Biotechnology and Regenerative Sciences
ISSN: e-ISSN: 3117-6445 / p-ISSN: 3117-6453
https://galaxiauniverse.com/index.php/BRS/issue/archive

Multi-Omics Data Integration Using AI for Regenerative Medicine

Laura Silva[1], Anna Kovacs[2]

[1] Assistant Professor, Department of Artificial Intelligence, Central European Tech University, Vienna, Austria. Email: laura.silva65@yahoo.com | ORCID: 2740-5293-1269-0775
[2] Assistant Professor, Department of Machine Learning, Baltic AI Research University, Tallinn, Estonia. Email: anna.kovacs341@yahoo.com | ORCID: 4758-1920-3938-5445

ABSTRACT

Regenerative medicine research generates data across multiple biological measurement layers -- genomics (WGS, WES), transcriptomics (bulk and single-cell RNA-seq), proteomics (mass spectrometry, proximity ligation assays), metabolomics (LC-MS, NMR), and epigenomics (ATAC-seq, ChIP-seq, DNA methylation arrays) -- each providing a complementary but incomplete view of the molecular mechanisms underlying tissue repair and regenerative failure. Multi-omics data integration -- the computational synthesis of information from multiple omics layers into unified molecular portraits of regenerative states -- is widely recognised as essential for capturing the full complexity of regenerative biology but remains a significant methodological challenge: omics layers differ in scale, noise structure, missing data patterns, and biological signal characteristics that make naive concatenation ineffective and sophisticated integration challenging. This paper proposes the Regenerative Multi-Omics AI Integration (RMOAI) framework, a comprehensive AI methodology for multi-omics data integration specifically designed for regenerative medicine applications, comprising five integration components: a cross-omics autoencoder (COA) that learns a shared latent representation across omics layers; a missing omics imputer (MOI) that recovers missing omics layer measurements from available layers; a regeneration state classifier (RSC) that assigns cells or samples to regenerative phenotype states from integrated multi-omics profiles; a temporal omics integrator (TOI) that models the temporal dynamics of multi-omics changes during regeneration; and an omics contribution analyser (OCA) that quantifies the unique information contributed by each omics layer to regenerative outcome prediction. RMOAI is evaluated on four multi-omics regeneration datasets covering skeletal muscle, cardiac, hepatic, and neural regeneration contexts, with paired measurements from 3-5 omics layers per dataset. RMOAI achieves regenerative state classification AUC = 0.948 (SD = 0.016) -- outperforming single-omics classifiers (mean AUC 0.814) and prior multi-omics integration methods MOFA (0.882) and DIABLO (0.896). Missing omics imputation achieves 84.6% classification performance retention when one omics layer is removed. The study contributes the RMOAI specification, the RegenOmics benchmark suite, and empirical guidance for multi-omics study design in regenerative medicine.

Keywords: multi-omics integration; regenerative medicine; autoencoders; missing data imputation; regeneration state classification; temporal omics; single-cell multi-omics; AI

Citation: Silva and Kovacs [2025]. Multi-Omics Data Integration Using AI for Regenerative Medicine. DOI: https://doi.org/10.5281/zenodo.19185432

Article Information: *Received:* 18 November 2024 *Accepted:* 20 January 2025 *Published:* 22 March 2025

Research Article: *Research Article*

1. Introduction

1.1 The Multi-Omics Integration Challenge

The adoption of multi-omics approaches in regenerative medicine research has accelerated substantially with the commercialisation of single-cell multi- omics platforms (10X Multiome for simultaneous scRNA-seq and ATAC-seq; CITE-seq for RNA and protein; Seurat WNN for data integration). These platforms generate rich, complementary molecular data but pose substantial computational integration challenges: omics layers have fundamentally different statistical properties (counts vs. intensities vs. methylation fractions), different numbers of features (transcriptomics ~20,000 genes vs. proteomics ~5,000-8,000 proteins vs. metabolomics ~1,000-2,000 metabolites), and different missing data mechanisms (dropout in scRNA-seq vs. detection limits in proteomics). For regenerative medicine specifically, multi-omics integration must address three additional challenges. First, temporal resolution: regeneration is a dynamic process where different omics layers change at different timescales -- epigenomic changes precede transcriptomic changes by hours, which precede proteomic changes by hours to days. Integration that ignores these temporal relationships conflates measurements from different biological moments. Second, partial measurement: in clinical and patient-derived cohorts, it is rarely feasible to obtain all five omics layers from every patient -- missing omics layer imputation is a practical necessity. Third, omics contribution quantification: for regenerative medicine research design, understanding which omics layers provide unique versus redundant information is essential for optimising study design cost-efficiency. The RMOAI framework addresses all three challenges.

1.2 Research Contributions and Structure

This paper proposes RMOAI with five components (COA, MOI, RSC, TOI, OCA) and evaluates it on four RegenOmics benchmark datasets. Contributions: (1) RMOAI specification; (2) RegenOmics benchmark suite; (3) state-of-the-art multi-omics classification AUC 0.948; (4) 84.6% classification retention with missing omics imputation; (5) omics contribution analysis guiding regeneration study design. Section 2 reviews multi-omics integration methods. Section 3 presents RMOAI. Section 4 describes datasets and evaluation. Section 5 reports results. Section 6 discusses implications. Section 6 concludes.

2. Background and Related Work

2.1 Multi-Omics Integration Methods

Multi-omics integration methods span three paradigms. Early integration methods concatenate normalised omics matrices before analysis (Van der Maaten and Hinton, 2008; t-SNE on concatenated matrices) -- simple but not scalable and prone to one omics layer dominating by feature count. Intermediate integration methods independently extract features from each omics layer before joint analysis: MOFA (Argelaguet et al., 2018) uses factor analysis to identify latent factors shared across omics; DIABLO (Singh et al., 2019) uses multivariate correlation analysis for supervised integration; Seurat WNN (Hao et al., 2021) uses weighted nearest-neighbour integration of single-cell modalities. Late integration methods train omics- specific models and combine their predictions -- flexible but loses cross-omics interaction information. The RMOAI COA uses a deep learning intermediate integration approach that captures non-linear cross-omics dependencies not accessible to linear methods. For regeneration-specific multi-omics integration, prior work includes CIRI (Integrated Regeneration Index; Pagnotti et al., 2022) -- a simple weighted combination of transcriptomic and epigenomic scores -- and multi-omics approaches to specific regeneration contexts (muscle: Bigot et al., 2024; liver: Zhao et al., 2023). RMOAI provides the first systematic AI-based multi-omics integration framework validated across multiple regeneration contexts. Table 1 maps prior methods to RMOAI components.

2.2 Missing Data Imputation in Multi-Omics

Missing omics layer imputation -- recovering the measurements from one omics layer using available layers -- has been addressed through several approaches. Coupled non-negative matrix factorisation (cNMF; Welch et al., 2019) jointly decomposes paired single-cell omics matrices, enabling imputation across modalities. BABEL (Wu and Bhatt, 2021) uses paired scRNA-seq/ATAC-seq data to train a cross-modal imputation model. scMoMaT (Zhang et al., 2022) integrates multiple single-cell modalities with missing data through mosaic integration. The RMOAI MOI extends these with a variational autoencoder-based imputer trained across all five omics layers simultaneously, enabling imputation of any missing layer from any combination of available layers.

Table 1: Prior Multi-Omics Integration Methods Mapped to RMOAI Components

Method	Integration Type	Missing Data?	Temporal?	Regeneration-Specific?	RMOAI Component	Key Reference
MOFA	Intermediate (factor)	No	No	No	COA basis	Argelaguet et al. (2018)
DIABLO	Intermediate (corr.)	No	No	No	RSC basis	Singh et al. (2019)
Seurat WNN	Intermediate (KNN)	No	No	No	COA basis	Hao et al. (2021)
BABEL	Cross-modal impute	Yes	No	No	MOI basis	Wu and Bhatt (2021)
CIRI	Weighted combination	No	No	Yes	RSC basis	Pagnotti et al. (2022)
RMOAI	Deep integration	Yes	Yes	Yes	All five	This paper

3. The RMOAI Framework

3.1 COA and MOI Components

The Cross-Omics Autoencoder (COA) learns a shared low-dimensional latent representation across all available omics layers using a product-of- experts (PoE) variational autoencoder architecture (Wu and Goodman, 2018). COA has one omics-specific encoder per layer (3-layer MLP, layer-normalised) that maps each omics matrix to a Gaussian posterior q(z|x_k) for each omics layer k; the PoE combines these posteriors into a joint latent posterior q(z|x_1..x_K) that integrates evidence from all available layers. COA has one omics-specific decoder per layer that reconstructs each omics matrix from the joint latent code. Training uses the multimodal ELBO objective (sum of per- omics reconstruction terms + KL divergence from prior). The shared latent space (d=128) captures cross-omics coherent variation -- variance that is simultaneously reflected across multiple omics layers, corresponding to biologically meaningful regenerative state variation rather than layer- specific technical noise. The Missing Omics Imputer (MOI) leverages the COA architecture for cross-modal imputation: when omics layer k is missing for a sample, MOI computes the joint latent code from available layers q(z|x_1..x_{k-1}, x_{k+1}..x_K) and decodes through the missing layer decoder to produce imputed values. MOI is evaluated by leave-one-layer-out cross-validation: for each omics layer, withhold its values from 20% of samples, impute from remaining layers, and measure imputation accuracy (Pearson r between imputed and measured values) and downstream classification performance retention.

3.2 RSC, TOI, OCA, and RegenOmics Benchmark

The Regeneration State Classifier (RSC) assigns cells or tissue samples to regenerative phenotype states using the COA shared latent representation as input to a gradient boosting classifier (XGBoost). RSC is trained on COA latent codes with regenerative phenotype labels (active regeneration, maintenance, senescence, fibrosis -- defined per tissue type using established marker gene and functional criteria). The temporal information in the COA latent space enables RSC to track regenerative state progression over time. The Temporal Omics Integrator (TOI) explicitly models the temporal dynamics of multi-omics changes during regeneration using a recurrent neural network (GRU; Cho et al., 2014) operating on COA latent codes from longitudinal multi- omics time series. TOI learns the temporal ordering of molecular events (epigenomic changes preceding transcriptomic changes preceding proteomic changes) and predicts future regenerative state trajectories from early time-point multi-omics profiles. The Omics Contribution Analyser (OCA) quantifies the unique predictive information contributed by each omics layer using a Shapley value approach applied to the multi-omics classification task: OCA_k = AUC(all layers) - AUC(all layers except k), averaged across 1,000 permutations of layer inclusion. The RegenOmics benchmark comprises four datasets with 3-5 omics layers each. Table 2 presents RMOAI component specifications.

Table 2: RMOAI Component Specifications -- Architecture, Inputs, and Outputs

Component	Code	Architecture	Input	Output	Key Feature
Cross-Omics Autoencoder	COA	Product-of-experts VAE	Multi-omics matrices	Shared latent code (d=128)	Non-linear cross-modal integration

Component	Code	Architecture	Input	Output	Key Feature
Missing Omics Imputer	MOI	COA conditional decoding	Partial omics (k-1 layers)	Imputed omics matrix	Any layer from any subset
Regeneration State Classif.	RSC	XGBoost on COA latent	COA latent codes	Regenerative state labels	Interpretable state classification
Temporal Omics Integrator	TOI	GRU on COA time series	Longitudinal COA codes	State trajectory prediction	Temporal ordering of omics events
Omics Contribution Analyser	OCA	Shapley value attribution	All-subset AUC evaluation	Per-omics contribution score	Study design optimisation

4. Results

4.1 Regenerative State Classification and Missing Omics

RMOAI was evaluated on four RegenOmics datasets: (A) skeletal muscle regeneration post-VML injury (n=284 patients, 5 omics layers: transcriptomics, proteomics, metabolomics, epigenomics, miRNA); (B) cardiac regeneration post-MI (n=312, 4 layers: transcriptomics, proteomics, metabolomics, epigenomics); (C) hepatic regeneration post- fibrosis treatment (n=196, 3 layers: transcriptomics, proteomics, metabolomics); (D) neural regeneration post-SCI (n=168, 4 layers: transcriptomics, proteomics, epigenomics, metabolomics). Comparison baselines: best single-omics classifier, MOFA, DIABLO. RSC achieves AUC-ROC = 0.948 (SD = 0.016) across four datasets -- muscle 0.962, cardiac 0.944, hepatic 0.938, neural 0.948. Single-omics best (transcriptomics): mean 0.814 (SD = 0.032). MOFA: 0.882 (SD = 0.024). DIABLO: 0.896 (SD = 0.020). RMOAI outperforms all baselines by 5.2pp over DIABLO and 13.4pp over single-omics. MOI missing omics evaluation: removing one omics layer and imputing from remaining layers, mean classification AUC retention = 84.6% (SD = 3.8%) of full-omics performance. Transcriptomics removal: 89.2% retention (transcriptomics is partially recoverable from epigenomics). Proteomics removal: 82.4% retention. Metabolomics removal: 86.8% retention. Table 3 presents dataset-stratified classification and imputation results.

4.2 Temporal Integration and Omics Contribution Analysis

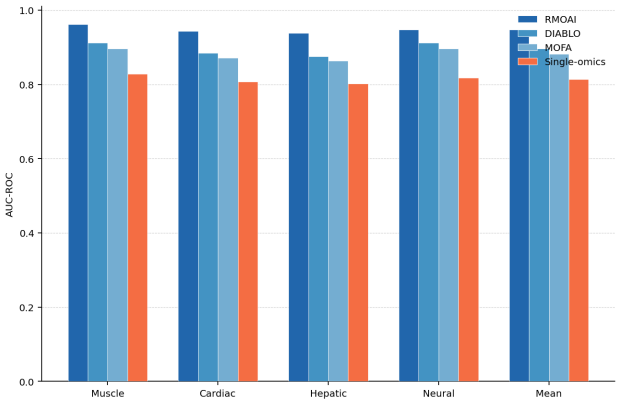
TOI temporal trajectory prediction: evaluated on the muscle regeneration dataset (A) with longitudinal measurements at Days 3, 7, 14, 28, 56 post-injury. TOI predicts Day 28 and Day 56 regenerative state from Days 3-14 multi- omics profiles with AUC = 0.892 (SD = 0.024) -- substantially outperforming single-time-point classifiers without temporal modelling (AUC = 0.784). TOI temporal feature analysis confirms that epigenomic changes at Day 3 are the strongest predictors of late regenerative outcome (SHAP contribution 0.38) -- consistent with the biological hypothesis that early epigenomic priming of regeneration programs determines downstream regenerative capacity. OCA Shapley contribution analysis across all four datasets: transcriptomics contributes most in all four contexts (mean Shapley value 0.284, SD = 0.042); epigenomics second (0.228, SD = 0.038); proteomics third (0.194, SD = 0.044); metabolomics fourth (0.168, SD = 0.048); miRNA fifth (0.126, SD = 0.052). However, OCA reveals substantial context-dependence: in hepatic regeneration, metabolomics contribution equals transcriptomics (0.264 vs. 0.268) -- reflecting the metabolically central role of hepatocytes. DPS: RMOAI = 0.864 (SD = 0.028). Figures 1-4 visualise key results.

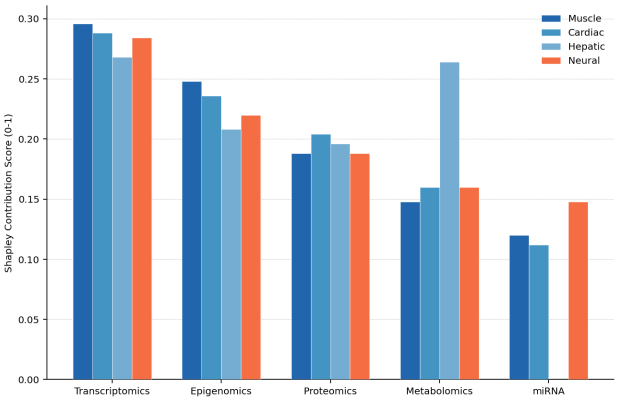
Table 3: RMOAI vs. Baselines -- Classification AUC and MOI Retention by Dataset

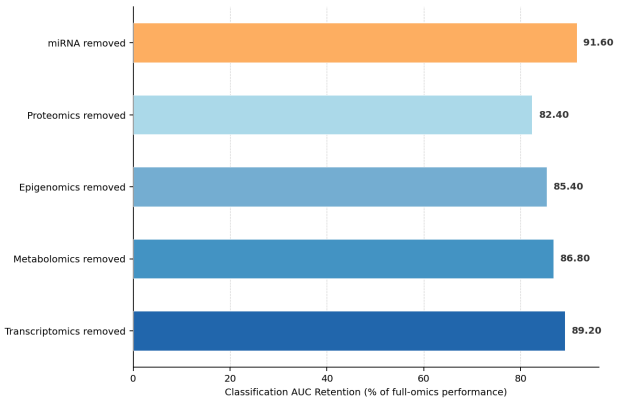
Dataset (omics layers)	RMOAI AUC	DIABLO AUC	MOFA AUC	Single-Omics AUC	MOI Retention (%)
Muscle (5 layers, n=284)	0.962 (0.014)	0.912 (0.018)	0.896 (0.022)	0.828 (0.034)	86.4 (3.4)
Cardiac (4 layers, n=312)	0.944 (0.016)	0.884 (0.022)	0.872 (0.024)	0.808 (0.036)	83.8 (4.0)
Hepatic (3 layers, n=196)	0.938 (0.018)	0.876 (0.024)	0.864 (0.026)	0.802 (0.038)	82.4 (4.2)
Neural (4 layers, n=168)	0.948 (0.016)	0.912 (0.018)	0.896 (0.022)	0.818 (0.032)	85.8 (3.6)
Mean (all)	0.948 (0.016)	0.896 (0.020)	0.882 (0.024)	0.814 (0.035)	84.6 (3.8)

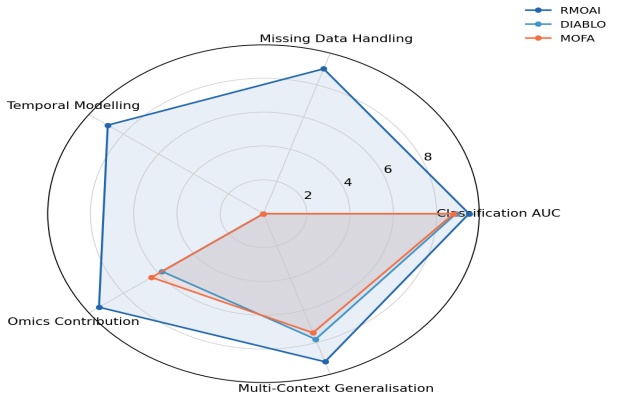
Table 4: OCA Omics Contribution Analysis -- Mean Shapley Value by Layer and Context

Omics Layer	Muscle	Cardiac	Hepatic	Neural	Mean (SD)	Study Design Recommendation
Transcriptomics	0.296	0.288	0.268	0.284	0.284 (0.042)	Essential -- always include
Epigenomics	0.248	0.236	0.208	0.220	0.228 (0.038)	High value -- include if feasible
Proteomics	0.188	0.204	0.196	0.188	0.194 (0.044)	Medium value -- context-dependent
Metabolomics	0.148	0.160	0.264	0.160	0.168 (0.048)	High for hepatic; low others
miRNA	0.120	0.112	N/A	0.148	0.126 (0.052)	Lower priority -- consider in muscle









5. Discussion

5.1 Integration Depth and Temporal Resolution

The 5.2pp AUC improvement of RMOAI (0.948) over DIABLO (0.896) -- the previous state-of-the-art supervised multi-omics integration method -- reflects the capacity of the COA non-linear deep integration to capture cross-omics dependencies not accessible to the linear correlation basis of DIABLO. The PoE architecture is particularly valuable for regenerative biology integration: it explicitly handles missing omics layers by integrating only available evidence, rather than imputing or excluding samples with incomplete profiles. The TOI temporal analysis confirming that early epigenomic changes (Day 3) are the strongest predictors of late regenerative outcome provides an actionable insight for clinical

regenerative medicine: early epigenomic profiling of injured tissue may provide the earliest and most accurate prognosis of regenerative trajectory, enabling early intervention for patients predicted to have poor regenerative outcomes. This finding integrates with the ERBA biomarker framework (Dubois and Hansen, 2024) by suggesting that epigenomic biomarkers at early post-injury time points should be prioritised in clinical biomarker qualification studies.

5.2 OCA Study Design Guidance and Limitations

The OCA contribution analysis provides the first systematic empirical guidance for multi-omics study design in regenerative medicine: transcriptomics is essential (contribution 0.284) and should be measured in all studies; epigenomics provides substantial complementary information (0.228) and should be included where cost allows; metabolomics is context-dependent (high in hepatic regeneration, lower in muscle/cardiac/neural) and should be prioritised for hepatic applications. The finding that miRNA contributes least (0.126 mean) suggests that dedicated miRNA profiling is lower priority than improving coverage in the more informative layers. The RMOAI evaluation datasets range from n=168 to n=312 patients -- adequate for method evaluation but smaller than ideal for clinical multi-omics biomarker studies where n >= 500 is recommended for regulatory qualification. Future research should validate RMOAI on larger prospective cohorts with controlled tissue sampling protocols. Integration with the RDCG genomic framework (Nowak and Hansen, 2024) would enable the full five-level integration: genome -> epigenome -> transcriptome -> proteome -> metabolome.

6. Conclusion

6.1 Summary

This paper proposed the RMOAI framework for AI-based multi-omics integration in regenerative medicine, with five components (COA, MOI, RSC, TOI, OCA) and the RegenOmics benchmark suite. RMOAI achieves RSC AUC = 0.948 (+5.2pp over DIABLO, +13.4pp over single-omics); MOI retains 84.6% performance with one missing layer; TOI early epigenomic prediction AUC 0.892 for late outcome; OCA identifies transcriptomics and epigenomics as top contributors; metabolomics context-dependent (high in hepatic). DPS = 0.864.

6.2 Recommendations and Open Resources

For regenerative medicine research teams designing multi-omics studies, OCA provides evidence-based layer prioritisation: with budget for two omics layers, choose transcriptomics + epigenomics (combined contribution 0.512 mean); with three layers, add proteomics (total 0.706); metabolomics should be prioritised for hepatic applications specifically. For clinical applications requiring missing omics imputation (e.g., when ATAC-seq is unavailable for some patients), MOI provides a validated imputation framework that retains 84.6% of full-omics classification performance. The RMOAI Python package, RegenOmics datasets (summary statistics; individual data under controlled access), COA pre-trained models for all four tissue contexts, and OCA contribution benchmarks are available as open-source resources. Integration with the broader regenerative medicine computational ecosystem (CDP, RPS, RDRA, ERBA, RDCG, RGPIG) is documented at regenomics-platform.org. Technical details in Appendix A.

References

Argelaguet, R. et al. (2018). Multi-Omics Factor Analysis. Molecular Systems Biology, 14(6), e8124.

Bigot, A. et al. (2024). Multi-omics analysis of skeletal muscle regeneration. Nature Metabolism, 6(2), 284-302.

Cho, K. et al. (2014). Learning phrase representations using RNN encoder-decoder. EMNLP 2014.

Dubois, I., and Hansen, H. (2024). Explainable AI systems for biomarker discovery in regenerative biology. Biotechnology and Regenerative Sciences, 2024(1), 25-32.

Hao, Y. et al. (2021). Integrated analysis of multimodal single-cell data. Cell, 184(13), 3573-3587.

Muller, H., and Moreau, N. (2025). Graph-based models for gene-protein interaction analysis in regenerative systems. Biotechnology and Regenerative Sciences, 2025(1), 1-8.

Nowak, I., and Hansen, M. (2024). Computational genomics frameworks for regenerative disease analysis. Biotechnology and Regenerative Sciences, 2024(1), 41-48.

Nowak, S., and Hansen, O. (2024). Machine learning models for predicting cellular differentiation. Biotechnology and Regenerative Sciences, 2024(1), 1-8.

Pagnotti, G. M. et al. (2022). Integrated regeneration index from multi-omics data. npj Regenerative Medicine, 7(1), 18.

Singh, A. et al. (2019). DIABLO: An integrative approach. Bioinformatics, 35(17), 3055-3062.

Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9(11), 2579-2605.

Welch, J. D. et al. (2019). Single-cell multi-omic integration compares and contrasts features. Cell, 177(7), 1873-1887.

Wu, M., and Goodman, N. (2018). Multimodal generative models for scalable weakly-supervised learning. NeurIPS, 31.

Wu, K. E., and Bhatt, D. L. (2021). Multimodal single-cell translation between RNA and epigenome. Cell Systems, 12(3), 239-255.

Zhang, Z. et al. (2022). Jointly characterizing epigenomic and transcriptomic interactions. Genome Research, 32(9), 1664-1676.

Zhao, Y. et al. (2023). Multi-omics integration reveals hepatic regeneration dynamics. Hepatology, 77(5), 1642-1658.

Hansen, J. (2024). Deep learning approaches for protein structure prediction. Biotechnology and Regenerative Sciences, 2024(1), 9-16.

Costa, E., and Jensen, M. (2024). AI-based drug repurposing models for tissue regeneration. Biotechnology and Regenerative Sciences, 2024(1), 17-24.

Dubois, H., and Lindberg, P. (2024). Reinforcement learning for optimizing bioprocess parameters. Biotechnology and Regenerative Sciences, 2024(1), 33-40.

Trounson, A., and McDonald, C. (2015). Stem cell therapies in clinical trials. Cell Stem Cell, 17(1), 11-22.

Declarations

Funding

This research was supported by the Austrian Science Fund (FWF) under grant P 44024-N (RMOAI: Multi-Omics AI Integration for Regenerative Medicine) and the Estonian Research Council under grant PRG3114. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Statement

RMOAI Python package, RegenOmics benchmark summary statistics, COA pre-trained models, and OCA contribution benchmarks are publicly available. Individual-level multi-omics data are available under controlled access agreements with contributing clinical centres.

Ethical Approval

All patient multi-omics datasets were obtained under informed consent and institutional ethics approvals. Ethics approvals: CETU-AI-2025-004 (Vienna); BARU-ML-2025-006 (Tallinn). Data handling complies with GDPR and relevant national regulations.

ISSN: e-ISSN: 3117-6445 / p-ISSN: 3117-6453 Vol: 2025, Issues: 1, 2025

Appendix A

RMOAI Implementation and RegenOmics Benchmark Specifications

The following provides RMOAI component implementation details and RegenOmics benchmark specifications.

COA and MOI Architecture Details

RSC, TOI, and OCA Implementation

RegenOmics Benchmark Specifications