NetREm Network Regression Embeddings reveal cell-type transcription factor coordination for gene regulation

Saniya Khullar[1,2], Xiang Huang[1], Raghu Ramesh[1,4], John Svaren[1,5], Daifeng Wang[1,2,3,*]

[1]Waisman Center, University of Wisconsin-Madison, Madison, WI, 53705, USA
[2]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, 53076, USA
[3]Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI, 53706, USA
[4]Comparative Biomedical Sciences Training Program, University of Wisconsin-Madison, Madison, WI, 53706, USA
[5]Department of Comparative Biosciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI, 53706, USA
[*]Corresponding author: daifeng.wang@wisc.edu

Abstract

Transcription factor (TF) coordination plays a key role in target gene (TG) regulation via protein-protein interactions (PPIs) and DNA co-binding to regulatory elements. Single-cell technologies facilitate gene expression measurement for individual cells and cell-type identification, yet the connection between TF coordination and TG regulation of various cell types remains unclear. To address this, we have developed a novel computational approach, Network Regression Embeddings (NetREm), to reveal cell-type TF-TF coordination activities for TG regulation. NetREm leverages network-constrained regularization using prior knowledge of direct and/or indirect PPIs among TFs to analyze single-cell gene expression data. We test NetREm by simulation data and benchmark its performance in 4 real-world applications that have gold standard TF-TG networks available: mouse (mESCs) and simulated human (hESCs) embryonic stem (ESCs), human hematopoietic stem (HSCs), and mouse dendritic (mDCs) cells. Further, we showcase NetREm's ability to prioritize valid novel TF-TF coordination links in human Peripheral Blood Mononuclear cell (PBMC) sub-types. We apply NetREm to analyze various cell types in both central (CNS) and peripheral (PNS) nerve system (NS) (e.g. neuronal, glial, Schwann cells (SCs)) as well as in Alzheimer's disease (AD). Our findings uncover cell-type coordinating TFs and identify new TF-TG candidate links. We validate our top predictions using Cut&Run and knockout loss-of-function expression data in rat/mouse models and compare results with additional functional genomic data, including expression quantitative trait loci (eQTL) and Genome-Wide Association Studies (GWAS) to link genetic variants (single nucleotide polymorphisms (SNPs)) to TF coordination. NetREm is open-source available on GitHub as a software package.

Introduction

Transcription factors (TFs) collaborate to activate or repress the expression of their target genes (TGs) by binding to specific DNA sequences known as TF binding sites (TFBSs) on regulatory elements (REs; enhancers, promoters). These REs are shaped by the combinatorial interactions of multiple TFs, forming transcriptional regulatory modules (TRMs) that govern transcription initiation (Guo and Gifford 2017; Wunderlich and Mirny 2009). TFs often work in concert, either as stable complexes or by enhancing the binding affinity of nearby TFs (Nie et al. 2020; Zhao 2023). Despite the common occurrence of TF-TF coordination, the complexities underlying this phenomenon are not yet fully understood (Ibarra et. al 2020). The coordinated activity of core TFs, among 100s expressed, defines a cell type's unique transcriptional landscape and identity (Lee et al. 2012; D'Alessio et al. 2015). Indeed, the TF binding grammar is complex and context-specific, influenced by proteins occupying shared REs and the chromatin structure (Bentsen et. al 2022).

Gene regulatory networks (GRNs: models of TG regulation by TFs) must incorporate the complex combinatorial coordination (nature: cooperativity, antagonism; type: direct, indirect) among TFs, which is facilitated by the positioning of TFBSs within REs (Berenson et al. 2023; Ouyang et al. 2009). Historically, classical models of TF cooperativity were predicated on direct protein-protein interactions (PPIs) since many TFs may form complexes (to increase their binding affinity and motif specificity (Sönmezer et al. 2021)) to co-regulate TGs (Wang et al. 2023). Based on recent studies, it is now understood that TFs may jointly coordinate TG transcription even without direct PPIs (Rao et al. 2021; Wang et al. 2009). Pioneer TFs, for instance, recognize and bind to their TFBSs even within closed chromatin, inducing local chromatin remodeling (e.g., nucleosome eviction), thereby exposing TFBSs and rendering DNA more amenable to subsequent binding by settler TFs (Sinha et al. 2023; Mayran and Drouin 2018). Tethered-binding mechanisms, involving coactivators/corepressors, allow TFs to regulate TGs cooperatively by recruiting intermediary proteins (e.g., p300/CBP, mediator complexes) (Spitz and Furlong 2012; Ibarra et. al 2020). Antagonistic coordination, via sequestration and/or competition for TFBSs, generates cell-state heterogeneity by driving opposing effects on epigenetic programs and TG expression; some TFs bind DNA only in the presence of cooperating TFs and absence of antagonistic TFs (Hu et al. 2022; Berenson et al. 2023).

Despite extensive single-cell data analyses (Mathys et. al 2019; Wang et. al 2018), the extent of TF-TF coordination in regulating TGs across various cell types remains unclear (Nie et al. 2020; Ibarra et. al 2020; Karczewski et al. 2011; Hannenhalli and Levy 2002). Understanding this is crucial for dissecting GRNs driving

cognitive, motor, and behavioral processes, often dysregulated in disease states. GRN-inference tools often reverse-engineer expression data to find coordinated patterns and gene-gene interactions, uncovering meaningful biological signals from noise (Wang et al. 2023; Campos et al. 2019). State-of-the-art (SOTA) tools (e.g. BEELINE(Pratapa et. al 2020), SCENIC(Aibar et. al 2017), GRNBoost(Moerman et al. 2019), Signac(Stuart et. al 2021), scGRNom(Jin et al. 2021)) create cell-type GRNs using various methods (e.g. co-expression, correlation r, info theory, machine learning, multi-omics integration) but often overlook interdependent TF-TF PPI networks (PPINs) crucial for TG regulation. Neglecting PPINs may lead these tools to disregard TF_i-TG pairs with uncorrelated expression (Zaborowski and Walther 2020), which may bury true biological relations, as TF_i could still regulate the TG through complex avenues, often coordinating synergistically or antagonistically with other TFs. Coordinating TFs are often co-expressed (Ouyang et al. 2009) and amplify noise in expression profiles (Nagamine et al. 2005; Perna et al. 2020; Parab et al. 2022). Without additional context, GRN-inference tools might remove truly causal correlated TFs that co-regulate TGs, excluding fundamental aspects of GRN mechanisms (Nicodemus and Malley 2009). Thus, these tools ought to incorporate various levels of expression regulation like PPIs for TFs as they can help reveal these dynamics of TF-TF coordination (Nagamine et al. 2005; Perna et al. 2020; Parab et al. 2022). Tools like BGRMI (Iglesias-Martinez et al. 2016) identify relevant TF complexes using PPIs but ignore inherent PPIN structures and indirect PPIs. RTNduals (Chagas et al. 2019) predicts TF co-regulatory behavior solely from expression data, which may not always yield outputs. SCINET (Mohammadi et. al 2019) reconstructs cell-type interactomes, while TF-Cluster (Nie et al. 2011) identifies coordinated TFs but avoids using prior info. Neither tool focuses on TG regulation.

Previous studies leveraged network-regularized (e.g. graph-regularized, network Lasso) regression models to identify disease-associated genes and networks, incorporating biological info as metadata (Li and Li 2008; Wang et al. 2015; Dirmeier et al. 2018; Li and Li 2010; Kim et al. 2013). These models, guided and constrained by prior knowledge, enhance their biological relevance; however, they are not tailored to elucidate TF-TF coordination in TG regulation. There is a growing need for these models to advance, assimilating info from regression data and prior network inputs to generate robust embeddings that may extract significant relationships, offering new insights in biology and beyond (Chu et al. 2023; Gharavi et al. 2021; Choy et al. 2019).

To bridge these gaps, we developed NetREm (Network Regression Embeddings), a novel computational framework to infer cell-type TF-TF coordination activities for TG regulation. Building on established network-regularized regression techniques, NetREm integrates multimodal data (e.g. TF binding profiles, direct/indirect TF-TF

PPIs, derived TF-TF colocalization, gene expression, chromatin interaction, scATAC-seq epigenomic markers) capturing intricate aspects of TG regulation. It constructs robust predictive models for TF-TG regulation (complementary GRNs) as well as TF-TF coordination. A distinct feature of NetREm is its innovative ability to generate network regression embeddings, which identify and quantify coordination among cell-type TFs for co-regulating individual TGs. Public databases like STRINGdb provide direct and/or indirect, organism-specific, global, undirected, cell-type-agnostic PPIs for >12k organisms (Szklarczyk et al. 2023). Despite PPINs having some incorrect PPIs (False Positives (FPs)) and being largely incomplete (Kotlyar et. al 2022), it is helpful to integrate TF-TF PPINs as prior info (McCalla et al. 2023; Li and Jackson 2015; Ghanbari et al. 2015; Imoto et al. 2003; Mukherjee and Speed 2008). This improves NetREm's predicted GRNs from expression data. NetREm also addresses the need for not only cell-type-specific annotation of known PPIs, but also discovery of new PPIs. It provides much-needed insights into TRMs, reveals cell-type- and disease- specific TF-TF PPINs, and aids interactome studies in uncovering disease gene properties and differential PPIN rewiring (Göös et al. 2022; Sevimoglu and Arga 2014). To demonstrate, we apply NetREm to simulation data and NS cell types: myelinating (mSCs) vs. non-myelinating (nmSCs) SCs, as well as in Alzheimer's disease (AD) vs. control states in neurons/glia. NetREm is an open-source tool for general-purpose use.

Results

Overview of Network Regression Embeddings (NetREm)

We developed NetREm as a robust, multi-omics, computational approach to build and integrate networks of TF-to-TG regulation and TF-TF coordination in a cell-type-specific manner (**Fig. 1**). NetREm is for applications where predictors form subnetworks to influence the outcome (Li and Liu 2022). NetREm uses single-cell or bulk gene expression data, a comprehensive PPIN (e.g. STRINGdb) that globally captures verified direct/indirect functional associations among TFs (Peng et al. 2017). Providing a prior, candidate GRN (e.g. initial TF-RE-TG links) is optional but recommended as it improves biological relevance of outputs with its initial feature selection. NetREm integrates this multi-modal data, applying a 2-step optimization process for each TG: **1.** It formulates a network-regularized regression problem, using the input PPIN as a priori info, to sift through N candidate TFs, to find those most likely to co-regulate TG. **2.** To solve this regression problem, it then employs Singular Value Decomposition (SVD) to create network regression embeddings, which are used in Lasso regression to predict TG expression. In the process, NetREm learns not only key TFs (coefficients c^*, a complementary GRN) but also a TG-specific TF-TF coordination network -

$100 \leq B \leq 100$, predicting cooperative (> 0) or antagonistic (< 0) TF-TF relations for regulating TG. It also annotates known PPIs, uncovering confident cell-type-specific TF-TF PPI subnetworks involved in TG regulation (Szklarczyk et al. 2023). By integrating genomic and foundational PPIs, NetREm uses transfer learning to discover unanticipated, biologically significant TF-TG links and novel functional TF-TF relations for future investigation (Zhang et al. 2013). Aggregating outputs across all TGs, yields cell-type-specific directed TF-TG regulatory and undirected TF-TF coordination ($-100 \leq \bar{B} \leq 100$) networks. These outputs support downstream analyses, like linking non-coding SNPs to potential regulatory roles via expression trait loci (eQTL) SNP-TG (i.e. eSNP-eTG) associations in our networks.

We demonstrate NetREm's versatility across diverse real-world scenarios by benchmarking our TF-TG networks against established gold standard GRNs in human (hESCs, HSCs) and mouse (mESCs, mDCs) models. We use diverse techniques to assess $\bar{B}$ across these settings, as well as additional human contexts: 9 PBMCs, pooled stem cells from GTEx (Eraslan et. al 2022), 4 CNS cell types (Lake et al. 2018). Further, we highlight NetREm's use of prior input GRNs, derived from multi-omics data, for specific contexts (cell types, diseases) in 2 human applications: 1. mSCs vs. nmSCs; 2. AD vs. controls in 8 neuron/glia cell types. The resulting TF-TG networks, integrated with TF-RE-TG annotations (**Methods**), yield enhanced, context-specific TF-RE-TG regulatory and coordination $\bar{B}$ networks.

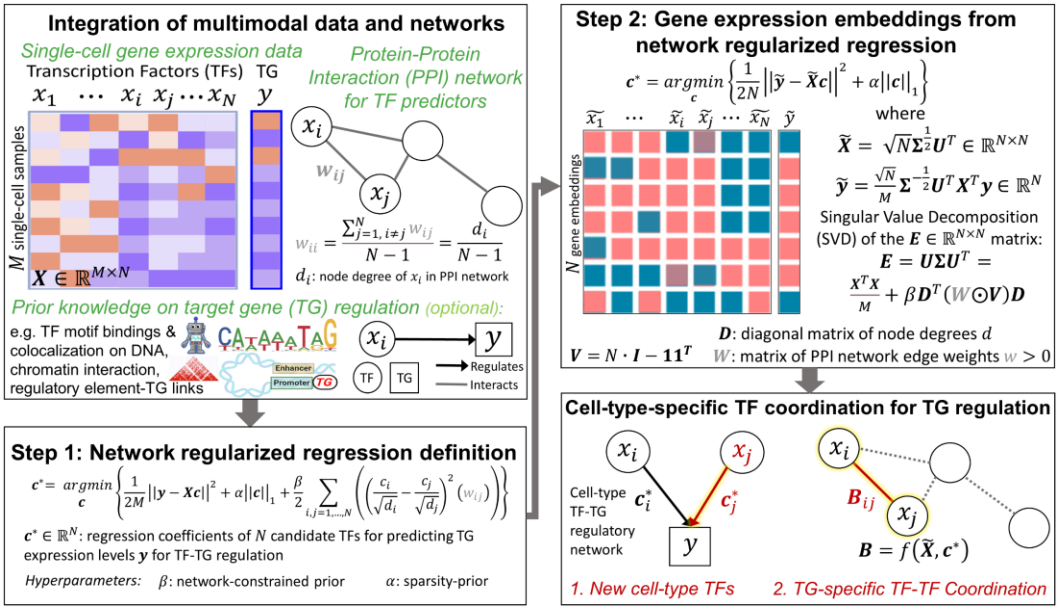


Figure 1 Overview of Network Regression Embeddings (NetREm) a multi-step, multi-omics computational framework to construct comprehensive cell-type-specific TF-TG regulatory and TF-TF coordination $\bar{B}$ networks. We apply this pipeline for each TG in a cell-type. NetREm integrates multimodal data: gene expression ($X \in \mathbb{R}^{M \times N}$, $y \in \mathbb{R}^M$) for M cells (samples), direct/indirect TF-TF PPIs with weights W, optional prior GRN information (e.g. TF-RE-TG links). This prior GRN helps select only relevant TF predictors for TG (from multiomics data) to reduce dimensionality of X: M (rows), N TFs (columns). X and y are standardized (mean μ: 0, standard deviation σ:1) across cells for TFs and TG. Goal: identify TFs, out of N candidates, whose expressions best predict TG expression y. **Step 1** sets up a PPIN-regularized regression problem to identify optimal TFs for a TG, guided by network β, sparsity α hyperparameters. **Step 2** solves this,

transforming (X, y) and PPIs to latent space gene expression embeddings ($\tilde{X} \in \mathbb{R}^{N \times N}$, $\tilde{y} \in \mathbb{R}^{N}$) by SVD on an $E = \frac{X^T X}{M} + \beta A$ matrix ($A = D^T (W \odot V) D$). E combines expression relations and PPIN information. $(\tilde{X}, \tilde{y})$ undergo Lasso regression (via model-type: Lasso (α is given) or LassoCV (α chosen by cross-validation (CV); default: no intercept) to predict optimal coefficients $c^* \in \mathbb{R}^N$ for TG. NetREm outputs 2 networks capturing different aspects of TG regulation that can be integrated. #1: links optimal TFs to TG by c^* (> 0: activator, < 0: repressor), in a directed TF-TG regulatory network, a complementary GRN. This likely uncovers novel cell-type TFs like x_j and reflects underlying biology of coordination among TFs. If we input a prior GRN, we may use it to help annotate our TF-TG links with biological metadata (e.g. TF-RE-TG network). #2: TG-specific undirected TF-TF coordination network B predicts indirect/direct relations among TFs to regulate TG. $TF_i - TF_j$ coordination B_{ij} shows cooperative (> 0 if both are co-activators or co-repressors) or antagonistic: (< 0 if 1 is repressor, other is activator) co-regulation of TG. B is a matrix of c^*-aware cosine similarity scores (**Methods**), a function of $\tilde{X}$ and c^*. NetREm thus identifies novel coordination among cell-type TFs for co-regulating TG. Results are stitched together across runs for all TGs to obtain final cell-type outputs.

Simulation study

We test NetREm on simulated single-cell gene expression data for 10,000 cells as a proof-of-concept, where: X (matrix of expression levels for N = 5 candidate TF predictors) and y (TG expression vector) are drawn from a normal distribution with dropouts to achieve ~40% sparsity, mimicking single-cell data (**Fig. S1A**). Train data (70%; M: 7,000 cells) standardizes ($X_{train}, y_{train}, X_{test}, y_{test}$) so each variable has ($\mu$ = 0, σ = 1). TFs 1 to 5 have expression with Pearson $r(TF, TG) \approx$ [0.9, 0.5, 0.4, -0.3, -0.8] with y, respectively, in train and test data (**Fig. 2A, Fig. S1B-C**). Our PPIN of known direct and/or indirect PPIs among TFs has strong w for $TF_1 - TF_2$ (0.8), $TF_4 - TF_5$ (0.95). We set w to $\eta = 0.01$ for missing PPIs, making our PPIN fully-connected. NetREm(β = 1, α = 0.1) predicts y based on X, subject to the PPIN constraint. It outputs 2 networks: TF-TG regulation (c^*) relates expression levels of TFs to TG expression (**Fig. 2B**) and TG-specific TF-TF coordination (B scores) (**Fig. 2C**) of behavior among TFs to regulate TG.

We compare NetREm to 4 default Scikit-Learn (Pedregosa et al. 2011) benchmark regression models (BRMs) fit with no intercept: Linear Regression and 3 regularization ones (ElasticNetCV, LassoCV, RidgeCV). TFs with lower ranks have higher $|c^*|$ and are more important. Absolute values of c^* for TF_2 and TF_4 are significantly higher in NetREm (**Fig. S1D, Table S1:** p < 2e-16), highlighting NetREm's grouped variable selection property that prioritizes them due to their strong corresponding PPIs with TF_1 and TF_5 that both strongly correlate with y (Li and Li 2008) (**Fig. S1E**). Unlike BRMs that favor TF_3 over TF_4 (since TF_3 has greater $|r|$ with y), NetREm prioritizes TF_4 over TF_3 since TF_4 strongly interacts with TF_5, while TF_3 has weak PPIs with other TFs. Sensitivity analysis (**Fig. 2D-E**) for fixed α and β confirms $|c^*|$ for TF_2 and TF_4 increase as their respective PPIs with TF_1 and TF_5 strengthen. NetREm's test MSE increases from 0.15 to 0.22 as β increases from 0.01 to 1 (**Fig. S1F**). Test MSEs for BRMs ≈0.14, which we can also achieve with NetREm(β = 0.01, LassoCV α); nonetheless, studies emphasize obtaining

more interpretable and context-driven features by incorporating network information into models, even if accuracy is sacrificed (Li and Liu 2022). **Fig. S1G** illustrates changes in B as β increases from 0.01 to 2.

One of NetREm's salient features is its ability to explicitly model and integrate PPIN structures/relations among TFs. By doing so, it identifies key TFs for predicting y, which may not exhibit strong r with y individually. NetREm also generates a coordination network among TFs that co-regulate TG, a capability beyond the scope of BRMs. Instead, BRMs, primarily focus on prediction accuracy by selecting c^* to best represent each TF's impact on y, typically treating TFs as "lone wolves", neglecting TF-TF interdependencies vital to GRNs. Our B network shows TF_1 and TF_2 cooperate with TF_3 weakly to increase y, and TF_4 and TF_5 collaborate to decrease it; TF_1-TF_2 and TF_4-TF_5 links are highly-confident. NetREm predicts novel direct/indirect links like TF_1-TF_5 activator-repressor antagonism: B = -97.9. TF_3 has weaker relations with TF_1 than TF_2 does (41.7 vs. 71.5) and the smallest $|c^*|$ = 3.4e-2.

We detail more simulations in **Fig. S2-5** and **Tables S2-3** in **Supp. Methods**. We show that results are stable and consistent for various expression sparsity levels and for $M \ll N$ cases as well, excess β over-constrains NetREm causing its predictions to suffer (bias-variance trade-offs), NetREm has more robust c^* estimates (less variable) and more accurate TF assignments, and it can capture complex TF-TF PPIs. Our toy data with a few TFs and 1 TG intuitively explains NetREm and its advantages.

Next, we benchmark NetREm in real-world settings: many TFs and TGs. We apply it step-by-step for each TG in a cell-type to learn B, TF-TG regulation, TF-TF coordination $\bar{B}$.

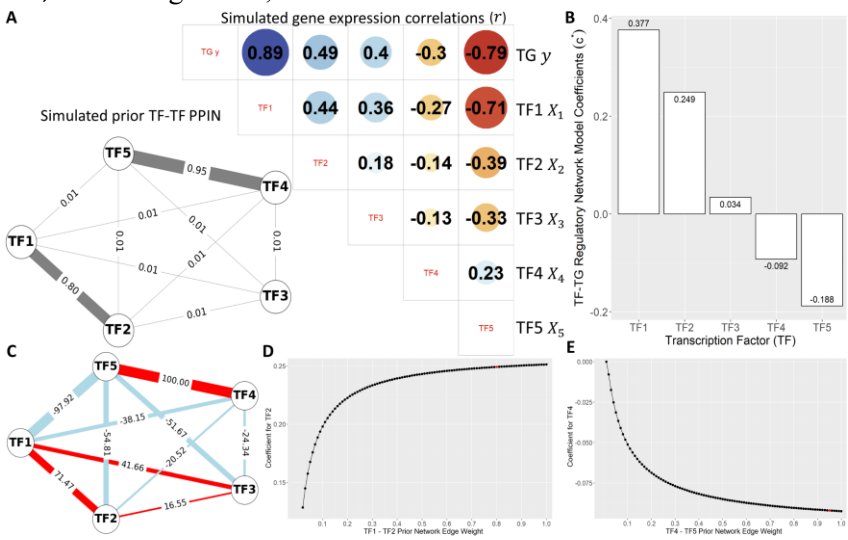


Figure 2 Simulation study of 5 TFs, 1 TG **A)** Bottom left: prior, undirected TF-TF PPIN with default weights (η = 0.01); strong experimentally-verified connections: TF_1- TF_2 (0.8); TF_4- TF_5 (0.95). Top right: Pearson r matrix in training data: $r(TF, TG) \approx$ [0.9, 0.5, 0.4, -0.3, -0.8]. Dot sizes: $|r|$: + (blue) or - (red) r. **B)** Coefficients c^* for TFs in TF-TG regulatory network, based on NetREm(β =1, α = 0.1, no intercept). Potential activators: TF_1 to TF_3. Repressors: TF_4 and TF_5. **C)** TG-specific TF-TF coordination network (scores: B) with red (antagonistic: -; activator-repressor links) and blue (cooperative:

+; links between co-activators or co-repressors). Functionally-valid coordination: TF_1- TF_2 and TF_4- TF_5; others are novel. **D)** Effects of varying TF_1-TF_2 w in original PPIN, from 0.01 to 1 in 0.01 increments, holding all else fixed. Respective TF_2 c^* increases monotonically in arc shape from 0.106 to 0.251. Red dot: PPIN w. **E)** Similar sensitivity analysis shows TF_4's $|c^*|$ increases, becoming more negative from 0 to -9.2e-2, as TF_4-TF_5 edge weight is perturbed.

Benchmarking NetREm with No Prior GRN Info

Here, we assess how effectively NetREm predicts cell-type TF-TG and $\bar{B}$ networks in real-world scenarios lacking prior GRN information (**Methods**, **Supp. File 5**). All TGs have the same N candidate TFs; for TGs that are TFs, N is reduced by 1. We anticipate that incorporating prior GRNs derived from multiomics data may enhance our performance by providing a tailored, context-specific list of promising TFs for each TG.

Evaluating TF-TG regulatory links

To evaluate our networks in terms of predicting regulatory TF-TG links as well as accurate TF roles (based on signed c^*: +/-), we compare NetREm with BRMs for predicting signed-TF-TG links. We use SERGIO (Dibaeinia and Sinha 2020) to input a signed hESC GRN and generate 6 realistic datasets (1,250 TGs, N = 207), varying M and noise %. Overall, NetREm has the highest precision, indicating its superior reliability in correctly identifying true TF-TG links, assigning activator/repressor roles, minimizing FPs (**Fig. S6**).

We benchmark our inferred TF-TG links using single-cell expression data for human HSCs and mice (mESCs, mDCs) (**Figs. S7-8**, **Tables S4-6**) against gold standard GRNs (McCalla et al. 2023; Zhang et al. 2023). Due to its use of biological information (Shojaie and Michailidis 2009), NetREm demonstrates higher sensitivity in identifying relevant biomarkers, though it shows lower specificity compared to Lasso and ElasticNet (Li and Li 2008). By adjusting β and α, we can fine-tune NetREm's behavior; increasing α might decrease sensitivity but enhance specificity. Our findings are consistent across updates in input PPINs, comparing well between mouse STRINGdb versions 11 (V11) and 12 in mESCs.

NetREm features a unique grouped variable selection mechanism that leverages PPIN structures to prioritize tightly-linked TF groups involved in known PPIs. This optimizes the selection of cell-type TFs that co-regulate TGs, enhancing NetREm's ability to identify complex TF-TF relations. We highlight this in HSCs. ElasticNet and Lasso fail to identify critical TF-TG links, for instance, missing regulation of *ATF2*, a key HSC TF (Ju et al. 2023). Conversely, Linear Regression and Ridge predict that *ATF2* is regulated by all N = 177 TFs, showing their alarming potential for FPs. However, NetREm(β = 10) identifies 8 TFs for *ATF2*, with 7 confirmed by gold standards. This pattern extends to other TGs like *BRD2, RNF167, DUSP2*; NetREm identifies groups of both validated and novel TFs involved in

biologically relevant PPIs for these TGs and more. For instance, 1 of 10 novel TFs for *RNF167* is TFAP4, which forms PPIs for adipogenesis and cell population proliferation regulation alongside 3 of 17 verified final TFs. NetREm effectively uses prior TF-TF PPIN knowledge to identify genuine TF-TF coordination for TG regulation, navigating through data noise and avoiding false correlations (r). It integrates r and PPIs among TFs, to achieve superior generalizability and consistency in identifying intricate TF-TF relationships that BRMs may struggle with.

Our TF-TG networks complement SOTA cell-type GRN tools. NetREm has comparable F1 Score, sensitivity, specificity, and accuracy with 10 tools benchmarked in mDCs (McCalla et al. 2023). No method outperforms others. However, NetREm uniquely infers B and cell-type $\bar{B}$, capabilities that other tools lack.

Evaluating TF-TF Coordination

We evaluate performance of our $\bar{B}$ in mESCs, mDCs, and PBMCs using V11 PPIN as input (proxy for outdated information), categorizing TF-TF pairs into 4 groups based on their status in V11 and in updated V12: absent in both, present in both (TPs), removed in V12 (FPs), discovered in V12 (FNs). Top TF-TF links have higher $|\bar{B}|$. Welch 1-sided tests (p-adj < 0.05) compare $|\bar{B}|$ values across groups (**Tables S7-12, Figs. S9-11**). For instance, in mESCs, NetREm often reflects known PPIs, flags FPs to remove, and uncovers biological truths, nominating promising candidate PPIs for follow-ups. Overall, NetREm prioritizes known TPs and has potential to flag future TF-TF PPIs (FNs) that are currently unknown.

Further, we benchmark NetREm against the RTNduals (Chagas et al. 2019) tool. In 12 of 13 human contexts (**Fig. S12**), our top k links using V11 input PPIN outperform RTNduals, with a higher % that are verified PPIs in V12 and other sources. This underscores NetREm's efficacy in identifying TPs and leveraging historical PPINs to predict previously unknown PPIs; this is encouraging as only a small number of ~130-650k estimated human PPIs are currently known (Sevimoglu and Arga 2014; Venkatesan et al. 2009; Yu et al. 2020).

To validate NetREm's capability to prioritize biologically relevant links, we leverage a Contextual PPI Database (CPPID) (Kotlyar et. al 2022) that annotates PPIs with >243 terms but lacks cell-type specificity (**Fig. S13**). By showing that our top links are indeed enriched for context-specific terms (e.g. NS-related in Microglia (Mic) and pooled SCs, immune-related in PBMCs), we suggest a potential extrapolation to cell-type specificity. This highlights NetREm's pioneering potential to discover cell-type TF-TF PPIs, addressing a crucial gap in existing global PPINs.

Gene regulatory links between TFs and TGs in human mSCs & nmSCs

With emerging new single-cell epigenomic data from many human tissues, we can model GRNs in novel contexts. We apply NetREm to analyze SCs, which play a pivotal role in maintaining, regenerating, myelinating, and supporting PNS neurons. SCs are derived from neural crest and exhibit tremendous flexibility not only in myelination but also in several other tissues; they also function as terminally differentiated cells that can reverse differentiation after nerve injury to aid nerve regeneration (Ma and Svaren 2018). Single-cell rodent NS studies reveal substantial diversity in SC differentiation status (Gerber et al. 2021; Yim et. al 2022). 2 main SC phenotypes (or sub-types) are: 1) mSCs associated with larger diameter axons (>1 micron); 2) nmSCs that wrap a bundle of smaller diameter axons (typically sensory axons) i.e. a Remak bundle. While mSCs envelop axons in myelin sheaths to enhance conduction speed, nmSCs support sensory axon function/interactions without forming myelin, contributing to overall nerve integrity.

Uncovering TF-TG regulatory mechanisms modulating cellular processes in mSCs and nmSCs may help us understand and treat debilitating nerve injuries, demyelinating disorders, and hereditary neuropathies. Mutations affecting SC function are the most prevalent cause of demyelinating genetic neuropathy Charcot-Marie-Tooth disease (CMT) (Tao et. al 2019) and some affect major transcriptional TFs of SC differentiation like EGR2 and SOX10 (Srinivasan et al. 2012; Fröb and Wegner 2021). Both TFs are co-expressed in myelinating SCs, colocalizing at several myelin REs (Jones et al. 2007; Poitelon et. al 2016). Mutations in EGR2 may disrupt cooperative SOX10 TFBS binding (LeBlanc et. al 2006) and its subsequent regulation of TGs. SC regulation also involves TEAD1 and other Hippo regulators to govern shared TGs and orchestrate PNS myelination (Srinivasan et al. 2012; Lopez-Anido et al. 2016). Since there are no TFs that are exclusively expressed by SCs, uncovering TF-TF coordination networks crucial for SC lineage maturation and differentiation into mSCs or nmSCs also holds significance (Ma and Svaren 2018). However, distinct transcriptional regulatory networks coordinating SC function and underlying states and cell fates require a variety of TFs beyond EGR2 and SOX10 (Hung et al. 2015).

In response, we apply NetREm to each TG in mSCs and nmSCs using single-cell data for human Dorsal Root Ganglion (DRG). We derive prior GRNs using multiomics data (details: **Methods**). To do this, we annotate open chromatin regions in adult SCs with known RE peak-to-TG links (Zhang et al. 2021) and use motif-based analysis to predict TFs that may associate with these REs. By removing low-expressed TFs in corresponding training gene expression data for mSCs and nmSCs in the human DRG, we create tailored prior mSC and nmSC GRNs. For each TG in an SC-type, we input its N TG-specific selected candidate TFs (from respective prior GRN of initial TF-RE-TG

links) to NetREm. In total, NetREm outputs 183,242 mSC and 277,541 nmSC total TF-TG links (**Supp. File 1**) comprising 221 TFs and 8,950 TGs in mSCs, 228 TFs and 5,207 TGs in nmSCs. Both share: 33,806 TF-TG links, 27,037 sign-TF-TG links, 3,841 TGs, 197 TFs. TF EGR2 is mSC-specific (Balakrishnan et al. 2021). We enhance our networks by overlaying them with prior GRN annotations, resulting in our finalized TF-RE-TG regulatory networks.

We examine NetREm's results for 8 core SC TFs that have validation (genome binding and loss-of-function (LOF)) data from rodent SCs: EGR2, NR2F2, RXRG, SOX10, SREBF1, STAT1, TEAD1, YY1. LOF TGs, whose expression varies upon TF knockdown, show their direct or indirect dependency on the TF (Nie et al. 2020). Valid direct TGs are a subset of LOF TGs with ChIP-seq evidence of nearby TF binding, suggesting direct regulation by the TF (Titsias et al. 2012); this dual confirmation enhances confidence in these direct TGs (Badia-i-Mompel et al. 2023). RXRG lacks LOF data. In **Fig. 3A** we provide counts of NetREm-predicted: direct, all LOF, and novel candidate TGs. We also report on eTGs with strong Tibial Nerve eQTL support (THE GTEX CONSORTIUM 2020) based on mapped SNP-TF-RE-TG predictions: instances where an eSNP strongly alters TF binding to a RE, influencing TG expression (**Methods**). For example, we predict 2,015 YY1 TGs in nmSCs (139 direct, 304 LOF only, 1,572 novel), where LOF and direct TGs are significant (hypergeometric $p < 0.05$); of these, 40 direct (28.8%), 75 LOF only (24.7%), 488 novel YY1 (31%) TGs are eTGs. In both SCs, our final TF-TG predictions for all LOF and direct TGs have higher sensitivity and F1 scores compared to LassoCV and ElasticNetCV and higher accuracy and specificity than GRNBoost2, Linear Regression, RidgeCV (**Fig. S14**).

We explore cases where NetREm accurately identifies a core SC TF as the top predictor (low rank) for its direct TG, even when training expression data shows a low r(TF, TG) (**Fig. 3B**). This is important as studies observe that in eukaryotes (unlike in prokaryotes), correlations r and mutual information among TFs and known TGs are not much higher than those between TFs and non-TGs (Zaborowski and Walther 2020; Escorcia-Rodríguez et al. 2023). For example, SOX10 weakly correlates with *SIPA1L2* (Tao et. al 2019)(candidate modifier TG for CMT type 1A) but predicted to have rank 17. Despite its relatively weak r = -0.14 with *MBP*, a major constituent of myelin sheaths, STAT1 is its top 10 TF. Although *APP* exhibits a higher r with STAT1 compared to *FBN1* (9.7e-2 vs. 10.8e-2), NetREm ranks STAT1 higher as an activator for *FBN1* than for *APP* (13 vs. 20). This aligns with findings (Gu et. al 2022) of STAT1 LOF impacts of -0.61 for *FBN1*, -0.51 for *APP*.

NetREm reveals biologically relevant signals, identifying novel TGs for TFs. RXRG's high regulatory activity in nmSCs (**Table S13**) is consistent with rodent/human studies (Gerber et al. 2021). All TGs across 3 groups (EGR2

mSC, SOX10 mSC and nmSC) are enriched in PNS-related terms (**Fig. 3C**). We find 2,316 SOX10 TGs (159 shared, 428 direct, 1,314 LOFs overall) with 29 direct and 103 LOFs overall in both SCs. **Fig. 3D** shows rat nerve epigenome tracks for 3 novel SOX10 candidate mSC TGs (*GPR157*, *CEP95*, *SCAMP5*) with strong SOX10 binding to SC regulatory regions in open chromatin (Lopez-Anido et. al 2015). In **Fig. 3E**, SNP rs55927047 enhances TEAD1 binding to its TFBS on *FOXN2*'s promoter to help activate *FOXN2* in adult mSCs. Orthologous rat nerve TEAD1 ChIP-seq peaks also overlap with this promoter. This eSNP correlates strongly (is in linkage disequilibrium (LD)) with rs79073127 that links to higher inflammatory polyneuropathy risk in Pan UK Biobank GWAS (Turley et al). *FOXN2* GWAS-eQTL colocalizes for this condition, with 75% probability (Wallace 2020). Both SNPs correlate with increased *FOXN2* expression. We highlight additional TFs in SCs, underscoring their significant roles despite limited validation. Notably, our 297 mSC TGs for MEIS2 (core TF in DRG sensory neurons (Roussel et al. 2022)) are prominent in PPI pathways like: PI3K-Akt signaling (crucial for PNS myelination (Ishii et al. 2021)), actin cytoskeleton organization (essential for PNS regeneration (Wang et al. 2018)) (**Fig. 3F**). These TGs have higher expression in pooled SCs compared to 4 CNS cell types (1-sided t, p-adj < 2e-6), a pattern absent in controls. **Fig. S15** showcases principal hub TFs like EGR1 (485 nmSC eTGs) and RXRA (971 eTGs), regulating the most eQTL-validated eTGs in SCs.

Figure 3 Gene regulatory links between TFs and TGs in human mSCs & nmSCs **A)** Bar plot categorizes TGs for core TFs in mSCs and nmSCs. Top: eTGs with eQTL validation. Bottom: original counts with star (*) for over-enriched TGs based on hypergeometric test (p-adj < 5%). **B)** Points are Direct TGs in mSCs. X-axis: r between core SC TF and direct TG in train data. Y-axis: rank of TF for regulating given TG, where smaller ranks imply greater $|c^*|$, strong relations. For simplicity, we show results where the TF is the top 5 to 20 for its direct TG. **C)** Hierarchically clustered, select gene enrichments for all EGR2 mSC, SOX10 mSC, SOX10 nmSC TGs; values: $-\log_{10}(q)$ **D)** Rat epigenome tracks for 3 novel mSC SOX10-activated TGs with SOX10 as a top predictor TF. Tracks: rat sciatic nerve ChIP-seq peaks, SOX10 PNS peaks, histone modifications associated with enhancers (H3K27ac immunoprecipitation sorted tags). Pink text: SOX10 binding in humans based on prior mSC GRN. **E)** Tibial Nerve eSNP rs55927047 (chromosome (chr) 2: 47,783,515 DNA change: C to G) in human *FOXN2* promoter (overlaps with orthologous TEAD1 rat ChIP-seq binding peaks) may strongly boost TEAD1 binding affinity to *FOXN2* promoter to activate ($c^* > 0$) *FOXN2*, a GWAS-eQTL colocalized biomarker for inflammatory polyneuropathy. *FOXN2* is a novel TEAD1 TG with support: TEAD1 rat ChIP-seq binding within 100 kilobases (kb) of its Transcription Start Site (TSS). **F)** Left: Boxplots compare relative expression percentiles for all MEIS2 mSC TGs in GTEx pooled SCs with those in 4 CNS cell types. Median percentile for MEIS2 mSC TGs in SCs overall is 71.4 versus ≤ 61.3 in controls (non-TGs for MEIS2 in mSCs). Right: enriched PPI paths for MEIS2 mSC TG proteins, which are important in SCs.

Coordination among TFs for gene regulation in human mSCs & nmSCs

Our test MSEs are significantly lower than those of Linear and Ridge BRMs (**Fig. 4A**, **Table S14**). NetREm predicts SC-type-specific coordination B for each TG. It also outputs 22,809 mSC and 24,795 nmSC non-zero direct/indirect TF-TF coordination $\bar{B}$ links (**Supp. File 2**). Notably, top CPPID contexts for strong mSC and/or nmSC $|\bar{B}|$ relate to DRG, brain (**Fig. S16A**). **Fig. 4B** shows 23 of 24 mSC-specific TFs in a mSC $\bar{B}$ network of 77 known PPIs, excluding novel links (w = 0.01) for simplicity (nmSCs: **Fig. S16B**); POU3F1-EGR2 mSC cooperativity is extremely strong ($\bar{B}$ = 99.14 percentile), JUNB-ATF4 ($\bar{B}$ = 96.98) interact in PNS neoplasms like Schwannomas. BNC2's regulatory activity in nmSCs may be attributed to the absence of its predicted repressor EGR2, ranked 10 of 20 mSC TFs for *BNC2*.

NetREm discovers and prioritizes novel TF-TF coordination links that are promising. 48 of our novel links in nmSCs (37 also in mSCs), comprising 30 TFs, are validated by strong SAINT scores of physical TF-TF binding in recent BioID/AP-MS human experiments (Göös et al. 2022) (**Fig. S16C**). RXRG, TEAD1, and YY1 co-regulate 366 nmSC (**Fig. S16D**) and 15 mSC TGs, suggesting their preferential coordination in nmSCs. RXRG links with STAT1, TEAD1, and YY1 are unknown in our input PPIN (**Fig. S17A**); these 4 core SC TFs co-regulate 174 TGs in nmSCs, *SETD2* in mSCs. RXRG strongly positively correlates with TEAD1 and YY1 (r = 0.5, 0.7) in nmSCs, but negatively in mSCs (-0.9, -0.7) in mice sciatic nerves (Gerber et al. 2021). STAT1 has 591 (Jaccard Similarity (JS): 0.21), YY1 has 599 (JS: 0.20), TEAD1 has 843 (JS: 0.26, significant) co-regulated nmSC TGs with RXRG (**Fig. S17B-C**). In nmSCs, RXRG and YY1 share 94 eSNPs, 95 eTGs, 88 eQTLs compared with 28, 54, 24 for RXRG and STAT1 (cooperate for 704 TGs (**Fig. S17D**), antagonistic for 104 (**Fig. S17E**)); RXRG- cooperation $\bar{B}$ is: TEAD1 (16.3), STAT1 (23.5), YY1 (24.1) (**Fig. S17F**).

To independently test our predicted TF-TF coordination for these core SC TFs, we use binding data from rat SCs in PNS, derived from ChIP-seq analysis of the active enhancer H3K27ac mark and ChIP-seq/Cut&Run read density assay (C&R) data of TF binding in nerve or S16 SC line. Of 15,864 ChIP-Seq H3K27 enhancer peaks shared between PNS and S16 lines, RXRG shares 3,450 and 2,017 peaks with YY1 and TEAD1 (**Fig. S17G**). 43.9% of RXRG peaks have YY1 binding (most colocalization among core SC TFs) and 25.7% of RXRG peaks have TEAD1 binding (**Fig. S17H**). Conversely, 24% of TEAD1 peaks and 28% of YY1 peaks have RXRG binding. C&R helps determine the extent of colocalized binding along REs. TEAD1 C&R centered across EGR2 peaks reveals TEAD1 colocalizes at ~40% of EGR2 TFBSs (**Fig. 4C**), supporting predicted EGR2-TEAD1 coordination in mSCs. YY1 C&R reads overlap ~70% when centered on RXRG peaks and SOX10 peaks. SC marker *CDH19* is preferentially expressed in nmSCs (Stratton et al. 2017) and has RXRG ChIP-seq binding nearby. Rs17799413 may associate with lower *CDH19* expression (slope: 2.4e-1), strongly altering binding of TEAD1 in both SCs and 7 core TFs (e.g. RXRG, YY1) in nmSCs. C&R sequencing reveals TEAD1 and RXRG binding peaks are 40 kb upstream of *CDH19*'s TSS (**Fig. 4D**). *NCAM1*, a YY1 LOF and SOX10 direct TG, codes for an adhesion molecule preferentially expressed in nmSCs (Martini and Schachner 1986). Our network predicts regulation by TEAD1, YY1, and RXRG in nmSCs. Rs10749999 associates with higher *NCAM1* expression (slope: 9e-2) and may boost YY1 and TEAD1 binding in nmSCs. Active SC enhancers ≈130 and 200 kb upstream of *NCAM1*'s TSS have TEAD1 and YY1 binding. *NCAM1* has TEAD1, RXRG, and YY1 binding at a promoter surrounding its gene locus. Rat H3K27Ac and TF ChIP-seq data shows RXRG co-regulates: 555 TGs with TEAD1, 352 with YY1, and 27 of 91 mapped STAT1 TGs. This further supports our predicted coordination by these 4 TFs, showing NetREm's prowess in identifying novel colocalizing TFs in the absence of current evidence of direct binding interactions from high-throughput studies of PPIs.

Computational methods help decipher functional impacts of SNPs on PPIs, aiding in uncovering disease risk genes for targeted precision therapies. Most non-coding, disease-associated SNPs alter human PPIs rather than protein properties like folding or stability (Cheng et al. 2021). Integrating TF-TG regulatory (c^*) and TF-TF coordination (B) networks with non-coding eSNP rs11663049 sheds insights on how PPIs associate with phenotype: polyneuropathy in diabetes (**Fig. 4E**). Dysregulated mitotic checkpoint regulators may lead to abnormal insulin signaling in diabetes (Choi et al. 2016). *CEP192* helps form mitotic spindles (Joukov et al. 2010) and colocalizes for this phenotype with 69% probability. The eSNP reduces the risk of this phenotype and *CEP192* expression and may strongly decrease binding of 6 activator TFs to *CEP192*'s promoter in nmSCs. *CEP192* is co-regulated by RXRG-STAT1 TG. RXRG-

368 NR4A2, NR4A2-THRB associate with PNS neoplasms. SOX10, with its dynamic, cell-type-specific cooperation,
369 works with these TFs; SOX family TFs achieve cell-type-specificity via partner TFs that facilitate TG regulation by
370 binding to nearby SOX TFBSs (Stevanovic et al. 2021). SOX10-NFIA cooperate in nmSCs ($\bar{B}$ = 81.3) but display
371 antagonism influencing glial lineage diversification (Glasgow et al. 2014) in mSCs ($\bar{B}$ = -78.7).

372 To illustrate differences in TG-specific TF-TF coordination networks B between SC sub-types, even for shared
373 TGs, we examine TG ART3 that colocalizes with neurofibromatosis (NF: characterized by the formation of NS-related
374 tumors) with a 68% probability (**Fig. 4F**).We predict an interacting DNA chromatin loop (Zhang et al. 2021) featuring
375 differing $\bar{B}$ between mSCs and nmSCs involving strong repressors and activators (Sharov et al. 2022) to regulate
376 *ART3*. 3 eSNPs link to lower *ART3* expression: NF-associated rs4859594 correlates with 2 regulatory SNPs
377 (rs6856681: intronic enhancer; rs34740469: promoter proximal RE) that may disrupt coordination networks by
378 strongly decreasing activator binding to *ART3's* REs, increasing it for repressors. In both SC types, common KLF9
379 coordinates with TFs at *ART3's* promoter, eagerly cooperating with: XBP1 (B = 4.84) in nmSCs, KLF6 (B = 31.3) in
380 mSCs. PPARA-NFIX (nmSCs) and NR3C2-PBX3 (mSCs) are antagonistic relations on the enhancer.

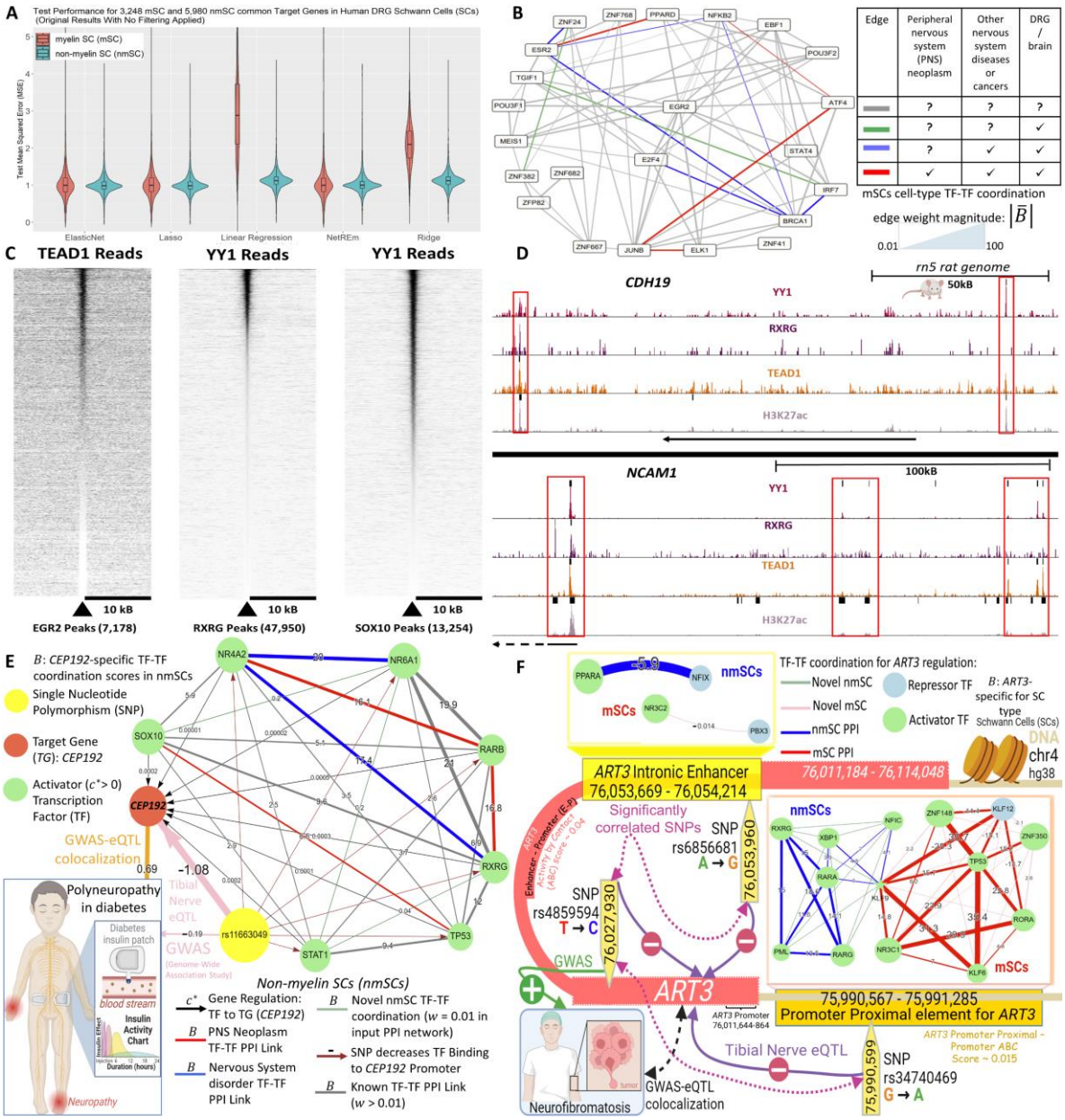


Figure 4 Coordination among TFs for gene regulation in human mSCs & nmSCs **A)** Density boxplots: NetREm outperforms Linear Regression and RidgeCV in both SCs, with lower test MSEs. NetREm predicts links for more TGs than ElasticNet and Lasso do. For mSCs, NetREm achieves median MSE: 0.95; nmSC: 1.1. **B)** Input PPIN subnetwork for only known TF-TF PPIs among 23 of 24 mSC-specific TFs only; artificial (novel) links excluded. Edges represent cell-type TF-TF cooperation ($\bar{B}$) across TGs in mSCs since $\bar{B} > 0$ for all. High $\bar{B}$ = stronger cooperation in mSCs. Edges annotated based on CPPID. **C)** C&R heatmaps show overlaps of TEAD1 centered on EGR2 peaks, YY1 C&R centered on RXRG peaks, YY1 C&R centered on SOX10 peaks. **D)** YY1, RXRG, and TEAD1 C&R tracks along with ChIP-seq H3K27ac tracks from rat S16 SC line. Boxes highlight enhancer regions where TFs colocalize across TGs *CDH19* (top tracks) and *NCAM1* (bottom tracks). Both TGs are impacted by Tibial Nerve eSNPs altering regulatory TF binding at their loci. **E)** *CEP192* in nmSCs GWAS-eQTL colocalizes with low risk of polyneuropathy in diabetes. Tibial Nerve eSNP (yellow) associates with low risk of this condition and low *CEP192* expression by potentially strongly disrupting bindings of 6 activators (green) to *CEP192* promoter. SOX10 cooperates with them. *CEP192*'s *B* links: novel (teal); known TF-TF PPIs colored based on CPPID annotations: PNS Neoplasms (red), NS disorders (blue), neither (grey). **F)** Tibial Nerve eSNPs potentially influence regulation of GWAS-eQTL colocalized neurofibromatosis TG *ART3* through activators and repressors forming distinct TRMs in mSCs and nmSCs along interacting REs (3D DNA loop). *ART3*'s (intronic) promoter has Activity

by Contact (ABC) scores of ~0.04 with intronic enhancer, ~0.015 with proximal promoter RE. TF-TF B edge widths are shown relative to other TF-TF B links in that given *ART3* RE (i.e. intronic enhancer, promoter proximal).

Prediction and comparative analysis of cell-type coordination among TFs for gene regulation across neuronal and glial cell types in Alzheimer's disease (AD)

Cell-type-specific TF-TF coordination networks $\bar{B}$ are crucial for neuronal functions like synapse plasticity and neurotransmission, and are disrupted in AD, leading to memory loss, neuroinflammation, cognitive decline (Mathew et al. 2022). Understanding how altered $\bar{B}$ impacts TG expression in AD is essential for identifying master regulators and developing targeted therapies (Wang et al. 2016). Signaling PPIs associated with dementia symptoms highlight the potential of targeting these altered PPIs (Mao et al. 2020) to delay AD progression (Vargas et al. 2018).

In response, we integrate multi-omics data to construct 16 context-specific prior GRNs for 8 cell types in AD and Controls. For each TG in a context, NetREm uses the N TG-specific candidate TFs from the respective prior GRN as input features, based on TFs that may associate with its REs to regulate it (**Methods**). Ultimately, NetREm generates 16 corresponding TF-RE-TG (TF-TG links: **Supp. File 3**) and $\bar{B}$ (**Supp. File 4**, word cloud: **Fig. S18A**) networks. Reverse engineering changes in networks across cell types may help illuminate molecular drivers of AD.

We explore TG-specific coordination B in Control vs. AD stages for 2 AD risk genes (Jia et al. 2020; Bossaerts et al. 2022) (**Fig. 5A**). *TMPRSS15* in Mic (t: 36.1, p-adj < 4.8e-283) and *ABCB5* in Inhibitory Neurons (InNs) (t: 22.6, p-adj < 4.5e-112) show notable increases in B in AD versus baseline Controls for the respective cell-type. Some TF-TF pairs strongly cooperate (i.e. $50 \leq B \leq 100$) exclusively in AD: ZBTB14-ZNF281, FLI1-TAL1 for *TMPRSS15*; ZNF331-ZNF354A, MYEF2-SOX2 for *ABCB5*. For *TMPRSS15*, known AD links IRF7-STAT3 and STAT3-STAT5B have strong antagonism (i.e. $-100 \leq B \leq -50$) in controls but cooperate in AD. **Fig. 5B** compares strong $\bar{B}$ among select TFs between conditions for Mic and InNs. IRF7-STAT3 is in Mic, MYEF2-SOX2 is in InNs only. AD link STAT3-STAT4 is antagonistic in Controls for *TMPRSS15* in Mic and in Control Mic overall (but cooperative for other 3 networks). RORA-ELK1 is cooperative in all 3 but antagonistic in Control Mic. FOSL2-BACH1 is antagonistic in Mic but cooperative in InNs. Indeed, RORA activity increases in InNs and Mic in AD (Acquaah-Mensah et al. 2015). AD-annotated links in AD InNs/Mic include ELK1-SPI1 and ELK1-STAT3.

Fig. 5C presents a multifaceted network weaving together TF-TG regulatory links, TF-TF coordination, phenotypes, SNPs. TF-TF coordination is discernibly stronger and positive in Control Mic (1-sided t), pointing to potentially disrupted cooperativity during AD. Our attention is drawn to regulation of *ANXA11*, a critical player in

diverse functions (e.g. apoptosis, neutrophil function) and signaling paths (e.g. MAPK, P53) (Mirsaeidi et al. 2016). Mutations in *ANXA11* are correlated with NS diseases (Wang et al. 2022) and high risk of inflammatory conditions like sarcoidosis (Smith et al. 2017). Our AD-Covid study (Khullar and Wang 2023) assigns *ANXA11* to a Hippocampal Control gene co-expression module. NetREm offers nuanced insights on non-coding SNPs for AD (rs11202929) and cognitive function (rs12412257), linked to more and less *ANXA11* expression, respectively, in Mic (resident CNS macrophage immune cells). Rs12412257 associates with reduced word interpolation ability, a measure of fluid intelligence and reasoning (Turley et al) and prognostic marker of AD (Eyigoz et al. 2020). Rs11202929 is protective against AD (GWAS slope < 0) and may enhance binding affinity of 12 cooperating activators to TFBSs on *ANXA11's* enhancer in open chromatin in adult Mic. SPI1 is a core Mic TF for AD genes (Rustenhoven et al. 2018). *ANXA11* has higher Mic expression in controls than in AD (t p-adj = 2.6e-68, $\log_2$(Fold Change) of (Control/AD) = 0.78; 668 AD, 676 controls). Our significant GWAS-eQTL colocalization analyses reveals that *ANXA11* expression positively associates with lower AD risk and better word interpolation ability. NetREm provides a powerful framework to deepen our understanding of complex GRNs and their implications across a spectrum of health conditions.

We further evaluate our $\bar{B}$. **Fig. S18B** compares $\bar{B}$ across 16 networks for 216 novel links validated in recent physical human experiments (Göös et al. 2022). In control ExNs and InNs, $r \approx 0.54$ and 0.45, respectively, between $\bar{B}$ and JS of ChIP-seq overlaps (proxy for cooperativity (Yu et al. 2015)) for 6 TFs in neural cells (**Fig. S19**).

For each cell-type, we build default Random Forest (RF) (**Fig. 5D)**, Logistic Regression, Naïve Bayes, and XGBoost models (Pedregosa et al. 2011) to detect TGs with altered B from Controls to AD that may predict TF-TF links annotated (CPPID) with neurodegenerative disease (**Fig. S20**). Input data consists of changes in B across TGs from Control to AD, for the given cell type. To tackle this positive unlabeled learning problem (Yang et al. 2014), we undersample class 0 so each cell-type has an equal count of TF-TF links in both classes (**Supp. File 5**). We evaluate performance via stratified 5-fold CV noting RF is the most optimal approach as all 8 RF cell-type models have average area under Receiver-Operator Characteristic (ROC) curve (AUC) ≥ 0.81. Across our models, the top 500 RF feature TGs are enriched for neurodegeneration, cell-type functions, immunity, intellectual disabilities, tauopathy (**Fig. 5E**). OPCs and InNs have the highest overlap (35 top TGs); disrupted InN signaling to OPCs may diminish myelination and CNS interneuron activity, and severely impair prefrontal cortical network functions and social cognitive behavior (Fang et al. 2022) (**Table S15**).

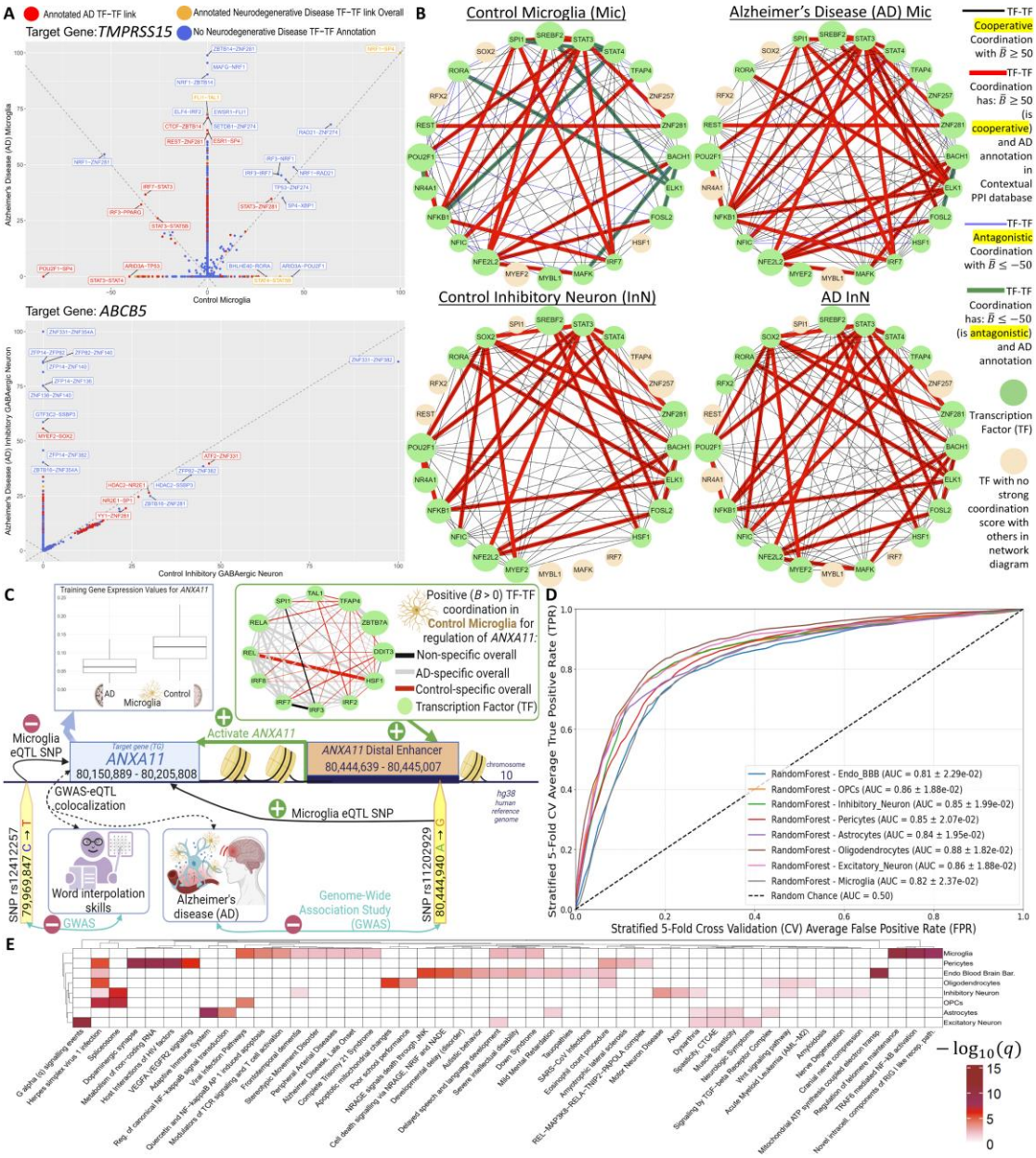


Figure 5 Prediction and comparative analysis of cell-type coordination among TFs for TG regulation across neuronal and glial cell types in AD. **A)** Coordination scores B for 2 AD risk TGs, *TMPRSS15* for Mic and *ABCB5* for InNs, in Control versus AD. Red and orange points are TF-TF links annotated by CPPID for AD and other neurodegenerative diseases, respectively. Blue points: no annotation for neurodegeneration. **B)** Circular network diagrams visualize strong cell-type direct/indirect TF-TF coordination in AD/Control Mic and InNs, focusing on light green TFs that exhibit potentially strong: cooperation ($50 \leq \bar{B} \leq 100$) or antagonism ($-100 \leq \bar{B} \leq -50$). Thick edges are AD-related in CPPID. Peach: TFs not expressed or lacking strong links with other TFs in this visual. **C)** 12 TFs that may cooperate to activate *ANXA11* in TF-TG regulatory network in Control Mic (higher expression for *ANXA11* than in AD). Widths for TF-TF links are B scores for *ANXA11* regulation in Controls and colors are based on statistical significance of TF-TF links across all TGs in both Mic networks. Light gray links: higher in AD; red: higher in Controls; black: not significant. A SNP correlating with lower AD susceptibility, increases binding of this TF-TF network, and links to higher *ANXA11* expression. Risk SNP for reduced cognitive ability (poor word interpolation skills) links to lower *ANXA11* expression. **D)** Average stratified 5-fold CV ROC curves for RF models predicting TF-TF links annotated in neurodegenerative diseases (class 1) or not (class 0) for balanced classes. **E)** Enrichment analysis for top 500 optimal genes identified by RF for each of 8 cell types. Hierarchical clustering is performed on rows (cell types) and columns (terms), and $-\log_{10}(q)$ values are shown.

Discussion

In this paper, we present NetREm, a computational multi-omics-based approach that employs network-regularized regression on single-cell gene expression data to predict cell-type coordination among TFs for TG regulation. NetREm addresses a major challenge in traditional studies of cell-type GRNs. Gene expression data, often nascent, sparse and noisy, fails to capture crucial GRN mechanisms such as TF binding to DNA, coordination among TFs/cofactors, and DNA accessibility, and typically provides weak signals for distinguishing TP from FP TF-TG links (Kim et al. 2023; Badia-i-Mompel et al. 2023). Sole reliance on expression data for GRN inference is therefore woefully inadequate, and perhaps even futile, often leading to unstable and inaccurate results.

Explicitly modeling direct and indirect TF-TF interactions can enhance GRN inference, enabling discovery of novel TF-TG links and key cell-type TFs (Skok Gibbs et al. 2022). Functionally-related TF predictors, like neighbors in scale-free feature networks (e.g. TF-TF PPINs), can coordinate synergistically or antagonistically in biological processes like TG regulation (Kong and Yu 2018). Nonetheless, traditional methods often miss such complex dynamics of TF-TF PPIs involved in GRNs (Yazaki et al. 2016).

SOTA GRN tools like SCENIC indirectly hint at TF-TF interactions by analyzing TFs that co-regulate multiple TGs. However, these tools primarily focus on TFs with strong motif binding, excluding other prior information. For instance, in our comparative analysis, SCENIC identifies many TFs in SCs (nmSCs: 640, mSCs: 522) but overlooks TEAD1 in both SCs, instead detecting 3 other TEAD family TFs. TFs like TEAD1, which exhibit relatively weak motif-binding signals, are drowned out. In contrast, by incorporating TF-TF PPINs, NetREm effectively captures essential GRN relations for core SC TFs like TEAD1. This underscores the importance of integrating comprehensive prior information like PPINs, in GRN predictions from expression regression, a capability that NetREm successfully implements (Dibaeinia and Sinha 2020).

NetREm reveals cell-type coordination among TFs, $\bar{B}$, with some mediated by physical and others by indirect (e.g. pioneer/settler models show TFBSs are often >50 bp apart in REs) PPIs (Martin et al. 2023). By weighing known direct/indirect PPIs in the context of TG regulation, NetREm helps characterize existing PPINs at a cell-type level (Johnson et al. 2021; Murtaza et al. 2022; Hsu et al. 2022). It also helps address the link prediction problem, flagging undiscovered PPIs for follow-ups (Singh and Vig 2017). The lack of cell-type PPI annotations, while a challenge, offers NetREm an opportunity to contribute to ongoing efforts to broaden understanding of PPIs and protein dynamics with its dual capacity to annotate known TF-TF PPIs and discover novel cell-type-specific ones (Yu et al. 2023).

NetREm predicts unprecedented, weighted, cell-type-specific TF-TF coordination networks $\bar{B}$ across various conditions, including both human and mouse contexts, even in the absence of prior GRNs. Our benchmarking shows our TF-TG regulatory networks not only perform competitively with SOTA GRNs but also uncover novel cell-type TFs that coregulate TGs; further, our $\bar{B}$ effectively prioritizes TP and FN TF-TF links.

Disrupted cell-type PPIs are critical in neurobiological disorders, since PPIs mediate neuronal functions (Mathew et al. 2022). We integrate multi-omics data, capturing various levels of TG regulation (e.g., scRNA-seq, scATAC-seq), to learn prior GRNs for NS cell-types for our SC and AD human applications. Detecting these candidate GRN TFs for TGs is key to inferring biologically significant cell-type TF-TG links (Zaborowski and Walther 2020; Zhang et al. 2023). Aligning NetREm's TF-TG links with prior GRNs helps us deduce TF-RE-TG links. NetREm uncovers novel TF-TF crosstalk for TG regulation in SCs and during AD in neurons/glia.

Insights derived from NetREm may contribute to advancing targeted therapies and regenerative medicine. We apply our predicted regulatory and coordination networks to trace how non-coding eSNPs may alter co-regulatory dynamics among TFs, potentially altering expression of disease-associated eTGs.

NetREm extends to any discipline where predictors exhibit a network structure that informs the outcome. It expands upon previous work in network regularized regression by learning and generating embeddings using SVD. In the future, we can incorporate nonlinear dimensionality reduction into NetREm to capture nonlinear patterns and regularize latent representations with prior information. We may see if any final TFs form homodimers or adapt NetREm to account for this. Additionally, regularization networks can integrate other information, like signaling pathways, to learn TF-TF coordination for TG regulation. Beyond expression regression, we can extend NetREm to other emerging single-cell omics like scATAC-seq to explore TF and chromatin interactions in open regions.

Methods

NetREm Methodology

Integrating multimodal data & networks in NetREm workflow

Preliminary definitions: Proteins (e.g. TFs), and TGs (*italics*) are represented by HGNC symbols. TF gene expression is a proxy for TF protein abundance, assuming high gene expression (quantifies mRNA abundance) translates to high protein expression. NetREm can be applied to both single-cell and bulk expression data (sample by gene). Bulk data represents pooled collections of cell lines or tissues, yielding averaged expression profiles, as seen in patients.

We start with single-cell gene expression data for M samples (individual cells) and $\mathcal{T}$ genes in a cell-type; scRNA-seq data is typically high-dimensional ($M \ll \mathcal{T}$), sparse, non-negative. We focus on G TGs, (i.e. $\{\mathrm{TG}_k\}_{k=1}^{G}$), with expression profiles $\{y_k\}_{k=1}^{G}$, respectively, and $\mathcal{N}$ potential cell-type TFs, where $G \leq \mathcal{T}$, $\mathcal{N} < \mathcal{T}$. If all TFs are master regulators: $\mathcal{T} = G + \mathcal{N}$; else (some TGs are TFs): $\mathcal{T} < G + \mathcal{N}$. We ensure a TG is not its own candidate TF.

NetREm can use prior GRN information that identifies $\{N_k\}_{k=1}^{G}$ respective promising candidate TFs for the set of G TGs where $N_k \leq \mathcal{N}$ and varies based on the optimal prior GRN TFs selected for the given TG_k. When prior GRN information is absent, TG_k (and other $G - 1$ TGs) will have the same fixed $N_k = \mathcal{N}$ candidate TFs; it is to be understood that if TG_k is also a TF, its self-TF is excluded, so it has $N_k = \mathcal{N} - 1$ candidate TFs.

These steps below are repeated for each of the G TGs. For simplicity, we explain the pipeline for predicting expression of a single TG. We use N to represent the # of its candidate TFs for this TG, and y for its true expression.

Optional (recommended) prior cell-type GRN information: Constructing candidate links from TFs to TGs may improve the quality of NetREm's solution via initial feature selection of N biologically meaningful TFs tailored for TG, where $N < \mathcal{N}$. Using diverse information on expression regulation guides the regression (Zaborowski and Walther 2020). We integrate prior GRNs (initial TF-RE-TG links) from multi-omics data in applications 6-7. When no prior GRN is used: if TG is a TF, it has $N = \mathcal{N} - 1$ candidate TFs, otherwise it has all $\mathcal{N}$ TFs as candidates.

PPI network (PPIN, network prior): Our comprehensive, weighted, undirected PPIN $\mathbb{W}$ illuminates biological interactions among proteins (network nodes) with strong functional association evidence in the organism. It inherently captures direct (e.g. complex formation, transient interactions) and/or indirect (e.g. participate in shared processes or bind by intermediate hidden partners (De Las Rivas and Fontanillo 2012)) PPIs. Alas, PPINs attribute weight $w > 0$ to all PPIs, even to those with antagonistic, competing roles in paths (Szklarczyk et al. 2023); each edge w accounts for uncertainty in $\mathbb{W}$ and is proportional to the probability that the 2 connected nodes interact (i.e. their integrative functional essentiality in $\mathbb{W}$) (Li and Liu 2022). For the TG, we subset $\mathbb{W}$ to obtain $W^{(0)}$ that captures known TF-TF PPIs, reflects potential structure/relation background information among TG's N candidate TF proteins, and is symmetric (i.e. $\left(W^{(0)}\right)^T = W^{(0)}$). Higher w_{ij} denotes more confidence that TF_i and TF_j partner, directly and/or indirectly, in processes like regulating DNA chromatin loops of interacting REs for TG regulation (Wang et al. 2021). Here, $w_{min} = \min\{w_{ij} \in W^{(0)}\}$ is the smallest confidence for known direct/indirect TF-TF links among N TFs. To enable NetREm to consider candidate TFs missing from $W^{(0)}$, we add artificial weight $0 < \eta < w_{min}$ for missing

pairwise edges; these novel edges may not exist (i.e. TFs truly do not coordinate: True Negatives) or are yet to be discovered (FNs). We obtain our final, fully-connected, TG-specific, TF-TF input PPIN: $W \in \mathbb{R}^{N \times N}$. W has $\frac{N(N-1)}{2}$ unitless, global PPIs with $w_{ij} > 0$ equal to: $w_{ij}^{(0)}$ for known and η for artificial TF-TF links. We do not consider self-loops; instead, we set $W_{ii} = \frac{d_i}{N-1}$ where TF_i's degree (connectivity) with other $N-1$ TFs is $d_i = \sum_{k \neq i} w_{ik} > 0$.

<u>*NetREm integrates input data:*</u> For the M samples, the $X \in \mathbb{R}^{M \times N}$ matrix contains expression data for N predictors, while $y^{(0)} \in \mathbb{R}^M$ is the expression vector for TG. We standardize each column of X ($X_{ij} \leftarrow \frac{X_{ij}-\mu_j}{\sigma_j}$) and y ($y_i \leftarrow \frac{y_i-\mu_y}{\sigma_y}$) by respective means μ_j, μ_y and standard deviations σ_i, σ_y, making them unitless. Then, each TF in X has $\overline{\mu_j} \approx 0$ and $\sigma_j \approx 1$; $\mu_y \approx 0$, $\sigma_y \approx 1$. Pairwise r are preserved. If $M < N$, X suffers the curse of dimensionality. With technological advances and the advent of large single-cell sequencing studies, we anticipate a boost in M so $M \gg N$ will soon be the norm (Cuomo et al. 2023), especially since $N \ll \mathcal{T}$ (relatively few genes are transcribed and translated to proteins that are TFs). NetREm identifies which of N TFs can predict TG expression y, considering PPIs among TFs. NetREm can comprehensively integrate multi-omics data and PPINs to discover key TFs and TF-TF coordination events for TG regulation in a cell-type-specific manner.

<u>Step 1: Network regularized regression (Problem Definition)</u>

Given gene expression data with M samples (rows) and N features (columns) represented as $X \in \mathbb{R}^{M \times N}$ (e.g. N candidate TF expression values for each M) and response $y \in \mathbb{R}^M$ (e.g. TG expression for each M), we want to learn a linear predictor $y \approx Xc^*$ for *TG.* By imposing prior TF-TF PPIN information as a regularization term, we develop a functional map highlighting connectivity patterns among N TFs for TG; this typically steers NetREm to favor groups of TFs with shared PPIN connectivity over isolated TFs, thereby enhancing its ability to capture biologically relevant PPIs (Li and Liu 2022). Coefficients $c^* \in \mathbb{R}^N$ represent the importance of TFs for regulating TG and are found by optimizing the following problem with objective function:

$$c^* = \underset{c}{\operatorname{argmin}}\, f(c) = \frac{1}{2M}\left|\left|y - Xc\right|\right|^2 + \alpha\left|\left|c\right|\right|_1 + \frac{\beta}{2}\sum_{i=1}^{N}\sum_{j=i}^{N} w_{ij}\left(\frac{c_i}{\sqrt{d_i}} - \frac{c_j}{\sqrt{d_j}}\right)^2 \qquad Eq\ (1)$$

Our 3 terms in equation Eq (1) are unitless and compatible for addition:

1 (data-fitting): ensures Xc is close to y, and $\frac{1}{2M}$ is a normalization factor to make it invariant to sample size M.

2 (sparsity-prior): favors a sparse solution (small # of non-zero c^*), helping simplify the model and boost reliability. **3** (network-prior): penalizes differences between c^* of connected TF nodes, normalized by their respective network centrality d and adjusted for their global (organism-based) PPI weights with other candidate TFs in TG-specific input W. Inspired by (Li and Li 2008), this approach allows for a more equitable representation of TFs, irrespective of their d; this network-oriented variant of Ridge L_2 penalty $\sum_{i=1}^{N} c_i^2$, promotes topology-aware c^* shrinkage and smoothing for neighboring TF_i and TF_j, with probability proportional to w_{ij}. It underscores the principle that strongly connected TFs likely perform shared functions, even if their influence on TG expression (c^* signs) differs. This recognizes the community structure in existing PPINs that groups proteins with similar biological roles with $w > 0$, not distinguishing between cooperative (+) and antagonistic (-) PPIs (Padi and Quackenbush 2015; Szklarczyk et al. 2023). NetREm leverages this refined understanding, offering a comprehensive perspective on the interplay of TFs in W.

We can tune 2 hyperparameter knobs: network-constrained prior, $\beta > 0$, decides strength of PPIN regularization penalty (applied 1st: higher β guides NetREm to prioritize TFs with strong PPIs); $\alpha \geq 0$ impacts $\mathcal{L}_1$ penalty (applied 2nd). NetREm, a PPIN-aware adaptation of ElasticNet, performs automatic variable selection based on expression and PPIN data, grouping and selecting strongly-connected TFs (emphasizing known TF-TF PPI subnetworks) in a spirit akin to ElasticNet. If $M \ll N$, ElasticNet and NetREm may still select $\leq N$ TFs as final; this addresses limitations of Lasso regression that may indiscriminately select only 1 TF from a group of highly correlated TFs and only $\leq M$ TFs if $M \ll N$ (Zou and Hastie 2005; Li and Li 2008).

Step 2: Gene expression embeddings from network regression

Our novel method transforms the original problem into a Lasso regression problem in a new space with cell-type-specific TF-TF interactions. 1st, we represent the network-prior term in a more compact matrix-vector form as:

$$\frac{\beta}{2}\sum_{i=1}^{N}\sum_{j=i}^{N} w_{ij}\left(\frac{c_i}{\sqrt{d_i}} - \frac{c_j}{\sqrt{d_j}}\right)^2 = \frac{\beta}{2} c^T A c$$

where $A \in \mathbb{R}^{N \times N}$, a symmetric normalized Laplacian matrix, is computed as:

$$A = D^T(W \odot V)D = \begin{bmatrix} 1 & -w_{12}/\sqrt{d_1 d_2} & \cdots & -w_{1N}/\sqrt{d_1 d_N} \\ -w_{21}/\sqrt{d_1 d_2} & 1 & \cdots & -w_{2N}/\sqrt{d_2 d_N} \\ \vdots & \vdots & \ddots & \vdots \\ -w_{N1}/\sqrt{d_1 d_N} & -w_{N2}/\sqrt{d_2 d_N} & \cdots & 1 \end{bmatrix},$$

where $D, V, I \in \mathbb{R}^{N \times N}$. $D = diag\left(1/\sqrt{d}\right)$ is a diagonal matrix with elements $1/\sqrt{d_i}$, $V = N \cdot I - 11^T$ is a constant matrix where $1 \in \mathbb{R}^{N \times 1}$ is an all 1 column vector and I is an identity matrix. $W \odot V$ is element wise multiplication of W and V. Using this matrix-vector representation, we reformulate $f(c)$ in Eq (1) as

$$f(c) = \frac{1}{2M}\left|\left|y - Xc\right|\right|^2 + \alpha\left|\left|c\right|\right|_1 + \frac{\beta}{2}c^T Ac$$

$$= \frac{1}{2M}(y^T y - 2y^T Xc + c^T X^T Xc) + \alpha\left|\left|c\right|\right|_1 + \frac{\beta}{2}c^T Ac$$

$$= \frac{1}{2M}c^T(X^T X + \beta MA)c - \frac{1}{M}y^T Xc + \alpha\left|\left|c\right|\right|_1 + \frac{1}{2M}y^T y\,.$$

We set $E = \frac{X^T X}{M} + \beta A \in \mathbb{R}^{N \times N}$, which is symmetric and positive semi-definite (psd) since A and Gram matrix $X^T X \in \mathbb{R}^{N \times N}$ are symmetric and psd and $\beta, M, N > 0$. Cell-type-specific data term $\frac{X^T X}{M}$ represents r among TFs in the expression data, the covariance matrix of X scaled by $\frac{1}{M}$. Cell-type-independent PPI term βA is based on global PPI W and is β for main-diagonals and is $-\beta \leq \beta A_{ij} < 0$ for $i \neq j$. Thus, $E_{ij} = r\left(x_i, x_j\right) - \beta\frac{w_{ij}}{\sqrt{d_i d_j}}$, which balances r with PPIN knowledge (for a given β value) and information $\left(\frac{w_{ij}}{\sqrt{d_i d_j}}\right)$.

$$f(c) = \frac{1}{2M}c^T(X^T X + M\beta A)c - \frac{1}{M}y^T Xc + \alpha\left|\left|c\right|\right|_1 + \frac{1}{2M}y^T y$$

$$= \frac{1}{2N}c^T \tilde{X}^T \tilde{X}c - \frac{1}{N}\tilde{y}^T \tilde{X}c + \alpha\left|\left|c\right|\right|_1 + \frac{1}{2M}y^T y$$

$$= \frac{1}{2N}\left|\left|\tilde{y} - \tilde{X}c\right|\right|^2 + \alpha\left|\left|c\right|\right|_1 + \frac{1}{2M}y^T y - \frac{1}{2N}\tilde{y}^T \tilde{y}\,, \qquad Eq\ (2)$$

where $\tilde{X} \in \mathbb{R}^{N \times N}$ and is symmetric (i.e. $\tilde{X} = \tilde{X}^T$) and $\tilde{y} \in \mathbb{R}^N$ satisfies:

$$\frac{1}{N}\tilde{X}^T \tilde{X} = \frac{1}{M}X^T X + \beta A \qquad Eq\ (3a)$$

$$\frac{1}{N}\tilde{y}^T \tilde{X} = \frac{1}{M}y^T X \qquad Eq\ (3b)$$

Finally, we reformulate Eq (2) as a conventional Lasso problem (by omitting constant term $\frac{1}{2M}y^T y - \frac{1}{2N}\tilde{y}^T \tilde{y}$) that we solve using existing standard Lasso or LassoCV (Pedregosa et al. 2011) regression solvers:

$$c^* = \underset{c}{\operatorname{argmin}}\, \tilde{f}(c) = \frac{1}{2N}\left|\left|\tilde{y} - \tilde{X}c\right|\right|^2 + \alpha\left|\left|c\right|\right|_1 \quad Eq\ (4)$$

To compute $\tilde{X}$ and $\tilde{y}$ we perform SVD on E expressed as: $E = U\Sigma U^T$ where $U, \Sigma \in \mathbb{R}^{N \times N}$. Here, U has left singular vectors and Σ is a diagonal matrix of non-negative singular values $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$ of E. Then, $E = U\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}U^T =$

$\left(\Sigma^{\frac{1}{2}}U^T\right)^T\left(\Sigma^{\frac{1}{2}}U^T\right)$. Based on Eq $(3a)$, $E = \frac{1}{N}\tilde{X}^T\tilde{X}$. Then, $\tilde{X}^T\tilde{X} = NE$. Hence, $\left(\sqrt{N}\Sigma^{\frac{1}{2}}U^T\right)^T\left(\sqrt{N}\Sigma^{\frac{1}{2}}U^T\right) = \tilde{X}^T\tilde{X}$. For improved stability, we set small singular values (e.g., $s_i < 10^{-6}s_{\max}$) to 0, resulting in truncated Σ_{trunc}, and similarly, we adjust inverse ${\Sigma_{\text{trunc}}}^{-1}$ setting inverse elements corresponding to small singular values to 0. By substituting ${\Sigma_{\text{trunc}}}^{-\frac{1}{2}}$ in place of $\Sigma^{-\frac{1}{2}}$, we effectively use truncated SVD, enhancing NetREm's robustness by excluding contributions from small singular values. Thus, X and y are transformed to a new latent space of gene expression embeddings that incorporate PPIN information: $\tilde{X} \in \mathbb{R}^{N\times N}$ and $\tilde{y} \in \mathbb{R}^N$, respectively, in Eq (4):

$$\tilde{X} = \sqrt{N}{\Sigma_{\text{trunc}}}^{\frac{1}{2}}U^T$$

$$\tilde{y} = \frac{\sqrt{N}}{M}{\Sigma_{\text{trunc}}}^{-\frac{1}{2}}U^TX^Ty$$

When $\beta = 0$, the transformation yields X's principal components (PCs) via SVD on X^TX, a trivial case without PPIN information; however, since we require $\beta > 0$, $\tilde{X}$ not only reflects its PCs but also includes PPIN structure, with βA added to $\frac{1}{M}X^TX$, creating an "embedding". This comprehensive approach captures both data patterns and PPI relations. We perform Lasso regression to solve Eq (4) for $\tilde{X}$ and $\tilde{y}$, determining optimal c^*. **Supp. File 5** provides more details.

Output 1: Identification of novel cell-type TFs for regulating TG

TF-TG regulatory network for TG:

Several TFs regulate transcriptional activity of TG in a cell-type at a certain time (Nie et al. 2011). Solving the network-regularized regression problem produces a vector of Lasso $c^* \in \mathbb{R}^N$ for N TFs predicting true TG expression y. We focus on $c^* \neq 0$, which represents N^* final TFs for TG out of N candidates, where $0 < N^* \leq N$. NetREm constructs a comprehensive directed TF-TG regulatory network (complementary GRN) of N^* edges, weighted by c^*. Here $|c_i^*|$, indicates the strength of the TF_i-TG link, measuring TF_i's relative importance in regulating TG. TF_i with $c_i^* > 0$ may activate TG and TF_j with $c_j^* < 0$ may repress TG transcription and subsequent expression. This tug-of-war between activators and repressors orchestrates TG regulation. Biological complexity enables certain TFs to have dual-function roles, alternating between activation and repression depending on context and signals (Skok Gibbs et al. 2022; Boyle and Després 2010). Given the competitive nature of TFBS binding, $N - N^*$ discarded TFs may lose to some of the N^* TFs (i.e. antagonistic relation), but we do not speculate on this. Overall, NetREm unearths novel cell-type-specific, coordinating TFs involved in TG regulation, providing a more nuanced view of TF-TG interactions.

We evaluate our performance in training and testing expression data, comparing predicted $\hat{\tilde{y}} \in \mathbb{R}^{N\times 1}$ to actual $\tilde{y}$ using metrics like Mean Square Error (MSE) = $\frac{1}{N}\sum_{v=1}^{N}\left(\widehat{\tilde{y}_v} - \tilde{y}\right)^2$. To achieve more accurate regulatory links, we

integrate multiomics data like TF-DNA-binding (Dibaeinia and Sinha 2020) in applications 6-7 to predict prior GRNs. We identify REs for TG and determine TFs likely to bind directly to or associate indirectly with TFBSs on these REs. We input N TG-specific candidate TFs to NetREm for TG. Then, we overlay NetREm's N^* TF-TG regulatory links for TG with this prior GRN (initial TF–RE –TG links for N TFs for TG). This helps us annotate our links with epigenomic information on REs. Ultimately, we isolate highly-confident final TF-RE-TG links for our final N^* TFs.

Cell-type TF-TG regulatory network*:*
We apply NetREm, iteratively, to each of the G TGs and weave together individual TF-TG links (details: **Supp. File 5**). We may narrow down links by retaining TFs with $|c^*| > c_{min}$ (min threshold, default: 0) and TGs meeting specific criteria (e.g. $MSE_{TG} < MSE_{max}$). Our cell-type-specific complementary GRN relates TFs to TGs they regulate, helping explain how cell-types establish and maintain cellular identity. We may annotate/validate this network by identifying eSNPs impacting TF binding with eQTL links to altered TG expression (Coetzee et al. 2015); when prior GRNs are used, we ensure SNPs fall in the same REs where TFs are predicted to bind, linking them to TG regulation.

Output 2: TF-TF coordination (direct/indirect interactions)

TG-specific cell-type coordination B:
NetREm helps fulfill the need for cell-type-specific proteome analysis by which proteins interact to carry out processes like TG regulation. Existing PPINs aggregate direct and/or indirect PPIs in an organism. This broad approach has limitations, as not all proteins are expressed in every cell or tissue type, and some may be aberrant in diseases (Padi and Quackenbush 2015). To overcome this, there are efforts to annotate global PPIs at various levels, including tissue-specific protein expression, cell-line-specific links, phenotype-based studies (e.g. CPPID) (Federico and Monti 2020). However, co-expression of TFs need not imply they interact in specific cell or tissue types (Gonzalez-Teran et al. 2022). Further, PPINs do not distinguish between cooperation and antagonism. This underscores the need for NetREm's 2nd output that predicts how TFs coordinate to regulate TG in the cell-type.

NetREm's 2nd output is a weighted and signed TG-specific TF-TF coordination network given by an adjacency matrix of coordination scores B, if $N^* \geq 2$ final TFs for TG. These scores are a function of both embeddings $\tilde{X}$ and c^*: $B = f(\tilde{X}, c^*)$. In our framework, $B_{ij} > 0$ suggests cooperativity (e.g. co-binding, pioneer-settler TF relations) and $B_{ij} < 0$ indicates antagonism (e.g. sequestration) between TF_i and TF_j for co-regulating TG.

We use c^* to predict the nature of interactions among N^* TFs for TG in a symmetric matrix $C \in \mathbb{R}^{N \times N}$. Here, $C_{ij} = sign(c_i, c_j) = \{1 \text{ if } c_i^* c_j^* > 0;\ -1 \text{ if } c_i^* c_j^* < 0;\ 0 \text{ otherwise}\}$ and $C_{ii} = 0$. RTNduals assesses coordinated behavior of 2 TFs by analyzing correlation distributions between them and their shared TGs; building on approaches like these, we use C to deduce relative coordination relations among TFs for TG, acknowledging TFs may exhibit antagonistic or cooperative interactions depending on TG and context. If $C_{ij} > 0$, both TFs likely cooperate, aiming to either upregulate or downregulate TG expression in unison; their combined synergistic net effect on the TG is stronger than their individual effects. Conversely, if $C_{ij} < 0$, both likely act antagonistically, with conflicting influences on TG expression; this activator-repressor antagonism weakens their combined effect compared to their individual impacts, potentially due to partially canceling each other's activities (Berenson et al. 2023). When $C_{ij} = 0$ (and $i \neq j$), at least 1 of the 2 TFs is not a final TF and we cannot ascertain their potential nature of interaction.

Earlier, we set $E = \frac{X^T X}{M} + \beta A$. The 1st term $\frac{X^T X}{M}$ represents the original normalized inner product space from column vectors $x_1, x_2, \ldots, x_N$ of X, where $x_i \in R^M$ represents TF_i's standardized expression levels across M cells in cell type. The 2nd term $A = D^T (W \odot V) D$ purely depends on TF-TF PPIN strengths $w > 0$ and can be retrieved from public databases. By $Eq\ (3a)$, $E = \frac{X^T X}{M} + \beta A = \frac{1}{N} \tilde{X}^T \tilde{X}$; we thus transform X and network-prior PPIN data to $\tilde{X}$ embedding data. This yields a new normalized inner-product space $\frac{1}{N} \tilde{X}^T \tilde{X}$ that helps depict and encode an aspect of cell-type TF-TF coordination scores for regulating TG. Since N is a scalar, we use $\left|\tilde{X}^T \tilde{X}\right| \in \mathbb{R}^{N \times N}$. For each $TF_i - TF_j$ pair, we divide $\left|\widetilde{X_i}^T \widetilde{X_j}\right|$, which is proportional to the extent of their potential coordination, by $\left|\left|\widetilde{X_i}\right|\right| \cdot \left|\left|\widetilde{X_j}\right|\right|$ to scale it. $\left|\left|\widetilde{X_i}\right|\right| = \sqrt{\sum_{z=1}^{M} (\widetilde{x_{iz}})^2} > 0$ is the Euclidean norm of TF_i's embedding $\widetilde{X_i}$. This essentially is their cosine similarity (cos) magnitude: $\left|\cos(\widetilde{X_i}, \widetilde{X_j})\right| = \frac{\left|\widetilde{X_i}^T \widetilde{X_j}\right|}{\left|\left|\widetilde{X_i}\right|\right| \cdot \left|\left|\widetilde{X_j}\right|\right|} \leq 1$. To learn coordination scores, we use coefficient-aware-cos metric: $B_{ij}^{(0)} = \left|\cos(\widetilde{X_i}, \widetilde{X_j})\right| \odot C$ for $i \neq j$ and $B_{ii}^{(0)} = 0$. We apply max absolute value scaling $\frac{B_{ij}^{(0)}}{\max(|B|)}$ where $\max(|B|)$ is the max magnitude of $\ell_N = \frac{N(N-1)}{2}$ scores. Our TG-specific TF-TF coordination $B \in \mathbb{R}^{N \times N}$ has: $B_{ij} = \frac{100 B_{ij}^{(0)}}{\max(|B|)}$ where $-100 \leq B_{ij} \leq 100$ if TF_i and TF_j are among N^* TFs (i.e. $c_i^* c_j^* \neq 0$), else 0. Of ℓ_N scores, $\ell_{N^*} = \frac{N^*(N^*-1)}{2}$ are $\neq 0$.

TFs with higher $|B|$ have stronger coordination for co-regulating TG. NetREm predicts B for known TF-TF PPIs (pairs with $w_{ij} > \eta$), uncovering meaningful, cell-type-specific PPI subnetworks. These documented PPIs have

established partnership for orchestrating biological processes. It also predicts B for novel, artificial PPI links ($w_{ij} = \eta$), flagging (high $|B|$) promising FN TF-TF links for follow-up investigation. TF-TF coordination can be direct (e.g. form complexes, tethering) or indirect (e.g. TFs may not interact physically but can modify local chromatin environments, facilitating binding of other TFs) (Srivastava and Mahony 2020). Typically, direct PPIs rely on genome-wide data, indirect (e.g. guilt-by-association) PPIs use genetic interaction data (Wang et al. 2009). While studies predict negatomes (proteins unlikely to interact physically) (Jha et al. 2022), they offer limited insights on indirect PPIs.

Cell-type coordination $\bar{B}$:
We define -100 ≤ $\bar{B}$ ≤ 100 (math details in **Supp. File 5**). Each non-zero score in $\bar{B}$ is unique. $|\overline{B_{ij}}|$ denotes relative strength of $TF_i - TF_j$ coordination and cell-type co-regulation (i.e. normalized interaction significance); the sign encapsulates their net observed behavior across G TGs: $\overline{B_{ij}} < 0$: antagonistic, $\overline{B_{ij}} > 0$: cooperate.

Real-world Datasets & Pre-processing

We apply NetREm for each TG in the cell-type for 7 main applications spanning 2 organisms (humans: 1, 2, 5-7; mouse: 3, 4). We run NetREm without prior GRNs for these 5 applications: **1.** Simulated hESC data. **2.** HSCs: self-renewing, long-lived cells in bone marrow that help produce blood cells. **3.** mESCs: derived from inner cell mass of early embryo; they are pluripotent as they can self-renew, develop, specialize, differentiate, and mature into any cell type. **4.** mDCs: immune cells that capture and present antigens to other immune cells. **5.** PBMCs containing 9 sub cell-types. In these 5 applications, we fix $N = \mathcal{N}$ candidate TFs for all TGs. If the TG is also a TF, we remove it from the set for the TG, so $N = \mathcal{N} - 1$. In applications 1-4, we use gold standards to hone our TGs and TFs so TF-TG regulatory links are comparable with ground truth (McCalla et al. 2023); we train models for TGs and for $\mathcal{N}$ TFs found in both gene expression and ground truth data. For PBMCs, we train NetREm on 9 cell types with 1,029 TFs (Lambert et al. 2018). We run NetREm on 2 applications using prior GRNs from multi-omics to define a custom set of N highly probable, candidate TFs for each TG (N differs across TGs): **5.** mSCs vs. nmSCs in PNS. **6.** AD and Control stages for 8 CNS cell-types. 4 Glial cells: Astrocytes, Oligo, OPCs, Mic; 2 Neurons: GABA-ergic InNs and Glutamatergic Excitatory Neurons (ExNs); 2 Vascular and Blood-Brain Barrier (BBB) cells: Pericytes, Endothelial (Endo.) BBB cells. **Table S16-17** breaks down the # of TGs, # of TFs if N is fixed (applications 1 – 5) or metrics if variable (applications 6 – 7). For instance, in hESCs, we run NetREm 1,250 times (1 for each TG), with N = 206-207 TFs for each TG. **Supp. File 5** details parameters and evaluation for our applications.

Single-cell gene expression data

Supp. File 5 provides preprocessing details for our 7 main applications and additional datasets. **1:** We randomly select 1,250 TGs and corresponding TFs from weighted and signed (+: activates; -: represses) ground truth GRN atlas from TF induction analysis (Sharov et al. 2022). This results in $\mathcal{N}$ = 207 TFs and 5,050 GRN links we input to SERGIO to simulate realistic single-cell data for 100 and 1k cells and 1,442 genes. We vary noise (30, 60, 90)%, retrieving 3 different synthetic expression datasets. **2:** We use (Buenrostro et al. 2018). **3:** We use (Tran et. al 2019) that reprograms mouse embryonic fibroblasts to embryonic-like induced pluripotent stem cells. **4:** We use normalized data (McCalla et al. 2023) from (Shalek et. al 2014) for >1.7k primary bone marrow DCs. **5:** We use public healthy donor data on 2.7k PBMCs based on (Satija et. al. 2024). **6:** We use (Avraham et al. 2022) for mSCs and nmSCs in DRG L4,5 regions from 5 donors. **7:** We use processed (Gupta et al. 2022) for 24 AD and 24 healthy humans for 8 cell types based on 80,660 droplet-based single-nucleus prefrontal cortex transcriptomes (Mathys et. al 2019).

Prior GRN reference info on TG regulation

In applications 6-7, we employ multi-omics and epigenetic data from open scATAC-seq chromatin regions mapped to TGs (i.e. peak-TG links) to identify potential interacting REs for TGs. We map sequence-specific TF motifs to REs, using Position Weight Matrix (PWM) databases to predict TFBSs, forming a motif-based GRN: direct TF-RE-TG candidate links. We prune this motif-based TF list based on relative TF expression (proxy for protein abundance) and motif matching scores. To capture overlooked TFs, we augment our pruned TF list by adding TFs with known PPIs and predictions of TF-TF colocalization, complexes, and/or Molecular Function similarity (Wu et al. 2021). This addresses limitations of GRN-inference tools that rely solely on accessible motif matches and may miss causal TFs for TGs (Zhang et al. 2023); as ~10% of the ~1.6k human TFs lack motif data and are traditionally excluded, using PPIs during GRN inference is recommended to incorporate these missing TFs (Badia-i-Mompel et al. 2023). Augmented TFs may bind to TFBSs directly (weak signals) or indirectly (associate with DNA-binding TFs) (Gordân et al. 2009; Sloan et al. 2016). Our final prior GRNs comprise these initial TF-RE-TG links. We input the N biologically-promising TG-specific candidate TFs to NetREm for the given TG. These subjective steps are detailed in **Supp. File 5**.

Organism-specific PPINs

We employ public STRINGdb (Szklarczyk et al. 2023) to construct human and mouse PPINs (details: **Supp. File 5**; we add more resources for humans) for proteins that partner through direct (physical binding, complex coexistence) and/or indirect (e.g., metabolic/signaling paths) PPIs (Oughtred et al. 2021). We scale average combined scores (many

evidence types) to assign weights: $0.01 < w \leq 1$. Any self-loops are ignored. For each TG in a context, we filter this partly-connected PPIN $\mathbb{W}$ to only keep its N TFs to yield $W^{(0)}$. We run NetREm, assigning an artificial weight $\eta = 0.01$ to missing edges so $0.01 \leq w \leq 1$ for $i \neq j$ to ensure numerical stability and propel discovery of novel TF-TF links; this yields our TG-specific final comprehensive fully-connected TF-TF PPIN (input W).

Data access

We implement NetREm as an open-source software package: GitHub.com/SaniyaKhullar/NetREm with details in **Supp. File 5**. **Table S18** lists resources. We use human hg38 and rat rn5 reference genomes.

Competing interest statement

None declared.

Acknowledgements

D.W. conceived and designed the study. S.K., X.H., and D.W. developed the methodology for NetREm. In particular, X.H. primarily formulated the mathematical models in Steps 1 and 2, added the adaptive pseudo-inverses adjustment, and verified the linear algebra used in the method. X.H. and S.K. provided mathematical justification for Steps 1 and 2. S.K. and X.H. designed methodology for TF-TF coordination scores (from network regression embeddings) and TF-TG regulatory networks and designed simulation studies. S.K., X.H., R.R., J.S., and D.W. contributed to writing the manuscript. S.K. performed benchmarking, computational experiments, bioinformatics analysis, and created all supplementary materials. S.K. and X.H. wrote open-source code for NetREm. R.R. and J.S. conducted experiments in rats, providing external data and analysis for Schwann cell validation. S.K., J.S., and D.W. analyzed the data. All authors have read and approved the final manuscript and are grateful to receive funding from these sources: National Institutes of Health, R01AG067025, RF1MH128695, R03NS123969, R21NS127432, R21NS128761 to D.W. and P50HD105353 to Waisman Center. S.K. is supported by an NLM training grant to Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359). Authors are grateful for feedback and advice from Kalpana Hanthanan Arachchilage, Ph.D., Jerome Choi, Noah Cohen Kalafut, Sean Chang, and Harrison Pantera. In addition, the authors appreciate feedback from Renu Poochie Khullar and Armaan Khullar for the manuscript.

References

Acquaah-Mensah GK, Agu N, Khan T, Gardner A. 2015. A Regulatory Role for the Insulin- and BDNF-Linked RORA in the Hippocampus: Implications for Alzheimer's Disease. *J Alzheimers Dis* **44**: 827–838.

Aibar et. al 2017. SCENIC: single-cell regulatory network inference and clustering | Nature Methods. https://www.nature.com/articles/nmeth.4463 (Accessed March 27, 2023).

Avraham O, Chamessian A, Feng R, Yang L, Halevi AE, Moore AM, Gereau RWI, Cavalli V. 2022. Profiling the molecular signature of satellite glial cells at the single cell level reveals high similarities between rodents and humans. *PAIN* **163**: 2348.

Badia-i-Mompel P, Wessels L, Müller-Dott S, Trimbour R, Ramirez Flores RO, Argelaguet R, Saez-Rodriguez J. 2023. Gene regulatory network inference in the era of single-cell multi-omics. *Nat Rev Genet* **24**: 739–754.

Balakrishnan A, Belfiore L, Chu T-H, Fleming T, Midha R, Biernaskie J, Schuurmans C. 2021. Insights Into the Role and Potential of Schwann Cells for Peripheral Nerve Repair From Studies of Development and Injury. *Front Mol Neurosci* **13**: 608442.

Bentsen et. al 2022. TF-COMB – Discovering grammar of transcription factor binding sites - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2001037022003051 (Accessed March 27, 2023).

Berenson A, Lane R, Soto-Ugaldi LF, Patel M, Ciausu C, Li Z, Chen Y, Shah S, Santoso C, Liu X, et al. 2023. Paired yeast one-hybrid assays to detect DNA-binding cooperativity and antagonism across transcription factors. *Nat Commun* **14**: 6570.

Bossaerts L, Cacace R, Van Broeckhoven C. 2022. The role of ATP-binding cassette subfamily A in the etiology of Alzheimer's disease. *Mol Neurodegener* **17**: 31.

Boyle P, Després C. 2010. Dual-function transcription factors and their entourage. *Plant Signal Behav* **5**: 629–634.

Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, Majeti R, Chang HY, Greenleaf WJ. 2018. Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173**: 1535-1548.e16.

Campos LM de, Cano A, Castellano JG, Moral S. 2019. Combining gene expression data and prior knowledge for inferring gene regulatory networks via Bayesian networks using structural restrictions. *Stat Appl Genet Mol Biol* **18**. https://www.degruyter.com/document/doi/10.1515/sagmb-2018-0042/html (Accessed November 17, 2023).

Chagas VS, Groeneveld CS, Oliveira KG, Trefflich S, de Almeida RC, Ponder BAJ, Meyer KB, Jones SJM, Robertson AG, Castro MAA. 2019. RTNduals: an R/Bioconductor package for analysis of co-regulation and inference of dual regulons. *Bioinformatics* **35**: 5357–5358.

Cheng F, Zhao J, Wang Y, Lu W, Liu Z, Zhou Y, Martin WR, Wang R, Huang J, Hao T, et al. 2021. Comprehensive characterization of protein-protein interactions perturbed by disease mutations. *Nat Genet* **53**: 342–353.

Choi E, Zhang X, Xing C, Yu H. 2016. Mitotic Checkpoint Regulators Control Insulin Signaling and Metabolic Homeostasis. *Cell* **166**: 567.

Choy CT, Wong CH, Chan SL. 2019. Embedding of Genes Using Cancer Gene Expression Data: Biological Relevance and Potential Application on Biomarker Discovery. *Front Genet* **9**. https://www.frontiersin.org/articles/10.3389/fgene.2018.00682 (Accessed January 2, 2024).

Chu X, Guan B, Dai L, Liu J, Li F, Shang J. 2023. Network embedding framework for driver gene discovery by combining functional and structural information. *BMC Genomics* **24**: 426.

Cuomo ASE, Nathan A, Raychaudhuri S, MacArthur DG, Powell JE. 2023. Single-cell genomics meets human genetics. *Nat Rev Genet* **24**: 535–549.

D'Alessio AC, Fan ZP, Wert KJ, Baranov P, Cohen MA, Saini JS, Cohick E, Charniga C, Dadon D, Hannett NM, et al. 2015. A Systematic Approach to Identify Candidate Transcription Factors that Control Cell Identity. *Stem Cell Rep* **5**: 763–775.

De Las Rivas J, Fontanillo C. 2012. Protein–protein interaction networks: unraveling the wiring of molecular machines within the cell. *Brief Funct Genomics* **11**: 489–496.

Dibaeinia P, Sinha S. 2020. SERGIO: A Single-Cell Expression Simulator Guided by Gene Regulatory Networks. *Cell Syst* **11**: 252-271.e11.

Dirmeier S, Fuchs C, Mueller NS, Theis FJ. 2018. netReg: network-regularized linear models for biological association studies. *Bioinformatics* **34**: 896–898.

Eraslan et. al 2022. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function | Science. https://www.science.org/doi/10.1126/science.abl4290 (Accessed April 5, 2024).

Escorcia-Rodríguez JM, Gaytan-Nuñez E, Hernandez-Benitez EM, Zorro-Aranda A, Tello-Palencia MA, Freyre-González JA. 2023. Improving gene regulatory network inference and assessment: The importance of using network structure. *Front Genet* **14**. https://www.frontiersin.org/articles/10.3389/fgene.2023.1143382 (Accessed November 24, 2023).

Eyigoz E, Mathur S, Santamaria M, Cecchi G, Naylor M. 2020. Linguistic markers predict onset of Alzheimer's disease. *eClinicalMedicine* **28**. https://www.thelancet.com/journals/eclinm/article/PIIS2589-53702030327-8/fulltext (Accessed September 6, 2023).

Fang L-P, Zhao N, Caudal LC, Chang H-F, Zhao R, Lin C-H, Hainz N, Meier C, Bettler B, Huang W, et al. 2022. Impaired bidirectional communication between interneurons and oligodendrocyte precursor cells affects social cognitive behavior. *Nat Commun* **13**: 1394.

Federico A, Monti S. 2020. Contextualized Protein-Protein Interactions. *Patterns* **2**: 100153.

Fröb F, Wegner M. 2021. Coordination of Schwann cell myelination and node formation at the transcriptional level. *Neural Regen Res* **17**: 1269–1270.

Gerber D, Pereira JA, Gerber J, Tan G, Dimitrieva S, Yángüez E, Suter U. 2021. Transcriptional profiling of mouse peripheral nerves to the single-cell level to build a sciatic nerve ATlas (SNAT) eds. D.E. Bergles, M.E. Bronner, L.A. Goff, and A. Hoke. *eLife* **10**: e58591.

Ghanbari M, Lasserre J, Vingron M. 2015. Reconstruction of gene networks using prior knowledge. *BMC Syst Biol* **9**: 84.

Gharavi E, Gu A, Zheng G, Smith JP, Cho HJ, Zhang A, Brown DE, Sheffield NC. 2021. Embeddings of genomic region sets capture rich biological associations in lower dimensions. *Bioinformatics* **37**: 4299–4306.

Glasgow SM, Zhu W, Stolt CC, Huang T-W, Chen F, LoTurco JJ, Neul JL, Wegner M, Mohila C, Deneen B. 2014. Mutual antagonism between Sox10 and NFIA regulates diversification of glial lineages and glioma subtypes. *Nat Neurosci* **17**: 1322–1329.

Gonzalez-Teran B, Pittman M, Felix F, Thomas R, Richmond-Buccola D, Hüttenhain R, Choudhary K, Moroni E, Costa MW, Huang Y, et al. 2022. Transcription factor protein interactomes reveal genetic determinants in heart disease. *Cell* **185**: 794-814.e30.

Göös H, Kinnunen M, Salokas K, Tan Z, Liu X, Yadav L, Zhang Q, Wei G-H, Varjosalo M. 2022. Human transcription factor protein interaction networks. *Nat Commun* **13**: 766.

Gordân R, Hartemink AJ, Bulyk ML. 2009. Distinguishing direct versus indirect transcription factor–DNA interactions. *Genome Res* **19**: 2090–2100.

Gu et. al 2022. GEO Accession viewer. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6466265 (Accessed April 11, 2024).

Guo Y, Gifford DK. 2017. Modular combinatorial binding among human trans-acting factors reveals direct and indirect factor binding. *BMC Genomics* **18**: 45.

Gupta C, Xu J, Jin T, Khullar S, Liu X, Alatkar S, Cheng F, Wang D. 2022. Single-cell network biology characterizes cell type gene regulation for drug repurposing and phenotype prediction in Alzheimer's disease. *PLOS Comput Biol* **18**: e1010287.

Hannenhalli S, Levy S. 2002. Predicting transcription factor synergism. *Nucleic Acids Res* **30**: 4278–4284.

Hsu Y-HH, Nacu E, Liu R, Pintacuda G, Kim A, Tsafou K, Petrossian N, Crotty W, Suh JM, Riseman J, et al. 2022. Using brain cell-type-specific protein interactomes to interpret genetic data in schizophrenia. 2021.10.07.21264568. https://www.medrxiv.org/content/10.1101/2021.10.07.21264568v2 (Accessed August 22, 2023).

Hu S, Metcalf E, Mahat DB, Chan L, Sohal N, Chakraborty M, Hamilton M, Singh A, Singh A, Lees JA, et al. 2022. Transcription factor antagonism regulates heterogeneity in embryonic stem cell states. *Mol Cell* **82**: 4410-4427.e12.

Hung HA, Sun G, Keles S, Svaren J. 2015. Dynamic Regulation of Schwann Cell Enhancers after Peripheral Nerve Injury *. *J Biol Chem* **290**: 6937–6950.

Ibarra et. al 2020. Mechanistic insights into transcription factor cooperativity and its impact on protein-phenotype interactions | Nature Communications. https://www.nature.com/articles/s41467-019-13888-7 (Accessed March 27, 2023).

Iglesias-Martinez LF, Kolch W, Santra T. 2016. BGRMI: A method for inferring gene regulatory networks from time-course gene expression data and its application in breast cancer research. *Sci Rep* **6**: 37140.

Imoto S, Higuchi T, Goto T, Tashiro K, Kuhara S, Miyano S. 2003. Combining microarrays and biological knowledge for estimating gene networks via Bayesian networks. *Proc IEEE Comput Soc Bioinforma Conf* **2**: 104–113.

Ishii A, Furusho M, Bansal R. 2021. Mek/ERK1/2-MAPK and PI3K/Akt/mTOR signaling plays both independent and cooperative roles in Schwann cell differentiation, myelination and dysmyelination. *Glia* **69**: 2429–2446.

Jha K, Saha S, Singh H. 2022. Prediction of protein–protein interaction using graph neural networks. *Sci Rep* **12**: 8360.

Jia L, Li F, Wei C, Zhu M, Qu Q, Qin W, Tang Y, Shen L, Wang Y, Shen L, et al. 2020. Prediction of Alzheimer's disease using multi-variants from a Chinese genome-wide association study. *Brain* **144**: 924–937.

Jin T, Rehani P, Ying M, Huang J, Liu S, Roussos P, Wang D. 2021. scGRNom: a computational pipeline of integrative multi-omics analyses for predicting cell-type disease genes and regulatory networks. *Genome Med* **13**: 95.

Johnson KL, Qi Z, Yan Z, Wen X, Nguyen TC, Zaleta-Rivera K, Chen C-J, Fan X, Sriram K, Wan X, et al. 2021. Revealing protein-protein interactions at the transcriptome scale by sequencing. *Mol Cell* **81**: 4091-4103.e9.

Jones EA, Jang S-W, Mager GM, Chang L-W, Srinivasan R, Gokey NG, Ward RM, Nagarajan R, Svaren J. 2007. Interactions of Sox10 and Egr2 in Myelin Gene Regulation. *Neuron Glia Biol* **3**: 377–387.

Joukov V, De Nicolo A, Rodriguez A, Walter JC, Livingston DM. 2010. Centrosomal protein of 192 kDa (Cep192) promotes centrosome-driven spindle assembly by engaging in organelle-specific Aurora A activation. *Proc Natl Acad Sci* **107**: 21022–21027.

Ju H, Yun H, Kim Y, Nam YJ, Lee S, Lee J, Jeong SM, Heo J, Kwon H, Cho YS, et al. 2023. Activating transcription factor-2 supports the antioxidant capacity and ability of human mesenchymal stem cells to prevent asthmatic airway inflammation. *Exp Mol Med* **55**: 413–425.

Karczewski KJ, Tatonetti NP, Landt SG, Yang X, Slifer T, Altman RB, Snyder M. 2011. Cooperative transcription factor associations discovered using regulatory variation. *Proc Natl Acad Sci* **108**: 13353–13358.

Khullar S, Wang D. 2023. Predicting brain-regional gene regulatory networks from multi-omics for Alzheimer's disease phenotypes and Covid-19 severity. *Hum Mol Genet* **32**: 1797–1813.

Kim D, Tran A, Kim HJ, Lin Y, Yang JYH, Yang P. 2023. Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data. *Npj Syst Biol Appl* **9**: 1–13.

Kim S, Pan W, Shen X. 2013. Network-based penalized regression with application to genomic data. *Biometrics* **69**: 582–593.

Kong Y, Yu T. 2018. A graph-embedded deep feedforward network for disease outcome classification and feature selection using gene expression data. *Bioinformatics* **34**: 3727–3737.

Kotlyar et. al 2022. IID 2021: towards context-specific protein interaction analyses by increased coverage, enhanced annotation and enrichment analysis | Nucleic Acids Research | Oxford Academic. https://academic.oup.com/nar/article/50/D1/D640/6424757 (Accessed July 14, 2023).

Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, Duong TE, Gao D, Chun J, Kharchenko PV, et al. 2018. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* **36**: 70–80.

Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. 2018. The Human Transcription Factors. *Cell* **172**: 650–665.

LeBlanc et. al 2006. Full article: Neuropathy-Associated Egr2 Mutants Disrupt Cooperative Activation of Myelin Protein Zero by Egr2 and Sox10. https://www.tandfonline.com/doi/full/10.1128/MCB.01689-06 (Accessed November 2, 2023).

Lee B-K, Bhinge AA, Battenhouse A, McDaniell RM, Liu Z, Song L, Ni Y, Birney E, Lieb JD, Furey TS, et al. 2012. Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. *Genome Res* **22**: 9–24.

Li and Li 2008. Network-constrained regularization and variable selection for analysis of genomic data | Bioinformatics | Oxford Academic. https://academic.oup.com/bioinformatics/article/24/9/1175/206444 (Accessed March 27, 2023).

Li C, Li H. 2010. VARIABLE SELECTION AND REGRESSION ANALYSIS FOR GRAPH-STRUCTURED COVARIATES WITH AN APPLICATION TO GENOMICS. *Ann Appl Stat* **4**: 1498–1516.

Li L, Liu Z-P. 2022. A connected network-regularized logistic regression model for feature selection. *Appl Intell* **52**: 11672–11702.

Li Y, Jackson SA. 2015. Gene Network Reconstruction by Integration of Prior Biological Knowledge. *G3 GenesGenomesGenetics* **5**: 1075–1079.

Lopez-Anido C, Poitelon Y, Gopinath C, Moran JJ, Ma KH, Law WD, Antonellis A, Feltri ML, Svaren J. 2016. Tead1 regulates the expression of Peripheral Myelin Protein 22 during Schwann cell development. *Hum Mol Genet* **25**: 3055–3069.

Lopez-Anido et. al 2015. GEO Accession viewer. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64703 (Accessed July 14, 2023).

Ma KH, Svaren J. 2018. Epigenetic Control of Schwann Cells. *The Neuroscientist* **24**: 627–638.

Mao Y, Fisher DW, Yang S, Keszycki RM, Dong H. 2020. Protein-protein interactions underlying the behavioral and psychological symptoms of dementia (BPSD) and Alzheimer's disease. *PLOS ONE* **15**: e0226021.

Martin V, Zhuang F, Zhang Y, Pinheiro K, Gordân R. 2023. High-throughput data and modeling reveal insights into the mechanisms of cooperative DNA-binding by transcription factor proteins. *Nucleic Acids Res* **51**: 11600–11612.

Martini R, Schachner M. 1986. Immunoelectron microscopic localization of neural cell adhesion molecules (L1, N-CAM, and MAG) and their shared carbohydrate epitope and myelin basic protein in developing sciatic nerve. *J Cell Biol* **103**: 2439–2448.

Mathew B, Bathla S, Williams KR, Nairn AC. 2022. Deciphering Spatial Protein–Protein Interactions in Brain Using Proximity Labeling. *Mol Cell Proteomics* **21**. https://www.mcponline.org/article/S1535-9476(22)00230-4/abstract (Accessed August 23, 2023).

Mathys et. al 2019. Single-cell transcriptomic analysis of Alzheimer's disease | Nature. https://www.nature.com/articles/s41586-019-1195-2 (Accessed March 27, 2023).

Mayran A, Drouin J. 2018. Pioneer transcription factors shape the epigenetic landscape. *J Biol Chem* **293**: 13795–13804.

McCalla SG, Fotuhi Siahpirani A, Li J, Pyne S, Stone M, Periyasamy V, Shin J, Roy S. 2023. Identifying strengths and weaknesses of methods for computational network inference from single-cell RNA-seq data. *G3 GenesGenomesGenetics* **13**: jkad004.

Mirsaeidi M, Gidfar S, Vu A, Schraufnagel D. 2016. Annexins family: insights into their functions and potential role in pathogenesis of sarcoidosis. *J Transl Med* **14**: 89.

Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J, Aerts S. 2019. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**: 2159–2161.

Mohammadi et. al 2019. Reconstruction of Cell-type-Specific Interactomes at Single-Cell Resolution: Cell Systems. https://www.cell.com/cell-systems/fulltext/S2405-4712(19)30383-7?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2405471219303837%3Fshowall%3Dtrue (Accessed March 27, 2023).

Mukherjee S, Speed TP. 2008. Network inference using informative priors. *Proc Natl Acad Sci* **105**: 14313–14318.

Murtaza N, Cheng AA, Brown CO, Meka DP, Hong S, Uy JA, El-Hajjar J, Pipko N, Unda BK, Schwanke B, et al. 2022. Neuron-specific protein network mapping of autism risk genes identifies shared biological mechanisms and disease-relevant pathologies. *Cell Rep* **41**: 111678.

Nagamine N, Kawada Y, Sakakibara Y. 2005. Identifying cooperative transcriptional regulations using protein–protein interactions. *Nucleic Acids Res* **33**: 4828–4837.

938 Nicodemus KK, Malley JD. 2009. Predictor correlation impacts machine learning algorithms: implications for
939 genomic studies. *Bioinformatics* **25**: 1884–1890.

940 Nie J, Stewart R, Zhang H, Thomson JA, Ruan F, Cui X, Wei H. 2011. TF-Cluster: A pipeline for identifying
941 functionally coordinated transcription factors via network decomposition of the shared coexpression
942 connectivity matrix (SCCM). *BMC Syst Biol* **5**: 53.

943 Nie Y, Shu C, Sun X. 2020. Cooperative binding of transcription factors in the human genome. *Genomics* **112**: 3427–
944 3434.

945 Oughtred R, Rust J, Chang C, Breitkreutz B, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F, et al. 2021.
946 The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical
947 interactions. *Protein Sci Publ Protein Soc* **30**: 187–200.

948 Ouyang Z, Zhou Q, Wong WH. 2009. ChIP-Seq of transcription factors predicts absolute and differential gene
949 expression in embryonic stem cells. *Proc Natl Acad Sci* **106**: 21521–21526.

950 Padi M, Quackenbush J. 2015. Integrating transcriptional and protein interaction networks to prioritize condition-
951 specific master regulators. *BMC Syst Biol* **9**: 80.

952 Parab L, Pal S, Dhar R. 2022. Transcription factor binding process is the primary driver of noise in gene expression.
953 *PLOS Genet* **18**: e1010535.

954 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg
955 V, et al. 2011. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12**: 2825–2830.

956 Peng X, Wang J, Peng W, Wu F-X, Pan Y. 2017. Protein–protein interactions: detection, reliability assessment and
957 applications. *Brief Bioinform* **18**: 798–819.

958 Perna S, Pinoli P, Ceri S, Wong L. 2020. NAUTICA: classifying transcription factor interactions by positional and
959 protein-protein interaction information. *Biol Direct* **15**: 13.

960 Poitelon et. al 2016. YAP and TAZ control peripheral myelination and the expression of laminin receptors in Schwann
961 cells | Nature Neuroscience. https://www.nature.com/articles/nn.4316 (Accessed September 7, 2023).

962 Pratapa et. al 2020. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic
963 data | Nature Methods. https://www.nature.com/articles/s41592-019-0690-6 (Accessed March 27, 2023).

964 Rao S, Ahmad K, Ramachandran S. 2021. Cooperative binding between distant transcription factors is a hallmark of
965 active enhancers. *Mol Cell* **81**: 1651-1665.e4.

966 Roussel J, Larcher R, Sicard P, Bideaux P, Richard S, Marmigère F, Thireau J. 2022. The autism-associated Meis2
967 gene is necessary for cardiac baroreflex regulation in mice. *Sci Rep* **12**: 20150.

968 Rustenhoven J, Smith AM, Smyth LC, Jansson D, Scotter EL, Swanson MEV, Aalderink M, Coppieters N, Narayan P,
969 Handley R, et al. 2018. PU.1 regulates Alzheimer's disease-associated genes in primary human microglia. *Mol*
970 *Neurodegener* **13**: 44.

971 Satija et. al. 2024. Analysis, visualization, and integration of spatial datasets with Seurat.
972 https://satijalab.org/seurat/articles/spatial_vignette (Accessed February 4, 2024).

973 Sevimoglu T, Arga KY. 2014. The role of protein interaction networks in systems biomedicine. *Comput Struct*
974 *Biotechnol J* **11**: 22–27.

975 Shalek et. al 2014. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation | Nature.
976 https://www.nature.com/articles/nature13437 (Accessed April 4, 2024).

977 Sharov AA, Nakatake Y, Wang W. 2022. Atlas of regulated target genes of transcription factors (ART-TF) in human
978 ES cells. *BMC Bioinformatics* **23**: 377.

979 Shojaie A, Michailidis G. 2009. Analysis of Gene Sets Based on the Underlying Regulatory Network. *J Comput Biol*
980 **16**: 407–426.

981 Singh KV, Vig L. 2017. Improved prediction of missing protein interactome links via anomaly detection. *Appl Netw*
982 *Sci* **2**: 2.

983 Sinha KK, Bilokapic S, Du Y, Malik D, Halic M. 2023. Histone modifications regulate pioneer transcription factor
984 cooperativity. *Nature* **619**: 378–384.

985 Skok Gibbs C, Jackson CA, Saldi G-A, Tjärnberg A, Shah A, Watters A, De Veaux N, Tchourine K, Yi R, Hamamsy
986 T, et al. 2022. High-performance single-cell gene regulatory network inference at scale: the Inferelator 3.0.
987 *Bioinformatics* **38**: 2519–2528.

988 Sloan CA, Chan ET, Davidson JM, Malladi VS, Strattan JS, Hitz BC, Gabdank I, Narayanan AK, Ho M, Lee BT, et al.
989 2016. ENCODE data at the ENCODE portal. *Nucleic Acids Res* **44**: D726–D732.

990 Smith BN, Topp SD, Fallini C, Shibata H, Chen H-J, Troakes C, King A, Ticozzi N, Kenna KP, Soragia-Gkazi A, et
991 al. 2017. Mutations in the vesicular trafficking protein annexin A11 are associated with amyotrophic lateral
992 sclerosis. *Sci Transl Med* **9**: eaad9157.

993 Sönmezer C, Kleinendorst R, Imanci D, Barzaghi G, Villacorta L, Schübeler D, Benes V, Molina N, Krebs AR. 2021.
994 Molecular co-occupancy identifies transcription factor binding cooperativity in vivo. *Mol Cell* **81**: 255-267.e6.

995 Spitz F, Furlong EEM. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*
996 **13**: 613–626.

997 Srinivasan R, Sun G, Keles S, Jones EA, Jang S-W, Krueger C, Moran JJ, Svaren J. 2012. Genome-wide analysis of
998 EGR2/SOX10 binding in myelinating peripheral nerve. *Nucleic Acids Res* **40**: 6449–6460.

999 Srivastava D, Mahony S. 2020. Sequence and chromatin determinants of transcription factor binding and the
1000 establishment of cell type-specific binding patterns. *Biochim Biophys Acta Gene Regul Mech* **1863**: 194443.

1001 Stevanovic M, Drakulic D, Lazic A, Ninkovic DS, Schwirtlich M, Mojsin M. 2021. SOX Transcription Factors as
1002 Important Regulators of Neuronal and Glial Differentiation During Nervous System Development and Adult
1003 Neurogenesis. *Front Mol Neurosci* **14**: 654031.

1004 Stratton JA, Kumar R, Sinha S, Shah P, Stykel M, Shapira Y, Midha R, Biernaskie J. 2017. Purification and
1005 Characterization of Schwann Cells from Adult Human Skin and Nerve. *eNeuro* **4**.
1006 https://www.eneuro.org/content/4/3/ENEURO.0307-16.2017 (Accessed October 18, 2023).

1007 Stuart et. al 2021. Single-cell chromatin state analysis with Signac | Nature Methods.
1008 https://www.nature.com/articles/s41592-021-01282-5 (Accessed March 27, 2023).

1009 Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo
1010 S, et al. 2023. The STRING database in 2023: protein–protein association networks and functional enrichment
1011 analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**: D638–D646.

1012 Tao et. al 2019. Variation in SIPA1L2 is correlated with phenotype modification in Charcot– Marie– Tooth disease
1013 type 1A - Tao - 2019 - Annals of Neurology - Wiley Online Library.
1014 https://onlinelibrary.wiley.com/doi/10.1002/ana.25426 (Accessed July 14, 2023).

1015 THE GTEX CONSORTIUM. 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues.
1016 *Science* **369**: 1318–1330.

1017 Titsias MK, Honkela A, Lawrence ND, Rattray M. 2012. Identifying targets of multiple co-regulating transcription
1018 factors from expression time-series by Bayesian model comparison. *BMC Syst Biol* **6**: 53.

1019 Tran et. al 2019. Defining Reprogramming Checkpoints from Single-Cell Analyses of Induced Pluripotency -
1020 ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2211124719305297 (Accessed March 9,
1021 2024).

1022 Turley et al. Executive Summary | Pan UKBB. https://pan-dev.ukbb.broadinstitute.org/docs/summary (Accessed April
1023 10, 2024).

1024 Vargas DM, De Bastiani MA, Zimmer ER, Klamt F. 2018. Alzheimer's disease master regulators analysis: search for
1025 potential molecular targets and drug repositioning candidates. *Alzheimers Res Ther* **10**: 59.

1026 Wallace C. 2020. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLOS*
1027 *Genet* **16**: e1008720.

1028 Wang et. al 2018. Comprehensive functional genomic resource and integrative model for the human brain | Science.
1029 https://www.science.org/doi/10.1126/science.aat8464?url_ver=Z39.88-
1030 2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed (Accessed March 27, 2023).

1031 Wang H, Huang B, Wang J. 2021. Predict long-range enhancer regulation based on protein–protein interactions
1032 between transcription factors. *Nucleic Acids Res* **49**: 10347–10368.

1033 Wang J, Liu Q, Sun J, Shyr Y. 2016. Disrupted cooperation between transcription factors across diverse cancer types.
1034 *BMC Genomics* **17**: 560.

1035 Wang Q, Guo M, Chen J, Duan R. 2023. A gene regulatory network inference model based on pseudo-siamese
1036 network. *BMC Bioinformatics* **24**: 163.

1037 Wang Y, Duan X, Zhou X, Wang R, Zhang X, Cao Z, Wang X, Zhou Z, Sun Y, Peng D. 2022. ANXA11 mutations are
1038 associated with amyotrophic lateral sclerosis–frontotemporal dementia. *Front Neurol* **13**: 886887.

1039 Wang Y, Hu X, Jiang X, He T, Yuan J. 2015. Predicting microbial interactions by using network-constrained
1040 regularization incorporating covariate coefficients and connection signs. In *2015 IEEE International*
1041 *Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 635–638.

1042 Wang Y, Shan Q, Pan J, Yi S. 2018. Actin Cytoskeleton Affects Schwann Cell Migration and Peripheral Nerve
1043 Regeneration. *Front Physiol* **9**: 23.

1044 Wang Y, Zhang X-S, Xia Y. 2009. Predicting eukaryotic transcriptional cooperativity by Bayesian network integration
1045 of genome-wide data. *Nucleic Acids Res* **37**: 5943–5958.

1046 Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. 2021. clusterProfiler 4.0: A
1047 universal enrichment tool for interpreting omics data. *The Innovation* **2**. https://www.cell.com/the-
1048 innovation/abstract/S2666-6758(21)00066-7 (Accessed August 25, 2023).

1049 Wunderlich Z, Mirny LA. 2009. Different gene regulation strategies revealed by analysis of binding motifs. *Trends*
1050 *Genet* **25**: 434–440.

1051 Yang P, Li X, Chua H-N, Kwoh C-K, Ng S-K. 2014. Ensemble Positive Unlabeled Learning for Disease Gene
1052 Identification. *PLOS ONE* **9**: e97079.

1053 Yim et. al 2022. Disentangling glial diversity in peripheral nerves at single-nuclei resolution | Nature Neuroscience.
1054 https://www.nature.com/articles/s41593-021-01005-1 (Accessed October 24, 2023).

1055 Yu D, Chojnowski G, Rosenthal M, Kosinski J. 2023. AlphaPulldown—a python package for protein–protein
1056 interaction screens using AlphaFold-Multimer. *Bioinformatics* **39**: btac749.

1057 Yu G, Wang L-G, He Q-Y. 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and
1058 visualization. *Bioinformatics* **31**: 2382–2383.

1059 Zaborowski AB, Walther D. 2020. Determinants of correlated expression of transcription factors and their target
1060 genes. *Nucleic Acids Res* **48**: 11347–11369.

1061 Zhang K, Hocker JD, Miller M, Hou X, Chiou J, Poirion OB, Qiu Y, Li YE, Gaulton KJ, Wang A, et al. 2021. A
1062 single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**: 5985-6001.e19.

1063 Zhang S, Pyne S, Pietrzak S, Halberg S, McCalla SG, Siahpirani AF, Sridharan R, Roy S. 2023. Inference of cell type-
1064 specific gene regulatory networks on cell lineages from single cell omic datasets. *Nat Commun* **14**: 3064.

1065 Zhang W, Wan Y, Allen GI, Pang K, Anderson ML, Liu Z. 2013. Molecular pathway identification using biological
1066 network-regularized logistic models. *BMC Genomics* **14**: S7.

1067 Zhao Y. 2023. TFSyntax: a database of transcription factors binding syntax in mammalian genomes. *Nucleic Acids Res*
1068 **51**: D306–D314.

1069 Zou H, Hastie T. 2005. Regularization and Variable Selection Via the Elastic Net. *J R Stat Soc Ser B Stat Methodol* **67**:
1070 301–320.

1071 An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders | Nature
1072 Genetics. https://www.nature.com/articles/s41588-018-0130-z (Accessed November 10, 2023a).

1073 Know Thy Friends, know ThyselF – Transcription Factor (TF) regulation resolved | Springer Nature Protocols and
1074 Methods Community. https://protocolsmethods.springernature.com/posts/know-thy-friends-know-thyself-
1075 transcription-factor-tf-regulation-resolved (Accessed March 27, 2023b).

1076 Noncoding RNA and Gene Expression | Learn Science at Scitable.
1077 https://www.nature.com/scitable/topicpage/regulation-of-transcription-and-gene-expression-in-1086/
1078 (Accessed November 17, 2023c).

1079 Systematic Functional Annotation of Somatic Mutations in Cancer: Cancer Cell. https://www.cell.com/cancer-
1080 cell/fulltext/S1535-6108(18)30021-7 (Accessed November 10, 2023d).

1081 Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders: Cell.
1082 https://www.cell.com/cell/fulltext/S0092-8674(15)00430-
1083 4?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867415004304%3Fs
1084 howall%3Dtrue (Accessed November 10, 2023e).

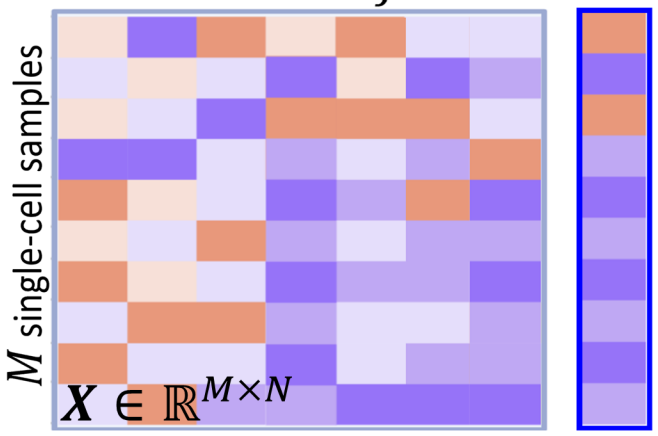
1085

Integration of multimodal data and networks

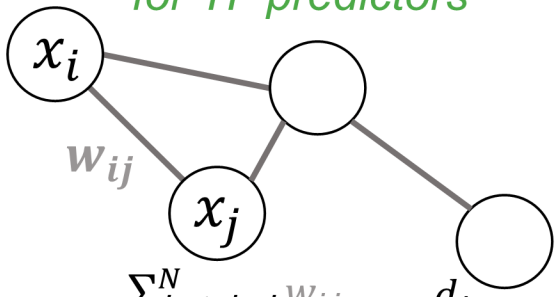
Single-cell gene expression data

Transcription Factors (TFs) TG

$x_1 \quad \cdots \quad x_i \quad x_j \cdots x_N \qquad y$



Protein-Protein Interaction (PPI) network for TF predictors

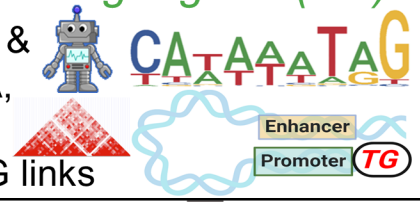


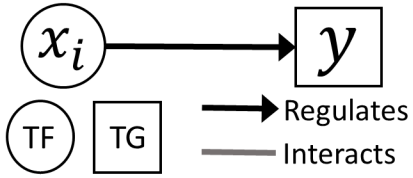
$$w_{ii} = \frac{\sum_{j=1,\, i \neq j}^{N} w_{ij}}{N-1} = \frac{d_i}{N-1}$$

d_i: node degree of x_i in PPI network

Prior knowledge on target gene (TG) regulation (optional):

e.g. TF motif bindings & colocalization on DNA, chromatin interaction, regulatory element-TG links





Step 1: Network regularized regression definition

$$\boldsymbol{c}^* = \underset{\boldsymbol{c}}{argmin} \left\{ \frac{1}{2M} \left|\left| \boldsymbol{y} - \boldsymbol{X}\boldsymbol{c} \right|\right|^2 + \alpha \left|\left| \boldsymbol{c} \right|\right|_1 + \frac{\beta}{2} \sum_{i,j=1,\ldots,N} \left(\left(\frac{c_i}{\sqrt{d_i}} - \frac{c_j}{\sqrt{d_j}} \right)^2 (w_{ij}) \right) \right\}$$

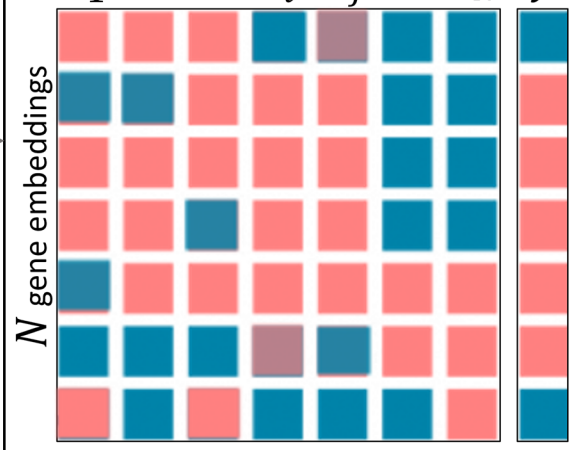
$\boldsymbol{c}^* \in \mathbb{R}^N$: regression coefficients of N candidate TFs for predicting TG expression levels $\boldsymbol{y}$ for TF-TG regulation

Hyperparameters: β: network-constrained prior α: sparsity-prior

Step 2: Gene expression embeddings from network regularized regression

$$\boldsymbol{c}^* = \underset{\boldsymbol{c}}{argmin} \left\{ \frac{1}{2N} \left|\left| \tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{c} \right|\right|^2 + \alpha \left|\left| \boldsymbol{c} \right|\right|_1 \right\}$$

$\tilde{x}_1 \quad \cdots \quad \tilde{x}_i \quad \tilde{x}_j \quad \cdots \quad \tilde{x}_N \qquad \tilde{y}$



where

$$\tilde{\boldsymbol{X}} = \sqrt{N}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{U}^T \in \mathbb{R}^{N \times N}$$

$$\tilde{\boldsymbol{y}} = \frac{\sqrt{N}}{M}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{U}^T\boldsymbol{X}^T\boldsymbol{y} \in \mathbb{R}^N$$

Singular Value Decomposition (SVD) of the $\boldsymbol{E} \in \mathbb{R}^{N \times N}$ matrix:

$$\boldsymbol{E} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^T = \frac{\boldsymbol{X}^T\boldsymbol{X}}{M} + \beta \boldsymbol{D}^T(\boldsymbol{W} \odot \boldsymbol{V})\boldsymbol{D}$$

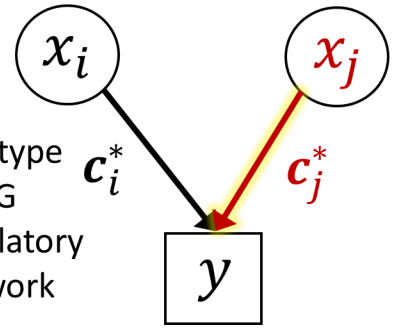
$\boldsymbol{V} = N \cdot \boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^T$

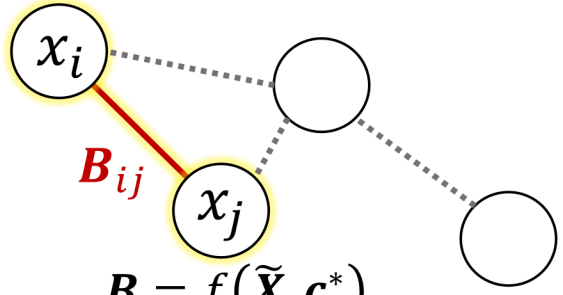
$\boldsymbol{D}$: diagonal matrix of node degrees d

$\boldsymbol{W}$: matrix of PPI network edge weights $w > 0$

Cell-type-specific TF coordination for TG regulation

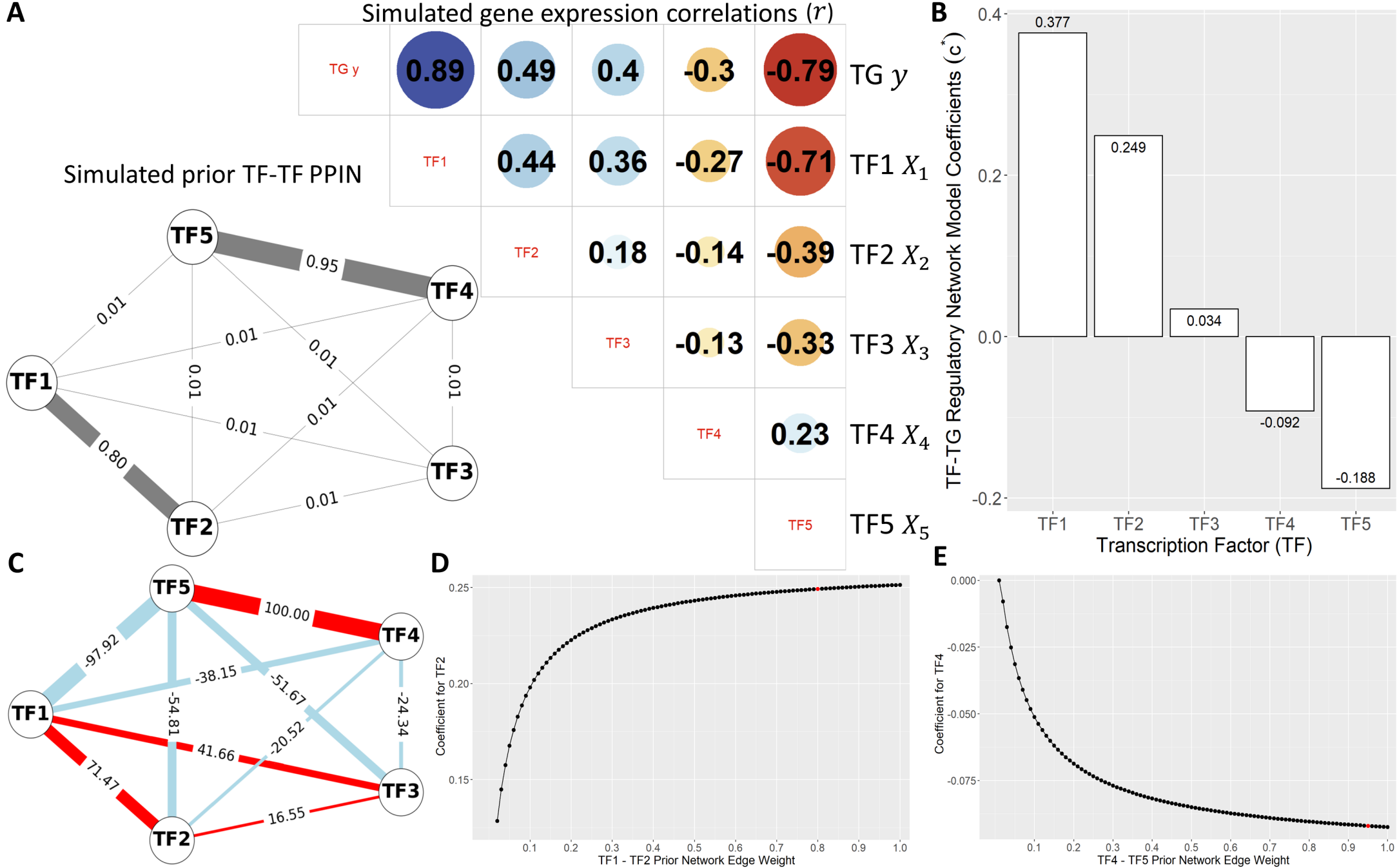
Cell-type TF-TG regulatory network

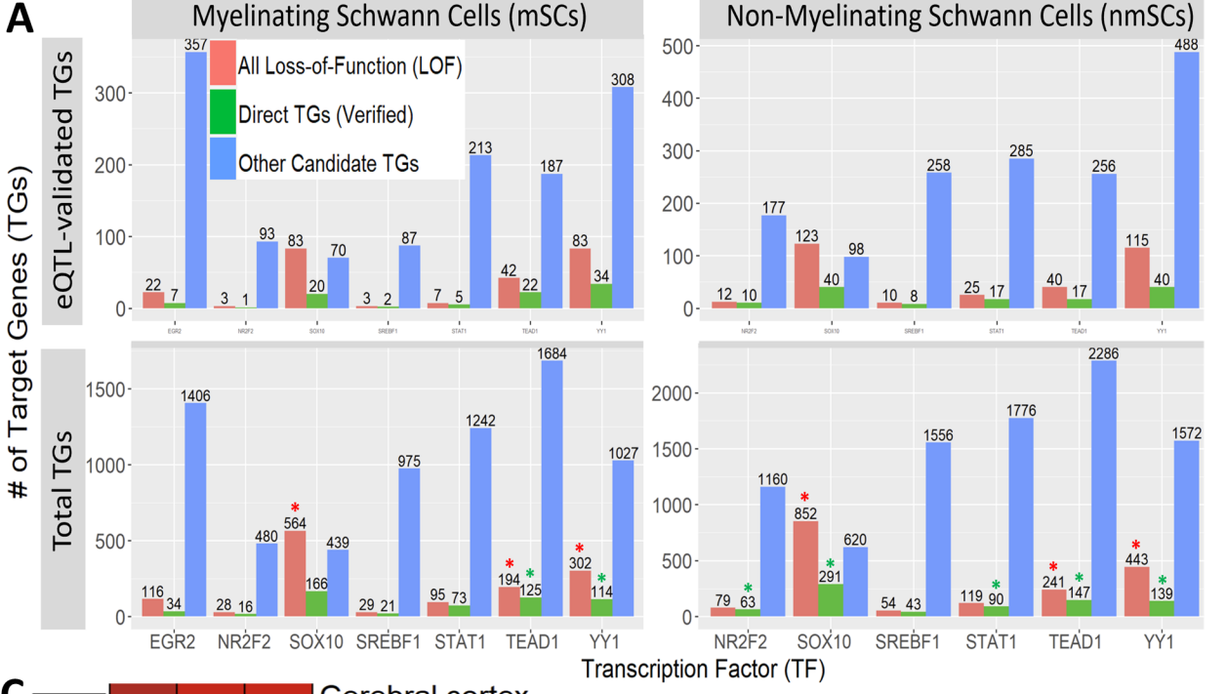


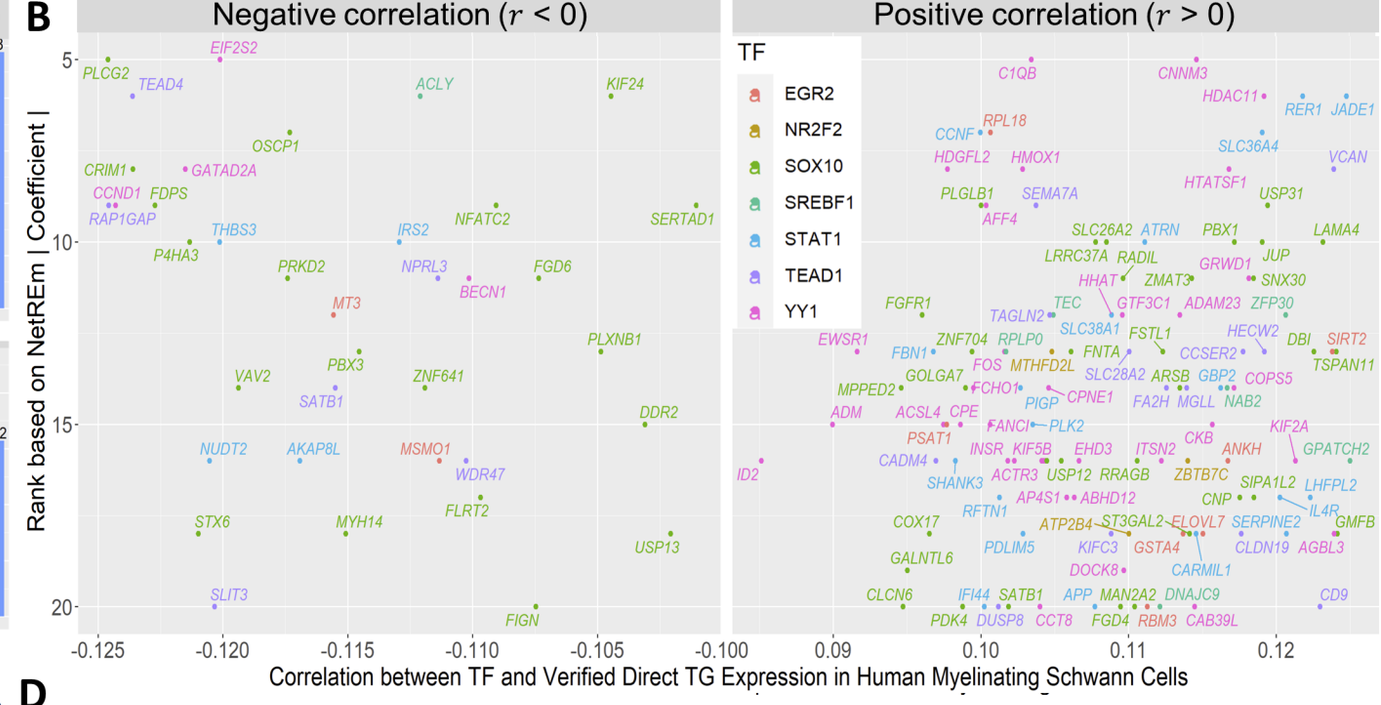


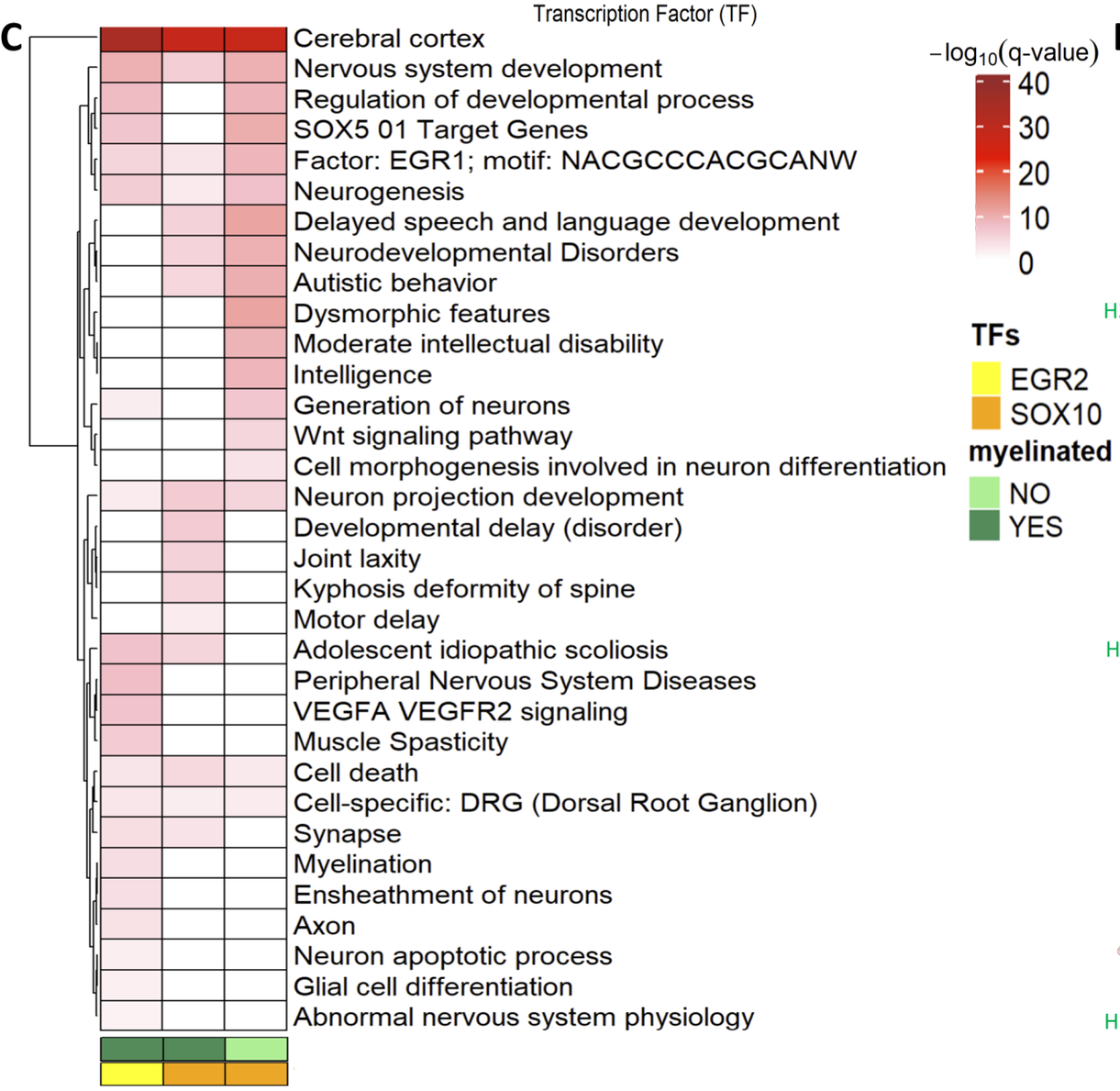
$$\boldsymbol{B} = f(\tilde{\boldsymbol{X}}, \boldsymbol{c}^*)$$

1. New cell-type TFs *2. TG-specific TF-TF Coordination*

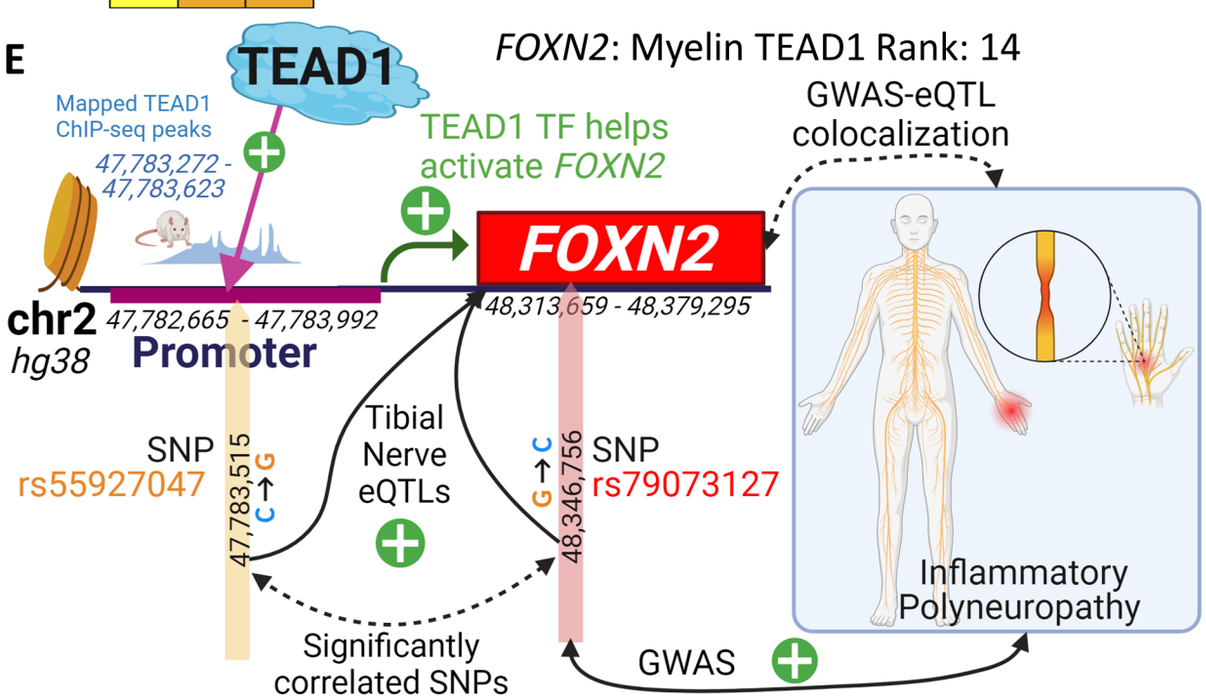

A
Simulated gene expression correlations (r)
TG y
0.89
0.49
0.4
-0.3
-0.79
TG y
TF1
0.44
0.36
-0.27
-0.71
TF1 X_1
TF2
0.18
-0.14
-0.39
TF2 X_2
TF3
-0.13
-0.33
TF3 X_3
TF4
0.23
TF4 X_4
TF5
TF5 X_5
Simulated prior TF-TF PPIN
TF5
TF4
TF1
TF3
TF2
0.95
0.80
0.01
B
TF-TG Regulatory Network Model Coefficients (c^*)
0.377
0.249
0.034
-0.092
-0.188
Transcription Factor (TF)
C
100.00
-97.92
-38.15
-51.67
-54.81
-24.34
41.66
-20.52
71.47
16.55
D
Coefficient for TF2
TF1 - TF2 Prior Network Edge Weight
E
Coefficient for TF4
TF4 - TF5 Prior Network Edge Weight

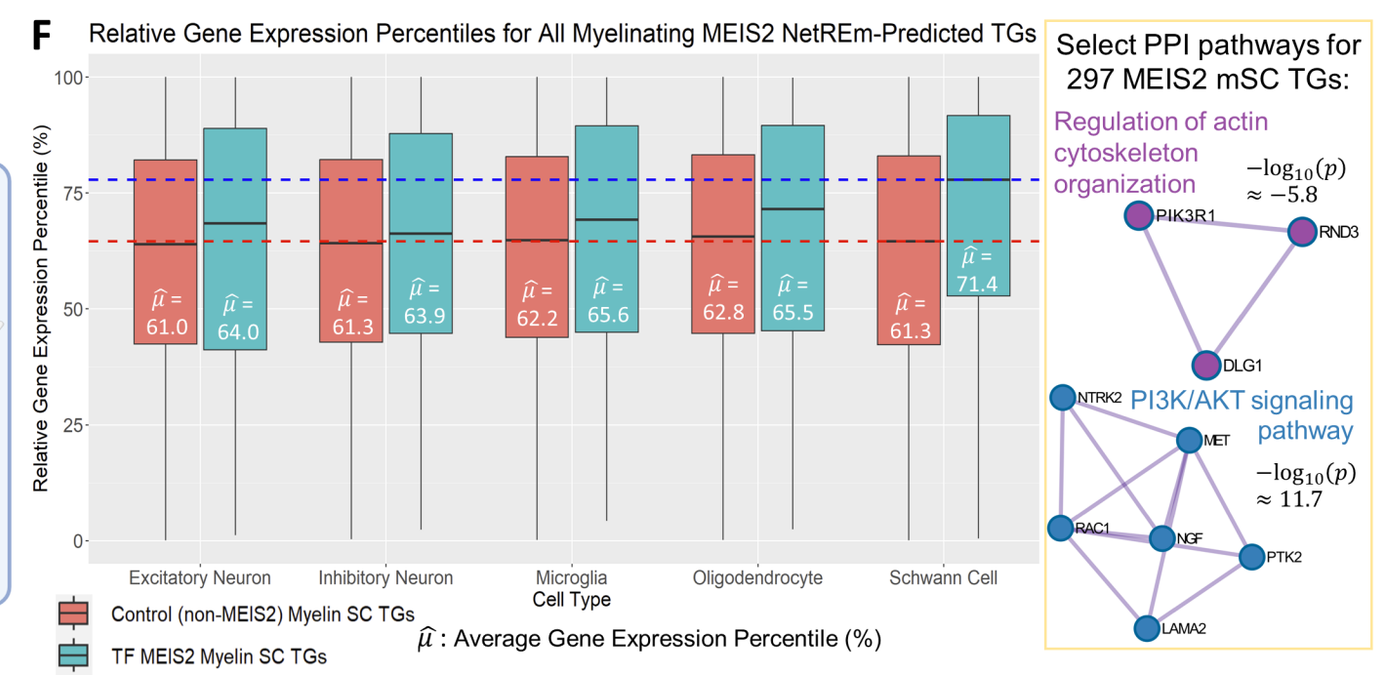

A
Myelinating Schwann Cells (mSCs)
Non-Myelinating Schwann Cells (nmSCs)
All Loss-of-Function (LOF)
Direct TGs (Verified)
Other Candidate TGs
eQTL-validated TGs
Total TGs
# of Target Genes (TGs)
EGR2
NR2F2
SOX10
SREBF1
STAT1
TEAD1
YY1
Transcription Factor (TF)

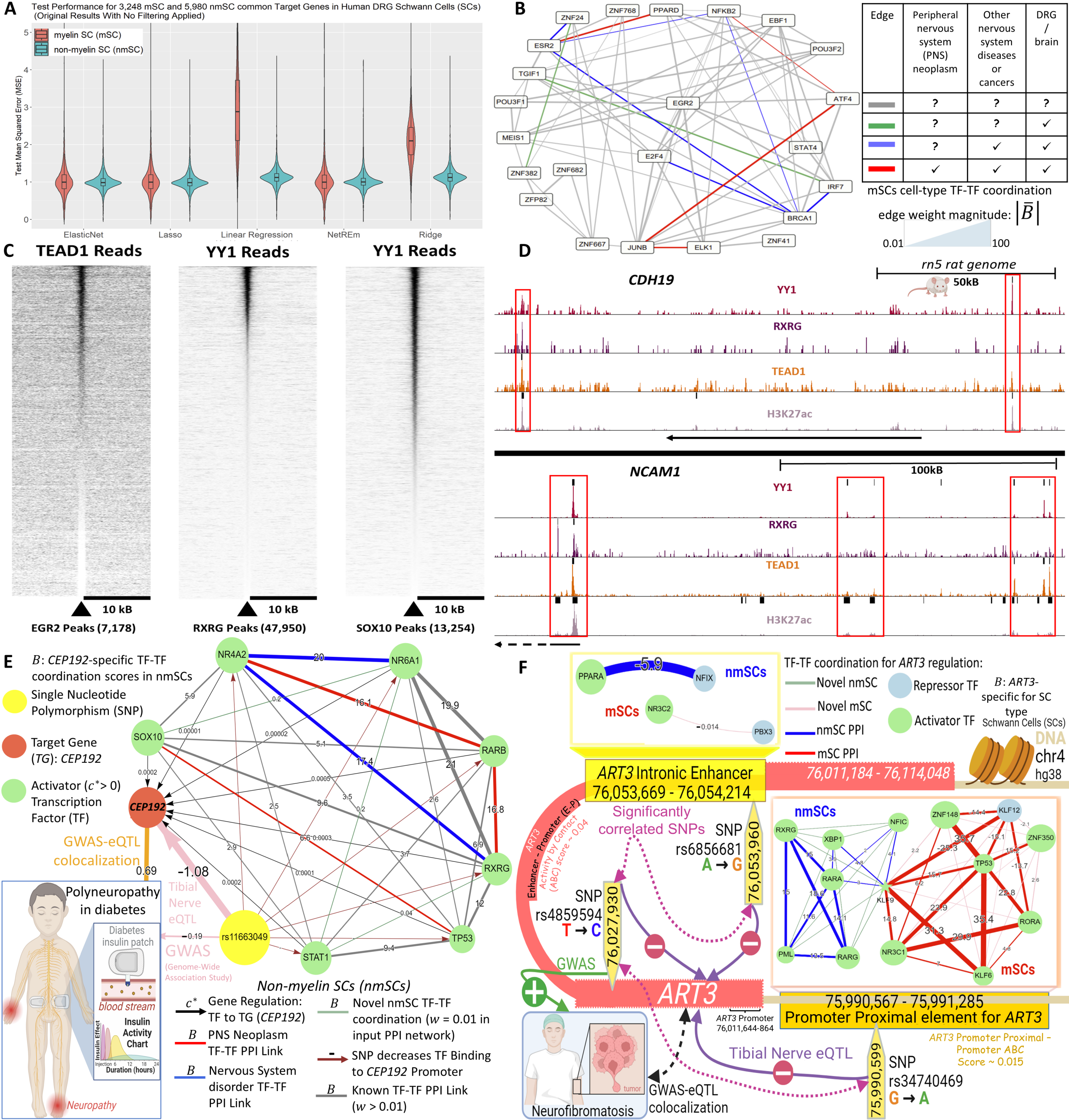

B
Negative correlation (r < 0)
Positive correlation (r > 0)
TF
EGR2
NR2F2
SOX10
SREBF1
STAT1
TEAD1
YY1
Rank based on NetREm | Coefficient |
Correlation between TF and Verified Direct TG Expression in Human Myelinating Schwann Cells

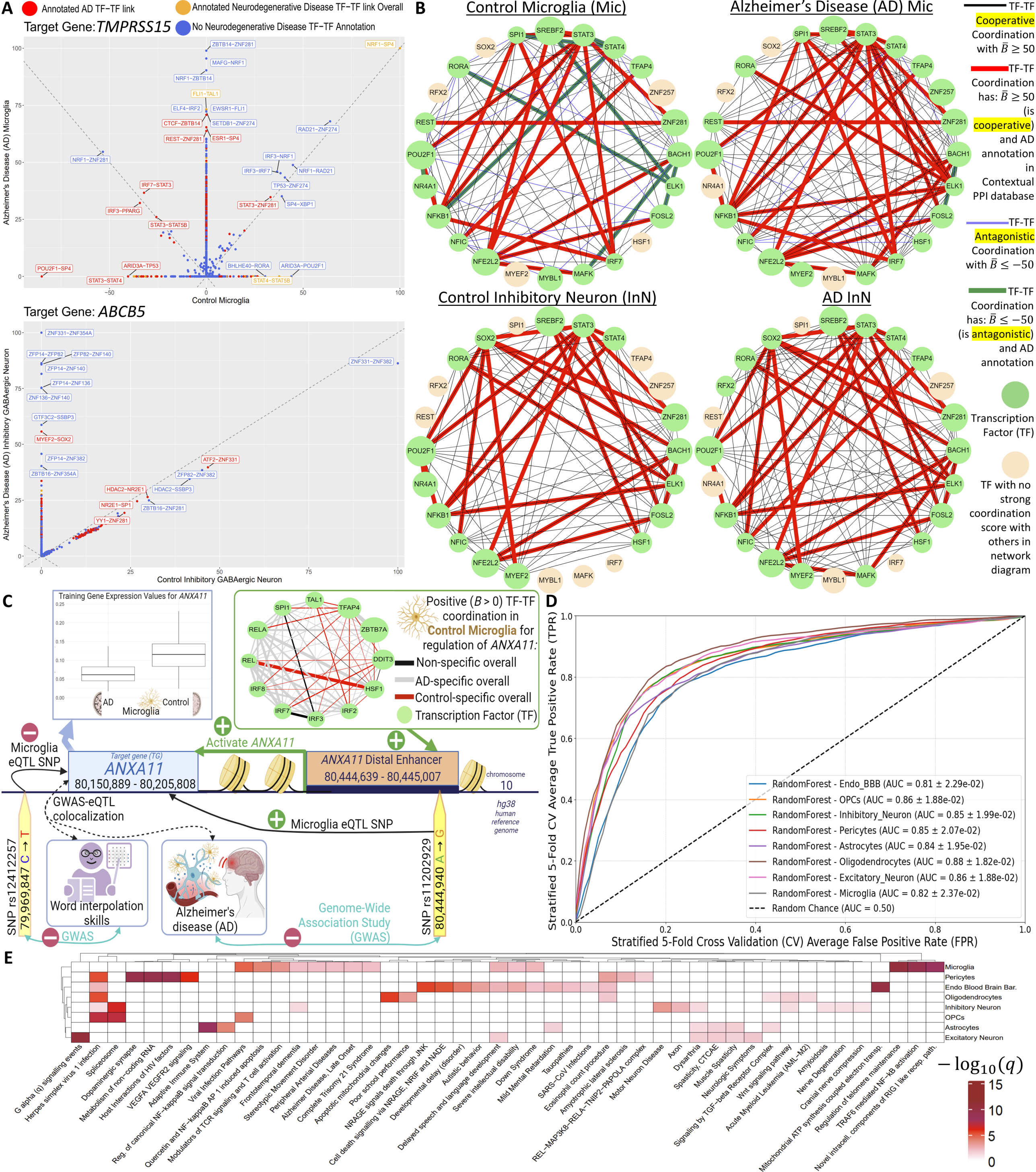

C
Cerebral cortex
Nervous system development
Regulation of developmental process
SOX5 01 Target Genes
Factor: EGR1; motif: NACGCCCACGCANW
Neurogenesis
Delayed speech and language development
Neurodevelopmental Disorders
Autistic behavior
Dysmorphic features
Moderate intellectual disability
Intelligence
Generation of neurons
Wnt signaling pathway
Cell morphogenesis involved in neuron differentiation
Neuron projection development
Developmental delay (disorder)
Joint laxity
Kyphosis deformity of spine
Motor delay
Adolescent idiopathic scoliosis
Peripheral Nervous System Diseases
VEGFA VEGFR2 signaling
Muscle Spasticity
Cell death
Cell-specific: DRG (Dorsal Root Ganglion)
Synapse
Myelination
Ensheathment of neurons
Axon
Neuron apoptotic process
Glial cell differentiation
Abnormal nervous system physiology
−log10(q-value)
TFs
EGR2
SOX10
myelinated
NO
YES

D
GPR157: Myelin SOX10 Rank: 2
Predicted to bind to GPR157 regulatory elements (REs) (e.g. promoter in humans (hg38): chr1:9,539,105-9,540,245)
SOX10 Sciatic Nerve (SN) Reads: Run 204
H3K27Ac Immunoprecipitation Sorted Tag Directory
CEP95: Myelin SOX10 Rank: 2
Predicted to bind to CEP95 REs (e.g. promoter in humans: chr17:64,506,140-64,507,288)
SCAMP5: Myelin SOX10 Rank: 12
Predicted to bind to SCAMP5 REs (e.g. enhancer in humans: chr15: 75,029,296-75,030,505)


E
FOXN2: Myelin TEAD1 Rank: 14
TEAD1
Mapped TEAD1 ChIP-seq peaks
47,783,272 - 47,783,623
TEAD1 TF helps activate FOXN2
GWAS-eQTL colocalization
FOXN2
chr2 hg38
47,782,665 - 47,783,992
Promoter
48,313,659 - 48,379,295
SNP rs55927047
47,783,515
Tibial Nerve eQTLs
SNP rs79073127
48,346,756
Inflammatory Polyneuropathy
Significantly correlated SNPs
GWAS


F
Relative Gene Expression Percentiles for All Myelinating MEIS2 NetREm-Predicted TGs
Relative Gene Expression Percentile (%)
Excitatory Neuron
Inhibitory Neuron
Microglia
Oligodendrocyte
Schwann Cell
Cell Type
Control (non-MEIS2) Myelin SC TGs
TF MEIS2 Myelin SC TGs
Select PPI pathways for 297 MEIS2 mSC TGs:
Regulation of actin cytoskeleton organization
PI3K/AKT signaling pathway


A
Test Performance for 3,248 mSC and 5,980 nmSC common Target Genes in Human DRG Schwann Cells (SCs)
(Original Results With No Filtering Applied)
myelin SC (mSC)
non-myelin SC (nmSC)
Test Mean Squared Error (MSE)
ElasticNet
Lasso
Linear Regression
NetREm
Ridge
B
ZNF24
ZNF768
PPARD
NFKB2
EBF1
ESR2
POU3F2
TGIF1
POU3F1
EGR2
ATF4
MEIS1
STAT4
E2F4
ZNF382
ZNF682
IRF7
ZFP82
BRCA1
ZNF667
JUNB
ELK1
ZNF41
mSCs cell-type TF-TF coordination
edge weight magnitude: |B̄|
0.01
100
Edge
Peripheral nervous system (PNS) neoplasm
Other nervous system diseases or cancers
DRG / brain
C
TEAD1 Reads
YY1 Reads
YY1 Reads
10 kB
10 kB
10 kB
EGR2 Peaks (7,178)
RXRG Peaks (47,950)
SOX10 Peaks (13,254)
D
rn5 rat genome
CDH19
50kB
YY1
RXRG
TEAD1
H3K27ac
NCAM1
100kB
YY1
RXRG
TEAD1
H3K27ac
E
B: CEP192-specific TF-TF coordination scores in nmSCs
Single Nucleotide Polymorphism (SNP)
Target Gene (TG): CEP192
Activator (c*> 0) Transcription Factor (TF)
NR4A2
NR6A1
SOX10
RARB
CEP192
RXRG
TP53
STAT1
rs11663049
GWAS-eQTL colocalization
0.69
-1.08
Tibial Nerve eQTL
-0.19
GWAS (Genome-Wide Association Study)
Polyneuropathy in diabetes
Diabetes insulin patch
blood stream
Insulin Activity Chart
Insulin Effect
Duration (hours)
Neuropathy
Non-myelin SCs (nmSCs)
Gene Regulation: TF to TG (CEP192)
PNS Neoplasm TF-TF PPI Link
Nervous System disorder TF-TF PPI Link
Novel nmSC TF-TF coordination (w = 0.01 in input PPI network)
SNP decreases TF Binding to CEP192 Promoter
Known TF-TF PPI Link (w > 0.01)
F
PPARA
NFIX
nmSCs
mSCs
NR3C2
PBX3
TF-TF coordination for ART3 regulation:
Novel nmSC
Novel mSC
nmSC PPI
mSC PPI
Repressor TF
Activator TF
B: ART3-specific for SC type
Schwann Cells (SCs)
DNA
chr4
hg38
ART3 Intronic Enhancer
76,053,669 - 76,054,214
76,011,184 - 76,114,048
Significantly correlated SNPs
SNP rs6856681 A → G
76,053,960
SNP rs4859594 T → C
76,027,930
Enhancer - Promoter (E-P) Activity by Contact (ABC) score ~ 0.04
GWAS
ART3
ART3 Promoter 76,011,644-864
nmSCs
mSCs
75,990,567 - 75,991,285
Promoter Proximal element for ART3
ART3 Promoter Proximal - Promoter ABC Score ~ 0.015
Tibial Nerve eQTL
SNP rs34740469 G → A
75,990,599
Neurofibromatosis
tumor
GWAS-eQTL colocalization


A
Annotated AD TF-TF link
Annotated Neurodegenerative Disease TF-TF link Overall
No Neurodegenerative Disease TF-TF Annotation
Target Gene: TMPRSS15
Alzheimer's Disease (AD) Microglia
Control Microglia
Target Gene: ABCB5
Alzheimer's Disease (AD) Inhibitory GABAergic Neuron
Control Inhibitory GABAergic Neuron
B
Control Microglia (Mic)
Alzheimer's Disease (AD) Mic
Control Inhibitory Neuron (InN)
AD InN
TF-TF Cooperative Coordination with $\bar{B} \geq 50$
TF-TF Coordination has: $\bar{B} \geq 50$ (is cooperative) and AD annotation in Contextual PPI database
TF-TF Antagonistic Coordination with $\bar{B} \leq -50$
TF-TF Coordination has: $\bar{B} \leq -50$ (is antagonistic) and AD annotation
Transcription Factor (TF)
TF with no strong coordination score with others in network diagram
C
Training Gene Expression Values for ANXA11
AD
Control
Microglia
Positive ($B > 0$) TF-TF coordination in Control Microglia for regulation of ANXA11:
Non-specific overall
AD-specific overall
Control-specific overall
Transcription Factor (TF)
Microglia eQTL SNP
Activate ANXA11
Target gene (TG)
ANXA11
80,150,889 - 80,205,808
ANXA11 Distal Enhancer
80,444,639 - 80,445,007
chromosome 10
hg38 human reference genome
GWAS-eQTL colocalization
Microglia eQTL SNP
SNP rs12412257 79,969,847 C → T
SNP rs11202929 80,444,940 A → G
Word interpolation skills
Alzheimer's disease (AD)
GWAS
Genome-Wide Association Study (GWAS)
D
Stratified 5-Fold CV Average True Positive Rate (TPR)
Stratified 5-Fold Cross Validation (CV) Average False Positive Rate (FPR)
RandomForest - Endo_BBB (AUC = 0.81 ± 2.29e-02)
RandomForest - OPCs (AUC = 0.86 ± 1.88e-02)
RandomForest - Inhibitory_Neuron (AUC = 0.85 ± 1.99e-02)
RandomForest - Pericytes (AUC = 0.85 ± 2.07e-02)
RandomForest - Astrocytes (AUC = 0.84 ± 1.95e-02)
RandomForest - Oligodendrocytes (AUC = 0.88 ± 1.82e-02)
RandomForest - Excitatory_Neuron (AUC = 0.86 ± 1.88e-02)
RandomForest - Microglia (AUC = 0.82 ± 2.37e-02)
Random Chance (AUC = 0.50)
E
Microglia
Pericytes
Endo Blood Brain Bar.
Oligodendrocytes
Inhibitory Neuron
OPCs
Astrocytes
Excitatory Neuron
$-\log_{10}(q)$