CardamomOT: a mechanistic optimal transport-based framework for gene regulatory network inference, trajectory reconstruction and generative modeling

Author: Yann Maugé[1], Elias Ventre[1].
1 - COMPO, Inria Méditerranée, Cancer Research Center of Marseille, Inserm UMR1068, CNRS UMR7258, UM105, Aix-Marseille Université, SouthROCK Center of Excellence in Pediatric Oncology Research, 13273 Marseille, France.
contact: elias.ventre@inria.fr.

Abstract

A key challenge in inferring gene regulatory networks (GRNs) governing cellular processes such as differentiation and reprogramming from experimental data lies in the impossibility of directly measuring protein dynamics at the single-cell level, which prevents establishing causal relationships between regulator activity and target responses. In earlier work, we introduced CARDAMOM, an algorithm that uses temporal snapshots of scRNA-seq data to calibrate a GRN-driven mechanistic model of gene expression. However, this method had several limitations: it could only rely on the relative ordering of time points rather than their exact labels, imposed restrictive quasi-stationary assumptions on protein dynamics, and depended on multiple hyperparameters.
Here, we present CardamomOT, a new method based on the same mechanistic model that jointly reconstructs the GRN and unobserved protein trajectories from the data within a mechanistic optimal transport framework. By incorporating exact time labels and priors on protein kinetic rates from the literature, and substantially reducing the number of required hyperparameters, our approach addresses these limitations and substantially improves the accuracy and robustness of GRN calibration. We validate our framework on both *in silico* and experimental datasets, demonstrating computational scalability and consistently improved performance over state-of-the-art methods in both GRN and trajectory reconstruction. In particular, CardamomOT accurately recovers velocity fields driving cellular trajectories and unobserved protein levels, alongside reliable GRN structures. We also show that these improvements make the calibrated mechanistic model suitable to be used as a generative model to predict cellular responses to unseen perturbations. To our knowledge, this is among the first methods to explicitly integrate mechanistic GRN inference, trajectory reconstruction, and simulation of realistic datasets into a unified framework for scRNA-seq time series analysis.

Keywords: gene regulation network, mechanistic learning, transcriptional bursting, trajectory inference, optimal transport, generative modeling.

Author summary

Predicting gene regulatory interactions and understanding how cellular trajectories respond to perturbations are central challenges in cell biology and bioinformatics, yet they have long been addressed as separate problems. Only a few recent approaches leverage temporal single-cell RNA sequencing data to jointly infer gene regulatory networks (GRNs) and cellular trajectories, including our previous method CARDAMOM.

In this work, we introduce CardamomOT, a new framework that integrates optimal transport—a mathematical theory that has become widely used in computational biology in recent years—within a mechanistic modeling approach. This allows us to link cells across timepoints while jointly reconstructing the underlying, unobserved protein trajectories that drive gene regulation. By explicitly modeling these latent dynamics, our method substantially improves the robustness and accuracy of GRN inference.

In contrast to many black-box approaches, CardamomOT is grounded in a biologically interpretable mechanistic model of gene expression, incorporating key processes such as transcriptional bursting, protein translation, and degradation. This framework naturally accommodates prior knowledge on RNA and protein kinetic rates available in the literature and remains consistent with the statistical properties of single-cell transcriptomic data.

We demonstrate on both simulated and experimental datasets that CardamomOT accurately infers GRNs and reconstructs cellular trajectories. Moreover, because it provides a fully calibrated generative model, it can be used to predict cellular responses to unseen perturbations by modifying the inferred regulatory interactions.

Introduction

Biological processes such as cellular differentiation and reprogramming are governed by time-dependent changes in gene expression. These dynamic processes can be partially observed at the single-cell level thanks to the development of single-cell RNA sequencing (scRNA-seq) technologies. Recent advances have enabled researchers to collect single-cell gene expression data from cell populations at multiple time points, revealing high and meaningful variability between individual cells [1, 2, 3, 4] that population-based approaches fail to capture due to averaging effects [5, 6].

The dynamics of gene expression in single cells are traditionally attributed to the underlying structure of gene regulatory networks (GRNs), which describe interactions between genes mediated by the production of regulatory proteins (e.g. transcription factors (TFs) and their upstream activators) and RNAs (e.g. miRNA, siRNA, etc.). Consequently, automated inference of GRNs from high-throughput data has become a central task in systems biology [7], driven by continuous improvements in experimental technologies. Methods such as GENIE3 [8] and its extension SCENIC [9] have become standard tools, and have been extended to account for temporal structure when available [10]. More broadly, it is now well established that GRN identifiability can be greatly enhanced by observing cellular trajectories at single-cell resolution [11].

In practice, however, such trajectories remain largely inaccessible at the genome-wide scale. Although emerging approaches (e.g., live-cell transcriptomics or Raman-based techniques [12, 13]) begin to enable longitudinal measurements, they currently lack the throughput, resolution, and temporal depth required for robust inference. Standard scRNA-seq protocols, by contrast, are destructive and preclude repeated measurements of the same cell.

As a result, most datasets consist of independent snapshots of cell populations at successive timepoints, treated as samples from a shared underlying dynamical process. This disconnect between the destructive nature of scRNA-seq and the need for longitudinal tracking poses a major challenge for GRN inference. To address it, optimal transport (OT)-based methods

[14] have been widely used to reconstruct population-level trajectories by linking cells across timepoints [15, 16, 17], effectively approximating cell dynamics under a least-action principle in gene expression space. A wide range of methods have subsequently been developed for building a GRN from such trajectories, from simple differential expression analysis [18] to flow matching [19, 20] and mechanistic neuralODEs [21]. However, most OT-based trajectory inference methods rely on oversimplified assumptions about the geometry of gene expression space and consider a squared-Euclidean OT cost, which amounts to working in a flat space. In the entropically regularized version of OT used in the aforementioned references, this is equivalent to assuming that individual cells follow a stochastic differential equation (SDE) with constant diffusion rate. Trajectory reconstruction then effectively consists in finding the best gradient drift reproducing the observed dynamics under an initial Brownian-motion prior in gene expression space [22], reflecting a lack of mechanistic prior knowledge [23]. Recent approaches propose to go beyond this Brownian prior by introducing new OT-derived dynamical problems [24, 25, 26, 27], but the geometric constraints remain only partially mechanistic, and realistic priors often come at a much higher computational cost (see Appendix B).

A second important challenge arises from the limited availability of single-cell proteomics. Despite notable progress in recent years [28], current technologies can quantify only a subset of proteins, and key regulators such as TFs often remain inaccessible. As a result, statistical modeling typically relies on scRNA data whose synthesis by transcriptional bursts makes it inherently stochastic and highly variable [29, 30], with non-Gaussian distributions [31, 32]. The variable relationship between mRNA and protein levels, due to transcriptional and translational regulation, makes it difficult to functionally interpret mRNA variations to build a GRN, particularly in time-series scRNA-seq datasets where the kinetic rates characterizing mRNA and protein dynamics can significantly impact the reconstructed GRN [33]. This challenge is generally ignored in the aforementioned OT-based methods, which operate in transcript space only and thus rely on a gene-only underlying model, without explicitly accounting for hidden variables such as proteins for which kinetic information is nevertheless available. Bertin et al. constitute a notable exception [21], using deep learning techniques to integrate latent variables such as additional genes and proxies for proteins. However, they do not explicitly model protein dynamics or kinetic rates, and the trajectory inference component still relies on a Brownian-motion prior without mechanistic constraints. To the best of our knowledge, the only class of methods directly addressing this challenge is based on mechanistic modeling of gene expression dynamics [34, 35, 36, 33, 37], all using multidimensional generalizations of the well-known two-state model [38].

In this paper, we present CardamomOT, which extends our previously published method CARDAMOM [33] and addresses both challenges simultaneously within a unified mechanistic optimal transport framework. Compared to CARDAMOM, CardamomOT introduces three key advances: (i) it explicitly reconstructs protein trajectories through an iterative OT-based procedure, abandoning the quasi-stationary approximation on protein dynamics; (ii) it directly incorporates exact time labels and prior knowledge on kinetic rates, increasing robustness by removing the need for time-dependent hyperparameters; and (iii) it produces a fully calibrated generative model of gene expression, enabling downstream analyses that go beyond GRN inference.

More precisely, CardamomOT calibrates the GRN-driven mechanistic model of gene expression introduced in [34, 33] from temporal snapshots of scRNA-seq data by jointly reconstructing hidden protein trajectories and a GRN capable of regenerating the observed dataset (Figure 1). This joint reconstruction is performed through an iterative expectation-maximization-like (EM-like) procedure alternating between protein trajectory inference — via an OT problem with a mechanistic, GRN-driven cost — and GRN calibration by regression on the reconstructed protein values, until convergence (Figure 1D). The method requires prior knowledge about kinetic rates of protein dynamics, can benefit from prior knowledge about kinetic rates of mRNA

dynamics, and can be further constrained by known regulatory interactions.
We validate CardamomOT on multiple tasks: GRN inference accuracy, reconstruction of cellular trajectories and associated velocity fields, data regeneration, and perturbation prediction. A central advantage of this approach is that the calibrated mechanistic model functions as a generative model: once trained, it can simulate new temporal snapshots that faithfully reproduce the observed data and can be used to predict the effect of unseen genetic perturbations directly *in silico* (Figure 1E). In particular, gene knockouts and overexpressions can be simulated by modifying protein levels and basal parameters in the GRN without retraining the model.
Benchmarks and applications are performed on an already published set of simulated datasets [33], an experimental dataset of sympathoadrenal differentiation, preprocessed with pseudotime in [39] from Kameneva et al. [40], and two well-known timestamped experimental datasets of differentiating cells: one characterizing the retinoic acid driven mouse embryonic stem cells differentiation, from Semrau et al. [3], and one representing the reprogramming of fibroblasts to induced pluripotent stem cells (iPSCs), from Schiebinger et al. [15]. We show, among other results, that CardamomOT correctly predicts the effect of the overexpression of *Obox6* and *Zfp42* on iPSC reprogramming efficiency [15] — an outcome that had been verified experimentally — using only temporal scRNA-seq data for inference and no prior knowledge of these factors' function.

Related works. On the GRN inference side, the mechanistic model used throughout this article was introduced together with the first method to infer a GRN from it in [34, 31], and later extended to temporal snapshots in CARDAMOM [33]. CARDAMOM, however, could not exploit experimental time labels beyond their relative ordering, relied on a quasi-stationary approximation for protein dynamics that implicitly constrained kinetic rates, did not reconstruct individual cell trajectories, and required time-dependent regularization hyperparameters that limited its robustness and prevented its use as a generative model.
On the trajectory inference side, several approaches have recently been proposed to go beyond the Brownian-motion prior. Reference Fitting [25] and GRIT [26] jointly infer cellular trajectories and GRNs using entropic OT by approximating the system dynamics as a linear Ornstein–Uhlenbeck process. This approach, however, is limited to linear systems and cannot effectively model non-linear dynamics. Wasserstein Lagrangian Flows (WLF) have been introduced to solve a dynamical OT-like problem by directly incorporating a potential energy constraint in the Lagrangian of the corresponding variational problem [24], but this comes at a high computational cost since all characteristics of the underlying process and its time-varying measure must be parametrized by neural networks. Our approach can be seen as an extension of Reference Fitting and GRIT to the case where the reference process is a non-linear mechanistic process, making it suitable for modeling complex behaviors. There is also a direct link with WLF, as CardamomOT can be interpreted as solving an OT-like variational problem where the Lagrangian is induced by a simplified mechanistic process on a well-chosen latent space. These correspondences are detailed in Appendix B.

1 Methods

1.1 Mechanistic modeling of cellular dynamics

Single-cell dynamics emerge from the stochastic interplay between gene activation, mRNA synthesis, and protein production. We adopt the hybrid two-state gene-expression model embedded in a regulatory network introduced in [34], which provides a parsimonious yet biologically grounded description of this interplay.

Biological model. Each gene $i \in \{1, \dots, n\}$ has a promoter that switches stochastically between an inactive and an active state. When active, transcription fires at rate $s_{0,i}$, producing mRNA molecules subsequently translated into proteins at rate $s_{1,i}$; both species degrade with rates $d_{0,i}$ and $d_{1,i}$, respectively. For an isolated gene, the switching rates are fixed constants $k_{\text{on},i}$ and $k_{\text{off},i}$. To encode a GRN $\theta = (\theta_{ij})$, the rates of gene i are replaced by protein-dependent functions, $k^{\theta}_{\text{on},i}(P)$, so that each entry θ_{ij} captures the sign, direction, and magnitude of the influence of gene j on gene i.

Recent experimental evidence suggests that regulation acts primarily through changes in activation frequency rather than inactivation rates [41, 42]. Accordingly, we take $k_{\text{off},i}$ to be independent of protein levels and focus on the *bursty* regime $k_{\text{on},i} \ll k_{\text{off},i}$, where transcription occurs in short, highly productive bursts of mean size $s_{0,i}/k_{\text{off},i}$ [43].

Following [31, 33], in all applications presented in this article we adopt a sigmoid parameterization for the activation rates

$$k^{\theta}_{\text{on},i}(P) = k_{0,i} + (k_{1,i} - k_{0,i})\, \sigma\left(\beta_i + \sum_{j=1}^{n} \theta_{ij} P_j\right), \qquad \sigma(u) = \left(1 + e^{-u}\right)^{-1}, \tag{1}$$

where $k_{0,i}$ and $k_{1,i}$ are the minimal and maximal burst frequencies of gene i, and β_i is its basal activity. We nevertheless emphasize that our method is directly applicable to a more general class of functions, that are already implemented in our Python package CardamomOT, as discussed in Section 3.

The full stochastic dynamics of mRNA $M_i(t)$ and protein $P_i(t)$ are then

$$M_i(t) \xrightarrow{k^{\theta}_{on,i}(P(t))} M_i(t) + \mathcal{E}\left(\frac{k_{\text{off},i}}{s_{0,i}}\right), \qquad \frac{dM_i}{dt} = -d_{0,i} M_i, \qquad \frac{dP_i}{dt} = s_{1,i} M_i - d_{1,i} P_i. \tag{2}$$

Simplified statistical model. When bursting and mRNA degradation are fast relative to protein dynamics, the conditional distribution of $M_i(t)$ given $P(t)$ reaches quasi-stationarity and is well approximated by a Negative Binomial (NB) distribution [31]:

$$M_i(t) \mid P(t) \sim \text{NB}\left(\frac{k^{\theta}_{\text{on},i}(P(t))}{d_{0,i}}, \frac{k_{\text{off},i}}{s_{0,i}}\right). \tag{3}$$

Assuming further that $k^{\theta}_{\text{on},i}(P)$ is effectively piecewise constant, taking values $\{\mathbf{k}_{z,i}\}_{z \in Z}$ associated with discrete promoter-activity *basins*, the marginal distribution of $M(t)$ at population level becomes an NB mixture:

$$M(t) \sim \sum_{z \in Z} \mu_t(z) \prod_{i=1}^{n} \text{NB}\left(\frac{\mathbf{k}_{z,i}}{d_{0,i}}, \frac{k_{\text{off},i}}{s_{0,i}}\right), \tag{4}$$

where μ_t is the distribution of cells among basins at time t. This mixture structure motivates using the basin labels z as a discrete, mechanistically interpretable representation of cell state.

Deterministic limit for proteins. In the limit where proteins evolve on a much slower timescale than mRNAs, the latter are well approximated by their conditional mean given P, and protein dynamics satisfy

$$\frac{dP}{dt} = s_1 \frac{k^{\theta}_{\text{on}}(P)\, s_0}{k_{\text{off}}\, d_0} - d_1 P(t),$$

where the notation is here vectorial. Setting $s_1 = d_1 d_0 k_{\text{off}}/(k_1 s_0)$ rescales protein levels to the unit interval, yielding

$$\frac{dP}{dt} = v_1(P) := d_1 \left(\frac{k^{\theta}_{\text{on}}(P)}{k_1} - P(t)\right), \tag{5}$$

which we use both for trajectory interpolation (Appendix A) and for the postprocessing calibration of kinetic rates (Section 1.2.3).

This normalization does not alter the qualitative dynamics of the system, but simply corresponds to a rescaling of protein concentrations. Since protein levels are not directly observed, their absolute scale is not identifiable and can be fixed without loss of generality. Importantly, this choice ensures that all protein variables P_i lie on a comparable scale, so that the interaction coefficients θ_{ij} in Eq. (1) have a consistent interpretation in terms of regulatory strength across genes.

We emphasize that in our method presented below, we systematically adopt a *metastable* perspective (i.e., cells spend most of their time near stable attractors and undergo rare stochastic transitions between them), as characterized in [44]: the deterministic model (5) captures intra-basin dynamics, with basins corresponding to attractors of the limiting system, while stochastic fluctuations govern rare transitions between these basins in experimentally relevant regimes.

The dynamics of transitions between basins, combined with (3), define a Hidden Markov Model in which basins follow a continuous-time Markov chain and generate mRNA levels according to the GRN-dependent distribution (3). We provide further details on this perspective in Appendix B, where we refer to this model as the *phenomenological model* (see also [44]).

Average behavior for mRNAs. In contrast to proteins, there is no simple deterministic limit that closes the dynamics directly in mRNA space: the burst frequency k_{on}^{θ} is controlled by proteins, so that M alone does not form a closed system. Nevertheless, given a protein state $P(t)$, we can define an average mRNA velocity based on burst size and frequency:

$$\frac{dM}{dt} = v_0(P) := k_{\text{on}}^{\theta}(P)\,\frac{s_0}{k_{\text{off}}} - d_0\,M(t), \tag{6}$$

which we use for evaluation of trajectory inference in Figure S7.

We emphasize that this effective field is in general highly history-dependent, because the promoter activation regimes are driven by proteins that evolve on a slower timescale and encode the underlying regulatory logic. Any velocity model formulated directly on mRNAs alone must therefore implicitly integrate over these hidden protein states, and thus over the mechanistic structure of the GRN, leading to an effective regulation that is no longer explicitly mechanistic [21]. CardamomOT keeps the dependence on P explicit and reconstructs protein trajectories alongside the GRN, so that both protein and mRNA dynamics remain grounded in the same mechanistic model.

1.2 Overview of CardamomOT

CardamomOT calibrates the mechanistic model above as a *generative model* for scRNA-seq time series. Starting from raw count matrices at successive timepoints $t_1 < \cdots < t_N$, it jointly infers the GRN θ^* and hidden protein trajectories $\{P_c(t)\}_{t=t_1,\ldots,t_N}$ for each cell c, then uses the calibrated model to simulate new data or predict perturbation effects. The method requires prior knowledge of protein degradation rates d_1, and can optionally incorporate mRNA degradation rates d_0 and known regulatory interactions θ^0.

1.2.1 Preprocessing: inferring the NB mixture and initializing basin labels

We first calibrate the NB mixture model (4) gene by gene across timepoints, following [33]. For each gene i and timepoint t_j, we fit $\text{NB}(a_{j,i}, b_i)$ with b_i shared across time, and set

$$\frac{k_{0,i}}{d_{0,i}} = \min_j a_{j,i}, \quad \frac{k_{1,i}}{d_{0,i}} = \max_j a_{j,i}, \quad \frac{k_{\text{off},i}}{s_{0,i}} = b_i.$$

This yields, for each cell c, a probability matrix Pr_c of size $|Z| \times n$ encoding the likelihood of belonging to each basin. Cells are initialized by maximum-likelihood assignment:

$$z_c^0 := \arg\max_z \prod_i \mathrm{Pr}_c(z, i). \tag{7}$$

If a prior GRN θ^0 is available, this assignment can already incorporate network information via the GRN-constrained update described in Step 3 below.

1.2.2 Iterative inference of GRN and protein trajectories

The core novelty of CardamomOT is to abandon the quasi-stationary approximation $P_c \approx k_{\mathrm{on}}^{\theta}(P_c)/k_1$ used in CARDAMOM, and instead reconstruct protein trajectories $\{P_c(t)\}_{t=t_1,\ldots,t_N}$ through explicit trajectory inference. Since protein dynamics are driven by the GRN, trajectory inference and GRN reconstruction are tightly coupled: the GRN shapes the interpolation cost, and the reconstructed trajectories in turn constrain the GRN. We therefore alternate between the two in an EM-like loop.

Step 1: Protein trajectory reconstruction via mechanistic OT. Given a current GRN θ and basin labels $\{z_c\}$, we seek, for each cell c_{t_j} at time t_j, its most likely descendant $c_{t_{j+1}}$ at t_{j+1} (i.e., the cell at the next timepoint that most plausibly derives from the same lineage). For each candidate cell c' observed at t_{j+1}, we compute a candidate protein vector $P_{c_{t_j},c'}$ by integrating the deterministic ODE (5) from $P_{c_{t_j}}$ towards a state consistent with c'; this interpolation accounts for a possible single mode switch along the interval (see Appendix A for details). The cost of the candidate is

$$C^j_{c_{t_j},c'} := \mathrm{loss}\Big(k_{\mathrm{on}}^{\theta}\Big(P_{c_{t_j},c'}(t_{j+1})\Big)\,,\, \mathbf{k}_{z_{c'}}\Big)\,, \tag{8}$$

where $\mathbf{k}_{z_{c'}}$ denotes the target burst-mode vector associated with the basin of c', and `loss` is typically taken as the cross-entropy, consistent with the sigmoid parametrization of k_{on}^{θ} in Eq. (1).
We then solve the entropically regularized OT problem

$$\pi_j := \arg\min_{\pi} \left\{ \sum_{c,c'} C^j_{c,c'}\, \pi(c,c') + \varepsilon_1\, \pi(c,c') \ln \frac{\pi(c,c')}{\hat{P}_{t_j}(c)\, \hat{P}_{t_{j+1}}(c')} \right\}, \tag{9}$$

subject to the marginal constraints $\sum_c \pi(c,c') = \hat{P}_{t_{j+1}}(c')$ and $\sum_{c'} \pi(c,c') = \hat{P}_{t_j}(c)$. Here $\hat{P}_{t_j}$ denotes the empirical distribution of cells at time t_j, and ε_1 plays the role of a temperature parameter, typically decreased linearly over iterations from $\sqrt{N_g}$ to 0, where N_g is the number of genes. $\sqrt{N_g}$ is chosen to counterbalance the typical scale of the cost matrix (8): at the first iterations, the number of incorrectly labeled genes scales as the square root of the total number of genes.
The optimal coupling π_j is then used to sample a descendant c' for each ancestor c_{t_j}, and we set $P_{c_{t_{j+1}}} := P_{c_{t_j},c'}$. Repeating this from t_1 to t_{N-1} yields full protein trajectories for all sampled cells.

Step 2: GRN update. Given the reconstructed protein trajectories, we update the GRN by minimizing the loss between GRN-predicted burst rates and mixture-derived basin modes over all cells and timepoints:

$$\theta^* = \arg\min_{\theta} \sum_c \mathrm{loss}\Big(k_{\mathrm{on}}^{\theta}(P_c),\, \mathbf{k}_{z_c}\Big) + \lambda \|\theta - \theta^0\|_1, \tag{10}$$

where $\mathbf{k}_{z_c}$ is the burst-mode vector associated with the basin of cell c. Compared to CARDAMOM, all timepoints now contribute jointly through their reconstructed protein values, and a simple $l1$ penalty with respect to the prior GRN θ^0 replaces the custom time-dependent regularization. The regularization strength λ thus controls the sparsity of the inferred network. We set it as $\lambda = 0.05 \cdot N_c/\sqrt{N_g}$, where N_c is the number of cells, reflecting the heuristic that the number of active regulators per gene scales at most as $\sqrt{N_g}$, and that a per-gene basin label attribution error of 5% is a reasonable target for reliable inference.

Step 3: Basin label refinement. Given the updated GRN θ^*, basin labels are refined by solving a GRN-constrained assignment problem that balances the per-cell NB log-likelihood with consistency between inferred proteins and candidate basin modes:

$$\gamma^* = \arg\min_{\gamma} \sum_{z,c} -\Big(\ln \mathrm{Pr}_c(z) + \varepsilon_2 \ln\Big(1 - \mathrm{loss}\Big(k^{\theta}_{\mathrm{on}}(P_c), \mathbf{k}_{z_c}\Big)\Big)\Big)\, \gamma(c,z), \tag{11}$$

subject to $\sum_c \gamma(c,z) = \sum_c \mathrm{Pr}_c(z)$ and $\sum_z \gamma(c,z) = 1$. Here, the quantity $\mathrm{loss}\big(k^{\theta}_{\mathrm{on}}(P_c), \mathbf{k}_{z_c}\big)$ is renormalized to be in $[0,1]$, such that the ln is well defined. The parameter ε_2 controls how strongly consistency with the mechanistic model influences label assignment; it is typically increased linearly over iterations from 0 to 0.5, so that it never completely overwhelms the NB likelihood.
Cells are then reassigned as $z_c = \arg\max_z \gamma^*(c,z)$.

Algorithm summary. Steps 1–3 are iterated until convergence, typically within 20–40 iterations:

Algorithm 1 CardamomOT

1: Infer NB mixture parameters; initialize basin labels $\{z_c^0\}$ via (7), or via (11) if a prior GRN is available.
2: **for** $k = 0, 1, \ldots$ until convergence **do**
3: Compute cost matrices $\{C^j\}$ via interpolation (Appendix A); solve N OT problems (9); sample L protein trajectories from the multimarginal coupling $\pi_1 \otimes \cdots \otimes \pi_N$. (Step 1)
4: Update $\theta^{(k+1)}$ by solving (10). (Step 2)
5: Update basin labels $\{z_c^{(k+1)}\}$ via (11). (Step 3)
6: **end for**

1.2.3 Postprocessing: kinetic rate recalibration via NeuralODEs

Once the GRN θ^* and protein trajectories $\{P^l(t)\}_{l=1,\ldots,L}$ have converged, we recalibrate kinetic rates so that simulations from the model reproduce the observed dynamics. This step can be interpreted as fitting the parameters of an SDE approximating the protein dynamics in the metastable regime, theoretically derived from the mechanistic model in [44] (Appendix C).
In practice, we use a two-step approximation:

1. We use the inferred GRN θ^* and inferred protein trajectories $\{P^l(t)\}$ to recalibrate the protein degradation rate d_1 using a NeuralODE approach on the deterministic limit (5) with teacher forcing [45], as is now standard for single-cell time series [46].

2. We recalibrate the mRNA degradation rate d_0 by fitting the variance of stochastic fluctuations around the deterministic trajectories, so that simulated expression noise matches the observed variability.

These refined kinetic parameters ensure that the model generates trajectories with realistic timescales and levels of stochasticity.

Overall, the method outputs: the inferred GRN θ^* (signed and directed); the reconstructed protein trajectories with basin labels and burst-mode vectors; and all parameters of the mechanistic model described in Section 1.1, together forming a calibrated generative model ready for simulation and *in silico* perturbation experiments.

2 Results

We validated the CardamomOT pipeline on both synthetic and experimental datasets, demonstrating its ability to jointly infer Gene Regulatory Networks (GRNs) and cellular trajectories from temporal snapshots of scRNA-seq data, and to generate realistic simulated datasets. The method was tested on five benchmark network topologies, previously introduced in [33], and three experimental datasets: (i) mouse embryonic stem cell (mESC) differentiation [3] (9 timepoints, $\sim$ 1,300 cells), (ii) sympathoadrenal differentiation ordered by pseudotime [39] (6 pseudo-timepoints, $\sim$ 1,800 cells), and (iii) mouse embryonic fibroblast (MEF) reprogramming to induced pluripotent stem cells (iPSCs) [15] (35 timepoints, $\sim$ 250,000 cells). In what follows, we present results that illustrate the complete CardamomOT workflow, from preprocessing to perturbation analysis, and show how the calibrated mechanistic model can be used as a comprehensive tool for dynamic single-cell analysis.

2.1 Simulation of temporal snapshots

To evaluate CardamomOT's performance, we first generated synthetic datasets using the HARISSA package [47], which simulates exact trajectories of the mechanistic model described in Section 1.1. We used five benchmark network topologies from [33]: (i) FN4, a 4-gene feedforward network; (ii) FN8, an 8-gene feedforward network; (iii) BN8, an 8-gene bifurcating network; (iv) CN5, a 5-gene cascade network; and (v) a tree topology with varying numbers of genes (5, 10, 20, 50, 100 genes).

For each network, we simulated *in vitro*-like perturbation experiments by first equilibrating the system for $t < 0$, then introducing a virtual stimulus gene fixed at maximal protein level at $t = 0$, triggering a transition toward a new stochastic steady state. Temporal snapshots were generated by independently sampling cells at predefined timepoints from $t = 0$ to $t = 96$ hours. The model parameters ($k_{0,i}$, $k_{1,i}$, $s_{0,i}/k_{\text{off},i}$, β_i, θ_{ij}) and degradation rates ($d_{0,i}$, $d_{1,i}$) for all genes used for these benchmarks are available online with the code repository.

This simulation framework allowed us to evaluate CardamomOT's performance under controlled conditions where the ground truth GRN and protein levels are known, as well as true cellular trajectories when simulating without killing cells at each timepoint. We then apply the method to experimental datasets, where validation is more challenging.

2.2 The statistical mixture model accurately captures both mechanistic parameters and biological cell states

The first step of CardamomOT (preprocessing) consists in calibrating the Negative Binomial (NB) mixture model described in Equation (4) for each gene and timepoint (Figure 1C). This step serves two critical purposes: (i) inferring the mechanistic parameters ($k_{0,i}$, $k_{1,i}$, $k_{\text{off},i}/s_{0,i}$) from the data, and (ii) clustering cells into discrete promoter-activity basins that reflect their underlying regulatory states.

We first validated this approach on the simulated datasets, where ground truth parameters are known. The inferred burst frequencies $k_{0,i}/d_{0,i}$, $k_{1,i}/d_{0,i}$ and the inferred burst size parameter $k_{\text{off},i}/s_{0,i}$ are reconstructed with good accuracy across all benchmark networks (Table S1), demonstrating that the NB mixture model can recover the mechanistic parameters underlying gene expression dynamics. In addition, CardamomOT precisely classifies cells into basins using

maximum likelihood for each gene, with an average of 85%–93% correct classification across the datasets (Table S1).

For experimental datasets, where ground truth is unavailable, we assessed the quality of the statistical model fit through multiple metrics:

- **Low-dimensional embeddings:** We performed UMAP dimensionality reduction on both the original scRNA-seq data and data simulated from the calibrated statistical model. The simulated data preserved the overall temporal structure of the experimental data in UMAP space (Figure 2A,B), indicating that the statistical model captures the essential features of the gene expression landscape. In all comparisons using UMAPs throughout the paper, a low-dimensional embedding is learned from the concatenated dataset and the coordinates are then used to plot each dataset separately.

- **Marginal distributions:** The inferred NB mixture models accurately reproduced the observed marginal distributions and temporal patterns of gene expression for all time-points, as illustrated in Figure 2E–H for selected genes exhibiting strong connectivity in the inferred GRN of each dataset (Figures S1–S3).

- **Correlations:** We assessed the ability of the NB mixture model to reproduce pairwise gene expression correlations observed in the reference data (Figure S5). The inferred model captures the overall correlation structure, with gene pair correlations in the NB mixture closely tracking those in the reference data across all three datasets. However, the strongest correlations tend to be slightly underestimated. This is consistent with a known limitation of the NB mixture framework: by assuming conditional independence of genes given the basin label, the model cannot capture residual correlations between genes that remain correlated within individual basins. As a result, intra-basin co-regulation between highly correlated genes is partially lost, leading to a mild attenuation of the highest correlation values.

- **Cell type identification:** We trained Random Forest classifiers on the experimental data using cell type annotations (when available) or clustering-based labels, and evaluated their performance on data simulated from the calibrated statistical model. The classifiers achieved high accuracy on reconstructing the relative proportions of cell types (Figure 2C,D), demonstrating that the inferred basin labels correspond to biologically meaningful cell states. This evaluation strategy, based on cross-dataset classification, has become standard in single-cell analysis [48] and provides an objective measure of how well the statistical model captures cell type-specific expression patterns. These results are particularly striking for the Kameneva and Schiebinger datasets, which exhibit non-trivial branching dynamics that are shown to be well distinguished by the statistical model.

Taken together, these results show that the preprocessing step of CardamomOT not only infers mechanistically interpretable parameters but also produces a discrete representation of cellular states (basin labels) that forms the foundation for the subsequent joint GRN and trajectory inference.

2.3 Benchmarking CardamomOT: performance, robustness and scalability of GRN inference

We next evaluated CardamomOT's ability to infer GRN structure from temporal snapshots. We compared CardamomOT against several state-of-the-art methods: (i) CARDAMOM [33], the previous version of our method; (ii) Reference Fitting (RF) [25], a recent optimal transport-based method that jointly infers trajectories and linear GRNs; (iii) GENIE3 [8], a widely used random forest-based method; and (iv) SINCERITIES [49], a method designed for temporal scRNA-seq data. For methods that infer undirected edges, we additionally compared against

PIDC [50] and Pearson correlation, used as a sanity check to ensure that good performance is not solely due to highly correlated data that would be easy to detect.
All methods were evaluated using the Area Under the Precision-Recall curve (AUPR), a standard metric for GRN inference that is robust to class imbalance (i.e., the fact that most gene pairs are not connected). Results were averaged over 10 independent simulations for each network topology.

CardamomOT achieves superior and scalable GRN inference across network topologies. On all benchmark networks, CardamomOT achieved consistently higher AUPR scores than all competing methods (Figure 3A). The improvement was particularly pronounced for the more complex networks (FN4, FN8, BN8, and trees with more than 10 genes), where CardamomOT's mechanistic modeling and trajectory inference provided a substantial advantage over methods that rely solely on correlation-based or quasi-static approximations. CardamomOT infers both the sign and direction of regulatory interactions, which is crucial for understanding causal relationships in GRNs. When comparing against methods that only infer undirected edges (Figure 3B), CardamomOT maintained its performance advantage, highlighting the value of mechanistic modeling for resolving regulatory directionality.
The convergence of the inference procedure is illustrated in Figure S8: the training loss decreases monotonically and stabilizes within 20–25 optimization steps across all benchmark networks, while the AUPR score rises rapidly in the first iterations and plateaus at its final value. This indicates that the iterative EM-like procedure converges reliably and that the reported AUPR scores correspond to fully converged models rather than intermediate states.
Moreover, the runtimes reported in Table S1 show that CardamomOT scales well to realistic dataset sizes: even for networks of 100 genes with 1,000 cells, inference completes in under one minute, and CardamomOT remains faster than GENIE3, a widely used reference for GRN inference. The true limiting factor for scaling to larger networks is therefore not computational cost but the identifiability of the network structure: with n genes, the number of GRN parameters grows as n^2, and the information available in temporal snapshots becomes insufficient to robustly constrain all interactions beyond a certain network size. This identifiability ceiling is further discussed in Section 3.

Iterative refinement significantly improves GRN inference accuracy. To assess the contribution of the iterative refinement procedure (Steps 1–3 in Section 1.2.2), we compared CardamomOT against several ablated versions: (i) a version that directly returns the network and trajectories inferred after the first round of the iterative procedure, i.e. the network computed from the trajectories associated with an OT coupling between basins; (ii) a version with random initialization of the OT coupling instead of the optimal transport-based initialization; and (iii) a version combining both modifications. The full CardamomOT method consistently outperformed ablated versions (i) and (iii) (Figure 4A–E), demonstrating that the iterative refinement procedure does contribute to improved inference accuracy. Remarkably, the iterative procedure converged to similar GRN structures regardless of initialization (compare "CardamomOT" and "random init" in Figure 4A–E), suggesting that the benchmark networks exhibit sufficient identifiability for the method to escape local optima.

CardamomOT is robust with respect to prior estimation of kinetic rates. A critical advantage of CardamomOT over CARDAMOM is that it directly incorporates protein degradation rates d_1 into the inference procedure (see Equation (5) and Step 1 of Section 1.2.2), allowing a much better approximation of protein trajectories than the quasi-static approximation of the previous method. To evaluate robustness to uncertainty in these rates, we ran CardamomOT on simulated data while systematically perturbing the input degradation rates from $\pm 20\%$ up to $\pm 400\%$ of their true values. The method maintained high AUPR scores (averaged across

FN4, FN8, BN8 and CN5) even with moderate perturbations (±50%), and significantly outperformed random guessing and competing benchmark methods even with very large perturbations (±200–400%, Figure 4F). This shows that while accurate kinetic parameters improve inference, CardamomOT can still recover meaningful GRN structure even when these parameters are moderately mis-specified—a common scenario in experimental applications.

GRN inference for experimental datasets. We applied CardamomOT to the three experimental datasets to obtain GRN structures (Figures S1–S3). For the Semrau and Kameneva datasets, all available cells were used directly on the subset of genes identified in [33] and [39], respectively. For the Schiebinger dataset, genes were selected as a combination of markers identified in the original study [15] and dynamically variable genes across timepoints, the latter being identified using a weighted measure combining per-gene entropy and Wasserstein distances between consecutive timepoint distributions, following [39].

The inferred GRNs are notably sparse (Figures S1A–S3A), with a small number of hub regulators concentrating most of the regulatory activity— a hallmark of biologically realistic network structure that is not always recovered by inference methods without explicit sparsity constraints. The most connected regulators are consistent with those identified in the original studies: for the Semrau dataset, *Hoxb2*, *Lamb1* and *Col4a2* emerge as top regulators (Figure S1C), in agreement with [3]; for the Kameneva dataset, *CHGA* dominates the network (Figure S2C), consistent with [39].

We note that the diagonal coefficients of the inferred GRN matrices, corresponding to autoregulatory interactions, are systematically non-negligible across all three datasets (Figures S1B–S3B). While often excluded in GRN inference frameworks [8], these terms naturally arise in our setting as a parsimonious explanation for bimodal gene expression dynamics. Their interpretation and potential biological relevance are discussed in detail in Section 3.

Following the simulation framework described above, a virtual stimulus gene is included in each inferred network to represent the experimental perturbation applied at $t = 0$. For the Semrau dataset, this corresponds to the addition of retinoic acid, which triggers ES cell differentiation [3]; for the Schiebinger dataset, it represents the doxycycline-induced expression of the OSKM reprogramming factors [15]. For the Kameneva dataset, where cells are ordered by pseudotime rather than experimental time and no clear exogenous stimulus is applied, a stimulus gene is nonetheless included to push cells out of their initial equilibrium state, but its contribution to the network is strongly penalized via the prior regularization so that it does not dominate the inferred regulatory interactions.

As an internal consistency check, the stimulus gene inferred for the Semrau dataset activates early pluripotency-exit regulators such as *Esrrb*, *Klf2* and *Sox2* while repressing *Klf4* and *Dnmt3a* (Figure S1F), broadly consistent with the known pro-differentiation effect of retinoic acid. For the Schiebinger dataset, the experimental protocol involves in reality two successive phases: an initial doxycycline-induced expression of the OKSM reprogramming factors (days 0–8), followed by withdrawal of doxycycline and transition to either 2i/LIF or serum/LIF medium (days 8–18) [15]. Since CardamomOT uses a single constant stimulus gene, this virtual node captures an averaged effect of these successive perturbations rather than their individual contributions. Accordingly, the inferred stimulus targets (Figure S3F) reflect early transcriptional responses to the reprogramming cocktail rather than specific OKSM targets, and should be interpreted with this simplification in mind.

These inferred networks serve as the foundation for the perturbation experiments described in Section 2.6.

Overall, these results establish CardamomOT as a robust and accurate method for GRN inference.

2.4 CardamomOT accurately reconstructs hidden protein vectors, cellular trajectories and velocity fields

Unlike existing trajectory inference methods that model dynamics directly on mRNA levels using simplified diffusion processes [15] or linear systems [25], CardamomOT constructs trajectories at the protein level, explicitly accounting for the highly bursty nature of transcription and the slower, more stable dynamics of proteins. We evaluate the benefits of this trajectory reconstruction on three complementary levels:

- the accuracy of the predicted **velocity fields** on cross-sectional data (Figure 5), evaluated in both transcript and protein space;
- the accuracy of the predicted **cell-to-cell couplings** on true trajectory data when simulated cells are not virtually killed at each timepoint (Figure S7);
- the similarity of the hidden protein levels reconstructed by the method compared to the simulated ones, which are not used by the algorithm (Figure S6).

Velocity fields on cross-sectional data. For the mechanistic model, the expected instantaneous velocity fields for mRNAs and proteins at a given state (M, P) are given by the velocity fields v_0 and v_1 described in Eqs. (6) and (5), respectively. For CardamomOT, both velocity fields are evaluated using the inferred parameters (θ, d_0, d_1, k_1) and the reconstructed protein trajectories. For RF [25], the mRNA velocity is given by the instantaneous drift of the calibrated linear Ornstein–Uhlenbeck process, $v_0^{\mathrm{RF}}(M) = MA + b$, where A and b are the inferred generator matrix and bias; the protein velocity is derived by applying the same linear operator to the protein values used as a proxy. We evaluate the agreement of each predicted velocity field with the mechanistic reference using a weighted cosine similarity, where weights are given by the product of the norms of the two velocity vectors.

RF's mRNA-based velocity field shows poor agreement with the mechanistic reference, with cosine similarities close to zero across all benchmark networks (Figure 5A). This reflects an intrinsic limitation: mRNA levels are dominated by stochastic transcriptional bursting, and a global linear model cannot capture the local, protein-dependent regulation that drives gene expression changes. The resulting velocity field is essentially a uniform flow in UMAP space (Figure 5E), bearing little resemblance to the structured, GRN-driven dynamics of the system. In contrast, CardamomOT achieves cosine similarities exceeding 0.8 across all networks (Figure 5A), and the corresponding UMAP stream plots (Figure 5C) show velocity fields with coherent, biologically interpretable structure.

The disparity between methods is also visible at the protein level. Proteins evolve more slowly and deterministically than mRNAs, making their dynamics less sensitive to transcriptional noise and therefore more tractable for trajectory inference. When RF's linear dynamics are applied to protein space, the agreement with the mechanistic reference clearly improves (cosine similarity $\sim$ 0.25–0.4, Figure 5B), suggesting that even simplified linear models can partially capture protein-level trends. The resulting velocity field is nevertheless still close to a uniform flow in UMAP space, and poorly reflects the true temporal evolution of cells (Figure 5F).

By contrast, CardamomOT— which directly infers protein trajectories and uses the nonlinear mechanistic model to predict protein velocities — achieves higher cosine similarities (again exceeding 0.75, Figure 5B), with protein UMAP stream plots (Figure 5D) showing well-structured trajectories consistent with the underlying GRN dynamics (in particular a correct orientation of the cycle across timepoint).

Coupling accuracy on true trajectory data. To complement this analysis, we also evaluated both methods on simulated datasets where true cell-to-cell correspondences are known, i.e. where individual cells are tracked across consecutive timepoints (Figure S7). In this setting,

we can directly assess the quality of the inferred couplings by measuring the expected distance between predicted and observed descendants, with the same metric as in [51]. This metric is computed separately for mRNA (with a $\log(1 + x)$ transformation to reduce the influence of highly expressed genes) and for proteins (already rescaled in $[0, 1]$.
CardamomOT achieves consistently lower descendant distances than RF on mRNA (Figure S7A) and proteins (Figure S7B), confirming that the mechanistic coupling provides more accurate cell-to-cell correspondences.

CardamomOT reconstructs protein trajectories from transcriptomic snapshots. A key feature of CardamomOT is its ability to infer protein trajectories that are never directly observed in scRNA-seq data. To evaluate this capability on simulated datasets, where ground truth protein values are available, we compared the reconstructed protein trajectories to the true simulated ones across all benchmark networks (Figure S6). The inferred protein UMAP embeddings closely reproduce the structure of the reference trajectories in all cases, capturing both the global topology of the differentiation process and the temporal ordering of cells. This demonstrates that CardamomOT can reliably recover hidden protein dynamics from transcriptomic snapshots alone.

2.5 The calibrated model accurately regenerates experimental data

A distinguishing feature of CardamomOT is its use as a *generative model*: once calibrated on temporal snapshots, the model (with its inferred GRN, mechanistic parameters and protein dynamics) should be able to simulate new datasets that closely resemble the original data. This generative capability is critical for two reasons: (i) it provides a stringent test of whether the model has captured the essential dynamics of the system, and (ii) it enables *in silico* experimentation through perturbation analysis (Section 2.6).
To validate this capability, we simulated new temporal snapshots starting from $t = 0$ and compared them to the original experimental data using the same metrics as in Section 2.2. Despite the complexity of the inference procedure, the fully calibrated model reproduces the experimental data with a quality comparable to the NB mixture model alone and, in some respects, improves upon it:

- **Low-dimensional embeddings:** Original and simulated cells overlap extensively in joint mRNA UMAP space (Figure 6A–F), confirming that the mechanistic model, driven by the inferred GRN and protein dynamics, faithfully captures the global structure of the gene expression landscape while exhibiting realistic stochastic variability.

- **Marginal distributions and correlations:** The marginal distributions and pairwise gene expression correlations of simulated data are in excellent agreement with those of the NB mixture model at all timepoints (Figures S4 and S5; > 0.95 correlation of gene pair correlations across all datasets), demonstrating that the GRN-driven simulation does not distort the statistical structure captured at the preprocessing stage. The mild underestimation of the strongest correlations observed in Figure S5 is thus inherited from the conditional independence assumption of the NB mixture, and not introduced by the mechanistic model.

- **Regulatory gene dynamics:** Temporal profiles of key regulators show excellent agreement between original and simulated data in both mean expression and variance (Figure 6J–O). This is a particularly stringent test: errors in GRN inference typically manifest as incorrect dynamics for highly connected genes [33], so this agreement provides strong evidence that the inferred GRN structure captures genuine regulatory relationships.

- **Cell type proportions:** Cell type compositions are highly consistent between original and simulated data across all timepoints (Figure 6G–I), confirming that the mechanistic model accurately reproduces the temporal evolution of major cell populations. Nevertheless, for the Kameneva and Schiebinger datasets, some intermediate or rare populations (in particular the "Intermediate" cells in Kameneva and the trophoblasts in Schiebinger) are less well captured, leading to noticeable discrepancies in their inferred proportions. These cell types correspond to subtle variations in expression patterns that are difficult to represent with a statistical model using only two modes per gene, as further discussed in Section 3.

These results demonstrate that the mechanistic model inferred by CardamomOT constitutes a faithful generative model of single-cell transcriptomic dynamics and supports its use for predictive perturbation analysis.

2.6 In silico perturbations predict experimental outcomes and reveal key regulatory genes

A central objective of GRN inference is to assess whether the inferred network can predict the effects of unseen genetic perturbations on cell fate decisions. The calibrated CardamomOT model enables such analyses by simulating the system under controlled perturbation scenarios. We considered two types of perturbations: (i) gene knockout (KO), where a gene's protein level is fixed at zero and its basal parameter β_i is set to -1000, and (ii) gene overexpression (OV), where a gene's protein level is fixed at one (in the rescaled units of Equation (5)) and its basal parameter β_i is set to $+1000$ (this value ensuring that it overwhelms the effect than any other regulatory interaction).
For each perturbation, we generated simulated temporal snapshots using the same time points and sampling scheme as in the original datasets. We quantified perturbation effects by measuring changes in cell-type proportions, with cell identities assigned using the Random Forest classifiers described in Section 2.2. To isolate the contribution of the inferred GRN, we compared experimental data to: (i) wild-type simulated data (*sim WT* in Figure 7); (ii) simulated data with the perturbation applied from $t = 0$ (*Sim perturb*); and (iii) wild-type simulated data in which only the expression level of the perturbed gene was replaced by its value under perturbation (*Sim single*).
Condition (iii) ensures that the observed effects are not trivially driven by the perturbed gene acting as a strong marker of a given cell type, which alone could induce an apparent shift of many cells toward that identity. To further quantify these differences, we performed χ^2 tests comparing the cell-type distributions of Sim perturb and Sim single, using reconstructed cell counts from the observed proportions. All tested perturbations yielded highly significant differences ($p < 10^{-3}$), supporting that the predicted effects are not solely driven by marginal changes in the perturbed gene expression but arise from the propagated regulatory interactions encoded in the inferred GRN.

ES differentiation (Semrau). Using the GRN inferred from the ES differentiation time series, we evaluated the effect of *Dnmt3a* overexpression. The model predicts that *Dnmt3a* OV substantially alters cell fate distribution, with a marked increase in the proportion of neuroectoderm cells at the expense of pluripotent cells compared to the wild-type simulation (Figure 7A). The predicted regulatory targets of *Dnmt3a* are predominantly inhibitory, including key pluripotency regulators such as *Esrrb*, *Zfp42* and *Klf4*, as well as lineage-associated factors such as *Hoxb2* and *Zic3*, consistent with a broad repressive role on the pluripotency transcriptional program.
This result is particularly interesting from a mechanistic perspective. Unlike transcription factors that directly contact the transcription machinery, *Dnmt3a* is a de novo DNA methyltrans-

ferase that acts indirectly on gene regulation through CpG methylation — a well-documented epigenetic mark generally associated with gene expression silencing [52]. The predominantly inhibitory interactions inferred by CardamomOT for *Dnmt3a* are consistent with its known epigenetic function, illustrating that the GRN model is capable of capturing indirect regulatory effects mediated by epigenetic mechanisms and is not restricted to direct physical interactions between transcription factors and their targets.

Sympathoadrenal differentiation (Kameneva). The analysis in [39] identified four important regulators: *CHGA*, *STMN2*, *CHGB* and *HMGB2*. We quantified the effect on the proportion of chromaffin cells and sympathoblasts, the principal cell types in sympathoadrenal development, associated with the perturbation of each of these genes alone or in combination. Among the tested perturbations, KO of *CHGA* led to a marked reduction in the proportion of sympathoblasts (Figure 7B). The combined perturbation consisting of *CHGA* KO and *STMN2* OV further amplified this effect, nearly abolishing the emergence of sympathoblasts in the simulations (Figure 7C). These results indicate that the inferred GRN captures non-trivial combinatorial effects of perturbations on lineage commitment.

MEF-to-iPSC reprogramming (Schiebinger). In [15], the authors reported that OV of *Obox6* and *Zfp42*, two transcription factors not typically included in standard reprogramming cocktails, enhances the efficiency of MEF-to-iPSC conversion. To assess whether CardamomOT recovers this behavior, we inferred a GRN from the reprogramming time series and simulated the effect of *Obox6* and *Zfp42* OV.

Both perturbations consistently increase the proportion of iPSC-like cells compared to control conditions (Figure 7D,E), in line with the experimental observations of [15], with a similar magnitude of effect (for *Obox6* OV we measure an increase from $\sim 20\%$ to $\sim 40\%$ of iPSC-like cells). Importantly, this effect emerges solely from the temporal scRNA-seq data used for inference, without incorporating any prior knowledge about the function of these factors.

The predicted regulatory targets provide a mechanistic interpretation of this behavior. *Zfp42* OV activates several pluripotency-associated genes, including *Esrrb* and *Dppa4* (Figure 7D), both well-known markers of the naive pluripotent state. *Obox6* OV similarly activates *Gata3*, *Lncenc1* and *Mycn*, and notably upregulates *Zfp42* itself (Figure 7E), suggesting a positive feedback loop between the two factors that may reinforce the iPSC-like state. Rather than inducing broad lineage rewiring, both perturbations thus appear to strengthen specific components of the pluripotency transcriptional program already present in the data.

Taken together, these perturbation analyses indicate that CardamomOT can be used not only to infer GRN structure but also to generate quantitative, mechanistically interpretable predictions of perturbation effects across distinct biological systems. While experimental validation remains essential, the consistency observed across multiple datasets supports the relevance of the inferred networks for studying cell fate control.

3 Discussion

Taken together, our results demonstrate that CardamomOT provides a complete pipeline for mechanistic analysis of temporal scRNA-seq data. Starting from raw count matrices and time-point labels, the method (i) infers mechanistic parameters and identifies discrete regulatory states (Section 2.2), (ii) reconstructs both GRN structure and hidden protein trajectories through an iterative refinement procedure (Sections 2.3 and 2.4), (iii) generates a calibrated model that can simulate realistic cellular dynamics (Section 2.5), and (iv) predicts the effects of unseen genetic perturbations (Section 2.6).

A unified framework for GRN inference and trajectory reconstruction Most existing approaches treat GRN inference and trajectory inference as separate problems: GRN inference methods typically assume static or quasi-static cellular states, while trajectory inference methods focus on reconstructing cell-to-cell transitions without explicitly modeling gene regulation. CardamomOT unifies these perspectives through an iterative procedure that alternates between reconstructing protein trajectories given a GRN (Step 1) and updating the GRN structure given the trajectories (Step 2), until convergence.

This iterative refinement is essential for achieving high inference accuracy: removing the iterative loop substantially degrades performance (Figure 4), confirming that the GRN and trajectories contain complementary information that must be jointly optimized. At the same time, the method converges to similar GRN structures regardless of initialization in our benchmarks (Figure 4A–E), suggesting that the temporal structure of the data provides sufficient constraints to escape local optima for the network topologies and temporal resolutions tested here.

The mechanistic foundation of CardamomOT also enables trajectory reconstruction at the protein level rather than the mRNA level. As shown in Section 2.4, mRNA-based velocity fields exhibit poor agreement with the underlying mechanistic dynamics due to the highly stochastic, bursty nature of transcription, whereas protein dynamics are slower and more deterministic, providing a more stable substrate for trajectory inference. By explicitly modeling the conditional distribution of mRNA given protein levels, CardamomOT captures both bursty transcriptional dynamics and relatively smooth, GRN-driven protein trajectories, achieving high cosine similarities with respect to the ground-truth velocity fields (Figure 5B).

Mechanistic modeling improves interpretability and generalizability Unlike black-box or purely data-driven approaches, CardamomOT is grounded in an explicit mechanistic model of gene expression derived from biophysical principles. This provides two key advantages. First, the inferred parameters—burst frequencies, burst sizes, and protein degradation rates—are directly interpretable in terms of molecular processes, enabling biological insights beyond network topology alone. Second, by incorporating prior knowledge about kinetic rates and known regulatory interactions, CardamomOT can leverage existing biological information to improve inference, particularly in data-limited regimes.

Robustness to uncertainty in kinetic parameters is a practical consideration for experimental applications. As shown in Figure 4F, CardamomOT maintains high inference accuracy even when protein degradation rates are perturbed by up to $\pm 50\%$ from their true values, and significantly outperforms baseline methods even with perturbations as large as ± 200–400%. This robustness arises because the iterative procedure jointly optimizes the GRN structure and the inferred trajectories, allowing the method to partially compensate for kinetic misspecification through adjustments in the network topology.

The same mechanistic framework allows CardamomOT to function as a generative model. Once calibrated on experimental data, the model can simulate new datasets that closely reproduce the original observations in terms of marginal distributions, low-dimensional embeddings, and cell-type proportions (Figure 6). This generative capability simultaneously provides a stringent test of whether the model has captured the essential dynamics of the system and enables *in silico* experimentation, including perturbation analysis and prediction of unseen cell fates.

Perturbation predictions and biological relevance of inferred GRNs A central objective of GRN inference is to predict how unseen genetic perturbations will alter cellular behavior. CardamomOT addresses this challenge by simulating perturbations directly within the calibrated mechanistic model. In the experimental systems analyzed here, the inferred GRNs successfully recapitulate known perturbation effects, such as the impact of transcription factor overexpression on reprogramming efficiency, and suggest additional regulatory relationships that could be tested experimentally.

These results indicate that CardamomOT not only infers network topology but also captures quantitative, functionally relevant regulatory relationships. Nonetheless, experimental validation remains essential, particularly for predicted regulatory interactions that have not been previously characterized.

Optimal transport as a mechanistic framework for trajectory inference The use of optimal transport for trajectory inference has become widespread in recent years, motivated by the idea that cells follow trajectories that minimize a cost in gene expression space. However, most OT-based methods implicitly assume that cells follow Brownian motion with a constant diffusion rate, corresponding to a Euclidean or squared-Euclidean cost. This assumption lacks biological justification and ignores the mechanistic constraints imposed by gene regulation [23]. CardamomOT extends this framework by introducing a mechanistic cost function that explicitly incorporates the GRN structure. The cost of transitioning between two cells is determined by how well the GRN-predicted burst rates match the observed burst modes, naturally accounting for the fact that gene expression dynamics are driven by protein-mediated regulation rather than passive diffusion. Our approach can be viewed as a generalization of Reference Fitting to the case where the reference process is a nonlinear mechanistic model instead of a linear Ornstein–Uhlenbeck process, making it suitable for modeling complex regulatory dynamics.
From a theoretical perspective, there is a direct connection between our approach and Wasserstein Lagrangian Flows, which incorporate potential energy constraints into the OT Lagrangian. CardamomOT implicitly defines a Lagrangian derived from the mechanistic model via stochastic calculus (see Appendix B), but avoids the computational burden of parametrizing all characteristics of the process and time-varying measures with neural networks. By discretizing the problem into a sequence of pairwise OT problems constrained by the mechanistic model, CardamomOT achieves computational efficiency while retaining mechanistic interpretability.

Scalability and identifiability In our experiments, the computational cost of CardamomOT was not the primary bottleneck: inference on networks of approximately 100 genes with tens of thousands of cells completed in about one hour. The true limiting factor is the identifiability of the network parameters. With n genes and a fully connected GRN, the number of parameters scales as n^2, *i.e* 10^4 for $n = 100$; beyond this scale, the information contained in temporal snapshots may be insufficient to robustly constrain the network structure.
Importantly, from a theoretical perspective, the convergence of our iterative procedure to a global optimum is not guaranteed, as the objective (17) is non-convex in general; the empirical convergence observed throughout our experiments should therefore be understood as convergence to a stationary point, whose quality depends on the temporal structure and richness of the data.
The stochasticity inherent in our iterative procedure nevertheless provides a natural framework to assess identifiability of this optimum. Indeed, the OT coupling sampled at each iteration introduces randomness analogous to the temperature parameter in simulated annealing, so that different initializations may in principle converge to different local optima. In practice, our benchmark experiments (Figure 4A–E) show that the method converges to similar GRN structures regardless of initialization, suggesting that the temporal data provide sufficient constraints for the networks tested here. Systematically exploring the variability of optimal networks across different initial couplings would therefore be a principled way to quantify identifiability as a function of network size and data richness, and represents a promising direction for future work.
Addressing identifiability for larger networks will require incorporating external biological constraints. Prior regulatory information—such as transcription factor binding databases, ChIP-seq data, or resources like OmniPath [53] can substantially reduce the effective parameter space. A prior network construction option using the Neko tool [54] via OmniPath is already available in the CardamomOT repository.

A complementary strategy concerns the structure of the basin space. While the number of theoretically possible basins for n genes is 2^n, biological systems occupy only a small fraction of this space: not all combinations of per-gene expression modes are accessible, as cell types correspond to coherent regulatory states. It is therefore both possible and desirable to introduce an intermediate step that restricts the basin space using prior biological information—such as known cell types or unsupervised clustering—combined with mixture-of-experts approaches to construct burst-mode vectors from per-gene univariate NB fits. Reducing the effective basin space in this way would simultaneously improve identifiability and computational efficiency.

Diagonal coefficients and auto-regulatory interactions A specificity of CardamomOT concerns the diagonal coefficients of the inferred GRN matrix, i.e. auto-regulatory interactions. As visible in the inferred networks on experimental datasets (Figures S1–S3), CardamomOT tends to infer non-negligible auto-activations for many variable genes. This is not a numerical artifact: in the mechanistic model, the combination of a positive auto-activation with a negative basal parameter β_i is the most parsimonious way to produce bistability for an isolated gene, and therefore to reproduce transitions between two expression modes. It is thus natural that for any gene exhibiting bimodal dynamics, the model partially attributes this variability to self-reinforcing regulation [35].

Importantly, this tendency is not purely a mathematical convenience. Positive auto-regulatory feedbacks are well documented as a mechanism for stabilizing cell states and enhancing the robustness of cell fate decisions [55]. In this sense, the auto-activations inferred by CardamomOT may reflect genuine biological regulatory structure rather than overfitting. We note that most GRN inference methods explicitly exclude diagonal coefficients from inference [7], which circumvents this question entirely at the cost of ignoring a potentially relevant class of interactions.

In CardamomOT, we deliberately choose not to apply a specific penalty on diagonal coefficients, allowing the model to use auto-activation as an explanation for observed bimodality when this is consistent with the data. In practice, however, we recommend that users of CardamomOT specify in advance whether auto-regulatory interactions are biologically plausible for their system of interest—for instance by setting the corresponding entries of the prior network θ^0 to zero—in order to prevent the model from systematically routing unexplained variability through this pathway.

Limitations and future directions Despite its strengths, the current version of CardamomOT has limitations that open avenues for future development.

Model assumptions. CardamomOT relies on several simplifying assumptions, including the two-state model of transcription, the bursty regime approximation ($k_{on} \ll k_{off}$), and the use of sigmoids to model GRN-dependent burst rates. While these assumptions are well supported by experimental observations [56, 57], they may not hold universally across all genes and cell types. Notably, an extension to softmax activation functions — which generalize the sigmoid to genes exhibiting more than two expression modes — is already implemented in the CardamomOT package. In practice, this extension introduces a preliminary model selection step in which the optimal number of modes per gene is determined from a user-specified maximum using an AIC criterion; the full inference pipeline is then adapted to handle gene-specific numbers of modes. This comes at the cost of reduced interpretability, since each mode transition is governed by a distinct regulatory network. Further extensions could incorporate even more flexible regulatory functions, such as Hill functions [58, 34] or more complex neural network-based parametrizations [59], albeit with further reductions in interpretability and identifiability.

Temporal resolution. CardamomOT requires temporal snapshots at multiple time points, and

its performance depends on the temporal resolution. The OT assumption—that between two consecutive timepoints, mode transitions are direct with no unobserved intermediate states—is well justified when timepoints are sufficiently close relative to regulatory timescales. For datasets with sparse or irregular sampling, rapid transient dynamics involving multiple mode transitions may be missed. Extending the framework to accommodate intermediate unobserved switches would improve robustness, albeit at an additional computational cost.

Proliferation and death. The current implementation of CardamomOT assumes that cells neither proliferate nor die between consecutive timepoints, so that the marginal constraints of the OT problem are strictly enforced. This assumption is reasonable for many differentiation datasets, but may break down in systems with significant proliferation or apoptosis. Incorporating growth and death rates into our framework is however conceptually straightforward within the optimal transport formalism. When proliferation and death rates can be estimated from external data — for instance from cell cycle markers or DNA content, as in [15] — they can be directly folded into the marginal constraints of the OT problem following the same strategy. When lineage tracing data are available alongside transcriptomic snapshots, the observed lineage tree can be used to deconvolve the contribution of proliferation from that of regulatory dynamics, following an approach analogous to [60]. In the absence of such prior information, unbalanced optimal transport [61] provides a natural relaxation of the marginal constraints, allowing the inferred coupling to account for unequal cell numbers across timepoints without requiring explicit knowledge of proliferation or death rates. These extensions lie beyond the scope of the present work, but are actively being integrated into the CardamomOT package.

Multi-omic extensions. Beyond the transport formalism itself, the mechanistic model underlying CardamomOT can also be enriched to accommodate richer data modalities. In its current form, it describes gene expression dynamics at the level of mRNA and protein, and is thus naturally suited to transcriptomic data, possibly augmented by proteomic measurements. Multi-omic single-cell technologies are now increasingly available, jointly profiling transcription and chromatin accessibility (e.g. scATAC-seq), DNA methylation, or histone modifications. These data provide complementary mechanistic information: chromatin accessibility reflects the regulatory state of promoters and enhancers, and is directly related to the promoter-switching dynamics modeled here; methylation provides information about longer-term epigenetic memory. Extending our mechanistic framework to jointly model transcription, chromatin dynamics, and other epigenomic layers is a major direction for future work. Our inference approach — grounded in the Schrödinger problem and relying on state-of-the-art OT tools — is well suited to be extended to such richer mechanistic models, and we anticipate that this line of development will substantially expand the scope and biological resolution of mechanistic single-cell inference.

In summary, CardamomOT provides a principled, mechanistically grounded framework for joint GRN inference, trajectory reconstruction, and generative modeling from temporal snapshots of single-cell transcriptomic data. By explicitly modeling the stochastic dynamics of gene expression and integrating optimal transport theory with mechanistic modeling, it significantly advances our ability to decode the regulatory programs that govern cellular differentiation and reprogramming. Beyond inference, the calibrated model functions as a digital twin of the biological system: it can simulate realistic single-cell datasets, predict the effects of unseen genetic perturbations, and generate mechanistically interpretable hypotheses about cell fate control — capabilities that go beyond what current state-of-the-art methods offer in a unified framework. We anticipate that CardamomOT will be most powerful in biological systems where prior knowledge is available to constrain the inference: systems with characterized protein kinetics, known regulatory modules, or existing network resources such as OmniPath [53]. As single-cell atlases of well-studied differentiation and reprogramming systems mature, such settings are increasingly

common, and CardamomOT could become a useful reference tool for mechanistic single-cell analysis, bridging the gap between data-driven trajectory inference and biophysically grounded models of gene regulation.

A Detailed interpolation procedure

We describe how the candidate protein vector $P_{c,c'}(t_{j+1})$ is computed for each pair of cells (c, c') at consecutive timepoints (t_j, t_{j+1}). The goal is to integrate the protein ODE (5) between t_j and t_{j+1}, while allowing for a single switch of the driving burst mode.

Switching time. For each source cell c and each gene i, we introduce a *switching parameter* $\alpha_{c,i} \in (0, 1)$ that encodes the fraction of the interval $[t_j, t_{j+1}]$ elapsed before the mode of gene i transitions from the basin of c to that of c'. Formally, the protein of gene i evolves under basin z_c on $[t_j,\, t_j + \alpha_{c,i}\Delta t]$ and under basin $z_{c'}$ on $[t_j + \alpha_{c,i}\Delta t,\, t_{j+1}]$, where $\Delta t = t_{j+1} - t_j$.

Explicit interpolation formula. Since (5) is a linear ODE with piecewise constant driving term, it can be integrated in closed form on each sub-interval. For gene i, denoting by $m_{c,i} := \mathbf{k}_{z_c,i}/k_{1,i}$ and $m_{c',i} := \mathbf{k}_{z_{c'},i}/k_{1,i}$ the normalized basin modes at t_j and t_{j+1}, respectively, we obtain

$$P_{c,c',i}(t_{j+1}) = m_{c',i} + \Big(m_{c,i} + \big(P_{c,i}(t_j) - m_{c,i}\big)e^{-d_{1,i}\,\alpha_{c,i}\Delta t} - m_{c',i}\Big)\, e^{-d_{1,i}(1-\alpha_{c,i})\Delta t}. \tag{12}$$

Use of observed mRNAs as mode proxies. To better reflect the stochasticity of mRNA in the full model (2), we replace the normalized basin modes $m_{c,i}$ by observed (normalized) mRNA counts $M_{c,i}\,\frac{s_{0,i}}{k_{1,i}k_{\mathrm{off},i}}$ as proxies for the true modes. This provides a richer, cell-specific signal while introducing only a mild approximation (gene expression noise and mode uncertainty are partially conflated). In practice, this proxy yields more realistic protein trajectories, especially when cells within the same basin show broad mRNA distributions.

Iterative update of switching parameters. The switching parameters $\alpha_{c,i}$ are updated at each iteration of the main loop. For a given cell c at time t_j:

1. For each gene i, we compute $P_{c,c',i}(t_{j+1}; \alpha)$ using (12) at a grid of candidate values $\alpha \in \{1/n_{\mathrm{pas}}, 2/n_{\mathrm{pas}}, \ldots, 1\}$.
2. We evaluate $k^{\theta}_{\mathrm{on},i}\big(P_{c,c'}(t_{j+1}; \alpha)\big)$ for each candidate α under the current GRN θ.
3. We set $\alpha_{c,i}$ to the smallest α such that the predicted burst rate is closer to the target mode $\mathbf{k}_{z_{c'},i}$ than to the source mode $\mathbf{k}_{z_c,i}$, i.e. the earliest time at which the trajectory commits to the new basin.

Genes for which the difference between source and target burst rates is very small (below a threshold τ) are left unchanged, as their switching time cannot be reliably identified from the data. This update scheme ensures that $\alpha_{c,i}$ progressively encodes the effective transition timing as the GRN estimate improves.

Remark. The interpolation described above can be interpreted as an approximate bridge for the mechanistic model (2): by conditioning on observed mRNA states at both endpoints and integrating the protein equation with a single mode switch at time $\alpha\Delta t$, we obtain a deterministic approximation of the most probable protein path between the two observations. The iterative update of α refines this approximation alongside the GRN.

B Link with other model learning approaches

We briefly explain how CardamomOT fits into the broader family of Schrödinger- and OT-based trajectory inference methods, and how it relates to Reference Fitting [25] and Wasserstein Lagrangian Flows [24].

B.1 The Schrödinger problem as a template for trajectory inference

Let Q be a reference path measure on gene expression trajectories over $[t_1, t_N]$. Given empirical marginals $\{\hat{P}_{t_j}\}_{j=1}^N$ from a time-stamped scRNA-seq dataset, the (multimarginal) Schrödinger problem seeks the path measure closest to Q that matches the observed marginals:

$$P^* := \underset{P:\, P_{t_j} = \hat{P}_{t_j}, \forall j}{\arg\min} \; \mathbf{H}(P \mid Q), \tag{13}$$

where $\mathbf{H}(P \mid Q)$ is the relative entropy (KL divergence) of P with respect to Q. When $Q = W^\sigma$ is a Brownian motion with diffusion σ, (13) reduces to the entropic OT problem used in [15]: its solution P^* describes the most likely cellular trajectories under a free-diffusion prior. The two-marginal version between consecutive timepoints yields the classical Sinkhorn algorithm. In the diffusion setting, (13) can be reformulated as a variational problem with Lagrangian

$$P^* = \underset{P:\, P_{t_j} = \hat{P}_{t_j}}{\arg\min} \int_0^T \mathcal{L}(x(t), \dot{x}(t))\, dP_t, \qquad \mathcal{L}(x, v) = \frac{1}{2}\|v\|^2, \tag{14}$$

which is the dynamic Benamou–Brénier formulation of OT [22]. The choice of reference process Q thus encodes a prior on cellular dynamics.

B.2 Reference Fitting and Wasserstein Lagrangian Flows

Reference Fitting (RF). Reference Fitting [25] replaces the Brownian reference by a linear Ornstein–Uhlenbeck (OU) process U^θ, whose drift $A^\theta x$ is parametrized by a network matrix θ. The corresponding problem reads

$$\theta^* = \arg\min_\theta \; \min_{P: P_{t_j} = \hat{P}_{t_j}} \mathbf{H}(P \mid U^\theta), \tag{15}$$

where the inner minimization is again a Schrödinger problem (solved pairwise with Gaussian Sinkhorn), and the outer minimization fits the linear drift. This restricts the dynamics to linear systems and does not capture nonlinear GRN behaviour.

Wasserstein Lagrangian Flows (WLF). Wasserstein Lagrangian Flows [24] instead modify the Lagrangian by adding a potential energy term $V(x)$:

$$P^* = \underset{P: P_{t_j} = \hat{P}_{t_j}}{\arg\min} \int_0^T \mathcal{L}(x(t), \dot{x}(t))\, dP_t, \qquad \mathcal{L}(x, v) = \frac{1}{2}\|v\|^2 + V(x), \tag{16}$$

where V acts as a prior on the energy landscape of cells. All components (time-varying measures and drift) are learned jointly through neural networks, providing a flexible but computationally demanding framework.

B.3 CardamomOT as a mechanistic Schrödinger problem

In CardamomOT, the reference process is the GRN-driven model (2), which belongs to the general class of Piecewise Deterministic Markov Processes (PDMPs). This nonlinear, non-diffusive process induces a Markovian path measure on the basin space via the phenomenological model introduced in Section 1.1: the basin label $z_c(t)$ follows a Markov chain whose transition rates are determined by θ through a complex relationship involving the computation of mean first passage times between basins for protein dynamics, as explored in detail in previous work [44]. We denote this path measure by Q^θ.
The Schrödinger problem on the basin space thus becomes

$$\theta^* = \arg\min_\theta \min_{Z:Z_{t_j}=\hat{Z}_{t_j}} \mathbf{H}(Z \mid Q^\theta), \tag{17}$$

where $\hat{Z}_{t_j}$ is the empirical distribution of basin labels at time t_j, obtained in the preprocessing step. This is structurally analogous to (15), with the linear reference U^θ replaced by the mechanistic Markov chain Q^θ on the space of basins, whose rates depend on the non-linear underlying dynamics on proteins.

B.4 The EM loop as an iterative minimizer of (17)

The alternating optimization in CardamomOT (Steps 1–3 of Section 1.2.2) can be viewed as a coordinate-descent algorithm for (17).
Minimizing over Z with fixed θ (Step 1) corresponds to computing a Schrödinger bridge with a structured reference process Q^θ [62], which provides a mechanistic justification for the cost in (8). Indeed, basin dynamics can be viewed as the evolution of the effective activation state $k^\theta_{\text{on}}(P(t))$, and are therefore driven by the underlying protein dynamics. In the metastable regime, where proteins follow the deterministic limit within each basin and undergo rare stochastic transitions between basins, this induces an effective dynamics on the space of basins whose associated action yields the cost in (8).
Minimizing over θ with fixed Z (Step 2) is a regression step that fits the GRN to basin transitions, equivalent to maximizing a path likelihood for Q^θ along sampled trajectories. Step 3 updates the empirical marginals $\hat{Z}_{t_j}$ by refining basin labels consistently with the current θ.
Together, these steps define an iterative scheme converging toward a stationary point of (17), analogous to Iterative Proportional Fitting (Sinkhorn) but with a mechanistic reference process on the latent basin space.

B.5 Summary

Method	Reference process Q	Lagrangian modification	Ambient space
WOT / Sinkhorn [15]	Brownian motion	None	mRNA
Reference Fitting [25]	Linear OU	None	mRNA
WLF [24]	Brownian motion	$V(x)$	mRNA
CardamomOT	**Non-linear PDMP 2**	**None**	**basin**

Table 1: Positioning of CardamomOT within the Schrödinger/variational OT family. The key distinction is the use of a mechanistic, non-diffusive reference process defined on the latent basin space rather than on mRNA space.

In summary, CardamomOT can be seen as a mechanistic generalization of Reference Fitting, using a nonlinear GRN-driven PDMP as reference instead of a linear SDE. The connection to WLF arises through the Lagrangian interpretation: the mechanistic GRN effectively defines an

energy landscape that constrains inferred trajectories, but here this landscape is derived from biophysical modeling rather than learned from data.

Code availability

The algorithm CardamomOT, as well as the codes for generating the figures of this article, are available on https://github.com/eliasventre/CardamomOT

Acknowledgment

We would like to especially thank Olivier Gandrillon for sharing the preprocessed dataset obtained from Kameneva et al. [40], critical reading of the manuscript and substantial help for testing and improving the Python package CardamomOT. We also thank all members of the Inria COMPO team, and of the Southrock consortium, for providing such stimulating working environment.

References

1. Mojtahedi M, Skupin A, Zhou J, Castano IG, Leong-Quong RY, Chang H, Trachana K, Giuliani A, and Huang S. Cell Fate Decision as High-Dimensional Critical State Transition. PLoS Biol 2016; 14:e2000640
2. Richard A, Boullu L, Herbach U, Bonnafoux A, Morin V, Vallin E, Guillemin A, Papili Gao N, Gunawan R, Cosette J, Arnaud O, Kupiec JJ, Espinasse T, Gonin-Giraud S, and Gandrillon O. Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process. PLoS Biol 2016; 14:e1002585
3. Semrau S, Goldmann JE, Soumillon M, Mikkelsen TS, Jaenisch R, and Oudenaarden A van. Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. Nat Commun 2017; 8:1–16
4. Stumpf P, Smith R, Lenz M, Schuppert A, Müller FJ, Babtie A, Chan T, Stumpf M, Please C, Howison S, Arai F, and MacArthur B. Stem Cell Differentiation as a Non-Markov Stochastic Process. Cell Systems 2017; 5:268–82
5. Mar JC. The rise of the distributions: why non-normality is important for understanding the transcriptome and beyond. Biophys Rev. 2019 :89–94
6. Coskun AF, Eser U, and Islam S. Cellular identity at the single-cell level. Mol Biosyst 2016; 12:2965–79
7. Sanguinetti G et al. Gene regulatory network inference: an introductory survey. *Gene Regulatory Networks*. Springer, 2019 :1–23
8. Huynh-Thu VA, Irrthum A, Wehenkel L, and Geurts P. Inferring regulatory networks from expression data using tree-based methods. PloS one 2010; 5:e12776
9. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J, et al. SCENIC: single-cell regulatory network inference and clustering. Nature methods 2017; 14:1083–6
10. Huynh-Thu VA and Geurts P. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. Scientific reports 2018; 8:3384
11. Qiu X, Rahimzamani A, Wang L, Ren B, Mao Q, Durham T, McFaline-Figueroa JL, Saunders L, Trapnell C, and Kannan S. Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. Cell systems 2020; 10:265–74
12. Chen W, Guillaume-Gentil O, Rainer PY, Gäbelein CG, Saelens W, Gardeux V, Klaeger A, Dainese R, Zachara M, Zambelli T, et al. Live-seq enables temporal transcriptomic recording of single cells. Nature 2022; 608:733–40
13. Kobayashi-Kirschvink KJ, Comiter CS, Gaddam S, Joren T, Grody EI, Ounadjela JR, Zhang K, Ge B, Kang JW, Xavier RJ, et al. Prediction of single-cell RNA expression profiles in live cells by Raman microscopy with Raman2RNA. Nature biotechnology 2024; 42:1726–34
14. Santambrogio F. Optimal transport for applied mathematicians. Vol. 87. Springer, 2015
15. Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, Gould J, Liu S, Lin S, Berube P, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. Cell 2019; 176:928–43
16. Lavenant H, Zhang S, Kim YH, and Schiebinger G. Towards a mathematical theory of trajectory inference. arXiv preprint arXiv:2102.09204 2021

17. Qiu X, Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, Pogson AN, Hein MY, Min KHJ, Wang L, et al. Mapping transcriptomic vector fields of single cells. Cell 2022; 185:690–711

18. Massri AJ, Greenstreet L, Afanassiev A, Berrio A, Wray GA, Schiebinger G, and McClay DR. Developmental single-cell transcriptomics in the Lytechinus variegatus sea urchin embryo. Development 2021; 148:dev198614

19. Tong A, Malkin N, Fatras K, Atanackovic L, Zhang Y, Huguet G, Wolf G, and Bengio Y. Simulation-free schr\" odinger bridges via score and flow matching. arXiv preprint arXiv:2307.03672 2023

20. Rimawi-Fine NE, Stecklov A, Nelson L, Blanchette M, Tong A, Zhang SY, and Atanackovic L. Simulation-free Structure Learning for Stochastic Dynamics. arXiv preprint arXiv:2510.16656 2025

21. Bertin P, Viviano JD, Tejada-Lapuerta A, Wang W, Bauer S, Theis FJ, and Bengio Y. A scalable gene network model of regulatory dynamics in single cells. arXiv preprint arXiv:2503.20027 2025

22. Léonard C. From the Schrödinger problem to the Monge–Kantorovich problem. Journal of Functional Analysis 2012; 262:1879–920

23. Fournié C, Ventre E, Herbach U, Baradat A, Gandrillon O, and Crauste F. Cell Trajectory Inference based on Schrödinger Problem and a Mechanistic Model of Stochastic Gene Expression. bioRxiv 2025 :2025–11

24. Neklyudov K, Brekelmans R, Tong A, Atanackovic L, Liu Q, and Makhzani A. A computational framework for solving wasserstein lagrangian flows. arXiv preprint arXiv:2310.10649 2023

25. Zhang SY. Joint trajectory and network inference via reference fitting. arXiv preprint arXiv:2409.06879 2024

26. Lamoline F, Haasler I, Karlsson J, Gonçalves J, and Aalto A. Dynamic gene regulatory network inference from single-cell data using optimal transport. Bioinformatics 2025; 41:btaf394

27. Guan V, Janssen J, Rahmani H, Warren A, Zhang S, Robeva E, and Schiebinger G. Identifying Drift, Diffusion, and Causal Structure from Temporal Snapshots. arXiv preprint arXiv:2410.22729 2024

28. Ahmad R and Budnik B. A review of the current state of single-cell proteomics and future perspective. Analytical and Bioanalytical Chemistry 2023; 415:6889–99

29. Zong C, So Lh, Sepúlveda LA, Skinner SO, and Golding I. Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene. Molecular systems biology 2010; 6:440

30. Ochiai H, Sugawara T, Sakuma T, and Yamamoto T. Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. Scientific reports 2014; 4:1–9

31. Herbach U. Modélisation stochastique de l'expression des gènes et inférence de réseaux de régulation. PhD thesis. Université de Lyon, 2018

32. Mar JC. The rise of the distributions: why non-normality is important for understanding the transcriptome and beyond. Biophysical reviews 2019; 11:89–94

33. Ventre E, Herbach U, Espinasse T, Benoit G, and Gandrillon O. One model fits all: combining inference and simulation of gene regulatory networks. PLoS Computational Biology 2023; 19:e1010962

34. Herbach U, Bonnaffoux A, Espinasse T, and Gandrillon O. Inferring gene regulatory networks from single-cell data: a mechanistic approach. BMC Systems Biology 2017; 11:105

35. Bonnaffoux A, Herbach U, Richard A, Guillemin A, Gonin-Giraud S, Gros PA, and Gandrillon O. WASABI: a dynamic iterative framework for gene regulatory network inference. BMC bioinformatics 2019; 20:1–19

36. Ventre E. Reverse engineering of a mechanistic model of gene expression using metastability and temporal dynamics. In Silico Biology 2022; 14:89–113

37. Fang M, Gorin G, and Pachter L. Trajectory inference from single-cell genomics data with a process time model. PLoS computational biology 2025; 21:e1012752

38. Peccoud J and Ycart B. Markovian modeling of gene-product synthesis. Theoretical population biology 1995; 48:222–34

39. Gandrillon O. Inferring and simulating a gene regulatory network for the sympathoadrenal differentiation from single-cell transcriptomics in human. F1000 Research 2025 :14:910

40. Kameneva P, Artemov AV, Kastriti ME, Faure L, Olsen TK, Otte J, Erickson A, Semsch B, Andersson ER, Ratz M, et al. Single-cell transcriptomics of human embryos identifies multiple sympathoblast lineages with potential implications for neuroblastoma origin. Nature genetics 2021; 53:694–706

41. Li GW and Xie XS. Central dogma at the single-molecule level in living cells. Nature 2011; 475:308–15

42. Rodriguez J and Larson DR. Transcription in Living Cells: Molecular Mechanisms of Bursting. Annu Rev Biochem 2020; 89:189–212

43. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, and Naef F. Mammalian genes are transcribed with widely different bursting kinetics. science 2011; 332:472–4

44. Ventre E, Espinasse T, Brehier CE, Calvez V, Lepoutre T, and Gandrillon O. Reduction of a stochastic model of gene expression: Lagrangian dynamics gives access to basins of attraction as cell types and metastability. bioRxiv 2020

45. Yeo GHT, Saksena SD, and Gifford DK. Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. Nature communications 2021; 12:3222

46. Huizing GJ, Samaran J, Capocefalo D, Audit A, Peyré G, and Cantini L. STORIES: learning cell fate landscapes from spatial transcriptomics using optimal transport. Nature Methods 2025 :1–10

47. Herbach U. Harissa: tools for mechanistic gene network inference from single-cell data. 2024

48. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, and Satija R. Comprehensive integration of single-cell data. cell 2019; 177:1888–902

49. Papili Gao N, Ud-Dean SM, Gandrillon O, and Gunawan R. SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. Bioinformatics 2018; 34:258–66

50. Chan TE, Stumpf MP, and Babtie AC. Gene regulatory network inference from single-cell data using multivariate information measures. Cell systems 2017; 5:251–67

51. Forrow A and Schiebinger G. LineageOT is a unified framework for lineage tracing and trajectory inference. Nature communications 2021; 12:4940

52. Bird A. DNA methylation patterns and epigenetic memory. Genes & development 2002; 16:6–21

53. Türei D, Korcsmáros T, and Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nature methods 2016; 13:966–7

54. Ruscone M, Tsirvouli E, Checcoli A, Turei D, Barillot E, Saez-Rodriguez J, Martignetti L, Flobak Å, and Calzone L. NeKo: a tool for automatic network construction from prior knowledge. PLOS Computational Biology 2025; 21:e1013300

55. Ferrell Jr JE. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. Current opinion in cell biology 2002; 14:140–8

56. Silva LGS da, Yvinec R, Prata G, Dhar V, Reinitz J, and Ramos AF. Two-state stochastic model of in vivo observations of transcriptional bursts. Brazilian Journal of Physics 2025; 55:150

57. Dar RD, Razooky BS, Singh A, Trimeloni TV, McCollum JM, Cox CD, Simpson ML, and Weinberger LS. Transcriptional burst frequency and burst size are equally modulated across the human genome. Proceedings of the National Academy of Sciences 2012; 109:17454–9

58. Ismail B. Exploring Logistic Functions as Robust Alternatives to Hill Functions in Genetic Network Modeling. arXiv preprint arXiv:2512.14325 2025

59. Huang Z, Luo S, Wang Z, Zhang Z, Jiang B, Nie Q, and Zhang J. Deep learning linking mechanistic models to single-cell transcriptomics data reveals transcriptional bursting in response to DNA damage. eLife 2026; 13:RP100623

60. Ventre E, Forrow A, Gadhiwala N, Chakraborty P, Angel O, and Schiebinger G. Trajectory inference for a branching SDE model of cell differentiation. arXiv preprint arXiv:2307.07687 2023

61. Chizat L, Peyré G, Schmitzer B, and Vialard FX. Scaling algorithms for unbalanced optimal transport problems. Mathematics of computation 2018; 87:2563–609

62. Baradat A and Ventre E. Convergence of the Sinkhorn algorithm when the Schrödinger problem has no solution. *Annales de la Faculté des sciences de Toulouse: Mathématiques.* Vol. 33. 5. 2024 :1297–371

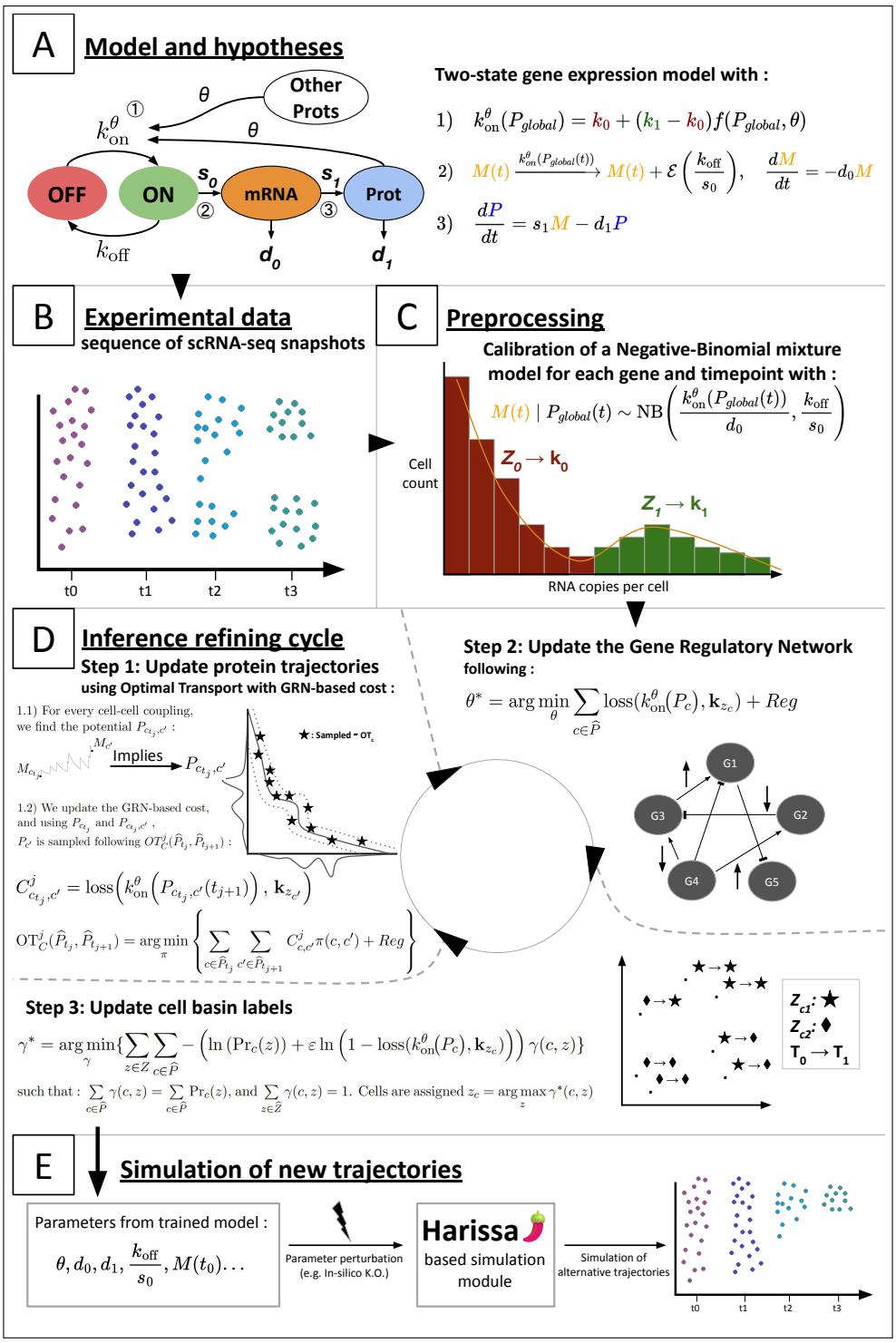

Figure 1: CardamomOT infers gene regulatory networks (GRNs) and protein trajectories from temporal snapshots of scRNA-seq data using a mechanistic optimal transport framework. (A) The two-state gene expression model underlying the method, in which promoter switching rates are modulated by a GRN matrix θ. (B) Input data consisting of independent scRNA-seq snapshots at successive timepoints. (C) Preprocessing step: a Negative Binomial mixture model is calibrated for each gene and timepoint, clustering cells into discrete promoter-activity basins. (D) Iterative EM-like inference cycle alternating between protein trajectory reconstruction via GRN-driven optimal transport (Step 1), GRN update by regression on reconstructed protein values (Step 2), and basin label refinement (Step 3). (E) The calibrated model is used as a generative module for simulating alternative trajectories and in silico perturbation experiments.

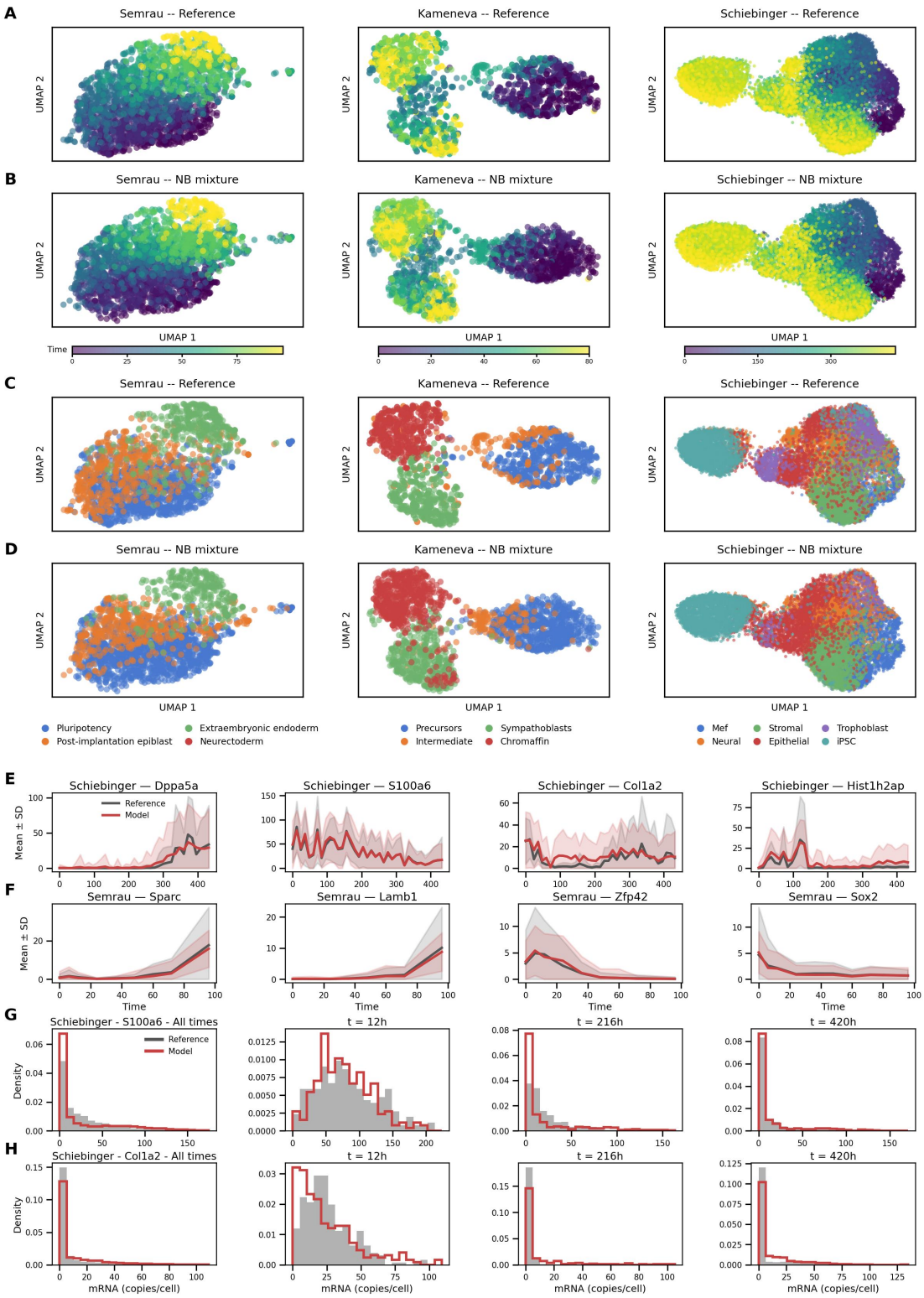

Figure 2: (A–D) UMAP embeddings of reference scRNA-seq data (A, C) and data simulated from the CardamomOT NB mixture model (B, D), colored by timepoint (A–B) and cell type (C–D), for the Semrau, Kameneva and Schiebinger datasets. (E–F) Temporal profiles of mean ± SD gene expression for selected key regulators, comparing reference data (gray) and model simulations (red), for the Schiebinger (E) and Semrau (F) datasets. (G–H) Marginal mRNA distributions at selected timepoints for S100a6 (G) and Col1a2 (H) in the Schiebinger dataset, showing close agreement between model and reference data across all timepoints.

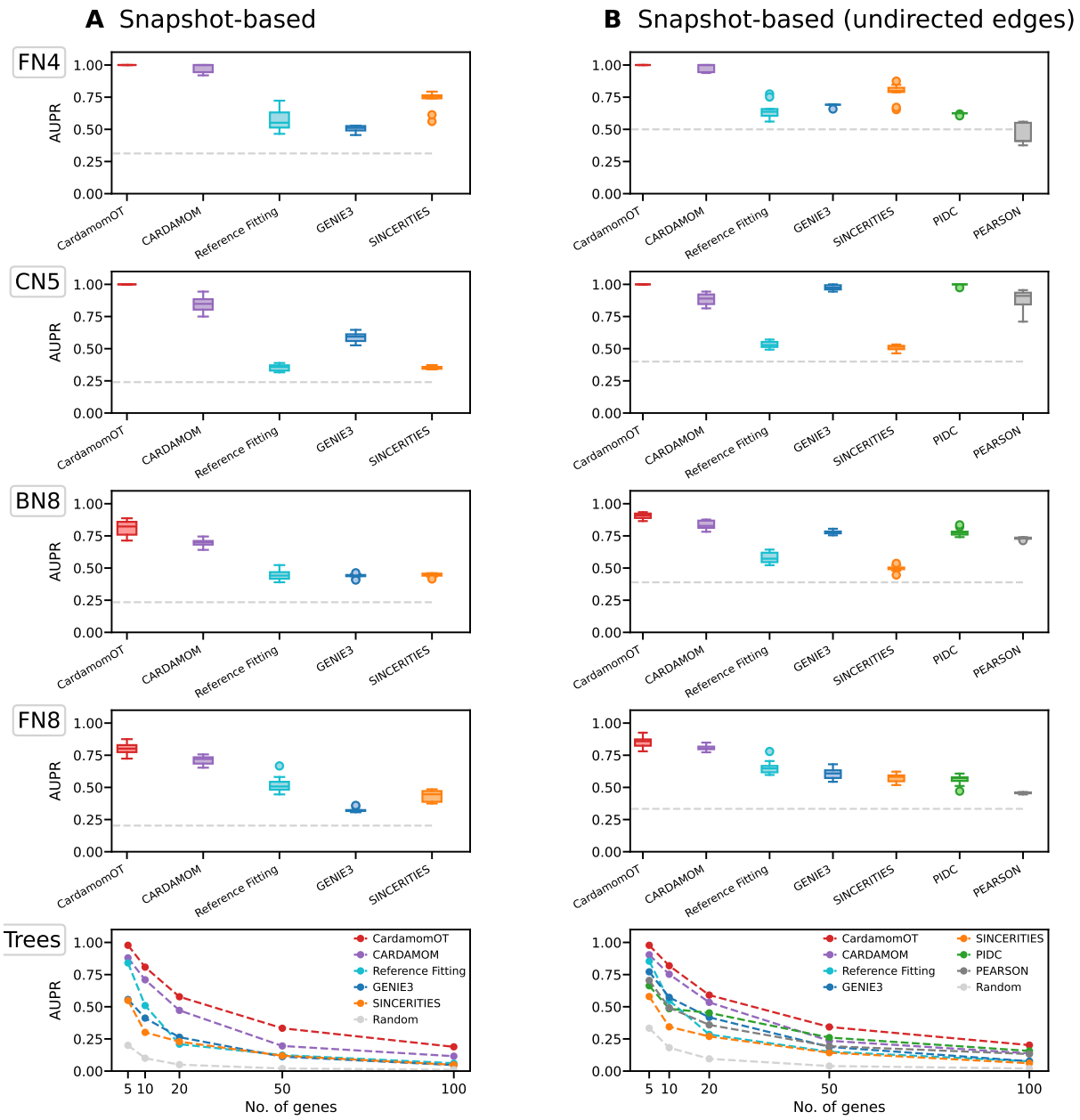

Figure 3: CardamomOT outperforms competing methods on GRN inference benchmarks. Performance is measured by the Area Under the Precision-Recall curve (AUPR), averaged over 10 independent simulations per network. The dashed gray line indicates the random baseline. (A) Directed GRN inference from temporal snapshots on four benchmark network topologies (FN4, CN5, BN8, FN8) and tree networks of increasing size (5–100 genes). (B) Same comparison for undirected edge inference, including additional correlation-based methods (PIDC, Pearson).

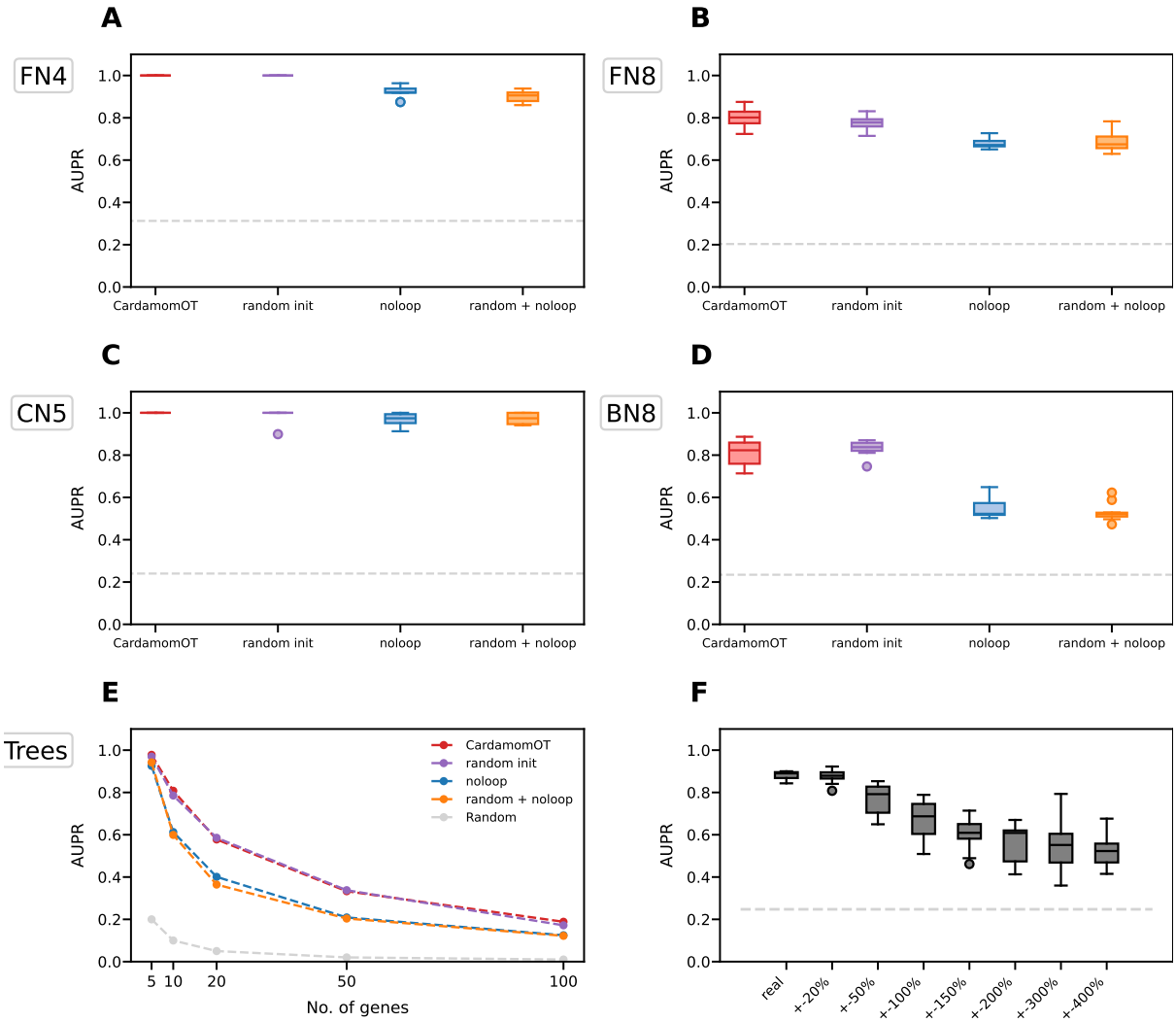

Figure 4: (A–E) Directed GRN inference performance (AUPR) comparing CardamomOT against three ablated variants: random OT initialization (random init), removal of the loop constraint (noloop), and their combination (random + noloop), on four benchmark networks and tree topologies of increasing size. (F) Robustness to uncertainty in protein degradation rates d_1: AUPR averaged across all four benchmark networks when input degradation rates are perturbed from ±20% to ±400% of their true values, compared to the performance obtained with the real rates.

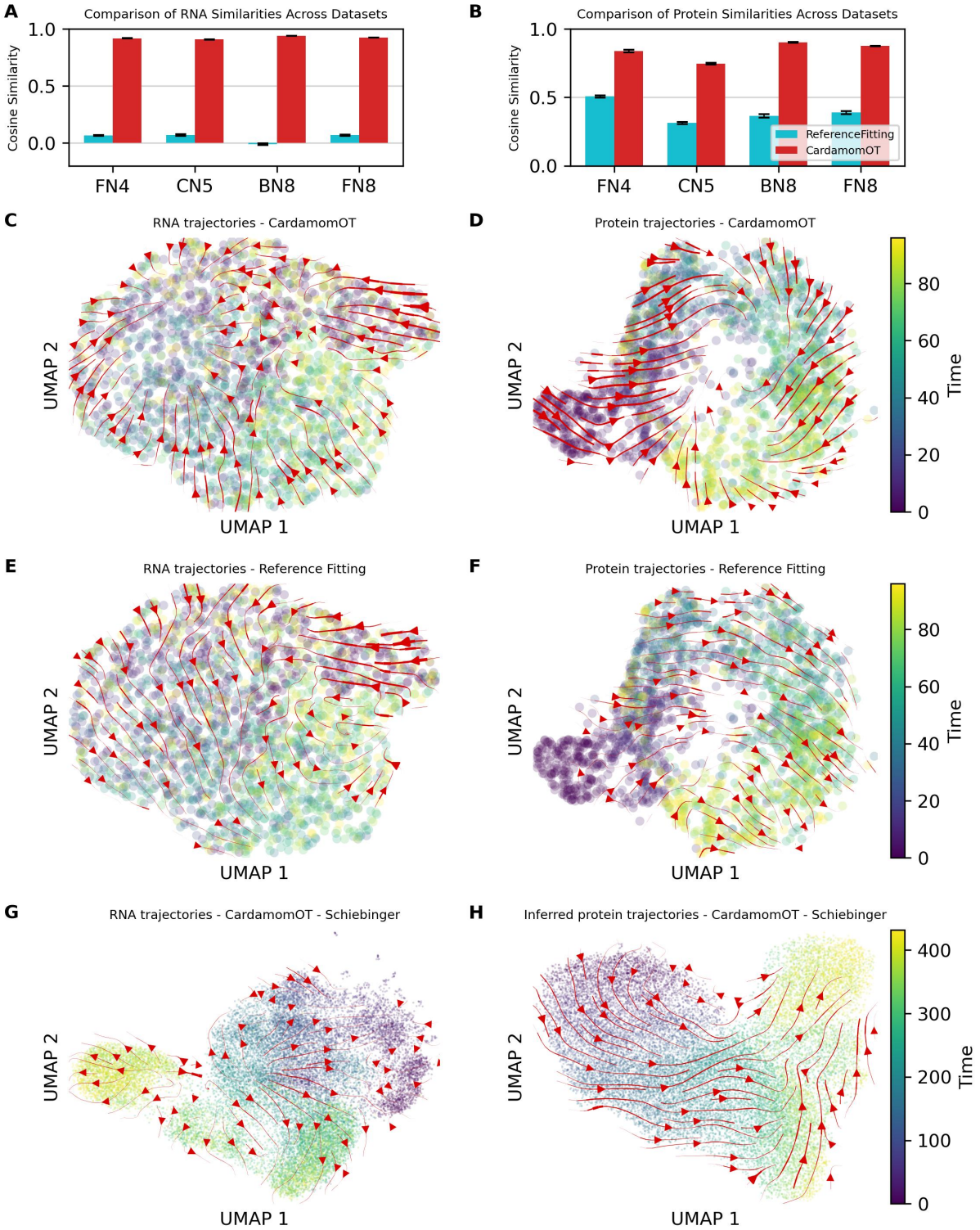

Figure 5: Evaluation of the mRNA and protein velocity fields reconstruction on cross-sectional data for CardamomOT and Reference Fitting. (A–B) Weighted cosine similarity between the velocity fields predicted by CardamomOT and Reference Fitting and the mechanistic ground-truth velocity field, for mRNA (A) and protein (B), on four benchmark networks (BN8, FN8, CN5, FN4). Bars show the mean over 5 independent simulations; error bars indicate the standard error of the mean. (C–F) UMAP embeddings of RNA (C,E) and protein (D,F) expression of CN5 benchmark cells, overlaid with velocity stream plots from CardamomOT (C–D) and Reference Fitting (E–F). (G–H) UMAP of RNA (G) and inferred protein (H) data from the Schiebinger dataset, with CardamomOT velocity stream plots.

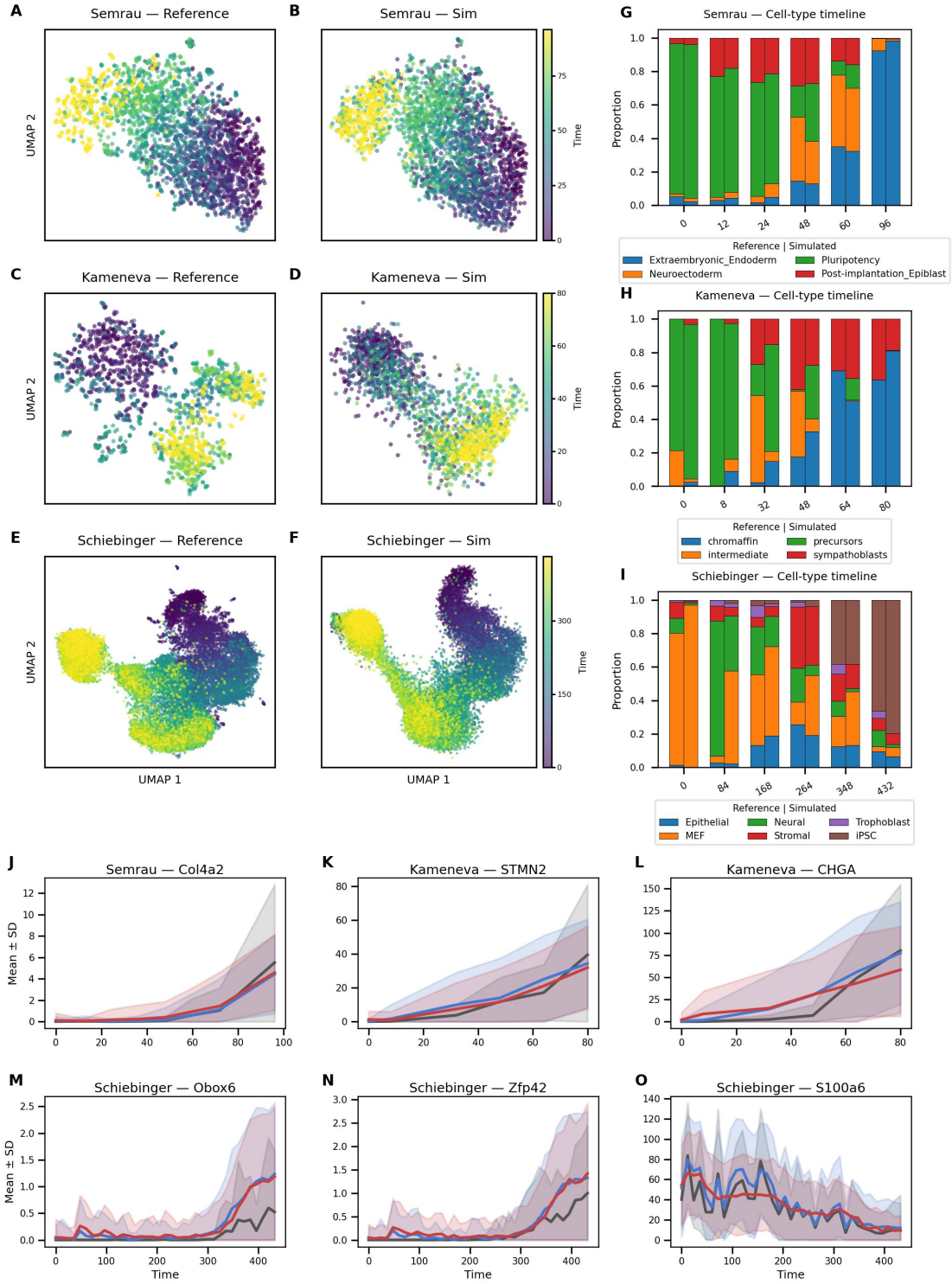

Figure 6: CardamomOT performs well as a generative model of protein and mRNA distributions. UMAPs of reference (A,C,E) and simulated (B,D,F) data for 3 differentiation datasets. Cell type composition for reference and simulated data across timepoints (G–I). Cell type composition was inferred for simulated data by a random forest model trained on reference data. (J–O) Temporal profiles of mean ± SD gene expression for selected key regulators, comparing reference data (gray), Negative Binomial reconstruction (blue) and model simulations (red), for Dnmt3a in the Semrau dataset (J), STMN2 and CHGA in the Kameneva dataset (K–L), and Obox6, Zfp42 and S100a6 in the Schiebinger dataset (M–O).

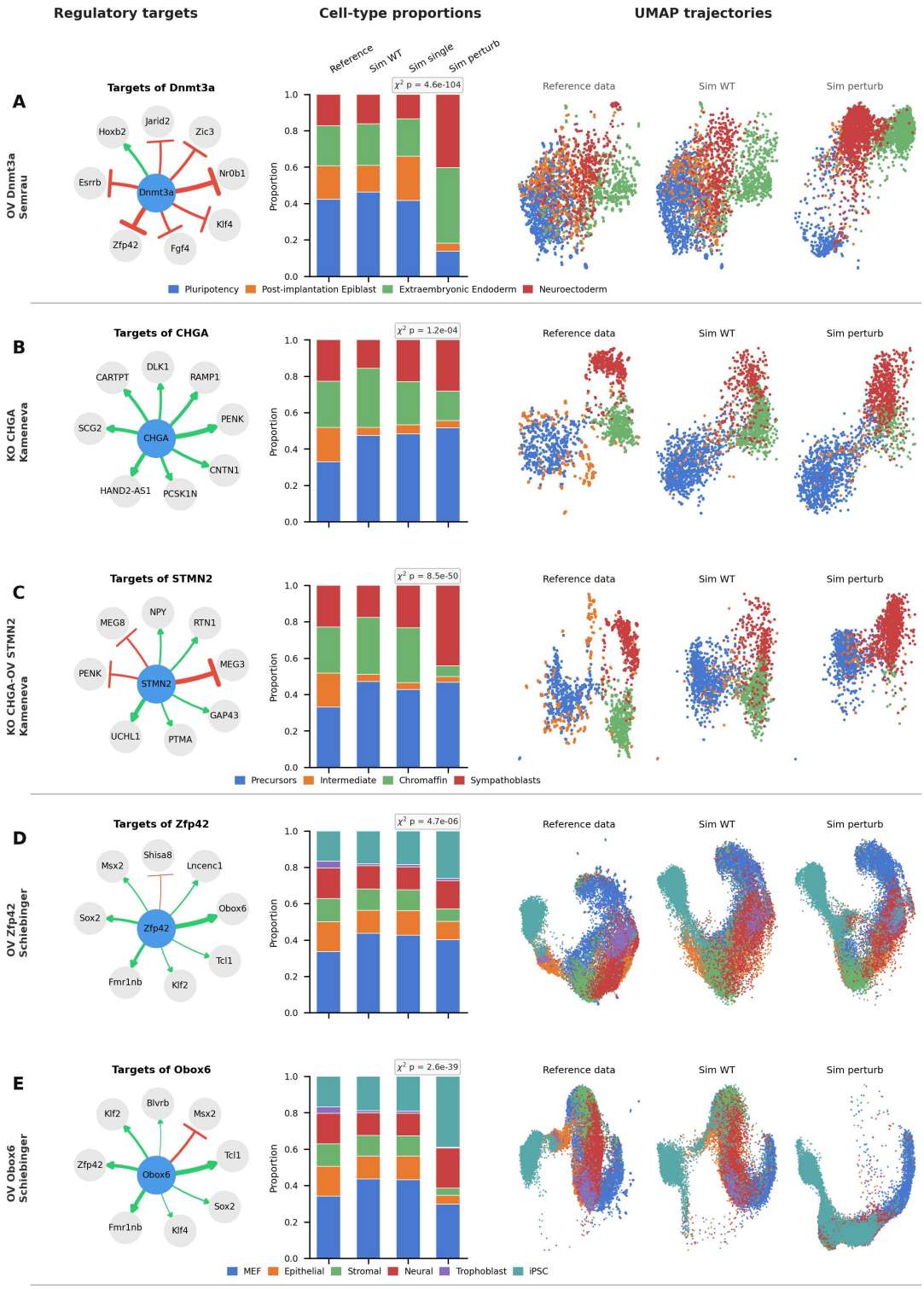

Figure 7: In silico perturbations using the inferred GRNs recapitulate experimentally observed effects. For each perturbation, three panels are shown: (left) the regulatory network centered on the perturbed gene, showing its top predicted targets with activation (green) and inhibition (red) edges; (middle) cell type proportions comparing reference data, wild-type simulation (Sim WT), single-cell simulation with only the perturbed gene modified (Sim single) and perturbed simulation (Sim perturb); (right) UMAP trajectories for reference data, wild-type and perturbed simulations, colored by cell type. Statistical differences between Sim perturb and Sim single were assessed using a χ^2 test on reconstructed cell counts from proportions; all comparisons yielded $p < 10^{-3}$.

4 Supplementary information

Table S1: Comparison of parameters for different models

Parameter	FN4		CN5		BN8		FN8	
	Ref	Inf	Ref	Inf	Ref	Inf	Ref	Inf
k_0/d_0	0	0.194±0.211	0	0.263±0.120	0	0.138±0.117	0	0.076±0.046
k_1/d_0	10	8.044±0.888	4	3.315±0.401	8	5.599±1.870	5	3.474±1.445
k_{off}/s_0	0.02	0.018±0.001	0.02	0.019±0.001	0.02	0.017±0.003	0.02	0.019±0.005
% of correctly attributed modes		92,9		85,2		91,3		92,6

Table S2: Related to Figure 3. Average runtime for inferring a network from datasets simulated with tree-like networks for which the results of the inference are represented in Figure 3, for the six algorithms that are used in the benchmark. Timings measured on a 16-GB RAM, 2.4 GHz Intel Core i5 computer.

Runtime	PIDC	SINCERITIES	CARDAMOM	Reference Fitting	**CardamomOT**	GENIE3
5 genes	0.02 s	0.01 s	0.10 s	2.71 s	4.71 s	9.02 s
10 genes	0.04 s	0.03 s	0.29 s	3.45 s	8.92 s	17.27 s
20 genes	0.06 s	0.06 s	0.72 s	4.05 s	17.18 s	31.17 s
50 genes	0.18 s	0.24 s	2.98 s	5.91 s	28.57 s	83.18 s
100 genes	0.75 s	0.77 s	9.14 s	10.06 s	55.08 s	159.42 s

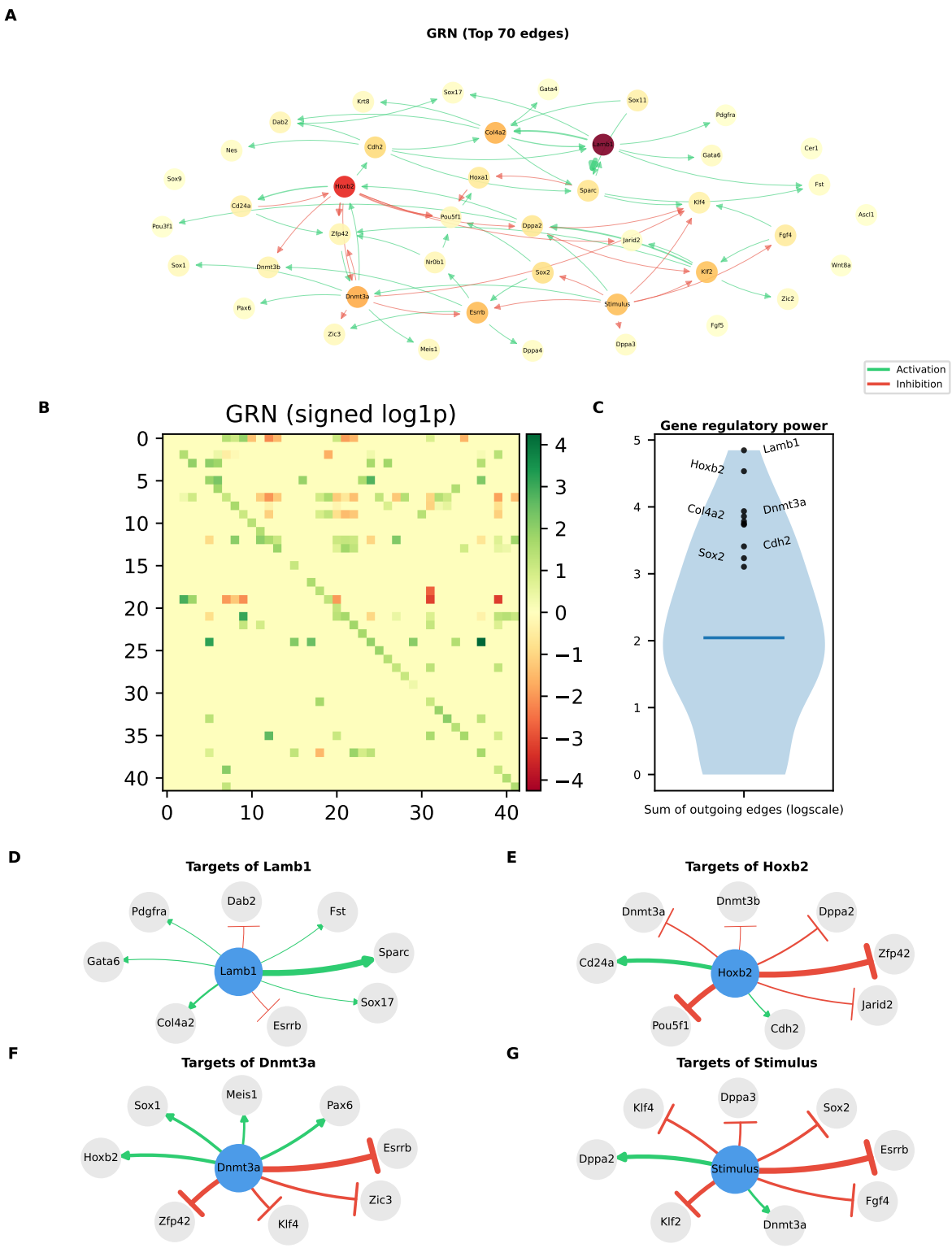

Figure S1: Representation of the inferred gene regulatory network for the Semrau dataset. (A) Graph showing the top 70 edges of the GRN by absolute regulation strength, limited to a max of 5 incoming regulating edges for visualization. (B) Signed log1p transformed full GRN matrix representing the strength of all genes on each other, with lines being regulators and columns being targets. (C) Violin plot of dataset genes ranked by the sum of all their outgoing edges, with top 10 genes labeled. (D–G) Small GRN representation of the top regulators *Hoxb2*, *Lamb1*, *Col4a2* and *Dnmt3a*, showing their top 8 regulatory targets.

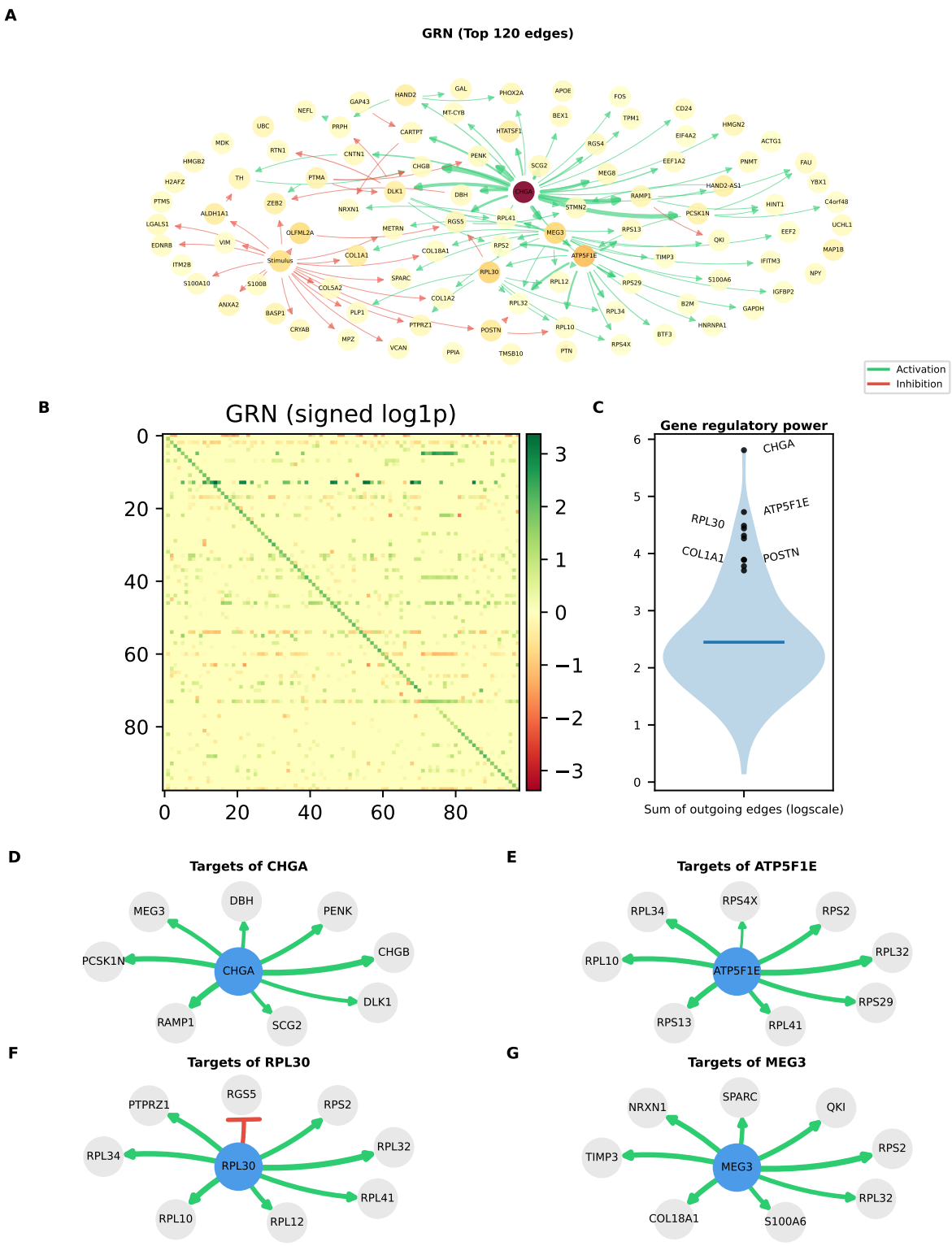

Figure S2: Representation of the inferred gene regulatory network for the Kameneva dataset. (A) Graph showing the top 80 edges of the GRN by absolute regulation strength, limited to a max of 5 incoming regulating edges for visualization. (B) Signed log1p transformed full GRN matrix representing the strength of all genes on each other, with lines being regulators and columns being targets. (C) Violin plot of dataset genes ranked by the sum of all their outgoing edges, with top 10 genes labeled. (D–G) Small GRN representation of the top regulators *RPL30*, *CHGA*, *MEG3* and *ATP5F1E*, showing their top 8 regulatory targets.

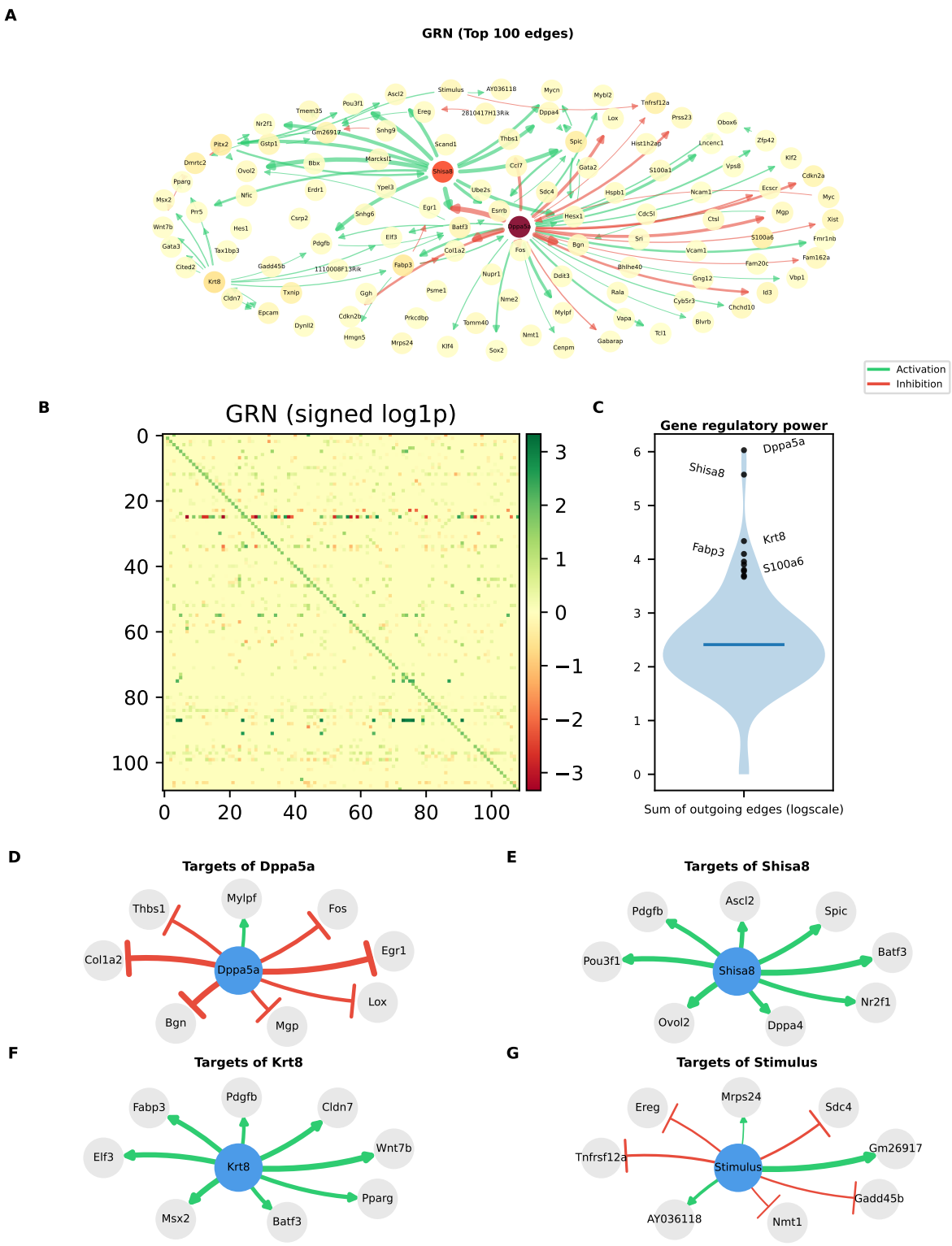

Figure S3: Representation of the inferred gene regulatory network for the Schiebinger dataset. (A) Graph showing the top 80 edges of the GRN by absolute regulation strength, limited to a max of 5 incoming regulating edges for visualization. (B) Signed log1p transformed full GRN matrix representing the strength of all genes on each other, with lines being regulators and columns being targets. (C) Violin plot of dataset genes ranked by the sum of all their outgoing edges, with top 10 genes labeled. (D–G) Small GRN representation of the top regulators *Dppa5a*, *Shisa8*, *Fabp3* and *S100a6*, showing their top 8 regulatory targets.

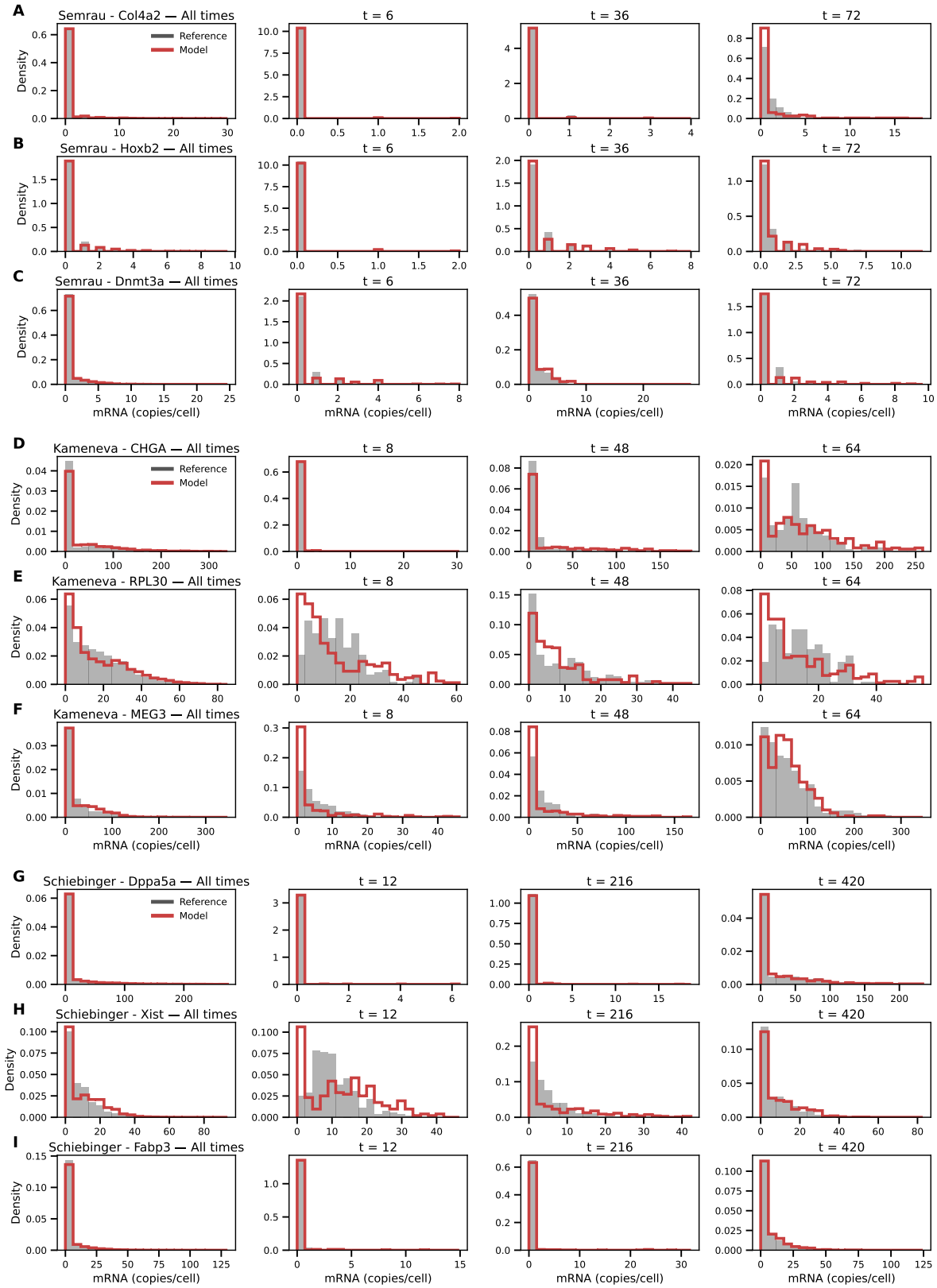

Figure S4: Marginal mRNA distributions for selected genes, showing (leftmost column) all timepoints joined and (remaining columns) three selected timepoints for each gene. Three of the main regulators were selected for the Semrau (A–C), Kameneva (D–F) and Schiebinger (G–I) datasets, namely *Col4a2*, *Hoxb2* and *Dnmt3a* for Semrau, *CHGA*, *RPL30* and *MEG3* for Kameneva, and *Dppa5a*, *Xist* and *Fabp3* for Schiebinger. Marginal distributions are shown for the reference data (gray) and the Negative Binomial mixture model (red).

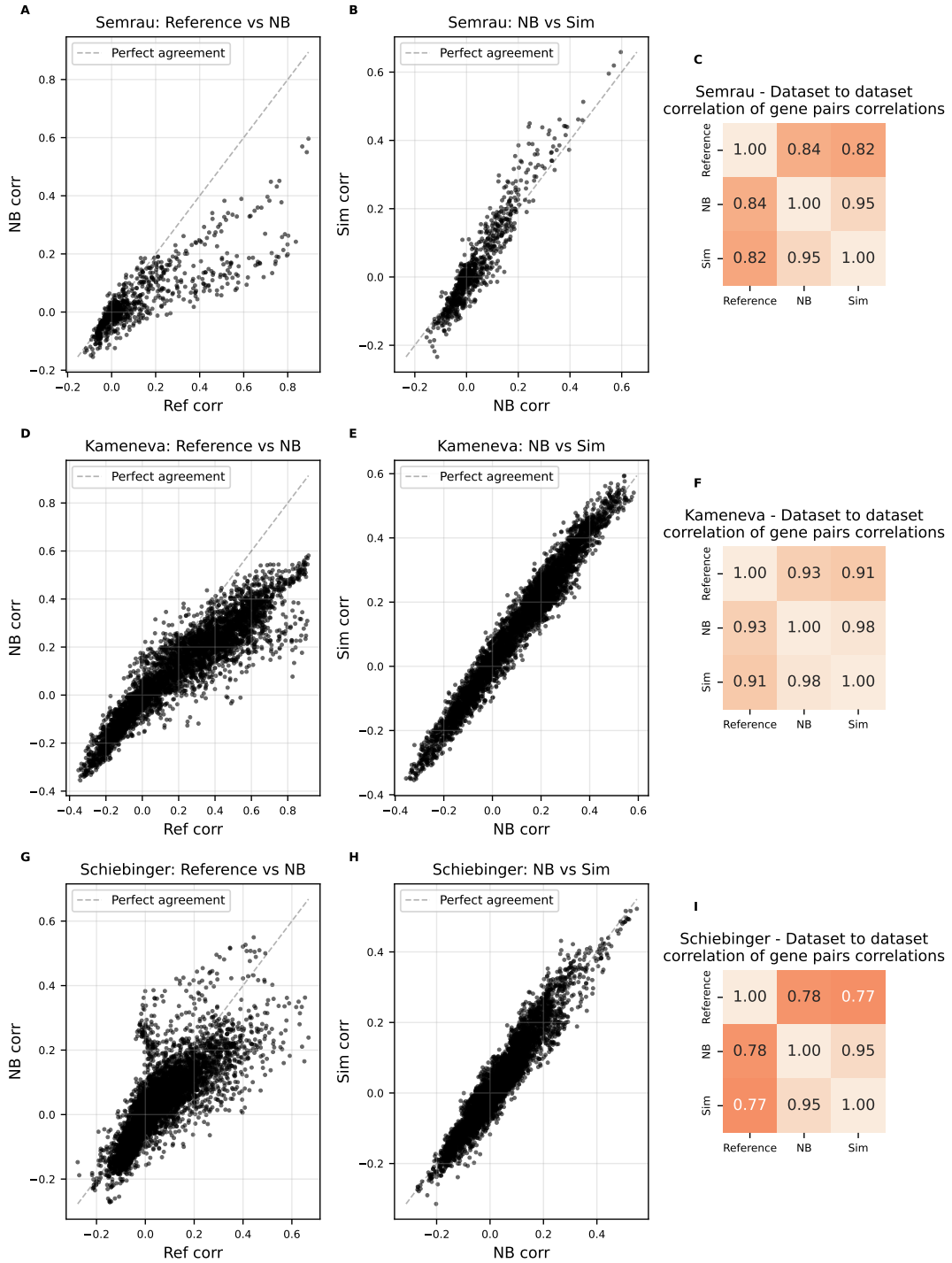

Figure S5: Pairwise gene expression correlation structure across reference, NB mixture and simulated datasets, for the Semrau (A–C), Kameneva (D–F) and Schiebinger (G–I) datasets. (A, D, G) Scatter plots of pairwise gene correlations computed from the reference data versus those computed from the NB mixture model, with the dashed line indicating perfect agreement. (B, E, H) Same comparison between the NB mixture model and the full generative model simulations. (C, F, I) Dataset-to-dataset correlation matrices summarizing the Pearson correlation between the vectors of all pairwise gene correlations across the three datasets (Reference, NB, Sim), providing a global measure of how well each model reproduces the correlation structure of the reference data.

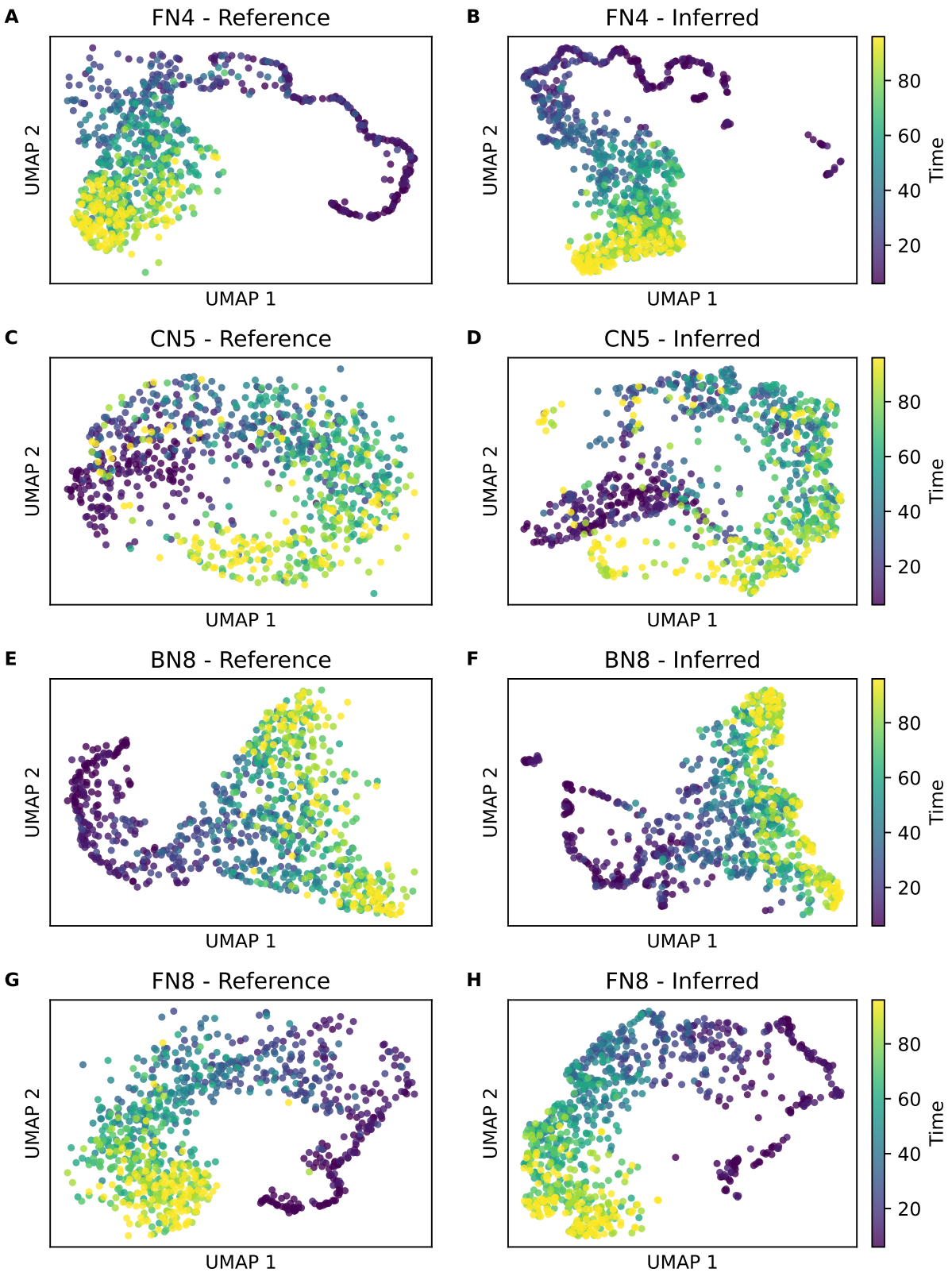

Figure S6: UMAPs of reference and CardamomOT inferred protein trajectories for the (A–B) FN4, (C–D) CN5, (E–F) BN8 and (G–H) FN8 datasets. (A, C, E, G) UMAPs of reference simulated protein trajectories are reproduced by (B, D, F, H) inferred protein trajectories from the reference protein-associated RNA data.

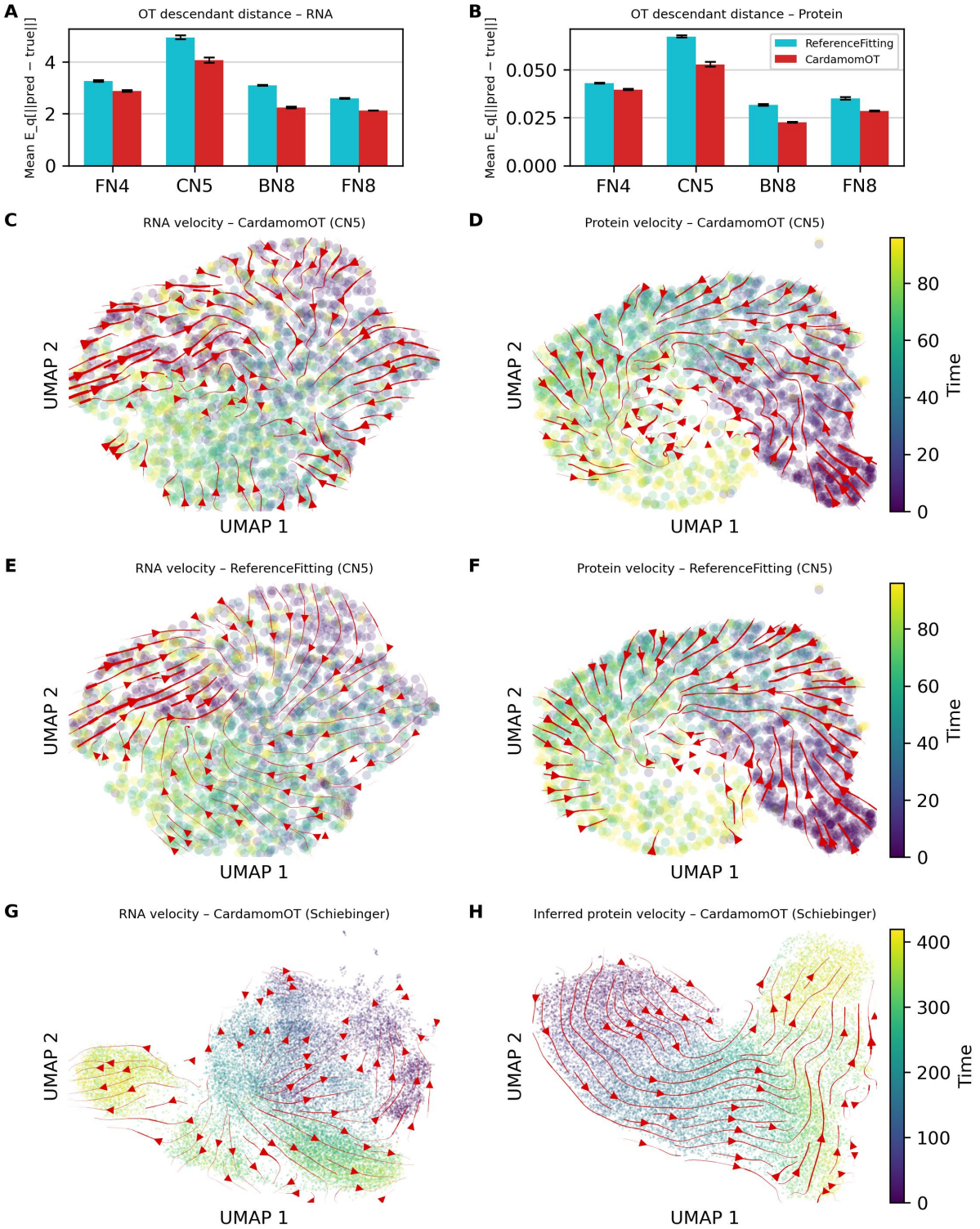

Figure S7: Evaluation of the cell-to-cell coupling accuracy on true trajectory data for CardamomOT and Reference Fitting. (A–B) OT descendant distance for mRNA (A) and protein (B): Bars show the mean over 5 independent simulations of true trajectories per network; error bars indicate the standard error. (C–F) UMAP embeddings of RNA (C,E) and protein (D,F) expression of CN5 benchmark cells simulated as true trajectories, with velocity stream plots derived from the inferred couplings of CardamomOT (C–D) and Reference Fitting (E–F). (G–H) UMAP of RNA (G) and inferred protein (H) data from Schiebinger et al. 2019 [15], with CardamomOT velocity stream plots derived from the inferred trajectories.

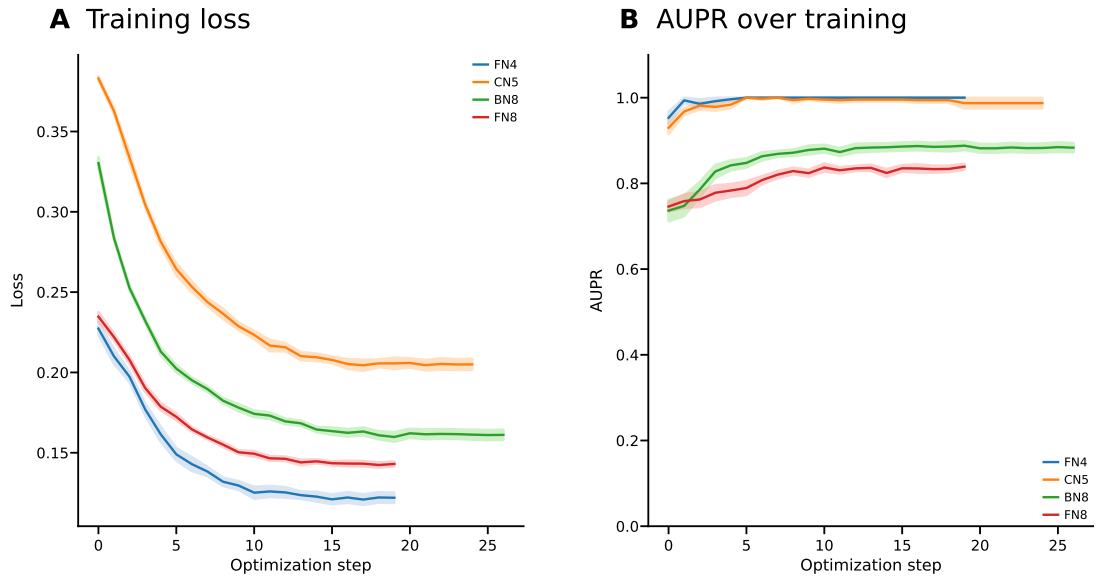

Figure S8: Evolution of the loss function and AUPR score during training for the networks FN4, CN5, BN8, and FN8 used in the Benchmark of Figures 3 and 4. (A) Mean training loss at each optimization step, averaged over 10 independent runs, with shaded bands indicating the uncertainty (standard error). (B) The corresponding evolution of the AUPR score over the same training steps and networks.