OmniSurv: A Six-Modality Generalist Model for Zero-Shot Cancer Survival Prediction and Molecular Profiling

Mobareji Abejide

Noblesville High School, Noblesville, Indiana

Motivation / Problem

Modern cancer AI usually depends on many disparate task-specific labeled datasets, supervised models, and validation pipelines for different problems (eg. mutation prediction, subtype classification)

Many published models are evaluated on cancers that also appear in training -> Strong performance may reflect memorizing cancer-specific patterns rather than learning the shared biological mechanisms that make tumors aggressive.

Survival offers a different possibility. Unlike specialized labels, survival outcomes are relatively common in research datasets and reflect the combined effects of many biological processes, including driver mutations, genomic instability, immune state, and therapy resistance.

Research Questions and Hypotheses

Q1: Can survival act as a global supervision signal for learning transferable tumor biology leveraging deep learning?

Q2: Can a model trained on some cancers generalize to entirely unseen cancers without being told the cancer type at inference?

Q3: Does the learned embedding capture broader biology, such as biomarkers, immune state, and latent subtypes, even without direct supervision on those tasks?

Hypothesis

Because survival reflects many underlying biological processes related to tumor aggressiveness, a multimodal model trained on survival should learn a general biological embedding that:

- **supports zero-shot survival prediction in unseen cancers,**
- **preserves meaningful molecular and clinical structure,**
- **and can be reused for downstream oncology tasks with simple probes.**

OmniSurv: Multimodal Survival-Supervised Representation Learning

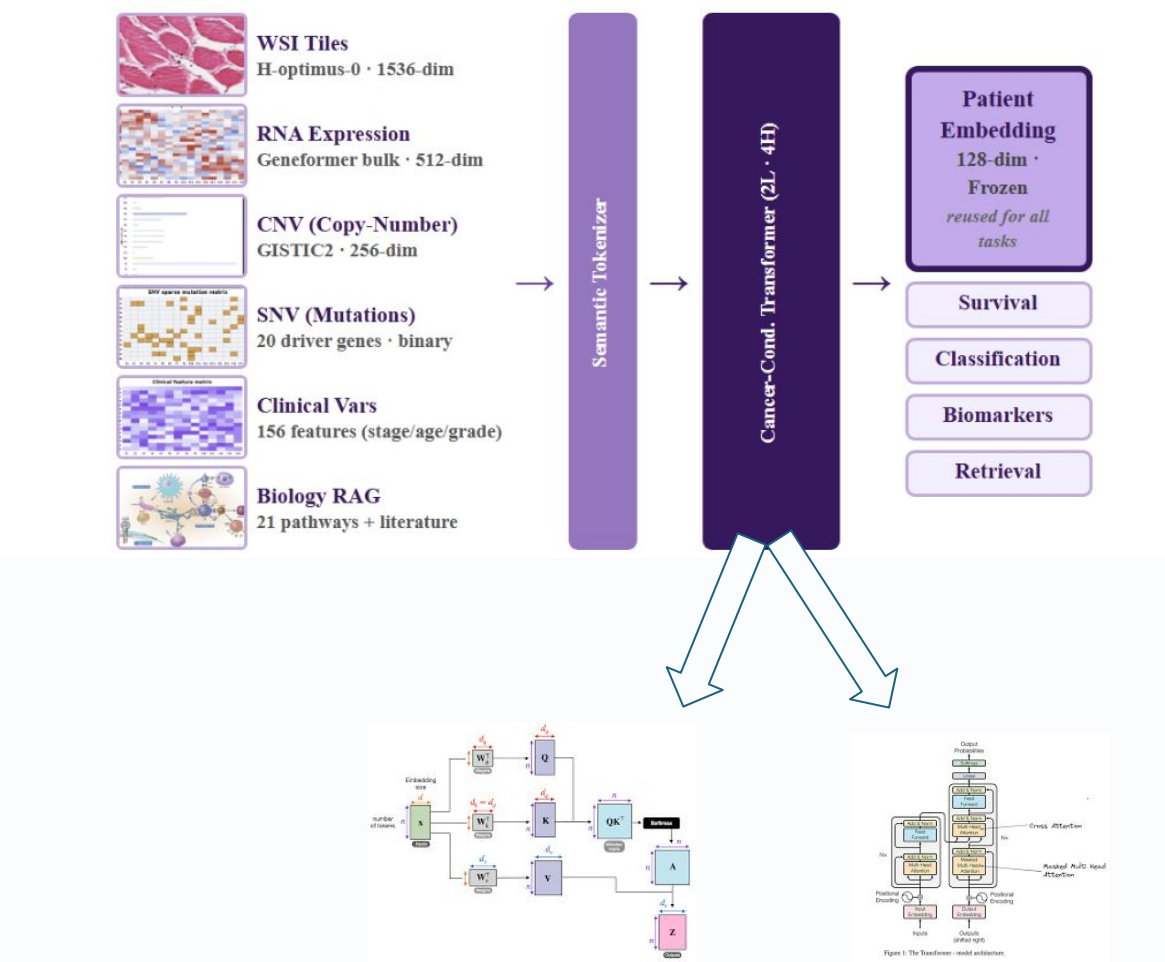
To test this idea, I built OmniSurv, a multimodal model that integrates six patient data modalities into a shared patient embedding:

- Pathology images to capture tumor morphology
- RNA expression to measure transcriptional activity
- Copy-number variation (CNV) to reflect genomic instability
- Somatic mutation (SNV) data to capture driver alterations
- Clinical variables to provide patient-level context
- Biology/text context from curated pathways and literature

Each modality is first encoded into a common token space. These tokens are then fused by cancer-conditioned transformer blocks, where:

- modality tokens first interact through self-attention,
- a context token then uses cross-attention to pool information across modalities,
- and the final context token becomes the patient embedding used for survival prediction and downstream tasks.

Training uses a stratified Cox survival loss, which compares patients within the same cancer during training so the model cannot rely on simple between-cancer survival shortcuts.



Evaluation Protocol

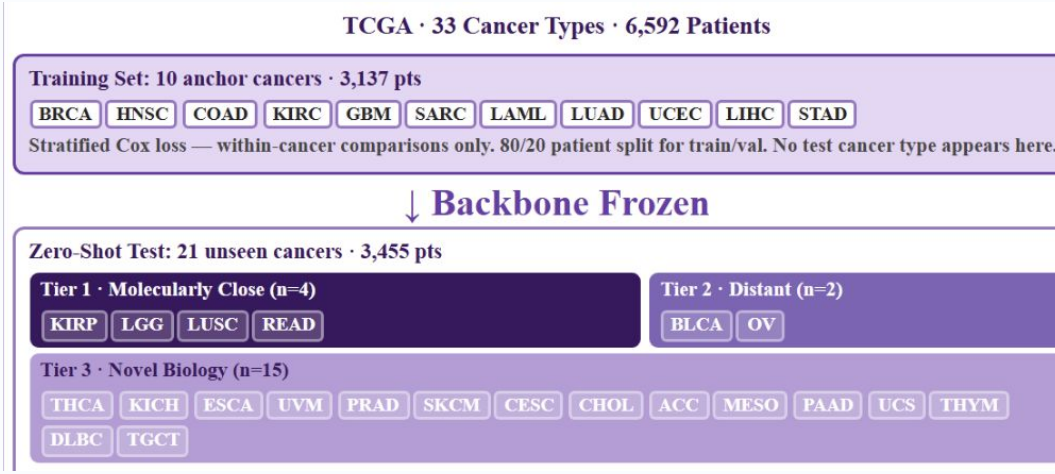
The central goal of this project is not just to predict survival on familiar data, but to test whether the model learns transferable tumor biology.

To do that, I used a cancer-type holdout protocol instead of standard pooled cross-validation:

- The model was trained on only 10 anchor cancers
- It was then evaluated on 22 completely unseen cancers
- At inference, the model was not given the cancer type

This creates a zero-shot setting where the model cannot rely on memorized disease-specific patterns. Instead, it must generalize using biological signals that are shared across cancers

To make this evaluation more honest, the primary metric is within-cancer C-index, which measures whether the model can correctly rank patient risk inside each unseen cancer, rather than benefiting from broad survival differences between cancer types.



$$\text{C-index} = \frac{\sum_{i,j} 1_{T_j<T_i} \cdot 1_{\eta_j>\eta_i} \cdot \delta_j}{\sum_{i,j} 1_{T_j<T_i} \cdot \delta_j}$$

Retrieved from Deotte, C. (2024). C-index calculation formula [Image]. Kaggle.
https://www.kaggle.com/competitions/equity-post-HCT-survival-predictions/discussion/554225

Results: OmniSurv Generalizes to Unseen Cancers for Survival

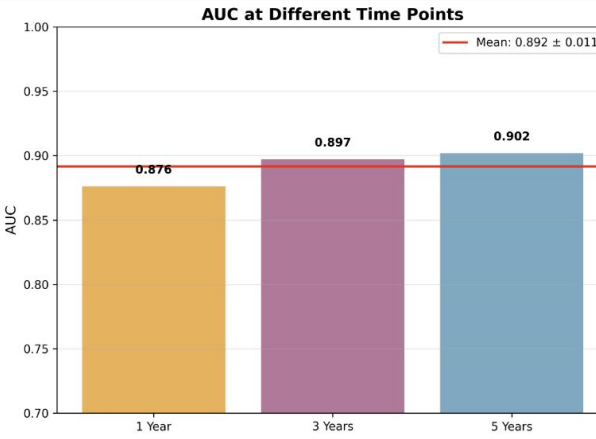
OmniSurv retained strong prognostic performance even when evaluated on cancers it had never seen during training.

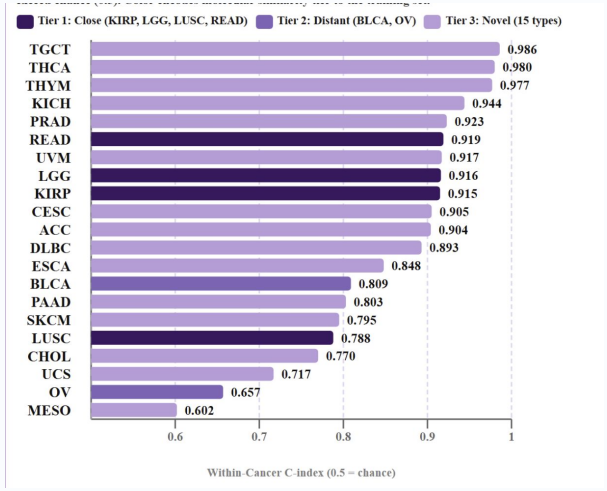
On the unseen-cancer test set:

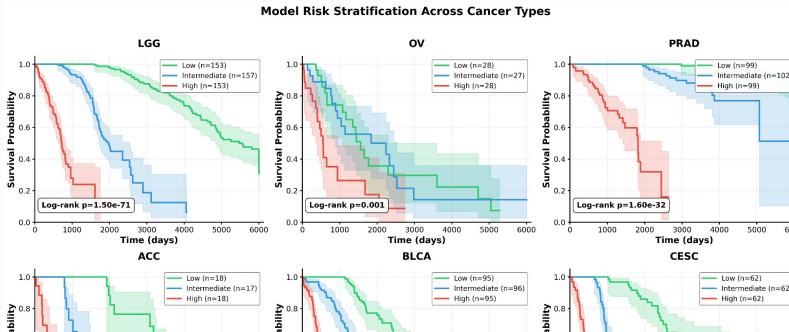
- the model achieved strong zero-shot survival discrimination
- it maintained high time-dependent AUC
- and its predictions separated patients into clearly different risk groups

This means the model was still able to identify which patients were more likely to do poorly, even in entirely new disease settings.

The most important result is that performance remained strong within individual unseen cancers, which supports the idea that the model learned transferable aggressiveness biology rather than only broad cancer-level survival differences, .







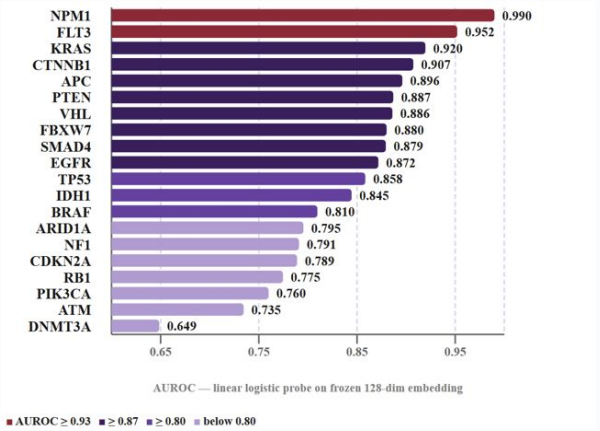
Results: The Embedding Captures Biology Beyond Survival

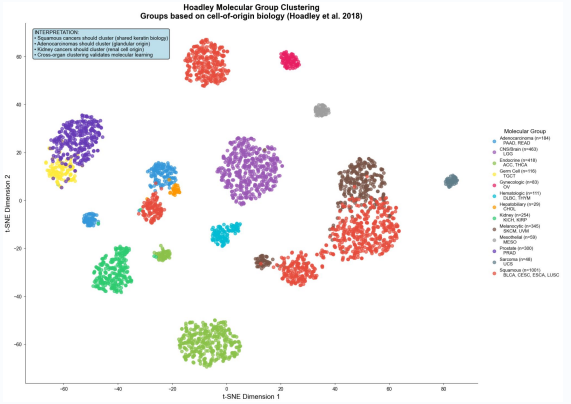
To probe deeper behind the survival score, I examined whether it helped the model learn a broader representation of tumor biology.

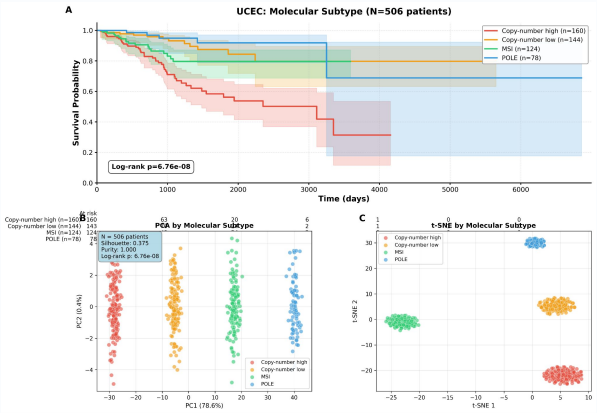
To test this, I froze the learned embedding and used simple downstream probes. Even without retraining the backbone, the embedding supported:

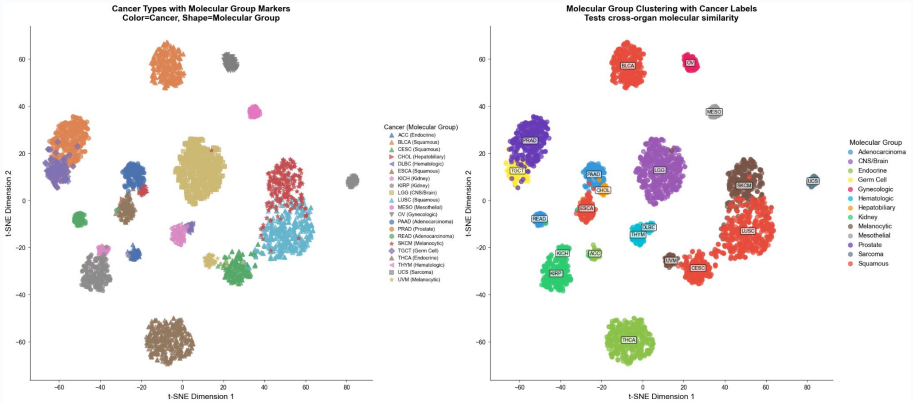
- biomarker prediction, including canonical driver-gene signals
- recovery of broader biological axes such as tumor mutational burden, immune state, and stemness
- subtype discovery, where clustering revealed groups with distinct survival and biological structure
- organization of cancers in a way that aligns with known pan-cancer molecular relationships

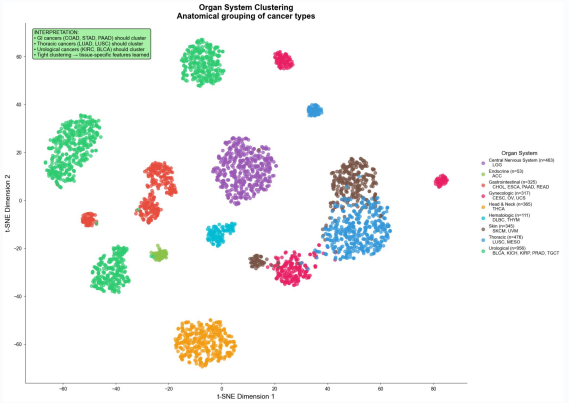
This is important because none of these tasks were used as the main training objective. They emerge from a model trained only on survival, suggesting that survival can act as a global teacher for learning reusable tumor biology.

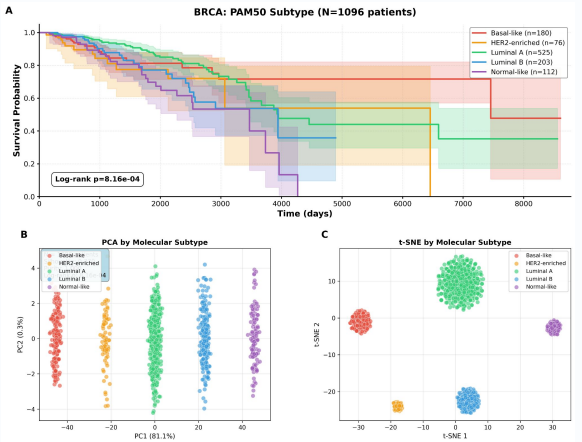












Conclusions and Limitations

Conclusions

These results support the idea that **survival can act as a global supervision signal** for learning transferable tumor biology.

Even though OmniSurv was trained only on survival outcomes, the learned embedding retained strong prognostic information, meaningful molecular and immune-related structure, and latent subtype organization.

This suggests the model learned more than shortcutting. Instead, it appears to have learned shared biological signals of tumor aggressiveness that transfer across diseases.

That is important because it points toward a different strategy for oncology AI where instead of training a separate model for every mutation, subtype, or task, it may be possible to train one general multimodal backbone and reuse it across many downstream problems.

Impact

- reduces dependence on many separate labeled datasets
- improves scalability across oncology
- may be especially useful for **rare, low-resource, and previously unseen cancers**
- provides a more honest benchmark for biological generalization

Limitations

Although these results are promising, this work has important limitations.

- The model was evaluated primarily on TCGA-derived data, so broader external and multi-institutional validation is still needed.
- Strong predictive performance does not by itself prove causal biological understanding. This is demonstrated by moderate and sometimes low alignment metrics with with labe biological datasets.

Future work will focus on:

- testing on more external cohorts and rarer cancers
- improving calibration and deployment safety
- strengthening robustness under missing or shifted modalities
- and refining the representation for clinical and biological interpretability

References

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 74(3), 229–263. https://doi.org/10.3322/caac.21834

Gatta, G., van der Zwan, J. M., Casali, P. G., Siesling, S., Dei Tos, A. P., Kunkler, I., Otter, R., Licitra, L., Mallone, S., Tavilla, A., Trama, A., & Capocaccia, R. (2011). Rare cancers are not so rare: The rare cancer burden in Europe. European Journal of Cancer, 47(17), 2493–2511. https://doi.org/10.1016/j.ejca.2011.08.008

Atun, R., Jaffray, D. A., Barton, M. B., Bray, F., Baumann, M., Vikram, B., Hanna, T. P., Knaul, F. M., Lievens, Y., Lui, T. Y. M., Milosevic, M., O'Sullivan, B., Rodin, D. L., Rosenblatt, E., Van Dyk, J., Yap, M. L., Zubizarreta, E., & Gospodarowicz, M. (2015). Expanding global access to radiotherapy. The Lancet Oncology, 16(10), 1153–1186. https://doi.org/10.1016/S1470-2045(15)00222-3

Elmore, S. N. C., Polo, A., Bourque, J.-M., Pynda, Y., van der Merwe, D., Grover, S., Hopkins, K., Zubizarreta, E., & Abdel-Wahab, M. (2021). Radiotherapy resources in Africa: An International Atomic Energy Agency update and analysis of projected needs. The Lancet Oncology, 22(9), e391–e399. https://doi.org/10.1016/S1470-2045(21)00351-X

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. Nature Medicine, 28(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0

Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge University Press. https://doi.org/10.1017/CBO9781107298019

Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. Array, 16, 100258. https://doi.org/10.1016/j.array.2022.100258

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., & Stuart, J. M. (2013). The Cancer Genome Atlas pan-cancer analysis project. Nature Genetics, 45(10), 1113–1120. https://doi.org/10.1038/ng.2764

Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., McKenna, M. T., Covington, K. R., Nokova, I., Beechem, J., Benz, C. C., & Stuart, J. M. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell, 173(2), 291–304. https://doi.org/10.1016/j.cell.2018.03.022

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Medical Research Methodology, 18(1), 24. https://doi.org/10.1186/s12874-018-0482-1

Chen, R. J., Lu, M. Y., Weng, W.-H., Chen, T. Y., Williamson, D. F. K., Manz, T., Shady, M., & Mahmood, F. (2021). Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. Proceedings of the IEEE/CVF International Conference on Computer Vision, 4015–4025. https://doi.org/10.1109/ICCV48922.2021.00398

Chen, R. J., Lu, M. Y., Williamson, D. F. K., Chen, T. Y., Lipkova, J., Noor, Z., Shaban, M., Shady, M., Williams, M., Joo, B., Noor, Z., & Mahmood, F. (2022). Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Cancer Cell, 40(8), 865–878. https://doi.org/10.1016/j.ccell.2022.07.004

Jaume, G., Vaidya, A., Chen, R. J., Williamson, D. F. K., Liang, P. P., & Mahmood, F. (2024). Modeling dense multimodal interactions between biological pathways and whole slide images for survival prediction. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11849–11861. https://doi.org/10.1109/CVPR52733.2024.01126

Vale-Silva, L. A., & Rohr, K. (2021). Long-term cancer survival prediction using multimodal deep learning. Scientific Reports, 11(1), 13505. https://doi.org/10.1038/s41598-021-92799-4

Cheerla, A., & Gevaert, O. (2019). Deep learning with multimodal representation for pancancer prognosis prediction. Bioinformatics, 35(14), i446–i454. https://doi.org/10.1093/bioinformatics/btz342

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., Williams, M., Oldenburg, L., Weishaupt, L. L., Wang, J. J., Vaidya, A., Le, L. P., Gerber, G., Sahai, S., Williams, W., & Mahmood, F. (2024). Towards a general-purpose foundation model for computational pathology. Nature Medicine, 30(3), 850–862. https://doi.org/10.1038/s41591-024-02857-3

Lu, M. Y., Chen, B., Williamson, D. F. K., Chen, R. J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L. P., Gerber, G., Parwani, A. V., Zhang, A., & Mahmood, F. (2024). A visual-language foundation model for computational pathology. Nature Medicine, 30(3), 863–874. https://doi.org/10.1038/s41591-024-02856-4

Filiot, A., Ghermi, R., Olivier, A., Jacob, P., Fidon, L., Kain, A. M., Saillard, C., & Schiratti, J.-B. (2023). Scaling self-supervised learning for histopathology with masked image modeling [Preprint]. medRxiv. https://doi.org/10.1101/2023.07.21.23292757

Chen, R. J., Chen, C., Li, Y., Chen, T. Y., Trister, A. D., Krishnan, R. G., & Mahmood, F. (2022). Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 16144–16155. https://doi.org/10.1109/CVPR52688.2022.01567

Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., Wong, C., Gero, Z., González, J., Gu, Y., Xu, Y., Wei, M., Wang, W., Ma, S., Wei, F., Yang, J., Li, C., Gao, J., Rosemon, J., … Poon, H. (2024). A whole-slide foundation model for digital pathology from real-world data. Nature, 630(8015), 181–188. https://doi.org/10.1038/s41586-024-07441-w

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. International Conference on Learning Representations. https://openreview.net/forum?id=YicbFdNTTy

Ilse, M., Tomczak, J., & Welling, M. (2018). Attention-based deep multiple instance learning. Proceedings of the 35th International Conference on Machine Learning, 80, 2127–2136. https://proceedings.mlr.press/v80/ilse18a.html

Bardes, A., Ponce, J., & LeCun, Y. (2022). VICReg: Variance-invariance-covariance regularization for self-supervised learning. International Conference on Learning Representations. https://openreview.net/forum?id=xm6YD62D1Ei

Oord, A. V. D., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748. https://arxiv.org/abs/1807.03748

Assran, M., Duval, Q., Misra, I., Bojanowski, P., Vincent, P., Rabbat, M., LeCun, Y., & Ballas, N. (2023). Self-supervised learning from images with a joint-embedding predictive architecture. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 15619–15629. https://doi.org/10.1109/CVPR52729.2023.01499