REDIM: DECODABLE MULTI-VIEW REPRESENTATION LEARNING OF RNA–PROTEIN DISCORDANCE WITH SPATIAL AND METABOLIC CONTEXT

Anonymous authors
Paper under double-blind review

ABSTRACT

Spatial multi-omics connects transcriptomic, proteomic, and metabolic organization within intact tissue, but spatial proteomics remains more limited in molecular coverage and availability than spatial transcriptomics. Predicting spatial protein abundance from RNA is essential for bridging this measurement gap. However, RNA is not a quantitative proxy for protein: translation, degradation, and metabolic state create context-dependent discordance across tissue regions. Existing predictors typically learn a direct RNA-to-protein mapping and absorb such deviations into undifferentiated prediction error. We introduce ReDiM, a decodable representation-learning model that casts context-dependent RNA–protein correction as a supervised, low-dimensional learning problem. ReDiM freezes a reference RNA-to-protein predictor, defines correction as the difference between observed and reference-expected protein abundance, and represents it in a training-fitted 16-dimensional protein basis. It predicts correction coordinates from three RNA-accessible views: reference protein expectation, spatially contextualized RNA, and RNA-derived reaction-level flux enrichment scores (FES). FES maps RNA onto a genome-scale metabolic network, enabling FES-associated corrections to be traced across protein, reaction, and pathway levels. A fixed decoder maps corrections to a common 31-protein space. Across held-out liver sections, amyotrophic lateral sclerosis (ALS) donors, and brain regions, training-derived bases reconstruct held-out discordance, and ReDiM improves prediction beyond the frozen reference. Its two contextual inputs play distinct roles: spatial RNA drives the recovery of protein rank structure, whereas FES further refines absolute prediction error, with larger gains in target spots that differ more from the training data. To characterize this metabolic contribution, we analyze FES attribution in ALS and identify protein- and pathway-level correction patterns that recur across donors. By learning transferable RNA–protein corrections from complementary spatial and metabolic context, ReDiM improves protein prediction in unseen biological domains. ReDiM thus recasts spatial RNA-to-protein prediction as learning a transferable, decodable correction rather than treating discordance as undifferentiated error.

1 INTRODUCTION

Spatial multi-omics aims to connect transcriptomic, proteomic, and metabolic states within intact tissue, revealing how molecular programs vary across cellular neighborhoods and anatomical regions (Longo et al., 2021; Moses & Pachter, 2022). Yet these molecular layers are measured at markedly different scales. Spatial transcriptomic assays can profile thousands of genes, whereas paired RNA–protein platforms such as SPOTS measure predefined protein panels at matched tissue locations (Ben-Chetrit et al., 2023). This imbalance makes spatial protein coverage a major bottleneck for spatial multi-omics. Accurate RNA-to-protein prediction could expand the utility of spatial transcriptomic atlases beyond RNA-level analysis, providing protein-level hypotheses even where matched proteomic measurements are unavailable. STProtein and DGAT use graph neural networks to combine RNA expression with spatial neighborhoods and directly predict protein abundance (Jiang et al., 2026; Wang et al., 2026). However, they do not explicitly isolate or decode the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

context-dependent correction to a reference RNA-to-protein mapping. The transfer problem extends beyond filling a missing modality: a model trained in one biological setting must remain useful across donors, diseases, and anatomical regions, where the quantitative RNA–protein relationship may shift.

The central dogma specifies the direction of genetic information flow, but it does not imply a quantitative or spatial correspondence between RNA and protein abundance. Protein abundance is determined not only by RNA availability but also by translation efficiency and protein turnover (Schwanhäusser et al., 2011; Vogel & Marcotte, 2012; Buccitelli & Selbach, 2020). These processes respond to cellular context: immune activation reshapes protein synthesis and degradation (Jovanovic et al., 2015), while nutrient-sensitive mTORC1 signaling couples translation to metabolic conditions (Thoreen et al., 2012). Similar RNA profiles can therefore produce different protein states across diseases, treatments, and tissue regions. At spot resolution, differences in cell-type composition provide an additional source of mismatch.

This discordance is not merely noise; it can encode a structured, context-dependent departure from the reference RNA–protein relationship. Existing RNA-to-protein predictors learn the observed protein profile directly, so deviations from the learned mapping are absorbed into the overall prediction error. This formulation obscures whether the correction is shared across samples, transferable to unseen domains, or traceable to biological context. The resulting machine-learning problem is therefore to disentangle a stable reference mapping from a context-dependent correction.

To address this problem, we introduce ReDiM, a multi-view model for learning predictable RNA–protein discordance, which factorizes spatial protein prediction into a frozen reference expectation and a learned context-dependent correction. First, an RNA-to-protein model is trained on paired reference data and then frozen. Second, RNA from a target spot is passed through this model to obtain the protein profile expected from the reference relationship, P_0. Third, paired target measurements are used to define discordance as the difference between the observed protein profile and this expectation, $D = P - P_0$. ReDiM represents D in a low-dimensional protein basis and learns to predict it from three views: P_0, spatially contextualized target RNA, and RNA-derived reaction-level flux enrichment scores (FES). A fixed decoder maps the predicted coordinates back to protein space, yielding the final protein prediction $\widehat{P} = P_0 + \widehat{D}$. At inference, ReDiM requires only RNA and spatial coordinates. This factorization turns what direct predictors treat as undifferentiated error into a supervised representation that can be evaluated for cross-domain transfer, decoded into protein-specific corrections, and traced to contextual inputs.

We evaluate ReDiM in three biologically distinct domain-transfer settings: across liver sections in alcohol-associated hepatitis, across donors in sporadic amyotrophic lateral sclerosis (ALS), and across anatomical regions in the brain. In each setting, an entire biological domain is held out, and all preprocessing transformations, protein-basis fitting, model training, and model-selection decisions are restricted to the training domains. This leave-one-domain-out design tests whether correction structure learned in one biological context remains useful in an unseen one. We evaluate not only final protein prediction but also the transferability and interpretability of the learned correction representation. Specifically, we examine whether a training-derived basis reconstructs held-out discordance, whether decoded corrections exhibit recurrent protein-space geometry across domains, and which contextual views drive prediction. Finally, we use branch-specific ablation and decoding to trace the FES-associated correction to proteins and pathways and test its recurrence across ALS donors. Together, these analyses evaluate ReDiM as a transferable and inspectable correction model, not only as an endpoint predictor.

We summarize our contributions as follows:

- We formulate spatial RNA-to-protein prediction as reference-relative correction learning, in which the difference between observed and reference-expected protein abundance becomes an explicit, low-dimensional, and decodable learning target.

- We develop ReDiM, which predicts correction coordinates from the expected protein profile, spatially contextualized RNA, and RNA-derived reaction-level FES, and maps them through a fixed decoder into a common protein space. The model requires only RNA and spatial coordinates at inference.

- We evaluate both endpoint prediction and the learned correction representation across nine held-out biological domains using reconstruction, protein-space geometry, component ablation, and attribution analyses.
- We characterize the contexts in which metabolic information provides complementary absolute-error refinement and trace its decoded contribution to recurrent protein and pathway patterns across ALS donors.

2 Related Work

RNA–protein measurement and prediction. CITE-seq established paired measurements of transcript abundance and surface-protein epitopes in single cells (Stoeckius et al., 2017), while SPOTS extended joint RNA–protein profiling to spatially matched tissue locations (Ben-Chetrit et al., 2023). These technologies have supported computational models of the relationship between the two modalities. For example, totalVI uses a probabilistic latent-variable model to jointly represent RNA and protein observations while accounting for modality-specific technical variation (Gayoso et al., 2021). More broadly, scVI and MOFA+ demonstrate how latent-variable models can learn compact representations of high-dimensional single-cell and multimodal measurements (Lopez et al., 2018; Argelaguet et al., 2020). Such approaches treat protein abundance as a directly predicted or jointly modeled modality; they do not explicitly separate a reference RNA-to-protein relationship from its context-dependent deviation. ReDiM instead freezes a reference RNA-to-protein mapping and learns the remaining correction as a supervised, low-dimensional, and decodable representation. This makes the correction itself, not only the final protein prediction, an object of cross-domain evaluation and attribution.

Spatial representation and domain transfer. Spatial models use tissue adjacency to distinguish molecular states that may appear similar without spatial context. SpaGCN and STAGATE established graph-based strategies for combining expression with spatial neighborhoods (Hu et al., 2021; Dong & Zhang, 2022). In spatial multi-omics, SpatialGlue learns integrated representations for identifying tissue domains across modalities (Long et al., 2024), whereas STProtein and DGAT use graph-based representations to predict spatial protein abundance from transcriptomic measurements (Jiang et al., 2026; Wang et al., 2026). Transfer-learning methods such as scArches instead use a model trained on an existing reference atlas to represent observations from a new query dataset in the same latent space (Lotfollahi et al., 2022). In ReDiM, the frozen RNA-to-protein predictor provides the reference anchor, while a spatial encoder learns the correction required in a new biological domain. A fixed protein-space decoder makes these corrections directly comparable across held-out domains.

3 Method

ReDiM models the difference between observed target protein abundance and the abundance expected from a frozen RNA-to-protein model. Let $X_i \in \mathbb{R}^G$ denote the RNA profile of spot i, where G is the number of RNA input features, and let $P_i \in \mathbb{R}^M$ denote its observed protein profile, where $M = 31$ in our experiments. We first fit a reference model f_{ref} on paired reference RNA and protein measurements and then freeze its parameters. For a target spot, the frozen model maps its RNA profile to a spot-specific expected protein profile, $P_{0,i} = f_{\mathrm{ref}}(X_i^{\mathrm{tar}})$.

We define RNA–protein discordance as the signed difference between the observed and expected protein profiles,

$$D_i = P_i - P_{0,i}. \tag{1}$$

ReDiM represents D_i in a training-derived low-dimensional protein basis and predicts its latent coordinates from the expected protein profile, spatially contextualized target RNA, and RNA-derived FES. At inference, ReDiM receives test spatial transcriptomics and spatial coordinates. The frozen reference model produces $P_{0,i}$ from the test RNA, and FES is derived from the same transcriptomic input. ReDiM then predicts $\widehat{D}_i$, decodes it in protein space, and obtains the final prediction as $\widehat{P}_i = P_{0,i} + \widehat{D}_i$. Test spatial proteomics is used only for evaluation. Figure 1 summarizes the complete procedure.

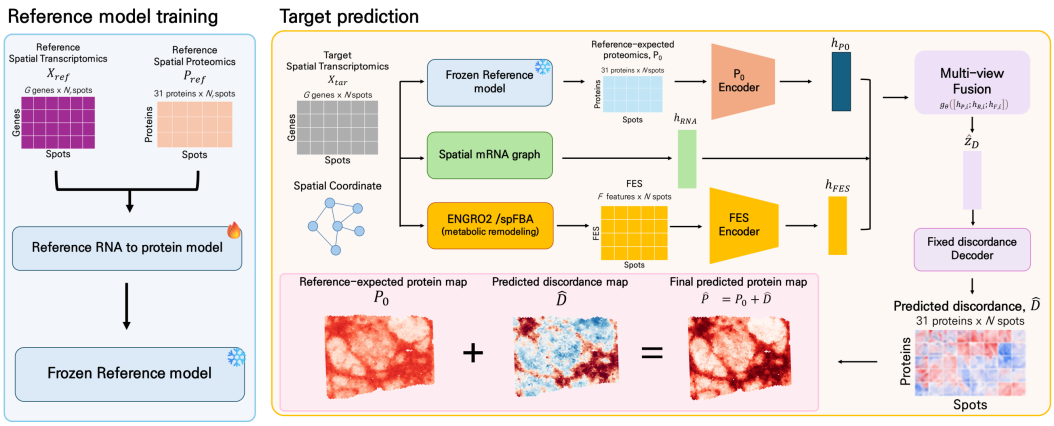


Figure 1: Overview of ReDiM. Paired reference RNA and protein measurements train an RNA-to-protein reference model, which is frozen before correction learning. Target RNA is processed in parallel by the frozen reference model, the spatial-RNA graph encoder using tissue coordinates, and the ENGRO2–spFBA workflow that produces RNA-derived FES. Encoded reference, spatial-RNA, and FES views are fused to predict 16-dimensional discordance coordinates. A fixed decoder maps the prediction to the common 31-protein space, and the decoded discordance is added to the reference expectation to obtain the final protein prediction. Held-out protein measurements are used only for evaluation.

3.1 Reference expectation and discordance

The reference model f_{ref} is a multilayer perceptron that maps reference RNA to the common measured protein panel. ALD and ALS use their designated paired reference data. Brain uses a region-matched reference model for each anatomical region. Once fitted, f_{ref} is frozen and is not updated while the discordance predictor is trained.

For each target spot i, we apply the frozen reference model to the target RNA profile, $P_{0,i} = f_{\text{ref}}(X_i^{\text{tar}})$. Unlike a reference-level average, $P_{0,i}$ is spot-specific and represents the protein profile expected from the RNA state of target spot i under the reference RNA–protein relationship. We define RNA–protein discordance as $D_i = P_i - P_{0,i}$. Thus, D_i captures the deviation of the observed target protein profile from the abundance expected under the reference relationship. Target protein measurements are used to construct D_i during training, but are not required at inference time.

3.2 Training-derived discordance basis

Protein deviations may vary coordinately across multiple markers. We therefore parameterize D_i in a compact coordinate system learned from training-domain discordance. Let $\mu_D \in \mathbb{R}^M$ and $W_D \in \mathbb{R}^{K\times M}$ denote the mean and the first K principal directions obtained by fitting PCA only on discordance profiles from the training target domains. The discordance representation is

$$z_{D,i} = (D_i - \mu_D)W_D^\top. \tag{2}$$

The fixed basis provides a common low-dimensional representation of coordinated protein deviations while retaining an explicit linear mapping back to the measured protein space. Given a predicted discordance representation $\widehat{z}_{D,i}$, the corresponding protein-space discordance is reconstructed as

$$\widehat{D}_i = \mu_D + \widehat{z}_{D,i}W_D. \tag{3}$$

This construction follows the standard PCA representation and reconstruction framework (Jolliffe, 2002).

3.3 Multi-view discordance predictor

Reference anchor and spatial RNA. An MLP E_P encodes $P_{0,i}$. The RNA transformation $R_i = T_R(X_i^{\text{tar}})$ produces a fixed 256-component representation. Let $s_i \in \mathbb{R}^2$ denote the spatial coordinate

of spot i. Within each tissue section, we construct a symmetric k-nearest-neighbor graph from these coordinates with $k = 8$. A two-head graph-attention encoder E_R then contextualizes R_i with its spatial neighbors (Veličković et al., 2018).

$$\begin{aligned} h_{P,i} &= E_P(P_{0,i}), \\ h_{R,i} &= E_R(R_i, G^{\mathrm{sp}}). \end{aligned} \tag{4}$$

RNA-derived metabolic context. Let F_i denote the FES vector computed from the RNA of spot i. ENGRO2 gene–protein–reaction relationships and the spFBA workflow transform RNA into 469 reaction-level enrichment channels (Di Filippo et al., 2022; Maspero et al., 2026). After outer-training filtering and transformation, T_F produces a fixed 32-component representation, which is encoded by an MLP E_F. FES is a network-informed transformation of RNA, not measured metabolite abundance or experimentally observed flux; complete construction details are provided in Appendix B.

$$h_{F,i} = E_F(T_F(F_i)) \,. \tag{5}$$

The fusion network g_θ concatenates the reference, spatial-RNA, and FES branch outputs and predicts the latent discordance coordinates. All three views are available without protein measurements at inference, and all molecular inputs ultimately derive from target RNA.

$$\widehat{z}_{D,i} = g_\theta([h_{P,i}; h_{R,i}; h_{F,i}]) \,. \tag{6}$$

3.4 Learning objective and held-out inference

The fused prediction is supervised in both latent discordance space and decoded protein space. Auxiliary heads g_R and g_F predict the same latent coordinates from the spatial-RNA and FES branches, respectively. Let ℓ denote Huber regression loss (Huber, 1964). We minimize

$$\mathcal{L} = \mathcal{L}_D + \alpha \mathcal{L}_Z + \beta_R \mathcal{L}_R + \beta_F \mathcal{L}_F. \tag{7}$$

Here $\mathcal{L}_D = \ell(\widehat{D}, D)$ is the decoded protein-space loss, $\mathcal{L}_Z = \ell(\widehat{z}_D, z_D)$ is the fused latent loss, and $\mathcal{L}_R = \ell(\widehat{z}_D^R, z_D)$ and $\mathcal{L}_F = \ell(\widehat{z}_D^F, z_D)$ apply the latent loss to the spatial-RNA and FES auxiliary heads. We use $\alpha = 0.5$ and $\beta_R = \beta_F = 0.25$. Together, these objectives align the latent prediction with its decoded protein-space correction. The auxiliary losses encourage the spatial-RNA and FES branches to retain predictive information during training, while the fused prediction produces the final correction at inference.

At inference, test spatial transcriptomics and spatial coordinates are the only measured inputs. The frozen reference model maps the test RNA to the expected protein profile $P_{0,i}$. The same test RNA and coordinates are used to construct the spatial RNA graph, while FES is derived from the test RNA. With all learned parameters and training-fitted transformations fixed, ReDiM predicts the latent discordance, decodes it into protein space as $\widehat{D}_i$, and produces the final protein prediction

$$\widehat{P}_i = P_{0,i} + \widehat{D}_i. \tag{8}$$

All preprocessing transformations, the discordance basis, and ReDiM parameters are fitted without using the held-out biological domain. The complete evaluation protocol, metric definitions, aggregation procedure, and support analysis are provided in Appendix C.

4 Results

4.1 A shared discordance geometry supports accurate cross-domain correction

We first tested the central premise of ReDiM: whether RNA–protein discordance has a compact structure that can be learned from available domains and retained in an unseen biological domain. We evaluated three distinct transfer regimes—two liver sections in ALD, three spinal-cord donors in ALS, and four anatomical regions in Brain. Each dataset uses its own models and training-derived basis, while every analysis is decoded into the same 31-protein output space. Dataset composition and evaluated domain sizes are reported in Appendix A.

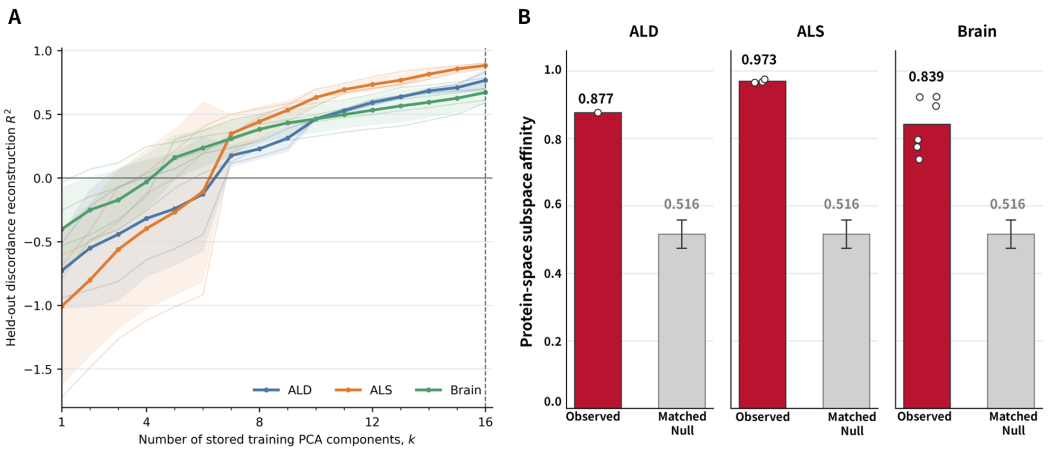


Figure 2: Discordance structure and cross-domain correction geometry. (A) Reconstruction R^2 obtained by projecting the observed discordance of each held-out domain onto PCA bases fitted only to the corresponding training domains. Lines show dataset means across held-out domains, and shaded bands indicate ± 1 SD. The vertical dashed line marks the $k = 16$ basis used by ReDiM. Held-out protein measurements are used only to compute this post hoc representation diagnostic and are not used for model fitting. (B) Protein-space subspace affinity between predicted corrections from pairs of held-out domains after decoding them into the common 31-protein space. Red bars show the mean observed affinity, and white circles show individual domain-pair comparisons. Gray bars show the matched-random expectation, with whiskers indicating the central 95% simulation interval.

Observed discordance is strongly reconstructible across domains. Before evaluating prediction, we asked whether the representation itself transfers. For each outer fold, PCA was fitted only to observed discordance in the training domains and then used to reconstruct observed discordance in the held-out domain. With $k = 16$ components, these training-derived bases reconstructed a substantial fraction of held-out variation: mean R^2 was 0.750 in ALD, 0.888 in ALS, and 0.649 in Brain (Figure 2A). Thus, even across different sections, donors, and brain regions, discordance is far from arbitrary; it occupies a compact protein-space structure that is sufficiently shared to support ReDiM's fixed, training-derived decoder. This is a representation diagnostic rather than a prediction result, and held-out proteins are used only for post hoc evaluation.

ReDiM recovers the shared geometry in its predicted corrections. After decoding predictions into the common 31-protein space, cross-domain subspace affinity was 0.877 in ALD, 0.973 in ALS, and 0.839 in Brain, compared with a matched-random expectation of 0.516 (Figure 2B). The advantage over random subspaces persisted as the comparison rank varied from $r = 4$ to $r = 16$ at a principal-cosine threshold of 0.7 (Figure S1). Figure 2A therefore establishes that observed discordance has transferable low-dimensional structure, while Figure 2B and Figure S1 show that ReDiM learns corrections aligned with that recurring protein-space geometry. End-to-end rank sensitivity is reported in Table A6.

We then tested whether representing discordance in this shared basis yields a practical predictive advantage. Under identical inputs and evaluation, we compared direct prediction of the 31-protein profile, a non-learned correction $P_0 + \mu_D$, direct prediction of the full 31-dimensional residual $D = P - P_0$, and ReDiM's prediction of 16 latent discordance coordinates through the fixed training-derived basis.

ReDiM was the only formulation ranked first or second in every dataset–metric combination across all six cells in Table 1. It achieved the lowest RMSE among learned models in ALD, ALS, and Brain. Relative to direct protein prediction, RMSE decreased by 0.020, 0.003, and 0.009, respectively; relative to direct 31-dimensional residual prediction, it decreased by 0.011, 0.002, and 0.008. ReDiM also achieved the highest ALD rank correlation and remained within 0.004 and 0.003 of the best correlation in ALS and Brain. The non-learned mean-residual correction had lower ALS RMSE, but sacrificed 0.032 in rank correlation relative to ReDiM and generalized poorly on the

Table 1: Matched target and representation ablation. Values are domain-equal mean ± SD across held-out domains after averaging three seeds. ρ is mean protein-wise Spearman correlation; RMSE pools held-out spot–protein errors. Bold and underlined values indicate the best and second-best result, respectively, for each dataset and metric.

Model	ALD $\rho \uparrow$	RMSE $\downarrow$	ALS $\rho \uparrow$	RMSE $\downarrow$	Brain $\rho \uparrow$	RMSE $\downarrow$
ReDiM $K = 16$	**0.499 ± 0.035**	**0.361 ± 0.046**	0.318 ± 0.078	0.332 ± 0.049	0.325 ± 0.045	**0.477 ± 0.215**
Direct 31D residual	0.489 ± 0.040	0.372 ± 0.031	**0.322 ± 0.079**	0.334 ± 0.053	0.322 ± 0.039	0.485 ± 0.226
Direct protein prediction	0.479 ± 0.022	0.381 ± 0.046	0.317 ± 0.067	0.335 ± 0.049	**0.328 ± 0.024**	0.486 ± 0.227
Mean residual correction	0.434 ± 0.028	0.389 ± 0.018	0.286 ± 0.078	**0.310 ± 0.045**	0.276 ± 0.068	0.509 ± 0.216

Table 2: Matched component ablation. P_0 is the frozen reference prediction, M1 adds the spatial-RNA correction, and M3 additionally incorporates RNA-derived FES. Values are domain-equal mean ± SD after averaging the three optimization seeds within each held-out domain. Bold and underlined values indicate the best and second-best result, respectively, for each dataset and metric.

Model	ALD $\rho \uparrow$	RMSE $\downarrow$	ALS $\rho \uparrow$	RMSE $\downarrow$	Brain $\rho \uparrow$	RMSE $\downarrow$
P_0	0.434 ± 0.028	0.400 ± 0.048	0.286 ± 0.078	**0.307 ± 0.045**	0.276 ± 0.068	0.513 ± 0.206
M1	**0.506 ± 0.037**	0.362 ± 0.048	0.310 ± 0.069	0.338 ± 0.051	**0.327 ± 0.037**	0.481 ± 0.220
M3	0.499 ± 0.035	**0.361 ± 0.046**	**0.318 ± 0.078**	0.332 ± 0.049	0.325 ± 0.045	**0.477 ± 0.215**

other datasets. Taken together, the matched comparison supports the low-dimensional, reference-relative parameterization as the most consistently strong representation across transfer settings.

4.2 Spatial and metabolic context play complementary roles

To identify the source of these gains, we compared three configurations under the same outer folds and optimization seeds. P_0 is the frozen reference prediction; M1 learns a correction from spatially contextualized RNA; and M3 adds the FES branch. Because each configuration is trained separately, the comparison measures the contribution of the complete configuration rather than an isolated test-time input removal.

Spatial RNA supplies the dominant rank-recovery signal. Moving from P_0 to M1 increased mean protein-wise Spearman correlation by 0.072 in ALD, 0.024 in ALS, and 0.051 in Brain. The gain appears in every dataset, showing that spatial context consistently helps recover how protein abundance is reordered across spots beyond the frozen RNA-to-protein reference.

FES adds a complementary absolute-error refinement. Relative to M1, M3 reduced global RMSE in every dataset summary, by 0.0008 in ALD, 0.0057 in ALS, and 0.0039 in Brain, while largely preserving rank correlation and further improving it in ALS. Relative to P_0, the complete M3 model retained net correlation gains of 0.065, 0.032, and 0.049 in ALD, ALS, and Brain. The two context views therefore make distinct contributions: spatial RNA establishes the major rank-structure correction, whereas FES sharpens the magnitude of the decoded protein correction.

The FES improvement was not driven by a single summary outlier. M3 achieved lower RMSE than M1 in 6 of 9 held-out domains—1 of 2 ALD sections, all 3 ALS donors, and 2 of 4 Brain regions—with the complete domain-level comparison shown in Figure S3. The effect is strongest and most consistent in ALS, but remains positive at the dataset-summary level in all three transfer settings.

Capacity-matched controls strengthen this interpretation. Canonical FES produced the lowest RMSE of all matched third-view models in ALD, ALS, and Brain (Table A5), so the result is not explained by merely adding another 32-dimensional RNA representation or an extra encoder. In a separately matched three-seed shuffle rerun, real FES outperformed shuffled FES in ALD (0.359 versus 0.363) and ALS (0.334 versus 0.338), tied at the reported precision in Brain (0.480), and was better in 7 of 9 held-out domains. The spot-specific FES contribution is therefore modest and dataset dependent, but reproducible enough to improve absolute prediction beyond both M1 and capacity-matched alternatives.

Table 3: Comparison with spatial protein-prediction methods. Columns report mean protein-wise Spearman correlation (ρ) and global RMSE. Bold and underlined values indicate the best and second-best result, respectively, for each dataset and metric.

Model	ALD $\rho \uparrow$	RMSE $\downarrow$	ALS $\rho \uparrow$	RMSE $\downarrow$	Brain $\rho \uparrow$	RMSE $\downarrow$
STProtein (Jiang et al., 2026)	0.360 ± 0.060	0.413 ± 0.043	0.249 ± 0.045	0.315 ± 0.038	0.223 ± 0.059	0.510 ± 0.223
totalVI (Gayoso et al., 2021)	$\mathbf{0.501 \pm 0.022}$	0.365 ± 0.037	0.294 ± 0.082	$\mathbf{0.298 \pm 0.034}$	0.172 ± 0.106	0.595 ± 0.275
DGAT (Wang et al., 2026)	0.369 ± 0.003	$\underline{0.362 \pm 0.080}$	$\underline{0.303 \pm 0.082}$	$\underline{0.306 \pm 0.060}$	$\underline{0.316 \pm 0.015}$	$\mathbf{0.475 \pm 0.212}$
ReDiM M3 (ours)	$\underline{0.499 \pm 0.035}$	$\mathbf{0.361 \pm 0.046}$	$\mathbf{0.318 \pm 0.078}$	0.332 ± 0.049	$\mathbf{0.325 \pm 0.045}$	$\underline{0.477 \pm 0.215}$

4.3 ReDiM delivers consistently strong cross-domain protein prediction

We next evaluated whether the complete ReDiM model translates these representation and component gains into competitive endpoint prediction. ReDiM M3 was compared with STProtein, totalVI, and DGAT using the same held-out biological domains and common 31-protein panel. Appendix E documents baseline implementation, training, and evaluation details.

ReDiM ranked first or second in five of the six dataset–metric comparisons in Table 3, the broadest top-two coverage among the evaluated methods. It achieved the highest protein-wise rank correlation in ALS (0.318) and Brain (0.325); in ALD, its correlation of 0.499 was only 0.002 below totalVI. ReDiM also achieved the lowest ALD RMSE (0.361) and the second-lowest Brain RMSE (0.477), only 0.002 above DGAT. ALS RMSE was the single exception, where totalVI and DGAT performed better. Overall, ReDiM combines leading rank recovery with consistently competitive absolute error across three distinct forms of biological domain transfer.

The external comparison and matched ablations support a coherent result: ReDiM is not dependent on one dataset or one metric to appear competitive. Its reference-relative representation yields consistently strong cross-domain protein prediction; spatial RNA accounts for the main recovery of protein ordering; and FES supplies additional absolute-error refinement. This complementarity motivates examining where the FES contribution is largest.

4.4 FES gains concentrate in target states with limited training support

We asked whether FES is most useful precisely where correction transfer is hardest. For each held-out spot, we measured the mean Euclidean distance between its reference-expected protein profile P_0 and its 50 nearest correction-training P_0 profiles, then formed within-domain support terciles (Figure 3A).

FES gains concentrate in under-supported target states. Although baseline P_0 error rose most clearly with support distance in Brain (Figure 3B), the poorly supported tercile showed the largest M3-over-M1 RMSE gain in all three datasets (Figure 3C). The same relationship held continuously: support distance was positively associated with FES gain in 8 of 9 held-out domains, with Brain tcx as the sole exception (Figure 3D).

This repeated pattern indicates that FES is most valuable when ReDiM must extend its learned correction beyond well-represented training states. The effect is not universal at every domain, but its consistency across datasets and in 8 of 9 continuous domain-level tests supports FES as a complementary signal for the most weakly supported targets.

4.5 FES attribution reveals recurrent correction patterns

FES attribution is recurrent across independent ALS donors. We compared each trained M3 model with the same model after zeroing the FES encoder output and decoded the resulting difference into protein space. Twelve of 31 proteins had the same attribution sign in all three donors, and 26 of 31 agreed in at least two of three donors (Figure 4A and Figure S2). Among the fully concordant proteins, BCL2, CD19, and CD4 were positive, whereas KRT5, CD68, HLA-DRA, and CD163 were negative. Donor-level attribution vectors showed partial but clear alignment (cosine similarity 0.22–0.51; mean 0.40; Figure 4B), indicating shared structure without erasing donor heterogeneity.

The recurrence extended from proteins to pathways. Across donors, energy, nucleotide, and glutathione-related gene sets showed coordinated positive and negative enrichment (Figure 4C), a pattern more consistent with selective redistribution than with indiscriminate scaling. These model

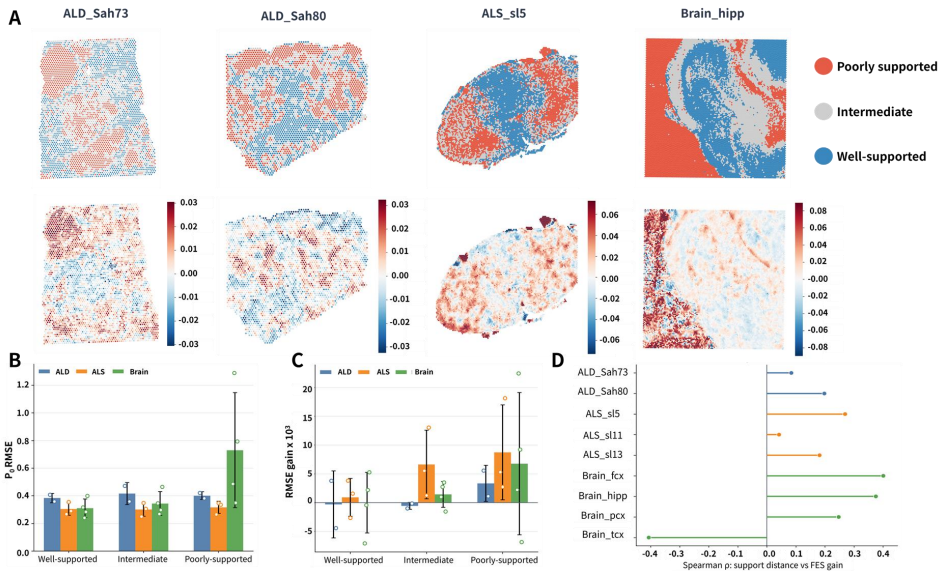


Figure 3: FES gain across training-support contexts. (A) Representative support terciles and spot-level $\mathrm{RMSE(M1)} - \mathrm{RMSE(M3)}$. (B) P_0 RMSE by support tercile. (C) Dataset-level M3-over-M1 gain by tercile; bars show domain-equal means, whiskers ± 1 SD, and points held-out domains. (D) Within-domain Spearman association between continuous support distance and FES gain.

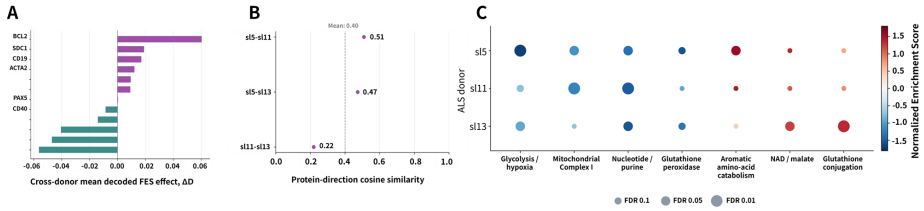


Figure 4: Recurrent ALS FES attributions. (A) Mean decoded effects for 12 proteins with a common sign across donors; Figure S2 shows all proteins. (B) Donor-vector cosine similarity. (C) Selected pathway enrichment. Color denotes NES and point size denotes FDR. Values are model attributions.

attributions align with immune-metabolic heterogeneity reported for the same ALS cohort (Zhang et al., 2026). They do not establish causal metabolic regulation, but they demonstrate that ReDiM's FES contribution can be decoded into biologically coherent, donor-recurrent correction patterns.

5 Conclusion

ReDiM reframes spatial RNA-to-protein prediction as learning a transferable and decodable correction to a frozen reference model. Across nine held-out liver sections, ALS donors, and brain regions, training-derived bases captured shared low-dimensional discordance structure, and ReDiM delivered consistently competitive protein prediction in unseen biological domains. Spatially contextualized RNA drove the recovery of protein rank structure, whereas RNA-derived FES provided complementary reductions in absolute error, with larger gains in target states that were poorly represented during training. Decoding the FES contribution further revealed recurrent protein- and pathway-level attribution patterns across ALS donors. Together, these results show that RNA–protein discordance is not merely undifferentiated prediction error, but contains transferable structure that can be learned, decoded, and exploited for spatial cross-modal prediction under biological domain shift.

Reproducibility Statement

Appendix A specifies the target-domain inventory and evaluation counts. Appendix B records FES construction, preprocessing, and fixed model settings. Appendix C gives evaluation, validation-

graph, and attribution details. Appendix D contains the matched design, metabolic-input, and rank-sensitivity analyses. Appendix E documents the external baselines, Appendix F contains supplementary figures, and Appendix G states the study limitations. The main component comparison uses the same P_0/M1/M3 lineage, a common 31-protein panel, fixed RNA PCA dimensionality of 256, fixed FES PCA dimensionality of 32, fixed discordance rank $K = 16$, and seeds 42, 43, and 44. Seeds and spots are not treated as independent biological replicates.

AI USE STATEMENT

We used large language models (LLMs) to assist with drafting and polishing the manuscript and with limited code-level troubleshooting, such as identifying implementation errors and suggesting possible fixes. The authors independently conducted the research ideation, method design, experiments, data analysis, and interpretation of results. All AI-assisted text and code suggestions were reviewed, tested, and verified by the authors. The authors take full responsibility for the final content of this work.

REFERENCES

R. Argelaguet et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology*, 21:111, 2020. doi: 10.1186/s13059-020-02015-1.

N. Ben-Chetrit et al. Integration of whole transcriptome spatial profiling with protein markers. *Nature Biotechnology*, 41:788–793, 2023. doi: 10.1038/s41587-022-01536-3.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57:289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.

C. Buccitelli and M. Selbach. mRNAs, proteins and the emerging principles of gene expression control. *Nature Reviews Genetics*, 21:630–644, 2020. doi: 10.1038/s41576-020-0258-4.

M. Di Filippo et al. INTEGRATE: Model-based multi-omics data integration to characterize multilevel metabolic regulation. *PLOS Computational Biology*, 18:e1009337, 2022. doi: 10.1371/journal.pcbi.1009337.

K. Dong and S. Zhang. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature Communications*, 13:1739, 2022. doi: 10.1038/s41467-022-29439-6.

A. Gayoso et al. Joint probabilistic modeling of single-cell multi-omic data with totalvi. *Nature Methods*, 18:272–282, 2021. doi: 10.1038/s41592-020-01050-x.

Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research*, 49:D325–D334, 2021. doi: 10.1093/nar/gkaa1113.

M. Gillespie et al. The Reactome pathway knowledgebase 2022. *Nucleic Acids Research*, 50: D687–D692, 2022. doi: 10.1093/nar/gkab1028.

I. Gulrajani and D. Lopez-Paz. In search of lost domain generalization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=lQdXeXDoWtI.

Y. Hao et al. Integrated analysis of multimodal single-cell data. *Cell*, 184:3573–3587.e29, 2021. doi: 10.1016/j.cell.2021.04.048.

J. Hu et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature Methods*, 18:1342–1351, 2021. doi: 10.1038/s41592-021-01255-8.

P. J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35: 73–101, 1964. doi: 10.1214/aoms/1177703732.

Z. Jiang, Y. Yuan, L. Hu, and W. Pang. STProtein: predicting spatial protein expression from multi-omics data. *arXiv preprint arXiv:2602.05811*, 2026. URL https://arxiv.org/abs/2602.05811.

I. T. Jolliffe. *Principal Component Analysis*. Springer, 2 edition, 2002. doi: 10.1007/b98835.

M. Jovanovic et al. Dynamic profiling of the protein life cycle in response to pathogens. *Science*, 347:1259038, 2015. doi: 10.1126/science.1259038.

M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, and M. Tanabe. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Research*, 49:D545–D551, 2021. doi: 10.1093/nar/gkaa970.

M. V. Kuleshov et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44:W90–W97, 2016. doi: 10.1093/nar/gkw377.

A. Liberzon et al. The molecular signatures database hallmark gene set collection. *Cell Systems*, 1: 417–425, 2015. doi: 10.1016/j.cels.2015.12.004.

Y. Long et al. Deciphering spatial domains from spatial multi-omics with SpatialGlue. *Nature Methods*, 21:1658–1667, 2024. doi: 10.1038/s41592-024-02316-4.

S. K. Longo, M. G. Guo, A. L. Ji, and P. A. Khavari. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nature Reviews Genetics*, 22:627–644, 2021. doi: 10.1038/s41576-021-00370-8.

R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, and N. Yosef. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15:1053–1058, 2018. doi: 10.1038/s41592-018-0229-2.

I. Loshchilov and F. Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

M. Lotfollahi et al. Mapping single-cell data to reference atlases by transfer learning. *Nature Biotechnology*, 40:121–130, 2022. doi: 10.1038/s41587-021-01001-7.

D. Maspero et al. Spatial FBA reveals heterogeneous warburg niches in renal tumors and lactate consumption in colorectal cancer. *npj Systems Biology and Applications*, 12:32, 2026. doi: 10.1038/s41540-026-00654-x.

L. Moses and L. Pachter. Museum of spatial transcriptomics. *Nature Methods*, 19:534–546, 2022. doi: 10.1038/s41592-022-01409-2.

NCBI Gene Expression Omnibus. GSE278662: Visium spatial transcriptomics and proteomics identifies novel hepatic cell populations and transcriptomic signatures of alcohol-associated hepatitis. Dataset, 2024. URL https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE278662.

J. D. Orth, I. Thiele, and B. O. Palsson. What is flux balance analysis? *Nature Biotechnology*, 28: 245–248, 2010. doi: 10.1038/nbt.1614.

F. Pedregosa et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. URL https://jmlr.org/papers/v12/pedregosa11a.html.

B. Schwanhäusser et al. Global quantification of mammalian gene expression control. *Nature*, 473: 337–342, 2011. doi: 10.1038/nature10098.

M. Stoeckius et al. Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods*, 14:865–868, 2017. doi: 10.1038/nmeth.4380.

A. Subramanian et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102:15545–15550, 2005. doi: 10.1073/pnas.0506580102.

C. C. Thoreen et al. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature*, 485:109–113, 2012. doi: 10.1038/nature11083.

L. van Olst et al. Microglial mechanisms drive amyloid-β clearance in immunized patients with alzheimer's disease. *Nature Medicine*, 31:1604–1616, 2025. doi: 10.1038/s41591-025-03574-1.

P. Veličković et al. Graph attention networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

C. Vogel and E. M. Marcotte. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13:227–232, 2012. doi: 10.1038/nrg3185.

H. Wang et al. DGAT: a dual-graph attention network for inferring spatial protein landscapes from transcriptomics. *Nature Communications*, 17:6702, 2026. doi: 10.1038/s41467-026-73114-z.

Z. Zhang et al. Integrated single-cell and spatial transcriptomic profiling in ALS uncovers peripheral-to-central immune infiltration and reprogramming. *Nature Neuroscience*, 29:1735–1748, 2026. doi: 10.1038/s41593-026-02300-5.

A DATASETS AND TARGET-DOMAIN EVALUATION

The collections are analyzed separately. Domain counts describe target evaluation units, not the number of independent cohorts or reference samples. RNA feature counts precede the training-fitted reductions. Evaluated spots are accumulated over outer held-out targets and are not training-sample counts.

Table A1: Dataset inventory and evaluation units.

Dataset	GEO	Target domains	RNA features	Evaluated spots
ALD	GSE278662	2 liver sections	37,082	7,299
ALS	GSE288365	3 spinal-cord donors	18,085	25,062
Brain	GSE263079	4 brain regions	18,085	57,344

Note. Dataset sources: ALD, NCBI GEO (NCBI Gene Expression Omnibus, 2024); ALS, Zhang et al. (2026); Brain, van Olst et al. (2025).

ALD targets are sections Sah73 and Sah80, ALS targets are donors sl5, sl11, and sl13, and Brain targets are fcx, hipp, pcx, and tcx from the lecanemab-treated AD analysis. ALD and ALS use their designated reference pipelines. Brain uses a region-matched P_0 fitted from nAD reference donors in the corresponding anatomy, while current-RNA PCA uses one basis fitted to the pooled outer-training regions. The superseded per-region current-RNA basis is not part of the canonical lineage. All datasets use the same ordered 31-protein CLR panel and aligned reaction coordinates. Brain regions must not be interpreted as four independent patients.

Table A2: Evaluated target-domain sizes.

Dataset	Domain	Evaluated spots	Proteins
ALD	Sah73	3,947	31
ALD	Sah80	3,352	31
ALS	sl5	8,272	31
ALS	sl11	11,634	31
ALS	sl13	5,156	31
Brain	fcx	14,336	31
Brain	hipp	14,336	31
Brain	pcx	14,336	31
Brain	tcx	14,336	31

All three optimization seeds are averaged within each target domain before dataset summaries are calculated. Repeated seeds, spatial spots, and pairwise comparisons increase the number of technical observations but do not increase the biological sample size. Region dependence and shared training folds also limit interpretation of across-domain SDs as population uncertainty.

B FES CONSTRUCTION AND IMPLEMENTATION

FES is an RNA-derived feature matrix computed with the existing ENGRO2 model and spFBA workflow. Flux balance analysis represents feasible reaction activity under stoichiometric and steady-state constraints (Orth et al., 2010). The published spFBA method samples these feasible flux distributions and summarizes relative reaction usage (Maspero et al., 2026). The raw output contains 469 reaction channels. Outer-training filtering retains 467 channels, followed by imputation, standardization, clipping, and a fixed 32-component PCA. FES depends on network, expression-processing, constraint, and sampling choices. It is neither measured metabolite abundance nor experimentally observed reaction flux.

Table A3: FES construction and fold-specific transformations.

Stage	Operation	Output	Control
1. Gene harmonization	Target RNA identifiers mapped to genes used by ENGRO2 GPR rules	Per-sample RNA `.h5ad`	No protein target used
2. Reaction activity	ENGRO2 GPR mapping (Di Filippo et al., 2022); OR and AND logic handled by spFBA	Reaction activity/constraints	Computed from RNA only
3. Flux inference	spFBA on the existing ENGRO2 network (Maspero et al., 2026)	469 FES channels	Same network and reaction order
4. Feature filtering	Invalid or invariant reaction channels filtered	467 channels retained	Fit from outer-training domains
5. Compression	Median imputation, scaling, clipping, and PCA	32 FES PCs	Parameters fit on outer-training domains
6. Model input	FES PCs encoded by an MLP and fused with P_0 and spatial RNA	h_F branch representation	Held-out proteins unavailable at inference

Reaction ordering is kept consistent with the ENGRO2 annotation. Each fold reuses its outer-training-fitted transformations for held-out inputs. Protein labels from the held-out target are excluded from feature construction and prediction. Transcriptome-derived within-sample operations are distinguished from parameters estimated across outer-training domains.

Table A4: Recorded model and optimization settings.

Component	Setting	Implementation
Reference anchor	Frozen non-spatial MLP	Frozen reference MLP; Brain uses region-matched reference models.
RNA transform	256 fixed PCs	Filtering, standardization, clipping, and PCA are fit per outer fold; dimensionality is not selected by validation.
Spatial graph	Coordinate kNN, $k = 8$	Two-head graph-attention encoder (Veličković et al., 2018). Graphs are constructed independently for the training, validation, and held-out test subsets; no edges connect different splits.
FES transform	469 raw $\rightarrow$ 467 retained $\rightarrow$ 32 fixed PCs	ENGRO2/spFBA features with training-fold preprocessing.
Discordance basis	$K = 16$ fixed	PCA is fit to outer-training D after centering by its training mean μ_D.
Fusion	Hidden width 128	Concatenated reference, spatial-RNA, and FES branch outputs.
Optimizer	AdamW (Loshchilov & Hutter, 2019)	Learning rate 10^{-3}; weight decay 10^{-4}; at most 300 epochs.
Early stopping	Patience 30	Inner-validation latent-discordance Huber loss.
Inner split	Spatial-block validation ($\sim 20\%$)	Spatially contiguous validation blocks are selected within the outer-training domains using tissue coordinates. Training and validation use separate induced spatial graphs, with no graph edges across the split.
Loss weights	$\alpha = 0.5$, $\beta_R = \beta_F = 0.25$	Protein-space $\mathcal{L}_D$ has unit weight; α weights latent $\mathcal{L}_Z$; β_R and β_F weight the auxiliary heads.
Seeds	42, 43, 44	Technical replicates, not biological replicates.

B.1 Data preprocessing and normalization

Common pipeline. All feature filtering parameters, scaling statistics, and PCA transformations are fitted using the complete set of outer-training domains and are then applied unchanged to the held-out domain. The inner-validation split is used for model selection and does not affect preprocessing estimation.

RNA. Counts are library-size normalized, log transformed, filtered, standardized using outer-training statistics, clipped, and projected onto 256 principal components. The PCA basis is fitted separately within each outer fold, while the retained dimensionality is fixed at 256 in every experiment and is not selected by inner validation. PCA and related preprocessing operations were implemented with scikit-learn (Pedregosa et al., 2011).

Protein. Isotype-control channels are removed before spot-wise centered-log-ratio transformation (Hao et al., 2021), producing the common 31-protein panel. Held-out-domain protein measurements are used only for evaluation and are unavailable to the model at inference.

RNA-derived metabolic context. RNA is mapped to the existing ENGRO2 network through the spFBA pipeline, yielding 469 reaction-level FES channels per spot. Outer-training-fold filtering retains 467 channels for the audited model inputs. These channels are imputed, standardized, clipped, and projected onto 32 fixed principal components. FES represents an RNA-derived metabolic estimate rather than measured metabolite abundance or experimentally observed flux.

C Evaluation details and attribution methods

Mean protein-wise Spearman correlation is computed across held-out spots for each of the 31 proteins and then averaged across proteins. Global RMSE pools all finite spot–protein errors within each held-out run. Technical seeds are averaged within a held-out domain, and biological domains receive equal weight in dataset summaries. FES gain is $\mathrm{RMSE}(\mathrm{M1}) - \mathrm{RMSE}(\mathrm{M3})$, so a positive value favors M3. Calibration and spatial statistics, where reported, are computed within held-out domains and are not treated as additional biological replicates.

To test whether discordance structure learned from the training domains transfers to an unseen domain, we fitted a PCA basis to the observed discordance in the outer-training domains and used it to reconstruct the observed discordance in the held-out domain. Figure 2A reports the resulting reconstruction R^2 as the number of retained components increases from $k = 1$ to $k = 16$. Because only the first 16 training-derived components were retained, this analysis evaluates reconstruction within the available basis and does not determine whether $k = 16$ is the full-rank or saturation point. Held-out protein measurements were used only to evaluate this representation after training; Figure 2A therefore measures the transferability of the discordance basis, not the accuracy of ReDiM's RNA-based predictions.

For the cross-domain subspace analysis, predicted corrections were decoded into the common 31-protein space, and their leading protein-space subspaces were compared using principal-angle cosines. Figure 2B reports subspace affinity relative to dimension-matched random subspaces. To assess sensitivity to the comparison rank, we varied the number of leading directions from $r = 4$ to $r = 16$ and calculated the fraction of principal cosines greater than or equal to 0.7 (Figure S1). The matched-random baseline was recomputed at each rank because the expected overlap between random subspaces increases with their dimensionality.

C.1 Validation and graph construction

Each outer fold holds out one complete biological domain for evaluation. Within the remaining domains, spatially contiguous blocks are reserved for inner validation. Training and validation spots are represented by separate induced spatial graphs, so message passing does not cross the split. The held-out domain is represented by an independent graph constructed only from its own RNA measurements and spatial coordinates. Inner validation is used for checkpoint selection based on latent-discordance Huber loss, while the held-out domain is not used for model selection, following the separation of model selection and domain-level testing emphasized in domain-generalization evaluation (Gulrajani & Lopez-Paz, 2021).

C.2 Exploratory ALS attribution and enrichment

The exploratory ALS attribution in Figure 4 starts from a frozen full M3 model. The FES branch hidden representation is set to zero without retraining, producing $\Delta z_{\mathrm{FES}} = z_{\mathrm{full}} - z_{h_F\text{-zero}}$. The fixed 16-component discordance PCA decoder maps this latent difference back to protein space as ΔD_{FES}. Donor-level vectors are summarized by pairwise cosine similarity. Figure 4A displays the 12 proteins whose direction agrees across all three donors. Figure S2 discloses the complete 31-protein result, including 12 proteins with 3/3 agreement, 14 with 2/3 agreement, and 5 with 1/3 agreement.

Gene-level attribution uses reaction occlusion. Each reaction's signed contribution is mapped through ENGRO2 gene–protein–reaction rules and divided equally when a reaction maps to multiple genes. The interpretation universe is restricted to the 497 ENGRO2 metabolic genes visible to the model. Donor-specific signed rankings are analyzed by preranked GSEA (Subramanian et al., 2005) with 1,000 permutations, seed 7, and gene sets of 3 to 500 in-universe genes. Multiple-testing

Table A5: Capacity-matched controls for RNA-derived metabolic context. Values report domain-equal global RMSE mean ± SD. M1 and M3 reproduce the canonical values reported in Table 2. The shuffled-FES row reports the matched shuffle rerun; its paired real-FES rerun means were 0.359, 0.334, and 0.480 for ALD, ALS, and Brain, respectively. Bold and underlined values indicate the best and second-best canonical result, respectively, for each dataset.

Third view	Trainable parameters	ALD RMSE ↓	ALS RMSE ↓	Brain RMSE ↓
None (M1)	108,320	<u>0.362 ± 0.048</u>	0.338 ± 0.051	0.481 ± 0.220
Additional RNA PC32	131,248	0.365 ± 0.044	<u>0.337 ± 0.052</u>	0.484 ± 0.216
ENGRO2-gene RNA PC32	131,248	0.366 ± 0.041	0.338 ± 0.053	<u>0.480 ± 0.219</u>
FES PC32 (M3)	131,248	**0.361 ± 0.046**	**0.332 ± 0.049**	**0.477 ± 0.215**
Shuffled FES PC32	131,248	0.363 ± 0.043	0.338 ± 0.054	0.480 ± 0.217

adjustment uses the Benjamini–Hochberg false-discovery-rate procedure (Benjamini & Hochberg, 1995). The four external collections are Hallmark (Liberzon et al., 2015), Reactome 2022 (Gillespie et al., 2022), GO Biological Process 2023 (Gene Ontology Consortium, 2021), and KEGG 2021 Human (Kanehisa et al., 2021), accessed through Enrichr (Kuleshov et al., 2016).

Functional themes are formed by average-linkage clustering of gene-set membership using Jaccard distance with a merge threshold of 0.9, then labeled from the member sets. This construction is based on gene membership rather than semantic similarity of pathway names. The enrichment analysis is exploratory and is used to summarize recurrent attribution patterns rather than to infer causal metabolic regulation.

D Additional design and sensitivity analyses

Appendix D reports two additional analyses under the same held-out-domain protocol used in the main evaluation. First, we compare FES with alternative 32-dimensional RNA-derived third views. Second, we evaluate sensitivity to the discordance rank. Unless stated otherwise, technical seeds are averaged within each held-out domain before biological domains receive equal weight.

D.1 Matched controls for the FES branch

FES is derived from target RNA through a separate ENGRO2–spFBA pipeline rather than measured as an independent modality, and M3 contains an additional encoder that is absent from M1. Its improvement could therefore arise from the extra trainable parameters or from supplying any second RNA representation, rather than from the metabolic-network transformation itself. We addressed these alternatives with two capacity-matched controls. Additional RNA PC32 supplies an extra 32-component transcriptome-wide RNA projection that is not used by the 256-component spatial-RNA branch, testing whether another RNA summary is sufficient. ENGRO2-gene RNA PC32 instead applies PCA to the expression of genes represented in ENGRO2, but does not apply the gene–protein–reaction mapping or reaction-network transformation; this tests whether selecting metabolic genes alone explains the result. FES PC32 uses the 32-component reaction-level representation used by M3. The three 32-dimensional third-view models share the same encoder architecture and 131,248 trainable parameters, whereas M1 contains 108,320. All configurations use the same outer-domain splits and optimization protocol. We report global RMSE because the main component analysis identifies absolute-error refinement as the primary contribution of FES.

Canonical FES produced the lowest RMSE in all three datasets (Table A5). The matched RNA controls use the same 32-dimensional input and encoder size, so the result is not explained by added capacity alone. In the separately matched shuffle rerun, shuffled FES increased RMSE relative to its real-FES counterpart in ALD (0.363 versus 0.359) and ALS (0.338 versus 0.334), while Brain was equal at the reported precision (0.480). Real FES was better in seven of nine held-out domains, supporting a modest spot-specific contribution that is strongest in ALD and ALS.

D.2 END-TO-END DISCORDANCE-RANK SENSITIVITY

We retrained the complete model with $K \in \{4, 8, 12, 16, 24, 31\}$; $K = 31$ is the largest possible rank for the shared protein panel. The sensitivity sweep used one fixed optimization seed across all nine held-out domains, whereas the canonical $K = 16$ results in Tables 1 and 2 average three seeds. Table A6 therefore reports the within-sweep change in Spearman correlation relative to $K = 16$ rather than repeating a second absolute $K = 16$ estimate.

Table A6: End-to-end sensitivity to discordance rank K. Values are changes in mean protein-wise Spearman correlation relative to $K = 16$ under a fixed optimization seed, reported as $\Delta\rho \times 10^3$; zero denotes the $K = 16$ reference within each dataset.

Rank K	$\Delta\rho \times 10^3$ ALD	$\Delta\rho \times 10^3$ ALS	$\Delta\rho \times 10^3$ Brain
4	-46	-45	-16
8	-3	-28	-19
12	0	-19	-10
16	0	0	0
24	-2	-1	-2
31	-11	-2	$+2$

Very small ranks reduced performance, particularly in ALD and ALS. Results were comparatively stable from $K = 12$ to $K = 31$: $K = 16$ matched the best ALD value at the reported precision, gave the strongest ALS result, and differed from the best Brain result by $\Delta\rho \times 10^3 = 2.0$. We therefore retain $K = 16$ as an intermediate rank that balances representation size and predictive performance rather than as a universally optimal value.

E EXTERNAL BASELINE IMPLEMENTATION

Common evaluation protocol. STProtein, DGAT, and totalVI were evaluated using the same nine outer-domain splits as ReDiM: two held-out ALD sections (Sah73 and Sah80), three held-out ALS donors (sl5, sl11, and sl13), and four held-out Brain regions (fcx, hipp, pcx, and tcx). All methods were evaluated on the same processed datasets and common 31-protein panel. Mean protein-wise Spearman correlation was computed across held-out spots for each protein and then averaged over the 31 proteins. Global RMSE was computed over all spot–protein entries in each held-out domain.

E.1 STPROTEIN

We used the official STProtein configuration with a SAGEConv encoder–decoder architecture. RNA preprocessing retained 4,000 highly variable genes and 30 RNA principal components, and spatial graphs used $k = 3$ nearest neighbors. Models were trained for 12,000 epochs using a learning rate of 10^{-4}, weight decay of 10^{-3}, RNA reconstruction weight 5.0, and protein-prediction weight 3.0. The final epoch was used according to the official training schedule, without outer-test-based checkpoint selection. One run was evaluated for each held-out domain.

E.2 DGAT

We used the official DGAT implementation without modifying the model architecture; only the data-loading and evaluation harness was adapted to the present outer-domain splits. The GAT encoder used hidden dimensions 2048, 1024, and 1024 with 16-head feature attention and dropout 0.3. Models were trained for up to 100 epochs with a 50-epoch warm-up and the official evidence-based stopping and learning-rate decay rules. Adam optimization used encoder learning rate 5×10^{-4} with weight decay 2×10^{-5}, decoder learning rate 10^{-4} with weight decay 10^{-5}, and gradient clipping at 1.0.

E.3 TOTALVI

We used `scvi.model.TOTALVI` from scvi-tools v1.3.3 with the package-default architecture. The base model was trained for 400 epochs using raw RNA counts and raw pro-

tein counts from the outer-training domains. The held-out domain was then registered using `TOTALVI.load_query_data` and adapted for 150 epochs using RNA only; held-out protein measurements were replaced by a zero placeholder and were not used during adaptation. Predictions were obtained with `get_normalized_expression` using 25 posterior samples, background-inclusive protein estimates, and equal weighting over training batches. We used seeds 0, 1, and 2; seed-level predictions were averaged within each held-out domain before domain-equal dataset summaries were computed.

F SUPPLEMENTARY FIGURES

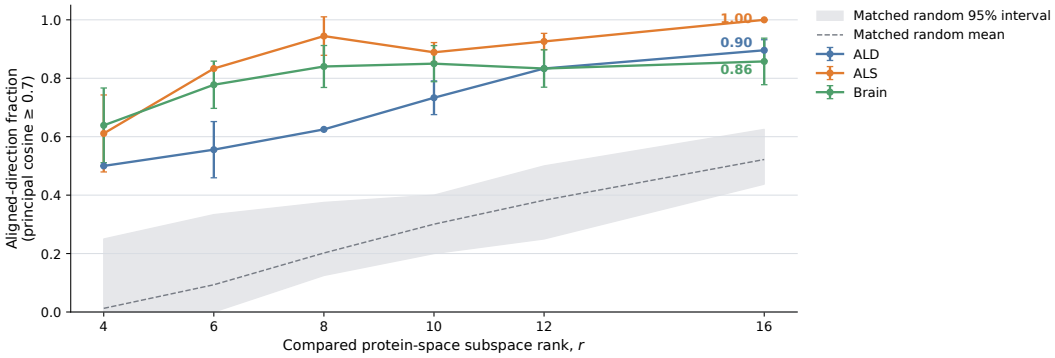


Figure S1: Cross-domain correction alignment across analysis ranks. For each domain pair and optimization seed, the leading r directions of the decoded predicted corrections were compared in the common 31-protein space, and the fraction of principal cosines greater than or equal to 0.7 was calculated. Colored points and lines show dataset means, and error bars indicate ± 1 SD across domain-pair–seed comparisons. The dashed line and shaded region show the matched-random mean and central 95% simulation interval. The observed aligned-direction fraction remains above the matched-random expectation from $r = 4$ to $r = 16$.

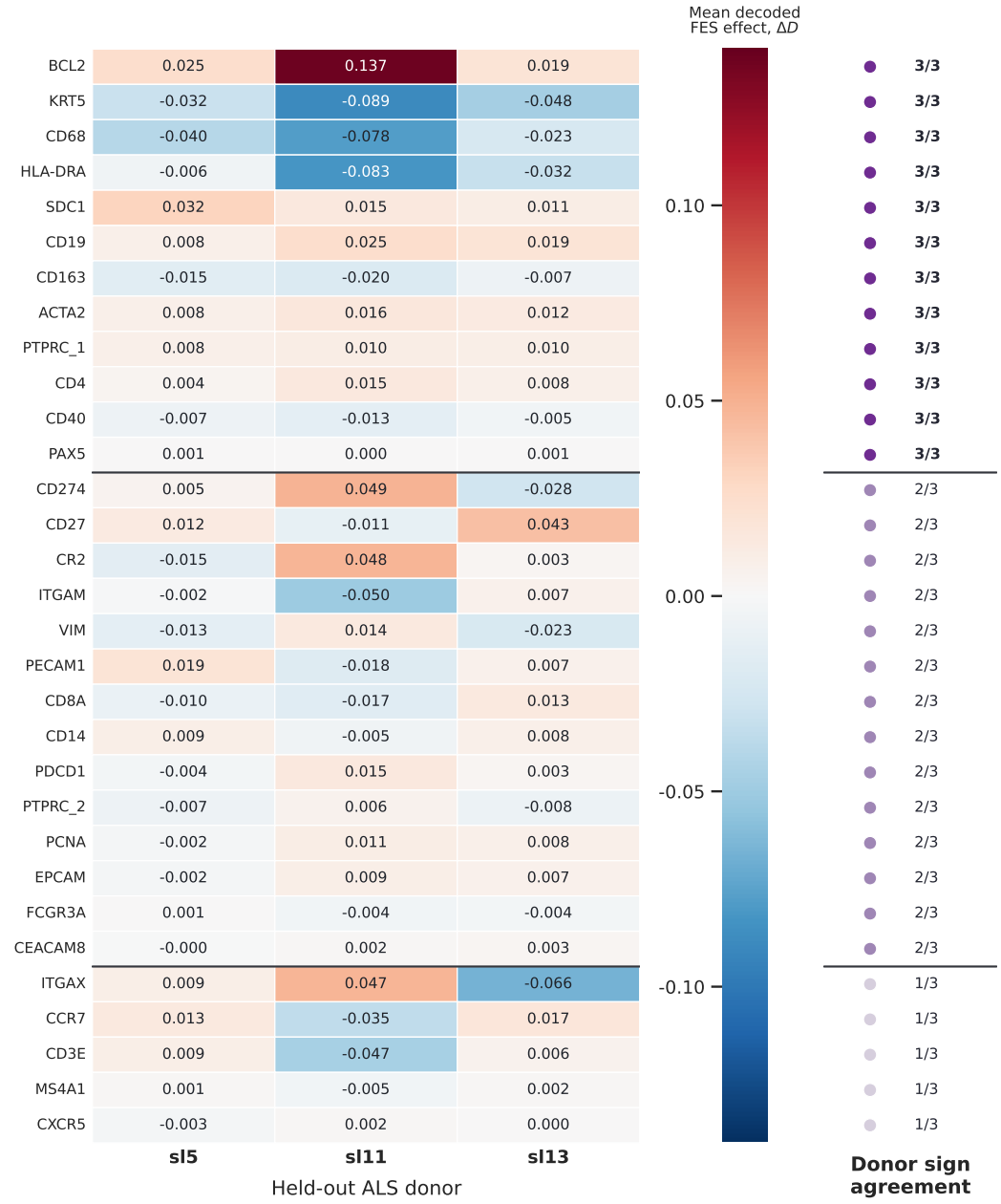


Figure S2: Complete ALS donor-level FES effects. All 31 measured proteins are shown for sl5, sl11, and sl13. Color represents the decoded FES correction ΔD for each donor. The right panel summarizes agreement in effect direction across donors: 12 proteins agree in 3/3 donors, 14 in 2/3 donors, and 5 in 1/3 donors. Figure 4A summarizes the 3/3 subset, while this figure reports the full panel.

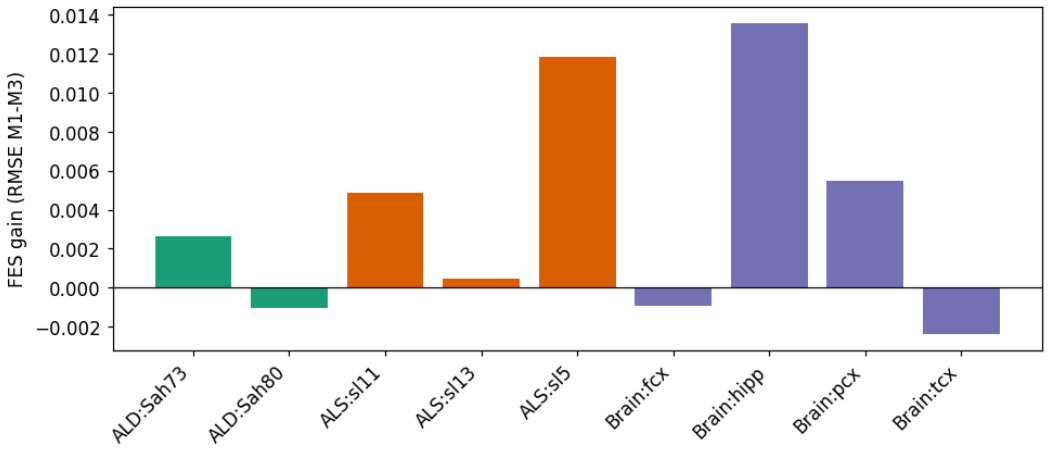


Figure S3: Domain-level FES gain across the nine held-out biological domains. Bars show $\mathrm{RMSE(M1)} - \mathrm{RMSE(M3)}$, so positive values indicate lower prediction error after adding the FES branch. Colors distinguish ALD, ALS, and Brain domains. The gain is positive in six of nine held-out domains, with the largest reductions in ALS sl5 and Brain hipp.

G LIMITATIONS

The experiments support a predictable reference-relative correction, with distinct effects on rank-based and absolute-error metrics. They do not establish that discordance is exclusively biological or that altered RNA–protein coupling is its sole cause; cell composition, spatial mixture, technical effects, and anchor misspecification remain possible contributors. FES is derived from RNA, and although the capacity-matched controls distinguish its contribution from added encoder capacity and alternative RNA summaries, they do not establish independently measured metabolic activity or causal regulation. Future integration with spatial metabolomics, direct flux measurements, or perturbation experiments could more directly test the biological interpretation of the inferred metabolic corrections. Only two to four target domains are evaluated per dataset, and regions or domain pairs should not be interpreted as independent donors. Broader evaluation across additional cohorts, donors, tissue regions, and disease states will be needed to establish the generality of the observed transfer patterns. The region-matched Brain reference and post hoc attribution analysis also limit stronger mechanistic conclusions. Similarly, the 16-component reconstruction diagnostic and random-subspace comparisons support the usefulness of the chosen representation but do not establish an optimal latent rank or independence from model-induced geometry. Future work could evaluate these properties under larger multi-cohort settings and prospectively defined representation analyses.