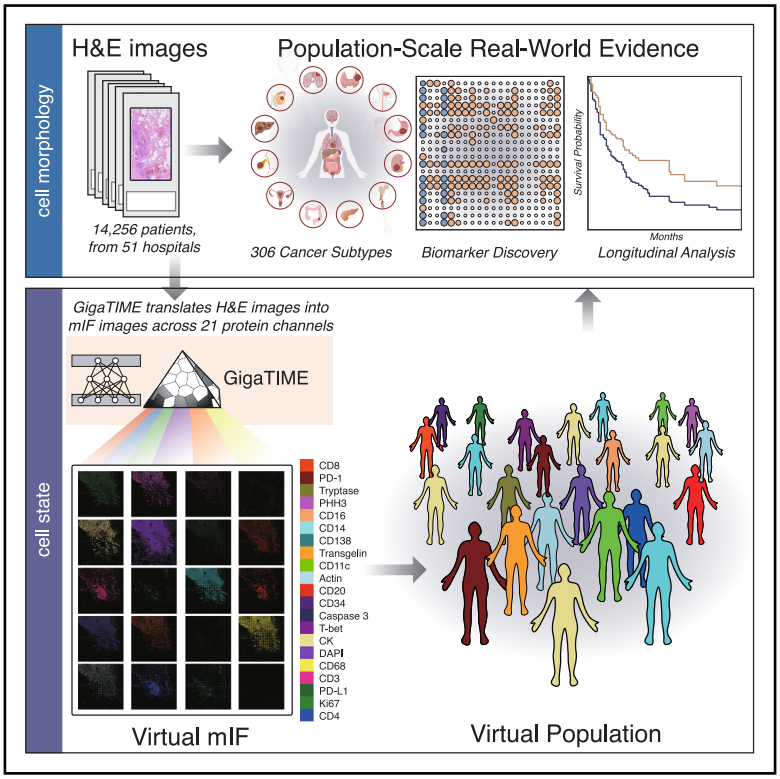
Cell

Multimodal AI generates virtual population for tumor microenvironment modeling

Graphical abstract



Authors

Jeya Maria Jose Valanarasu, Hanwen Xu, Naoto Usuyama, ..., Carlo Bifulco, Sheng Wang, Hoifung Poon

Correspondence

carlo.bifulco@providence.org (C.B.), swang@cs.washington.edu (S.W.), hoifung@microsoft.com (H.P.)

In brief

GigaTIME leverages multimodal AI to generate virtual multiplex immunofluorescence (mIF) profiles from standard H&E slides, enabling comprehensive tumor immune microenvironment modeling across a large (>14,000) and diverse patient population. This virtual approach unlocks new opportunities for large-scale clinical discoveries that were previously hindered by the scarcity of mIF data.

Highlights

- *GigaTIME* uses multimodal AI to translate H&E pathology slides to spatial proteomics
- *GigaTIME* generates a virtual population with cell states from routine H&E slides
- Virtual population enables large-scale clinical discovery and patient stratification
- Virtual population reveals new spatial and combinatorial protein activation patterns

Valanarasu et al., 2026, Cell *189*, 1–15
January 22, 2026
https://doi.org/10.1016/j.cell.2025.11.016

CellPress

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

Cell

CellPress
OPEN ACCESS

Article

Multimodal AI generates virtual population for tumor microenvironment modeling

Jeya Maria Jose Valanarasu,[1,7] Hanwen Xu,[1,2,7] Naoto Usuyama,[1,7] Chanwoo Kim,[1,2,7] Cliff Wong,[1] Peniel Argaw,[1] Racheli Ben Shimol,[4] Angela Crabtree,[4] Kevin Matlock,[4] Alexandra Q. Bartlett,[4] Jaspreet Bagga,[1] Yu Gu,[1] Sheng Zhang,[1] Tristan Naumann,[1] Bernard A. Fox,[4] Bill Wright,[5] Ari Robicsek,[5,6] Brian Piening,[3,4] Carlo Bifulco,[3,4,8,*] Sheng Wang,[1,2,8,*] and Hoifung Poon[1,8,9,*]

[1]Microsoft Research, Redmond, WA, USA
[2]School of Computer Science and Engineering, University of Washington, Seattle, WA, USA
[3]Providence Genomics, Portland, OR, USA
[4]Earle A. Chiles Research Institute, Providence Cancer Institute, Portland, OR, USA
[5]Providence Research Network, Renton, WA, USA
[6]Arclight Intelligence, Newton, MA, USA
[7]These authors contributed equally
[8]Senior author
[9]Lead contact
*Correspondence: carlo.bifulco@providence.org (C.B.), swang@cs.washington.edu (S.W.), hoifung@microsoft.com (H.P.)
https://doi.org/10.1016/j.cell.2025.11.016

SUMMARY

The tumor immune microenvironment (TIME) critically impacts cancer progression and immunotherapy response. Multiplex immunofluorescence (mIF) is a powerful imaging modality for deciphering TIME, but its applicability is limited by high cost and low throughput. We propose *GigaTIME*, a multimodal AI framework for population-scale TIME modeling by bridging cell morphology and states. *GigaTIME* learns a cross-modal translator to generate virtual mIF images from hematoxylin and eosin (H&E) slides by training on 40 million cells with paired H&E and mIF data across 21 proteins. We applied *GigaTIME* to 14,256 patients from 51 hospitals and over 1,000 clinics across seven US states in Providence Health, generating 299,376 virtual mIF slides spanning 24 cancer types and 306 subtypes. This virtual population uncovered 1,234 statistically significant associations linking proteins, biomarkers, staging, and survival. Such analyses were previously infeasible due to the scarcity of mIF data. Independent validation on 10,200 TCGA patients further corroborated our findings.

INTRODUCTION

The tumor immune microenvironment (TIME) plays a critical role in cancer progression, influencing tumor growth, invasion, metastasis, and response to cancer therapies by affecting tumor immune surveillance and evasion.[1–3] The TIME is a highly complex spatial ecosystem consisting of cancer cells, and diverse non-malignant cell types, including immune cells, cancer-associated fibroblasts (CAFs), endothelial cells (ECs), pericytes, and other cell types, embedded in an altered extracellular matrix.[4] Immunohistochemistry (IHC) visualizes the activation of a specific protein, which offers an important tool for unveiling key cell states in TIME study. For example, PD-L1 IHC staining identifies the activation of the suppressive tumor immune checkpoint PD-L1, which is a widely used predictor for response to checkpoint inhibitor therapies. A critical limitation of IHC is that the activations are evaluated one protein at a time, each on a separate tissue sample. This shortcoming is particularly problematic for tumor microenvironment modeling, where the study of the complex interplay among tumor and immune cells requires the simultaneous evaluation of a variety of proteins. Multiplex immunofluorescence (mIF) has emerged as a powerful alternative, enabling co-localized, multi-channel protein profiling on the same tissue while preserving the spatial architecture.[5–8]

Despite the promise, mIF remains considerably expensive for large-scale study, due to the substantial costs of reagents, specialized equipment, and computational infrastructure, combined with the labor-intensive workflows for staining, imaging, and data processing.[8,9] Consequently, existing mIF datasets are extremely scarce, which significantly limit their applicability in clinical discovery and translation. By contrast, hematoxylin and eosin (H&E) images are routinely generated in clinical workflows at low cost for studying tissue structure and cell morphology.[10,11] While an H&E image does not explicitly reveal cell states, the spatial configuration of cells it reveals can shed light on their individual states. Such patterns might not be obvious for human experts but are potentially discernible using advanced multimodal AI.[12–16] Recent advances in foundation

models further amplify this potential, as AI has demonstrated superior performance by pretraining from a large collection of pathology images.[17–19] Such advances in AI suggest the plausibility of learning pathological features indicative of spatially resolved protein activations.

In this paper, we propose *GigaTIME*, a multimodal AI framework that enables population-scale TIME study by learning to generate virtual mIF images from readily available H&E images. *GigaTIME*'s cross-modal translator was pretrained using 40 million cells with paired H&E and mIF slides for the same tissues. We applied *GigaTIME* to generate virtual mIF images for cancer patients from 51 hospitals and over a thousand clinics across seven US states at Providence Health, yielding 299,376 virtual mIF whole-slide images across 21 protein channels. This enabled us to create a large and diverse virtual population of 14,256 patients across 24 cancer types and 306 cancer subtypes, with virtual mIF slides and key clinical attributes such as biomarkers, staging, and survival information. In turn, this enables us to conduct population-scale clinical discovery, identifying 181, 453, and 600 statistically significant TIME protein-biomarker associations at the pan-cancer, cancer-type, and subtype levels, respectively, all adjusted for multiple testing, forming a comprehensive protein-biomarker association collection for TIME. The virtual population also enabled patient stratification predictive of staging and survival, with the *GigaTIME* signature integrating all virtual protein activations outperforming individual virtual proteins in patient stratification, underscoring the importance of multiplexed analysis. We demonstrated how a virtual population can support spatially informed metrics, such as sharpness and entropy. We further identified combinatorial protein channels that revealed synergistic associations. For example, our *GigaTIME* analysis found that the combination of CD138 and CD68 outperform each protein alone in predicting all representative clinical biomarkers, with statistically significant difference for 13 out of 20 biomarkers, suggesting their combined role in antibody-mediated tumor mechanisms.

To evaluate *GigaTIME*'s generalizability, we applied it to generate an independent virtual population of 10,200 tumors from the Cancer Genome Atlas (TCGA). We observe significant concordance across the virtual populations from Providence and TCGA, which attained a Spearman correlation of 0.88 for virtual protein activations across cancer subtypes. The two populations also uncover a significant overlap of TIME protein-biomarker associations (Fisher's exact test; $p < 2 \times 10^{-9}$). On the other hand, the Providence virtual population yielded 33% more significant associations than TCGA, highlighting the value of large and diverse real-world data for clinical discovery.

GigaTIME offers a promising AI framework for scaling tumor microenvironment modeling by learning to translate readily available H&E images into highly informative virtual mIF images, thus paving the way for advancing precision immuno-oncology through population-scale TIME analysis and discovery. To facilitate future research in tumor microenvironment modeling, we will release our pretrained *GigaTIME* model, as well as our in-house dataset of 40 million cells from de-identified paired H&E and mIF slides across 21 protein channels.

Table 1. TIME markers used in our study and their cellular expression

TIME marker	Typical cellular expression
CD4	T helper
CD138	Plasma cells
CD68	Macrophages
CD14	Monocytes
Tryptase	Mast cell
Caspase 3	Cells undergoing apoptosis
PHH3	Mitotic spindle
Ki67	Proliferating cells (normal and tumor)
T-bet	NK cells, T cells ($CD3^+$, of $CD4^+$ or $CD8^+$ lineages)
Transgelin	Fibroblasts; smooth muscle
CD3	All T cells
CK	Epithelial cell lineage
Actin	Myofibroblast
DAPI	Nuclear marker
CD8	Cytotoxic T cells
CD20	B cells
CD16	CD16a: NK cells; CD16b: neutrophils & macrophages
CD34	Vascular marker (also Hematopoietic precursors, capillary endothelium)
PD-1	T cells, B cells
PD-L1	Antigen presenting cells (APCs), tumor cells, vascular endothelium
CD11c	Dendritic cells (antigen presenting)

Table listing all tumor immune microenvironment (TIME) markers included in our analysis, along with their predominant cellular localization and biological function.

RESULTS

GigaTIME generates a virtual population of multiplex immunofluorescence

GigaTIME enables the generation of a large and diverse virtual population of mIF images from population-scale H&E slides. We first experimentally acquired 441 mIF images from 21 H&E-stained slides across 21 protein channels (see STAR Methods; Figure S1; Table 1). These paired H&E and mIF slides were processed through a computational pipeline that includes image registration and cell segmentation, resulting in a dataset comprising 40 million cells with paired H&E and mIF slides (Figures 1A and S2).

We divided the paired data into training, development, and held-out test sets (see STAR Methods). To translate H&E images into mIF ones, *GigaTIME* was trained on the training paired data using a patch-based encoder-decoder architecture built on NestedUNet.[20] The model inputs an H&E image patch and outputs 21 corresponding mIF patches, one for each protein channel. These channel-specific patches are subsequently stitched together to reconstruct the whole-slide mIF images, enabling spatially resolved, slide-level protein activation profiling.

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

CellPress
OPEN ACCESS

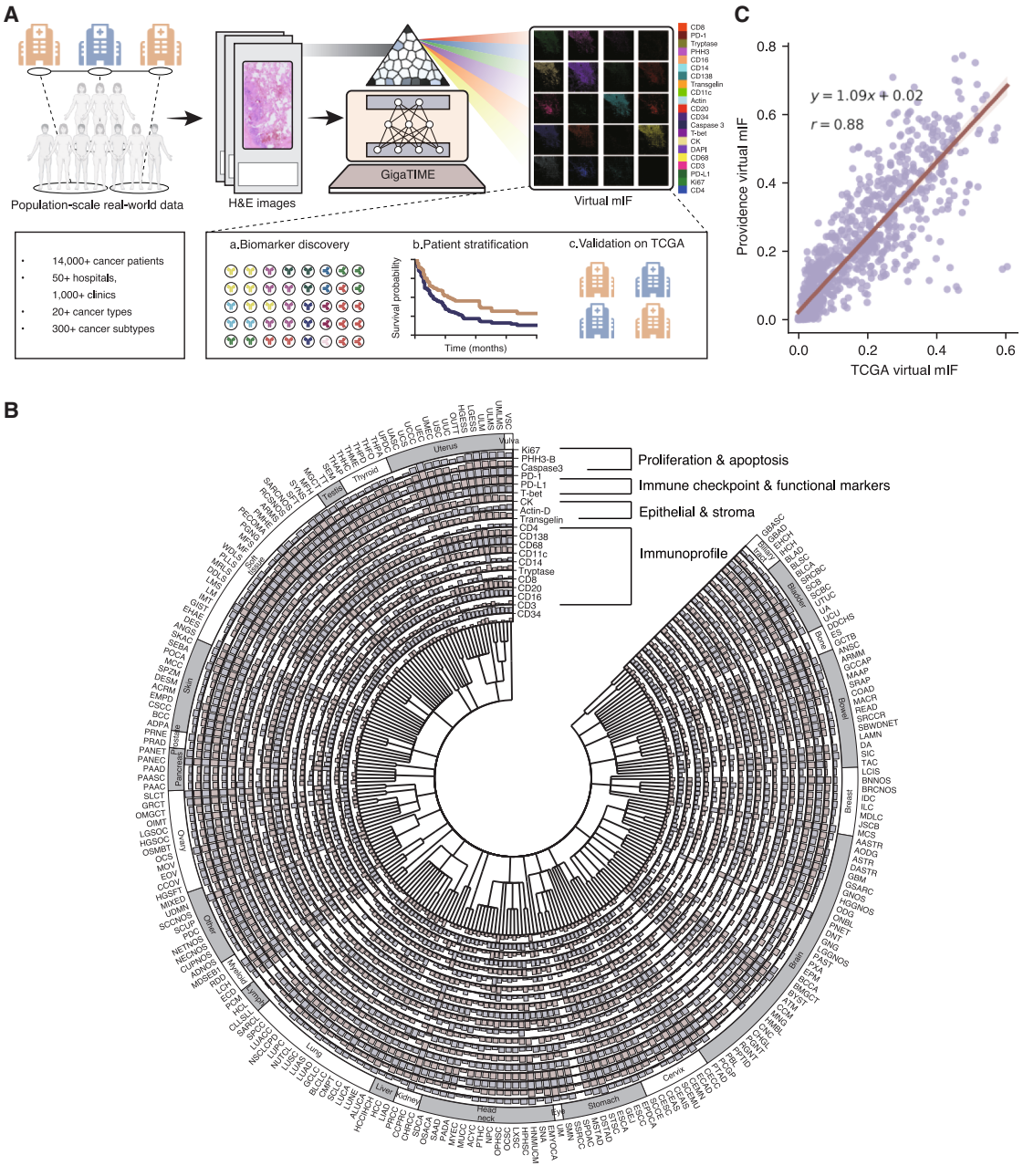


Figure 1. *GigaTIME* enables population-scale tumor immune microenvironment analysis

(A) *GigaTIME* inputs a hematoxylin and eosin (H&E) whole-slide image and outputs multiplex immunofluorescence (mIF) across 21 protein channels. By applying *GigaTIME* to 14,256 patients, we generated a virtual population with mIF information, leading to population-scale discovery on clinical biomarkers and patient stratification, with independent validation on TCGA.

(B) Circular plot visualizing a tumor immune microenvironment (TIME) spectrum encompassing the *GigaTIME*-translated virtual mIF activation scores across different protein channels at the population scale, where each channel is represented as an individual circular bar chart segment. The inner circle encodes OncoTree, which classifies 14,256 patients into 306 subtypes across 24 cancer types. The outer circle groups these activations by cancer types, allowing visual comparison across major categories.

(C) Scatter plot comparing the subtype-level *GigaTIME*-translated virtual mIF activations between TCGA and Providence virtual populations. Each dot denotes the average activation score of a protein channel among all tumors of a cancer subtype.

See also Figures S1, S2, S3, and S4.

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

Effectively, given a protein channel, *GigaTIME* outputs a binary classification label for each pixel indicating whether it is activated for the given protein channel. We can then tally the count of activated pixels, as well as activation density score (proportion of activated pixels), for any image patch or whole slide.

We then applied *GigaTIME* to a large and diverse real-world dataset consisting of 14,256 whole-slide H&E images collected from 51 hospitals and over 1,000 clinics across seven US states at Providence Health, encompassing 24 cancer types and 306 cancer subtypes (Figure S3). Using our trained model, we generated 299,376 virtual mIF whole-slide images for these patients. This enables us to construct a large-scale and diverse multimodal virtual population with H&E and virtual mIF images, along with clinical attributes, such as biomarkers, staging, and survival status.

As a proof of concept, we quantified a protein activation density score for each mIF image, defined as the proportion of activated pixels. These scores were aggregated across tumors of the same subtype using mean pooling, yielding a spectrum of mIF-based TIME features across cancer subtypes (Figure 1B). To evaluate the robustness of our approach, we further applied *GigaTIME* to 10,200 tumors from TCGA, generating 214,200 virtual mIF whole-slide images across 21 channels (Figure S4). A high degree of concordance was observed between the aggregated activation scores derived from the Providence-based and TCGA-based virtual populations (Figure 1C), underscoring the generalizability and reliability of *GigaTIME*.

GigaTIME translates whole-slide H&E images to mIF

We first evaluated *GigaTIME*'s performance in translating standard H&E whole-slide images into mIF images. To benchmark this translation task systematically, we compared *GigaTIME* with CycleGAN,[21] a widely used image translation model frequently applied in virtual staining tasks.[22] Evaluation was conducted using three standard metrics across different levels of granularity (pixel, cell, and slide) to capture both local and global translation fidelity. We also evaluated an average activation baseline by first computing the average positivity rate for each protein channel in the gold mIF data, and then using that as the sampling probability to generate virtual positive pixels for the given channel. At the pixel level, we treated the translation problem as a segmentation task and used the Dice score to evaluate fine-grained spatial concordance between the measured mIF and *GigaTIME*-translated virtual mIF. During evaluation, we skipped all blank regions at the 8×8 level to minimize potential risk for them to inflate the scores.

GigaTIME significantly outperformed CycleGAN on 15 out of 21 protein channels (no statistically significant difference for the remaining 6 channels), highlighting the importance of acquiring paired H&E-mIF data and learning directly from such paired data (Figure 2A). We also observed that the average activation baseline performed very poorly. For example, the test Dice score for average activation baseline was only 0.12 for DAPI, compared to 0.72 attained by *GigaTIME*. The Dice scores for average activation baseline in other protein channels were even worse.

To assess performance at the cell level, we used an 8×8-pixel window and counted the number of activated pixels in each window for each channel. Pearson correlation was then computed between predicted and ground-truth counts across the windows. *GigaTIME* achieved significantly higher correlations than CycleGAN, which performed close to random, suggesting that CycleGAN fails to recover coherent cell-level patterns (Figure 2B). In particular, *GigaTIME* achieved a Pearson correlation of 0.59 on the DAPI channel, which binds DNA and marks nuclei, indicating its capability in identifying and localizing individual cells. By contrast, CycleGAN attains a Pearson correlation of 0.03. The average activation baseline is even worse, attaining zero or negative scores (Figure S8).

These results verified that *GigaTIME* indeed learned non-trivial generalizable information about tissue and protein structures. For evaluating global spatial patterns, we implemented a slide-level metric inspired by immunoscore,[23] a validated scoring method that was originally developed for colorectal cancer and has been used in association studies of patient outcomes and treatment response. We implemented a conceptually similar version by computing the ratio of activated pixels within each 256×256 patch, and then calculated the Spearman correlation for patch-wise activation ratios between the measured mIF and *GigaTIME*-translated virtual mIF. This metric provides a broader view that complements pixel- and cell-level evaluations. *GigaTIME* achieved a Spearman correlation of 0.98 for DAPI and 0.56 for all channels, whereas CycleGAN yielded near-zero correlation for all channels (Figures 2C and S5), validating the necessity of training on paired H&E and mIF data for accurate cross-modal translation. Finally, qualitative comparisons on representative whole-slide image patches further illustrate the agreement between the measured mIF and the *GigaTIME*-translated virtual mIF (Figure 2D).

To further assess the generalizability of *GigaTIME*, we conducted additional analyses based on newly generated paired data to directly evaluate supervised virtual mIF generation on unseen cancer types. Specifically, we curated tumor cores from tissue microarrays (TMAs) of breast and brain, which were not included in the training data. These cohorts encompass a wide spectrum of histological subtypes and clinical stages. In addition to the differences in cancer types, stages, and histologies, the new mIF data comprise many small, cylindrical tissue samples separated by blank regions, which were quite different from the original whole slides with large contiguous tissue regions. Despite all these major differences, *GigaTIME* has demonstrated remarkable generalizability in Dice scores and Pearson correlations (Figures S11 and S12). It also continued to outperform CycleGAN and random imputation by a wide margin. We also observed that the ordering of translation quality is broadly conserved across the cancer types.

We also provide stratified results in Figure S13 based on subcellular localization: nuclear (DAPI, PHH3, Ki-67, and T-bet), surface (CD4, CD138, CD11c, CD14, CD3, CD8, CD20, CD16, CD34, PD-1, and PD-L1), and cytoplasmic (CD68, tryptase, caspase-3, transgelin, cytokeratin, and actin). Generally, nuclear proteins have better translation quality than cytoplasmic and surface proteins. This makes sense as nuclear proteins, such as DAPI and Ki-67, generally exhibit compact structures with clear histological boundaries. By contrast, cytoplasmic and surface proteins often exhibit diffuse or heterogeneous spatial patterns, making them inherently more challenging to predict.

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

CellPress
OPEN ACCESS

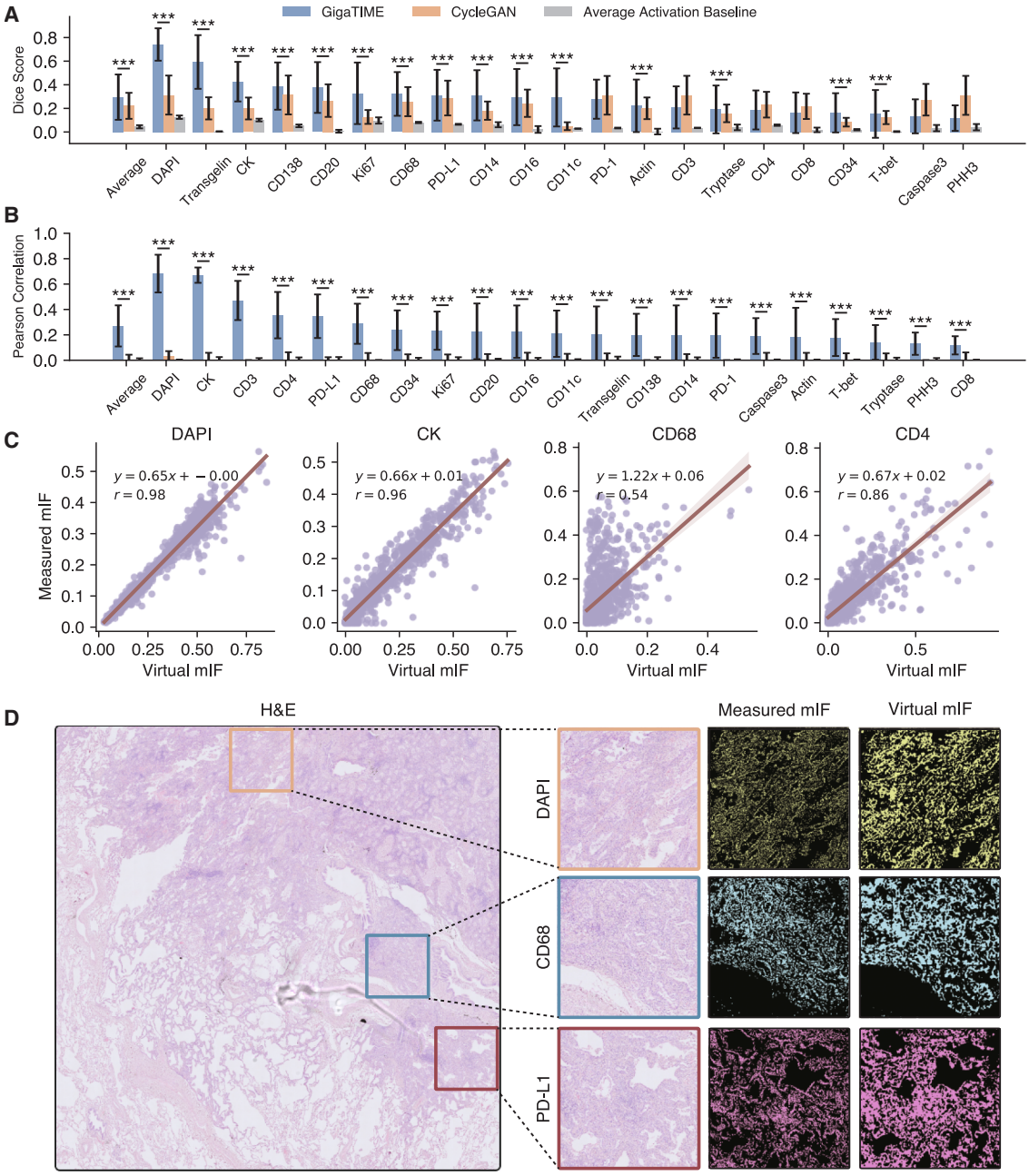


Figure 2. *GigaTIME* enables translation from H&E to mIF images

(A and B) Bar plot comparing *GigaTIME* and CycleGAN on the translation performance in terms of Dice score (A) and Pearson correlation (B) with error bars denoting standard deviation and significance levels from two-tailed t tests represented by *p ≤ 0.05, **p ≤ 0.01, and ***p ≤ 0.001.

(C) Scatter plots comparing the activation density of the translated mIF and the ground-truth mIF across four channels.

(D) Qualitative results for a sample H&E whole-slide image from our held-out test set with zoomed-in visualizations of the measured mIF and *GigaTIME*-translated mIF for DAPI, PD-L1, and CD68 channels.

See also Figures S5, S8, S9, S10, S11, S13, and S14.

Virtual population enables large-scale discovery of protein-biomarker associations

After validating *GigaTIME*'s performance in translating whole-slide H&E images into mIF images, we next examined how the virtual mIF population can enable large-scale clinical discovery. Specifically, our virtual population identified 1,234 statistically significant associations between 21 *GigaTIME*-translated virtual protein channels and 20 clinical biomarkers at pan-cancer, cancer-type, and cancer-subtype levels (Figure 3A). Statistical significance was determined by

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

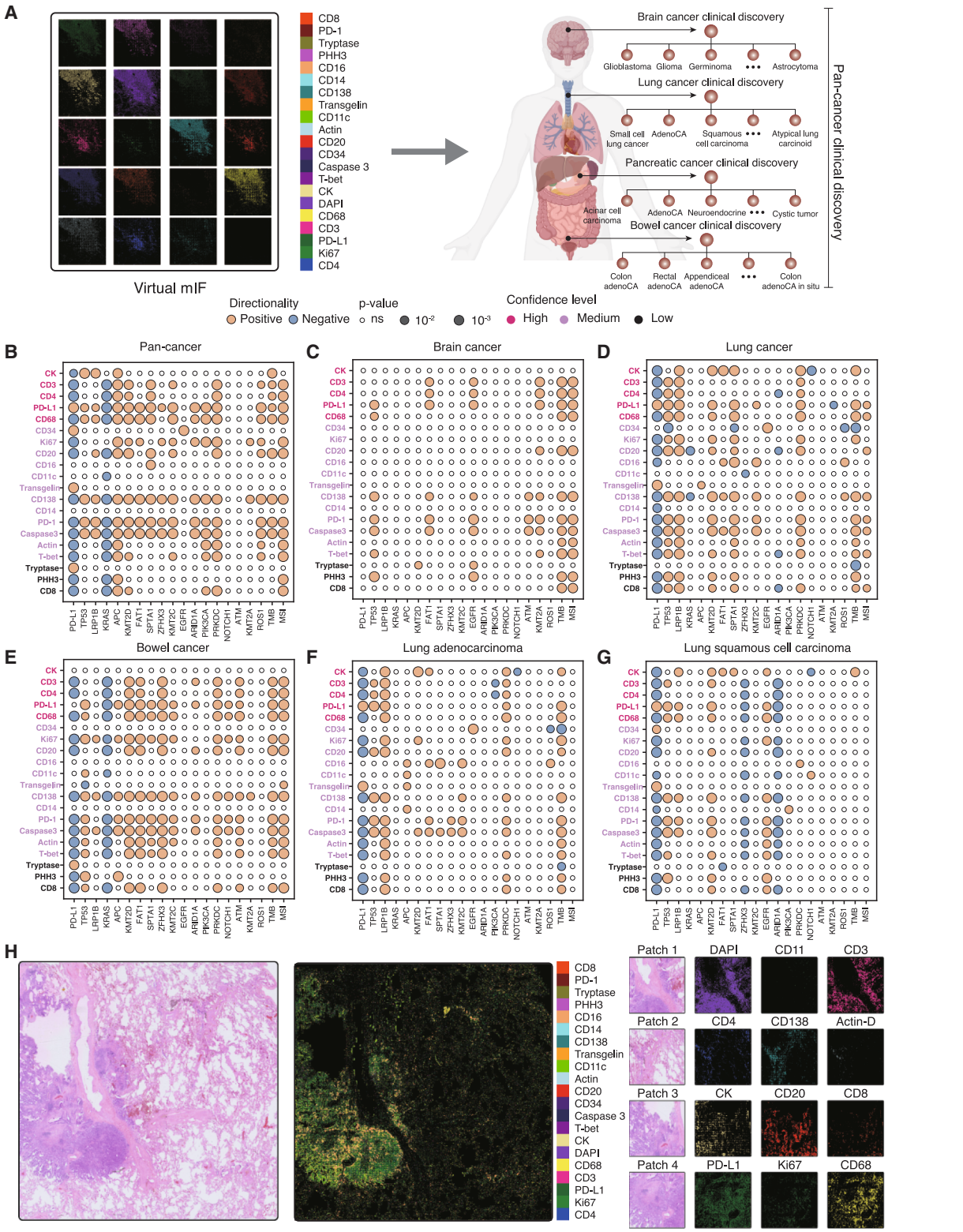


Figure 3. *GigaTIME* identifies novel TIME protein vs. biomarker associations at pan-cancer, cancer-type, and cancer-subtype levels

(A) *GigaTIME* generates a virtual population of 14,256 with virtual mIF by translating available H&E images to mIF images, enabling pan-cancer, cancer-type, and cancer-subtype levels of biomedical discovery.

(legend continued on next page)

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

comparing protein activation densities between biomarker-positive and biomarker-negative groups using *t* test, with *p* values adjusted for multiple hypothesis testing (see STAR Methods).

At the pan-cancer level, we identified 175 significant protein-biomarker associations (Figure 3B). Many of these findings are supported by existing literature. For instance, both high tumor mutational burden (TMB-H) and microsatellite instability-high (MSI-H) genotypes exhibited strong associations with increased activation in TIME-related channels such as CD138, CD20, CD68, and CD4, consistent with the well-described effects of antigenic-mediated immune activation.[24,25] Among genomic alterations, KMT2D mutations show strong positive correlations with multiple immune markers, including CD3, CD8, and CD20, suggesting KMT2D mutations are associated with enhanced immune infiltration at a pan-cancer level.[26] Conversely, KRAS mutations were negatively associated with markers of immune infiltration, such as CD3 and CD8 at the pan-cancer level, reflecting an immune-excluded phenotype. Interestingly, KRAS mutations also associated negatively with PD-L1 expression, despite established mechanisms by which KRAS signaling promotes PD-L1 expression via the ERK pathway.[27,28] Further investigation is needed to determine whether such correlations are causal or merely stem from confounding effects.

The clinically reported PD-L1 biomarker, assessed via IHC, was positively correlated with virtual PD-L1 channel activation, indicating strong concordance between virtual population-based predictions and clinically observed protein expression. Moreover, PD-L1 IHC showed negative associations with several TIME markers, including CD3, CD8, and CD20, reflecting the role of the PD-1/PD-L1 checkpoint in establishing immune blockade.[29] Additionally, PD-L1 expression was inversely correlated with proliferation markers (Ki-67 and PHH3) and the apoptosis marker caspase 3. One explanation for these findings is that PD-L1 expression is associated with less proliferative tumor states that demonstrate resistance to apoptosis. Alternatively, it is also possible that immune cells are simply less proliferative when there is high PD-L1.

At the cancer-type level, we observed a high number of associations in brain (Figure 3C), lung (Figure 3D), and bowel cancers (Figure 3E), which are also the most represented cancer types in our dataset (Figure S3). Specifically, *GigaTIME* identified 64 protein-biomarker associations in brain cancer, 137 in lung cancer, and 175 in bowel cancer. Many of these associations were specific to particular cancer types: for example, 48 pairs were unique to lung cancer, while 30 were unique to brain cancer, highlighting *GigaTIME*'s ability to uncover cancer-type-specific associations. In brain cancer, we found a strong correlation between T-bet and TP53 mutations, as well as between caspase 3 and KMT2A alterations. Neither of these is detected at the pan-cancer level, which might stem from the unique immune microenvironment of central nervous system malignancies.[30]

At the cancer-subtype level, the virtual population revealed histology-specific associations that are often difficult to detect in smaller cohorts (Figures 3F and 3G). For example, in lung adenocarcinoma, we observed stronger associations between PRKDC mutations, a potential target for enhancing the activity of checkpoint inhibitors[31,32] and immune response markers compared with lung squamous cell carcinoma, underscoring the importance of histological context in interpreting biomarker-genomic associations. In several cases, the subtype-level analysis revealed unexpected patterns that might warrant further investigation. For instance, in lung squamous cell carcinoma (LUSC), high tumor mutational burden (TMB-H) was not associated with an increased predicted immune response, deviating from the canonical link between TMB and enhanced immunogenicity. This is potentially related to the inferior outcomes in this subset of tumors for immune checkpoint inhibitors compared to those with non-squamous disease.[33] Similarly unique to LUSC is the negative association between ARID1A mutations and *GigaTIME*-translated activation of immune infiltration markers, a finding that may mechanistically be related to the reported worse clinical outcomes in this subset of patients.[34]

To further illustrate the power of *GigaTIME*, we highlighted a representative patient diagnosed with lung adenocarcinoma from the Providence data. The H&E image and its corresponding *GigaTIME*-translated virtual mIF images are shown in Figure 3H. The virtual mIF image patches exhibit diverse spatial activation patterns across multiple immune and tumor markers, demonstrating the ability of *GigaTIME* to capture distinct cellular functions and states. This patient harbored TMB-H and exhibited high PD-L1 and CD68 activation in the virtual mIF. This aligns with our identified associations in pan-cancer (Figure 3B), lung cancer (Figure 3D), and lung adenocarcinoma (LUAD)-specific (Figure 3F) analyses, further supporting the biological fidelity and clinical utility of our virtual population.

Virtual population enables large-scale discovery on pathological stages and patient stratification

In addition to clinical biomarkers, our virtual population enables systematic investigation of pathological stages and patient stratification through *GigaTIME*-translated multiplex protein activations. At the pan-cancer level, we identified significant associations between protein channels and pathological stages (Figure 4A). Consistent with prior studies,[35–38] primary tumor size (T stage) exhibited positive correlations with immune checkpoint markers such as PD-L1 and PD-1, as well as immune infiltration markers including CD68 and CD138. By contrast, nodal status (N stage) showed limited associations with these immune markers. The pathological stage group, which integrates TNM classifications, reflected a mixed pattern that aligns with these individual staging findings.

(B–G) Correlation analysis between protein channels in virtual mIF and patient biomarkers reveal TIME protein-biomarker associations at pan-cancer level (B), cancer-type level (C–E), and cancer-subtype level (F and G). Circle size denotes significance strength. Circle color denotes the directionality in which the correlation occurs. Channel color denotes high, medium, and low confidence based on Pearson correlations evaluated using the test set.
(H) A case study showcasing the activation maps across different virtual mIF channels for a H&E slide in our virtual population, and virtual mIF of sample patches from this slide.

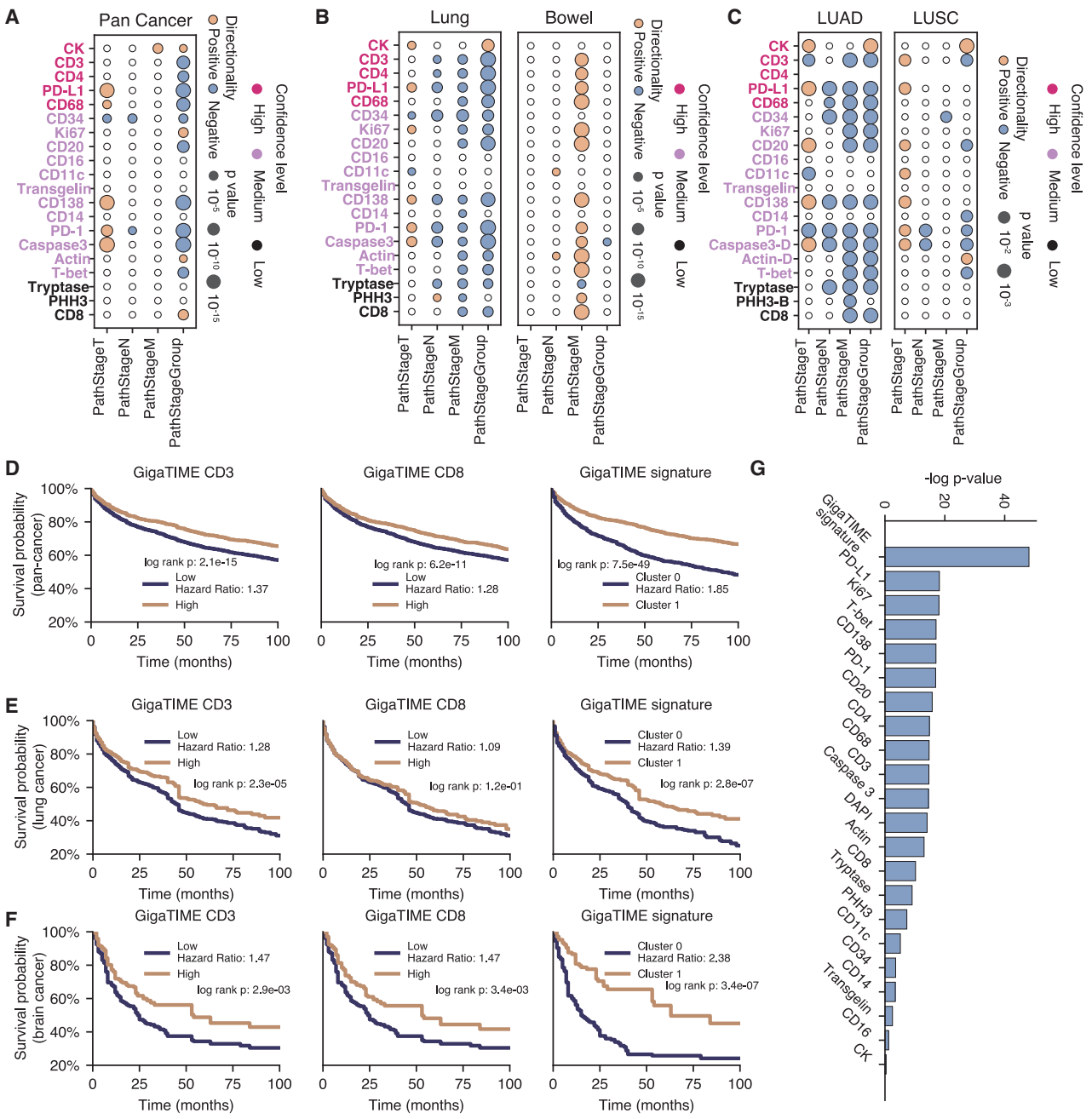


Figure 4. *GigaTIME* enables effective patient stratification across pathological stages and survival groups

(A–C) Correlation analysis between virtual mIF and pathological stages at pan-cancer level (A), cancer-type level (B), and cancer-subtype level (C). Circle size denotes significance strength. Circle color denotes the directionality in which the correlation happens. Channel color denotes high, medium, and low confidence based on Pearson correlations evaluated using test set.

(D–F) Survival analysis on lung cancer by using virtual CD3, virtual CD8, and virtual *GigaTIME* signature (all 21 *GigaTIME* protein channels) to stratify patients at a pan-cancer level (D) and cancer-type level: lung (E) and brain (F).

(G) Bar plot comparing pan-cancer patient stratification performance in terms of survival log-rank *p*-values among virtual *GigaTIME* signature and individual virtual protein channels.

See also Figure S6.

At the cancer-type level, these protein-stage associations varied substantially across cancer types (Figure 4B). Lung cancer largely mirrored the pan-cancer patterns, with the notable exception of N stage demonstrating a negative correlation with both immune checkpoint and infiltration markers, including PD-L1, PD-1, CD3, and CD138. Interestingly, bowel cancer

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

exhibited a unique pattern in which metastatic status (M stage) positively correlated with an increased *GigaTIME*-translated activation of PD-L1 and other immune markers, deviating from trends observed in the pan-cancer analysis, a pattern possibly related to the unique features of the portal/hepatic metastatic spread of this tumor type.[39,40]

Further subtype-level analysis in lung cancer revealed nuanced differences between LUAD and LUSC (Figure 4C). While both subtypes showed broadly consistent staging patterns, LUAD displayed a notably stronger negative correlation between nodal involvement and immune marker expression, particularly for CD34, PD-1, and PD-L1. This suggests that LUAD with nodal metastases may exhibit more pronounced immune evasion at the primary tumor site, highlighting the importance of histology approaches in developing predictive signatures for staging in clinical applications.

Finally, to further evaluate the clinical relevance of virtual population, we examined whether the virtual mIF can help stratify patients by survival outcomes. Across pan-cancer cohorts (Figures 4D and S6), as well as within specific cancer types (Figures 4E and 4F), *GigaTIME*-translated virtual protein activations were sufficient to separate patients into subgroups with significantly different survival trajectories. More importantly, integrating all 21 virtual protein channels into a composite *GigaTIME* signature yielded even more superior stratification (Figure 4G), underscoring the complementary signals across channels, validating the utility of a virtual population with mIF.

Independent validation using a virtual population from TCGA

Next, we validated the biomarker associations identified in the Providence virtual population using an independent virtual population generated from TCGA. Specifically, for each TCGA tumor with H&E slides, we applied *GigaTIME* to generate the corresponding virtual mIF images, resulting in a virtual population of 10,200 patients with 214,200 virtual mIF images across 21 protein channels. We then identified the significant TIME protein-biomarker associations and compared them with that for the Providence virtual population (Figure 5A).

Note that there are substantial differences between the Providence and TCGA tumors. The Providence dataset represents a diverse real-world population in standard clinical settings, whereas TCGA primarily includes early-stage, untreated tumors from surgical resection. Moreover, not all biomarkers in the Providence dataset are available in TCGA and vice versa.

Interestingly, despite all these differences, we observed encouraging agreements in a number of our findings, which bode well for the generalizability of *GigaTIME*. We have already reported the general concordance of virtual mIF activations in Providence and TCGA virtual populations, with a Spearman correlation of 0.88 (Figure 5A). Moreover, there are 80 protein-biomarker associations that are statistically significant in both Providence and TCGA, which is a highly significant enrichment (Fisher's exact test; $p < 2 \times 10^{-9}$), underscoring the generalizability and robustness of *GigaTIME*.

Meanwhile, it is also notable that the Providence virtual population yielded 33% more significant pan-cancer associations than TCGA, highlighting the value of large-scale real-world data for clinical discovery. Moreover, the Providence virtual population also uncovered substantially more associations at the cancer-type level, such as lung cancer (Figure 5B), as well as at the cancer-subtype level, such as LUAD (Figure 5C). By contrast, TCGA identified few significant associations at such granular levels (Figures 5B and S7).

To further validate *GigaTIME*'s generalizability in TCGA, we examined translated mIF images from different samples in TCGA, focusing on biomarker-associated differences in protein expression (Figure 5D). One case harbored a SPTA1 mutation, while the other did not. The resulting virtual mIF images displayed markedly different CD20 activation patterns, consistent with the association between CD20 and SPTA1 mutations. Similar visual distinctions were observed in case studies involving KMT2D and PD-L1, as well as TP53 and CD138, further reinforcing virtual population's ability to capture biologically meaningful associations between biomarkers and spatial protein expression. When the biomarker status is not mutated, it can be seen that the corresponding protein channel appears nearly black due to the very sparse predicted signal in these regions, which aligns with the expected low expression in non-mutated tissues.

Virtual population uncovers interesting spatial and combinatorial protein activation patterns

While activation density has been a widely used score to aggregate signals across a whole-slide image, it does not reflect the more sophisticated global patterns as revealed by spatial technologies such as mIF. By generating virtual mIF whole-slide images, *GigaTIME* enables the exploration of more sophisticated metrics that can capture richer spatial patterns. Such metrics can reveal the distribution, organization, and higher-order patterns of protein activations, offering deeper insights into the tumor immune microenvironment.

To this end, we evaluated three standard spatially aware metrics: entropy, signal-to-noise ratio (SNR), and sharpness. Entropy quantifies the complexity and non-uniformity of pixel distributions, reflecting the heterogeneity of spatial patterns in an image. SNR signifies the strength of the true biological signal relative to background noise, highlighting feature reliability. Sharpness assesses the prevalence of edges and fine structures, which is critical for identifying subtle yet meaningful patterns in protein localization.

We applied these spatially aware metrics to our Providence virtual population. We found that they often uncovered stronger associations with specific clinical biomarkers compared with density (Figures 6A–6C). In total, entropy, SNR, and sharpness yielded higher correlation strength than density in 89, 63, and 79 TIME protein-biomarker pairs, respectively, while being comparable for the rest.

In addition to sophisticated spatial patterns, we also investigated whether combining virtual protein activations could enhance detection of biomarker associations. As a preliminary exploration, we assessed pairwise virtual protein combinations using the OR logical operation (i.e., a pixel is activated if it is activated for either virtual protein) and evaluated their correlations with clinical biomarkers (Figures 6D and 6E). This approach uncovered numerous biomarker associations that were much stronger for the combination than individual virtual proteins.

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

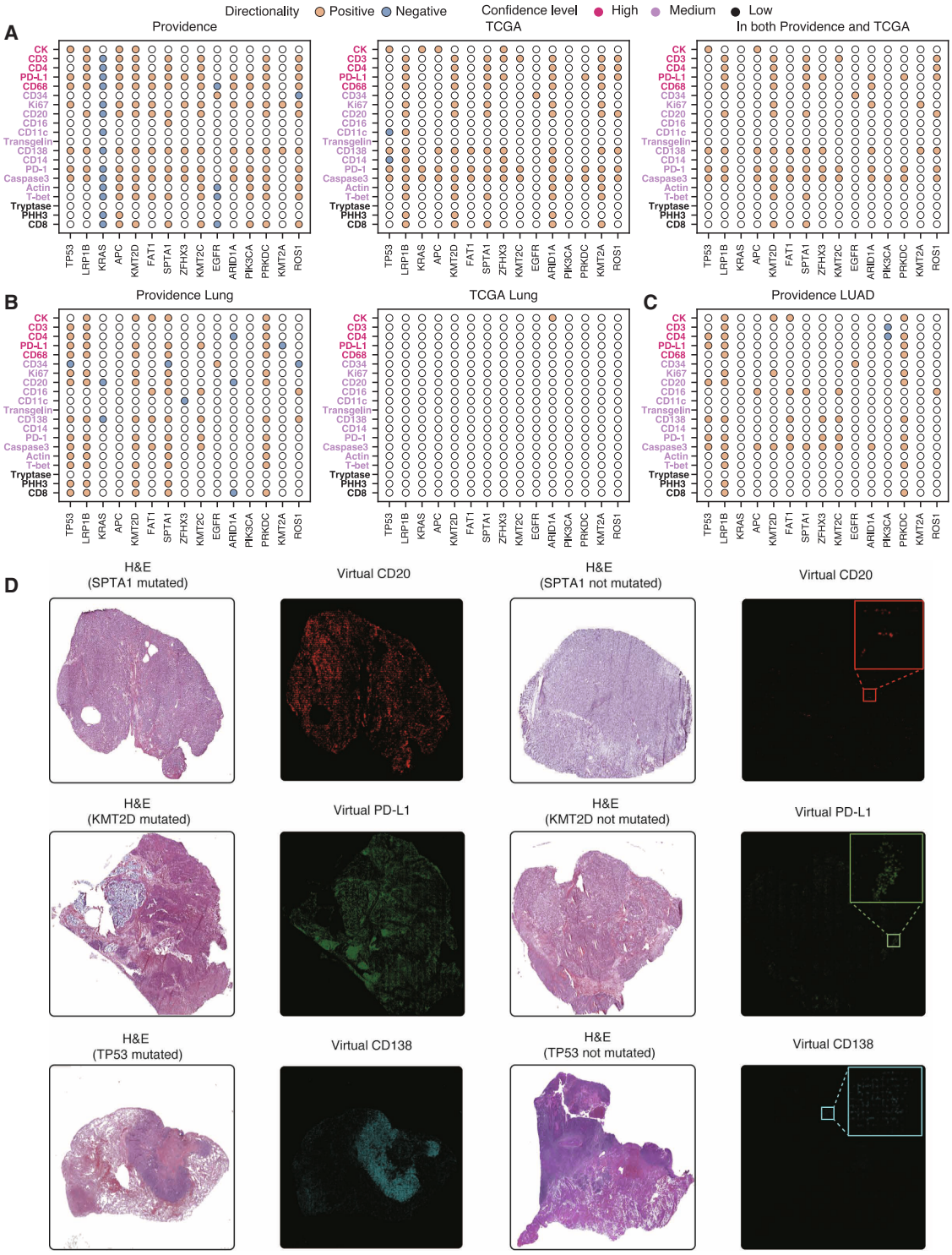


Figure 5. Independent validation on a virtual population from TCGA

(A) Grid charts showing significantly correlated pan-cancer *GigaTIME* protein-biomarker pairs in Providence (left), TCGA (middle), and both (right).

(B) Grid charts showing significantly correlated *GigaTIME* protein-biomarker pairs for lung cancer in Providence and TCGA.

(C) Grid chart showing significantly correlated *GigaTIME* protein-biomarker pairs for LUAD in Providence. Channel color denotes high, medium, and low confidence based on Pearson correlations evaluated using test set.

(D) Case studies with visualizations of H&E slides and the corresponding virtual mIF activations for the pair of a *GigaTIME* protein channel and a biomarker (mutated/non-mutated), where the patient with the given mutation demonstrates much higher activation scores for that *GigaTIME* protein channel.

See also Figure S7.

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

CellPress
OPEN ACCESS

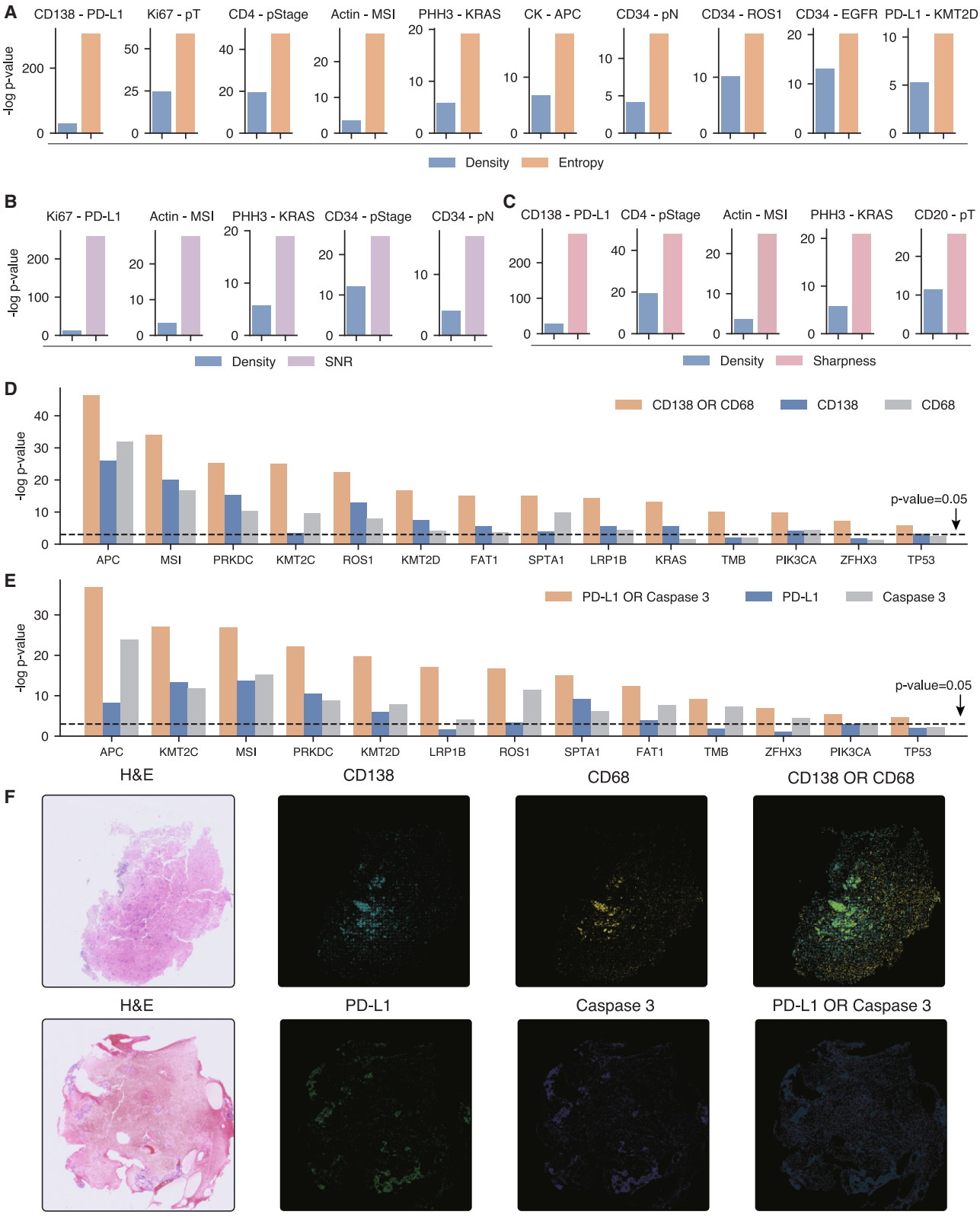


Figure 6. *GigaTIME* uncovers interesting spatial and combinatorial virtual mIF patterns

(A–C) Bar plots comparing virtual mIF activation density with spatial metrics on identifying TIME protein-biomarker correlations. We investigated three spatial metrics based on entropy (A), signal-to-noise ratio (SNR) (B), and sharpness (C).

(legend continued on next page)

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

For example, combining the virtual activations for CD138 (a plasma cell marker) and CD68 (a macrophage marker), two distinct yet complementary immune proteins, resulted in much stronger associations with multiple biomarkers than either protein alone (Figure 6C). The co-localization of $CD138^+$ plasma cells with $CD68^+$ macrophages may signify an active immune attack on the tumor via antibody-mediated mechanisms, as previous studies have shown that a robust local antibody presence can stimulate macrophages to target and destroy antibody-coated tumor cells.[41] A similar phenomenon was observed for the combination of PD-L1 (an immune checkpoint ligand) and caspase 3 (an apoptosis marker). This pairing captures the interplay between immune evasion and cell death. In lung cancer, PD-L1 is known to inhibit apoptotic pathways, including suppression of caspase 3 and caspase 8 activation.[42] Similarly, in ovarian cancer, downregulation of PD-L1 has been shown to increase caspase 3 activity and apoptotic cell death.[43] Finally, case studies also qualitatively demonstrate how combinations of CD138/CD68 and PD-L1/caspase 3 channels reveal complementary protein activations in the output virtual mIF (Figure 6F). These findings support the utility of combinatorial analysis, where integrating biologically complementary activation signals leads to more informative and clinically relevant insights.

DISCUSSION

GigaTIME offers a multimodal AI framework for scaling TIME modeling by learning to translate readily available H&E images into mIF images. *GigaTIME* was trained on a newly generated cross-modal dataset of 40 million cells with paired H&E and mIF slides and subsequently applied to H&E slides from Providence Health with 51 hospitals and over a thousand clinics across seven US states, generating 299,376 virtual mIF whole-slide images across 21 protein channels. Comprehensive analysis on the resulting virtual population of 14,256 cancer patients yielded a TIME spectrum across 24 cancer types and 306 cancer subtypes, as well as a TIME atlas with 1,234 statistically significant associations between virtual protein channels and key clinical biomarkers. *GigaTIME* uncovered interesting spatial and combinatorial protein activation patterns, and the *GigaTIME* signature of virtual proteins facilitated effective patient stratification predictive of cancer staging and survival. Independent validation on TCGA demonstrated the generalizability of *GigaTIME*, while also highlighted the value of diverse real-world data such as Providence for clinical discovery.

In our exploration of *GigaTIME* virtual populations, multiple biological features of the tumor immune response emerged. Intriguingly, the extent of local tumor invasion/pathological tumor (T) stage appeared to be associated, at a pan-cancer level, with increased virtual PD-L1 activation, as well as more complex virtual protein activation patterns indicative of a coordinated immune response. As tumors progress from localized to locoregional or advanced metastatic disease, this relationship appeared to invert: higher-stage locoregional and metastatic disease seem to be driven primarily by alternative immune evasion mechanisms, rather than active immune inhibition mediated by PD-L1-associated pathways. Notably, this dynamic was also reflected in cell-intrinsic features, including diminished predicted cleaved caspase 3 expression, possibly indicating evasion from immune-induced apoptotic pathways. A key commonality across all these findings is the apparent coordinated nature of the immune response, where protein activations for T and B cells, plasma cells, macrophages, and dendritic cells markers show significant cross-correlation. This reinforced the notion that effective immunotherapies should not target isolated cell compartments but instead adopt a comprehensive approach engaging multiple immune cell types to achieve robust anti-tumor immunity. The importance of the coordinated nature of the TIME was further emphasized by the superiority of our *GigaTIME* signature combining all protein channels over single channels in predicting survival. Associations between molecular alterations and *GigaTIME* virtual protein activations included some well-known genomic relationships, such as the positive correlations with TMB and MSI, while others are less well described, such as KMT2D mutations. We also observed associations between decreased immunogenicity and mutations in oncogenes such as KRAS. Interestingly, subset analysis further reinforced the importance of the histopathological context on the effect of mutational alterations on TIME, with distinctive differences in their impact across different cancer types and subtypes, such as LUAD and LUSC.

Limitations of the study

Our initial results are promising, but much more remains to be explored. Our association study was conducted on 14,256 patients from 51 hospitals and over 1,000 health centers. To our knowledge, this already comprises by far the largest population-scale study for virtual mIF. Still, the majority of our patients are from the western United States. There is substantial opportunity for further enhancing the geographic and ethnic diversity in the patient cohort to increase the coverage of under-represented populations. It is remarkable that our results on *GigaTIME* show that H&E slides indeed contain significant signals for spatial proteomics. However, not all proteins are equal, and AI is not magic. It is expected that some proteins might not manifest prominently in the morphology and are thus untranslatable based on H&E alone. Our results already unveil substantial variations in the translation quality across protein channels (Figure S5). This variability likely stems from three factors: the heterogeneous H&E tumor/normal tissue architecture, the varying frequency of positive events in the training datasets, and marker-specific technical challenges such as non-specific binding patterns, varying expression levels, and differential sensitivity to tissue processing methods. Systematic quantification of this variability can identify protein channels with major growth opportunities in translational performance, thus helping to guide future data generation. In future work, we plan to explore more protein channels and assess their translation quality, ultimately

(D and E) Bar plots comparing single channel and combinatorial channel (using the OR logical operation) in biomarker associations for two *GigaTIME* virtual protein pairs: CD138/CD68 (D) and PD-L1/caspase 3 (E), demonstrating substantially improved associations for the combination.
(F) Case studies visualizing the virtual mIF activation maps of individual channels (CD138, CD68; PD-L1, caspase 3) and their combinations.

compiling a comprehensive atlas for virtual mIF. A key motivation for *GigaTIME* is to unveil the "tumor microenvironment grammar" underlying the elaborate patterns of cell-to-cell interactions. In future work, we plan to integrate cell segmentation models[44–46] into *GigaTIME* training and inference to model cell-to-cell interactions.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Hoifung Poon (hoifung@microsoft.com).

Materials availability

The study did not generate new unique reagents.

Data and code availability

GigaTIME is accessible at https://aka.ms/gigatime_code, including the model weights and relevant source code.

ACKNOWLEDGMENTS

We thank Flaticon for open source icons, which we used in figures of this paper.

AUTHOR CONTRIBUTIONS

Conceptualization, C.B., S.W., and H.P.; methodology, J.M.J.V., N.U., S.W., and H.P.; code development, J.M.J.V., H.X., N.U., and C.K.; data infrastructure and pipeline, J.M.J.V., H.X., N.U., C.K., C.W., P.A., Y.G., S.Z., T.N., and J.B.; data acquisition, R.B.S., A.C., K.M., A.Q.B., B.A.F., B.W., A.R., and B.P.; real-world data testing and validation, J.M.J.V. and H.X.; writing—original draft, J.M.J.V., C.B., S.W., and H.P.; review & editing, all authors; visualization, J.M.J.V. and H.X.; supervision, C.B., S.W., and H.P.

DECLARATION OF INTERESTS

C.B. serves on the scientific advisory board of BioAI, PictorLabs, Lunaphore, and SironaDx; has a consultant and advisory relationship at Roche and Merk; and is an inventor of the following patents: US10552960B2, AU2015265811B2, and US12094105B2.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
  - Details of acquiring paired training and testing dataset for *GigaTIME*
  - Details of registration and preprocessing on acquired mIF images
  - Details of training *GigaTIME* as cross-modal translator model
  - Details of baseline comparisons
  - Details of metrics to evaluate the translation performance
  - Details of Providence data for generating virtual population
  - Details of generating virtual population data
  - Details of biomarker discovery using virtual population
  - Details of pathological staging and surival analysis using virtual population
  - Details of the TCGA-based virtual population
  - Details of the spatially informed metrics
  - Details of the combinatorial analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

Received: April 24, 2025
Revised: August 20, 2025
Accepted: November 10, 2025

REFERENCES

1. Nagarsheth, N., Wicha, M.S., and Zou, W. (2017). Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. Nat. Rev. Immunol. *17*, 559–572. https://doi.org/10.1038/nri.2017.49.
2. Binnewies, M., Roberts, E.W., Kersten, K., Chan, V., Fearon, D.F., Merad, M., Coussens, L.M., Gabrilovich, D.I., Ostrand-Rosenberg, S., Hedrick, C.C., et al. (2018). Understanding the tumor immune microenvironment (time) for effective therapy. Nat. Med. *24*, 541–550. https://doi.org/10.1038/s41591-018-0014-x.
3. Junttila, M.R., and De Sauvage, F.J. (2013). Influence of tumour microenvironment heterogeneity on therapeutic response. Nature *501*, 346–354. https://doi.org/10.1038/nature12626.
4. De Visser, K.E., and Joyce, J.A. (2023). The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. Cancer Cell *41*, 374–403. https://doi.org/10.1016/j.ccell.2023.02.016.
5. Rivest, F., Eroglu, D., Pelz, B., Kowal, J., Kehren, A., Navikas, V., Procopio, M.G., Bordignon, P., Pérès, E., Ammann, M., et al. (2023). Fully automated sequential immunofluorescence (seqif) for hyperplex spatial proteomics. Sci. Rep. *13*, 16994. https://doi.org/10.1038/s41598-023-43435-w.
6. 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., and Lim, T.K.H. (2020). Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. Cancer Commun. (Lond) *40*, 135–153. https://doi.org/10.1002/cac2.12023.
7. Hoyt, C.C. (2021). Multiplex immunofluorescence and multispectral imaging: forming the basis of a clinical test platform for immuno-oncology. Front. Mol. Biosci. *8*, 674747. https://doi.org/10.3389/fmolb.2021.674747.
8. Lin, J.R., Chen, Y.A., Campton, D., Cooper, J., Coy, S., Yapp, C., Tefft, J.B., McCarty, E., Ligon, K.L., Rodig, S.J., et al. (2023). High-plex immunofluorescence imaging and traditional histology of the same tissue section for discovering image-based biomarkers. Nat. Cancer *4*, 1036–1052. https://doi.org/10.1038/s43018-023-00576-1.
9. Lin, J.R., Wang, S., Coy, S., Chen, Y.A., Yapp, C., Tyler, M., Nariya, M.K., Heiser, C.N., Lau, K.S., Santagata, S., et al. (2023). Multiplexed 3d atlas of state transitions and immune interaction in colorectal cancer. Cell *186*, 363–381.e19. https://doi.org/10.1016/j.cell.2022.12.028.
10. De Matos, L.L., Trufelli, D.C., De Matos, M.G.L., and da Silva Pinhal, M.A. (2010). Immunohistochemistry as an important tool in biomarkers detection and clinical practice. Biomark. Insights *5*, 9–20. https://doi.org/10.4137/BMI.S2185.
11. Wang, S., Rong, R., Yang, D.M., Fujimoto, J., Yan, S., Cai, L., Yang, L., Luo, D., Behrens, C., Parra, E.R., et al. (2020). Computational staining of pathology images to study the tumor microenvironment in lung cancer. Cancer Res. *80*, 2056–2066. https://doi.org/10.1158/0008-5472.CAN-19-1629.
12. de Haan, K., Zhang, Y., Zuckerman, J.E., Liu, T., Sisk, A.E., Diaz, M.F.P., Jen, K.Y., Nobori, A., Liou, S., Zhang, S., et al. (2021). Deep learning-based transformation of h&e stained tissues into special stains. Nat. Commun. *12*, 4884. https://doi.org/10.1038/s41467-021-25221-2.
13. Zhang, Y., de Haan, K., Rivenson, Y., Li, J., Delis, A., and Ozcan, A. (2020). Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue. Light Sci. Appl. *9*, 78. https://doi.org/10.1038/s41377-020-0315-y.
14. Bouteldja, N., Klinkhammer, B.M., Schlaich, T., Boor, P., and Merhof, D. (2022). Improving unsupervised stain-to-stain translation using self-supervision and meta-learning. J. Pathol. Inform. *13*, 100107. https://doi.org/10.1016/j.jpi.2022.100107.

15. Zhang, R., Cao, Y., Li, Y., Liu, Z., Wang, J., He, J., Zhang, C., Sui, X., Zhang, P., Cui, L., et al. (2022). Mvfstain: multiple virtual functional stain histopathology images generation based on specific domain mapping. Med. Image Anal. *80*, 102520. https://doi.org/10.1016/j.media.2022.102520.

16. Pati, P., Karkampouna, S., Bonollo, F., Compérat, E., Radić, M., Spahn, M., Martinelli, A., Wartenberg, M., Kruithof-de Julio, M., and Rapsomaniki, M. (2024). Accelerating histopathology workflows with generative ai-based virtually multiplexed tumour profiling. Nat. Mach. Intell. *6*, 1077–1093. https://doi.org/10.1038/s42256-024-00889-5.

17. Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., Wong, C., Gero, Z., González, J., Gu, Y., et al. (2024). A whole-slide foundation model for digital pathology from real-world data. Nature *630*, 181–188. https://doi.org/10.1038/s41586-024-07441-w.

18. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F.K., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al. (2024). Towards a general-purpose foundation model for computational pathology. Nat. Med. *30*, 850–862. https://doi.org/10.1038/s41591-024-02857-3.

19. Wang, X., Zhao, J., Marostica, E., Yuan, W., Jin, J., Zhang, J., Li, R., Tang, H., Wang, K., Li, Y., et al. (2024). A pathology foundation model for cancer diagnosis and prognosis prediction. Nature *634*, 970–978. https://doi.org/10.1038/s41586-024-07894-z.

20. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., and Liang, J. (2018). UNet++: A nested U-net architecture for medical image segmentation. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support (2018) *11045*, 3–11. https://doi.org/10.1007/978-3-030-00889-5_1.

21. Zhu, J.Y., Park, T., Isola, P., and Efros, A.A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In 2017 IEEE International Conference on Computer Vision (ICCV) (IEEE), pp. 2242–2251. https://doi.org/10.1109/ICCV.2017.244.

22. Latonen, L., Koivukoski, S., Khan, U., and Ruusuvuori, P. (2024). Virtual staining for histology by deep learning. Trends Biotechnol. *42*, 1177–1191. https://doi.org/10.1016/j.tibtech.2024.02.009.

23. Pagès, F., Mlecnik, B., Marliot, F., Bindea, G., Ou, F.S., Bifulco, C., Lugli, A., Zlobec, I., Rau, T.T., Berger, M.D., et al. (2018). International validation of the consensus immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet *391*, 2128–2139. https://doi.org/10.1016/S0140-6736(18)30789-X.

24. Llosa, N.J., Cruise, M., Tam, A., Wicks, E.C., Hechenbleikner, E.M., Taube, J.M., Blosser, R.L., Fan, H., Wang, H., Luber, B.S., et al. (2015). The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. Cancer Discov. *5*, 43–51. https://doi.org/10.1158/2159-8290.CD-14-0863.

25. Sun, Z., Li, G., Shang, D., Zhang, J., Ai, L., and Liu, M. (2022). Identification of microsatellite instability and immune-related prognostic biomarkers in colon adenocarcinoma. Front. Immunol. *13*, 988303. https://doi.org/10.3389/fimmu.2022.988303.

26. Wang, D.X., Long, J.Y., Li, R.Z., Zhang, D.L., Liu, H., Liu, J., Tian, J.C., Li, H., Liu, J., Zhao, H.T., et al. (2024). Mutation status of the kmt2 family associated with immune checkpoint inhibitors (icis) therapy and implicating diverse tumor microenvironments. Mol. Cancer *23*, 15. https://doi.org/10.1186/s12943-023-01930-8.

27. Gao, Z., Chen, J.F., Li, X.G., Shi, Y.H., Tang, Z., Liu, W.R., Zhang, X., Huang, A., Luo, X.M., Gao, Q., et al. (2022). Kras acting through erk signaling stabilizes pd-l1 via inhibiting autophagy pathway in intrahepatic cholangiocarcinoma. Cancer Cell Int. *22*, 128. https://doi.org/10.1186/s12935-022-02550-w.

28. Chen, N., Fang, W., Lin, Z., Peng, P., Wang, J., Zhan, J., Hong, S., Huang, J., Liu, L., Sheng, J., et al. (2017). Kras mutation-induced upregulation of pd-l1 mediates immune escape in human lung adenocarcinoma. Cancer Immunol. Immunother. *66*, 1175–1187. https://doi.org/10.1007/s00262-017-2005-z.

29. Liu, J., Chen, Z., Li, Y., Zhao, W., Wu, J., and Zhang, Z. (2021). Pd-1/pd-l1 checkpoint inhibitors in tumor immunotherapy. Front. Pharmacol. *12*, 731798. https://doi.org/10.3389/fphar.2021.731798.

30. Ham, S.W., Jeon, H.Y., Jin, X., Kim, E.J., Kim, J.K., Shin, Y.J., Lee, Y., Kim, S.H., Lee, S.Y., Seo, S., et al. (2019). Tp53 gain-of-function mutation promotes inflammation in glioblastoma. Cell Death Differ. *26*, 409–425. https://doi.org/10.1038/s41418-018-0126-3.

31. Tan, K.T., Yeh, C.N., Chang, Y.C., Cheng, J.H., Fang, W.L., Yeh, Y.C., Wang, Y.C., Hsu, D.S.S., Wu, C.E., Lai, J.I., et al. (2020). Prkdc: new biomarker and drug target for checkpoint blockade immunotherapy. J. Immunother. Cancer *8*, e000485. https://doi.org/10.1136/jitc-2019-000485.

32. Ma, W., Tan, X., Xie, Z., Yu, J., Li, P., Lin, X., Ouyang, S., Liu, Z., Hou, Q., Xie, N., et al. (2024). P53: A key target in the development of osteoarthritis. Mol. Biotechnol. *66*, 1–10. https://doi.org/10.1007/s12033-023-00736-9.

33. Wang, Y., Safi, M., Hirsch, F.R., Lu, S., Peters, S., Govindan, R., Rosell, R., Park, K., and Zhang, J.J. (2025). Immunotherapy for advanced-stage squamous cell lung cancer: the state of the art and outstanding questions. Nat. Rev. Clin. Oncol. *22*, 200–214. https://doi.org/10.1038/s41571-024-00979-8.

34. Fontana, B., Gallerani, G., Salamon, I., Pace, I., Roncarati, R., and Ferracin, M. (2023). Arid1a in cancer: Friend or foe? Front. Oncol. *13*, 1136248. https://doi.org/10.3389/fonc.2023.1136248.

35. Baguley, B.J., Skinner, T.L., Leveritt, M.D., and Wright, O.R.L. (2017). Nutrition therapy with high intensity interval training to improve prostate cancer-related fatigue in men on androgen deprivation therapy: a study protocol. BMC Cancer *17*, 1. https://doi.org/10.1186/s12885-016-3022-6.

36. He, X., Peng, Y., He, G., Ye, H., Liu, L., Zhou, Q., Shi, J., Fu, S., Wang, J., Zhou, Z., et al. (2023). Increased co-expression of pd1 and tim3 is associated with poor prognosis and immune microenvironment heterogeneity in gallbladder cancer. J. Transl. Med. *21*, 717. https://doi.org/10.1186/s12967-023-04589-3.

37. Wang, M.D., Xiang, H., Zhang, L., and Wang, C. (2022). Integration of ov6 expression and cd68+ tumor-associated macrophages with clinical features better predicts the prognosis of patients with hepatocellular carcinoma. Transl. Oncol. *25*, 101509. https://doi.org/10.1016/j.tranon.2022.101509.

38. Malek-Hosseini, Z., Jelodar, S., Talei, A., Ghaderi, A., and Doroudchi, M. (2017). Elevated syndecan-1 levels in the sera of patients with breast cancer correlate with tumor size. Breast Cancer *24*, 742–747. https://doi.org/10.1007/s12282-017-0773-0.

39. Zeng, X., Ward, S.E., Zhou, J., and Cheng, A.S.L. (2021). Liver immune microenvironment and metastasis from colorectal cancer-pathogenesis and therapeutic perspectives. Cancers *13*, 2418. https://doi.org/10.3390/cancers13102418.

40. Ciner, A.T., Jones, K., Muschel, R.J., and Brodt, P. (2021). The unique immune microenvironment of liver metastases: challenges and opportunities. In Semin. Cancer Biol., *71* (Elsevier), pp. 143–156. https://doi.org/10.1016/j.semcancer.2020.06.003.

41. Laumont, C.M., Banville, A.C., Gilardi, M., Hollern, D.P., and Nelson, B.H. (2022). Tumour-infiltrating b cells: immunological mechanisms, clinical impact and therapeutic opportunities. Nat. Rev. Cancer *22*, 414–430. https://doi.org/10.1038/s41568-022-00466-1.

42. Gong, X., Li, X., Jiang, T., Xie, H., Zhu, Z., Zhou, F., and Zhou, C. (2017). Combined radiotherapy and anti–pd-l1 antibody synergistically enhances antitumor effect in non–small cell lung cancer. J. Thorac. Oncol. *12*, 1085–1097. https://doi.org/10.1016/j.jtho.2017.04.014.

43. Huang, R., Nakamura, B., Senguttuvan, R., Li, Y.J., Martincuks, A., Bakkar, R., Song, M., Ann, D.K., Rodriguez-Rodriguez, L., and Yu, H. (2025). A critical role of intracellular pd-l1 in promoting ovarian cancer progression. Cells *14*, 314. https://doi.org/10.3390/cells14040314.

44. Zhao, T., Gu, Y., Yang, J., Usuyama, N., Lee, H.H., Kiblawi, S., Naumann, T., Gao, J., Crabtree, A., Abel, J., et al. (2025). A foundation model for joint segmentation, detection and recognition of biomedical objects across nine modalities. Nat. Methods *22*, 166–176. https://doi.org/10.1038/s41592-024-02499-w.

45. Greenwald, N.F., Miller, G., Moen, E., Kong, A., Kagel, A., Dougherty, T., Fullaway, C.C., McIntosh, B.J., Leow, K.X., Schwartz, M.S., et al. (2022). Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nat. Biotechnol. *40*, 555–565. https://doi.org/10.1038/s41587-021-01094-0.

46. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nat. Methods *18*, 100–106. https://doi.org/10.1038/s41592-020-01018-x.

47. Gatenbee, C.D., Baker, A.M., Prabhakaran, S., Swinyard, O., Slebos, R.J.C., Mandal, G., Mulholland, E., Andor, N., Marusyk, A., Leedham, S., et al. (2023). Virtual alignment of pathology image series for multi-gigapixel whole slide images. Nat. Commun. *14*, 4502. https://doi.org/10.1038/s41467-023-40218-9.

48. Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). Cell detection with star-convex polygons. In Medical Image Computing and Computer Assisted Intervention—MICCAI 2018, A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds. (Springer International Publishing), pp. 265–273. https://doi.org/10.1007/978-3-030-00934-2_30.

49. Weigert, M., and Schmidt, U. (2022). Nuclei instance segmentation and classification in histopathology images with stardist. In The IEEE International Symposium on Biomedical Imaging Challenges (ISBIC) https://doi.org/10.1109/ISBIC56247.2022.9854534.

50. Otsu, N. (1975). A threshold selection method from gray-level histograms. Automatica *11*, 23–27.

51. Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015, N. Navab, J. Hornegger, W. Wells, and A. Frangi, eds. (Springer, Cham), pp. 234–241.

52. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., et al. (2021). CHAOS Challenge—combined (CT-MR) healthy abdominal organ segmentation. Med. Image Anal. *69*, 101950. https://doi.org/10.1016/j.media.2020.101950.

53. Valanarasu, J.M.J., and Patel, V.M. (2022). Unext: MLP-based rapid medical image segmentation network. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, L. Wang, Q. Dou, P.T. Fletcher, S. Speidel, and S. Li, eds. (Springer, Cham), pp. 23–33. https://doi.org/10.1007/978-3-031-16443-9_3.

54. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (IEEE), pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

55. Cancer; Genome; Atlas; Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The cancer genome atlas pan-cancer analysis project. Nat. Genet. *45*, 1113–1120. https://doi.org/10.1038/ng.2764.

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Python 3.11	Python Software Foundation	https://www.python.org; RRID:SCR_008394
Pytorch 2.4	Pytorch Foundation	https://pytorch.org/; RRID:SCR_018536
VALIS 1.1	Gatenbee et al.[47]	https://valis.readthedocs.io/en/latest/
Scikit-Learn 1.5	Scikit-Learn Community	https://scikit-learn.org/stable/; RRID:SCR_002577
Other		
COMET Device	Lunaphore	https://lunaphore.com/products/comet/

METHOD DETAILS

Details of acquiring paired training and testing dataset for *GigaTIME*

The first step in building a virtual population for tumor microenvironment modeling is the collection of a high-quality paired dataset of H&E and mIF slides to train a robust cross-modal model. For training the *GigaTIME* cross-modal translator, we curated twenty-one matched pairs of H&E slides and generated the corresponding mIF data using the COMET multiplex immunofluorescence platform,[5] with all processing performed in-house. Whole slide images (WSIs), obtained through conventional brightfield microscopy, comprised an area on average measuring 20 mm $\times$ 20 mm (approximately $\sim$100k $\times$ 100k pixels). mIF images concentrated on a specific subregion of interest, measuring 10 mm $\times$ 10 mm (40k $\times$ 40k pixels), captured via the COMET platform. Given the high resolution and rich cellular content of the slides, we estimate approximately 40 million cells with their corresponding activation labels. The protein markers selected for this study cover a broad spectrum of TIME markers including immune function profiling and tumor characterization. These TIME markers include: a) PD-1, PD-L1, T-bet, Ki67, PHH3 and Caspase 3, which respectively represent immune checkpoint activity, T-cell differentiation, proliferation, and apoptosis; b) immune phenotypic markers such as CD3, CD4, CD8, CD14, CD16, CD11c, CD68, CD20 and CD138, identifying T-cell, myeloid, B-cells and plasma cells/ late-stage B-cell differentiation; and c) microenvironment markers such as Cytokeratin (CK), Tryptase and CD34, Transgelin, and Actin characterize epithelial cells, mast cells, endothelial cells, and smooth muscle/myofibroblast/pericytes, while DAPI serves as a pan-nuclear marker for cell identification. This selection provides a comprehensive framework for tumor microenvironment modeling: immune activation, tumor phenotypes, stromal components, and cellular dynamics, facilitating a pathology-driven, immune landscape profiling analysis (Figure S1; Table 1). Our training data comprises of data collected from 21 patients across different stages of lung adenocarcinoma. Each of these patients were chosen to make sure they had diverse biomarker activations.

Additionally, to test the generalization of *GigaTIME*, we acquired some tissues and their mIF panels for additional cancer types. We curated tumor cores from tissue microarrays (TMAs) derived from whole-slide sections, representing two additional distinct cancer types: breast and brain. These cohorts encompass a wider spectrum of histological subtypes and clinical stages. The breast set predominantly consists of infiltrating duct carcinoma (IDC), with additional metaplastic carcinoma (MBCA) and sarcomatoid carcinoma (SARC) cases, spanning stages IIA to IIIC. The brain cohort includes lower-grade astrocytoma (grades 2–3; ASTR) alongside high-grade glioblastoma multiforme (grade 4; GBM). Specifically, the tissues were collected from 12 patients for breast and 20 patients for brain cohorts. Similar to training data, we used the COMET multiplex immunofluorescence (mIF) platform to generate mIF panels for the same set of protein markers.

Details of registration and preprocessing on acquired mIF images

Since the mIF images in our *GigaTIME* training data represents only a smaller subregion of the corresponding H&E whole slides, mapping its coordinates to those within the H&E WSI is necessary. To achieve this alignment, we utilized VALIS,[47] a robust, automated registration tool developed for large-scale pathology images. VALIS supports both H&E brightfield and fluorescence images, making it suitable for our dataset. For registration, we used the DAPI marker (nuclei marker) and grayscale-converted H&E images as references. VALIS implements a multi-step registration pipeline to ensure robust alignment, including tissue region filtering, image normalization, rigid registration, non-rigid registration, and micro-registration.

VALIS first picks only the tissue regions and masks out the background so the registration process happens only on relevant areas. Rigid registration is then performed on low-resolution versions of the images to learn linear transformations for initial global alignment where the image size is set at 1024 pixels. Subsequently, non-rigid registration is performed to refine the alignment by learning local,

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

non-linear transformations using higher-resolution images of size 8192 pixels. Finally, a second non-rigid registration step, called micro-registration, is performed at a high resolution of 16384 pixels to capture micro-features that were unresolved in previous steps.

The registered mIF images are divided into patches of size 128 × 128 micrometers which equals 556 × 556 pixels, and the patch coordinates are warped into the H&E image space to identify and extract corresponding regions. This process generated an initial set of 417,501 pairs of patches. Post-tiling, we filtered out background patch pairs, defined as > 99% black pixels or channel variance < 1, reducing the dataset to 218,214 patches. While registration results from VALIS are accurate for most patches, occasional failures were observed, potentially due to severe tissue deformations, large lighting condition differences, or other artifacts.[47] To detect these failures, we performed cell segmentation using StarDist on the DAPI channel for multiplex protein imaging and on the RGB H&E images, respectively.[48,49] The nuclei masks obtained from DAPI were dilated by 7 pixels to approximate whole-cell masks. We then calculated the Dice coefficient between the masks from H&E and multiplex images to quantify how they aligned well. We manually annotated 100 pairs as either successful or unsuccessful, and using this reference, we determined a threshold Dice score of 0.2 for distinguishing failures. All patch pairs with Dice coefficients below the threshold were excluded. This filtering resulted in a final dataset of 49,401 pairs with high-quality, accurately registered patch pairs. As the mIF channels are continuous signals, we use Otsu thresholding[50] to binarize the mIF to have only values of 0 and 1. Note that each of these patches contain at least 200 cells, resulting in around 40 M overall and 10 M high-quality registered cells (post high-quality filtering), each with 21 mIF activations. We use these 10 M cells for training our *GigaTIME* cross-translator model. This entire workflow of registration and preprocessing is illustrated in Figure S2.

Details of training *GigaTIME* as cross-modal translator model

We used NestedUNet[20,51] as the *GigaTIME* cross-translator model which processes H&E patches to predict virtual mIF images across 21 TIME markers. The NestedUNet architecture follows an encoder-decoder framework. The encoder consists of multiple convolutional blocks that encode the input data into a condensed latent feature representation. The decoder then reconstructs the 21 mIF markers from this latent feature. There are five hierarchical levels of feature channels $\{32, 64, 128, 256, 512\}$, capturing increasingly abstract semantic features while progressively reducing spatial resolution by factors of two via max pooling. The encoder produces feature maps of sizes $H \times W \times 32$, $H/2 \times W/2 \times 64$, $H/4 \times W/4 \times 128$, $H/8 \times W/8 \times 256$, and $H/16 \times W/16 \times 512$ where H and W are height and width of the image which is 256 in our case. The decoder reconstructs high-resolution outputs using bilinear upsampling followed by concatenation with feature maps from both the encoder and previous decoder nodes, forming densely nested skip pathways. This design enables effective fusion of low-level spatial details with high-level semantic context. Additionally, nested skip connections between all encoder and decoder blocks facilitate rich feature transfer (see Figure S9 for ablation study). We also experimented with *GigaTIME* training using a state-of-the-art pretrained vision transformer (ViT) giant model with over a billion parameters. Despite that this larger model leads to a slight improvement in the Dice score (Figure S14), it would incur about an order of magnitude increase in compute. Since our population-scale study already required about five thousand GPU hours in total compute, we decided to not use this expensive pretrained ViT-giant model version of *GigaTIME*.

We use 12 slides for training, 4 as the development set and keep out 5 slides as our test set. We used patches of size 256× 256 to train the model. Note that these patches are sub-regions from the 128 × 128 micrometer patches we had registered. We used a combination of Dice and binary cross-entropy (BCE) loss to train the model. The Dice score is a measure of overlap between the ground truth (GT) and the prediction, which is defined as:

$$Dice_c = \frac{2\sum y_c \hat{y}_c}{\sum y_c + \sum \hat{y}_c}$$

where y is the ground truth patch of mIF and $\hat{y}$ is the predicted *GigaTIME*-translated virtual mIF for protein channel c. The multi-class Dice loss is defined as:

$$L_{Dice} = 1 - \frac{1}{C}\sum_{c=1}^{C} Dice_c$$

where C represents the total number of channels. The multi-class BCE loss is defined as:

$$L_{BCE} = \frac{1}{C}\sum_{c=1}^{C} -(y_c \log(\hat{y}_c) + (1 - y_c)\log(1 - \hat{y}_c))$$

where y is the ground truth patch of mIF and $\hat{y}$ is the *GigaTIME*-translated virtual mIF for protein channel c, and C represents total number of protein channels in our training data. The loss we used to train the model is similar to leading segmentation literature[52,53] and is defined as follows:

$$L_{model} = L_{Dice} + 0.5 * L_{BCE}$$

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

We also used augmentations like random rotation, flipping, and contrast adjustments while training. The model was trained for 250 epochs using Adam optimizer with a learning rate of 0.0001. We used 8 A100 GPUs and a batch size of 16 per GPU while training. *GigaTIME* cross-translator model is relatively small with only 9.16 M parameters aiding in faster training and inference. We also attach the training and validation curves in Figure S10 which demonstrates stable training and convergence.

Details of baseline comparisons

We used CycleGAN for baseline comparison as it is a widely recognized baseline for unpaired image-to-image translation in histopathology.[16,22] Specifically, we followed prior standard methodology,[16,22] and trained CycleGAN models[21] from scratch on the same in-house H&E and mIF training data. In particular, we trained one CycleGAN model for every protein channel individually resulting in 21 different models for our baseline. CycleGAN was used in its canonical form, without architectural modifications. We utilized the development set to determine the learning rate, ensuring stable training.

To further validate the robustness of our evaluation metrics, we also created an average activation baseline. In particular, for each protein channel we calculated the average positivity ratio from the actual mIF data. We then used this value as the sampling ratio to assign positive pixels in non-blank tissue regions.

Details of metrics to evaluate the translation performance

We used three evaluation metrics to quantify the quality of *GigaTIME*-translated mIF: Dice score, Pearson correlation, and patch-wise activation comparison. Dice score is a similarity coefficient commonly used in segmentation tasks. It calculates the ratio of overlap between prediction and ground truth with that of the sum of the individual number of prediction and ground truth pixels. The Dice score is calculated between the prediction and the ground truth, and we report the mean Dice across all the patches of the test set for all the 21 protein channels. During the evaluation, we skipped all blank regions at the 8×8 level to ensure that they didn't artificially inflate the scores.

Next, we determine the activation counts for each window creating an activation matrix for a single patch and then compute the Pearson correlation between the *GigaTIME*-translated and actual mIF activation matrices. We report the mean Pearson for each TIME marker across all the patches in the test set. By counting the number of activations in each window, we effectively apply a form of local averaging. This approach can reduce the influence of small-scale noise and highlight larger-scale patterns at the cell level. Consequently, the resulting correlation may improve if the predictions accurately capture the overall structure of the ground truth, even when small pixel-level misalignments exist. This level of precision can be sufficient for specific pathology studies, as in some cases, simply identifying the activation of a particular marker is adequate. On the other hand, the Dice metric is sensitive to the size of objects in the masks, meaning larger objects contribute more significantly to the score. As a result, smaller activations, commonly observed in mIF, may be underrepresented or overlooked. We also use a performance metric to quantify the density of mIF at the patch level. To do this, we calculate the activation density for a specific channel for both *GigaTIME*-translated virtual mIF and ground truth mIF within each corresponding patch (256×256) and report the spearman correlation.

Details of Providence data for generating virtual population

To unleash *GigaTIME* for population-scale TIME modeling, it is essential to first assemble a diverse and richly annotated dataset that includes biomarkers, H&E slides, and comprehensive patient information. Our Providence data , curated from real-world patient data collected across 51 Providence Health hospitals and over 1,000 clinics in the USA meets this need. It includes H&E slides from 14,256 patients spanning multiple cancer types, each paired with corresponding genomic sequencing data, offering a unique and large-scale resource for TIME modeling clinical insight generation. For patients with multiple H&E slides, we select the most tumor-representative slide using a ResNet model[54] trained to classify whether a slide contains a tumor. The ResNet is trained with a binary cross-entropy loss, leveraging pathology reports and immuno-histochemistry (IHC) orders to identify tumor-positive slides. Tumor-negative slides are selected from patients with multiple slides by choosing those without corresponding IHC tests. While these labels are not gold-standard, we validated them through spot checks by expert pathologists to ensure quality. The trained ResNet model was also evaluated on a test set of 200 slides (100 positive, 100 negative), achieving 87% accuracy, demonstrating strong performance. We then applied this ResNet model to score all H&E slides from 14,256 patients, selecting the slide with the highest tumor score as the representative slide for further analysis. Additional patient data, including OncoTree ID, survival time, censorship information, and staging attributes, were extracted from electronic health records and reports. Since Providence data is not curated based on specific selection criteria, it closely emulates population-scale real-world data, capturing a diverse cancer landscape with 24 cancer types and 306 unique OncoTree IDs, reflecting 306 cancer subtypes (Figure S3). All genomics biomarkers were captured using a single 523-gene comprehensive genomic profiling (CGP) assay developed in-house at Providence (Portland, OR). The assay uses hybrid capture technology based on TruSight Oncology 500 Illumina (Illumina, Inc, SanDiego, CA) kits and sequencing on a NovaSeq 6000 platform. The assay captures and sequences coding regions of 523 genes, includes RNA-seq of 23 genes to identify fusions, and provides tumor mutational burden (TMB) and microsatellite instability (MSI) assessment. Analysis was performed using the TruSight Oncology 500 application with STAR-Fusion. All cases were interpreted by our expert team following AMP/CAP/ASCO guidelines, utilizing resources like OncoKB, ClinVar, gnomAD, and COSMIC databases. The assay is validated according to CAP/CLIA standards.

Please cite this article in press as: Valanarasu et al., Multimodal AI generates virtual population for tumor microenvironment modeling, Cell (2026), https://doi.org/10.1016/j.cell.2025.11.016

Details of generating virtual population data

To generate virtual mIF whole-slide images using *GigaTIME* for any provided H&E slide, we first tile the whole slide image into patches of 128 × 128 micrometers which translates to 256 × 256 pixels. We perform inference using our *GigaTIME* cross-translator model across all of these patches but also consider overlapping patches to avoid any edge artifacts. We use a sliding window technique with a stride size of 128 pixels across both height and width and keep the patches of resolution as 256 × 256. Then, we use our model to predict virtual mIF for all these patches. We then stitch back the patches together to obtain the slide-level virtual mIF. We conduct inference using 325 V100 GPUs, with each GPU processing a single slide. Please note that the inference time significantly depends on the size of the slide, and it can vary considerably but on average it can take anytime between 20 minutes to 2 hours for a single slide processed in one GPU. We set the batch size (number of patches processed in each GPU) to 12. This large-scale inference results in 299,376 virtual mIF whole-slide images across 21 protein channels unlocking population-scale TIME modeling.

Details of biomarker discovery using virtual population

Access to a rich combination of genomic biomarkers, clinical attributes, survival information, and TIME modeling using *GigaTIME* across a diverse patient population enables clinical discovery. By identifying associations between protein channels in the *GigaTIME*-translated virtual mIF and various biomarkers or clinical attributes, we aim to uncover novel insights that could deepen our understanding of cancer biology and potentially inform future diagnostic and therapeutic strategies. To this end, we analyzed the association between the protein channels in virtual mIF and various biomarkers derived from using genomics sequencing: PD-L1, TP53, LRP1B, KRAS, APC, KMT2D, FAT1, SPTA1, ZFHX3, KMT2C, EGFR, ARID1A, PIK3CA, PRKDC, NOTCH1, ATM, KMT2A, ROS1, TMB, and MSI. Each of these biomarkers offer insights into prognosis, treatment response, and tumor biology—ranging from immune evasion (PD-L1), genomic instability (TP53, ATM), oncogenic signaling (KRAS, EGFR) and epigenetic regulation (KMT2D/KMT2A/KMT2C, ARID1A). Using a two-sided t-test, we assessed the correlations between the individual protein channels in *GigaTIME* and the biomarkers and determined the directionality by comparing category-specific average scores. To ensure statistical robustness, we applied Benjamini-Hochberg (BH) correction for multiple testing, reducing the false discovery rate and enhancing result reliability.

Details of pathological staging and surival analysis using virtual population

We also assess the associations among clinical staging attributes, including PathStage T (tumor size and extent), N (lymph node involvement), M (distant metastasis), and Group (overall stage classification). PathStage T ranges from T1 (small, localized tumor) to T4 (extensive tumor invasion). PathStage N reflects nodal spread, from N0 (no involvement) to N3 (extensive lymph node metastasis). PathStage M classifies distant metastasis as M0 (none) or M1 (present). The Group stage integrates these factors into a comprehensive classification (I–IV), guiding prognosis and treatment decisions. For clinical staging attributes with more than two categories, we discretized the predicted scores into multiple bins with equal sample sizes, computed the contingency table, and applied the Chi-squared test for statistical significance. We also use BH correction to ensure statistical robustness. We calculated Pearson correlation coefficients to find the directionality of the associations between protein channels and clinical staging attributes. To visualize the correlations, we construct a grid chart mapping the relationships between biomarkers, staging clinical attributes, and the virtual mIF in *GigaTIME*.

Access to survival information in our Providence data enables us to perform comprehensive survival analysis using the *GigaTIME* framework. This allows us to evaluate the prognostic relevance of virtual mIF derived using *GigaTIME* and uncover associations between protein channels and patient outcomes. We first extract the overall survival time of the patients, censorship data and also use the oncotree ID to stratify the patients based on cancer type. This results in a subset of 7,219 patients for whom all necessary data is available. We then conduct survival analysis for each individual protein channel in *GigaTIME* at pan-cancer level. To do this, we first stratify the patient cohort into two groups—high and low activation—based on the median activation density of each channel. For each group, we applied the Kaplan-Meier estimator to model survival probability over time and fitted a Cox proportional hazards model to compute the hazard ratio, quantifying the risk of events between groups. To assess statistical significance, we performed a log-rank test to determine whether there were significant differences between the Kaplan-Meier (KM) survival curves of the two groups. For performing survival analysis on the entire 21 protein channels from *GigaTIME*, we first create a 21 feature vector which we term as *GigaTIME* signature. We then apply K-means clustering to this feature vector to group patients into distinct clusters. For direct comparison with KM curves on individual protein channels, we set k as 2 but also explore stratification beyond binary grouping by varying k from 2 to 7 (Figure S6). Given the presence of multiple cancer types in our dataset, we also perform cancer type–specific survival analysis. This allows us to assess the prognostic relevance within individual cancer types, capturing disease-specific patterns and survival trends in a tumor-context-aware manner.

Details of the TCGA-based virtual population

We make use of The Cancer Genome Atlas Program (TCGA) dataset[55] to validate the results we obtain using *GigaTIME*. TCGA data contains 28,402 H&E slides from 10,200 patients. As not all of these slides are tumor slides, we use the TCGA barcode to identify slides which are samples of solid tumors and only perform our analysis on those tumor slides. We use our *GigaTIME* cross-translator model to obtain the virtual mIF across all TIME markers. We then use the biomarker information present in TCGA dataset to perform association analysis. Similar to our analysis in the Providence data, we used a two-sided t-test, where correlations between the

individual protein channels and the biomarkers were assessed. We determined the directionality by comparing category-specific average scores. To maintain consistency and ensure statistical robustness, we applied Benjamini-Hochberg (BH) correction. The biomarkers present in TCGA and Providence differ and do not completely overlap; therefore, our analysis includes only the biomarkers common to both datasets: KRAS, LRP1B, KMT2D, TP53, FAT1, PIK3CA, ROS1, KMT2C, ARID1A, EGFR, KMT2A, SPTA1, PRKDC, APC, and ZFHX3.

Details of the spatially informed metrics

We use spatial features such as Signal-to-Noise Ratio (SNR), entropy, and sharpness to provide deeper insights into image structure and quality compared to simple features like activation density. SNR quantifies the strength of meaningful signal compared to noise, providing a measure of image clarity. Considering each pixel of a binary image be a Bernoulli random variable $X \in \{0,1\}$ with $P(X = 1) = p$ and $P(X = 0) = 1 - \mathrm{p}$ and treating the mean pixel value as the "signal" and the pixel-wise fluctuations about the mean as "noise", the signal–to–noise ratio (SNR) can be written as the ratio of the mean to the standard deviation:

$$SNR = \frac{\mu}{\sigma} = \frac{p}{\sqrt{p(1 - \mathrm{p})}}$$

The SNR tends to infinity when the image is uniform ($p = 0$ or $p = 1$) and is minimized when the image has a nuanced composition of 0s and 1s ($p = \frac{2}{2}$). Entropy measures the information content of an image based on the distribution of pixel intensities. It is calculated as:

$$H = -\sum_{i=1}^{N} p(x_i)\mathrm{logp}(x_i)$$

where $p(x_i)$ is the probability of pixel intensity x_i occurring in the image, N is the number of possible intensity levels. Higher entropy indicates a complex and diverse texture, which is useful for detecting heterogeneous regions in biological images, while lower entropy suggests more uniform, redundant regions. Sharpness quantifies the clarity of edges and fine structures in an image, which is crucial for distinguishing important boundaries. It is computed using the gradient magnitude, given by:

$$S = \sum_{x,y} \left|\frac{\partial I}{\partial x}\right| + \left|\frac{\partial I}{\partial y}\right|$$

where $I(x,y)$ is the pixel intensity at location (x,y), $\frac{\partial I}{\partial x}$ and $\frac{\partial I}{\partial y}$ represent the intensity gradients in the horizontal and vertical directions. A sharper image contains more high-frequency details, while a blurry image exhibits weaker gradients.

Details of the combinatorial analysis

As we generate virtual mIF whole-slide maps for individual protein channels within the *GigaTIME* framework, it opens up new possibilities to explore the combinatorial effects of multiple markers. For the combinatorial analysis, we apply simple logical operator OR to individual protein channels of the virtual mIF data in *GigaTIME* to create composite activation maps. We then compute the activation density from these combined channels to perform a t-test for assessing statistical significance. This approach enables us to evaluate whether combining multiple protein channels yields stronger and more meaningful associations with biomarkers and clinical attributes.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was conducted in Python 3.11 using machine learning libraries like scikit-learn. Our reported results in terms of dice score and pearson correlation includes the standard deviation as error bars in the plots Figures 2A and 2B. Also, statistical significance across models was assessed using two-tailed t-tests, comparing GigaTIME with baselines for each biomarker. Significance levels were indicated as follows: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Similarly, in our protein channel - biomarker/staging association tables found through Figures 3B–3G, 4A–4C, and 5A–5C, we report p values from a two-sided t test to assess the correlation between the individual protein channel and biomarker/staging. We also applied a Benjamini-Hochberg (BH) correction for multiple testing. The survival analysis was done with the packages lifelines, scikit-survival, and statsmodels. In Figures 4D–4G, survival probabilities were estimated using the Kaplan–Meier estimator, and differences between groups were evaluated using the log-rank test. Hazard ratios (HRs) were computed using the Cox proportional hazards model.

ADDITIONAL RESOURCES

Our code and model can be found at https://aka.ms/gigatime_code.

Supplemental figures

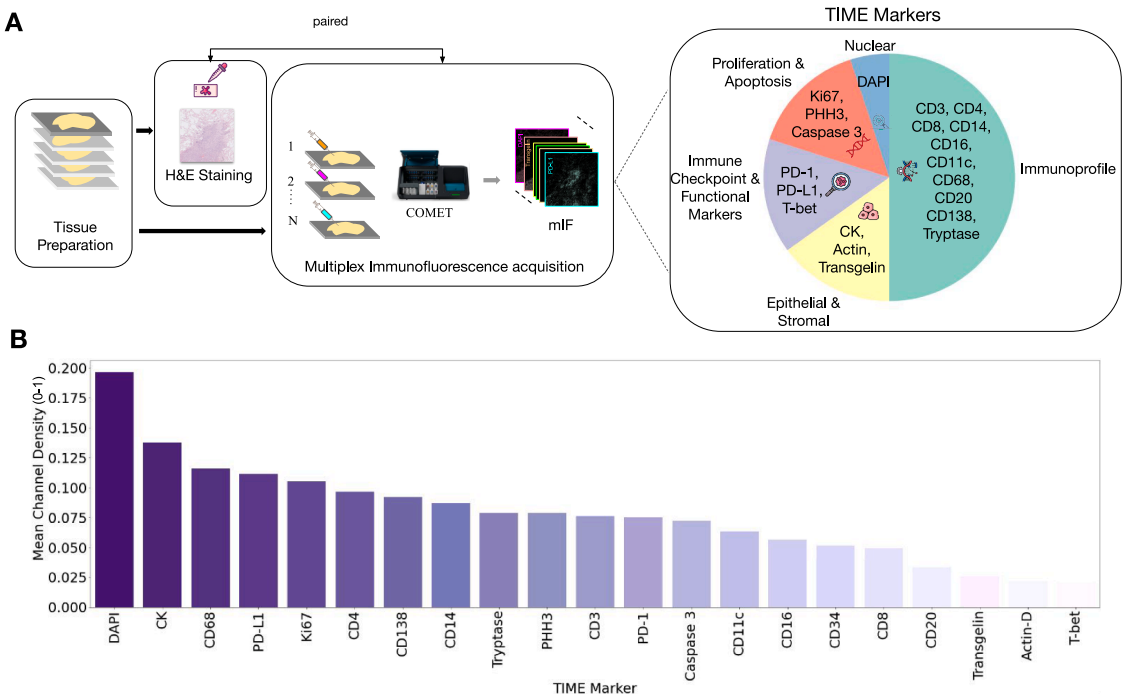


Figure S1. Data acquisition and channel distribution in *GigaTIME* training data, related to Figure 1A

(A) Flowchart illustrating the pipeline for obtaining paired H&E slides and mIF data in the *GigaTIME* training data, including the acquisition process using Lunaphore COMET hyperplexing device for various TIME markers.

(B) Bar chart displaying the mean channel intensity (scaled between 0 and 1) for each TIME marker in the training dataset. This visualization provides an overview of the activation distribution across all protein channels, highlighting relative expression levels and variability.

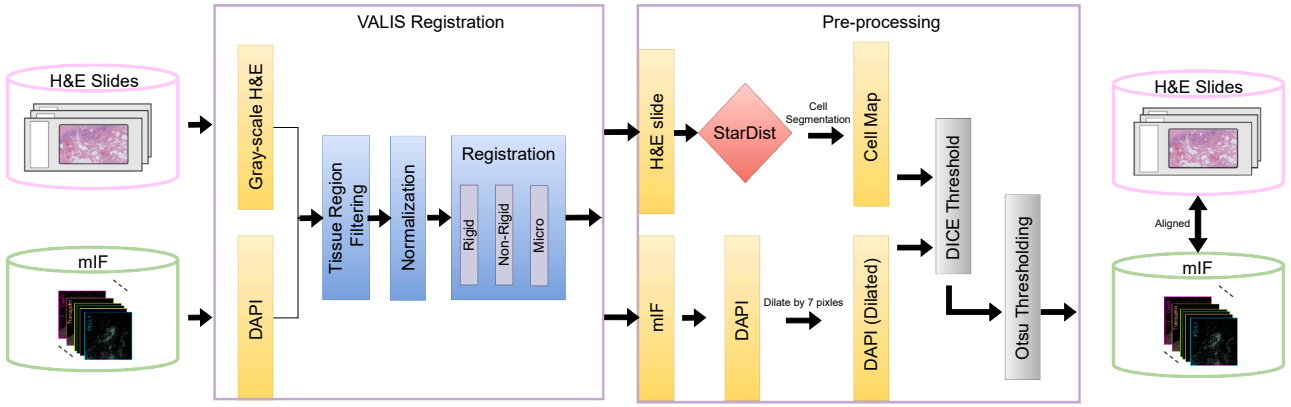


Figure S2. Registration and pre-processing workflow for *GigaTIME* training data, related to Figure 1

Flowchart illustrating the alignment process of H&E and mIF data from the *GigaTIME* training dataset. First, the slides are registered using VALIS. Following registration, pre-processing is performed by computing the Dice threshold similarity between the DAPI channel of the mIF image and the cell map extracted from the H&E slide using StarDist to get the pixel aligned pairs of H&E slides and mIF images.

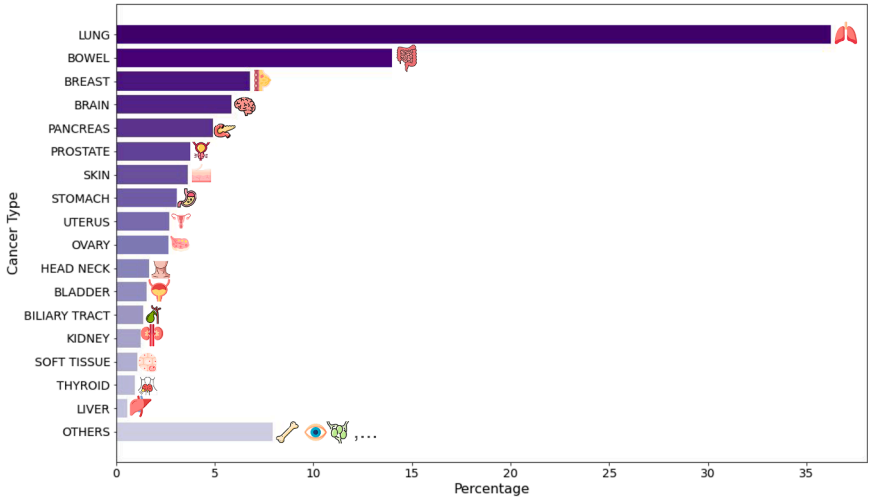


Figure S3. Providence data distribution by cancer type, related to Figure 1

Horizontal bar chart showing the percentage distribution of different cancer types found in our Providence data sorted in descending order by their occurence.

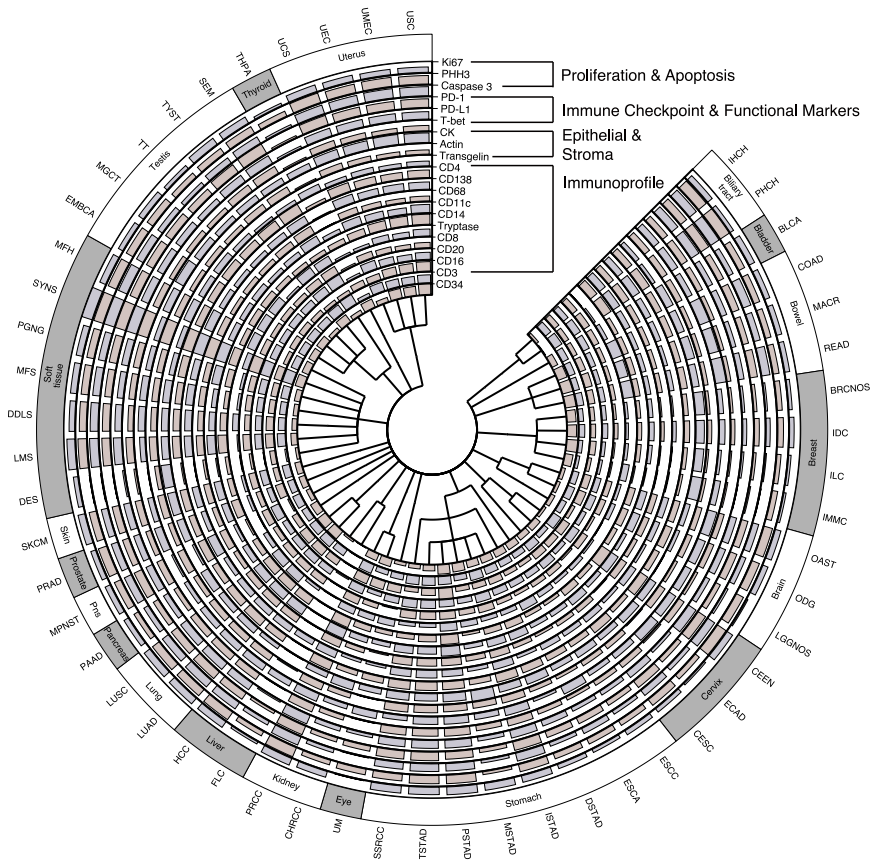


Figure S4. Circular plot for TCGA-based virtual population based on *GigaTIME* activations, related to Figure 1

Circular plot visualizing a TIME spectrum encompassing the activation scores of 21 protein channels from the TCGA-based virtual population, where each channel is represented as an individual circular bar chart segment. The inner circle encodes OncoTree, which classifies 10,200 patients into 52 cancer subtypes. The outer circle groups these activations by cancer types, allowing visual comparison across major categories.

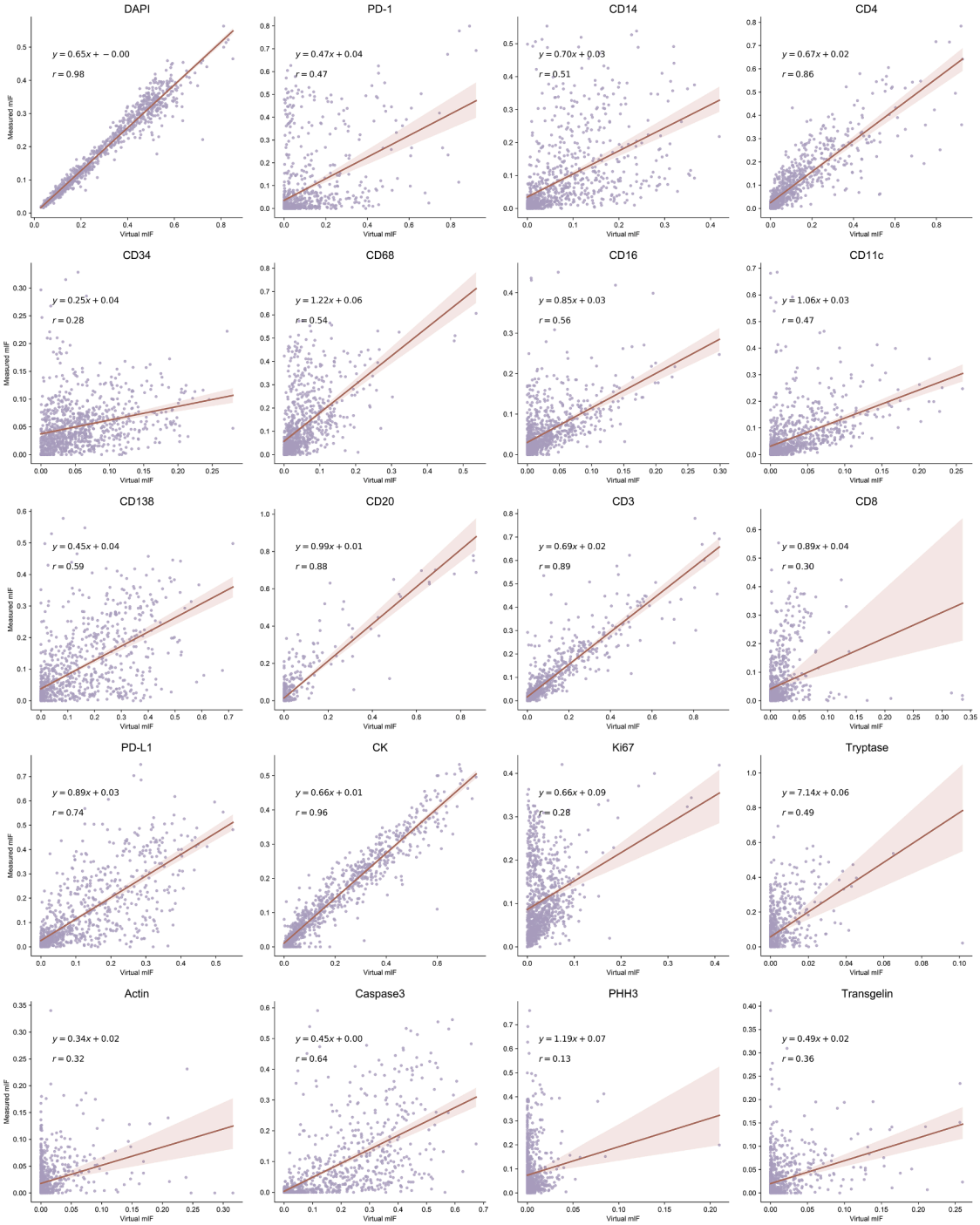


Figure S5. Scatter plots for all mIF channels in *GigaTIME*, related to Figure 2

Scatter plots comparing the activation density of the *GigaTIME*-translated mIF and the ground-truth-measured mIF across all the channels on the test set.

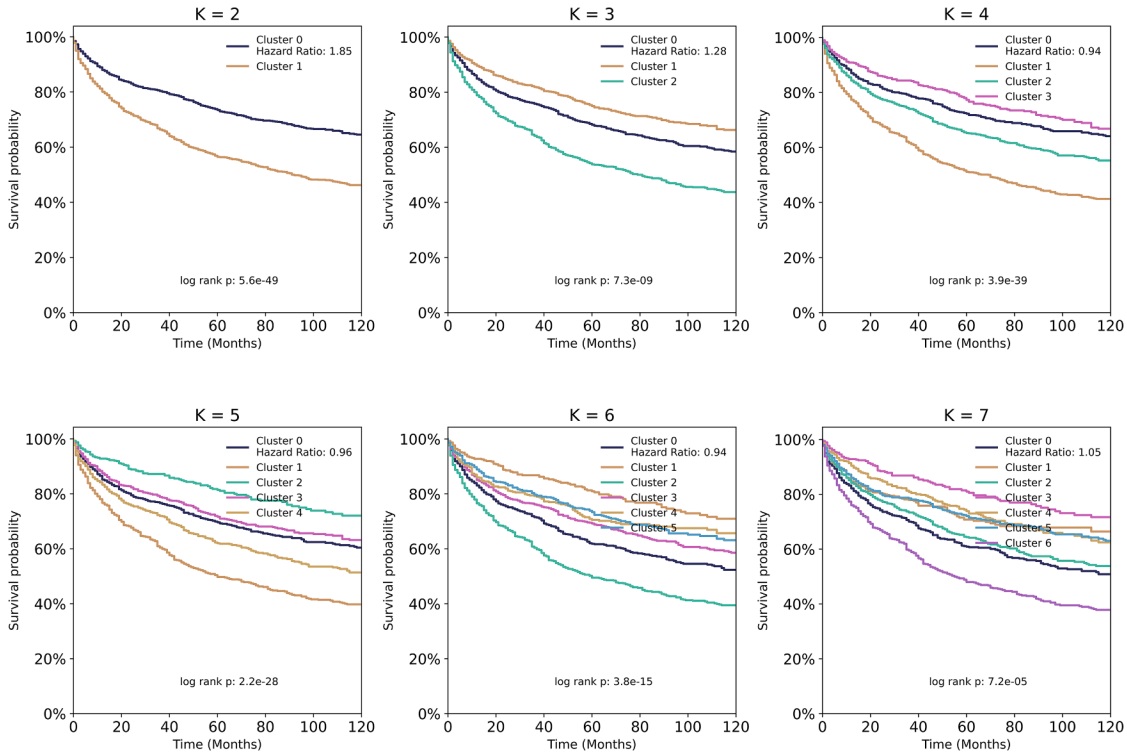


Figure S6. Survival analysis of *GigaTIME* signature stratifying patients into different numbers of clusters, related to Figure 4

Each subplot presents Kaplan-Meier (KM) survival curves based on the 21-dimensional *GigaTIME* signature, clustered using K-means into varying numbers of groups (from K = 2 to K = 7) along with hazard ratios and log-rank *p* values.

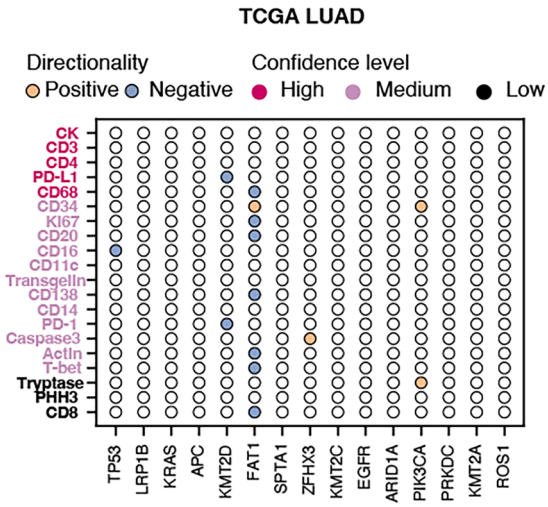


Figure S7. Grid chart showing TIME protein vs. biomarker associations for TCGA LUAD cancer patients, related to Figure 5

Correlation analysis between protein channels in virtual mIF and patient biomarkers reveal TIME protein-biomarker associations at LUAD cancer-subtype level for the TCGA patient cohort. Circle color denotes the directionality in which the correlation occurs.

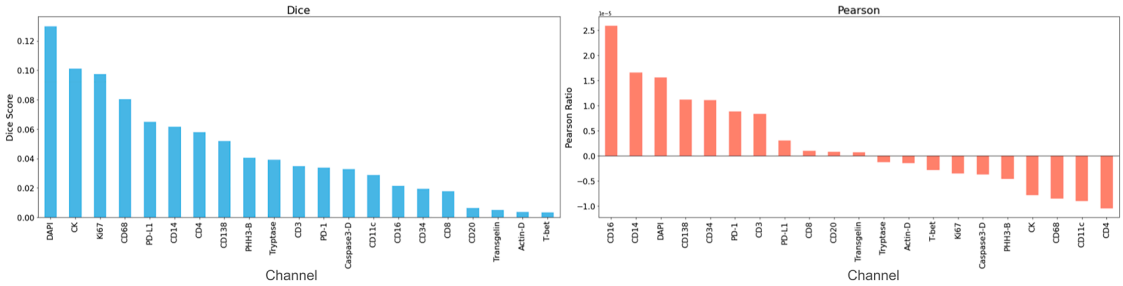


Figure S8. Average activation baseline comparison, related to Figure 2

Bar charts illustrating the Dice score (left) and Pearson correlation (right) across all protein channels involved in our study for the average activation baseline. The positivity ratio was fixed to match the average channel-wise positivity ratio observed in the measured mIF data.

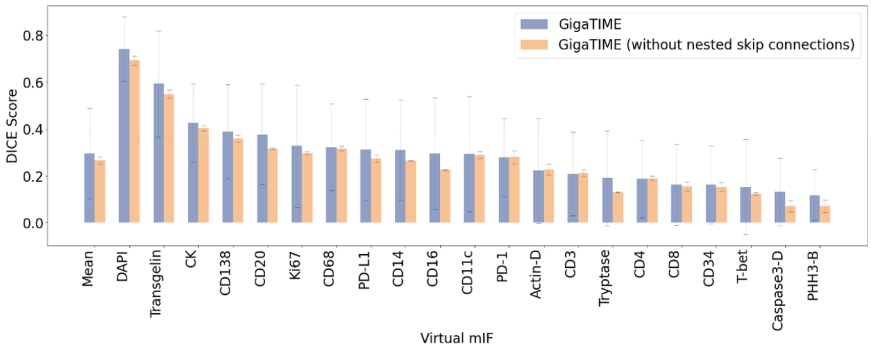


Figure S9. Ablation study of *GigaTIME* with and without nested skip connections, related to Figure 2 and STAR Methods

Bar chart illustrating the comparison of Dice scores of *GigaTIME* trained with and without nested skip connections across all the protein channels involved in our study. Inclusion of skip connections consistently improves performance across all channels.

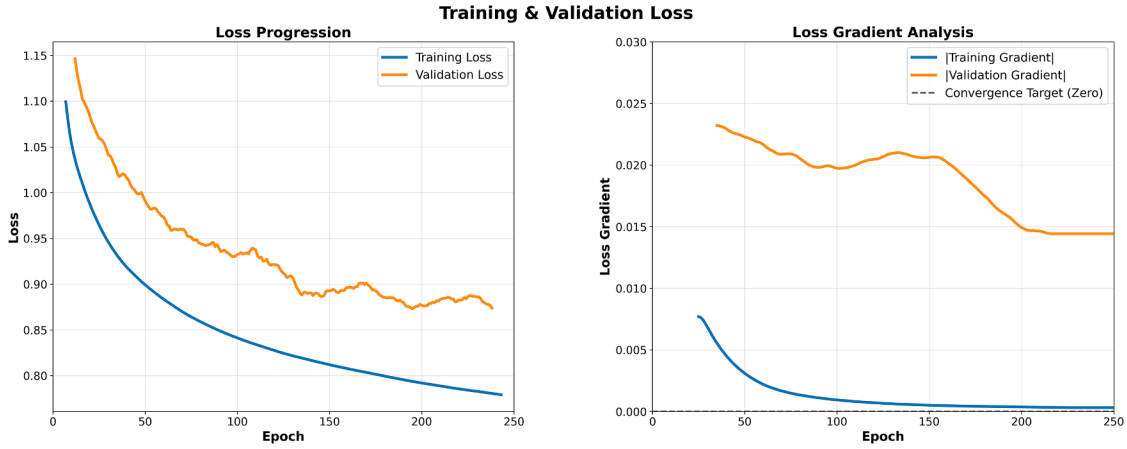


Figure S10. Training and validation loss analysis for *GigaTIME* cross-modal translator, related to Figure 2 and STAR Methods

Loss curves showing the training and validation curves while training *GigaTIME*. Blue and orange curves represent training and validation curves respectively. The left-hand side shows loss progression, while the right side shows loss gradient.

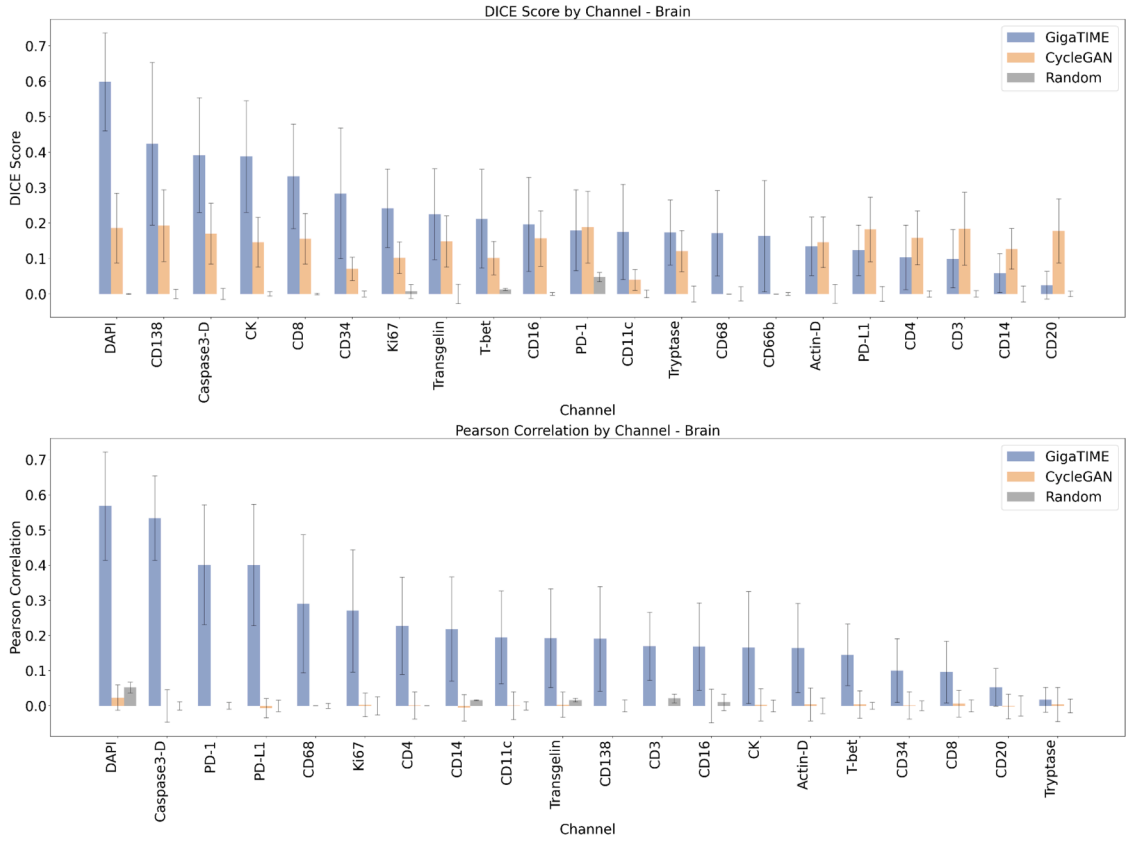


Figure S11. *GigaTIME* performance on brain tissue cohort based on Dice scores and Pearson ratios, related to Figure 2

Bar charts showing the performance of *GigaTIME* using Dice score (top) and Pearson correlation (bottom) across all the channels involved in our study for brain tissue cohort. We show the performance for *GigaTIME* and CycleGAN as well as average activation baseline.

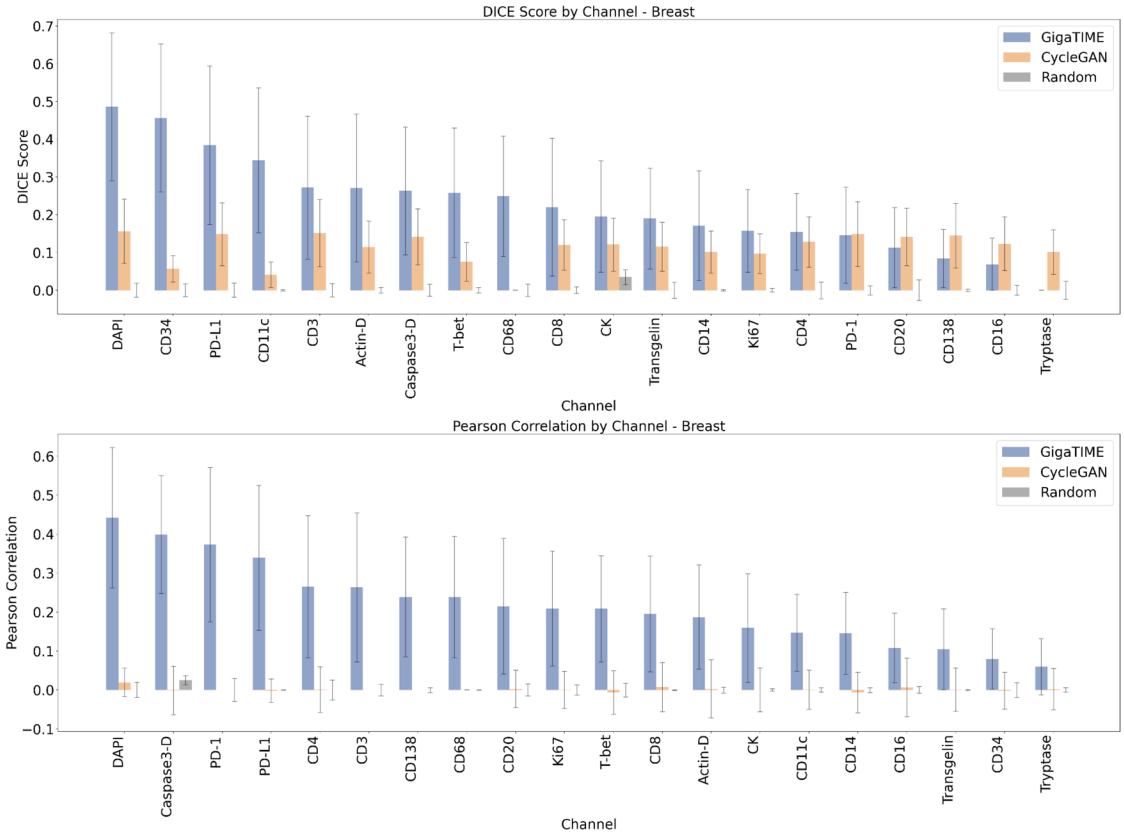


Figure S12. *GigaTIME* performance on breast tissue cohort based on Dice scores and Pearson ratios, related to Figure 2

Bar charts showing the performance of *GigaTIME* using Dice score (top) and Pearson correlation (bottom) across all the channels involved in our study for breast tissue cohort. We show the performance for *GigaTIME* and CycleGAN as well as average activation baseline.

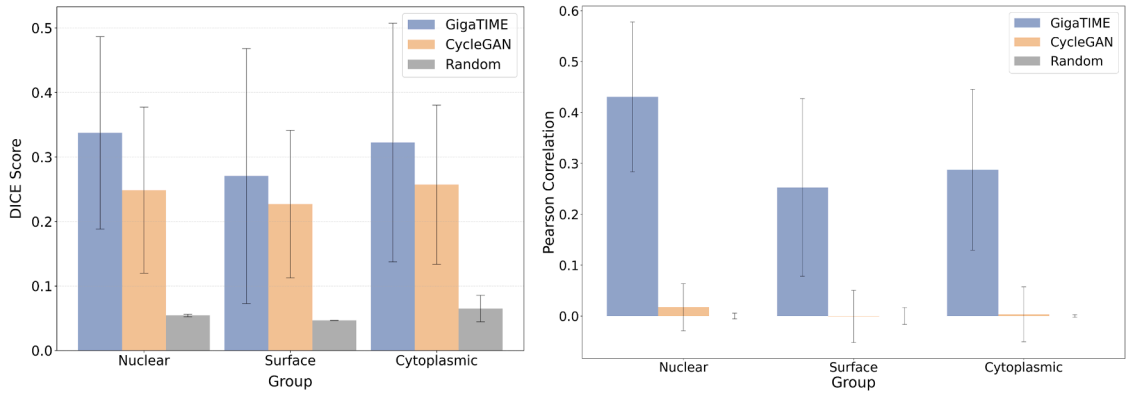


Figure S13. Performance stratified across nuclear, surface, and cytoplasmic markers, related to Figure 2

Bar charts show the performance of *GigaTIME*, CycleGAN, and average activation baseline using Dice scores (left) and Pearson correlation (right) across markers stratified by subcellular localization into nuclear, surface, and cytoplasmic markers.

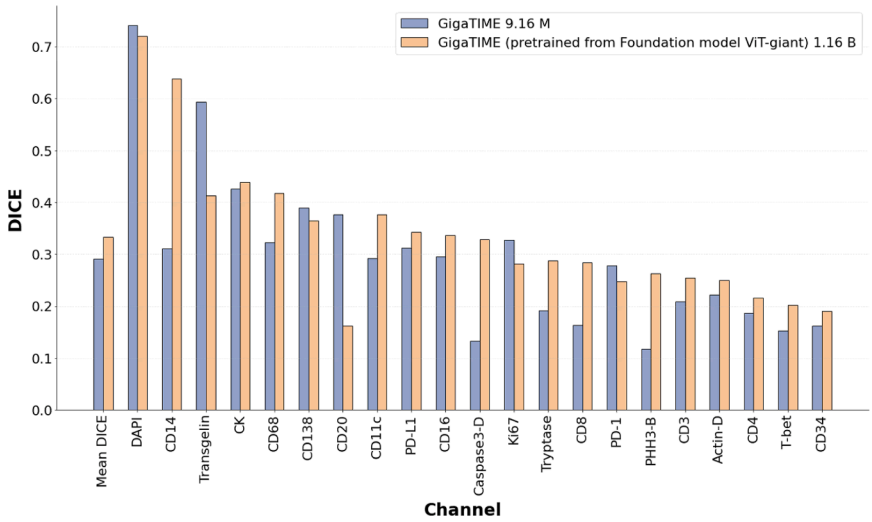


Figure S14. Comparison of *GigaTIME* with a translator built on top of a pretrained model, related to Figure 2 and STAR Methods

Bar chart comparison in terms of Dice score between *GigaTIME* with a nested UNet backbone and a ViT-giant backbone pretrained from a foundation model across all markers.[17]