelligence–based Detection of FGFR3 Mutational Status Directly from Routine Histology in Bladder Cancer: A Possible Preselection for Molecular Testing?" In: *European Urology Focus* 8.2 (2022), pp. 472–479.

[87] Ming Y. Lu et al. "A Multimodal Generative AI Copilot for Human Pathology". In: *Nature* (June 2024), pp. 1–3. ISSN: 1476-4687. DOI: 10.1038/s41586-024-07618-3.

[88] Ming Y Lu et al. "A visual-language foundation model for computational pathology". In: *Nature Medicine* 30.3 (2024), pp. 863–874.

[89] Ming Y Lu et al. "AI-based pathology predicts origins for cancers of unknown primary". In: *Nature* 594.7861 (2021), pp. 106–110.

[90] Ming Y Lu et al. "Data Efficient and Weakly Supervised Computational Pathology on Whole Slide Images". In: *Nature Biomedical Engineering* (2020).

[91] Ming Y Lu et al. "Data-efficient and weakly supervised computational pathology on whole-slide images". In: *Nature biomedical engineering* 5.6 (2021), pp. 555–570.

[92] Ming Y. Lu et al. "Visual Language Pretrained Multiple Instance Zero-Shot Transfer for Histopathology Images". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023, pp. 19764–19775.

[93] Elo Madissoon et al. "A spatially resolved atlas of the human lung characterizes a gland-associated immune niche". In: *Nature Genetics* 55.1 (2023), pp. 66–77.

[94] Silas Maniatis et al. "Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis". In: *Science* 364.6435 (2019), pp. 89–93.

[95] Luca Marconato et al. "SpatialData: an open and universal data framework for spatial omics". In: *Nat. Methods* (Mar. 2024), pp. 1–5. ISSN: 1548-7105. DOI: 10.1038/s41592-024-02212-x.

[96] Maja Marklund. *Prostate needle biopsies pre- and post-ADT: Count matrices, histological-, and Androgen receptor immunohistochemistry images*. Version V1. 2022. DOI: 10.17632/mdt8n2xgf4.1. URL: https://data.mendeley.com/datasets/mdt8n2xgf4/1.

[97] Vivien Marx. "Method of the Year: spatially resolved transcriptomics". In: *Nature methods* 18.1 (2021), pp. 9–14.

[98] Olivier Mauduit et al. "Spatial transcriptomics of the lacrimal gland features macrophage activity and epithelium metabolism as key alterations during chronic inflammation". In: *Front. Immunol.* 13 (Oct. 2022), p. 1011125. ISSN: 1664-3224. DOI: 10.3389/fimmu.2022.1011125.

[99] Kristen R Maynard et al. "Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex". In: *Nature neuroscience* 24.3 (2021), pp. 425–436.

32

[100] D. W. McKellar et al. "Large-scale integration of single-cell transcriptomic data captures transitional progenitor states in mouse skeletal muscle regeneration". In: *Communications Biology* 4.1 (2021), p. 1280. DOI: 10.1038/s42003-021-02756-8.

[101] David W. McKellar et al. "Spatial mapping of the total transcriptome by in situ polyadenylation". In: *Nat. Biotechnol.* 41 (Apr. 2023), pp. 513–520. ISSN: 1546-1696. DOI: 10.1038/s41587-022-01517-6.

[102] Rohit Mehra et al. "Identification of GATA3 as a Breast Cancer Prognostic Marker by Global Gene Expression Meta-analysis". In: *Cancer Res.* 65.24 (Dec. 2005), pp. 11259–11264. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-05-2495.

[103] Maxime Meylan et al. "Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer". In: *Immunity* 55.3 (Mar. 2022), 527–541.e5. ISSN: 1074-7613. DOI: 10.1016/j.immuni.2022.02.001.

[104] Reza Mirzazadeh et al. *Human ileum, Visium*. Version V1. 2021. DOI: 10.17632/v8s9nz948s.1. URL: https://data.mendeley.com/datasets/v8s9nz948s/1.

[105] Reza Mirzazadeh et al. "Spatially resolved transcriptomic profiling of degraded and challenging fresh frozen samples". In: *Nat. Commun.* 14.509 (Jan. 2023), pp. 1–16. ISSN: 2041-1723. DOI: 10.1038/s41467-023-36071-5.

[106] Acadia H. M. Moeyersoms et al. "Spatial Transcriptomics Identifies Expression Signatures Specific to Lacrimal Gland Adenoid Cystic Carcinoma Cells". In: *Cancers* 15.12 (June 2023), p. 3211. ISSN: 2072-6694. DOI: 10.3390/cancers15123211. eprint: 37370820.

[107] Raktim Kumar Mondol et al. "hist2RNA: An Efficient Deep Learning Architecture to Predict Gene Expression from Breast Cancer Histopathology Images". In: *Cancers* 15.9 (Apr. 2023), p. 2569. ISSN: 2072-6694. DOI: 10.3390/cancers15092569.

[108] Taku Monjo et al. "Efficient prediction of a spatial transcriptomics profile better characterizes breast cancer tissue sections without costly experimentation". In: *Sci. Rep.* 12.4133 (Mar. 2022), pp. 1–12. ISSN: 2045-2322. DOI: 10.1038/s41598-022-07685-4.

[109] Lambda Moses and Lior Pachter. "Museum of spatial transcriptomics". In: *Nat. Methods* 19 (May 2022), pp. 534–546. ISSN: 1548-7105. DOI: 10.1038/s41592-022-01409-2.

[110] Michelli F. Oliveira et al. "Characterization of immune cell populations in the tumor microenvironment of colorectal cancer using high definition spatial profiling". In: *bioRxiv* (2024). DOI: 10.1101/2024.06.04.597233. eprint: https://www.biorxiv.org/content/early/2024/06/05/2024.06.04.597233.full.pdf. URL: https://www.biorxiv.org/content/early/2024/06/05/2024.06.04.597233.

[111] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. "Representation learning with contrastive predictive coding". In: *arXiv preprint arXiv:1807.03748* (2018).

[112] Maxime Oquab et al. "Dinov2: Learning robust visual features without supervision". In: *arXiv preprint arXiv:2304.07193* (2023).

[113] Cantin Ortiz et al. "Molecular atlas of the adult mouse brain". In: *Sci. Adv.* 6.26 (June 2020). ISSN: 2375-2548. DOI: 10.1126/sciadv.abb3446.

[114] Hubert Pakula et al. "Distinct mesenchymal cell states mediate prostate cancer progression". In: *Nat. Commun.* 15.363 (Jan. 2024), pp. 1–21. ISSN: 2041-1723. DOI: 10.1038/s41467-023-44210-1.

[115] Giovanni Palla et al. "Squidpy: a scalable framework for spatial omics analysis". In: *Nat. Methods* 19 (Feb. 2022), pp. 171–178. ISSN: 1548-7105. DOI: 10.1038/s41592-021-01358-2.

[116] Minxing Pang, Kenong Su, and Mingyao Li. "Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors". In: *bioRxiv* (Nov. 2021), p. 2021.11.28.470212. eprint: 2021.11.28.470212. URL: https://doi.org/10.1101/2021.11.28.470212.

[117] S. M. Parigi et al. "The spatial transcriptomic landscape of the healing mouse intestine following damage". In: *Nature Communications* 13.1 (2022), p. 828. DOI: 10.1038/s41467-022-28423-2.

[118] Bálint Ármin Pataki et al. "HunCRC: annotated pathological slides to enhance deep learning applications in colorectal cancer screening". In: *Sci. Data* 9.370 (June 2022), pp. 1–7. ISSN: 2052-4463. DOI: 10.1038/s41597-022-01450-y.

33

[119] Duy Pham et al. "Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues". In: *Nat. Commun.* 14.7739 (Nov. 2023), pp. 1–25. ISSN: 2041-1723. DOI: 10.1038/s41467-023-43120-6.

[120] Jill Pilet et al. "Preneoplastic liver colonization by 11p15.5 altered mosaic cells in young children with hepatoblastoma". In: *Nat. Commun.* 14 (2023). DOI: 10.1038/s41467-023-42418-9.

[121] Md Mamunur Rahaman, Ewan K. A. Millar, and Erik Meijering. "Breast cancer histopathology image-based gene expression prediction using spatial transcriptomics data and deep learning". In: *Sci. Rep.* 13.13604 (Aug. 2023), pp. 1–11. ISSN: 2045-2322. DOI: 10.1038/s41598-023-40219-0.

[122] Anjali Rao et al. "Exploring tissue architecture using spatial transcriptomics". In: *Nature* 596 (Aug. 2021), pp. 211–220. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03634-9.

[123] Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, pp. 27–30. DOI: 10.1109/CVPR.2016.91.

[124] Jing Ren et al. "Tumor protein D52 promotes breast cancer proliferation and migration via the long non-coding RNA NEAT1/microRNA-218-5p axis". In: *Annals of Translational Medicine* 9.12 (2021).

[125] Oliver Lester Saldanha et al. "Self-supervised attention-based deep learning for pan-cancer mutation prediction from histopathology". In: *NPJ Precision Oncology* 7.1 (2023), p. 35.

[126] A. Schäbitz et al. "Spatial transcriptomics landscape of lesions from non-communicable inflammatory skin diseases". In: *Nat. Commun.* 13.7729 (Dec. 2022), pp. 1–13. ISSN: 2041-1723. DOI: 10.1038/s41467-022-35319-w.

[127] Benoît Schmauch et al. "A deep learning model to predict RNA-Seq expression of tumours from whole slide images". In: *Nature Communications* 11.1 (2020).

[128] Zhuchen Shao et al. "Transmil: Transformer based correlated multiple instance learning for whole slide image classification". In: *Advances in Neural Information Processing Systems* 34 (2021).

[129] Julio Silva-Rodríguez et al. "Going deeper through the Gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection". In: *Comput. Methods Programs Biomed.* 195 (Oct. 2020), p. 105637. ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2020.105637.

[130] Andrew H. Song et al. "Artificial intelligence for digital and computational pathology". In: *Nature Reviews Bioengineering* (2023). ISSN: 2731-6092. DOI: 10.1038/s44222-023-00096-8. URL: https://doi.org/10.1038/s44222-023-00096-8.

[131] Fabio A. Spanhol et al. "A Dataset for Breast Cancer Histopathological Image Classification". In: *IEEE Trans. Biomed. Eng.* 63.7 (Oct. 2015), pp. 1455–1462. DOI: 10.1109/TBME.2015.2496264.

[132] *Spotiphy: generative modeling in single-cell spatial whole transcriptomics*. [Online; accessed 16. May 2024]. Jan. 2024. DOI: 10.5281/zenodo.10520022.

[133] Patrik L Ståhl et al. "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics". In: *Science* 353.6294 (2016), pp. 78–82.

[134] Patrik L. Ståhl et al. "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics". In: *Science* 353.6294 (July 2016), pp. 78–82. ISSN: 0036-8075. DOI: 10.1126/science.aaf2403.

[135] Alberto Valdeolivas et al. "Charting the Heterogeneity of Colorectal Cancer Consensus Molecular Subtypes using Spatial Transcriptomics". In: *bioRxiv* (Jan. 2023), p. 2023.01.23.525135. eprint: 2023.01.23.525135. URL: https://doi.org/10.1101/2023.01.23.525135.

[136] Annika Vannan et al. "Image-based spatial transcriptomics identifies molecular niche dysregulation associated with distal lung remodeling in pulmonary fibrosis". In: *bioRxiv* (2023). DOI: 10.1101/2023.12.15.571954. eprint: https://www.biorxiv.org/content/early/2023/12/17/2023.12.15.571954.full.pdf. URL: https://www.biorxiv.org/content/early/2023/12/17/2023.12.15.571954.

[137] Bastiaan S Veeling et al. "Rotation Equivariant CNNs for Digital Pathology". In: (June 2018). arXiv: 1806.03962 [cs.CV].

34

[138] Marco Vicari et al. "Spatial multimodal analysis of transcriptomes and metabolomes in tissues". In: *Nat. Biotechnol.* (Sept. 2023), pp. 1–5. ISSN: 1546-1696. DOI: 10.1038/s41587-023-01937-y.

[139] Eva Gracia Villacampa et al. "Genome-wide spatial expression profiling in formalin-fixed tissues". In: *Cell Genom.* 1.3 (Dec. 2021), p. 100065. ISSN: 2666-979X. DOI: 10.1016/j.xgen.2021.100065. eprint: 36776149.

[140] Karin E. de Visser and Johanna A. Joyce. "The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth". In: *Cancer Cell* 41.3 (Mar. 2023), pp. 374–403. ISSN: 1535-6108. DOI: 10.1016/j.ccell.2023.02.016.

[141] Andrew P. Voigt et al. "Gene Expression Within a Human Choroidal Neovascular Membrane Using Spatial Transcriptomics". In: *Invest. Ophthalmol. Visual Sci.* 64.13 (Oct. 2023), p. 40. ISSN: 1552-5783. DOI: 10.1167/iovs.64.13.40.

[142] Eugene Vorontsov et al. "A foundation model for clinical-grade computational pathology and rare cancers detection". In: *Nature medicine* (2024), pp. 1–12.

[143] Sophia J. Wagner et al. "Transformer-based biomarker prediction from colorectal cancer histology: A large-scale multicentric study". In: *Cancer Cell* 41.9 (Sept. 2023). ISSN: 1535-6108.

[144] Hongyi Wang et al. "M2ORT: Many-To-One Regression Transformer for Spatial Transcriptomics Prediction from Histopathology Images". In: *arXiv* (Jan. 2024). DOI: 10.48550/arXiv.2401.10608. eprint: 2401.10608.

[145] Shuo Wang et al. "Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning". In: *European Respiratory Journal* 53.3 (2019).

[146] Xiyue Wang et al. "Transformer-based unsupervised contrastive learning for histopathological image classification". In: *Medical Image Analysis* 81 (2022), p. 102559.

[147] Xiyue Wang et al. "TransPath: Transformer-Based Self-supervised Learning for Histopathological Image Classification". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2021, pp. 186–195.

[148] Jerry Wei et al. "A Petri Dish for Histopathology Image Analysis". In: *International Conference on Artificial Intelligence in Medicine*. Springer. 2021, pp. 11–24.

[149] F. Alexander Wolf, Philipp Angerer, and Fabian J. Theis. "SCANPY: large-scale single-cell gene expression data analysis". In: *Genome Biol.* 19.1 (Dec. 2018), pp. 1–5. ISSN: 1474-760X. DOI: 10.1186/s13059-017-1382-0.

[150] Ido Wolf et al. "FOXA1: Growth inhibitor and a favorable prognostic factor in human breast cancer". In: *International journal of cancer* 120.5 (2007), pp. 1013–1022.

[151] Yi Xiao and Dihua Yu. "Tumor microenvironment as a therapeutic target in cancer". In: *Pharmacol. Ther.* 221 (May 2021), p. 107753. ISSN: 0163-7258. DOI: 10.1016/j.pharmthera.2020.107753.

[152] Ronald Xie et al. "Spatially Resolved Gene Expression Prediction from Histology Images via Bi-modal Contrastive Learning". In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh et al. Vol. 36. Curran Associates, Inc., 2023, pp. 70626–70637. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/df656d6ed77b565e8dcdfbf568aead0a-Paper-Conference.pdf.

[153] Feng Xu et al. "Predicting axillary lymph node metastasis in early breast cancer using deep learning on primary tumor biopsy slides". In: *Frontiers in oncology* 11 (2021), p. 759007.

[154] Hanwen Xu et al. "A whole-slide foundation model for digital pathology from real-world data". In: *Nature* 630 (June 2024), pp. 181–188. ISSN: 1476-4687. DOI: 10.1038/s41586-024-07441-w.

[155] Meilin Xue et al. "Schwann cells regulate tumor cells and cancer-associated fibroblasts in the pancreatic ductal adenocarcinoma microenvironment". In: *Nat. Commun.* 14.4600 (July 2023), pp. 1–18. ISSN: 2041-1723. DOI: 10.1038/s41467-023-40314-w.

[156] Jiahui Yu et al. "CoCa: Contrastive Captioners are Image-Text Foundation Models". In: 2022.

[157] Daiwei Zhang et al. "Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology". In: *Nat. Biotechnol.* (Jan. 2024), pp. 1–6. ISSN: 1546-1696. DOI: 10.1038/s41587-023-02019-9.

35

[158] Yuqing Zhang, Giovanni Parmigiani, and W. Evan Johnson. "ComBat-seq: batch effect adjustment for RNA-seq count data". In: *NAR Genomics Bioinf.* 2.3 (Sept. 2020), lqaa078. ISSN: 2631-9268.

[159] Edward Zhao et al. "Spatial transcriptomics at subspot resolution with BayesSpace". In: *Nat. Biotechnol.* 39 (Nov. 2021), pp. 1375–1384. ISSN: 1546-1696.

[160] Weiqin Zhao et al. "Hist2Cell: Deciphering Fine-grained Cellular Architectures from Histology Images". In: *bioRxiv* (Feb. 2024), p. 2024.02.17.580852.

[161] Jinghao Zhou et al. "iBOT: Image BERT Pre-Training with Online Tokenizer". In: *International Conference on Learning Representations (ICLR)* (2022).

[162] Eric Zimmermann et al. "Virchow2: Scaling Self-Supervised Mixed Magnification Models in Pathology". In: 2024.

36