

ORIGINAL INVESTIGATION | COMPUTATIONAL ONCOLOGY

Integrative Multi-Omics Learning for Cancer Subtype Classification in Incomplete Molecular Cohorts

Christina Curtis[1], Andrea Califano[2], Dana Pe'er[3], Eytan Ruppin[4], Sohrab Shah[3], Olivier Elemento[5], Benjamin Haibe-Kains[6], Han Liang[7], Mert Sabuncu[8], Eli Papaemmanuil[3], Rehan Akbani[7], Quaid Morris[6], Anil Korkut[7], Iman Hajirasouliha[5], John N. Weinstein[7], Christina Leslie[3], Naim U. Rashid[9], Nikolaus Schultz[3], Ye Zheng[7], Richard Moffitt[10], Andrea Sboner[5], Wenyi Wang[7], Xiaosong (Jonathan) Wang[11], Jeffrey H. Chuang[12], Samie Jaffrey[5], Jing Wang[7], Anguraj Sadanandam[13], Benjamin G. Neel[14]

[1] Department of Biomedical Data Science, Stanford University, Stanford, California, USA
[2] Department of Systems Biology, Columbia University, New York, New York, USA
[3] Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York, USA
[4] National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA
[5] Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, New York, USA
[6] Department of Computer Science, University of Toronto, Toronto, Ontario, Canada
[7] Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA
[8] Department of Biomedical Engineering, Cornell University, Ithaca, New York, USA
[9] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[10] Department of Biomedical Informatics, Emory University, Atlanta, Georgia, USA
[11] Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[12] The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA
[13] Division of Molecular Pathology, The Institute of Cancer Research, London, United Kingdom
[14] Perlmutter Cancer Center, NYU Langone Health, New York, New York, USA

Corresponding author: Christina Curtis, Department of Biomedical Data Science, Stanford University, Stanford, California, USA

Word count: approximately 3,000 (excluding abstract, tables, figure legends, and references).

Figures: 3.

Tables: 2.

References: 99.

KEY POINTS

Question Can a computational framework integrate incomplete multi-omics profiles and transfer subtype information to small target cohorts?

Findings In this benchmark evaluation adapted from a published source report, a four-way 10-shot transfer experiment reported 67.5% accuracy, 83.7% AUROC, 73.3% precision, and 67.6% class-weighted F1 for mRNA-to-mRNA transfer.

Meaning Multi-omics reconstruction and distribution-aware transfer may support cancer subtype classification when molecular measurements and labeled samples are limited; independent reproduction and validation are needed.

ABSTRACT

Importance Cancer subtype classification can be constrained by incomplete molecular profiling and small labeled cohorts.

Objective To describe a computational framework that combines missing-modality reconstruction, multi-omics integration, and cross-cohort meta-learning for cancer subtype diagnosis.

Design Computational analysis based on the published CancerMap framework, using publicly available cancer molecular datasets.

Setting Public gastric, lung, and breast cancer molecular datasets.

Participants Publicly available tumor samples with annotated molecular subtype labels from gastric, lung, and breast cancer datasets.

Methods The framework uses modality-specific feature encoders, low-rank multimodal fusion, contrastive representation learning, masked reconstruction, class-weighted subtype classification, and model-agnostic meta-learning with a category-level distributional contrastive term.

Main Outcomes and Measures Subtype classification accuracy, AUROC, precision, class-weighted F1, missing-modality reconstruction performance, and association of an integrated score with overall survival.

Results For the reported mRNA-to-mRNA four-way 10-shot transfer experiment, accuracy was 67.5%, AUROC 83.7%, precision 73.3%, and class-weighted F1 67.6%. The reported multivariable association between high vs low Integrated CancerMap Score and overall survival was HR 1.70 (95% CI, 1.07–2.70; P = .03). These values are reproduced as reported in the source report and have not been independently recalculated here.

Conclusions and Relevance The described approach combines reconstruction and cross-cohort adaptation to address incomplete molecular profiles and limited labeled samples. Its predictive and prognostic performance requires independent code-level reproduction, validation in external cohorts, and confirmation of all numerical results before clinical interpretation.

INTRODUCTION

Cancer is molecularly heterogeneous, and tumors arising in the same tissue can differ in clinical presentation, treatment response, and outcome.[1] Molecular subtype classification can therefore support precision oncology.[4–7] However, computational methods that integrate genomic, transcriptomic, and epigenomic measurements often assume that every patient has complete measurements across all modalities and that large labeled cohorts are available.[11–14] These assumptions are difficult to satisfy in routine clinical datasets.

Removing patients with missing modalities can discard useful samples and further reduce cohort size.[14–17] Imputation can preserve incomplete cases, but reconstruction methods may require large complete training datasets.[18–21] Transfer learning and few-shot learning offer a way to use external cohorts,[23–25] although differences in sample collection and cohort composition can produce negative transfer.[26] Gastric cancer is emphasized because of its molecular heterogeneity and tumor-microenvironment variation.[29–32] The four gastric molecular subtypes considered are Epstein-Barr virus (EBV), microsatellite instability (MSI), genomically stable (GS), and chromosomal instability (CIN).[34] Lung and breast cancer cohorts provide additional evaluation settings.

METHODS

Study Design and Data Sources

CancerMap uses TCGA-STAD and GSE62254 as gastric cancer cohorts, TCGA-LUAD, TCGA-LUSC, and CPTAC resources79,80 as lung cancer cohorts, and TCGA-BRCA as the breast cancer cohort (Table 1).; TCGA-LUAD, TCGA-LUSC, and CPTAC resources[79,80] as lung cancer cohorts; and TCGA-BRCA as the breast cancer cohort (Table 1). Gastric subtype labels were EBV, MSI, GS, and CIN. Breast cancer labels followed PAM50 categories.[81] Three molecular modalities were described: DNA methylation, miRNA expression, and mRNA expression. Only samples with known subtype labels were retained.

Expression values were log2 transformed. Low-variance features were removed using thresholds of 0.2 for methylation, 0.1 for miRNA, and 0.8 for mRNA; a set of features recommended in prior pan-cancer work[82] was then added back. The final feature counts reported were 3,287 methylation features, 702 miRNA features, and 4,089 mRNA features (Figure 2B). miRNA and mRNA values were rescaled to [0, 1]; methylation values were already in that range. Cohort-specific sample counts and preprocessing procedures are described.

Table 1. Cohorts and molecular data described in the source report

Cohort	Cancer type	Subtype labels	Omics available
TCGA-STAD	Gastric	EBV, MSI, GS, CIN	Methylation, miRNA, mRNA
GSE62254	Gastric	EBV, MSI, GS, CIN	mRNA only
TCGA-LUAD, TCGA-LUSC, CPTAC	Lung	Adenocarcinoma, squamous cell carcinoma	Methylation, miRNA, mRNA
TCGA-BRCA	Breast	PAM50: Luminal A/B, Basal-like, HER2-enriched, Normal-like	Methylation, miRNA, mRNA

Shared Patient Feature Encoder

Each modality is processed by a modality-specific feed-forward feature extractor. A constant feature is appended to each embedding so that modality-specific information remains represented during fusion. The embeddings are combined by low-rank multimodal fusion,[83] which captures cross-modality interactions while reducing the computational cost of a full tensor transformation (Figure 1).

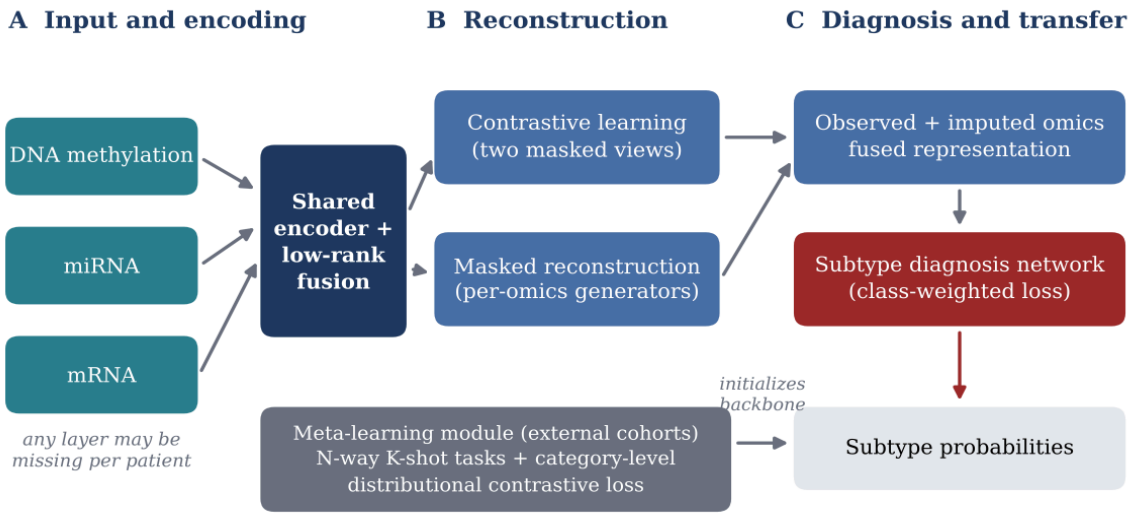


Figure 1. Conceptual architecture of the CancerMap framework. Observed molecular modalities are encoded and fused; contrastive learning and modality-specific generators support reconstruction of missing measurements; the diagnosis network predicts molecular subtype; and a meta-learning module transfers information from external cohorts to a small target cohort. This is a newly drawn conceptual schematic based on the methods description, not a reproduction of a source figure.

$$\mathbf{z} = \sum_{r=1}^{R} \mathbf{w}_r \odot \left(\backslash\text{bigodot}_{m=1}^{M} \mathbf{U}_{r,m}^{T} [\mathbf{h}_m; 1] \right)$$

Low-rank multimodal fusion (notation-explicit representation)

Here, Σ_r is the embedding for modality, U_r, is the modality-specific factor for rank component r, w_r is the corresponding output factor, and $\odot$ denotes element-wise multiplication across modality-transformed vectors. This equation expresses the low-rank fusion operation described in the source text; implementation-specific dimensions should be checked against the code.

Missing-Modality Reconstruction

Two randomly masked views of each complete patient are passed through a shared encoder and projection head. Contrastive learning encourages the two views of the same patient to have similar representations while separating representations from other patients.[85] A second objective randomly masks complete omics layers and trains modality-specific generators to reconstruct the hidden values using the remaining

modalities. For naturally incomplete patients, reconstruction loss is calculated only for modalities that were actually observed during training.

$$\mathscr{L}_{\text{con}} = -\frac{1}{2B}\sum_{i=1}^{B}\left[\log\frac{\exp(\text{sim}(u_i, v_i)/\tau)}{\sum_{j\neq i}\exp(\text{sim}(u_i, v_j)/\tau)} + \log\frac{\exp(\text{sim}(v_i, u_i)/\tau)}{\sum_{j\neq i}\exp(\text{sim}(v_i, u_j)/\tau)}\right]$$

Temperature-scaled contrastive objective (symmetric form)

B is the batch size, u_i and v_i are the projected representations of the two masked views, sim is cosine similarity, and τ is the temperature. The contrastive objective compares representations generated from two masked views of the same patient, encouraging similar representations while distinguishing different patients.

$$L_gen = (1/|\Omega|)\,\Sigma_(i,m)\in\Omega\ ||x_i - \hat{x}_i||^2 \quad (3)$$

Masked reconstruction loss

Ω contains patient–modality pairs used as reconstruction targets. The source describes mean squared error for masked modalities; exact normalization should be verified against code.

Subtype Diagnosis

Observed modalities are retained, and missing modalities are replaced with generator outputs. The integrated representation is passed to a diagnosis network that returns subtype probabilities. Class-weighted cross-entropy is used to address subtype imbalance.

$$\mathscr{L}_{\text{gen}} = \frac{1}{|\Omega|}\sum_{(i,m)\in\Omega}\left\|x_{i,m} - \widehat{x}_{i,m}\right\|^2$$

Class-weighted diagnosis loss

N is the number of training samples, C is the number of subtype classes, and n_c is the training count for class c.

$$\mathscr{L}_{\text{diag}} = -\frac{1}{N}\sum_{i=1}^{N} w_{y_i}\log p_\theta(y_i \mid x_i), \qquad w_c = \frac{N}{Cn_c}$$

Base-learner objective

The source report gives both loss weights as 1.

Cross-Cohort Knowledge Transfer

The meta-learning module follows a model-agnostic meta-learning strategy.[86,87] External cohorts are sampled into N-way K-shot tasks[88,97] with support and query sets. The inner loop adapts model parameters using support-set loss; the outer loop updates the shared initialization using query-set performance. A category-level contrastive term compares encoded subtype distributions across external and target cohorts. Kernel density estimation[89] is used to estimate distributions, and Jensen–Shannon divergence is used as a symmetric, bounded measure of distributional difference.

$$\theta'_t = \theta - \alpha\nabla_\theta\mathscr{L}_{\text{support},t}(\theta)$$

Inner-loop adaptation

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \left[\sum_{t} \mathscr{L}_{\text{query},t}(\theta'_{t}) + \lambda_{\text{CLC}} \mathscr{L}_{\text{CLC}} \right]$$

Outer-loop meta-update

$$\mathrm{JS}(P \parallel Q) = \frac{1}{2}\mathrm{KL}(P \parallel M) + \frac{1}{2}\mathrm{KL}(Q \parallel M), \quad M = \frac{1}{2}(P+Q)$$

Jensen–Shannon divergence

The category-level contrastive objective aligns representations of matching molecular subtypes across cohorts while distinguishing representations belonging to different subtypes

Experimental Design and Evaluation

For standard supervised comparisons, an 80:20 training–test split was described. Methods included k-nearest neighbors, random forest,[90] AE-XGBoost,[91] MOMA,[15] MOGONET,[14] DCP,[20] and APADC[21] in the follow-up comparisons. Missingness was simulated at 0%, 25%, 50%, and 75% of samples. Few-shot transfer used TCGA-STAD as the external cohort and GSE62254 as the target cohort, with a 4-way 10-shot target training subset. Transfer settings included multi-omics-to-mRNA and mRNA-to-mRNA. Each experiment was described as repeated 10 times, with accuracy, AUROC, precision, and class-weighted F1 averaged across runs.

A complete report should specify random seeds, stratification, sample counts, whether feature selection and scaling were fitted only on training data, handling of overlapping patients, confidence intervals, and statistical tests. These details are necessary to assess reproducibility and potential information leakage.

RESULTS

Standard Supervised Classification

CancerMap demonstrated competitive classification performance across the evaluated metrics in the TCGA-STAD. Under simulated missingness, performance declined for all methods, with larger declines described for methods relying on sample-similarity calculations. Methods designed for incomplete data were reported to retain an advantage as missingness increased. At the subtype level, EBV and CIN were described as easier to classify than GS; GS and CIN were the most frequent confusion pair.[35]

Detailed comparative estimates are not presented because the complete numerical results were not independently reproduced for this example.

Few-Shot Transfer

In the four-way 10-shot transfer setting, classification performance decreased compared with training on the full target cohort. Meta-learning approaches demonstrated better performance than pretraining-only methods. In the mRNA-to-mRNA setting, the reported CancerMap metrics were 67.5% accuracy, 83.7% AUROC, 73.3% precision, and 67.6% class-weighted F1 (Figure 2A). The figure displays only these four reported values and does not imply a comparison with methods for which exact values were not supplied.

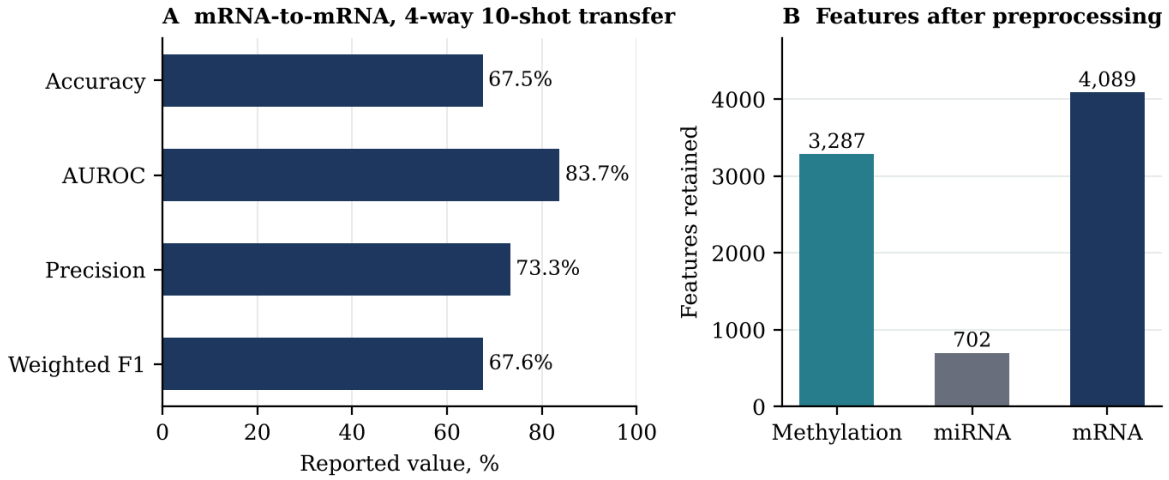


Figure 2. Reported performance and data dimensionality. (A) Performance in the mRNA-to-mRNA 4-way 10-shot transfer setting. (B) Number of features retained per molecular modality after preprocessing. Values are reproduced from the source report; no run-level uncertainty estimates were available.

Effect of Omics Type and Integration

Performance increased as additional modalities were integrated, with three modalities outperforming two and two outperforming one. Among single modalities, mRNA performed best and miRNA least well. Joint optimization of modalities was reported to outperform independent modality training followed by late fusion. These findings are presented qualitatively because complete numerical estimates and uncertainty measures were not independently reproduced for this example.

Molecular Feature Interpretation

Feature importance was defined as the increase in diagnosis loss after permuting a feature across test samples, averaged over 10 random trials. mRNA features showed higher importance than miRNA features. The top reported molecular features (Table 2) included CD9, CXCR4, KLK6, HLA-B, MUC1, MET, and several miRNAs.[36–43] These features were identified in the original study and are presented as exploratory computational findings rather than independently validated biomarkers.

Table 2. Top-ranked molecular features reported in the source report

Omics layer	Top 10 reported features
DNA methylation (gene)	CD9; C19orf51; ENC1; RHPN2; CXCR4; TRAF3IP2; ZFPM1; SLC5A5; ARHGEF7; HLA-F
mRNA expression	THEMIS; FBXL16; KLK6; HLA-B; ACE2; CFTR; CYP4F22; HAAO; MUC1; MET
miRNA expression	hsa-mir-1305; hsa-mir-5698; hsa-mir-488; hsa-mir-942; hsa-mir-219a-2; hsa-let-7e; hsa-mir-7-2; hsa-mir-1271; hsa-mir-944; hsa-mir-501

Stemness and Prognosis

The mRNA-based stemness index (mRNAsi)[57] was used to examine the relationship between subtype predictions and tumor stemness. mRNAsi differed across gastric molecular subtypes and was lowest in GS tumors.[65] A subtype score within GS correlated negatively with mRNAsi ($r = -0.353$).

An Integrated CancerMap Score (ICS) was constructed from subtype probabilities, with sigmoid-transformed scores constrained to (0, 1) and weights reflecting the reported prognostic ordering of subtypes (Figure 3A).

$$ICS = exp(ICS_raw),\ \ ICS_raw = s_CIN + (1 - s_MSI) + 2s_GS + (1 - s_EBV) \quad (9)$$

Integrated CancerMap Score

Here s_CIN, s_MSI, s_GS, and s_EBV are sigmoid-transformed subtype scores. The reported cutoff was 16.2.[68]

ICS correlated negatively with mRNAsi (r = −0.601). Overall survival differed between ICS groups at the 16.2 cutoff (P = .04), whereas disease-free survival did not (P = .16); disease-free survival data were available for 201 of 415 patients. In a multivariable Cox model, high vs low ICS was associated with overall survival (HR, 1.70; 95% CI, 1.07–2.70; P = .03; Figure 3B). The complete covariate set and univariate estimates were not available here.

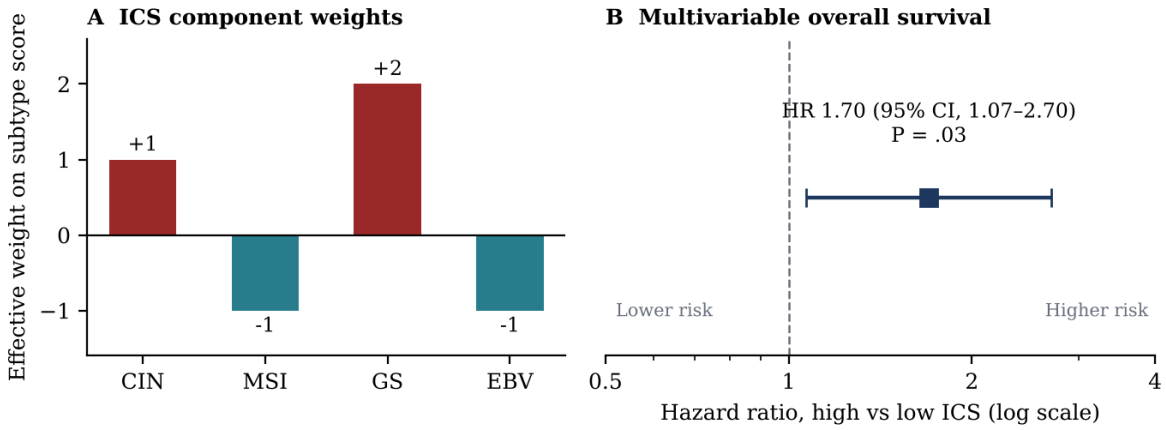


Figure 3. Integrated CancerMap Score (ICS). (A) Effective weight of each subtype score in the raw ICS formula. (B) Reported multivariable association between high vs low ICS and overall survival: HR 1.70 (95% CI, 1.07–2.70); P = .03. Panel B shows the single estimate available and is not a reconstruction of the full Cox model.

Lung and Breast Cancer

Additional evaluations in lung and breast cancer datasets demonstrated the framework's applicability beyond gastric cancer. These findings suggest that multi-omics integration and cross-cohort transfer may be useful across different cancer types.

DISCUSSION

CancerMap combines low-rank multimodal fusion, contrastive representation learning, masked reconstruction, subtype classification, and distribution-aware meta-learning. Together, these components address incomplete molecular measurements, limited labeled cohorts, and differences between external and target datasets. These components address distinct but related challenges: incomplete molecular measurements, small labeled cohorts, and differences between external and target datasets. The reported few-shot metrics and prognostic association provide hypotheses for further computational validation, but they should not be interpreted as evidence of clinical readiness.

Limitations

Several limitations are important. First, imputed values are model estimates and should not be treated as equivalent to measured molecular data. Second, randomly simulated missingness may not reflect clinical testing patterns, which can be informative rather than random. Third, cohort differences may reflect platforms, batch effects, patient selection, and annotation practices. Fourth, GS and CIN were reported to be difficult to distinguish, indicating that aggregate metrics may conceal subtype-specific weaknesses. Finally, the prognostic associations are observational and require independent validation before any clinical use.

The fusion architecture also does not incorporate prior biological relationships among genes, miRNAs, and mRNAs; incorporating such information may improve interpretability. The described transfer strategy is limited to datasets from the same cancer type; cross-cancer transfer remains a possible future direction but would require separate evaluation.

CONCLUSIONS

The CancerMap framework integrates incomplete multi-omics profiles and adapts subtype classifiers across cohorts with limited labeled samples. Benchmark performance and an association between the integrated score and overall survival were reported. Independent reproduction, external validation, transparent reporting of missingness assumptions, and confirmation of all numerical results are required before the method can support clinical decisions.

The code used to implement the CancerMap framework and reproduce the analyses described in this study is available at the CancerMap SourceForge repository: https://sourceforge.net/projects/cancer-map/

REFERENCES

1. Soerjomataram I, Bray F. Planning for tomorrow: global cancer incidence and the role of prevention 2020–2070. Nature Reviews Clinical Oncology. 2021; 18(10):663–672. https://doi.org/10.1038/s41571-021-00514-z
2. Nia HT, Munn LL, Jain RK. Physical traits of cancer. Science. 2020; 370(6516):eaaz0868. https://doi.org/10.1126/science.aaz0868
3. Kristensen VN, Lingjærde OC, Russnes HG, Vollan HKM, Frigessi A, Børresen-Dale AL. Principles and methods of integrative genomic analyses in cancer. Nature Reviews Cancer. 2014; 14(5):299–313. https://doi.org/10.1038/nrc3721
4. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell. 2014; 158(4):929–944. https://doi.org/10.1016/j.cell.2014.06.049
5. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. The Lancet. 2020; 396(10251):635–648. https://doi.org/10.1016/S0140-6736(20)31288-5
6. Park S, Hong TH, Hwang S, Heeke S, Gay CM, Kim J, et al. Comprehensive analysis of transcription factor-based molecular subtypes and their correlation to clinical outcomes in small-cell lung cancer. EBioMedicine. 2024; 102. https://doi.org/10.1016/j.ebiom.2024.105062
7. Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, et al. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. The Lancet Oncology. 2018; 19(1):40–50. https://doi.org/10.1016/S1470-2045(17)30904-X
8. Chen R, Yang L, Goodison S, Sun Y. Deep-learning approach to identifying cancer subtypes using high-dimensional genomic data. Bioinformatics. 2020; 36(5):1476–1483. https://doi.org/10.1093/bioinformatics/btz769
9. Sirinukunwattana K, Domingo E, Richman SD, Redmond KL, Blake A, Verrill C, et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. Gut. 2021; 70(3):544–554. https://doi.org/10.1136/gutjnl-2019-319866
10. Yang H, Wei Q, Li D, Wang Z. Cancer classification based on chromatin accessibility profiles with deep adversarial learning model. PLoS Computational Biology. 2020; 16(11):e1008405. https://doi.org/10.1371/journal.pcbi.1008405
11. Kang M, Ko E, Mersha TB. A roadmap for multi-omics data integration using deep learning. Briefings in Bioinformatics. 2022; 23(1):bbab454. https://doi.org/10.1093/bib/bbab454
12. Benkirane H, Pradat Y, Michiels S, Cournède PH. CustOmics: A versatile deep-learning based strategy for multi-omics integration. PLoS Computational Biology. 2023; 19(3):e1010921. https://doi.org/10.1371/journal.pcbi.1010921
13. Zheng X, Wang M, Huang K, Zhu E. Global and cross-modal feature aggregation for multi-omics data classification and application on drug response prediction. Information Fusion. 2024; 102:102077. https://doi.org/10.1016/j.inffus.2023.102077

14. Wang T, Shao W, Huang Z, Tang H, Zhang J, Ding Z, et al. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. Nature Communications. 2021; 12(1):3445. https://doi.org/10.1038/s41467-021-23774-w
15. Moon S, Lee H. MOMA: a multi-task attention learning algorithm for multi-omics data interpretation and classification. Bioinformatics. 2022; 38(8):2287–2296. https://doi.org/10.1093/bioinformatics/btac080
16. Rappoport N, Shamir R. NEMO: cancer subtyping by integration of partial multi-omic data. Bioinformatics. 2019; 35(18):3348–3356. https://doi.org/10.1093/bioinformatics/btz058
17. Rawal S, Gupta S, Singh S. Predicting missing values in a dataset: challenges and approaches. International Journal of Recent Research Aspects. 2017; 4(3):34–38.
18. Wang X, Yu G, Wang J, Zain AM, Guo W. Lung cancer subtype diagnosis using weakly-paired multi-omics data. Bioinformatics. 2022; 38(22):5092–5099. https://doi.org/10.1093/bioinformatics/btac643
19. Zhao C, Liu A, Zhang X, Cao X, Ding Z, Sha Q, et al. CLCLSA: Cross-omics linked embedding with contrastive learning and self attention for integration with incomplete multi-omics data. Computers in Biology and Medicine. 2024; 170:108058. https://doi.org/10.1016/j.compbiomed.2024.108058
20. Lin Y, Gou Y, Liu X, Bai J, Lv J, Peng X. Dual contrastive prediction for incomplete multi-view representation learning. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022; 45(4):4447–4461.
21. Xu J, Li C, Peng L, Ren Y, Shi X, Shen HT, et al. Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering. IEEE Transactions on Image Processing. 2023; 32:1354–1366. https://doi.org/10.1109/TIP.2023.3243521
22. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. Journal of the Royal Society Interface. 2018; 15(141):20170387. https://doi.org/10.1098/rsif.2017.0387
23. Ming W, Li F, Zhu Y, Bai Y, Gu W, Liu Y, et al. Predicting hormone receptors and PAM50 subtypes of breast cancer from multi-scale lesion images of DCE-MRI with transfer learning technique. Computers in Biology and Medicine. 2022; 150:106147. https://doi.org/10.1016/j.compbiomed.2022.106147
24. Mahajan K, Sharma M, Vig L. Meta-dermdiagnosis: Few-shot skin disease identification using meta-learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops; 2020. p. 730–731.
25. Qiu YL, Zheng H, Devos A, Selby H, Gevaert O. A meta-learning approach for genomic survival analysis. Nature Communications. 2020; 11(1):6350. https://doi.org/10.1038/s41467-020-20167-3
26. Pan SJ, Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering. 2009; 22(10):1345–1359. https://doi.org/10.1109/TKDE.2009.191
27. Yu T. AIME: Autoencoder-based integrative multi-omics data embedding that allows for confounder adjustments. PLoS Computational Biology. 2022; 18(1):e1009826. https://doi.org/10.1371/journal.pcbi.1009826
28. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. Biotechnology Advances. 2021; 49:107739. https://doi.org/10.1016/j.biotechadv.2021.107739
29. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2021; 71(3):209–249.
30. Jeong HY, Ham IH, Lee SH, Ryu D, Son SY, Han SU, et al. Spatially distinct reprogramming of the tumor microenvironment based on tumor invasion in diffuse-type gastric cancers. Clinical Cancer Research. 2021; 27(23):6529–6542. https://doi.org/10.1158/1078-0432.CCR-21-0792
31. Kumar V, Ramnarayanan K, Sundar R, Padmanabhan N, Srivastava S, Koiwa M, et al. Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer. Cancer Discovery. 2022; 12(3):670–691. https://doi.org/10.1158/2159-8290.CD-21-0683
32. Bian S, Wang Y, Zhou Y, Wang W, Guo L, Wen L, et al. Integrative single-cell multiomics analyses dissect molecular signatures of intratumoral heterogeneities and differentiation states of human gastric cancer. National Science Review. 2023; 10(6):nwad094. https://doi.org/10.1093/nsr/nwad094
33. Van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research. 2008; 9(11):2579–2605.
34. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014; 513(7517):202. https://doi.org/10.1038/nature13480
35. Lee JE, Kim KT, Shin SJ, Cheong JH, Choi YY. Genomic and evolutionary characteristics of metastatic gastric cancer by routes. British Journal of Cancer. 2023; 129(4):672–682. https://doi.org/10.1038/s41416-023-02338-3
36. Yasumoto K, Koizumi K, Kawashima A, Saitoh Y, Arita Y, Shinohara K, et al. Role of the CXCL12/CXCR4 axis in peritoneal carcinomatosis of gastric cancer. Cancer Research. 2006; 66(4):2181–2187. https://doi.org/10.1158/0008-5472.CAN-05-3393
37. Hashimoto I, Koizumi K, Tatematsu M, Minami T, Cho S, Takeno N, et al. Blocking on the CXCR4/mTOR signalling pathway induces the anti-metastatic properties and autophagic cell death in peritoneal disseminated gastric cancer cells. European Journal of Cancer. 2008; 44(7):1022–1029. https://doi.org/10.1016/j.ejca.2008.02.043
38. Xiang Z, Zhou Z, Xia G, Zhang X, Wei Z, Zhu J, et al. A positive crosstalk between CXCR4 and CXCR2 promotes gastric cancer metastasis. Oncogene. 2017; 36(36):5122–5133. https://doi.org/10.1038/onc.2017.108
39. Luo M, Deng X, Chen Z, Hu Y. Circular RNA circPOFUT1 enhances malignant phenotypes and autophagy-associated chemoresistance via sequestrating miR-488-3p to activate the PLAG1-ATG12 axis in gastric cancer. Cell Death & Disease. 2023; 14(1):10. https://doi.org/10.1038/s41419-022-05506-0
40. Yang D, Shi M, You Q, Zhang Y, Hu Z, Xu J, et al. Tumor-and metastasis-promoting roles of miR-488 inhibition via HULC enhancement and EZH2-mediated p53 repression in gastric cancer. Cell Biology and Toxicology. 2023; 39(4):1341–1358. https://doi.org/10.1007/s10565-022-09760-y
41. Fujita M, Furukawa Y, Tsunoda T, Tanaka T, Ogawa M, Nakamura Y. Up-regulation of the ectodermal-neural cortex 1 (ENC1) gene, a downstream target of the β-catenin/T-cell factor complex, in colorectal carcinomas. Cancer Research. 2001; 61(21):7722–7726.
42. Cui Y, Yang J, Bai Y, Li Q, Yao Y, Liu C, et al. ENC1 facilitates colorectal carcinoma tumorigenesis and metastasis via JAK2/STAT5/AKT axis-mediated epithelial mesenchymal transition and stemness. Frontiers in Cell and Developmental Biology. 2021; 9:616887. https://doi.org/10.3389/fcell.2021.616887
43. Than B, Linnekamp J, Starr T, Largaespada DA, Rod A, Zhang Y, et al. CFTR is a tumor suppressor gene in murine and human intestinal cancer. Oncogene. 2016; 35(32):4191–4199. https://doi.org/10.1038/onc.2015.483
44. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9(1):1–13. https://doi.org/10.1186/1471-2105-9-559
45. Zeng Z, Ma Y, Hu L, Liu P, Tan B, Wang Y, et al. OmicVerse: A single pipeline for exploring the entire transcriptome universe. bioRxiv. 2023; p. 2023–06.
46. Moreira AM, Pereira J, Melo S, Fernandes MS, Carneiro P, Seruca R, et al. The extracellular matrix: an accomplice in gastric cancer development and progression. Cells. 2020; 9(2):394. https://doi.org/10.3390/cells9020394
47. Wang YY, Li L, Zhao ZS, Wang YX, Ye ZY, Tao HQ. L1 and epithelial cell adhesion molecules associated with gastric cancer progression and prognosis in examination of specimens from 601 patients. Journal of Experimental & Clinical Cancer Research. 2013; 32:1–11. https://doi.org/10.1186/1756-9966-32-66
48. Liang Q, Yao X, Tang S, Zhang J, Yau TO, Li X, et al. Integrative identification of Epstein–Barr virus–associated mutations and epigenetic alterations in gastric cancer. Gastroenterology. 2014; 147(6):1350–1362. https://doi.org/10.1053/j.gastro.2014.08.036

49. Chen X, Li Y, Deng L, Wang L, Zhong W, Hong J, et al. Cardiovascular involvement in Epstein–Barr virus infection. Frontiers in Immunology. 2023; 14:1188330. https://doi.org/10.3389/fimmu.2023.1188330
50. Kim B, Byun SJ, Kim YA, Kim JE, Lee BL, Kim WH, et al. Cell cycle regulators, APC/β-catenin, NF-κB and Epstein-Barr virus in gastric carcinomas. Pathology. 2010; 42(1):58–65. https://doi.org/10.3109/00313020903356392
51. Chang MS, Kim DH, Roh JK, Middeldorp JM, Kim YS, Kim S, et al. Epstein-Barr virus-encoded BARF1 promotes proliferation of gastric carcinoma cells through regulation of NF-κB. Journal of Virology. 2013; 87(19):10515–10523. https://doi.org/10.1128/JVI.00955-13
52. Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JI, Kim K, et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. Nature Medicine. 2018; 24(9):1449–1458. https://doi.org/10.1038/s41591-018-0101-z
53. Tokunaga R, Zhang W, Naseem M, Puccini A, Berger MD, Soni S, et al. CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation–a target for novel cancer therapy. Cancer Treatment Reviews. 2018; 63:40–47. https://doi.org/10.1016/j.ctrv.2017.11.007
54. Griffith JW, Sokol CL, Luster AD. Chemokines and chemokine receptors: positioning cells for host defense and immunity. Annual Review of Immunology. 2014; 32:659–702. https://doi.org/10.1146/annurev-immunol-032713-120145
55. Jing Jj, Wang Zy, Li H, Sun Lp, Yuan Y. Key elements involved in Epstein–Barr virus-associated gastric cancer and their network regulation. Cancer Cell International. 2018; 18(1):1–12. https://doi.org/10.1186/s12935-018-0637-5
56. Xia T, O'Hara A, Araujo I, Barreto J, Carvalho E, Sapucaia JB, et al. EBV microRNAs in primary lymphomas and targeting of CXCL-11 by ebv-mir-BHRF1-3. Cancer Research. 2008; 68(5):1436–1442. https://doi.org/10.1158/0008-5472.CAN-07-5126
57. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell. 2018; 173(2):338–354. https://doi.org/10.1016/j.cell.2018.03.034
58. Nallasamy P, Nimmakayala RK, Parte S, Are AC, Batra SK, Ponnusamy MP. Tumor microenvironment enriches the stemness features: The architectural event of therapy resistance and metastasis. Molecular Cancer. 2022; 21(1):1–25. https://doi.org/10.1186/s12943-022-01682-x
59. Yang L, Shi P, Zhao G, Xu J, Peng W, Zhang J, et al. Targeting cancer stem cell pathways for cancer therapy. Signal Transduction and Targeted Therapy. 2020; 5(1):8. https://doi.org/10.1038/s41392-020-0110-5
60. Salomonis N, Dexheimer PJ, Omberg L, Schroll R, Bush S, Huo J, et al. Integrated genomic analysis of diverse induced pluripotent stem cells from the progenitor cell biology consortium. Stem Cell Reports. 2016; 7(1):110–125. https://doi.org/10.1016/j.stemcr.2016.05.006
61. Daily K, Ho Sui SJ, Schriml LM, Dexheimer PJ, Salomonis N, Schroll R, et al. Molecular, phenotypic, and sample-associated data to describe pluripotent stem cell lines and derivatives. Scientific Data. 2017; 4(1):1–10. https://doi.org/10.1038/sdata.2017.30
62. Yi L, Huang P, Zou X, Guo L, Gu Y, Wen C, et al. Integrative stemness characteristics associated with prognosis and the immune microenvironment in esophageal cancer. Pharmacological Research. 2020; 161:105144. https://doi.org/10.1016/j.phrs.2020.105144
63. Chen D, Liu J, Zang L, Xiao T, Zhang X, Li Z, et al. Integrated machine learning and bioinformatic analyses constructed a novel stemness-related classifier to predict prognosis and immunotherapy responses for hepatocellular carcinoma patients. International Journal of Biological Sciences. 2022; 18(1):360. https://doi.org/10.7150/ijbs.66913
64. Mao D, Xu R, Chen H, Chen X, Li D, Song S, et al. Cross-talk of focal adhesion-related gene defines prognosis and the immune microenvironment in gastric cancer. Frontiers in Cell and Developmental Biology. 2021; 9:716461. https://doi.org/10.3389/fcell.2021.716461
65. Sohn BH, Hwang JE, Jang HJ, Lee HS, Oh SC, Shim JJ, et al. Clinical significance of four molecular subtypes of gastric cancer identified by the cancer genome atlas project. Clinical Cancer Research. 2017; 23(15):4441–4449. https://doi.org/10.1158/1078-0432.CCR-16-2211
66. Guo R, Chu A, Gong Y. Identification of cancer stem cell-related biomarkers in intestinal-type and diffuse-type gastric cancer by stemness index and weighted correlation network analysis. Journal of Translational Medicine. 2020; 18(1):1–14. https://doi.org/10.1186/s12967-020-02587-3
67. Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma: an attempt at a histo-clinical classification. Acta Pathologica Microbiologica Scandinavica. 1965; 64(1):31–49. https://doi.org/10.1111/apm.1965.64.1.31
68. Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. Computational Statistics & Data Analysis. 2003; 43(2):121–137. https://doi.org/10.1016/S0167-9473(02)00225-6
69. Gu Y, Chen Y, Wei L, Wu S, Shen K, Liu C, et al. ABHD5 inhibits YAP-induced c-Met overexpression and colon cancer cell stemness via suppressing YAP methylation. Nature Communications. 2021; 12(1):6711. https://doi.org/10.1038/s41467-021-26967-5
70. Li C, Wu JJ, Hynes M, Dosch J, Sarkar B, Welling TH, et al. c-Met is a marker of pancreatic cancer stem cells and therapeutic target. Gastroenterology. 2011; 141(6):2218–2227. https://doi.org/10.1053/j.gastro.2011.08.009
71. Li Y, Li A, Glas M, Lal B, Ying M, Sang Y, et al. c-Met signaling induces a reprogramming network and supports the glioblastoma stem-like phenotype. Proceedings of the National Academy of Sciences. 2011; 108(24):9951–9956. https://doi.org/10.1073/pnas.1016912108
72. Yashiro M, Nishii T, Hasegawa T, Matsuzaki T, Morisaki T, Fukuoka T, et al. A c-Met inhibitor increases the chemosensitivity of cancer stem cells to the irinotecan in gastric carcinoma. British Journal of Cancer. 2013; 109(10):2619–2628. https://doi.org/10.1038/bjc.2013.638
73. Bahrami A, Shahidsales S, Khazaei M, Ghayour-Mobarhan M, Maftouh M, Hassanian SM, et al. C-Met as a potential target for the treatment of gastrointestinal cancer: Current status and future perspectives. Journal of Cellular Physiology. 2017; 232(10):2657–2673. https://doi.org/10.1002/jcp.25794
74. Li D, Liu X, Liu T, Liu H, Tong L, Jia S, et al. Neurochemical regulation of the expression and function of glial fibrillary acidic protein in astrocytes. Glia. 2020; 68(5):878–897. https://doi.org/10.1002/glia.23734
75. Vazquez-Santillan K, Melendez-Zajgla J, Jimenez-Hernandez L, Martinez-Ruiz G, Maldonado V. NF-κB signaling in cancer stem cells: a promising therapeutic target? Cellular Oncology. 2015; 38:327–339. https://doi.org/10.1007/s13402-015-0236-6
76. Fu T, Ji K, Jin L, Zhang J, Wu X, Ji X, et al. ASB16-AS1 up-regulated and phosphorylated TRIM37 to activate NF-κB pathway and promote proliferation, stemness, and cisplatin resistance of gastric cancer. Gastric Cancer. 2021; 24:45–59. https://doi.org/10.1007/s10120-020-01096-y
77. Ding C, Tang W, Wu H, Fan X, Luo J, Feng J, et al. The PEAK1–PPP1R12B axis inhibits tumor growth and metastasis by regulating Grb2/PI3K/Akt signalling in colorectal cancer. Cancer Letters. 2019; 442:383–395. https://doi.org/10.1016/j.canlet.2018.11.014
78. Chao X, Zao J, Xiao-Yi G, Li-Jun M, Tao S. Blocking of PI3K/AKT induces apoptosis by its effect on NF-κB activity in gastric carcinoma cell line SGC7901. Biomedicine & Pharmacotherapy. 2010; 64(9):600–604. https://doi.org/10.1016/j.biopha.2010.08.008
79. Whiteaker JR, Halusa GN, Hoofnagle AN, Sharma V, MacLean B, Yan P, et al. CPTAC Assay Portal: a repository of targeted proteomic assays. Nature Methods. 2014; 11(7):703–704. https://doi.org/10.1038/nmeth.3002
80. Edwards NJ, Oberti M, Thangudu RR, Cai S, McGarvey PB, Jacob S, et al. The CPTAC data portal: a resource for cancer proteomics research. Journal of Proteome Research. 2015; 14(6):2707–2713. https://doi.org/10.1021/pr501254j
81. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of Clinical Oncology. 2009; 27(8):1160. https://doi.org/10.1200/JCO.2008.18.1370
82. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell. 2018; 173(2):291–304. https://doi.org/10.1016/j.cell.2018.03.022

83. Liu Z, Shen Y, Lakshminarasimhan VB, Liang PP, Zadeh AB, Morency LP. Efficient Low-rank Multimodal Fusion With Modality-Specific Factors. In: Proceedings of Annual Meeting of the Association for Computational Linguistics; 2018. p. 2247–2256.
84. Yao Y, Lv Y, Tong L, Liang Y, Xi S, Ji B, et al. ICSDA: a multi-modal deep learning model to predict breast cancer recurrence and metastasis risk by integrating pathological, clinical and gene expression data. Briefings in Bioinformatics. 2022; 23(6):bbac448. https://doi.org/10.1093/bib/bbac448
85. Chen T, Sun Y, Shi Y, Hong L. On Sampling Strategies for Neural Network-Based Collaborative Filtering. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2017. p. 767–776.
86. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning; 2017. p. 1126–1135.
87. Hospedales T, Antoniou A, Micaelli P, Storkey A. Meta-learning in neural networks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2021; 44(9):5149–5169.
88. Zintgraf L, Shiarli K, Kurin V, Hofmann K, Whiteson S. Fast context adaptation via meta-learning. In: International Conference on Machine Learning; 2019. p. 7693–7702.
89. Weglarczyk S. Kernel density estimation and its application. ITM Web of Conferences. 2018; 23:00037. https://doi.org/10.1051/itmconf/20182300037
90. Breiman L. Random forests. Machine Learning. 2001; 45:5–32. https://doi.org/10.1023/A:1017934522171
91. Ma B, Meng F, Yan G, Yan H, Chai B, Song F. Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data. Computers in Biology and Medicine. 2020; 121:103761. https://doi.org/10.1016/j.compbiomed.2020.103761
92. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biology. 2020; 21:1–17. https://doi.org/10.1186/s13059-020-02015-1
93. Liang PP, Deng Z, Ma MQ, Zou JY, Morency LP, Salakhutdinov R. Factorized contrastive learning: Going beyond multi-view redundancy. Advances in Neural Information Processing Systems. 2024; 36.
94. Bardes A, Ponce J, Lecun Y. VICReg: Variance-Invariance-Covariance Regularization For Self-Supervised Learning. In: ICLR 2022 International Conference on Learning Representations; 2022.
95. Yang H, Chen R, Li D, Wang Z. Subtype-GAN: a deep learning approach for integrative cancer subtyping of multi-omics data. Bioinformatics. 2021; 37(16):2231–2237. https://doi.org/10.1093/bioinformatics/btab109
96. Du JH, Cai Z, Roeder K. Robust probabilistic modeling for single-cell multimodal mosaic integration and imputation via scVAEIT. Proceedings of the National Academy of Sciences. 2022; 119(49):e2214414119. https://doi.org/10.1073/pnas.2214414119
97. Vinyals O, Blundell C, Lillicrap T, Wierstra D, et al. Matching networks for one shot learning. Advances in Neural Information Processing Systems. 2016; 29.
98. Wang X, Wang X, Jiang B, Luo B. Few-Shot Learning Meets Transformer: Unified Query-Support Transformers for Few-Shot Classification. IEEE Transactions on Circuits and Systems for Video Technology. 2023. https://doi.org/10.1109/TCSVT.2023.3282777
99. Xie J, Long F, Lv J, Wang Q, Li P. Joint distribution matters: Deep brownian distance covariance for few-shot classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022. p. 7972–7981.