ARO: Aligned Representation learning for multi-Omics data

Amogh Singh [*1] **Yash Shah** [*1] **Chiara D'Ercoli** [*2] **Arash Mehrjou** [3 4] **Patrick Schwab** [3] **Timothy Jones** [1] **Pietro Liò** [1]

Abstract

The high cost of functional molecular assays, and prevalence of missing modalities and unmatched samples in computational biology, create significant barriers to comprehensive multi-omic profiling, essential for capturing and reasoning over molecules, cells, tissues, and organisms. This work proposes a model that learns meaningful representations from multi-omics cancer data supporting the reconstruction of missing and unpaired modalities. Contrary to increasingly complex, larger models, e.g. Foundation Models (FMs), ARO prioritizes practical applicability in limited or incomplete data settings. ARO optimally reconstructs missing modalities (MSE of 0.166 on validation and of 0.168 on test data), with its learned latent embeddings enabling a downstream cancer classification task. Our findings indicate that analyzing diverse molecular layers as a single integrated system offers a reliable and cost-efficient approach, reducing dependence on large-scale experimental testing, while still supporting multi-omic exploration in limited data settings.

1. Introduction

A common bottleneck in computational biology is the limited availability of paired, complete multi-omics datasets necessary to fully characterize biological processes. Prohibitively expensive assays bottleneck the acquisition of high-quality multi-omic data, as they require large sample volumes, and are harder to find in multiple paired modalities (Banerjee et al., 2023). Transcriptomic measurements constitute the most accessible data modality; however, they provide only a partial view of cellular function, not fully reflecting downstream biological processes shaped by post-translational modifications (Fernie & Stitt, 2012) and metabolic flux. This makes the integration of multiple omic layers essential for comprehensive understanding of disease mechanisms and therapeutic response, also in the context of rare diseases (Subramanian et al., 2020).

FMs and multi-omics data integration approaches have revolutionized the field by enabling self-supervised learning on several datasets that capture biological principles and derive joint representations of data. State-of-the-art FMs exploit information from omics modalities including imaging data or structural motifs (Yang et al., 2021; Osseni et al., 2022). However, by compressing these biologically rich layers into rigid structural abstractions or secondary imaging proxies, such architectures often overlook the dynamic signaling networks and regulatory feedback mechanisms that underpin cellular homeostasis and organisms representation.

Although existing approaches to multi-omics integration, such as Denoising Autoencoders (DAE) and Multi-Kernel Learning (MKL) have proven effective (Yao et al., 2025), they are typically tailored to narrow, task-specific settings. Moreover, they assume complete, well-aligned multi-omics data (Huang et al., 2025), or focus on robustness to missing subsets of modalities, rather than complete modality recovery (Lepe-Soltero et al., 2025).

Despite these advances, cross-modal inference, such as inferring proteomics or metabolomics from transcriptomics, is still an open problem. To address these limitations, this work aims to present the following contributions:

1. Introducing a model able to process a combination of transcriptomic, proteomic and metabolomic layers that does not require paired modality samples, and capable of mapping accessible transcriptomic profiles to predicted proteomic and metabolomic spaces.

2. Showing that ARO learns a shared latent representational space from which it can correctly infer data characteristics.

3. Demonstrating that when testing the proposed model on several downstream tasks, ARO achieves 100% ac-

[1]Department of Computer Science, University of Cambridge, Cambridge, UK [2]DIAG, University of Rome "Sapienza", Rome, Italy [3]GSK.ai [4]Max Plank Institute for Intelligent Systems. Correspondence to: Amogh Singh <as3600@cl.cam.ac.uk>, Yash Shah <ys562@cl.cam.ac.uk>, Chiara D'Ercoli <cd956@cl.cam.ac.uk>, <dercoli@diag.uniroma1.it>.

Proceedings of the ICML 2026 3rd Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences, Seoul, Korea. 2026.

curacy on classifying tumorous samples and 72.4% on laterality classification.

2. Background and Preliminaries

Current computational modeling research of omics-modalities focuses on developing FMs exploiting the transformer architecture (Vaswani et al., 2017), aimed at extracting and interpreting patterns from omics data, rather than reconstructing missing modalities or completing multi-omics profiles. These models are typically restricted to specific data modalities, including uni-modal (Ji et al., 2021; Yang et al., 2024; Dalla-Torre et al., 2025), single-cell (Cui et al., 2024; Chevalier et al., 2025), protein structure–informed (Zhao et al., 2026), or imaging datasets (Yang et al., 2021).

Recent multi-modal approaches for drug response prediction (Zhao et al., 2026) and other task-specific applications (Osseni et al., 2022) have begun integrating heterogeneous data sources through fusion techniques and transformers. However, these methods often combine fundamentally different modalities (e.g. structural and sequence data) rather than focusing on multi-omics alignment within the same-modality (e.g. tabular).

To ensure biologically coherent integration of modalities, some approaches leverage domain priors such as knowledge graphs and molecular pathways, while others redirect their focus on cross-species integration (Galkin et al., 2024; Huang et al., 2025). Additionally, recent work has leveraged Autoencoders (AEs) specialized to each modality to target omics-representation (Lepe-Soltero et al., 2025).

3. Data Extraction and Pre-processing

ARO reconstructs missing and unpaired modalities from a single observed input. Specifically, it enables the reconstruction of proteomic and metabolomic data from transcriptomic measurements. Unlike existing approaches, ARO employs a single autoencoder shared across modalities; does not require pretraining on large external datasets, and does not incorporate biological priors, pathway information, or auxiliary imaging data. Instead, through careful pre-processing (modality alignment, handling of missingness, and normalization), a unified latent representation (a single AE), and masking, it enables accurate reconstruction of masked data.

The study leverages two publicly available multi-omics cancer datasets: (1) a glioblastoma cohort (Wang et al., 2021), including gene expression (FPKM), proteome, phosphoproteome, and metabolome, as well as (2) a renal cell cancer cohort (Li et al., 2024), including RNA-seq expression (TPM), protein abundance, and metabolite compound abundance. For both cohorts, multi-omics profiles are obtained for tumor and normal paired samples.

Each modality is represented as a feature matrix, and these matrices are concatenated along the feature dimension to form a unified dataset of size $115 \times 74{,}002$, where each feature corresponds to a distinct molecular entity spanning transcriptomic, proteomic, and metabolomic measurements, and each row to a patient. Feature identifiers are harmonized to ENSG gene IDs via g:Profiler (Reimand et al., 2007), and KEGG metabolite IDs via MetaboAnalyst with stripped ENSEMBL IDs version suffixes. Features with zero variance (e.g. all-zero columns obtained after NaN imputation) are removed. Train-validation-test splits are performed at the patient level considering a 70/15/15 ratio.

Missing value handling and normalization is also performed. NaN positions are recorded into binary masks and filled with per-feature medians from the training-set. Next, winsorization reduces the influence of extreme outliers, by clamping values to the 1st and 99th percentiles (computed on the training data). Respecting the distinct magnitudes of each omic modality, Z-score normalization is utilized to perform per-feature standardization using training-set mean and standard deviation. Log transform (log1p) is applied after clamping to non-negative values. A second round of median imputation addresses NaNs introduced by Scipy's Z-score standardization (stable values for specific gene types, result in low variance features that cause a division by zero). The resulting unified matrix is then used as input for the modality masking step and subsequent model training.

Regarding normalization, we acknowledge that the two cohorts considered originate from different transcriptomic quantification pipelines (FPKM in (Wang et al., 2021) and TPM in (Li et al., 2024)). Rather than harmonizing the datasets by converting them to a common expression metric, we adopted an approach more aligned with machine learning practice than with traditional bioinformatics workflows. Specifically, normalization was performed independently for each feature, allowing each gene to be standardized according to its own distribution within the respective dataset. Furthermore, the primary objective of this work is methodological rather than bioinformatic. Our aim was not to construct a rigorously harmonized multi-cohort expression dataset, but rather to investigate whether an autoencoder-based framework is able to learn meaningful latent representations from heterogeneous omics data under minimal domain-specific preprocessing assumptions.

We would like to emphasize that care is taken to prevent leakage of any information regarding the validation or test set distributions into the training data.

4. Architecture

ARO is a masked autoencoder designed to learn robust representations from high-dimensional multi-omics data, as

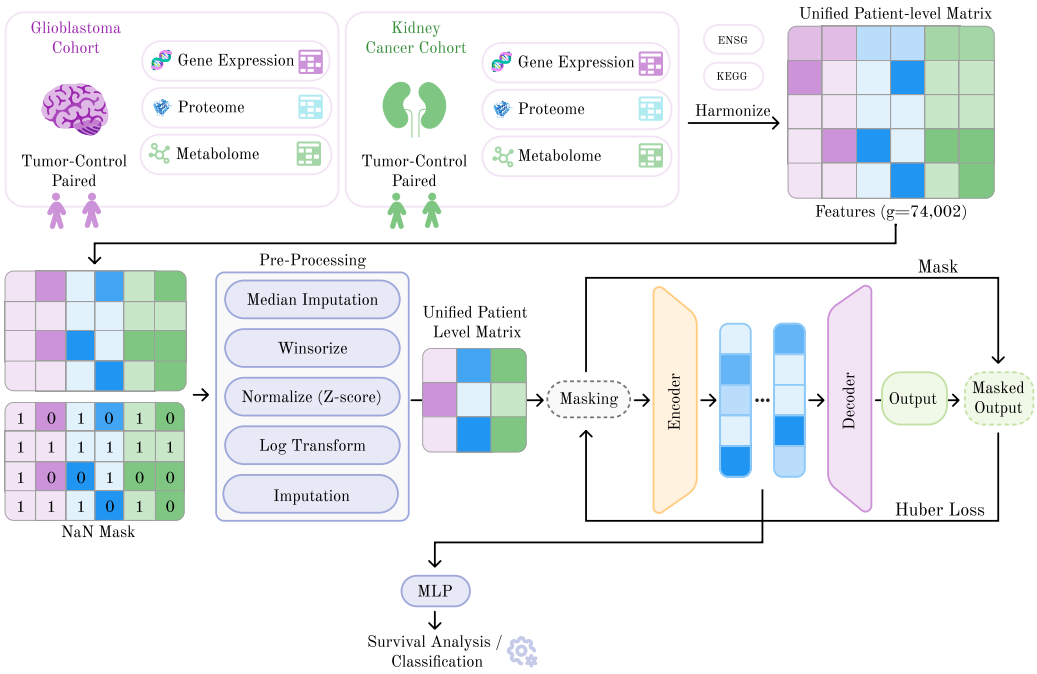


Figure 1. Overview of ARO. Multi-omics cancer datasets are first harmonized and aggregated. Next, pre-processing involving median imputation, Winsorization, normalization, and log transformation is performed. Features are then optionally masked (shown by dotted outlines) and passed to the model. The loss is computed, and learned latent representations can be used for downstream tasks.

illustrated in Figure 1. It employs a multi-layer perceptron (MLP)-based (Rumelhart et al., 1986) encoder/decoder, and a fully-connected hidden layer stack with ReLU nonlinearities. Details are provided in Appendix A.

Two setups are used for training ARO: (1) masking all omics modalities in the input apart from transcriptomics (ARO MASKED), and (2) omitting input masking (ARO UNMASKED). Input masking allows ARO to reconstruct the other two omic-modalities using only transcriptomics.

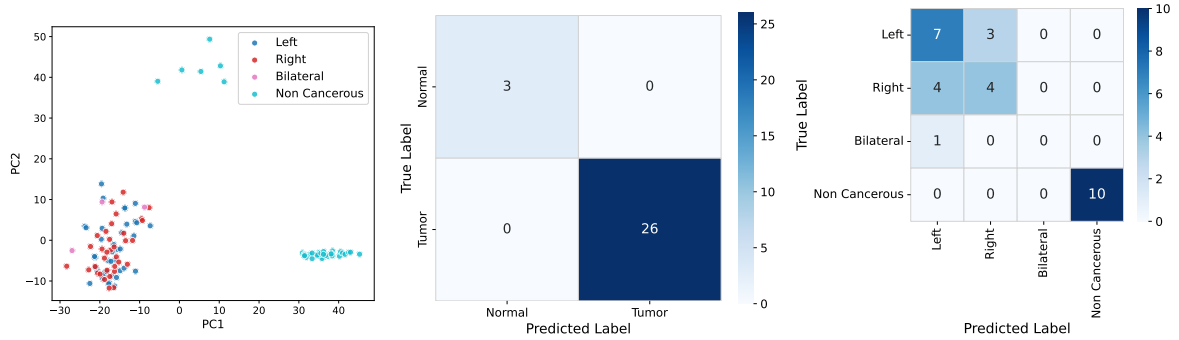
During training, a random subset of features (10%) is dropped at each forward pass by setting their values to zero. The model is then optimized to reconstruct the original input via the decoder. Reconstruction quality is optimized using the Huber loss (Huber, 1992) (defined in Equation (1)), with missing values (NaNs) excluded from this loss via a binary mask, ensuring that the model is not penalized for imputing unobserved data. Optimization and hyperparameter tuning details are provided in Appendix B. Performance is evaluated through Mean Squared Error (MSE), together with downstream classification metrics and confusion matrices.

5. Experiments and Results

Reconstruction ARO is designed to reconstruct omic profiles in both masked and unmasked settings, where the masked setting involves inferring the remaining omic modalities solely from transcriptomic input. Given the combination of imputation and data handling during the preprocessing step, we ensured that the model did not exhibit any bias toward trivial solutions–such as predicting median-valued outputs. In particular, since the reconstruction loss is computed only over observed values (non-masked), and the masking pattern varies across samples and batches, the model is discouraged from converging to a trivial statistic. This is empirically demonstrated by the fact that the reconstructed outputs preserve the underlying distributional structure of the data rather than collapsing to mean-like predictions.

Reconstruction performance of masked and unmasked ARO compared to PCA, is presented in Tables 1 and 2 and the full results are in Tables 4 and 5. Although PCA achieves superior reconstruction accuracy, its performance on downstream tasks, discussed subsequently, is substantially weaker.

After identifying the framework proposed in Lepe-Soltero et al. as the closest approach to ARO, we aimed at comparing the two methods. However, due to the high dimensionality of our data, the MODIS architecture scales to approximately 22B parameters, resulting in out-of-memory (OOM) errors and rendering complete training infeasible. Constructing a reduced-scale version of MODIS was not a



(a) *(b)* *(c)*

Figure 2

Figure 3. **ARO performance** in discriminating each omics-modality within the latent space, visualized using PCA (Figure 2a) as well as across downstream tasks including binary cancer classification (Figure 2b) and cancer laterality prediction (Figure 2c).

meaningful option. The high dimensionality of our dataset primarily arises from the large number of gene features rather than the number of samples–therefore, reducing the model size would have required discarding a substantial portion of the omics set, fundamentally altering the problem setting and compromising validity of the comparison. In contrast, ARO remains computationally tractable across a plausible parameter range (≈50M–700M parameters), with model size showing limited impact on reconstruction performance, thereby enabling more efficient training in high-dimensional regimes. Our limited MODIS training achieved an MSE loss of 2.53, 0.55 and 0.43 for transcriptomics, proteomics, and metabolomics respectively. This further credits the robust performance of ARO compared to existing work.

Table 1. Reconstruction performance using MSE Loss averaged over 5 runs

	TRAIN	VALIDATION	TEST
ARO MASKED	0.046	0.14	0.14
ARO UNMASKED	0.064	0.15	0.15
PCA	$3.1e-30$	0.14	0.14

Table 2. Reconstruction performance using Huber Loss averaged over 5 runs

	TRAIN	VALIDATION	TEST
ARO MASKED	0.022	0.070	0.068
ARO UNMASKED	0.032	0.071	0.07
PCA	$1.5e-30$	0.068	0.065

Shared representation Embeddings extracted from the latent space of our trained models are used to generate PCA (Figure 2a) and t-SNE plots (Figure 5) that show distinct clusters of cancerous and non-cancerous samples, and of cancer laterality. The clear separation between sample groups indicates that the learned embeddings effectively capture the underlying structure of the original omics data while retaining sufficient discriminatory information to distinguish between clinical states.

Downstream analysis For evaluating the performance of ARO on downstream tasks, the embeddings from the hidden layer of the ARO UNMASKED model are used with an MLP to perform binary classification (cancer/non-cancer) and multi-label classification (cancer laterality). ARO performs well on the binary classification task, achieving 100% accuracy on test data, and it accurately distinguishes non-cancerous from cancerous samples, as shown in Figures 2b and 2c respectively. Although ARO demonstrated ability to infer tumor laterality, with an accuracy of 72.4%, performance on this task was comparatively weaker—an outcome that aligns with biological expectations, as predicting laterality requires more subtle and complex biological signals (i.e.: spatial or micro-environmental) that cannot be inferred from omics profiles alone. In comparison, using the projections from PCA for downstream tasks gives 72.41% and 58.62% accuracy on the binary and multi-label classification tasks respectively, and the corresponding confusion matrices are shown in Figure 6b and Figure 6a.

6. Limitations and Future Work

Although the data are highly feature-dimensional, the number of training samples remains limited. This setting may constrain the model's ability to generalize to unseen data distributions, highlighting the need for future validation on external cohorts, possibly also including other diseases.

ARO's low-capacity architecture enables effective training with limited samples, reflecting the intended scope of the model. The focus of this work is not on architectural innovation, but on robust learning from aligned multi-omics

matrices, in contrast to heterogeneous multi-modal settings such as imaging or spatial transcriptomics. The learned associations remain statistical regularities and carry no claims of biological plausibility. Further analysis would be necessary to substantiate potential biological meaning.

Finally, an extension of ARO could include the integration of generative settings, such as variational autoencoders (VAEs), to enable the full synthesis of proteomic and metabolomic profiles conditioned on transcriptomic inputs.

7. Conclusion

ARO captures meaningful biological structure across heterogeneous omics modalities and enables reconstruction of missing modalities from a single input. By training ARO to reconstruct masked inputs, robust shared representations are learned. Notably, even when trained solely on the reconstruction objective, the learned latent representations exhibit well-separated clustering structure using linear projections (PCA) and non-linear techniques (t-SNE). The obtained embeddings enable accurate discrimination between cancerous and non-cancerous samples and qualitative performance on cancer laterality classification using lightweight classifiers such as MLPs.

These results highlight the promise of masked reconstruction as a foundation for multi-omics representation learning, motivating future work to evaluate generalization across datasets and disease types, as well as to explore generative extensions of ARO, such as Large Language Models (LLMs) through the obtained representations.

The potential of this work lies in learning a unified representation across multiple omic modalities within a single latent space, instead of combining separate representations from each modality after training. Given the increasing availability of open biomedical datasets and the growing interest in computational biology modeling, we believe that these developments may contribute to more comprehensive solutions for paired multi-omics data collection.

Software and Data

This study leverages publicly available multi-omics data from two distinct cancer cohorts: (1) glioblastoma (Wang et al., 2021) and (2) renal cell carcinoma (Li et al., 2024), to evaluate the model's capability for cross-omic translation and latent representation alignment. The complete source code is available at https://github.com/ccdderc/ARO.

Impact Statement

The broader impact of this work lies at the interface of machine learning and computational biology. By enabling the reconstruction of entire omics modalities—such as proteomics and metabolomics—from more readily available transcriptomic data, the proposed approach addresses the limitation of paired dataset scarcity in multi-omics research and offers the potential to expand the utility of existing data resources. However, reconstructed modalities should be interpreted as approximations rather than direct biological measurements. Care must be taken to validate results in clinical or high-stakes applications, to avoid over-reliance on inferred data.

Contributions

Amogh Singh contributed to the multi-omic data selection process, development of the model, collection of the comparison metrics and the analysis to validate the performance on downstream tasks. **Yash Shah** also contributed to the implementation of the primary autoencoder approach, including the data pre-processing pipeline, and explored alternative approaches, namely the use of variational autoencoders and transformers. **Chiara D'Ercoli** also contributed to the multi-omic data selection process, including data acquisition and preparation and efforts to source additional data, literature review, drafting, writing, and finalizing the manuscript.

References

Banerjee, J., Taroni, J. N., Allaway, R. J., Prasad, D. V., Guinney, J., and Greene, C. Machine learning in rare disease. *Nature methods*, 20(6):803–814, 2023.

Bourlard, H. and Kamp, Y. Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics*, 59(4):291–294, 1988.

Chevalier, A., Ghosh, S., Awasthi, U., Watkins, J., Bieniewska, J., Mitrea, N., Kotova, O., Shkura, K., Noble, A., Steinbaugh, M., et al. Teddy: a family of foundation models for understanding single cell biology. *arXiv preprint arXiv:2503.03485*, 2025.

Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., and Wang, B. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature methods*, 21(8):1470–1480, 2024.

Dalla-Torre, H., Gonzalez, L., Mendoza-Revilla, J., Lopez Carranza, N., Grzywaczewski, A. H., Oteri, F., Dallago, C., Trop, E., De Almeida, B. P., Sirelkhatim, H., et al. Nucleotide transformer: building and evaluating robust foundation models for human genomics. *Nature Methods*, 22(2):287–297, 2025.

Fernie, A. R. and Stitt, M. On the discordance of metabolomics with proteomics and transcriptomics: coping with increasing complexity in logic, chemistry, and

network interactions scientific correspondence. *Plant Physiology*, 158(3):1139–1145, 2012.

Galkin, F., Naumov, V., Pushkov, S., Sidorenko, D., Urban, A., Zagirova, D., Alawi, K. M., Aliper, A., Gumerov, R., Kalashnikov, A., Mukba, S., Pogorelskaya, A., Ren, F., Shneyderman, A., Tang, Q., Xiao, D., Tyshkovskiy, A., Ying, K., Gladyshev, V. N., and Zhavoronkov, A. Precious3gpt: Multimodal multi-species multi-omics multi-tissue transformer for aging research and drug discovery. *bioRxiv*, 2024. doi: 10.1101/2024.07.25.605062. URL https://www.biorxiv.org/content/early/2024/07/25/2024.07.25.605062.

Huang, Y., Su, X., Ullanat, V., Moon, I., Liang, I., Clegg, L., Olabode, D., Johnson, R., Ho, N., Gibbs, M., et al. Multimodal ai predicts clinical outcomes of drug combinations from preclinical data. *arXiv preprint arXiv:2503.02781*, 2025.

Huber, P. J. Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution*, pp. 492–518. Springer, 1992.

Ji, Y., Zhou, Z., Liu, H., and Davuluri, R. V. Dnabert: pre-trained bidirectional encoder representations from transformers model for dna-language in genome. *Bioinformatics*, 37(15):2112–2120, 2021.

Lepe-Soltero, D., Artières, T., Baudot, A., and Villoutreix, P. Modis: multi-omics data integration for small and unpaired datasets. *arXiv preprint arXiv:2503.18856*, 2025.

Li, G. X., Chen, L., Hsiao, Y., Mannan, R., Zhang, Y., Luo, J., Petralia, F., Cho, H., Hosseini, N., da Veiga Leprevost, F., et al. Comprehensive proteogenomic characterization of rare kidney tumors. *Cell Reports Medicine*, 5(5), 2024.

Osseni, M. A., Tossou, P., Laviolette, F., and Corbeil, J. Mot: a multi-omics transformer for multiclass classification tumour types predictions. *BioRxiv*, pp. 2022–11, 2022.

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. g: Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic acids research*, 35(suppl_2):W193–W200, 2007.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. Multi-omics data integration, interpretation, and its application. *Bioinformatics and biology insights*, 14:1177932219899051, 2020.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Wang, L.-B., Karpova, A., Gritsenko, M. A., Kyle, J. E., Cao, S., Li, Y., Rykunov, D., Colaprico, A., Rothstein, J. H., Hong, R., et al. Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer cell*, 39 (4):509–528, 2021.

Yang, K. D., Belyaeva, A., Venkatachalapathy, S., Damodaran, K., Katcoff, A., Radhakrishnan, A., Shivashankar, G., and Uhler, C. Multi-domain translation between single-cell imaging and sequencing data using autoencoders. *Nature communications*, 12(1):31, 2021.

Yang, Y., Sun, S., Yang, S., Yang, Q., Lu, X., Wang, X., Yu, Q., Huo, X., and Qian, X. Structural annotation of unknown molecules in a miniaturized mass spectrometer based on a transformer enabled fragment tree method. *Communications Chemistry*, 7(1):109, 2024.

Yao, X., Wang, T., Yang, Q., Wang, J., Qi, Y., Xu, T., Wei, Z., Cui, Y., Cao, H., and Yun, K. Multi-omics data integration for improved cancer subtyping via denoising autoencoder-based multi-kernel learning. *Genes*, 16(11): 1246, 2025.

Zhao, Y., Tennant, J., Yang, J., Guo, Z., Whang, Y., and Sui, N. Deepdtf: Dual-branch transformer fusion for multi-omics anticancer drug response prediction. *arXiv preprint arXiv:2603.24265*, 2026.

Appendix

A. Model Architecture

The architecture of ARO is shown in Figure 1. The model is implemented as a fully-connected feedforward autoencoder. The encoder consists of an input dropout layer ($p = 0.1$), followed by a linear projection from 74,002 to 2,048 dimensions and three fully connected hidden layers of dimension 2,048. Each hidden layer is followed by a ReLU activation and dropout ($p = 0.0$). The result is a 2,048-dimensional hidden representation of modalities.

The decoder reconstructs the original feature space through a final linear layer projection from the latent representation back to the original dimensions, followed by a softplus activation to enforce non-negative outputs after preprocessing and log-space normalization.

The training pipeline is implemented using PyTorch Lightning, with Hydra for configuration management and Weights & Biases for experiment tracking and hyperparameter tuning. The model evaluated in Sec. 5 has a total of ∼303.9M trainable parameters and the outline is reported in Table 3.

Table 3. Parameter numbers of each block of the autoencoder.

LAYER	SHAPE	PARAMETERS
INP_LIN.WEIGHT	(2048, 71094)	145,600,512
INP_LIN.BIAS	(2048,)	2048
ENCODER_LAYERS.0.WEIGHT	(2048, 2048)	4,194,304
ENCODER_LAYERS.0.BIAS	(2048,)	2048
ENCODER_LAYERS.1.WEIGHT	(2048, 2048)	4,194,304
ENCODER_LAYERS.1.BIAS	(2048,)	2048
ENCODER_LAYERS.2.WEIGHT	(2048, 2048)	4,194,304
ENCODER_LAYERS.2.BIAS	(2048,)	2048
OUT_LIN.WEIGHT	(71094, 2048)	145,600,512
OUT_LIN.BIAS	(71094,)	71094

B. Optimization Details and Hyperparameter Tuning

Optimization is performed using AdamW with a learning rate of 10^{-5} and no weight decay. A ReduceLROnPlateau scheduler is employed with a reduction factor of 0.1 and patience of 10 epochs. Training is performed for 200 epochs with a batch size of 80 using mixed bfloat16 precision (fp32 as fallback on unsupported hardware) and JIT compilation via torch.compile on CUDA-enabled devices. No batch normalization is employed.

Hyperparameter tuning was performed via Bayesian Optimization through Weights & Biases. The parameters optimized for are: dropout configuration, hidden dimensions, learning rate, modality-specific masking, number of hidden layers, batch normalization usage and weight decay—and the results are shown in Figure 4.

During training, the model optimizes for the Huber loss, defined as:

$$L_\delta(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{if } |y - \hat{y}| \le \delta \\ \delta\left(|y - \hat{y}| - \frac{1}{2}\delta\right) & \text{otherwise} \end{cases} \tag{1}$$

C. Autoencoders

The model exploits the Autoencoder (AE) architecture (Bourlard & Kamp, 1988). AEs are unsupervised neural networks designed to learn efficient data encodings by forcing the model to reconstruct the input from a compressed latent representation (i.e. the "bottleneck"). An autoencoder consists of both an encoder and a decoder network, whereby the encoder first compresses the input $x \in \mathbb{R}^n$ into a latent space representation $h \in \mathbb{R}^m$. The decoder then maps this latent representation back to the original input space to produce a reconstruction $y \in \mathbb{R}^n$:

$$h = f_e(x) = s_e(W_e x + b_e) \tag{2}$$

$$y = f_d(h) = s_d(W_d h + b_d) \tag{3}$$

where $s_{\{e,d\}}$ are the encoder/decoder activation functions, $W_{\{e,d\}} \in \mathbb{R}^{m \times n}$ are the encoder/decoder weight matrices, and $b_{\{e,d\}} \in \mathbb{R}^m$ are the encoder/decoder biases.

The objective of the AE is to minimize the reconstruction loss L_δ (Huber loss in our case, defined in Eq. 1) over a dataset of N samples, expressed as:

$$\arg\min J\left(W_{\{e,d\}}, b_{\{e,d\}}\right) = \frac{1}{N} \sum_{i=1}^{N} L_\delta \left(x^{(i)}, y^{(i)}\right) \tag{4}$$

D. Results

Table 4. Reconstruction performance using MSE Loss averaged over 5 runs

	TRAIN	VALIDATION	TEST
ARO MASKED	$0.046 \pm 56.0e-04$	$0.14 \pm 31.0e-05$	$0.14 \pm 21.0e-05$
ARO UNMASKED	$0.064 \pm 31.0e-05$	$0.15 \pm 21.0e-05$	$0.15 \pm 24.0e-05$
PCA	$3.1e-30$	0.14	0.14

Table 5. Reconstruction performance using Huber Loss averaged over 5 runs

	TRAIN	VALIDATION	TEST
ARO MASKED	$0.022 \pm 28.0e-04$	$0.070 \pm 5.6e-05$	$0.068 \pm 20.0e-05$
ARO UNMASKED	$0.032 \pm 16.0e-05$	$0.071 \pm 10.0e-05$	$0.07 \pm 6.1e-05$
PCA	$1.5e-30$	0.068	0.065

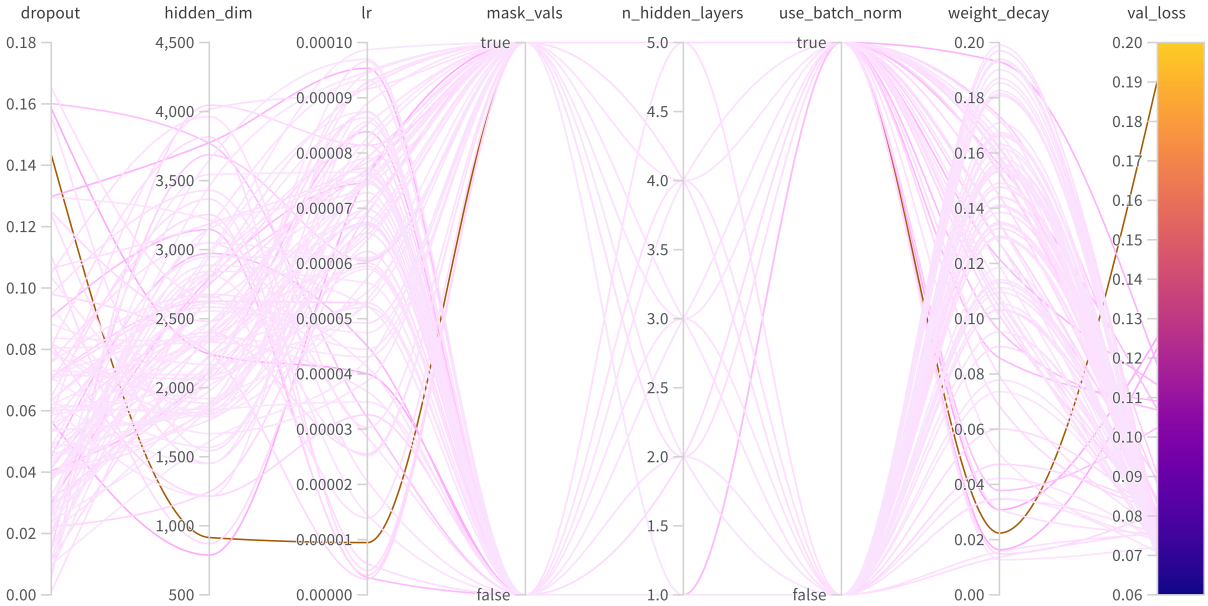
E. Comparison

Confusion matrix of the downstream analysis using PCA is given in Figure 6.

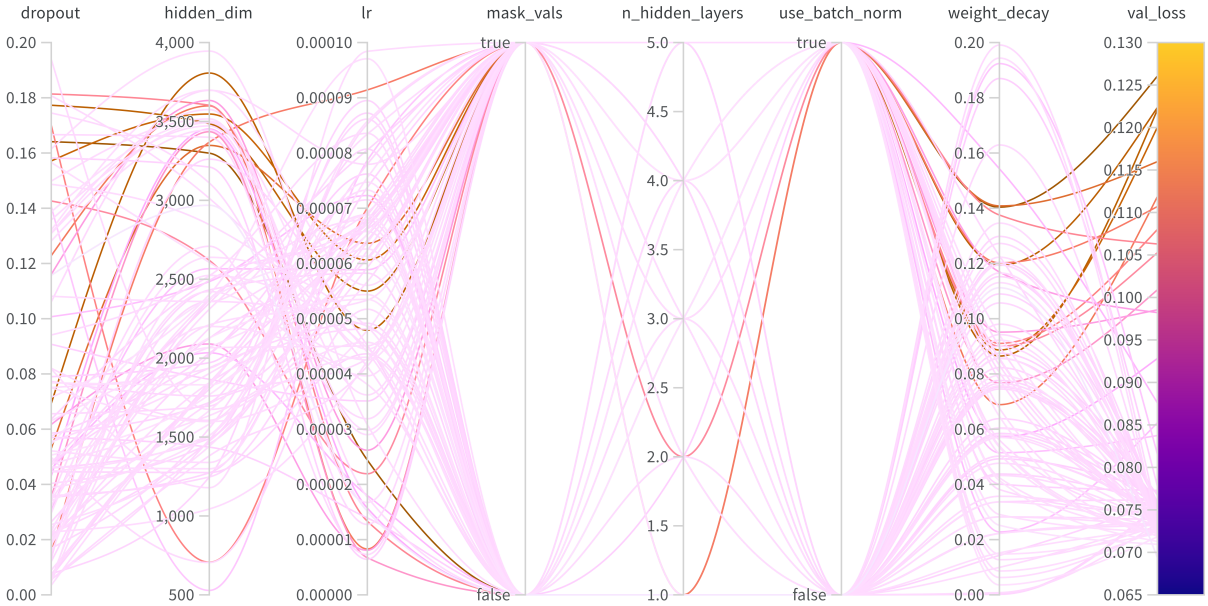
F. Dataset Breakdown

Table 6. **Details of the Multi-omics** datasets considered and analysed in the study.

DATA SET	SAMPLES	FEATURE COUNT
(LI ET AL., 2024) - TRANSCRIPTOMICS	108	45886
(LI ET AL., 2024) - PROTEOMICS	109	10998
(LI ET AL., 2024) - METABOLOMICS	87	67
(WANG ET AL., 2021) - TRANSCRIPTOMICS	240	26941
(WANG ET AL., 2021) - PROTEOMICS	333	12309
(WANG ET AL., 2021) - METABOLOMICS	36	216



(a) Using dropout on input



(b) Not using dropout on input

Figure 4. Iterations for hyperparameter tuning performed over validation loss, optimizing for dropout configuration, hidden dimensions, learning rate, modality-specific masking, number of hidden layers, batch normalization usage and weight decay. The configuration with the lowest validation loss is chosen from each setup.

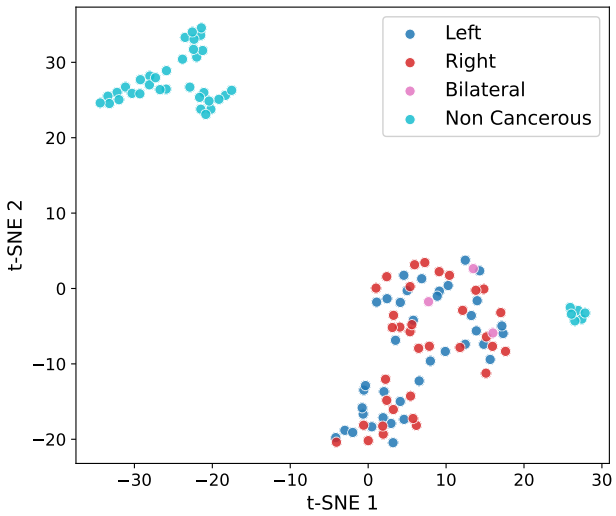
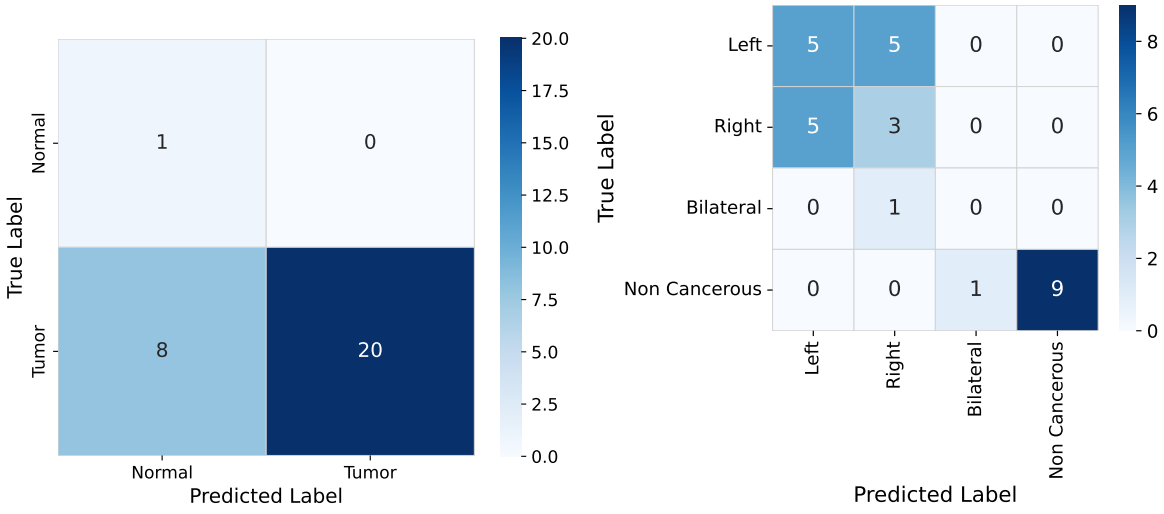


Figure 5. TSNE cluster of the learned embeddings from ARO



(a) Binary classification task

(b) Multi-label classification task

Figure 6. Confusion matrices of downstream tasks using PCA projections.