

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| ☐ | ☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ A description of all covariates tested |
| ☒ | ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	The packages used for data collection and processing were Python v3.10.12(https://www.python.org/), openslide-python v1.3.1 (https://github.com/openslide/openslide-python), Pillow v10.3.0 (https://github.com/python-pillow/Pillow), and scikit-learn v1.5.0 (https://github.com/scikit-learn/scikit-learn) for WSI processing.
-----------------	--

Data analysis

Our codes are available at <https://github.com/lilab-stanford/MUSK>. We used Python (version 3.10.12) for all experiments and analyses in the study. For pretraining, we utilized 64 NVIDIA V100 GPUs each with 32-GB memory using PyTorch (version 2.0.1, CUDA 11.0) and the Accelerator from the accelerate library (<https://huggingface.co/docs/accelerate/en/index>). This setup enabled the PyTorch code to run across any distributed configuration for multi-GPU, task-agnostic pretraining. The first pretraining step with masked modeling took about 240 hours using 64 V100 GPUs. The second pretraining step with contrastive learning took about 10 hours using 64 V100 GPUs. All downstream experiments were conducted on a single 48-GB NVIDIA RTX6000 GPU. The multimodal transformer is based on implementations from torchscale (<https://github.com/microsoft/torchscale>) and the BEiT3 library (<https://github.com/microsoft/unilm/tree/master/beit3>). The evaluation of downstream tasks, including VQA, cross-modal retrieval, image-to-image retrieval, and image classification (both zero-shot and supervised), was conducted within the unified framework CLIP Benchmark (https://github.com/LAION-AI/CLIP_benchmark). We employed Hugging Face's Transformers (v4.38.1) and Torchmetrics (v0.7.3) for evaluation metrics (e.g., accuracy, AUC). The clinical tasks, including melanoma relapse prediction, prognosis, and immunotherapy prediction, were built on PyTorch Lightning (v2.2.1) (<https://lightning.ai/docs/pytorch/stable/>), allowing for fast and scalable evaluation with easy configuration. OpenSlide (v4.3.1) and openslide-python (v1.3.1) facilitated WSI processing, while Scikit-learn (v1.5.0) provided machine learning model evaluation metrics and logistic regression models. Numpy (v1.26.4), Pandas (v2.2.2), Pillow (v10.3.0), Opencv-python (v4.9.0.80), Matplotlib (v3.9.0), and Seaborn (v0.13.2) were utilized for data handling, image processing, and visualization. Implementations of other visual-language models like PLIP, BiomedCLIP, OpenAICLIP, and Quilt1M were sourced from the Hugging Face model hub (<https://huggingface.co/models>).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The histopathology and clinical data for TCGA used in this study are publicly available at The National Cancer Institute Genomic Data Commons Portal: <https://portal.gdc.cancer.gov/>. The cBioPortal: <https://www.cbioportal.org/>. The pathology report for TCGA can be found at <https://github.com/tatonetti-lab/tcga-path-reports>. Quilt1M is available at <https://github.com/wisdomizeogwo/quilt1m>. PathAsst is available at <https://huggingface.co/datasets/jamessy/PathCap>. PathVQA is available at <https://huggingface.co/datasets/flaviagiammarino/path-vqa>, BookSet is available at https://warwick.ac.uk/fac/cross_fac/tia/data/arch, and PubmedSet is available at https://warwick.ac.uk/fac/cross_fac/tia/data/arch. PatchCamelyon can be accessed at <https://patchcamelyon.grand-challenge.org/>. The NCT-CRC-HE-100K dataset is available at <https://zenodo.org/record/1214456>. SICAPv2 can be downloaded from <https://www.biicancer.com/sicapv2-dataset>. The Osteo dataset is available at <https://zenodo.org/record/3626656>. RenalCell can be downloaded from <https://www.kaggle.com/mnassrib/renal-cell-carcinoma-rcc>. SkinCancer is accessible at <https://www.isic-archive.com/>. The LC25000 dataset is available for download at <https://www.kaggle.com/andrewmvd/lung-and-colon-cancer-histopathological-images>. PanNuke can be accessed at https://warwick.ac.uk/fac/cross_fac/tia/data/pannuke. The UniToPatho dataset is available at <https://zenodo.org/record/4295977>. WSSS4LUAD can be downloaded from <https://github.com/sjin010/WSSS4LUAD>. The BRACS datasets for 3 and 6 classes are available for download at <https://www.bracs.icar.cnr.it/download/>. The Visiomel data base (consortium SFP, SFD, CNPath) can be requested from the respective data science challenge organizers <https://www.drivendata.org/competitions/148/visiomel-melanoma/>. PathMMU is available at <https://huggingface.co/datasets/jamessy/PathMMU>. BCNB is available at <https://bcnb.grand-challenge.org/>. MUV-IDH is available at <https://doi.org/10.25493/WQ48-ZGX>. The data for patients in the immunotherapy cohorts are subject to controlled access since they contain sensitive information about patient privacy. Requests for access may be submitted to the corresponding author (rli2@stanford.edu) with a brief research plan and require consent to a data use agreement. We aim to respond to all data access requests within 4 weeks. Data usage is restricted to non-commercial academic research purposes.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

We used the term sex to denote the biological attribute. Our findings apply to both sexes. Sex is considered in the study design and was determined based on self-reporting in the electronic health records. We reported the aggregated results in all sexes because the reviewing pathologists are blinded to patients' sex when reviewing their pathology samples for this research project.

Reporting on race, ethnicity, or other socially relevant groupings

No covariates regarding race, ethnicity, or other social groupings were used or analyzed in this study.

Population characteristics

The population characteristics of the study participants are provided in Extended Data Figure 5, Supplementary Fig. 8, Supplementary Fig. 9, Supplementary Table 11 and 12.

Recruitment

We retrospectively collected clinical data and pathology slides for previously treated patients at Stanford University Medical Center. The inclusion criteria were advanced (metastatic or recurrent) non-small cell lung cancer or gastroesophageal cancer (originating from the esophagus, gastroesophageal junction, or stomach) treated with anti-PD1 or anti-PD-L1 immune checkpoint blockade between 2018 and 2023, with an available H&E-stained tissue section from a pre-treatment core needle or surgical tumor biopsy. Patients meeting the inclusion criteria were identified using the STAnford Research Repository. Informed consent was waived for this retrospective analysis.

Ethics oversight

The study was approved by the Institutional Review Board of Stanford University (protocol #67432).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A formal sample size calculation was not conducted in this study. Instead, sample size was determined empirically by using all available samples in a given benchmark dataset or by collecting all available data according to the pre-specified inclusion/exclusion criteria for a patient cohort. The data points in each benchmark dataset or clinical cohort vary from 100 to several thousands, which provided sufficient statistical power.
Data exclusions	All data points with available pathology images, pathology reports (when relevant), and clinical information were used in this study. We identified regions with insufficient tissue as background areas and filtered them out during pretraining and fine-tuning.
Replication	Findings were reproduced by rerunning the analyses with the same settings. All metrics reported were reproducible. The replication of experiment results was confirmed through 5-fold cross-validation or 10 independent experiments.
Randomization	All patients in the development cohorts were divided randomly into training, validation sets.
Blinding	Investigators were blinded to the labels of the test sets before the models were finalized.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study	n/a	Involved in the study
☒	☐ Antibodies	☒	☐ ChIP-seq
☒	☐ Eukaryotic cell lines	☒	☐ Flow cytometry
☒	☐ Palaeontology and archaeology	☒	☐ MRI-based neuroimaging
☒	☐ Animals and other organisms		
☒	☐ Clinical data		
☒	☐ Dual use research of concern		
☒	☐ Plants		

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A