# HistoPlexer: Histopathology-based Protein Multiplex Generation using Deep Learning

Sonali Andani[1,2,4,10], Boqi Chen[1,2,6,9,10], Joanna Ficek-Pascual[1,2], Simon Heinke[1], Ruben Casanova[8], Bernard Hild[3], Bettina Sobottka[3], Bernd Bodenmiller[8], Tumor Profiler Consortium, Viktor H Koelzer[3,4,5*], Gunnar Rätsch[1,2,6,7*]

[1]Department of Computer Science, ETH Zurich, Zurich Switzerland.
[2]Swiss Institute of Bioinformatics, Zurich Switzerland.
[3]Department of Pathology and Molecular Pathology, University Hospital Zurich, University of Zürich, Zurich Switzerland.
[4]Computational and Translational Pathology Group, Department of Biomedical Engineering, University of Basel, Basel Switzerland.
[5]Institute of Medical Genetics and Pathology, University Hospital Basel, Basel Switzerland.
[6]AI Center, ETH Zurich, Zurich Switzerland.
[7]Medical Informatics Unit, University Hospital Zurich, Zurich Switzerland.
[8]Department of Quantitative Biomedicine, University of Zurich, Zurich Switzerland.
[9]Computer Vision Laboratory, Dept. of Inf. Tech. and Electrical Eng., ETH Zurich, Zurich Switzerland.
[10]These authors contributed equally: Sonali Andani, Boqi Chen.

*Corresponding author(s). E-mail(s): viktor.koelzer@usb.ch; gunnar.raetsch@inf.ethz.ch;

**Abstract**

Multiplexed imaging technologies provide crucial insights into interactions between tumors and their surrounding tumor microenvironment (TME), but their widespread adoption is limited by cost, time, and tissue availability. We introduce HistoPlexer, a deep learning (DL) framework that generates spatially-resolved

protein multiplexes directly from histopathology images. HistoPlexer employs the conditional generative adversarial networks with custom loss functions that mitigate slice-to-slice variations and preserve spatial protein correlations. In a comprehensive evaluation on metastatic melanoma samples, HistoPlexer consistently outperforms existing approaches, achieving superior Multiscale Structural Similarity Index and Peak Signal-to-Noise Ratio. Qualitative evaluation by domain experts demonstrates that the generated protein multiplexes closely resemble the real ones, evidenced by Human Eye Perceptual Evaluation error rates exceeding the 50% threshold for perceived realism. Importantly, HistoPlexer preserves crucial biological relationships, accurately capturing spatial co-localization patterns among proteins. In addition, the spatial distribution of cell types derived from HistoPlexer-generated protein multiplex enables effective stratification of tumors into immune hot versus cold subtypes. When applied to an independent cohort, incorporating additional features from HistoPlexer-generated multiplexes enhances the performance of the DL model for survival prediction and immune subtyping, outperforming the model reliant solely on Hematoxylin & Eosin (H&E) image features. By enabling the generation of whole-slide protein multiplex from the H&E image, HistoPlexer offers a cost- and time-effective approach to understanding the TME, and holds promise for advancing precision oncology.

# 1 Introduction

Tumors are complex systems that obtain hallmark traits by creating a supportive tumor microenvironment (TME) which facilitates tumorigenesis and metastasis [1, 2]. Understanding cancer cell interactions with this surrounding tissue provides insights into disease progression and therapeutic response [3–5]. Multiplexed immunohistochemistry and immunofluorescence (mIHC/IF) technologies, such as Imaging Mass Cytometry (IMC), allow for spatially-resolved quantification of up to 40 protein markers, offering comprehensive insights into tumor-TME interactions [4, 6, 7]. These technologies facilitate analysis of spatial cell distribution, phenotype co-localization, and interactions in cellular communities—promising factors for clinical decision-making [4, 5, 8, 9]. However, IMC is limited by low throughput, high cost, and coverage restricted to small Region-of-Interests (RoIs), hindering its broader clinical adoption.

In contrast, Hematoxylin & Eosin (H&E) staining remains the gold standard for cancer diagnosis in clinical practice due to its low-cost, high throughput, and coverage of entire tissue sections. H&E images reveal crucial morphological features of tissue organization that aid in cancer grading, proliferation assessment, and staging [10]. Recent advances in Deep Learning (DL) have shown that these features can inform the prediction of protein markers. For instance, several studies have successfully predicted single markers such as pan-cytokeratin for pancreatic cancer [11], HER2 for breast cancer [12], and Ki-67 for neuroendocrine and breast cancers [13] directly from H&E images. Only a few studies have attempted a multiplexed prediction, with a focus, however, solely on either tumor [14, 15] or immune markers [16], limiting their utility for investigation of tumor-TME interactions. In addition, these studies either

employ separate models for each marker [14, 16] or lack quantitative validation on the advantages of multiplexed prediction with a single model [15, 16].

To address these limitations, we introduce HistoPlexer, a DL model that generates protein multiplexes from H&E images. HistoPlexer simultaneously predicts 11 markers, consisting of both tumor and immune markers, which enables an integrative visualization of tumor-host interactions. We train HistoPlexer on metastatic samples from the Tumor Profiler Study (TuPro) [17] using paired H&E and IMC images from serial sections. Through quantitative evaluation, we demonstrate the importance of simultaneous marker prediction through improved model performance and enhanced spatial co-localization of markers. We validate the biological relevance of generated IMC images through cell-typing and immune phenotyping analyses, particularly in characterizing immune-hot (inflamed) and immune-cold (excluded/desert) tumors based on CD8+ T-cell distributions. We also demonstrate out-of-distribution generalizability of HistoPlexer on samples from the human skin cutaneous melanoma (SKCM) study of The Cancer Genome Atlas (TCGA) project [18].

Our results show that HistoPlexer generates high-quality IMC images that closely align with real data distributions. These generated multiplexes enable precise immune phenotyping through spatial analysis of tumor-immune cell interactions, particularly in distinguishing immune-hot and cold subtypes. We also demonstrate that simultaneously predicting multiple protein markers preserves biologically meaningful relationships among them. Furthermore, by augmenting H&E Whole-Slide Images (WSIs) with generated IMC multiplex, HistoPlexer improves both survival and immune subtype prediction on the TCGA-SKCM dataset, indicating its potential to aid clinical decisions.

# 2 Results

## 2.1 HistoPlexer: a toolkit for histopathology-based protein multiplex generation

The HistoPlexer is a generative model based on conditional GAN (cGAN) which predicts spatially-resolved profiles of multiple proteins simultaneously from a single input H&E image. The model is trained on paired H&E and multiplexed IMC image patches (Figure 1A) extracted from aligned H&E and IMC RoIs. During training, the H&E patches are fed into the *translator* G, which learns to generate protein multiplexes (*i.e.*, IMC images) based on the tissue morphology from high-resolution H&E images. The generated IMC image patches, along with the input H&E image patches, are fed to the *discriminator* D to produce a realness score, which produces a realness score indicating how closely the generated IMC patches resemble ground truth (GT) IMC patches (Fig. 1B(i)). The translator and discriminator is trained adversarially using a least squares Generative Adversarial Network (GAN) loss, such that the generated IMC image patches are able to fool the discriminator to classify it as real. Besides the GAN loss, we incorporate two additional losses to ensure pixel-level and patch-level consistency between the generated and GT IMC images. The pixel-level consistency loss calculates the $L_1$ distance between the generated and GT IMC images. However, since the H&E and GT IMC images are obtained from serial sections of the tissue block,

there is a degree of spatial displacement of tissue organization between consecutive slices (termed slice-to-slice variations). While registered at the structural level after template-matching, consecutive slides obtained from real-world diagnostic material are not pixel-level aligned. To account for these differences, we adopt the Gaussian Pyramid loss [12], which relaxes the alignment constraint by evaluating the similarity between the generated and GT IMC images at multiple scales (Fig. 1B(ii)). For patch-level consistency, we utilize a patch-wise contrastive loss to ensure that corresponding patches in the generated and GT IMC images are closer in the embedding space than distant ones (Fig. 1B(iii)). We further incorporate adaptive weights for different patches based on their proximity to GT following [19].

We build our HistoPlexer framework using a multimodal metastatic melanoma dataset generated by the Tumor Profiler Study [17]. Each patient was characterized by multiple modalities, including H&E and IMC images. RoIs of 1 $mm^2$ were selected on each H&E WSI based on visual inspection by a pathology expert and IMC data was generated for those RoIs on a consecutive section of the same tumor block. Using template matching [20], we created a paired dataset of 336 H&E and IMC RoIs from 78 patients. We focus on predicting 11 protein markers that are essential for characterizing the tumor and its surrounding TME. These include tumor markers (MelanA, S100, gp100, SOX10), immune markers (CD3, CD8a, CD20, CD16, CD31), and antigen-presentation markers (HLA-ABC, HLA-DR).

## 2.2 HistoPlexer generates accurate and realistic protein multiplex.

We benchmark the HistoPlexer against Pix2pix [21] and PyramidP2P [12], evaluating each method in two settings: multiplex (MP) and singleplex (SP). In the MP setting, a single model is trained to predict all markers simultaneously, whereas in the SP setting, separate models are trained to predict each marker individually, after which the predictions are stacked for a (pseudo-)multiplexed output. All models are trained on 231 and tested on 105 RoIs.

We evaluate the quality of generated IMC images using Multiscale Structural Similarity Index (MS-SSIM) [22] for perceptual similarity at multiple scales and Peak Signal-to-Noise Ratio (PSNR) [23] for pixel-level distortion. Our results show that the HistoPlexer model trained in the MP setting achieves the highest MS-SSIM and PSNR values (refer Table 1), suggesting greater similarity to GT IMC images generated from consecutive tissue sections. Additionally, models in the MP setting consistently outperforms those in the SP setting across all methods, demonstrating that simultaneous prediction of all markers enhances performance by effectively capturing inter-marker correlations. The performance of individual markers for the HistoPlexer-MP model is presented in Table S1.

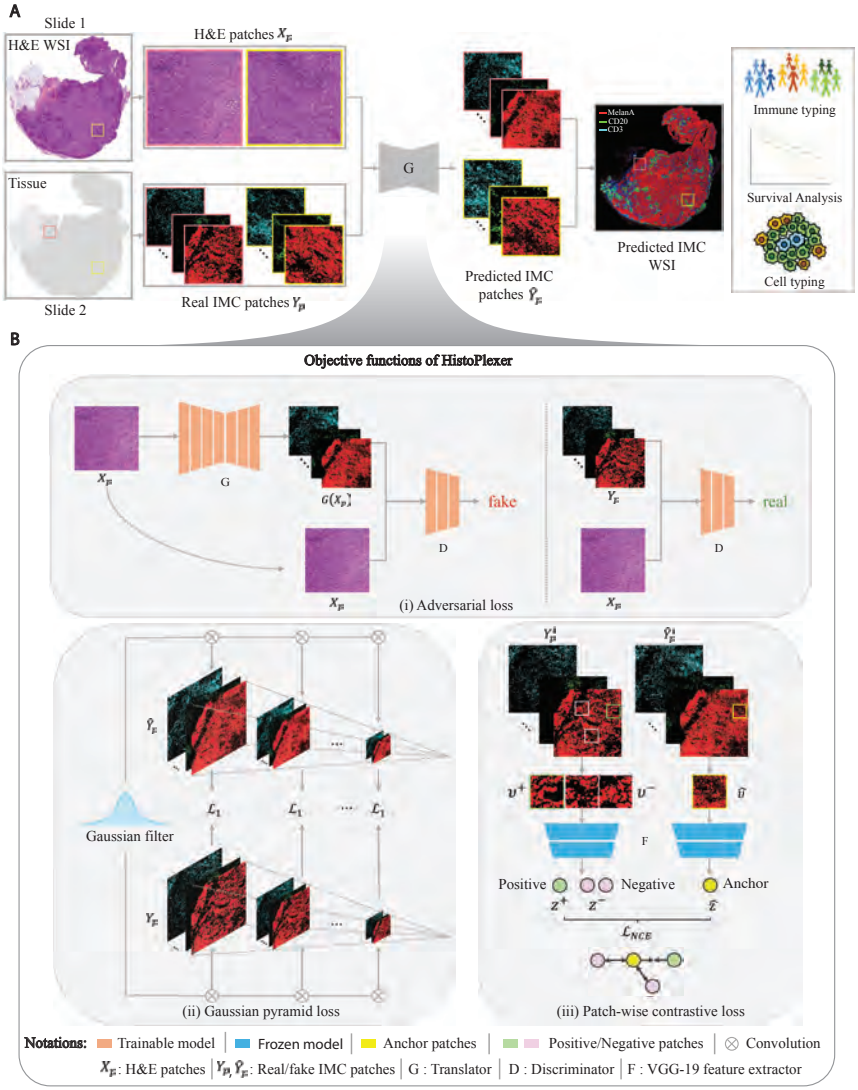



**Fig. 1 Overview of HistoPlexer architecture. (A)** The HistoPlexer consists of a translator G that takes H&E and IMC images as input and predicts protein multiplexes from morphology information encoded in the H&E images, ultimately generating protein multiplex on the WSI level from H&E input. **(B)** The objective functions of HistoPlexer contain the GAN adversarial loss, gaussian pyramid loss with average L1 score across scales and patch-wise contrastive loss with anchor from generated IMC and positive and negative from GT IMC.

We further qualitatively evaluate the generated IMC images by comparing them with the GT (Fig. 2A and Supplementary Fig. S1) and observe good alignment in global patterns. However, pixel-level correspondence is not expected due to the inherent slice-to-slice variations. In a few cases, we observe slight confusion between CD20 and CD3/CD8a markers. For instance, in the bottom-right region of Fig. 2A (ii), there

exists an overexpression of CD20 and an underexpression of CD3 and CD8a markers. This may stem from the highly similar and visually indistinguishable morphology of B- and T- cells in H&E images, leading to confusion between their markers (CD20 for B-cells and CD3/CD8a for T-cells) [24].

To quantify the perceived realism of generated IMC images, we employ the Human Eye Perceptual Evaluation (HYPE) framework [25] where experts evaluate pairs of IMC images (real or generated) for specific markers alongside their corresponding H&E images. Given that H&E staining reveals distinct nuclear and tissue morphology patterns crucial for identifying tumor regions and lymphocytes [24], we created two evaluation sets: tumor-associated markers (MelanA, S100, gp100, SOX10) and lymphocyte markers (CD20, CD3, CD8a). For each set, two pathology experts assessed 250 image pairs, with an equal distribution of real and generated images. The image pairs were created using RoIs from test set, with data augmentation through small translations and rotations. The evaluation yields mean HYPE scores of 41.8%($\pm$0.3%) for lymphocyte markers and 42.8%($\pm$0.6%) for tumor markers. The generated images achieved HYPE scores of 61.6% ($\pm$1.3%) and 72.8% ($\pm$1.1%), indicating that the majority (>50%) were perceived as real by domain experts, demonstrating their high perceived realism.

Next, we go beyond pixel-level evaluation by identifying relevant cell types. We use GT cell-type annotations from the GT IMC training set, following [8], and train a Random Forest classifier [26] based on average marker expression per cell to classify them into five classes: tumor cells, B-cells, CD8+ T-cells, CD4+ T-cells, and others. This classifier is then applied to both GT and generated IMC images from the test set to obtain cell-type maps (Fig. 2B). We visualize RoIs from the tumor center and the tumor front at the tumor–TME interface and examine spatial patterns based on immune subtype labels. We observe that immune "hot" tumors, characterized by high immune cell infiltration, show strong interactions between tumor and CD8+ T-cells (Fig.2B(i)), whereas immune "cold" tumors, with low immune presence, display minimal immune cell interaction, especially in the tumor center (Fig.2B(ii)). Immune "cold" RoIs at the tumor front similarly exhibit sparse or clustered immune cells with little interaction with tumor cells (Fig.2B(iii), (iv), (v)). The strong alignment between predicted and GT cell-type maps, as well as their spatial organization, suggests that

| | Method | MS-SSIM ↑ | PSNR ↑ |
|---|---|---|---|
| MP | Pix2pix [21] | 0.278$_{\pm 0.004}$ | 13.747$_{\pm 0.122}$ |
| | PyramidP2P [12] | 0.284$_{\pm 0.004}$ | 13.894$_{\pm 0.172}$ |
| | HistoPlexer | **0.299**$_{\pm 0.003}$ | **14.162**$_{\pm 0.076}$ |
| SP | Pix2pix [21] | 0.260$_{\pm 0.002}$ | 13.015$_{\pm 0.009}$ |
| | PyramidP2P [12] | 0.263$_{\pm 0.015}$ | 13.216$_{\pm 0.482}$ |
| | HistoPlexer | 0.279$_{\pm 0.002}$ | 13.353$_{\pm 0.038}$ |

**Table 1** Comparison of Model Performance against benchmarks using MS-SSIM and PSNR for multiplex (MP) and singleplex (SP) settings. ↑ arrow indicates higher values are better.

HistoPlexer effectively captures morphological features in H&E images relevant for predicting cell types using IMC data.

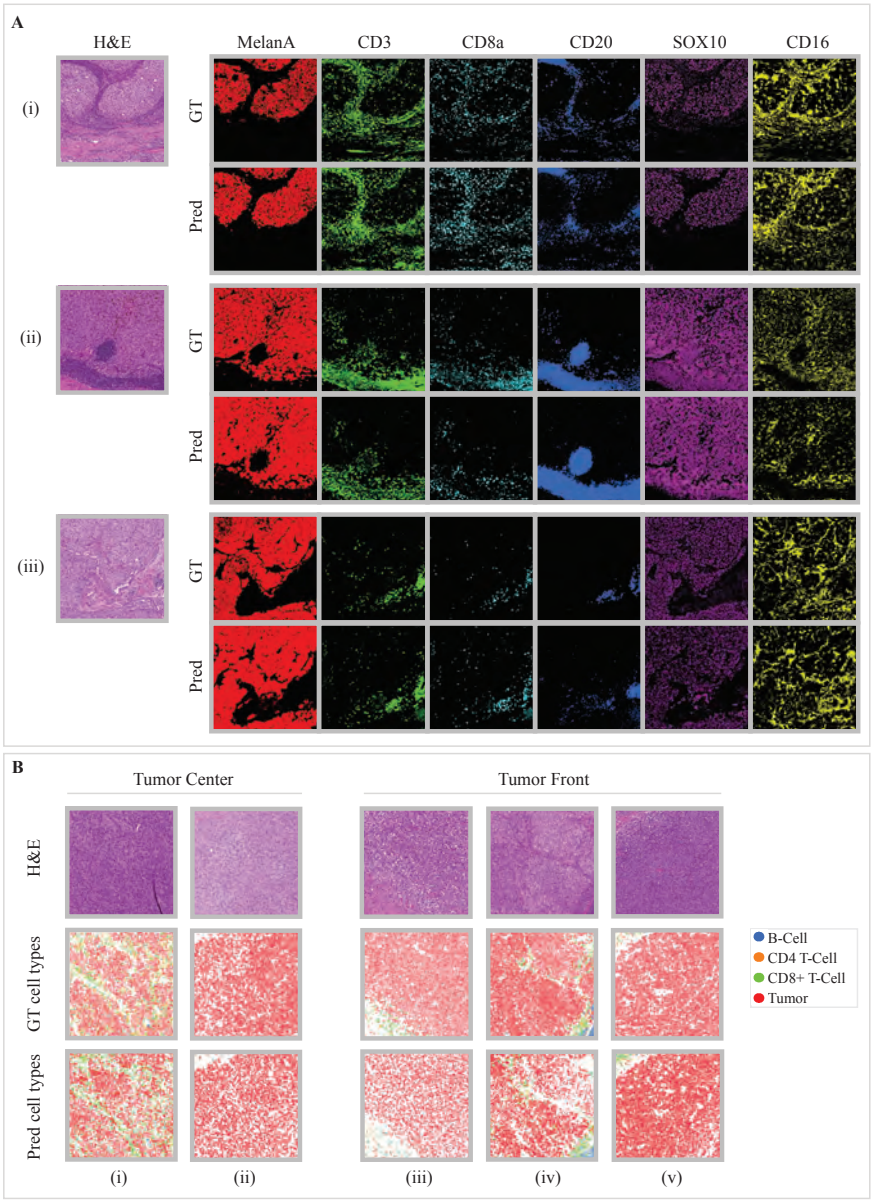



**Fig. 2 Qualitative RoI-level assesment of HistoPlexer**. **A** H&E (first column) and expression profiles of individual markers: MelanA, CD3, CD8a, CD20, SOX10 and CD16 (from second to last column). Top row: ground-truth (GT) expression profiles; bottom row: predicted (Pred) expression profiles. **B** Cell-typing results: H&E (first row), GT and predicted cell types (middle and bottom row) in RoIs grouped by their location within the tissue: "Tumor Center" and "Tumor Front".

## 2.3 HistoPlexer preserves spatial co-localization patterns

As importance of spatial patterns has been previously shown by [27, 28], we assess the spatial co-localization patterns by quantifying the correlation between two or more proteins markers simultaneously expressed within a given region. For each protein pair, we compute the Spearman's Correlation Coefficient (SCC) between the two proteins and average the correlation across RoIs, considering only pairs with strong positive ($> 0.15$) or strong negative ($< -0.15$) correlation in GT IMC images. We then compare the SCC obtained from GT and generated IMC multiplex.

As shown in Fig. 3A(i), the Multiplex (MP) model's predictions align more closely with the GT than those of the Singleplex (SP) model in terms of pairwise SCC, especially for protein pairs involving CD-based immune markers such as CD16:HLA-DR, CD3:HLA-ABC and CD16:CD8a, which are sparsely represented in the training data. We hypothesize these sparse markers lack sufficient tissue context for the SP model to generate accurate predictions. In contrast, the MP model benefits from learning inter-marker correlations by predicting all markers simultaneously. Leveraging auxiliary tissue morphology information from abundant markers, it enhances the prediction of both sparse markers and co-localization patterns. However, for a few protein pairs (CD3:CD8a and CD20:CD3), the SCC in MP exceeds that of the GT. This is likely due to the similar morphological features of CD8+ T-cells (a subset of CD3 T-cells) and CD3 T-cells, as well as of B-cells (CD20) and CD3 T-cells in H&E images [24], which can lead to the overprediction of sparse markers and, consequently, co-localization patterns. We further quantify spatial co-localization by measuring the Mean Square Error (MSE) between the SCC values from GT and generated IMC data across all test RoIs (Fig.3A(ii)). Compared to the SP model, the MP model achieves an MSE that is approximately an order of magnitude lower, which reinforces our hypothesis. A comparison of HistoPlexer with Pix2Pix[21] and PyramidP2P [12] baselines is provided in Supplementary Fig. S2A.

To explore spatial patterns beyond protein pairs, we visualize the expression profiles using t-SNE embeddings of cells from both GT and generated IMC multiplex, following [29]. We observe a good correspondence between t-SNE from both GT and generated IMC multiplex (Fig.2.3B). For instance, cells that are positive for CD3 and CD8a are at the same time negative for CD31, gp100 and MelanA. This is in line with their biological function, as CD3 and CD8a are expressed on T-cells but not on endothelium (CD31) or tumor cells (gp100 and MelanA). Full t-SNE plots for all markers are shown in Supplementary Fig. S2.

In conclusion, our quantitative and qualitative results suggest that the spatial co-localization patterns in GT can be effectively replicated using the generated IMC images. These spatial patterns are preserved across tissue sections, thus offering a robust evaluation metric that mitigates the impact of slice-to-slice variations.

## 2.4 HistoPlexer enables multiplexed proteomics profiling on the WSI-level.

HistoPlexer enables the generation of IMC images from H&E WSIs of up to 100,000×100,000 pixels, allowing for the simultaneous visualization of multiple protein

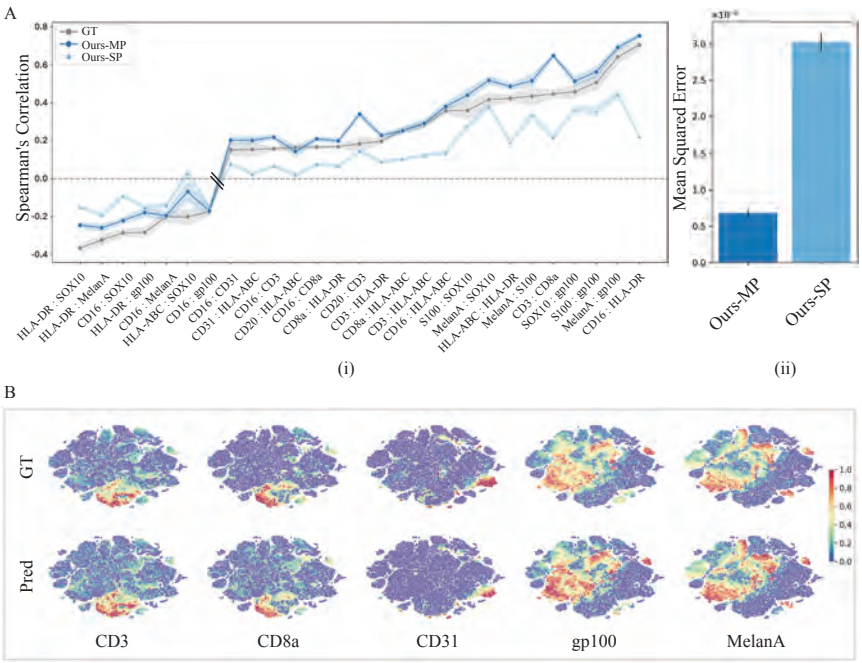



**Fig. 3** **A**(i) Spearman's correlation coefficients between protein pairs, comparing the ground truth (GT) with both singleplexed (SP) and multiplexed (MP) predictions of the HistoPlexer. The pairs on the X-axis are ordered by increasing Spearman's correlation in the GT. **A**(ii) Mean squared error between the GT and predicted Spearman's correlation coefficients, comparing the SP and MP predictions of the HistoPlexer. **B** Joint t-SNE visualization of protein co-localization patterns for selected markers: CD3, CD8a, CD31, gp100 and MelanA. The color represents protein expression.

markers across entire tissue sections. This capability provides a comprehensive view of tumor and TME interactions at the WSI level. Since GT IMC data is available only for RoIs, we use Ultivue's InSituPlex® technology to obtain multiplexed WSIs using the Immuno8 and MDSC FixVue™ panels. These panels include key markers, such as SOX10 for tumors, HLA-DR for antigen presentation, and CD3/CD8a for T-cell profiling, which are shared with the generated protein multiplex. Figure 4 provides a qualitative comparison between the generated IMC and Ultivue multiplex at the WSI level. In both cases, a strong correspondence in global structures and hotspot regions is observed across all markers. In Fig. 4(ii), while there is good alignment for CD3 and SOX10 markers, discrepancies appear for CD8A and HLA-DR, particularly along the tissue periphery (e.g., the bottom-left border). These differences are likely due to slice-to-slice variations between H&E and Ultivue images, which lead to slight shifts in tissue boundaries.

## 2.5 HistoPlexer facilitates immune phenotyping

We showcase the utility of HistoPlexer by stratifying immune subtypes according to the spatial distribution of CD8+ T-cells obtained using only H&E images from TuPro metastatic melanoma samples. Fig.5A illustrates the integrative visualization

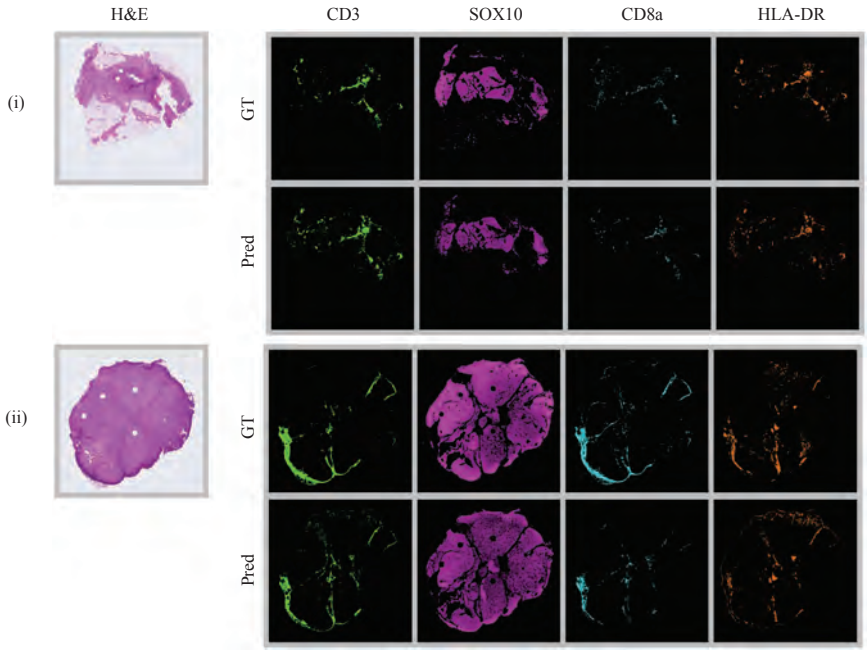



**Fig. 4 Qualitative WSI-level assessment of HistoPlexer**. H&E (first column) and expression profiles of individual markers: CD3, SOX10, CD8a and HLA-DR (from second to last column). Top row: GT expression profiles from Ultivue images; bottom row: predicted (pred) expression profiles on WSI level both samples in (i) and (ii).

of predicted tumor and CD8+ T-cells on H&E WSIs. In immune-hot cases, characterized by substantial CD8+ T-cell infiltration and typically better immunotherapy responses [30, 31], we observe the presence of both attacker tumor cells and infiltrating CD8+ defender T-cells within the tumor region, indicating active immune response. Conversely, immune-cold cases show minimal or no CD8+ T-cell infiltration in the tumor area, which generally correlates with poor immunotherapy outcomes. Building upon the immune subtype classification approach developed in [5], we further obtain intratumoral (iCD8) and stromal (sCD8) CD8+ T-cell densities in tumor center compartment after localizing CD8+ T-cells using HistoPlexer. For this, we annotated the tumor center compartment and segmented it into an intratumoral and stromal regions using HALO$^{AI}$ platform across 34 TuPro metastatic melanoma samples.

Fig. 5B(i) shows stratification of immune subtypes using iCD8 and sCD8 densities measured per $\mu m^2$. We observe that immune desert cases exhibit very low iCD8 and sCD8 density, indicating the presence of only rare or isolated CD8+ T-cells. Immune excluded cases also show very low iCD8 density but slightly higher sCD8 density compared to immune desert cases, suggesting some CD8+ T-cells have reached the stroma but not the intratumoral regions. Inflamed cases display high densities of both iCD8 and sCD8, indicating the presence of CD8+ T-cells in the stromal compartment and, most importantly, their infiltration into intratumoral regions. These observations align with the findings in [5], demonstrating the utility of our model. When assessing the

clinical relevance in distinguishing immune-hot (inflamed) and immune-cold (excluded and desert) cases, we find that both iCD8 and sCD8 densities are lower in immune-cold and higher in immune-hot cases (Fig. 5B(ii)). Additionally, we trained a random forest classifier to differentiate immune-hot and -cold cases and achieved F1 score of 0.873 (SD 0.006) and macro-average AUROC of 0.845 (SD 0.047) over 5-fold cross-validation. In conclusion, we demonstrate the capability of the HistoPlexer for immune phenotyping, which has potential implications for treatment recommendations.

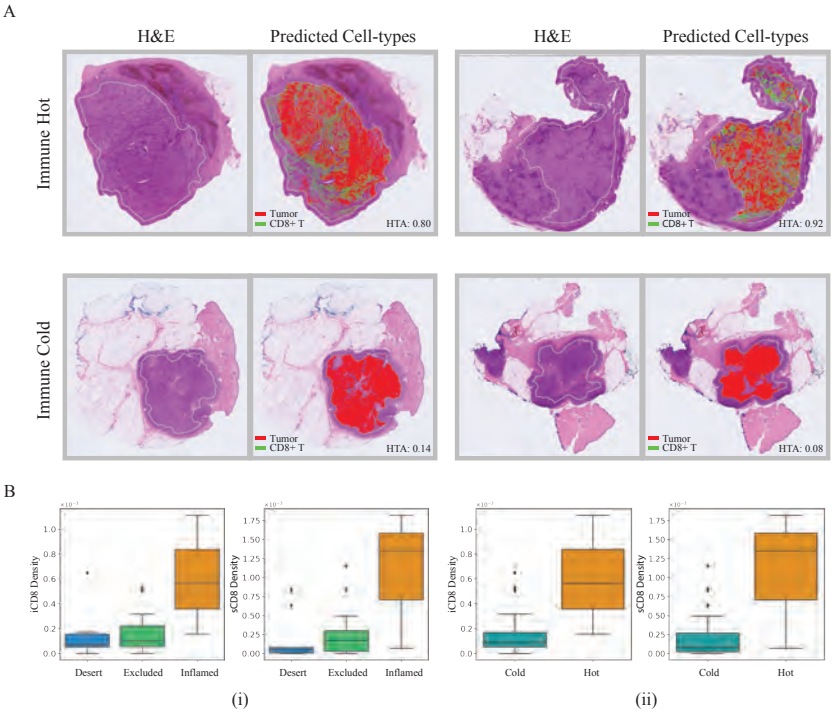



**Fig. 5 Immune phenotyping using HistoPlexer**. **A** H&E image along with overlay of predicted tumor and CD8+ T-cells within tumor center region using HistoPlexer model for two immune hot and two immune cold cases from TuPro metastatic melanoma cohort. **B**(i) Box plot of intratumoral (iCD8) and stromal (sCD8) CD8+ T-cell densities in tumor center compartment, stratified by immune desert, excluded and inflamed classes. **B**(ii)Box plot of intratumoral (iCD8) and stromal (sCD8) CD8+ T-cell densities in tumor center compartment, stratified by immune hot and cold classes.

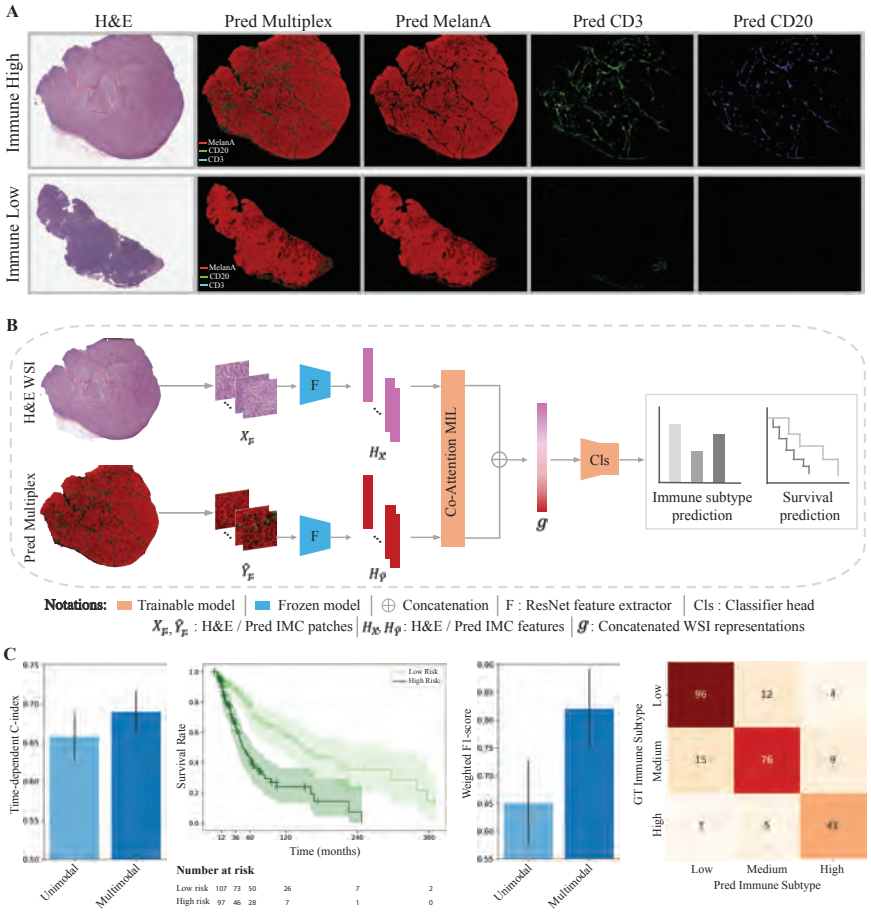
## 2.6 HistoPlexer generalizes to independent patient cohort data

We evaluate the generalizability of the HistoPlexer model on Out-of-Distribution (OOD) data from an independent TCGA-SKCM cohort [18]. Fig. 6A displays the generated protein multiplex at the WSI level, along with expression profiles for three markers: tumor-associated MelanA, T-cell marker CD3, and B-cell marker CD20. In the immune-high sample, we observe higher expression and tumor infiltration of

CD3 and CD20 markers, contrasting with the minimal or absent expression in the immune-low case, where immune labels are based on RNAseq expression [32].

Next, we assess the utility of generated IMC in augmenting clinical outcome prediction using expression profiles from MelanA, CD3 and CD20 markers due to their known prognostic significance [33, 34]. We encode the H&E and generated IMC WSIs using pretrained feature extractors. The features are input to an attention-based Multiple Instance Learning (MIL) predictor [35]. We train the MIL predictor under two settings: (1) the unimodal setting, where only H&E features are input to the predictor and (2) the multimodal setting, where features extracted from the corresponding H&E and predicted IMC patches are first aggregated via a co-attention layer [36], and the bag-level representations of H&E and predicted IMC WSIs after the MIL pooling layer are concatenated before fed into the classification head (Fig. 6).

We perform two clinically relevant tasks: immune subtype and survival prediction. For the survival prediction, we use the disease-specific survival from patients' metadata as it provides a more accurate representation of the patient's disease status [37]. For the immune subtype prediction, we classify the patients into three immune subgroups: low, intermediate and high with ground-truth labels obtained using Bulk RNA-seq expression data [32]. Overall, we observe the predictive performance of the multimodal setting to be superior to that of the unimodal setting for both tasks. Specifically, for the survival prediction task, incorporating features from predicted IMC images leads to an improvement of 3.18% in average time-dependent C-index [38] over 5-fold cross-validation. We further visualize the Kaplan-Meier survival curves for the multimodal setting, in which patients are separated into two groups of low-risk and high-risk based on predicted risk scores (Definition in 4.6). The logrank statistical significance test to determine if the separation between low and high-risk groups is statistically significant (p-value $= 5.05 \times 10^{-7}$). For the immune subtyping task, using features from both modalities demonstrates an improvement of 17.02% in terms of average weighted F1 score over 5-fold cross-validation. These results demonstrate not only the generalizability of the HistoPlexer to OOD samples, but also the clinical utility of the generated protein expression profiles by HistoPlexer in augmenting clinical decisions.




(i) Survival prediction TCGA-SKCM (ii) Immune subtype prediction TCGA-SKCM

**Fig. 6 OOD generalization. (A)** Two examples (immune-high and -low) from the TCGA-SKCM cohort, showing H&E images (first column), predicted protein multiplexes (second row) as well as expression profiles of MelanA, CD3 and CD20 markers (last three colums). **(B)** Model architecture for multimodal survival and immune subtype prediction. **(C)(i)** Survival prediction results, displaying time-dependent c-index scores (left) and Kaplan-Meier survival curves for the multimodal setting, with separation of low- and high-risk groups (right).; **(C)(ii)** Immune subtype prediction results, showing the weighted F1 score (left) and confusion matrix (right) for classification into low, intermediate, and high immune subtypes.

# 3 Discussion

In this study, we introduce HistoPlexer, a generative model that enables prediction of a high order (11) of multiplexed protein expression profiles, including both tumor and immune markers, directly from H&E images. Our approach addresses the challenge of predicting multiplexed IMC data, where individual protein markers lack the structural details available in conventional Immunohistochemistry (IHC) images. By simultaneously predicting multiple proteins, our model successfully captures sparse markers

and preserves biologically meaningful relationships, as validated through spatial correlation analysis of protein co-localization patterns. Our comprehensive evaluation demonstrates that the multiplexed prediction approach consistently outperforms singleplex alternatives, evidenced by higher MS-SSIM and PSNR values, and lower MSE of protein co-localization SCC compared to GT. Notably, the domain experts found the generated IMC images highly realistic, with HYPE error rates of 61.6% and 72.8% for lymphocyte and tumor markers, respectively, supporting the quality of our predictions.

The clinical utility of HistoPlexer is demonstrated through two key applications. First, HistoPlexer enables immune phenotyping at WSI level by quantifying spatial patterns using intratumoral (iCD8) and stromal (sCD8) CD8+ T-cell densities in the tumor center compartment. We found the spatial patterns in concordance with state-of-the-art approach [5], showcasing the utility of our model. We also successfully stratify patients into clinically actionable immune hot and cold subtypes. This capability is particularly valuable for immunotherapy decisions, where understanding the spatial distribution of CD8+ T-cells is crucial. Second, HistoPlexer shows generalizability to OOD data through evaluation on the independent TCGA-SKCM cohort. The integration of HistoPlexer-generated protein expression profile features with H&E features consistently improves the performance of DL-based predictive models in both survival (3.18% increase in time-dependent C-index) and immune subtype prediction (17.02% increase in weighted F1 score), demonstrating the potential of HistoPlexer in augmenting clinical decision-making.

The study has some limitations. First, in some cases the model confuses between T-cells CD3/CD8a and B-cell CD20 markers which have similar morphological features. While this is not an issue for many downstream tasks such as survival and immune subtype prediction, for more fine-grained analyses, such as distinguishing between closely related cellular subsets, our model may face limitations. Thus, it is a priority for future work to refine the model's ability to accurately distinguish between these finer subsets of cells. Second, we showed possibility to obtain major cell-types such as Tumor, B-cells, CD8+ T-cells and CD4+ T-cells. This set could be further extended to include more sparse cell-types such as endothelial cells by obtaining a larger training cohort. Third, for multimodal training on the TCGA-SKCM dataset, we used MelanA, CD3 and CD20 markers from generated protein multiplex. The choice of these lineage markers was based on their high level of information content for lymphocyte subpopulations and identification of tumor cells, however, this set could be potentially extended to study the importance of other markers towards survival and immune subtyping tasks. Lastly, due to slice-to-slice variations in data, we focused on the model's utility in downstream tasks rather than strict pixel-level correspondence.

HistoPlexer opens several promising research directions. First, expanding the framework to additional protein markers and cancer types could uncover valuable insights into disease mechanisms and treatment responses without requiring additional tissue material or incurring significant costs. By utilizing HistoPlexer on existing H&E images from clinical trials and population cohorts, it could support high-throughput workflows and offer comprehensive insights into spatial biology patterns correlated with clinical responses and epidemiological trends. Second, by making the Ultivue InSituPlex® dataset generated for this study publicly available, we invite researchers

to explore novel diffusion models for multiplexed protein marker generation, particularly those that account for slice-to-slice variations. Third, integrating generated protein multiplex with other molecular data modalities holds potential for enhancing our understanding of tumor biology and improving patient stratification, thereby supporting personalized treatment strategies. Finally, as computational pathology continues to advance, tools like HistoPlexer will play an increasingly important role in bridging the gap between routine histological analysis and advanced molecular profiling, ultimately contributing to more precise and personalized cancer treatment strategies.

In conclusion, HistoPlexer represents a significant advance in computational pathology, enabling the cost-effective generation of protein multiplexes from clinically established histology slides. Our promising results support further efforts toward clinical application, with the potential to transform cancer diagnosis and treatment planning for more personalized patient care.

# 4 Methods

## 4.1 Datasets and preprocessing

### 4.1.1 Tumor Profiler dataset

We build our HistoPlexer framework using a subset of highly multi-modal metastatic melanoma dataset generated by the Tumor Profiler Study (TuPro) [17]. Each patient was characterised using multiple technologies, including Digital Pathology and IMC. A total of six RoIs of 1 $mm^2$ were selected on each H&E WSI, three within tumor center and three at the tumor front (intersection of tumor and TME). IMC data was generated for those six RoIs on a consecutive section of the same tumor block. The IMC data was generated at a resolution of 1µm/pixel and H&E images were scanned at a resolution of 0.25 µm/pixel. Therefore, RoIs of 1 $mm^2$ are represented by 1000 pixels for IMC data and 4000 pixels for H&E images. Since the paired data was generated by visually choosing RoIs, in many cases a considerable positional shift and rotation between the specified H&E regions and the resulting IMC regions can be observed. This was overcome by using template matching [39], resulting in a paired dataset of 336 H&E and IMC ROIs from 78 patients for training and testing model performance.

IMC profiling was performed using a panel of 40 antibodies, from which 11 have been selected for this study based on the biological function of the corresponding proteins as well as high signal–to–noise ratio. The proteins targeted by the 11 antibodies include cell-type markers, such as tumor markers (MelanA, gp100, S100, SOX10), lymphocyte markers (CD20, CD16, CD3, CD8a) and an endothelial marker (CD31). Moreover, two functional markers corresponding to proteins involved in antigen presentation (HLA-ABC, HLA-DR) are included in the protein set.

The raw IMC images were processed with CellProfiler software for cell segmentation [40]. The protein counts extracted from the images have been first clipped to 99.9% per protein to exclude outliers ad then transformed using the $arcsinh$-function with cofactor one [41]. In order to exclude background noise, we apply OTSU

thresholding [42] with kernel size three and sigma three and the threshold, separating signal from background, determined per sample using all available RoIs. The resulting data per protein is first centered and standardized and then subjected to min-max-transformation, all using data statistics based on the train set only.

The data is split at the patient level into train and test set, stratified by immune phenotype (inflamed, immune excluded, and immune desert). The stratification ensures the representation of both tumor and immune cells in each set. The patient-level splitting guarantees that all RoIs from a given patient belong to only one set, preventing undesired information flow. The resulting train and test sets consist of 231 and 105 RoIs, respectively. During model training, RoIs are chosen at random and a tile of size 1024×1024 from H&E image and a corresponding IMC region of 256×256 is extracted.

For WSIs predictions, tissue segmentation is performed on the input H&E WSI by using OTSU thresholding [42]. Each segmented tissue region is then divided into tiles of size 1024×1024 pixels. The tiles undergo stain normalization using the Macenko method [43] to minimize staining variability and maintain color consistency across images. The generated IMC tiles are then stitched together to obtain WSI level IMC multiplex.

### 4.1.2 Ultivue dataset

For qualitative evaluation of HistoPlexer on WSIs, we employed Ultivue InSituPlex® technology to obtain multiplexed images using the Immuno8 and MDSC FixVue™ panels. The Immuno8 panel focuses on immune landscape characterization with markers such as CD3, CD4, CD8, CD68, PD-1, PD-L1, FoxP3, and PanCK/SOX10. The MDSC panel identifies myeloid-derived suppressor cells using markers CD11b, CD14, CD15, and HLA-DR. Ultivue images were acquired at a resolution of 0.325 µm/pixel. For evaluation, we used CD3, SOX10, CD8a, and HLA-DR markers to assess visual similarity between the generated protein multiplex and Ultivue images.

Paired H&E and Ultivue WSIs were generated by first staining H&E on one tissue section, followed by acquiring Immuno8 and MDSC data on consecutive sections for 10 samples. A tonsil tissue was included with each sample as a positive control. Image registration between H&E and Ultivue WSIs was performed using an unsupervised multimodal method [44], leveraging the DAPI nuclear stain in Ultivue for alignment with H&E images. Both Ultivue and generated IMC images underwent min-max normalization and histogram equalization. Additionally, adaptive thresholding was applied to Ultivue images to reduce noise and extract true signal. Regions with false signals, particularly those corresponding to hemorrhage, bleeding, or erythrocytes in H&E, were manually annotated and excluded from analysis.

Upon acceptance, we plan to publicly release the H&E and Ultivue images, their alignment matrices, and annotated excluded regions. The dataset could serve as a valuable baseline for the field.

### 4.1.3 TCGA-SKCM

Diagnostic WSIs of SKCM were downloaded from the TCGA database[1] for a total of 472 cases. Clinical data of SKCM samples including age, gender, sample type (primary tumor/metastatic) and disease-specific survival were also downloaded. For the survival prediction, we discarded cases where the diagnostic WSIs are of low resolution or the disease-specific survival data is missing, leaving 360 cases in total. For the immune subtype prediction, we kept a total of 257 cases where immune subtype labels are available. For each task, we randomly split the cases stratified by age, gender and sample type to create 5-fold cross-validation with a 4:1 ratio of training-validation sets.

## 4.2 HistoPlexer architecture

The HistoPlexer is based on cGAN which takes an H&E image as input condition and generates multiplexed IMC images where each corresponds to a spatially-resolved protein expression profile. The *translator* of the HistoPlexer is a fully convolutional U-Net [45] which consists of an encoder and a decoder. The encoder comprises six downsampling blocks, each with a convolution layer of stride 2 and kernel size 3. The decoder comprises of five upsampling blocks, each with nearest neighbor interpolation, followed by convolution layer of stride 1 and kernel size 3. Each layer is followed by a batch-norm layer and ReLU activation. The *discriminator* consists of six blocks, each with a convolution layer followed by a spectral normalization layer and ReLU activation. We use patches extracted from template-matched pairs of H&E and IMC RoIs to train the HistoPlexer and optimize the model with three objectives: an adversarial loss to enforce image-level consistency, a Gaussian pyramid loss to enforce pixel-level consistency, and a patch-wise contrastive loss to enforce patch-level consistency.

**Adversarial loss:** We use the least square loss proposed in LSGAN [46] as our adversarial loss, and the $0-1$ coding scheme where 0 and 1 are the labels for generated (*i.e.*, fake) and real IMC images, respectively. We also adopt the multi-scale gradient approach [47], which allows simultaneous gradient propagation at multiple scales (*i.e.*, resolutions). Considering a set of scales $\{s \in S\}$, the multi-scale adversarial losses for the translator $G$ and discriminator $D$ are formulated as:

$$\begin{aligned}
\mathcal{L}_G^{\text{adv}} &= \frac{1}{|S|}\mathbb{E}_{\mathbf{x}_p \sim X_p}\left[\left(D(G^{(s)}(\mathbf{x}_p)|\mathbf{x}_p) - 1\right)^2\right], \\
\mathcal{L}_D^{\text{adv}} &= \frac{1}{|S|}\sum_{s\in S}\left[\mathbb{E}_{\substack{\mathbf{x}_p \sim X_p \\ \mathbf{y}_p \sim Y_p}}\left[(D(\mathbf{y}_p|\mathbf{x}_p) - 1)^2\right] + \mathbb{E}_{\mathbf{x}_p \sim X_p}\left[(D(G^{(s)}(\mathbf{x}_p)|\mathbf{x}_p))^2\right]\right].
\end{aligned} \tag{1}$$

where $X_p = \{\mathbf{x}_p \in X_{\text{RoI}}\}$ and $Y_p = \{\mathbf{y}_p \in Y_{\text{RoI}}\}$ denote paired training patches sampled from template-matched H&E and IMC RoIs, respectively; $G^{(s)}(\cdot)$ and $D(\cdot)$ denote the mapping functions parameterized by the translator (at the output scale $s$) and discriminator, respectively; and $|\cdot|$ denotes the cardinality of a set.

[1] https://portal.gdc.cancer.gov/

**Gaussian pyramid loss:** We also implement a pixel-level $L_1$ loss as in [21]. Since our H&E and GT IMC images are not pixel-aligned, we relax the constraint on pixel-to-pixel correspondence by calculating the $L_1$ loss at multi-resolution representations of the generated and GT IMC images [12], termed as Gaussian pyramid loss [12]. More specifically, a Gaussian pyramid is constructed through iterative Gaussian smoothing and downsampling. Each level of resolution, termed as an octave, comprises a series of images with increasing degrees of smoothness. Transition between resolutions is achieved by downsampling the image at the highest smoothness level of the current octave to initiate the next:

$$\mathbf{y}_{p,1}^{r+1} = \text{Downsample}\Big(\mathbf{y}_{p,\#\text{gs}}^{r}\Big)$$

where $\#gs$ denotes the number of Gaussian smoothing at one resolution. Note that for the generated IMC images, we only compute the Gaussian pyramid on the final output scale. Considering a set of resolutions $\{r \in R\}$, the Gaussian pyramid loss is a weighted sum of $L_1$ loss computed on the primary layer of each octave, formulated as:

$$\mathcal{L}^{\text{gp}} = \sum_{r \in R} w_r \mathbb{E}_{\substack{\mathbf{x}_p \sim X_p \\ \mathbf{y}_p \sim Y_p}} \left\| \mathbf{y}_{p,1}^{r} - \hat{\mathbf{y}}_{p,1}^{r} \right\|_1, \tag{2}$$

where $\hat{\mathbf{y}}_p$ denotes the generated IMC image patches, $r$ denotes the resolution level, and $w_r$ is the weight of the $L_1$ loss at that level.

**Patch-wise contrastive loss:** We further incorporate a patch-wise contrastive loss, inspired by [19]. More specifically, we first extract multi-layer features using a pretrained feature encoder and apply a transformation via a small projection head (*e.g.*, a Multi-layer Perceptron) on the extracted features to enrich their expressiveness [48]. Then, we randomly select a set of pixel locations for each feature layer. By aggregating selected patch features from each layer, we can obtain two feature sets for the generated and GT IMC images, respectively.

Let $\hat{z}_l^i$ denote the anchor feature of the $i$-th patch of the generated IMC image, extracted from the $l$-th layer of the feature encoder; while $z_l^i$ and $\bar{z}_l^i$ denote the positive and negative features of the corresponding patch (*i.e.*, at the same pixel location) and the collection of non-corresponding patches (*i.e.*, at different pixel locations), extracted from the same layer, respectively. Our patch-wise contrastive loss is defined as:

$$\mathcal{L}^{\text{contrast}} = \mathop{\mathbb{E}}_{\substack{\mathbf{x}_p \sim X_p \\ \mathbf{y}_p \sim Y_p}} \frac{1}{\#\text{layer}} \frac{1}{\#\text{patch}} \sum_{l=1}^{\#\text{layer}} \sum_{i=1}^{\#\text{patch}} w_t(\hat{z}_l^i, z_l^i) \ell_{\text{InfoNCE}}(\hat{z}_l^i, z_l^i, \bar{z}_l^i), \tag{3}$$

where

$$\ell_{\text{InfoNCE}}(z, z^+, z^-) = -\log \frac{\exp(z \cdot z^+/\tau)}{\exp\left(z \cdot z^+/\tau\right) + \sum_{n=1}^{N} \exp\left(z \cdot z_n^-\right)/\tau)}$$

is the InfoNCE objective [49], and

$$w_t(\hat{z}_l^i, z_l^i) = \left(1 - g\left(\frac{t}{T}\right)\right) \times 1.0 + g\left(\frac{t}{T}\right) \times h\Big(\text{sim}(\hat{z}_l^i, z_l^i)\Big)$$

is the adaptive patch weight [19]. Here, #layer and #patch denote the number of layers and patches from which we extract features; $t$ and $T$ denote the current and total training steps; $h(\cdot)$ denotes some weighting function; and $\text{sim}(\cdot)$ is some similarity measurement.

While the HistoPlexer translator outputs the prediction of all selected IMC markers, we encounter a practical limitation when employing a pre-trained feature encoder, which often requires an RGB image as input. To circumvent this, we first extract each channel (*i.e.*, marker) of the output IMC image and replicate it along the channel dimension to create a pseudo RGB image. We then pass each of them to the feature encoder. The final patch-wise contrastive loss is the sum of that of each channel.

The total losses for $G$ and $D$ are formulated as,

$$\begin{aligned} \mathcal{L}_G &= \mathcal{L}_G^{adv} + \lambda_{\text{gp}}\mathcal{L}^{\text{gp}} + \lambda_{\text{contrast}}\mathcal{L}^{\text{contrast}} \\ \mathcal{L}_D &= \mathcal{L}_D^{adv} + \lambda_{R_1} R_1 \end{aligned} \tag{4}$$

where

$$R_1 = \mathop{\mathbb{E}}_{\substack{\mathbf{x}_p \sim X_p \\ \mathbf{y}_p \sim Y_p}} \|\nabla_{\mathbf{y}} D(\mathbf{y}_p | \mathbf{x}_p)\|_2^2$$

is the gradient penalty [50], and $\lambda_{\text{gp}}$, contrast and $\lambda_{R_1}$ are the weights for the Gaussian pyramid loss, patch-wise contrastive loss and gradient penalty, respectively.

**Implementation and training details:** The model is trained for 100 epochs using ADAM optimizer [51] with momentum parameters $\beta 1 = 0.5$ and $\beta 2 = 0.999$ with learning rates 0.004 and 0.0008 for translator and discriminator networks, respectively. The weights are initialized using Xavier initialization. The batch size is set to 16 and the patch size to 256 for IMC and 1024 for H&E images, to accommodate for the higher resolution of the latter. We increase the generalization capabilities of the model by adopting data augmentation, including color augmentation, random flipping, small translations, and rotations. We employ the least-squares GAN objective. The weights for loss terms is as follows: $\lambda_{\text{gp}}$=5.0, $\lambda_{\text{contrast}}$=1.0 and $\lambda_{R_1}$=1.0.

## 4.3 Evaluation metrics

To evaluate the quality of generated images, we use two widely adopted metrics: PSNR and MS-SSIM.

PSNR is used to measure the reconstruction quality by quantifying the ratio between the maximum possible signal power and the power of corrupting noise. It is expressed in decibels (dB), with higher values indicating better image quality. The PSNR is calculated as:

$$\text{PSNR} = 10 \log_{10}\left(\frac{L^2}{\text{MSE}}\right) \tag{5}$$

where $L$ is the dynamic range of the pixel values (e.g., 255 for 8-bit images), and MSE represents the Mean Squared Error between the original image $I$ and the generated image $I'$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(I(i) - I'(i)\right)^2 \tag{6}$$

MS-SSIM extends the traditional SSIM metric by incorporating multiple scales to capture structural differences at various resolutions. The SSIM between two images $I$ and $I'$ is defined as:

$$\text{SSIM}(I, I') = \frac{(2\mu_I\mu_{I'} + C_1)(2\sigma_{II'} + C_2)}{(\mu_I^2 + \mu_{I'}^2 + C_1)(\sigma_I^2 + \sigma_{I'}^2 + C_2)} \tag{7}$$

where $\mu_I$ and $\mu_{I'}$ are the means, $\sigma_I^2$ and $\sigma_{I'}^2$ are the variances, and $\sigma_{II'}$ is the covariance between the two images. $C_1$ and $C_2$ are small constants to stabilize the division. In MS-SSIM, SSIM is computed at multiple scales, and the final score is a weighted product of SSIM values across these scales:

$$\text{MS-SSIM}(I, I') = \prod_{j=1}^{M}\left(\text{SSIM}_j(I, I')\right)^{\alpha_j} \tag{8}$$

where $M$ is the number of scales and $\alpha_j$ is weighting factor at scale $j$. Higher MS-SSIM values indicate better perceptual similarity.

These metrics provide a comprehensive assessment of both pixel-level accuracy (PSNR) and perceptual similarity (MS-SSIM) of the generated images. Frechet Inception Distance (FID) and Kernel Inception Distance (KID) are widely used metrics for evaluating the quality of generated images, however they are less effective on small datasets as they rely on mean and covariance of a cohort. Hence they are not used when evaluating HistoPlexer.

To quantify the evaluation by domain experts, we use HYPE score which measures the error rate at which humans mistake generated images for real ones or vice versa. It is defined as:

$$\begin{aligned}
\text{HYPE} &= \left(\frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}\right) \times 100 \\
\text{HYPE}_{\text{fake}} &= \left(\frac{\text{FP}}{\text{TN} + \text{FP}}\right) \times 100 \\
\text{HYPE}_{\text{real}} &= \left(\frac{\text{FN}}{\text{TP} + \text{FN}}\right) \times 100
\end{aligned} \tag{9}$$

where TP is the number of True Positives, TN is the number of True Negatives, FP is the number of False Positives and FN is the number of False Negatives. $\text{HYPE}_{fake}$ and $\text{HYPE}_{real}$ are the error rates for generated and real images, respectively.

## 4.4 HistoPlexer for cell-level analysis

### 4.4.1 Pseudo-cells

Since spatial analyses of IMC data typically rely on cell-level readouts, we create pseudo-single-cell data by extracting circular regions of 10 µm diameter around nuclei coordinates for both input H&E and GT IMC images. Protein expression is averaged across pixels within each pseudo-cell for individual markers. Nuclei coordinates for H&E images are obtained using the HoVer-Net model [24], while nuclei coordinates and cell-type labels for GT IMC multiplexes are derived using Ilastik [52] and CellProfiler [40], as described in [8]. For simplicity, we refer to pseudo-cells as "cells" in the following text.

### 4.4.2 Cell-typing

We use a Random Forest (RF) classifier [26] to categorize cells based on the average expression of 11 markers from the HistoPlexer. The classifier distinguishes between tumor cells, B-cells, CD8+ T-cells, CD4+ T-cells, and other cells. Training is performed using the scikit-learn library [53], with hyperparameters (100 base estimators, maximum tree depth of 30) selected based on the lowest out-of-bag error. The model achieves a macro-averaged F1 score of 0.81 on an internal test set. We then apply the trained RF classifier to both GT and generated protein expression data to produce cell type maps for cells in test set.

### 4.4.3 t-SNE on cell level marker expression

To explore spatial patterns beyond pairwise protein interactions, we conduct a low-dimensional embedding analysis of cell-level marker expression. Following the approach commonly used for mass cytometry data [54], we subsample 1,000 cells per RoI from both GT and generated IMC, resulting in total 2,000 cells per RoI. A joint t-SNE dimensionality reduction (two dimensions, perplexity of 50, and 1,000 iterations) is then applied. For visualization, protein abundance is scaled and clipped at the 99th percentile, and the t-SNE plots are colored according to the scaled protein expression [54].

## 4.5 Annotations for Immune phenotyping

To stratify samples into immune subtypes based on the spatial distribution of CD8+ T-cells, we used annotated regions as established in [5]. Our dataset included 109 metastatic melanoma H&E WSIs from the TuPro cohort, with metastatic sites in lymph nodes, soft tissue, brain, and other distant locations. The primary region for immune-subtyping, termed "Tumor Center", comprises entirely tumor tissue, which was manually defined as a continuous tumor mass excluding a 500$\mu$m margin from the tumor–non-tumor boundary. This "Tumor Center" was further segmented into two regions: the "Intratumoral Tumor" region, consisting of dense clusters of malignant melanocytes without stromal presence, and the "Intratumoral Stromal" region, which includes extracellular matrix (typically desmoplastic) interwoven within the

tumor cell mass but free from malignant melanocytes. These regions were automatically classified using a DL model implemented on the HALO$^{AI}$ platform, trained with selected H&E WSIs regions. Tissue classification was conducted at 0.30$\mu$m/pixel resolution with a minimum object size threshold of 50$\mu m^2$. Excluded regions—such as preexisting lymphatic tissue, large adipose and muscle regions, artifacts, necrosis, hemorrhage, and background—were omitted from the analysis. Ultimately, we analyzed 34 samples with the highest quality tissue classifications from the HALO$^{AI}$ model predictions. Supplementary Fig. S3 shows an example H&E WSI with region annotation and classification.

## 4.6 MIL-based Clinical Outcome Prediction

**Attention-based MIL for survival and immune subtype prediction:** MIL is a weakly-supervised learning method for set-based data structures. In MIL, an input $X$ is a bag (*i.e.*, permutation-invariant set) of instances $X = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$, where $N$ denotes the number of instances in the bag. Given a classification task with $K$ classes, the goal is to learn a function $\mathcal{F}$ from $M$ training pairs $\{(X^{(m)}, \mathbf{y}^{(m)})\}_{m=1}^{M}$ that maps $X$ to a bag-level label $\mathbf{y} \in K$ without knowing label $\mathbf{y}_i \in K$ for each instance in the bag. In our context, the input is a WSI and the instances denote the extracted patches. More specifically, we follow the embedding-based MIL approach [35] and extract a feature vector $\mathbf{h}_i = h(\mathbf{x}_i) \in \mathbb{R}^d$ from each patch. Then, an attention-pooling operator aggregates the patch features $\mathbf{h}_{i=1:N}$ to a single WSI-level representation [35]

$$\mathbf{g} = g(\mathbf{h}_i) = \sum_{i=1}^{N} a_i \mathbf{h}_i,$$

where

$$a_i = \frac{\exp\{\mathbf{w}^\top(\tanh(\mathbf{V}\mathbf{h}_i) \odot \eta(\mathbf{U}\mathbf{h}_i))\}}{\sum_{j=1}^{N} \exp\{\mathbf{w}^\top(\tanh(\mathbf{V}\mathbf{h}_j) \odot \eta(\mathbf{U}\mathbf{h}_j))\}}$$

is the gated attention [35]. Here, $\mathbf{w} \in \mathbb{R}^{L\times 1}$, $\mathbf{V} \in \mathbb{R}^{L\times D}$, $\mathbf{U} \in \mathbb{R}^{L\times D}$ are learnable parameters with hidden dimension $L$, $\odot$ is element-wise multiplication, and $\eta(\cdot)$ denotes the Sigmoid function. Finally, a classifier $f(\cdot)$ maps the WSI-level representation to a WSI-level label $\hat{\mathbf{y}} \in K$.

The end-to-end prediction takes the following general form:

$$\hat{\mathbf{y}} = \mathcal{F}(X) = f\Big(g\Big(\{h(\mathbf{x}_i) : \mathbf{x}_i \in X\}\Big)\Big) . \tag{10}$$

For survival prediction, we model the time-to-event distributions as an ordinal regression task with right censored data (*i.e.*, patient death is unobserved until last known follow-up). Following [36], we define discrete time intervals and model each interval using an independent neuron in the output layer. More specifically, we partition the continuous time scale into non-overlapping time intervals $[t_{j-1}, t_j), j \in [1, \cdots, J]$ based on the quartiles of survival time values, denoted as $\mathbf{y}_j$. The continuous time-to-event $t^{(m)}$ for each patient is then replaced by a discrete time label $\mathbf{y}_j^{(m)}$,

where

$$\mathbf{y}_j^{(m)} = \mathbf{y}_j \quad \text{if } t^{(m)} \in [t_{j-1}, t_j) \text{ for } j \in \{0, \cdots, J\}.$$

The problem then simplifies to classification where each patient is defined by a triplet $(\mathbf{g}^{(m)}, \mathbf{y}_j^{(m)}, c^{(m)})$. Here, $\mathbf{g}$ is the aggregated bag features; $c$ is the censorship status where $c = 0$ if the death of the patient is observed and $c = 1$ otherwise; and $\mathbf{y}_j$ is the discrete time GT label. We adopt the negative log-likelihood survival loss [55] for modal optimization, formulated as:

$$\begin{aligned}\mathcal{L}_{\text{surv}}\Big(\{X^{(m)}, \mathbf{y}_j^{(m)}, c^{(m)}\}_{m=1}^{M}\Big) = \\ \sum_{i=1}^{M}\Bigg( &- c^{(m)} \log(f_{\text{surv}}(\mathbf{y}_j^{(m)}|\mathbf{g}^{(m)})) \\ &+ (1 - c^{(m)}) \log(f_{\text{surv}}(\mathbf{y}_j^{(m)} - 1|\mathbf{g}^{(m)})) \\ &+ (1 - c^{(m)}) \log(f_{\text{hazard}}(\mathbf{y}_j^{(m)}|\mathbf{g}^{(m)}))\Bigg),\end{aligned} \tag{11}$$

where $f_{\text{harzard}}(\mathbf{y}_j|\mathbf{g}) = \text{Sigmoid}(\hat{\mathbf{y}}_j)$ is the discrete hazard function and $f_{\text{surv}}(\mathbf{y}_j|\mathbf{g}) = \prod_{k=1}^{j}\left(1 - f_{\text{hazard}}(\mathbf{y}_k|\mathbf{g})\right)$ is the discrete survival function. Finally, the patient-level risk is defined as the negative sum of all logits [37], which enables the identification of distinct risk groups and the stratification of patients.

For immune subtype prediction, we adopt the cross-entropy loss defined as:

$$\mathcal{L}_{\text{ce}} = -\sum_{m=1}^{M}\sum_{k=1}^{K}\mathbf{y}_k^{(m)} \log\left(\hat{\mathbf{y}}_k^{(m)}\right). \tag{12}$$

**Multimodal fusion via co-attention mechanism:** To fuse the patch features from different modalities, we adopt the co-attention mechanism proposed in [36]. More specifically, given the H&E feature bag $\mathbf{H} \in \mathbb{R}^{N\times d}$ and IMC feature bag $\mathbf{P} \in \mathbb{R}^{N\times d}$, we guide the feature aggregation of $\mathbf{H}$ using $\mathbf{P}$ by calculating the cross-attention:

$$\begin{aligned}\hat{\mathbf{H}} &= \text{Softmax}\Big(\frac{\mathbf{W}_q\mathbf{P}\mathbf{H}^\top\mathbf{W}_k^\top}{\sqrt{d}}\Big)\mathbf{W}_v\mathbf{H} \\ &= \mathbf{A}_{P\to H}\mathbf{W}_v\mathbf{H},\end{aligned} \tag{13}$$

where $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v \in \mathbb{R}^{d\times d}$ are learnable weights and $\mathbf{A}_{P\to H} \in \mathbb{R}^{N\times N}$ is the co-attention matrix. Intuitively, the co-attention measures the pairwise similarity for how much an H&E instance $\mathbf{h}_i$ attend to the IMC instance $\mathbf{p}_i$ for $i \in N$. Similarly, we can guide the feature aggregation of $\mathbf{P}$ using $\mathbf{H}$ via $\mathbf{A}_{H\to P}$. Each co-attention guided feature bag is input to an attention-based MIL module, which outputs an aggregated WSI-level representation. We concatenate the WSI-level representations from multiple modalities and project it back to the original feature dimension $d$ via a linear layer, resulting in a multimodal WSI-level representation. Then, a classifier $f(\cdot)$ uses this representation to predict the output label $\hat{\mathbf{y}}$.

**Implementation and training details:** We adopt the original implementation of attention-based MIL on GitHub[2] and modify it for survival prediction based on the code for SurvPath[3]. We implement the co-attention mechanism based on the original implementation of MCAT[4]. Each WSI is cropped to $256 \times 256$ non-overlapping patches at $20\times$ magnification to create bags, where patches with more than 10% non-tissue area are discarded. We use ResNet18 [56] pretrained on pathology-specific datasets using self-supervised learning [57] to extract features from H&E patches and ResNet50 pretrained on ImageNet [58] to extract features from IMC patches. Since ResNet18 requires three-channel input, we concatenate IMC images of three different protein markers along the channel dimension: one tumor marker (MelanA) and two immune markers (CD8 and CD20). The dimension of extracted features is 512 for both H&E and IMC patches. We run the survival and immune subtype prediction for 5-fold cross-validation. The model hyperparameters are set as: Adam optimizer with initial learning rate of $1e^{-4}$ (survival) and $5e^{-5}$ (immune subtype), a ReduceLROnPlateau scheme based on validation loss for scheduling, and a mini-batch size of 1. The model is trained for 100 epochs with early stopping based on validation loss (survival) and weighted F1-score (immune subtype).

**Computational requirements.** The data processing and model training was done on NVIDIA A100 40GB GPU. The DL models were trained using pytorch (1.13.1). The pipeline was implemented in Python (3.8.12).

**Data Availability.** Data and material from the Tumor Profiler study are available to members of the international Tumor Profiler Research Consortium. Requests for sharing of all data and material should be addressed to the corresponding author and include a scientific proposal. Depending on the specific research proposal, the Tumor-Profiler consortium will determine when, for how long, for which specific purposes, and under which conditions the requested data can be made available, subject to ethical consent. The multiplexed WSIs images for Immuno8 and MDSC FixVue™ panels from Ultivue InSituPlex® technology, along with paired H&E images will be made available upon acceptance of publication. The H&E WSIs for TCGA-SKCM were downloaded via GDC data portal (https://portal.gdc.cancer.gov/).

**Code Availability.** The source code for HistoPlexer is available at https://github.com/ratschlab/HistoPlexer.

**Acknowledgements.** We gratefully acknowledge funding from the Tumor Profiler Initiative and the Tumor Profiler Center (to V.H.K., G.R., B.B.). The Tumor Profiler study is jointly funded by a public-private partnership involving F. Hoffmann-La Roche Ltd., ETH Zurich, University of Zurich, University Hospital Zurich, and University Hospital Basel. We also acknowledge funding from the Swiss Federal Institutes of Technology strategic focus area of personalized health and related technologies project 2021-367 (to G.R., V.H.K., S.A.), the ETH AI Center (to G.R. and B.C.), ETH core funding (to G.R.), UZH core funding (to V.H.K) and funding by the Promedica Foundation grant F-87701-41-01 (to V.H.K). B.B. was funded by two SNSF project

[2] https://github.com/AMLab-Amsterdam/AttentionDeepMIL
[3] https://github.com/mahmoodlab/SurvPath
[4] https://github.com/mahmoodlab/MCAT

grants (#310030_205007: Analysis of breast tumor ecosystem properties for precision medicine approaches and #316030_213512: Cellular-resolution high-performance mass spectrometric imaging of biological samples), an NIH grant (UC4 DK108132), the CRUK IMAXT Grand Challenge, and the European Research Council (ERC) under the European Union's Horizon 2020 Program under the ERC grant agreement no. 866074 ("Precision Motifs").

We thank Ilario Scapozza, University of Zurich, Switzerland, for supporting the visual assessment evaluation. We appreciate the contribution of Flavia Pedrocchi towards image registration of pathology images. We express our gratitude towards Lucy Godson, University of Leeds, UK, and Jeremie Nsengimana, Newcastle University, UK, for sharing the immune subtypes for TCGA-SKCM data.

**Author Contributions.** G.R., V.H.K, S.A. and J.F.P. conceived the study. G.R., V.H.K, S.A., B.C. and J.F.P. participated in designing the study and reviewing the literature. V.H.K and G.R. supervised the study. V.H.K., S.A., J.F.P., B.S. and R.C. collected the data. S.A., J.F.P., B.C. and S.H. contributed to code writing, model training, and data analysis. V.H.K. and G.R. contributed to expert review and data interpretation. B.C., S.A., V.H.K., and G.R. designed the figures. S.A. and B.C. drafted the manuscript. T.P.C., V.H.K, R.C., B.B. and B.S. contributed to providing data and tumor material used in this study. B.H. contributed towards annotation tasks and qualitative assessment of data. All authors had access to the result data presented in the final manuscript and all authors read and approved the final manuscript. The decision to submit was made by all authors. V.H.K. and G.R. contributed equally to this work and share senior authorship.

**Competing Interests.** V.H.K. reports being an invited speaker for Sharing Progress in Cancer Care (SPCC) and Indica Labs; advisory board of Takeda; sponsored research agreements with Roche and IAG, all unrelated to the current study. V.H.K. is a participant of a patent application on the assessment of cancer immunotherapy biomarkers by digital pathology; a patent application on multimodal deep learning for the prediction of recurrence risk in cancer patients, and a patent application on predicting the efficacy of cancer treatment using deep learning. G.R. and J.F.P. are participants of a patent application on matching cells from different measurement modalities which is not directly related to the current work. Moreover, G.R. is cofounder of Computomics GmbH, Germany, and one of its shareholders. B.B. has co-founded Navignostics, a spin-off company of the University of Zurich developing precision oncology diagnostics, and is one of its shareholders and a board member.

**Approval from ethics committee.** The ethics committee of the "Swiss Association of Research Ethics Committees" gave ethical approval for the data from the Tumor Profiler Study used in this work. The Tumor Profiler Study is an approved, observational clinical study (BASEC: 2018-02050, 2018-02052, 2019-01326, 2024-01428).

**TUMOR PROFILER CONSORTIUM.** Rudolf Aebersold[5], Melike Ak[33], Faisal S Al-Quaddoomi[12,22], Silvana I Albert[10], Jonas Albinus[10], Ilaria Alborelli[29], Sonali Andani[9,22,31,36], Per-Olof Attinger[14], Marina Bacac[21], Daniel Baumhoer[29], Beatrice Beck-Schimmer[44], Niko Beerenwinkel[7,22], Christian Beisel[7], Lara Bernasconi[32], Anne Bertolini[12,22], Bernd Bodenmiller[11,40], Ximena Bonilla[9], Lars Bosshard[12,22], Byron Calgua[29], Ruben Casanova[40], Stéphane Chevrier[40], Natalia Chicherova[12,22], Ricardo Coelho[23], Maya D'Costa[13], Esther Danenberg[42], Natalie R Davidson[9], Monica-Andreea Dragan[7], Reinhard Dummer[33], Stefanie Engler[40], Martin Erkens[19], Katja Eschbach[7], Cinzia Esposito[42], André Fedier[23], Pedro F Ferreira[7], Joanna Ficek-Pascual[1,9,16,22,31], Anja L Frei[36], Bruno Frey[18], Sandra Goetze[10], Linda Grob[12,22], Gabriele Gut[42], Detlef Günther[8], Pirmin Haeuptle[3], Viola Heinzelmann-Schwarz[23,28], Sylvia Herter[21], Rene Holtackers[42], Tamara Huesser[21], Alexander Immer[9,17], Anja Irmisch[33], Francis Jacob[23], Andrea Jacobs[40], Tim M Jaeger[14], Katharina Jahn[7], Alva R James[9,22,31], Philip M Jermann[29], André Kahles[9,22,31], Abdullah Kahraman[22,36], Viktor H Koelzer[36,41], Werner Kuebler[30], Jack Kuipers[7,22], Christian P Kunze[27], Christian Kurzeder[26], Kjong-Van Lehmann[2,4,9,15], Mitchell Levesque[33], Ulrike Lischetti[23], Flavio C Lombardo[23], Sebastian Lugert[13], Gerd Maass[18], Markus G Manz[35], Philipp Markolin[9], Martin Mehnert[10], Julien Mena[5], Julian M Metzler[34], Nicola Miglino[35,41], Emanuela S Milani[10], Holger Moch[36], Simone Muenst[29], Riccardo Murri[43], Charlotte KY Ng[29,39], Stefan Nicolet[29], Marta Nowak[36], Monica Nunez Lopez[23], Patrick GA Pedrioli[6], Lucas Pelkmans[42], Salvatore Piscuoglio[23,29], Michael Prummer[12,22], Prélot, Laurie[9,22,31], Natalie Rimmer[23], Mathilde Ritter[23], Christian Rommel[19], María L Rosano-González[12,22], Gunnar Rätsch[1,6,9,22,31], Natascha Santacroce[7], Jacobo Sarabia del Castillo[42], Ramona Schlenker[20], Petra C Schwalie[19], Severin Schwan[14], Tobias Schär[7], Gabriela Senti[32], Wenguang Shao[10], Franziska Singer[12,22], Sujana Sivapatham[40], Berend Snijder[5,22], Bettina Sobottka[36], Vipin T Sreedharan[12,22], Stefan Stark[9,22,31], Daniel J Stekhoven[12,22], Tanmay Tanna[7,9], Alexandre PA Theocharides[35], Tinu M Thomas[9,22,31], Markus Tolnay[29], Vinko Tosevski[21], Nora C Toussaint[12,22], Mustafa A Tuncel[7,22], Marina Tusup[33], Audrey Van Drogen[10], Marcus Vetter[25], Tatjana Vlajnic[29], Sandra Weber[32], Walter P Weber[24], Rebekka Wegmann[5], Michael Weller[38], Fabian Wendt[10], Norbert Wey[36], Andreas Wicki[35,41], Mattheus HE Wildschut[5,35], Bernd Wollscheid[10], Shuqing Yu[12,22], Johanna Ziegler[33], Marc Zimmermann[9], Martin Zoche[36], Gregor Zuend[37]

[1]AI Center at ETH Zurich, Andreasstrasse 5, 8092 Zurich, Switzerland, [2]Cancer Research Center Cologne-Essen, University Hospital Cologne, Cologne, Germany, [3]Cantonal Hospital Baselland, Medical University Clinic, Rheinstrasse 26, 4410 Liestal, Switzerland, [4]Center for Integrated Oncology Aachen (CIO-A), Aachen, Germany, [5]ETH Zurich, Department of Biology, Institute of Molecular Systems Biology, Otto-Stern-Weg 3, 8093 Zurich, Switzerland, [6]ETH Zurich, Department of Biology, Wolfgang-Pauli-Strasse 27, 8093 Zurich, Switzerland, [7]ETH Zurich, Department of Biosystems Science and Engineering, Mattenstrasse 26, 4058 Basel, Switzerland, [8]ETH Zurich, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 1-5/10, 8093 Zurich, Switzerland, [9]ETH Zurich, Department of Computer Science, Institute of Machine Learning, Universitätstrasse 6, 8092 Zurich,

Switzerland, [10]ETH Zurich, Department of Health Sciences and Technology, Otto-Stern-Weg 3, 8093 Zurich, Switzerland, [11]ETH Zurich, Institute of Molecular Health Sciences, Otto-Stern-Weg 7, 8093 Zurich, Switzerland, [12]ETH Zurich, NEXUS Personalized Health Technologies, Wagistrasse 18, 8952 Zurich, Switzerland, [13]F. Hoffmann-La Roche Ltd, Grenzacherstrasse 124, 4070 Basel, Switzerland, [14]F. Hoffmann-La Roche Ltd, Grenzacherstrasse 124, 4070 Basel, Switzerland, , [15]Joint Research Center Computational Biomedicine, University Hospital RWTH Aachen, Aachen, Germany, [16]Life Science Zurich Graduate School, Biomedicine PhD Program, Winterthurerstrasse 190, 8057 Zurich, Switzerland, [17]Max Planck ETH Center for Learning Systems, , [18]Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany, [19]Roche Pharmaceutical Research and Early Development, Roche Innovation Center Basel, Grenzacherstrasse 124, 4070 Basel, Switzerland, [20]Roche Pharmaceutical Research and Early Development, Roche Innovation Center Munich, Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany, [21]Roche Pharmaceutical Research and Early Development, Roche Innovation Center Zurich, Wagistrasse 10, 8952 Schlieren, Switzerland, [22]SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland, [23]University Hospital Basel and University of Basel, Department of Biomedicine, Hebelstrasse 20, 4031 Basel, Switzerland, [24]University Hospital Basel and University of Basel, Department of Surgery, Brustzentrum, Spitalstrasse 21, 4031 Basel, Switzerland, [25]University Hospital Basel, Brustzentrum & Tumorzentrum, Petersgraben 4, 4031 Basel, Switzerland, [26]University Hospital Basel, Brustzentrum, Spitalstrasse 21, 4031 Basel, Switzerland, [27]University Hospital Basel, Department of Information- and Communication Technology, Spitalstrasse 26, 4031 Basel, Switzerland, [28]University Hospital Basel, Gynecological Cancer Center, Spitalstrasse 21, 4031 Basel, Switzerland, [29]University Hospital Basel, Institute of Medical Genetics and Pathology, Schönbeinstrasse 40, 4031 Basel, Switzerland, [30]University Hospital Basel, Spitalstrasse 21/Petersgraben 4, 4031 Basel, Switzerland, [31]University Hospital Zurich, Biomedical Informatics, Schmelzbergstrasse 26, 8006 Zurich, Switzerland, [32]University Hospital Zurich, Clinical Trials Center, Rämistrasse 100, 8091 Zurich, Switzerland, [33]University Hospital Zurich, Department of Dermatology, Gloriastrasse 31, 8091 Zurich, Switzerland, [34]University Hospital Zurich, Department of Gynecology, Frauenklinikstrasse 10, 8091 Zurich, Switzerland, [35]University Hospital Zurich, Department of Medical Oncology and Hematology, Rämistrasse 100, 8091 Zurich, Switzerland, [36]University Hospital Zurich, Department of Pathology and Molecular Pathology, Schmelzbergstrasse 12, 8091 Zurich, Switzerland, [37]University Hospital Zurich, Rämistrasse 100, 8091 Zurich, Switzerland, [38]University Hospital and University of Zurich, Department of Neurology, Frauenklinikstrasse 26, 8091 Zurich, Switzerland, [39]University of Bern, Department of BioMedical Research, Murtenstrasse 35, 3008 Bern, Switzerland, [40]University of Zurich, Department of Quantitative Biomedicine, Winterthurerstrasse 190, 8057 Zurich, Switzerland, [41]University of Zurich, Faculty of Medicine, Zurich, Switzerland, [42]University of Zurich, Institute of Molecular Life Sciences, Winterthurerstrasse 190, 8057 Zurich, Switzerland, [43]University of Zurich, Services and Support for Science IT, Winterthurerstrasse 190, 8057 Zurich, Switzerland, [44]University of Zurich, VP Medicine, Künstlergasse 15, 8001 Zurich, Switzerland

# References

[1] Hanahan, D., Weinberg, R.A.: Hallmarks of cancer: the next generation. cell

**144**(5), 646–674 (2011)

[2] Hanahan, D.: Hallmarks of cancer: new dimensions. Cancer discovery **12**(1), 31–46 (2022)

[3] Egeblad, M., Nakasone, E.S., Werb, Z.: Tumors as organs: complex tissues that interface with the entire organism. Developmental cell **18**(6), 884–901 (2010)

[4] Jackson, H.W., Fischer, J.R., Zanotelli, V.R.T., Ali, H.R., Mechera, R., Soysal, S.D., Moch, H., Muenst, S., Varga, Z., Weber, W.P., Bodenmiller, B.: The single-cell pathology landscape of breast cancer. Nature **578**(7796), 615–620 (2020) https://doi.org/10.1038/s41586-019-1876-x

[5] Sobottka, B., Nowak, M., Frei, A.L., Haberecker, M., Merki, S., Levesque, M.P., Dummer, R., Moch, H., Koelzer, V.H.: Establishing standardized immune phenotyping of metastatic melanoma by digital pathology. Laboratory investigation **101**(12), 1561–1570 (2021)

[6] Ptacek, J., Locke, D., Finck, R., Cvijic, M.-E., Li, Z., Tarolli, J.G., Aksoy, M., Sigal, Y., Zhang, Y., Newgren, M., Finn, J.: Multiplexed ion beam imaging (mibi) for characterization of the tumor microenvironment across tumor types. Laboratory Investigation **100**(8), 1111–1123 (2020)

[7] Tan, W.C.C., Nerurkar, S.N., Cai, H.Y., Ng, H.H.M., Wu, D., Wee, Y.T.F., Lim, J.C.T., Yeong, J., Lim, T.K.H.: Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. Cancer Communications **40**(4), 135–153 (2020)

[8] Windhager, J., Zanotelli, V.R.T., Schulz, D., Meyer, L., Daniel, M., Bodenmiller, B., Eling, N.: An end-to-end workflow for multiplexed image processing and analysis. Nature Protocols **18**(11), 3565–3613 (2023)

[9] Jin, M.-Z., Jin, W.-L.: The updated landscape of tumor microenvironment and drug repurposing. Signal Transduction and Targeted Therapy **5**(1), 166 (2020) https://doi.org/10.1038/s41392-020-00280-x

[10] Fischer, A.H., Jacobson, K.A., Rose, J., Zeller, R.: Hematoxylin and eosin staining of tissue and cell sections. Cold spring harbor protocols **2008**(5), 4986 (2008)

[11] Burlingame, E.A., McDonnell, M., Schau, G.F., Thibault, G., Lanciault, C., Morgan, T., Johnson, B.E., Corless, C., Gray, J.W., Chang, Y.H.: Shift: speedy histological-to-immunofluorescent translation of a tumor signature enabled by deep learning. Scientific Reports **10**(1), 17507 (2020) https://doi.org/10.1038/s41598-020-74500-3

[12] Liu, S., Zhu, C., Xu, F., Jia, X., Shi, Z., Jin, M.: Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix. In: Proceedings of

the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1815–1824 (2022)

[13] Liu, S., Zhang, B., Liu, Y., Han, A., Shi, H., Guan, T., He, Y.: Unpaired stain transfer using pathology-consistent constrained generative adversarial networks. IEEE transactions on medical imaging **40**(8), 1977–1989 (2021)

[14] Pati, P., Karkampouna, S., Bonollo, F., Compérat, E., Radić, M., Spahn, M., Martinelli, A., Wartenberg, M., Kruithof-de Julio, M., Rapsomaniki, M.: Accelerating histopathology workflows with generative ai-based virtually multiplexed tumour profiling. Nature Machine Intelligence, 1–17 (2024)

[15] Zhang, R., Cao, Y., Li, Y., Liu, Z., Wang, J., He, J., Zhang, C., Sui, X., Zhang, P., Cui, L., *et al.*: Mvfstain: multiple virtual functional stain histopathology images generation based on specific domain mapping. Medical Image Analysis **80**, 102520 (2022)

[16] Zhou, Z., Jiang, Y., Sun, Z., Zhang, T., Feng, W., Li, G., Li, R., Xing, L.: Virtual multiplexed immunofluorescence staining from non-antibody-stained fluorescence imaging for gastric cancer prognosis. Ebiomedicine **107** (2024)

[17] Irmisch, A., Bonilla, X., Chevrier, S., Lehmann, K.-V., Singer, F., Toussaint, N.C., Esposito, C., Mena, J., Milani, E.S., Casanova, R., *et al.*: The tumor profiler study: Integrated, multi-omic, functional tumor profiling for clinical decision support. Cancer cell **39**(3), 288–293 (2021)

[18] Guan, J., Gupta, R., Filipp, F.V.: Cancer systems biology of tcga skcm: efficient detection of genomic drivers in melanoma. Scientific reports **5**(1), 7857 (2015)

[19] Li, F., Hu, Z., Chen, W., Kak, A.: Adaptive supervised patchnce loss for learning h&e-to-ihc stain translation with inconsistent groundtruth image pairs. arXiv preprint arXiv:2303.06193 (2023)

[20] Culjak, I., Abram, D., Pribanic, T., Dzapo, H., Cifrek, M.: A brief introduction to opencv. In: 2012 Proceedings of the 35th International Convention MIPRO, pp. 1725–1730 (2012). IEEE

[21] Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)

[22] Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, vol. 2, pp. 1398–1402 (2003). Ieee

[23] Jain, A.K.: Fundamentals of digital image processing. Prentice-Hall google schola **2**, 1375–1382 (1989)

[24] Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N.: Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical image analysis **58**, 101563 (2019)

[25] Zhou, S., Gordon, M., Krishna, R., Narcomey, A., Fei-Fei, L.F., Bernstein, M.: Hype: A benchmark for human eye perceptual evaluation of generative models. Advances in neural information processing systems **32** (2019)

[26] Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)

[27] Mondello, P., Fama, A., Larson, M.C., Feldman, A.L., Villasboas, J.C., Yang, Z.-Z., Galkin, I., Svelolkin, V., Postovalova, E., Bagaev, A., *et al.*: Lack of intrafollicular memory cd4+ t cells is predictive of early clinical failure in newly diagnosed follicular lymphoma. Blood cancer journal **11**(7), 130 (2021)

[28] Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., Samaras, D., Shroyer, K.R., Zhao, T., Batiste, R., *et al.*: Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. Cell reports **23**(1), 181–193 (2018)

[29] Chevrier, S., Levine, J.H., Zanotelli, V.R.T., Silina, K., Schulz, D., Bacac, M., Ries, C.H., Ailles, L., Jewett, M.A.S., Moch, H., Broek, M., Beisel, C., Stadler, M.B., Gedye, C., Reis, B., Pe'er, D., Bodenmiller, B.: An immune atlas of clear cell renal cell carcinoma. Cell **169**(4), 736–74918 (2017) https://doi.org/10.1016/j.cell.2017.04.016

[30] Herbst, R.S., Soria, J.-C., Kowanetz, M., Fine, G.D., Hamid, O., Gordon, M.S., Sosman, J.A., McDermott, D.F., Powderly, J.D., Gettinger, S.N., *et al.*: Predictive correlates of response to the anti-pd-l1 antibody mpdl3280a in cancer patients. Nature **515**(7528), 563–567 (2014)

[31] Ji, R.-R., Chasalow, S.D., Wang, L., Hamid, O., Schmidt, H., Cogswell, J., Alaparthy, S., Berman, D., Jure-Kunkel, M., Siemers, N.O., *et al.*: An immune-active tumor microenvironment favors clinical response to ipilimumab. Cancer Immunology, Immunotherapy **61**, 1019–1031 (2012)

[32] Godson, L., Alemi, N., Nsengimana, J., Cook, G.P., Clarke, E.L., Treanor, D., Bishop, D.T., Newton-Bishop, J., Gooya, A., Magee, D.: Immune subtyping of melanoma whole slide images using multiple instance learning. Medical Image Analysis **93**, 103097 (2024)

[33] Pfannstiel, C., Strissel, P.L., Chiappinelli, K.B., Sikic, D., Wach, S., Wirtz, R.M., Wullweber, A., Taubert, H., Breyer, J., Otto, W., *et al.*: The tumor immune microenvironment drives a prognostic relevance that correlates with bladder cancer subtypes. Cancer immunology research **7**(6), 923–938 (2019)

[34] Wouters, M.C., Nelson, B.H.: Prognostic significance of tumor-infiltrating b cells

and plasma cells in human cancer. Clinical Cancer Research **24**(24), 6125–6135 (2018)

[35] Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning, pp. 2127–2136 (2018). PMLR

[36] Chen, R.J., Lu, M.Y., Weng, W.-H., Chen, T.Y., Williamson, D.F., Manz, T., Shady, M., Mahmood, F.: Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4025 (2021)

[37] Jaume, G., Vaidya, A., Chen, R.J., Williamson, D.F., Liang, P.P., Mahmood, F.: Modeling dense multimodal interactions between biological pathways and histology for survival prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11579–11590 (2024)

[38] Antolini, L., Boracchi, P., Biganzoli, E.: A time-dependent discrimination index for survival data. Statistics in medicine **24**(24), 3927–3944 (2005)

[39] Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools (2000) https://doi.org/10.1038/s41374-020-0417-4

[40] McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., *et al.*: Cellprofiler 3.0: Next-generation image processing for biology. PLoS biology **16**(7), 2005970 (2018)

[41] Crowell, H.L., Chevrier, S., Jacobs, A., Sivapatham, S., Bodenmiller, B., Robinson, M.D., Consortium, T.P., *et al.*: An r-based reproducible and user-friendly preprocessing pipeline for cytof data. F1000Research **9**(1263), 1263 (2020)

[42] Otsu, N.: A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics **9**(1), 62–66 (1979)

[43] Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E.: A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: from Nano to Macro, pp. 1107–1110 (2009). IEEE

[44] Nan, A., Tennant, M., Rubin, U., Ray, N.: Drmime: Differentiable mutual information and matrix exponential for multi-resolution image registration. In: Medical Imaging with Deep Learning, pp. 527–543 (2020). PMLR

[45] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234–241 (2015). Springer

[46] Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2794–2802 (2017)

[47] Karnewar, A., Wang, O.: Msg-gan: Multi-scale gradients for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7799–7808 (2020)

[48] Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607 (2020). PMLR

[49] Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)

[50] Mescheder, L., Geiger, A., Nowozin, S.: Which training methods for gans do actually converge? In: International Conference on Machine Learning, pp. 3481–3490 (2018). PMLR

[51] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

[52] Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., *et al.*: Ilastik: interactive machine learning for (bio) image analysis. Nature methods **16**(12), 1226–1232 (2019)

[53] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., *et al.*: Scikit-learn: Machine learning in python. the Journal of machine Learning research **12**, 2825–2830 (2011)

[54] Wagner, J., Rapsomaniki, M.A., Chevrier, S., Anzeneder, T., Langwieder, C., Dykgers, A., Rees, M., Ramaswamy, A., Muenst, S., Soysal, S.D., *et al.*: A single-cell atlas of the tumor and immune ecosystem of human breast cancer. Cell **177**(5), 1330–1345 (2019)

[55] Zadeh, S.G., Schmid, M.: Bias in cross-entropy-based training of deep survival networks. IEEE transactions on pattern analysis and machine intelligence **43**(9), 3126–3137 (2020)

[56] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

[57] Ciga, O., Xu, T., Martel, A.L.: Self supervised contrastive learning for digital histopathology. Machine Learning with Applications **7**, 100198 (2022)