



OPEN

A graph neural network model for inferring interindividual variation from experimental biological data

Fuminori Kawano

Interindividual variation in biological responses to physiological stimuli is a widely recognized phenomenon. However, effective computational tools for identifying the individual-specific mechanisms remain limited. We present the bioreaction–variation network, a graph neural network (GNN) model designed to infer hidden molecular and physiological relationships underlying such variation in experimental biological data. To ensure applicability at a laboratory scale, the model was trained on a domain-specific corpus constructed from approximately 65 K published studies containing the keyword “skeletal muscle”. The architecture comprises five layers with a multi-head attention mechanism and a multi-layer perceptron, enabling the model to capture both local topological features and directional dominance between connected nodes. The GNN was trained to learn relationships from experimental models to target features, as well as among target features. Using real experimental input consisting of differential gene expression data from mouse skeletal muscle subjected to acute exercise, the model successfully inferred individualized networks, identifying both common and unique paths across individuals based on input experimental context. These results demonstrate the model’s capacity to extract interpretable, individual-specific biological connectivity patterns. The proposed framework serves as a proof of concept for customizable, context-based GNN inference designed to address biological variation at the individual level.

Keywords Interindividual variation, Artificial intelligence, Machine learning, Deep learning

Deep understanding of interindividual variation, including sex differences, in responses to physiological stimuli such as exercise, nutrition, aging, and pathological conditions is a critical step toward promoting lifelong health and enhancing the quality of human life. The Dunedin Multidisciplinary Health and Development Study, a longitudinal birth cohort study of individuals born in 1972–1973, has revealed substantial variability in the pace of aging among participants^{1,2}. It further suggests that early-life psychological and nutritional exposures may induce long-term alterations in epigenetic mechanisms that contribute to interindividual differences in aging trajectories. Despite these findings, the processes by which such interindividual variation emerge and develop across the lifespan remain poorly understood. While regular physical activity is widely recognized for its beneficial effects on human health, it is also well documented that considerable variability exists in the outcomes of exercise interventions. For instance, Bamman et al.³ demonstrated that individuals exhibit heterogeneous responses to resistance training and can be categorized as extreme-, modest-, or non-responders, with these variations not attributable to sex or age. Similar variability has been reported in response to endurance training. Ross et al.⁴ reviewed data demonstrating that maximal oxygen consumption increased by only 7% in the lowest responder, whereas the highest responder exhibited an increase of 118% following 24 weeks of treadmill-based endurance training. Furthermore, Bonafiglia et al.⁵ quantitatively identified meaningful interindividual differences in trainability by calculating the standard deviation of individual response (SD_{IR}). They proposed that true variability exists when the SD_{IR} exceeds the smallest worthwhile change, defined as 0.2 times the standard deviation of a control group. Using this criterion, they confirmed significant interindividual variation in the adaptive increases of skeletal muscle citrate synthase activity and capillary density after 4 weeks of cycling exercise training. Collectively, these studies confirm the existence of substantial biological diversity in response to external stimuli. Nevertheless, the underlying mechanisms that give rise to such interindividual variation remain largely elusive. This gap in knowledge may be attributed to the inherent complexity of the biological systems involved, which arises from their multifactorial and person-specific nature, posing a substantial challenge to identifying the determinants of individual responsiveness.

Graduate School of Health Science, Matsumoto University, 2095-1 Niimura, Matsumoto City, Nagano 390-1295, Japan. email: kawano@t.matsu.ac.jp

Recent advances in artificial intelligence (AI) have enabled *in silico* analyses to uncover complex patterns and high-dimensional relationships across biological datasets. In bioinformatics, deep learning models trained on genome-wide multi-omics data have been successfully applied to infer the regulatory logic of transcriptional control and to identify key molecular drivers of disease. For example, Xi et al.⁶ developed a neural network that integrates transcription factor expression with chromatin accessibility at *cis*-regulatory elements using a multi-head self-attention mechanism, enabling accurate prediction of gene expression and classification of cell states from single-cell multi-omics data, as well as the identification of candidate regulatory factors in type 2 diabetes. In single-cell and spatial omics, recent advances in AI-driven methods have established a range of graph- and generative-model approaches. Graph neural networks (GNNs) have been applied to quantify network rewiring across conditions, prioritize candidate regulators, and capture both short- and long-range cell–cell interactions in spatial transcriptomics^{7,8}. Multi-view and path-based frameworks have enabled the integration of inter- and intra-cellular signaling, allowing gene-expression imputation and the inference of ligand–gene regulatory relationships at single-cell resolution⁹. In addition, graph-based models using single-cell ATAC-seq data have been developed to reconstruct higher-order 3D chromatin compartment structures, while dual-topology graph convolutional networks have improved unsupervised clustering of heterogeneous transcriptomes^{10,11}. Beyond GNNs, diffusion-based generative models have been introduced to generate virtual transcriptomic profiles, offering applications in drug-response prediction¹². Together, these advances illustrate how deep learning architectures are increasingly capable of modeling both the structural and functional variation underlying complex cellular systems. While these AI-based models provide significant insights into cellular-level variation and regulation, a deeper understanding of interindividual biological variations, particularly in response to physiological stimuli, requires a broader, macro-scale modeling approach. Such a model must integrate molecular networks and physiological parameters with environmental influences that encompass both biological and physiological factors, enabling individualized inference on a mechanistic basis.

To investigate the biological mechanisms underlying interindividual variation, it is essential to move beyond static representations of biological systems and consider the dynamic, context-dependent relationships among molecular and physiological factors. While canonical pathway databases such as KEGG have been widely used to interpret high-throughput data by identifying enriched signaling pathways, these static maps often fail to reconcile inconsistencies observed in empirical studies. For instance, experimentally validated changes in gene expression or protein abundance may not align with expected pairwise interactions suggested by public databases. In several cases, only one component of a canonical interaction pair shows significant regulation, whereas the other remains unchanged, suggesting that the activation of biological pathways is conditional and context-specific. Such discrepancies likely stem from heterogeneity in experimental conditions, including differences in tissue types, environmental stimuli, or subject-specific physiological states. To address these limitations, we propose that a dynamic biological network, in which multiple, potentially conflicting factor relationships inferred from diverse experimental contexts are allowed to coexist and interact, is required for accurately modeling individual-specific biological responses. In this study, we present a novel deep learning architecture based on a GNN, designed to infer individualized mechanisms of “bioreaction-variation” — a term that refers to the interindividual variation in biological responses to physiological stimuli. Our GNN framework integrates both molecular-level and physiological-level nodes and is capable of identifying the most plausible mechanistic pathways that explain observed data under specific experimental contexts. Furthermore, the model is also designed to be implementable at a laboratory scale, with a minimized structure that facilitates targeted inference for specific tissues or physiological functions. This design enables context-aware inference of individualized biological networks, thereby offering a scalable and mechanistically interpretable approach for elucidating complex interindividual variation.

Methods

Overview of bioreaction-variation network

The present study introduces a bioreaction-variation network constructed using a GNN framework designed to capture the relationships between experimental models and corresponding physiological or biological parameters, as well as the interactions among these parameters. The GNN architecture comprises two core components: 1) a Model-to-Target interaction layer, which learns the associations between experimental conditions and observed outcomes, and 2) a Target-to-Target interaction layer, which models the interrelationships among the measured parameters themselves. All training data were curated from published experimental studies, with each parameter represented as a change or difference induced by a specific experimental intervention. This structure enables the model, upon input of new experimental data, including detailed conditions and observed parameter changes, to infer relevant biological pathways that mechanistically link experimental conditions and observed physiological responses, enabling individualized interpretation based on patterns learned from previously published studies.

Training data collection

A total of 65,096 published studies were obtained from PubMed Central. These articles were identified by first searching PubMed with the keyword “skeletal muscle”, and subsequently filtering for those available as free full text in PubMed Central. The main texts of the selected articles were processed using the GPT-4o-mini model to extract relevant experimental contexts. Each study was summarized into a structured format following a fixed JSON schema (Table 1). Each JSON entry represented a discrete experimental unit, capturing the details of the experimental model, the physiological or molecular parameters analyzed, and the direction of change.

Input data

The input data refer to practical datasets used to perform inference with the trained model. In this study, model performance was initially validated using virtual input data constructed to simulate biologically plausible

Attribute	Description
PMID	Article ID provided in PubMed
Species	The species used in the experiment using the common name
Age	The age or weeks old at the time of the experiment
Sex	The biological sex of the subject
Biosample_main	The type of tissue or cells used in the experiment
Biosample_detail	Additional details about the biosample, such as tissue or cell name
Experiment_type	The experimental level, e.g., in vivo, in vitro, ex vivo, case study
Model_main	The main physiological stimulus used in the experiment
Model_detail1	The first level of details, such as specific stimulus type
Model_detail2	Further details, such as intensity, dosage, duration, or frequency
Model_detail3	Additional specific experimental conditions if necessary
Timepoint	Chronological point at which the sample was collected for analysis
Targets	
Target	The specific name of the measured factor
Molecule_type	Specify the type of molecule, e.g., protein, mRNA
Analysis_main	The primary measurement method, e.g., western blotting, PCR
Analysis_detail	Additional details about the measurement method
Relation	The type of observed change, e.g., increase, decrease
Change	The degree of change in percentage or fold change
Significance	The statistical significance using p-values
Control	The control group used for comparison in “relation”

Table 1. Attributes and descriptions for data collection. Descriptions shown in this table are excerpts from the actual prompts used in the GPT-4o-mini model for attribute formatting. Note: The attribute was filled by “none”, if applicable answer did not meet in the article.

scenarios (see Supplementary data.zip online). Additionally, inference was conducted using real experimental data (see Supplementary data.zip online), derived from RNA sequencing results of individual mouse skeletal muscle samples. These datasets were originally reported in our previous publication¹³ and have been deposited on the official journal website associated with that article. To assess interindividual variability, fold changes in gene expression between exercised and non-exercised control mice were calculated across all possible pairwise combinations of individuals. For example, expression profiles of exercise #1 were compared independently to those of control #1, #2, and #3. Genes exhibiting differential responses to exercise were selected based on the following criterion: at least one exercised mouse showed an average fold change in gene expression greater than twofold when compared to all control mice, while the remaining exercised mice exhibited either changes below the twofold threshold or responses in the opposite direction. A total of 15 genes met this criterion and were used for individualized inference in downstream analyses (see Supplementary Fig. S1 online).

Embedding

To encode experimental context into vector representations, BioBERT (ver. 1.1)¹⁴ was used to tokenize the descriptive elements of each study. Graph construction was performed using PyTorch Geometric (ver. 2.6.1)¹⁵, in which nodes were defined based on the “model main” categories (model nodes) and individual “target” entries (target nodes) extracted from each publication. To support node-level learning, node embeddings were also aggregated into mean vectors representing the average features across identical model or target nodes, thereby capturing generalizable node representations across the dataset. Further details on the embedding procedures are provided in the Supplementary Methods online.

GNN model learning

GNN training was conducted using a five-layer architecture (Fig. 1). Full implementation details, including preprocessing scripts and model training code, are available via the GitHub repository (<https://github.com/fumikawano-lab/Bioreaction-Variation-Network>). To capture Model-to-Target interactions, the first layer employed a multi-head Graph Attention Convolution (GATConv) mechanism^{16,17} to compute attention weights (attn_W) for each target node. Model and target features, initially encoded as 768-dimensional BioBERT-based embeddings, were linearly transformed into 2,048-dimensional representations composed of 8 attention heads with 256 hidden dimensions per head.

$$\tilde{\mathbf{m}}_i = \text{BN}_m(\mathbf{W}_m \mathbf{m}_i), \tilde{\mathbf{t}}_j = \text{BN}_t(\mathbf{W}_t \mathbf{t}_j), \quad (1)$$

where $\mathbf{m}_i \in \mathbb{R}^{768}$ denotes the feature of a model node i , and $\mathbf{t}_j \in \mathbb{R}^{768}$ denotes the feature of a target node j . After reshaping, each is split into H attention heads of dimension d .

$$\mathbf{u}_{ij} = \mathbf{e}_{ij}[1 : 768] - \mathbf{e}_{ij}[769 : 1536], \mathbf{a}_{ij} = \mathbf{W}_e \mathbf{u}_{ij}, \quad (2)$$

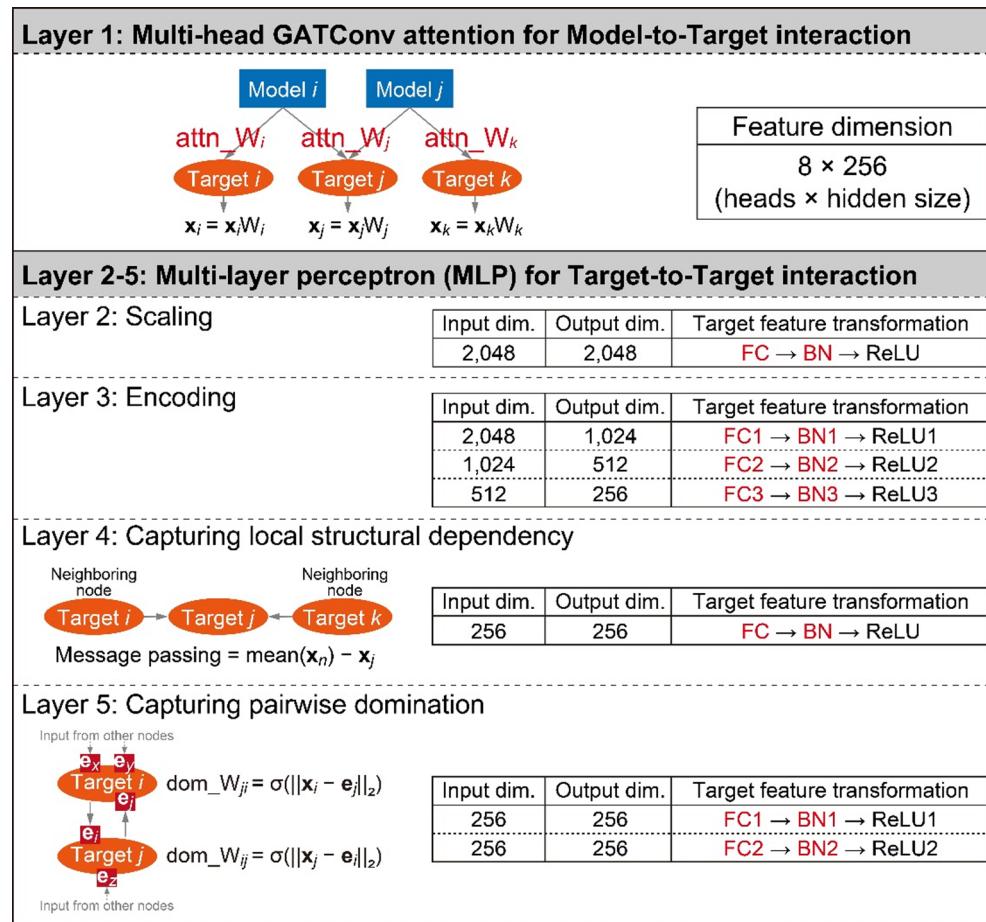


Fig. 1. GNN model training overview. The graph neural network (GNN) model was trained to capture both Model-to-Target and Target-to-Target interactions through a structured, layer-wise architecture. Layer 1 employed a multi-head Graph Attention Convolution (GATConv) to compute attention weights (attn_W) for each Target node based on associated Model nodes. These attention weights are learnable parameters that allow the model to assign dynamic importance to each model-target interaction. The mechanism operated across 8 attention heads, each with a 256-dimensional hidden space. Here, x denotes the 768-dimensional feature vector of a Target node, aggregated as the mean of all features assigned to the same target entity. Layers 2 to 5 focused on Target-to-Target interactions through a multi-layer perceptron (MLP). Layer 2 applied a fully connected (FC) transformation with batch normalization (BN) and ReLU activation, preserving the feature dimension at 2,048. Layer 3 sequentially encoded the feature vector via three FC-BN-ReLU blocks, progressively reducing the dimensionality from 2,048 to 256. Layer 4 captured local structural dependencies by computing message passing, defined as the difference between the mean feature vector of neighboring nodes (\bar{x}_n) and the current node's feature (x_j). This layer also included an FC-BN-ReLU block. Layer 5 inferred pairwise domination between connected Target nodes. The domination weight (dom_W) was calculated as the sigmoid of the feature distance between the mean feature of a target node (\bar{x}_i) and the incoming feature vector from another node (e_j). Two FC-BN-ReLU blocks were used in this layer as well. All learnable parameters including FC, BN, and attn_W are shown in red.

where $\mathbf{e}_{ij} \in \mathbb{R}^{1536}$ denotes the edge attribute vector between model node i and target node j .

$$\begin{aligned}\alpha_{ij}^{(h,m \rightarrow t)} &= \frac{\exp \left(\text{PReLU} \left(\left\langle \mathbf{m}_i^{(h)}, \mathbf{a}_{ij}^{(h)} \right\rangle \right) / \tau_{mt} \right)}{\sum_{h'} \exp \left(\text{PReLU} \left(\left\langle \mathbf{m}_i^{(h')}, \mathbf{a}_{ij}^{(h')} \right\rangle \right) / \tau_{mt} \right)}, \\ \alpha_{ij}^{(h,m \rightarrow n)} &= \frac{\exp \left(\text{PReLU} \left(\left\langle \mathbf{m}_i^{(h)}, \mathbf{a}_{ij}^{(h)} \right\rangle \right) / \tau_{mn} \right)}{\sum_{h'} \exp \left(\text{PReLU} \left(\left\langle \mathbf{m}_i^{(h')}, \mathbf{a}_{ij}^{(h')} \right\rangle \right) / \tau_{mn} \right)}, \\ \alpha_{ij}^{(h,t)} &= \frac{\exp \left(\text{PReLU} \left(\left\langle \mathbf{t}_j^{(h)}, \mathbf{a}_{ij}^{(h)} \right\rangle \right) / \tau_t \right)}{\sum_{h'} \exp \left(\text{PReLU} \left(\left\langle \mathbf{t}_j^{(h')}, \mathbf{a}_{ij}^{(h')} \right\rangle \right) / \tau_t \right)}.\end{aligned}\quad (3)$$

Three types of attention coefficients are computed. The model-to-target attention ($m \rightarrow t$) quantifies the contribution of a model node to a target node via edge attributes. The model-to-target feature reinforcement ($m \rightarrow n$) indicates how strongly a model node should amplify the original feature of the target node. The target self-attention (t) captures the degree to which the target node's own feature is emphasized relative to the edge attributes.

$$\begin{aligned}\mathbf{u}_{ij}^{(h)} &= \alpha_{ij}^{(h,m \rightarrow t)} \mathbf{a}_{ij}^{(h)} + \alpha_{ij}^{(h,m \rightarrow n)} \mathbf{t}_j^{(h)} + \alpha_{ij}^{(h,t)} \mathbf{t}_j^{(h)}, \\ \mathbf{z}_{ij} &= \left[\mathbf{u}_{ij}^{(1)}; \mathbf{u}_{ij}^{(2)}; \dots; \mathbf{u}_{ij}^{(H)} \right] \in \mathbb{R}^{Hd_h}.\end{aligned}\quad (4)$$

Each edge-level message \mathbf{z}_{ij} is formed by, for every head h , taking a weighted sum of the projected target feature and the edge attribute using the three attention coefficients, and then concatenating the H head-wise results.

$$\mathbf{x}_j^{(1)} = \frac{1}{|\mathcal{E}_{MT}(j)|} \sum_{(i,j) \in \mathcal{E}_{MT}(j)} \mathbf{z}_{ij}. \quad (5)$$

For each target node j , the updated node feature $x_j^{(1)}$ is the simple mean of the edge-level messages z_{ij} over all incoming edges $(i,j) \in \mathcal{E}_{MT}$.

Subsequent layers (Layers 2–5) implemented a multi-layer perceptron (MLP) to capture Target-to-Target interactions.

$$\mathbf{x}_j^{(2)} = \text{ReLU} (\text{BN}_1^{(2)} (\mathbf{W}_1^{(2)} \mathbf{x}_j^{(1)})), \mathbf{W}_1^{(2)} \in \mathbb{R}^{2048 \times 2048}. \quad (6a)$$

The second layer scales all target node features while preserving their dimensionality, ensuring consistent scaling across nodes without altering the representation size.

$$\begin{aligned}\mathbf{h}_j^{(3,1)} &= \text{ReLU} \left(\text{BN}_1^{(3)} \left(\mathbf{W}_1^{(3)} \mathbf{x}_j^{(2)} \right) \right), \mathbf{W}_1^{(3)} \in \mathbb{R}^{1024 \times 2048}, \\ \mathbf{h}_j^{(3,2)} &= \text{ReLU} \left(\text{BN}_2^{(3)} \left(\mathbf{W}_2^{(3)} \mathbf{h}_j^{(3,1)} \right) \right), \mathbf{W}_2^{(3)} \in \mathbb{R}^{512 \times 1024}, \\ \mathbf{X}_j &= \text{ReLU} \left(\text{BN}_3^{(3)} \left(\mathbf{W}_3^{(3)} \mathbf{h}_j^{(3,2)} \right) \right), \mathbf{W}_3^{(3)} \in \mathbb{R}^{256 \times 512}.\end{aligned}\quad (6b)$$

The third layer encodes the feature stepwise from 2,048 to 256 dimensions, producing the final representation \mathbf{X}_j used in subsequent layers.

The fourth layer was designed to model local topological structure via message passing. Message passing in the training process was defined as the difference between a given target node's feature and the average feature of its input neighbors, as derived from the original training data.

$$\begin{aligned}\boldsymbol{\mu}_j &= \frac{1}{|\mathcal{N}_{TT}(j)|} \sum_{i \in \mathcal{N}_{TT}(j)} \mathbf{X}_i, \Delta \mathbf{x}_j = \text{ReLU} (\text{BN}_x (\mathbf{W}_x (\boldsymbol{\mu}_j - \mathbf{X}_j))), \\ \mathbf{M}_j &= \mathbf{X}_j + \Delta \mathbf{x}_j.\end{aligned}\quad (7)$$

Here, $\mathcal{N}_{TT}(j)$ denotes the set of target nodes directly connected to target node j within the target-target graph. The term $\boldsymbol{\mu}_j$ represents the mean feature vector of the neighbors of j , and $\Delta \mathbf{x}_j$ is the predicted message passing update, obtained as the difference between the neighbor mean and the node's own representation after transformation and normalization. Finally, \mathbf{M}_j is defined as the updated representation of node j , where \mathbf{x}_j is augmented with the message passing it would receive if the entire network were activated.

$$\begin{aligned}\bar{\mathbf{e}}_j &= \frac{1}{|\mathcal{N}(j)|} \sum_{i \in \mathcal{N}(j)} (\mathbf{e}_{ij}[1:768] - \mathbf{e}_{ij}[769:1536]), \\ \widehat{\Delta \mathbf{x}}_j &= \text{ReLU} \left(\text{Norm} \left(\mathbf{W}_e^{(\text{orth})} \bar{\mathbf{e}}_j \right) \right).\end{aligned}\quad (8)$$

The teacher signal $\widehat{\Delta \mathbf{x}}_j$ is derived directly from edge attributes. Specifically, the averaged edge difference vector $\bar{\mathbf{e}}_j$ is linearly transformed using a weight matrix $\mathbf{W}_e^{(\text{orth})}$, which is initialized as an orthogonal matrix and kept fixed during training. This transformation also projects the teacher signal into a 256-dimensional space, thereby matching the dimensionality of the model-predicted update $\Delta \mathbf{x}_j$. This design ensures that the teacher signal provides a stable and unbiased reference for message passing, independent of the model's parameter updates.

The fifth layer modeled pairwise dominance among target nodes using a domination weight.

$$d_{i \rightarrow j} = \|\mathbf{A}_j - \mathbf{A}_{i \rightarrow j}\|_2, W_{i \rightarrow j} = \sigma(\beta(d_{i \rightarrow j} - d_0)). \quad (9)$$

Here, $W_{i \rightarrow j}$ denotes the domination weight, representing how strongly the feature representation of node j depends on the input from node i .

$$\begin{aligned}\overline{W}_j &= \frac{1}{|\mathcal{N}(j)|} \sum_{i \in \mathcal{N}(j)} W_{i \rightarrow j}, \\ \mathbf{D}_j^{\text{teacher}} &= \text{ReLU} \left(\text{Std} \left(\mathbf{W}_d^{(\text{orth})} (\overline{W}_j \cdot \mathbf{A}_j) \right) \right).\end{aligned}\quad (10)$$

For each target node j , the domination weights from all connected neighbors $i \in \mathcal{N}(j)$ are computed and averaged to form the aggregated weight \overline{W}_j . The resulting $\mathbf{D}_j^{\text{teacher}}$ provides the teacher signal for domination, serving as the fixed reference in the supervised learning of dominance.

$$\mathbf{D}_j^{\text{out}} = \text{ReLU} (\text{BN}_2 (\mathbf{W}_2 \text{ReLU} (\text{BN}_1 (\mathbf{W}_1 \mathbf{M}_j))). \quad (11)$$

The domination output is the final model prediction of the GNN model learning, obtained by transforming the updated node representation. This vector represents the learned dominance distribution over the connected nodes.

$$\text{Dom}(i \rightarrow j) := (\mathbf{D}_j^{\text{out}})_i, i \in \mathcal{N}(j). \quad (12a)$$

$$i \succ j \text{ iff } (\mathbf{D}_j^{\text{out}})_i > (\mathbf{D}_j^{\text{out}})_j. \quad (12b)$$

Here, the pairwise domination relation $\text{Dom}(i \rightarrow j)$ is defined as the i -th component of the model-predicted domination output vector for the target node j , $\mathbf{D}_j^{\text{out}}$. In practice, for each target node j , the domination scores for all connected neighbors $i \in \mathcal{N}(j)$ are computed and employed during network exploration; an oriented edge $i \rightarrow j$ is preferred when $\text{Dom}(i \rightarrow j)$ exceeds $\text{Dom}(j \rightarrow i)$.

The training loss was computed as the mean squared error (MSE) between the model-predicted outputs and the corresponding teacher signals: the message passing updates in the fourth layer (message passing loss) and the final dominance vectors in the fifth layer (domination loss).

$$\mathcal{L}_{\text{message}} = \frac{1}{|\mathcal{T}|} \sum_{j \in \mathcal{T}} \|\Delta \mathbf{x}_j - \widehat{\Delta \mathbf{x}}_j\|_2^2. \quad (13)$$

$$\mathcal{L}_{\text{dom}} = \frac{1}{|\mathcal{T}|} \sum_{j \in \mathcal{T}} \|\mathbf{D}_j^{\text{out}} - \mathbf{D}_j^{\text{teacher}}\|_2^2. \quad (14)$$

$$\mathcal{L}_{\text{total}} = 2.0 \mathcal{L}_{\text{message}} + 0.1 \mathcal{L}_{\text{dom}}. \quad (15)$$

The learning was formulated under the assumption that the entire network is activated, such that every node receives input signals as if all physiological stimuli were simultaneously present. Under this condition, the message passing loss evaluates the accuracy of the predicted influence from proximal biological factors, which enables the model to reconstruct local subgraph structures. Importantly, when the whole network is assumed to be activated, the node features obtained after message passing can be interpreted as reflecting the inherent positioning of each node within the biological system. The domination loss further evaluates the consistency of edge directionality, providing a quantitative criterion for embedding competitive up- and downstream relationships among parameters that were simultaneously analyzed within individual studies. By jointly optimizing these two losses, the model learns node features and directional connectivity in a manner that reflects their functional positions in the global biological network.

Model optimization was performed using Adam optimizer¹⁸ with a learning rate of 0.001, and backpropagation was carried out over 50 or 200 total epochs. To mitigate gradient bias, total loss was computed as a weighted sum of the two loss components: message passing loss $\times 2.0$ and domination loss $\times 0.1$ (Eq. 15). Final trained models

are publicly available at Google Cloud Storage (https://storage.googleapis.com/skeletal_muscle/sm_v1/gnn_model/gnn_model_final.pt for 200-epoch model and [gnn_model_final_50epoch.pt](https://storage.googleapis.com/skeletal_muscle/sm_v1/gnn_model_50epoch.pt) for 50-epoch model).

Network inference

To infer individual-specific hidden pathways, network traversal was performed based on user-provided input data comprising experimental model metadata and observed target parameters. These inputs, formatted consistently with the training data structure, were embedded into 768-dimensional feature vectors using BioBERT, following the same preprocessing pipeline used during model training. To initiate the inference process, cosine similarity was computed between the input model vector and all model embeddings in the training dataset. The top five most similar model nodes were selected, and their directly connected target nodes were identified as primary targets, serving as start nodes for subsequent traversal (Step 1 in Fig. 2). The aim was to reconstruct viable paths from these start nodes to goal nodes, which corresponded to the target features present in the user-provided input. In Step 2, all possible directed paths between the identified start and goal nodes were enumerated based on the GNN-learned edge connectivity between target nodes. This yielded a candidate path space representing potential mechanistic routes through the network. In Step 3, path refinement was performed by selecting the most contextually relevant edges at each intermediate node. When multiple edges existed between the same node pair (i.e., same source and target nodes but differing feature contexts), the edge whose source feature was most similar to the preceding target node, as determined by cosine similarity, was selected. This ensured feature continuity across the reconstructed path.

In Step 4, message passing was performed from each primary node through candidate paths, and the propagated features at the goal nodes were compared with the input goal features. The subsets of paths that minimized this discrepancy were iteratively selected using a greedy beam-style procedure. The final output corresponds to the best combination of paths whose averaged predictions most accurately reproduced the input goal features. Mathematically, this procedure can be described as follows.

For each primary node T_i with feature vector $\mathbf{x}_{T_i} \in \mathbb{R}^d$, an attention-based scalar weight was applied to initialize the node representation:

$$\hat{\mathbf{x}}_{T_i} = \alpha_{T_i} \mathbf{x}_{T_i}, \alpha_{T_i} \in \mathbb{R}. \quad (16)$$

Here, $\alpha_{T_i} \in \mathbb{R}$ denotes the attention-derived weight, and $\hat{\mathbf{x}}_{T_i}$ represents the adjusted primary feature used to start the propagation.

Message passing along a candidate path $p : v_0 \rightarrow v_1 \rightarrow \dots \rightarrow v_L (v_0 = T_i, v_L = g)$ was implemented as a multiplicative update. At each step $l = 1, \dots, L$, the node feature was updated according to

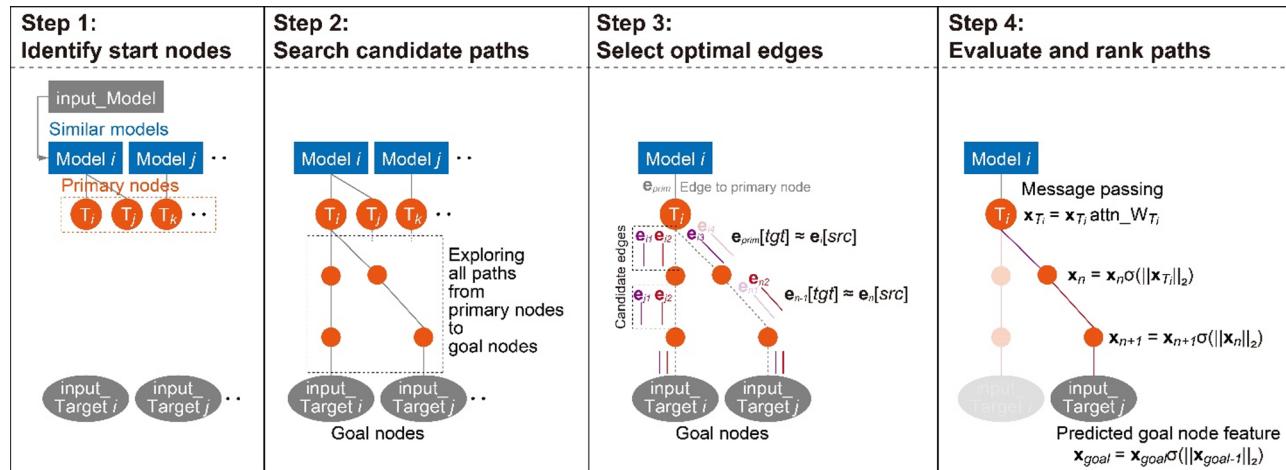


Fig. 2. Network inference overview. Step 1: The top five models in the training dataset most similar to the input model (based on model feature similarity) were identified. The target nodes directly connected to these models were then designated as primary nodes, serving as the starting points for network inference. Step 2: Using the trained GNN model, all candidate paths were explored from the primary nodes (start nodes) toward the input target nodes, which served as the goal nodes. Step 3: For each source–target node pair along the paths, the optimal edge was selected from a pool of candidate edges. The edge selection criterion required that the source node of the current edge closely matched the target node of the previous edge in the sequence (i.e., $e_{n-1}[tgt] \approx e_n[src]$), thereby ensuring topological continuity toward the goal node. Step 4: Message passing was applied along each candidate path. For the first target node, the attention weight (attn_W) from the connected model node was applied. Subsequent nodes received propagated signals scaled by the sigmoid-transformed Euclidean norm of the previous node's feature (i.e., $\sigma(\|\mathbf{x}_n\|_2)$). The predicted feature of each goal node (\mathbf{x}_{goal}) was obtained at the end of the path. The best combination of paths was selected for each goal node by minimizing the Euclidean distance between the predicted node feature and the actual input feature across all possible path combinations.

$$\begin{aligned}\mathbf{x}_{v_l}^{(l)} &= s_l \mathbf{t}_{e_l}, \\ s_l &= c\sigma(\alpha(\|\mathbf{x}_{v_{l-1}}^{(l-1)}\|_2 - \beta)).\end{aligned}\quad (17)$$

In this formulation, $e_l = (v_{l-1} \rightarrow v_l, m_l)$ is the selected edge, $\mathbf{t}_{e_l} \in \mathbb{R}^d$ is its target-side feature, and s_l is a scalar intensity determined by the norm of the preceding node feature. The constants $c > 0$, $\alpha > 0$, and $\beta \in \mathbb{R}$ regulate the scaling and nonlinearity applied through the sigmoid function.

The prediction obtained at the goal node g from a single path p was expressed as

$$\hat{\mathbf{x}}_g(p) = \mathbf{x}_{v_L}^{(L)}. \quad (18)$$

This corresponds to the propagated feature obtained after traversing the entire path.

During the best-combination search, subsets of candidate paths were iteratively selected for each goal node. For a subset $S_g \subseteq \mathcal{P}_g$, the aggregated prediction was defined as

$$\bar{\mathbf{x}}_g(S_g) = \frac{1}{|S_g|} \sum_{p \in S_g} \hat{\mathbf{x}}_g(p). \quad (19)$$

This aggregation involves only the paths in the selected subset S_g and serves as the basis for evaluating agreement with the input goal feature.

The discrepancy between the aggregated predictions and the input goal features was quantified by the loss

$$\mathcal{L}(S) = \sum_{g \in G_{\text{goal}}} \|\bar{\mathbf{x}}_g(S_g) - \mathbf{y}_g\|_2^2, \quad (20)$$

where \mathbf{y}_g is the input goal feature and $S = \{S_g\}_{g \in G_{\text{goal}}}$ represents the collection of selected subsets across all goals.

The subsets were updated by adding the candidate path p that minimized the squared error of the already-selected paths $q \in S_g^{(t)}$ augmented with p , at each iteration:

$$\begin{aligned}p^* &= \arg \min_{p \in \mathcal{P}_g \setminus S_g^{(t)}} \left\| \frac{1}{|S_g^{(t)}| + 1} \left(\sum_{q \in S_g^{(t)}} \hat{\mathbf{x}}_g(q) + \hat{\mathbf{x}}_g(p) \right) - \mathbf{y}_g \right\|_2^2, \\ S_g^{(t+1)} &= S_g^{(t)} \cup \{p^*\}.\end{aligned}\quad (21)$$

This procedure was repeated until no further improvement was achieved or the maximum beam width K_g was reached. The final subsets for each goal, denoted $S_g^{(\text{final})}$, were collected into

$$S^* = \{S_g^{(\text{final})} \mid g \in G_{\text{goal}}\}, \quad (22)$$

which was defined as the best combination. This output represents the path subsets whose averaged predictions most accurately reproduced the input goal features.

Steps 3 and 4 were embedded in a genetic algorithm framework using the DEAP library¹⁹. Each candidate path generated in Step 2 was treated as an individual. Fitness was evaluated as a weighted combination of loss and diversity, the latter reflecting the uniqueness of nodes and path topology. Path selection employed the Elitist Non-dominated Sorting Genetic Algorithm (NSGA-II), with mutation applied to 20% of individuals per generation. Mutation involved random substitution of an edge within a path, followed by structural repair using the same edge selection strategy described in Step 2 and 3. If the mutation resulted in higher loss, the original path configuration was retained. Given that the initial path candidates were exhaustively constructed and optimized during Steps 2 and 3, the evolutionary loop was iterated twice, solely to confirm that no further improvement in fitness metrics could be achieved. All source code and execution details for the network inference procedure are available at the associated GitHub repository (<https://github.com/fumikawano-lab/Bioreaction-Variation-Network>).

Analysis of individualized networks

To identify individual-specific pathways, inferred GNN outputs for each exercised mouse ($n=3$) were compared against all non-exercised controls ($n=3$). For each exercised mouse, individual networks were constructed by aggregating all predicted paths obtained from pairwise comparisons with each control mouse. A common network was then defined as the set of paths shared across all three individual networks, while unique networks were obtained by subtracting the common network from each individual's network. To assess overall reconstruction accuracy, message passing loss was calculated for each goal node and averaged across the network. In addition, among non-primary and non-goal nodes, the most frequently occurring source node across all inferred networks was identified. For this node, the contribution of each connected target node was assessed by referencing the message passing loss of the specific path (i.e., evolutionary algorithm-derived individual) in

which the corresponding edge was included. This provided a direct measure of each target node's involvement in the reconstruction of input-derived goal node features within individualized networks.

Results

Overview of training data

The training graph was constructed from a total of 27,155 model nodes and 84,723 target nodes. These nodes were interconnected by 383,225 model-to-target edges and 2,475,502 target-to-target edges. Node frequency distributions are summarized in the Supplementary data.xlsx online. Briefly, the most frequently represented experimental models included exercise, high-fat diet feeding, sarcopenia, aging, muscle injury, electrical stimulation, type 2 diabetes, Duchenne muscular dystrophy, diabetes, and isometric contraction. Frequently analyzed target nodes encompassed IL-6, myogenin, MyoD, TNF α , PGC-1 α , creatine kinase, insulin, myosin heavy chain, MuRF1, and atrogin1.

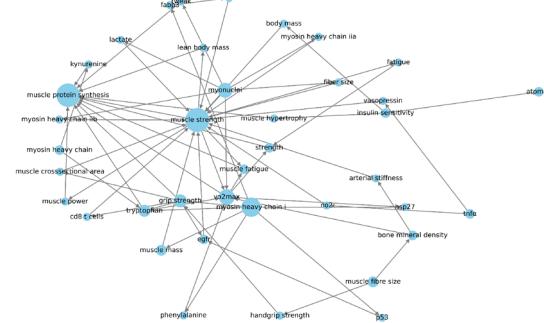
During model training, the domination loss decreased more rapidly than the message passing loss (see Supplementary Fig. S3 online), a trend that was accounted for by applying differential weighting to these two components in the total loss function. When equal weights were assigned, the message passing loss plateaued early in the training process, limiting further optimization. Although the domination loss approached a minimum around 50 epochs, the message passing loss continued to decline steadily, reducing the difference between the two losses from 0.269 at epoch 50 to 0.228 at epoch 200. These results indicate that training for 200 epochs did not result in overfitting and that both 50-epoch and 200-epoch models were retained for downstream inference to assess potential differences in generalization.

Validation of model with virtual input data

To evaluate the inference performance of the trained GNN model, two types of virtual input datasets were constructed that differed only in the biological sex of the subjects, with all other experimental conditions held constant. In the networks inferred from the model trained over 50 epochs, key nodes such as myosin heavy chain I, myonuclei, and VO₂ max were frequently identified in both male and female datasets; however, the surrounding edge structures differed noticeably between sexes (Fig. 3). Similar nodes were also retrieved from

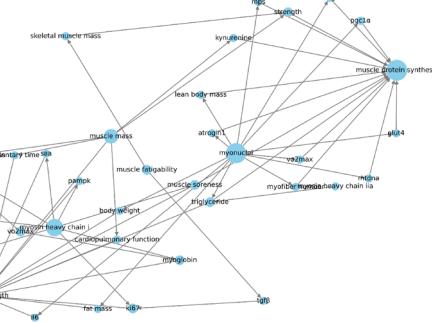
50 epochs (male)

Mean dist. = 0.0882



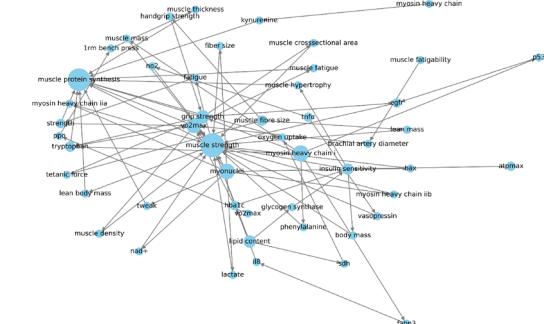
200 epochs (male)

Mean dist. = 0.1044



50 epochs (female)

Mean dist. = 0.0871



200 epochs (female)

Mean dist. = 0.0998

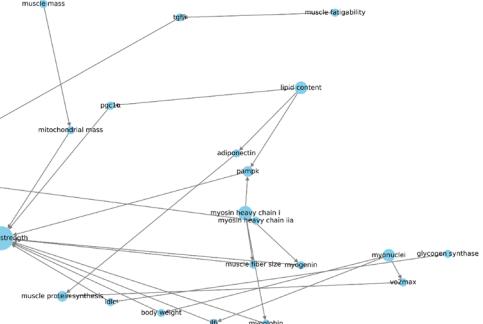


Fig. 3. Model validation using virtual input data. Network graphs illustrate all inferred paths based on virtual input data in which the only differing attribute was the biological sex of the subjects. Model training was performed for either 50 or 200 epochs. The message passing loss was calculated for the best combination of inferred paths associated with each goal node, and these values were averaged to produce the mean distance (mean dist.).

the model trained over 200 epochs, but no clear improvement was observed in the mean distance (average message passing loss) compared to the 50-epoch model, suggesting that inference performance had already stabilized by that point.

Network inference with real experimental data

To investigate individualized transcriptional responses to exercise, we reanalyzed skeletal muscle RNA sequencing data from exercised and non-exercised mice ($n=3$ each) previously reported in our study¹³. Differentially expressed genes between exercised and non-exercised conditions were used to construct input features, which were then inferred through the trained GNN models (trained for either 50 or 200 epochs). Because the contextual input for the experimental model (e.g., exercise protocol, species, tissue) was identical across the three exercised mice, the same primary nodes were consistently identified. However, the inferred networks showed marked interindividual variation in the structure and composition of intermediate and goal nodes (Fig. 4). These differences were reflected in the mean message passing loss, which varied across individuals. Notably, inference using the 200-epoch trained model yielded lower average loss values in all individuals compared to the 50-epoch model, indicating improved network reconstruction performance.

Given the superior performance of the 200-epoch model, we proceeded to identify common and unique network paths among individuals using this model. A path was defined as a complete sequence of connected nodes from a primary node to a goal node. A total of 27 paths were commonly found across all three individuals, involving key nodes such as AKT, creatine kinase, FOXO3A, mitochondrial morphology, OPA1, and UQCRC2 as intermediate nodes (Fig. 5a, b). Unique networks were obtained by subtracting this common network from each individual's network (Fig. 5c). For example, Exer #1 exhibited 65 unique paths, of which 21 overlapped with Exer #2 and 22 with Exer #3 (Fig. 5a). Exer #2 and Exer #3 shared 35 paths (Fig. 5a), indicating that, among the three exercised mice analyzed, these two mice shared similar characteristics in the individual-specific regulatory mechanisms governing transcriptional responses to exercise. PageRank centrality was calculated to analyze the nodes structuring the GNN shown in Fig. 4 (see Fig. 6 and Supplementary data.xlsx online). OPA1 was the most frequently observed node across individuals. IL-6 and MFN2 were highly ranked in Exer #1 and Exer #2, whereas glutamate and PGC-1 α showed higher centrality in Exer #3. COXIV was also frequently detected, being common to Exer #1 and Exer #3. Hierarchical clustering further revealed that these three individuals were differentially classified based on their centralities.

To further explore key regulatory factors potentially explaining interindividual variation, we focused on intermediate source nodes, excluding primary and goal nodes, and identified those with the highest frequency across all individuals. UQCRC2 emerged as the most recurrent source node in the inferred networks. This high recurrence indicated that UQCRC2 contributed most consistently to distinguishing the hidden network structures across individuals. Figure 7 illustrates the target nodes connected to UQCRC2 in each individual. UQCRC2 formed unique edges to citrate synthase, RPS6, cytochrome c, ANXA2, UQCRC1, CD36, MYH7, MYHC2A, and SIRT3 exclusively in Exer #1. While Exer #2 and Exer #3 shared several downstream nodes, edges from UQCRC2 to triglycerides and free fatty acids were specific to Exer #2, whereas P70S6K, CKMT2, ACTN3, and MYOZ1 were strongly associated with Exer #3. All edges identified from both common and unique paths are listed in the Supplementary data.xlsx online, along with the message passing loss of the paths in which each edge is included.

Discussion

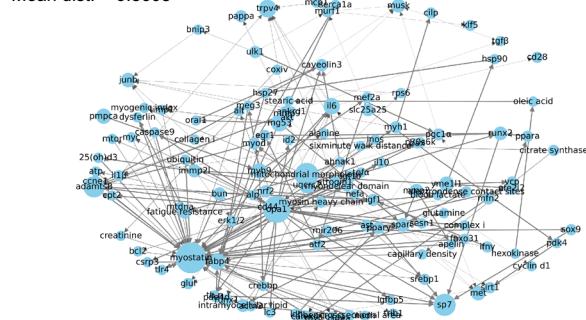
This study presents a GNN model capable of inferring latent networks that reflect individual-specific responses to physiological stimuli. The model was trained on a dataset constructed from published studies retrieved using the keyword “skeletal muscle”. As of July 2025, approximately 258,000 articles are indexed under this keyword, of which 104,000 are available as free full-text articles via PubMed Central. Our training dataset included 65,096 of these articles, covering more than one-fourth of the accessible literature on skeletal muscle.

The predominant experimental models extracted from these publications—such as exercise, high-fat diet, sarcopenia, aging, muscle injury, electrical stimulation, type 2 diabetes, and Duchenne muscular dystrophy—indicate that the dataset reflects biologically relevant contexts closely tied to skeletal muscle research. Accordingly, the extracted parameters and target nodes largely comprised skeletal muscle-related molecules and physiological outcomes. Notably, the model was designed to integrate both molecular and physiological parameters. Among the physiological parameters, non-muscle-related parameters such as VO_2 max, blood glucose, and plasma hormone levels were also included, allowing the reconstruction of networks that reflect both intracellular signaling and systemic physiological outcomes. Rather than focusing solely on inferring molecular cascades, the model learned connections between parameters that frequently co-occur across diverse experimental contexts. This architecture enables the exploration of novel, individualized routes from experimental models to observed phenotypes. This capability is likely supported by the model's global network structure, which connects parameters across distinct publications through shared nodes and feature similarities. As demonstrated by the inference with virtual input data (Fig. 3), even a single biological difference, such as sex, led to the identification of markedly distinct hidden networks. These results highlight the model's ability to adaptively explore individualized paths.

The model's inference behavior was influenced by the domination weight learned during training, as edge selection was strongly biased by the learned domination-based relationships between nodes. One key difference between the 50- and 200-epoch models appears to lie in the edge directionality and path structure, which became increasingly refined with extended training. Validation using virtual input data demonstrated that the 50-epoch model was sufficient to reconstruct the network, likely due to the simplicity of the input, which included only two goal nodes (Fig. 3). In contrast, inference using real biological input data showed improved performance with the 200-epoch model, presumably because the input was more complex, containing unchanged or noisy

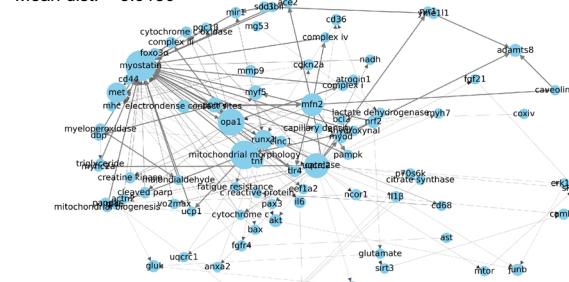
Exer #1 (50 epochs)

Mean dist. = 6.8608



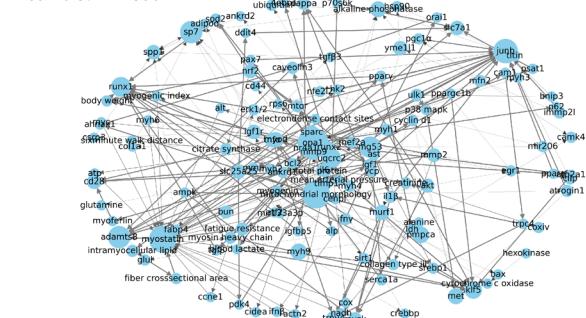
Exer #1 (200 epochs)

Mean dist. = 5.9180



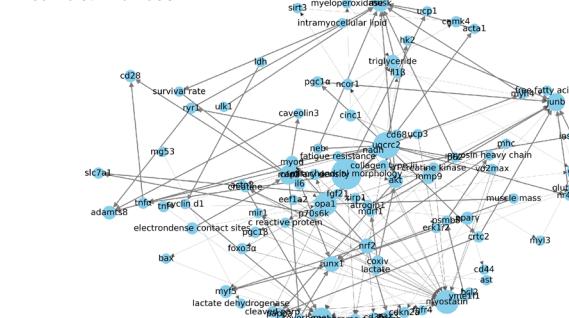
Exer #2 (50 epochs)

Mean dist. = 4.8934



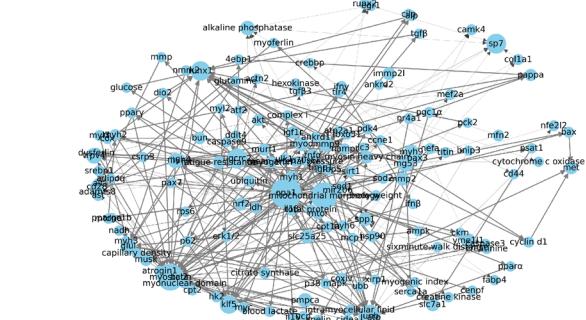
Exer #2 (200 epochs)

Mean dist. = 3.7908



Exer #3 (50 epochs)

Mean dist. = 3.400



Exer #3 (200 epochs)

Mean dist. = 3.1984

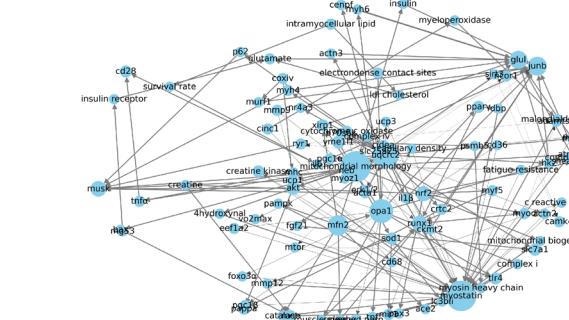


Fig. 4. Network inference using real experimental data. Network graphs illustrate all inferred paths for each exercised mouse (Exer #1–3), based on reanalyzed RNA sequencing data from individual tibialis anterior muscle samples obtained in our previous study¹³. A single bout of treadmill running was conducted after a 4-week training protocol. Gene expression profiles were compared between three exercised mice and three non-exercised controls. See the Methods section for details on the individual comparisons. Model training was performed for either 50 or 200 epochs. The message passing loss was calculated for the best combination of inferred paths associated with each goal node, and these values were averaged to produce the mean distance (mean dist.).

parameters, thereby requiring deeper learning to refine edge relevance and directionality (Fig. 4). The 200-epoch model successfully identified intermediate factors, such as AKT, creatine kinase, FOXO3A, mitochondrial morphology, OPA1, and UQCRC2, as common nodes among individuals from the input of 15 gene expression results (Fig. 5). To further evaluate the model performance, we also compared the results of inference using this model with those obtained from DAVID functional annotation analysis²⁰. With same 15 input genes, only terms such as ATP-binding and positive regulation of transcription by RNA polymerase II were identified, both with low enrichment scores (1.38 and 1.35, respectively) in GO terms and the UniProt Knowledgebase (Supplementary data.xlsx online). No results were retrieved from KEGG pathways. Because enrichment tools

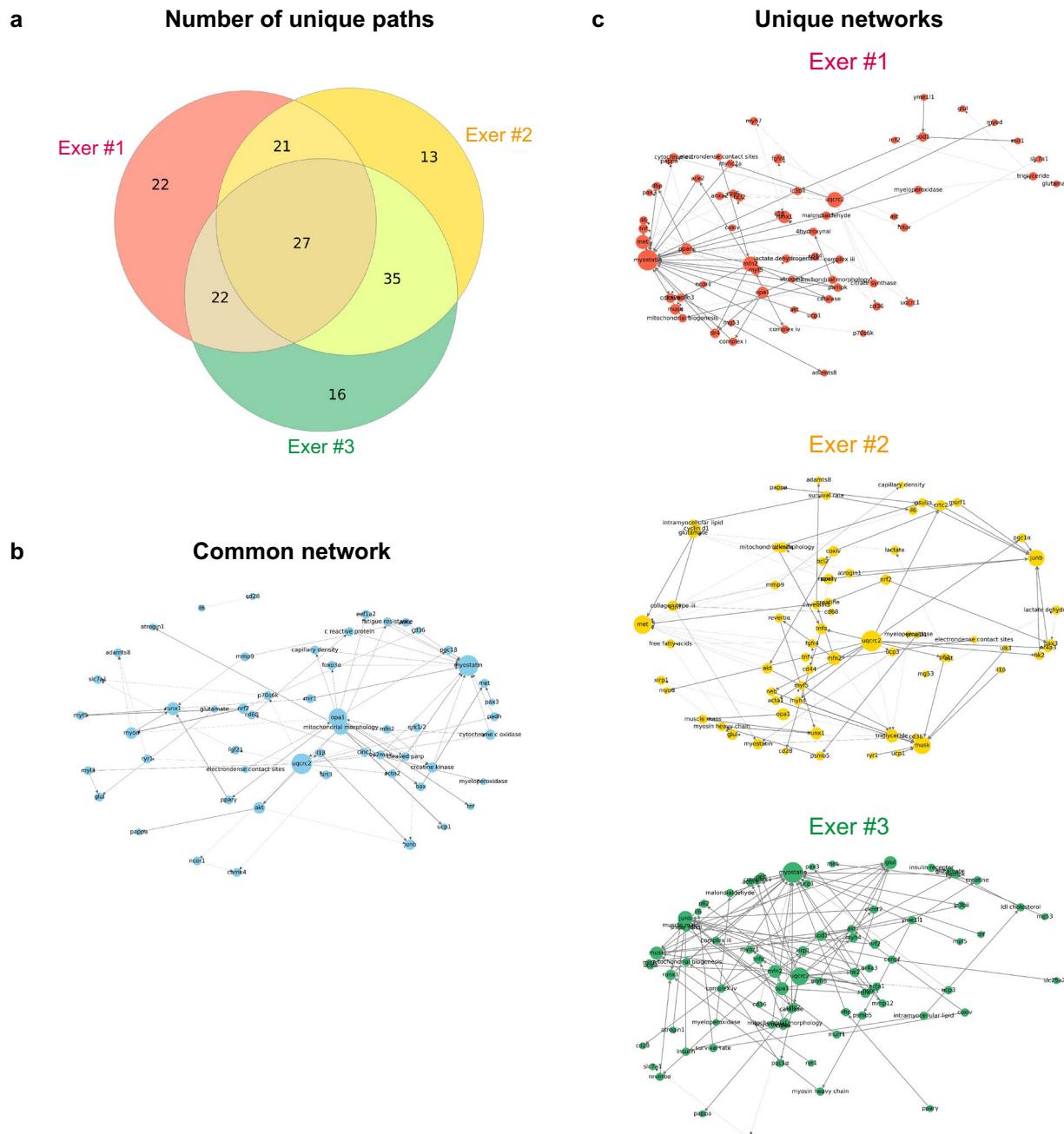


Fig. 5. Extraction of common and individual-specific networks. (a) Venn diagram showing the number of shared and unique inferred paths among the three exercised mice (Exer #1–3). (b) The common network represents the overlapping paths found in all three individuals. (c) Unique networks for each mouse were extracted from the total inferred networks presented in Fig. 4 by identifying individual-specific paths.

are designed to identify related cellular systems from sets of differentially expressed factors, they cannot capture interindividual variation, identify molecular factors beyond the input genes, or trace paths originating from physiological stimuli. These observations collectively suggest that the present GNN functions as a scientifically interpretable, small-scale inference model tailored to skeletal muscle biology, with sufficient capacity to support individualized network inference across a tissue-specific corpus of published studies.

PageRank centrality indicated that OPA1 was the most relevant factor associated with the mechanisms underlying interindividual variation in responses to exercise, despite being directly connected to only two types of nodes across the common and individual-specific networks. In contrast, UQCRC2 was connected to 36 types of nodes, although its centrality was lower (0.016–0.023 vs. 0.058–0.085 for OPA1). This apparently paradoxical result suggests that OPA1 is linked to a node functioning as a highly connected hub. Indeed, in the common

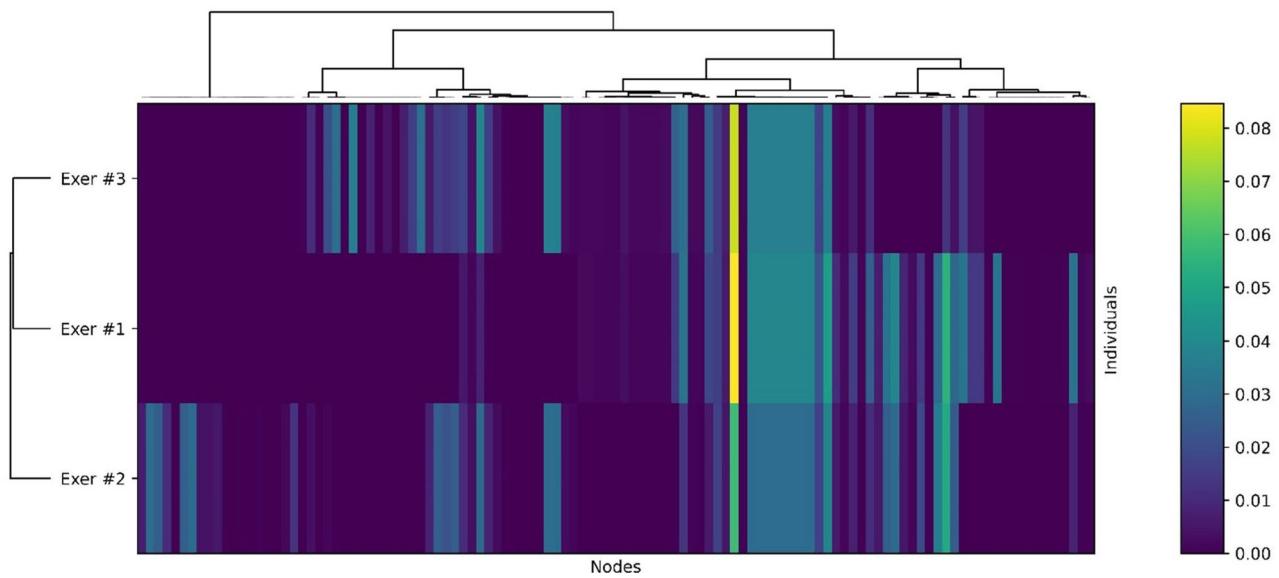


Fig. 6. Clustering of node PageRank centralities across individuals. For each individual-specific network shown in Fig. 4, PageRank centrality was computed from path node sequences, with path weights determined by losses relative to the input-derived goal node features. Individuals were clustered by Jensen–Shannon divergence between their PageRank distributions (excluding the primary and goal nodes) using hierarchical clustering. Nodes absent from an individual's network were assigned a centrality of zero. The three exercised mice (Exer #1–3) were each assigned to separate clusters, indicating mutually divergent network structures.

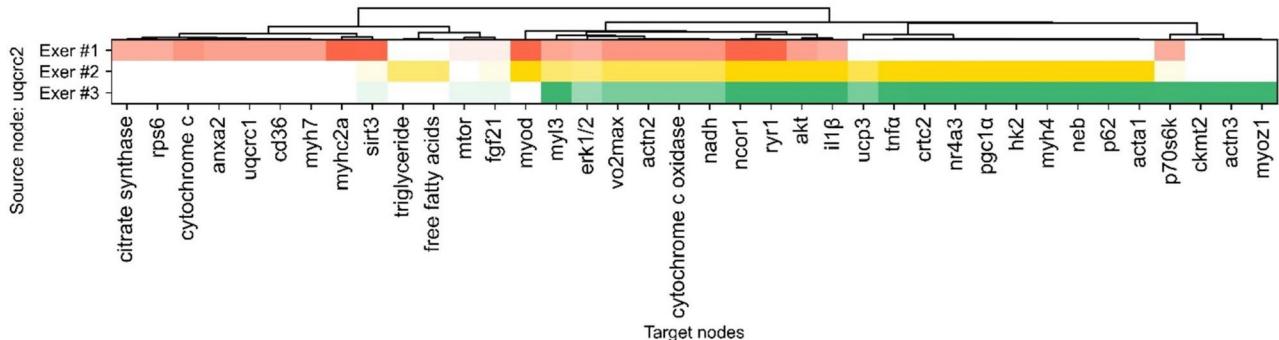


Fig. 7. Variation of UQCRC2-associated edges. Heat map illustrating the target nodes connected to UQCRC2 across the three exercised mice (Exer #1–3). Targets were re-ordered by hierarchical clustering. Color intensity reflects the message passing loss of the path including the corresponding edge, with darker blocks indicating lower loss.

network, OPA1 is connected to MFN2, one of the primary nodes, which itself links to multiple nodes across all individuals and exhibits high centrality (0.047 in Exer #1, 0.039 in Exer #2, and 0.037 in Exer #3). This local loop likely contributed to the elevated centralities of both OPA1 and MFN2. Taken together, these findings suggest that UQCRC2 represents a key intermediate factor potentially explaining interindividual variation in transcriptional responses to acute exercise.

UQCRC2 (ubiquinol-cytochrome c reductase core protein 2) is a component of mitochondrial respiratory chain complex III, essential for the formation of mitochondrial supercomplexes involving complexes I, III, and IV^{21,22}. Missense mutations in UQCRC2 have been linked to impaired mitochondrial respiration and inherited human disorders^{23–25}. Notably, a previous study reported that high volume of high-intensity interval training significantly increased UQCRC2 protein expression in skeletal muscle, contributing to enhanced ATP production through more efficient formation of respiratory complexes²⁶. In the individualized networks inferred by the model, UQCRC2 exhibited variable interactions with distinct downstream nodes across mice, suggesting that its local network context contributed differently to each individual's gene expression response (Fig. 7). Importantly, the model's use of domination-based weighting does not imply that UQCRC2 lies upstream in a canonical signaling pathway. Rather, the model hypothesizes directional relationships in which UQCRC2 functions as a dominant influence relative to its connected nodes within the inferred context. For instance, CKMT2, ACTN3, and MYOZ1 were uniquely associated with UQCRC2 in the Exer #3 (Fig. 7), implying that these factors might

be directly or indirectly modulated by UQCRC2, even though the precise molecular mechanisms linking them remain unknown. This interpretive framework reflects the core capability of the present GNN model: to propose biologically plausible, yet hypothetical, directional associations based on learned patterns across multiple studies. While the inferred relationships should be interpreted as hypothesis-generating rather than confirmatory, the model successfully produced skeletal muscle-specific network structures that align with experimental outcomes and capture individual-level diversity. Taken together, these findings support the utility of this GNN as a domain-focused inference tool with strong contextual validity for skeletal muscle biology.

Limitations and future directions

This study introduces a GNN model designed to infer bioreaction-variation networks within a skeletal muscle-specific corpus. The skeletal muscle model represents one example within the broader framework of bioreaction-variation networks. Models for other organs can be constructed using similar procedures. In fact, separating networks by organ is preferable, since the same molecule may play different roles in different tissues. We also emphasize the need for tools that can integrate these organ-specific networks into an inter-organ framework. The primary aim of this study was not to propose a universally optimized model, but rather to demonstrate a customizable model architecture that can be implemented and adapted at a laboratory scale to address specific biological questions. The model presented here serves as a proof-of-concept example of such a design. However, the current implementation has several limitations as shown below:

- The model requires input data that represent differential states between experimental and control conditions in order to infer individualized networks. Consequently, datasets comprising only baseline or resting conditions, without any comparative group, are not applicable for inference with this model. Therefore, the extraction of individual-specific features through inference reflects relative interindividual variation within the cohort, including the control subjects used for comparison.
- Experimental models or analytical parameters for which relevant edges are scarce in the training dataset may lack sufficient connectivity to support robust inference. In such cases, the model may fail to generate appropriate or biologically meaningful network predictions.
- Because numerical embeddings generated by BioBERT do not directly reflect the magnitude of biological changes in a quantitatively interpretable form, it was necessary to incorporate information on the direction of change, specifically by labeling each relation as either an increase or a decrease. This approach allowed the bioreaction-variation described in the contexts to be translated into vector space. However, capturing such directional changes at the individual level cannot rely on statistical significance testing, which is generally not applicable to single-subject data. In the present study, RNA sequencing data were used, and a twofold change threshold, commonly employed in transcriptomic analyses, was applied to assign directional labels. For other data types or analytical contexts, effect size metrics such as Cohen's d may provide a more appropriate basis for estimating individual-level deviations from the comparison group.

Data availability

The datasets generated and/or analysed during the current study are available in the GitHub repository, <https://github.com/fumikawano-lab/Bioreaction-Variation-Network>. The location of each PyTorch graph dataset on private Google Cloud Storage is also provided in the repository.

Received: 29 July 2025; Accepted: 6 October 2025

Published online: 12 November 2025

References

1. Belsky, D. W. et al. Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. *Elife* **9**, e54870. <https://doi.org/10.7554/eLife.54870> (2020).
2. Belsky, D. W. et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. *Elife* **11**, e73420. <https://doi.org/10.7554/eLife.73420> (2022).
3. Bammann, M. M., Petrella, J. K., Kim, J. S., Mayhew, D. L. & Cross, J. M. Cluster analysis tests the importance of myogenic gene expression during myofiber hypertrophy in humans. *J. Appl. Physiol.* **102**, 2232–2239. <https://doi.org/10.1152/japplphysiol.00024.2007> (2007).
4. Ross, R. et al. Precision exercise medicine: Understanding exercise response variability. *Br. J. Sports Med.* **53**, 1141–1153. <https://doi.org/10.1136/bjsports-2018-100328> (2019).
5. Bonafoglia, J. T. et al. Examining interindividual differences in select muscle and whole-body adaptations to continuous endurance training. *Exp. Physiol.* **106**, 2168–2176. <https://doi.org/10.1113/EP089421> (2021).
6. Xi, X. et al. A mechanism-informed deep neural network enables prioritization of regulators that drive cell state transitions. *Nat. Commun.* **16**, 1284. <https://doi.org/10.1038/s41467-025-56475-9> (2025).
7. Duan, Z. et al. iHerd: an integrative hierarchical graph representation learning framework to quantify network changes and prioritize risk genes in disease. *PLoS Comput. Biol.* **19**, e1011444. <https://doi.org/10.1371/journal.pcbi.1011444> (2023).
8. Duan, Z. et al. Impeller: A path-based heterogeneous graph learning method for spatial transcriptomic data imputation. *Bioinformatics* **40**, btae339. <https://doi.org/10.1093/bioinformatics/btae339> (2024).
9. Duan, Z., Xu, S., Lee, C., Riffle, D. & Zhang, J. iMIRACLE: An iterative multi-view graph neural network to model intercellular gene regulation from spatial transcriptomic data. In *Proceedings of the ... ACM International Conference on Information & Knowledge Management. ACM International Conference on Information and Knowledge Management* 538–548. <https://doi.org/10.1145/3627673.3679574> (2024).
10. Duan, Z. et al. scENCORE: Leveraging single-cell epigenetic data to predict chromatin conformation using graph embedding. *Brief. Bioinform.* **25**, bbae096. <https://doi.org/10.1093/bib/bbae096> (2024).
11. Zhang, Y., Feng, X., Wang, Y. & Shi, K. Deep learning powered single-cell clustering framework with enhanced accuracy and stability. *Sci. Rep.* **15**, 4107. <https://doi.org/10.1038/s41598-025-87672-7> (2025).
12. Sefer, E. DRGAT: Predicting drug responses via diffusion-based graph attention network. *J. Comput. Biol.* **32**, 330–350. <https://doi.org/10.1089/cmb.2024.0807> (2025).

13. Ohsawa, I. & Kawano, F. Chronic exercise training activates histone turnover in mouse skeletal muscle fibers. *FASEB J.* **35**, e21453. <https://doi.org/10.1096/fj.202002027RR> (2021).
14. Lee, J. et al. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **36**, 1234–1240. <https://doi.org/10.1093/bioinformatics/btz682> (2020).
15. Fey, M. & Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. In *ICLR*. <https://doi.org/10.48550/arXiv.1903.02428> (2019).
16. Veličković, P. et al. Graph attention networks. In *ICLR*. <https://doi.org/10.48550/arXiv.1710.10903> (2018).
17. Zhou, J. et al. Graph neural networks: A review of methods and applications. *AI Open* **1**, 57–81. <https://doi.org/10.1016/j.aiopen.2021.01.001> (2020).
18. Kingma, D. P. & Ba, J. L. Adam: A method for stochastic optimization. *ICLR* <https://doi.org/10.48550/arXiv.1412.6980> (2015).
19. De Rainville, F.-M., Fortin, F.-A., Gardner, M.-A., Parizeau, M. & Gagné, C. DEAP: A Python framework for evolutionary algorithms. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation (GECCO 2012)* 85–92. <https://doi.org/10.1145/2330784.2330799> (2012).
20. Huang, D. W. et al. DAVID Bioinformatics Resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* **35**, W169–175. <https://doi.org/10.1093/nar/gkm415> (2007).
21. Letts, J. A., Fiedorczuk, K. & Sazanov, L. A. The architecture of respiratory supercomplexes. *Nature* **537**, 644–648. <https://doi.org/10.1038/nature19774> (2016).
22. Wu, M., Gu, J., Guo, R., Huang, Y. & Yang, M. Structure of mammalian respiratory supercomplex I(1)III(2)IV(1). *Cell* **167**, 1598–1609. <https://doi.org/10.1016/j.cell.2016.11.012> (2016).
23. Burska, D. et al. Heterozygous missense mutation in UQCRC2 associated with severe encephalomyopathy, mitochondrial complex III assembly defect and activation of mitochondrial protein quality control. *Biochim. Biophys. Acta Mol. Basis Dis.* **1867**, 166147. <https://doi.org/10.1016/j.bbadic.2021.166147> (2021).
24. Gaillard, P. et al. UQCRC2 mutation in a patient with mitochondrial complex III deficiency causing recurrent liver failure, lactic acidosis and hypoglycemia. *J. Hum. Genet.* **62**, 729–731. <https://doi.org/10.1038/jhg.2017.22> (2017).
25. Miyake, N. et al. Mitochondrial complex III deficiency caused by a homozygous UQCRC2 mutation presenting with neonatal-onset recurrent metabolic decompensation. *Hum. Mutat.* **34**, 446–452. <https://doi.org/10.1002/humu.22257> (2013).
26. Granata, C. et al. High-intensity training induces non-stoichiometric changes in the mitochondrial proteome of human skeletal muscle without reorganisation of respiratory chain content. *Nat. Commun.* **12**, 7056. <https://doi.org/10.1038/s41467-021-27153-3> (2021).

Author contributions

F.K. conceptualized the study, designed the framework of the graph neural network model, wrote all code, validated the model, analyzed the data, drafted the manuscript, and finalized it.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-23320-4>.

Correspondence and requests for materials should be addressed to F.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025