Zuse Institute Berlin

Takustr. 7
14195 Berlin
Germany

N. Alexia Raharinirina, Felix Peppert, Max von Kleist, Christof Schütte, Vikram Sunkara[1]

Inferring Gene Regulatory Networks from Single Cell RNA-seq Temporal Snapshot Data Requires Higher Order Moments

Funded by DFG Center of Excellence MATH+ IN-B1

[1] 0000-0002-4940-8344

ZIB Report 20-25 (September 2020)

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30-84185-0
Telefax: +49 30-84185-125

E-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Inferring Gene Regulatory Networks from Single Cell RNA-seq Temporal Snapshot Data Requires Higher Order Moments

N. Alexia Raharinirina[1], Felix Peppert[1], Max von Kleist[3], Christof Schütte[1,2], and Vikram Sunkara[1]

[1]Modelling and Simulation of Complex Processes, Zuse Institute Berlin, Berlin, Germany.
[2]Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany.
[3]MF1 Bioinformatics, Methods Development and Research Infrastructure, Robert Koch Institute, Berlin, Germany
[1]sunkara@mi.fu-berlin.de

September 22, 2020

Abstract

Due to the increase in accessibility and robustness of sequencing technology, single cell RNA-seq (scRNA-seq) data has become abundant. The technology has made significant contributions to discovering novel phenotypes and heterogeneities of cells. Recently, there has been a push for using single– or multiple scRNA-seq snapshots to infer the underlying gene regulatory networks (GRNs) steering the cells' biological functions. To date, this aspiration remains unrealised.

In this paper, we took a bottom-up approach and curated a stochastic two gene interaction model capturing the dynamics of a complete system of genes, mRNAs, and proteins. In the model, the regulation was placed upstream from the mRNA on the gene level. We then inferred the underlying regulatory interactions from only the observation of the mRNA population through time.

We could detect signatures of the regulation by combining information of the mean, covariance, and the skewness of the mRNA counts through time. We also saw that reordering the observations using pseudo-time did not conserve the covariance and skewness of the true time course. The underlying GRN could be captured consistently when we fitted the moments up to degree three; however, this required a computationally expensive non-linear least squares minimisation solver.

There are still major numerical challenges to overcome for inference of GRNs from scRNA-seq data. These challenges entail finding informative summary statistics of the data which capture the critical regulatory information. Furthermore, the statistics have to evolve linearly or piece-wise linearly through time to achieve computational feasibility and scalability.

Keywords: Markov chains, Chemical Master Equation, single cell, RNA sequencing, Time course snapshots, Moment equations.

1 Introduction

There is growing interest to understand the degree to which cell to cell variation in a population contributes to biological processes such as stem cell differentiation and disease progression (1; 2; 3; 4; 5; 6). This heterogeneity of phenotypes is created by various regulatory mechanisms occurring within the cell where the products of gene expression modulate the life-cycle of proteins (e.g, transcription, post-processing, translation, transport, degradation etc.). The emergence of single cell RNA sequencing technology (scRNA-seq), the extraction of the transcriptome of individual cells, has helped immensely in detecting and delineating heterogeneities in cells (4; 7; 8; 9; 10; 11; 12). Furthermore, with advances in machine learning and mRNA metabolic tagging, scRNA-seq has given new insights into cellular development and disease pathogenesis (7; 10; 11; 12; 13; 14).

In light of these advances, the development of inferring the underlying *gene regulatory network* (GRN)–which drives cellular decisions–is lagging behind (15). That is in practice, using scRNA-seq data, we can confidently answer how many cellular phenotypes there are and identify their defining transcriptomic signatures, however, the mechanism by which the transcriptome maintains its phenotype or transitions between phenotypes, is dubious (4; 8).

Cellular function is dependent on the cell's transcriptomic signature, where the proteins translated from the mRNA form signalling pathways; which perform cellular function and then in a feedback loop; regulate the mRNA transcription to then translate proteins (1; 16; 17). The process of a gene affecting the expression of another gene is referred to as *gene regulation*, and the collection of all gene regulatory interactions (e.g. in a cell) forms a gene regulatory network (GRN) (8; 12; 16; 18; 19; 20; 21). Unlike protein-protein interactions, where educts are converted into products, gene regulation interactions are more illusive. A gene regulates another gene through its downstream protein complexes, which affect the rate of transcription of the gene being regulated. That is, gene regulation physically occurs on the DNA level. In particular, "gene A regulates gene B" means that gene A either up-regulates (promotes) or down-regulates (inhibits) the rate of transcription of gene B. The fact that scRNA-seq only captures mRNA, while gene regulation interactions take place up or down-stream from the mRNA, constitutes a major hurdle for inferring GRNs from scRNA-seq data (22; 23).

Recently, a trend has emerged to use multiple temporal scRNA-seq snapshots to capture the underlying GRNs of cells (8; 12; 15; 24; 25; 26; 27). Current GRN inference methods using temporal snapshot data include Mutual information (MI) methods (25; 28) and SINCERITIES (26). MI methods infer non-directed edges representing the amount of information shared between the genes (28; 29). SINCERITIES infers directed edges between each two genes by using the temporal change in gene distribution and partial correlation analysis (26). In some instances, for example, in the absence of temporal snapshot data, the data are reordered along a theoretical trajectory representing a dynamical process experienced by cells, a technique referred to as pseudo-time ordering (augmentation) (13; 14; 30; 31). Algorithms such as SCODE use pseudo-time ordered scRNA-seq data to infer a GRN by using a system of ordinary differential equations representing the change in gene expression through the pseudo-time trajectory (32). The rapid increase in the number of GRN inference methods has motivated the development of comprehensive comparative frameworks. A recent paper proposed the BEELINE framework to evaluate the performance of twelve GRN inference methods (15). The authors of the paper concluded that most algorithms struggled to predict the ground truth GRNs and speculated that the low performance was due to the insufficient resolution in the scRNA-seq data.

Rather than proposing another method, the focus of this paper is to dissect and identify some key stumbling blocks for inferring GRNs from scRNA-seq data. There are two key components to a GRN inference method, the first is *the statistic*, the second is *the minimisation problem*. The statistic is a function of the data which is intended to contain the information of the regulation. The minimisation problem is what finds the GRN among the space of all possible GRNs which matches the statistics of the data best. In this paper, we take a bottom-up approach and highlight key challenges in these two core components even in the most ideal scenario.

In our bottom-up approach, we begin by establishing three simple GRNs. We model the GRNs as Markov-jump processes according to the standard model and place the regulation up-stream from the mRNA (17; 33). The three models are a no-interaction GRN, a mono-interaction GRN, and lastly a double-interaction GRN. The models were chosen to have nested information, that is, each former model is fully contained as subset of the next model. We generate synthetic scRNA-seq data using these models. Prior to inference, we wish to understand the effect which transformations, such as statistics and pseudo-timing, have on the raw data. Hence, we investigate the changes in the central moments (i.e. mean, covariance, and skewness) of the data with the change in models. Furthermore, we juxtapose the central moments, of the raw data and the pseudo-time augmentation, to understand possible loss of information between the two. Then we infer the GRNs of the data from the three models using the MI method, SINCERITIES, the linear moment based method, and the non-linear moment based method. We finish by demonstrating the computational limits of the linear least squares and non-linear least squares methods–the core minimisation methods used in GRN inference–with respect to the handling of large time interval lengths between snapshot data.

Inferring GRNs from scRNA-seq data is still an open problem, and through the use of simple examples, we shed light on some core biological and computational issues in the current inference methods. We demonstrate that these obstacles can be overcome on a small scale, however, generalising to a whole transcriptome will require further research.

2 Results

2.1 Stochastic Two Gene Interaction Models

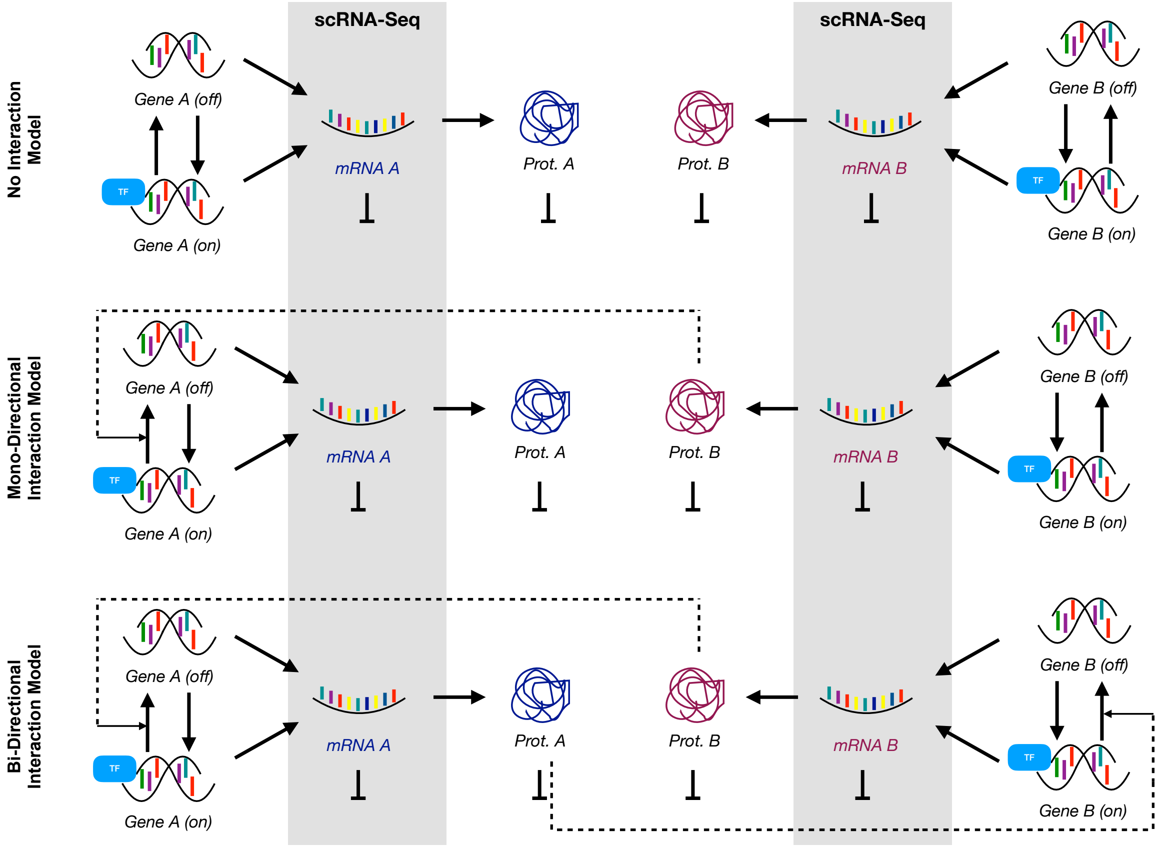


Figure 1: The three GRN models of interest comprising two Genes A and B and their corresponding mRNA and proteins. From top to bottom, model schematics for: **No-I**: No interaction model in which none of the genes are affected by the protein product of the other genes . **Mono-I**: Mono-directional interaction in which the switching off of Gene A is actively up regulated by the protein product of Gene B but the regulation of Gene B is independent of gene A and it's downstream products. **Bi-I**: Bi-directional interaction in which the protein product of each gene upregulates the switching off of the other gene. Only the scRNA-Seq data (grey shaded part of the models) will be used in the GRN inference methods.

To investigate the role of moments in unravelling regulatory reactions from scRNA-seq data, we constructed three simple two-gene GRN models (see Methods 4.1, 4.5). Our first GRN model was a simple no interaction (No-I) two gene model, where each gene, Gene A and Gene B, can be in one of two discrete states, on or basal (off state), and can switch between these states via a constant propensity. The gene is then transcribed into mRNA at a constant rate depending on the state of the gene. The transcribed mRNA then undergoes translation and the respective protein is synthesised (Fig. 1 Top). The mRNA and proteins undergo degradation proportional to their respective populations. In the No-I model, the downstream products associated to their respective gene are not correlated across genes. Our second GRN model was a mono-directional interaction (Mono-I) model, that is, it had the same reactions as the No-I model with the exception of an interaction where Protein B actively upregulates the switching off of Gene A (Fig. 1 Middle). In this scenario, Gene A and its downstream products are affected by the regulation of Gene B, however, Gene B is not affected by any downstream products of Gene A. Lastly, our third GRN model was the bi-directional interaction (Bi-I) model, where protein A upregulates the switching off of Gene B and vice versa, protein B upregulates the switching off of Gene A (Fig. 1 Bottom). In the Bi-I model, all products in the system are correlated.

2.1.1 Covariance and Skewness can aid in Detecting Regulatory Pathways

The three models were simulated using the Stochastic Simulation Algorithm (SSA) (see Methods 4.6). Only the mRNA expression counts from the simulations were extracted for regulatory inference, to mimic scRNA-seq data.

In the No-I model, we observed that both the time course of the mean expression of both mRNAs (A and B) increased identically until the time horizon (Fig. 2 a, Sup. Fig. A a). Due to there being no interactions across genes, as expected, the samples at any fixed time point exhibited near zero covariance between the mRNA expression counts (Fig. 2 a,e). In the Mono-I model, we observed that at early time points the mRNAs' mean expression increased similarly, then, the mean expression of mRNA A started to plateau while the mean expression of mRNA B continued to rise, and had a similar time course as the mRNA B in the No-I model (Fig. 2 b, Sup. Fig. A b). We observed in the time course, that the mRNAs had a negative covariance between them (Fig. 2 b,e). Lastly, in the Bi-I model, we observed that the mean expression time course of the mRNAs increased identically–like in the No-I model. The expression distribution was found to also have a negative covariance structure, however, upon inspecting the distribution of a snapshot at $T = 60$ sec, we saw that the distribution was very symmetric and was shaped like a waning crescent (Fig. 2 c-e).

To understand the origin of the crescent shape, we compared the time course of the skewness of mRNA A and mRNA B in the three GRN models. We found that in both the Bi-I and No-I model, the mRNA A was positively skewed and followed the identical time course. Furthermore, in the Bi-I model, mRNA B was also skewed similarly to mRNA A. That is, all downregulated mRNAs in the models exhibited similar skewness (Fig. 2 f, Sup. Fig. A c-d).

In summary, comparing only the mean time course of the three GRN models, we could not distinguish the underlying regulatory reactions between the No-I and Bi-I model. Similarly, the covariance could distinguish that Mono-I and Bi-I had some 'negative' interaction occurring, relative to the No-I model. However, the direction of the interactions was unclear. When we compared the skewness of the mRNAs, we could see that in the Mono-I interaction, mRNA A was being affected, and a similar effect was also acting on both mRNA A and B in the Bi-I model. Hence, the regulatory information was not in one statistic, but rather distributed over at least three statistics: the mean, the covariance, and the skewness.

2.1.2 Pseudo-time augmented snapshots do not recapitulate the skewness in the original data

When multiple snapshots are unavailable, pseudo-time based augmentations of the data are used to infer the underlying GRNs. To study if the pseudo-time augmentation preserves the moments, we removed all the true time labels within each of the three GRN model's data, and augmented the expression counts with a diffusion map based pseudo-time (see Section 4.9). In the time course of the central moments of the pseudo-time augmented data, we observed that both the No-I and Bi-I model's mean expression of the mRNAs had a similar trend, like in the true time course (Fig. 2 g-i, Sup. Fig. A e). With respect to the covariance, we found that pseudo-time augmented data had negative covariance in the Mono-I and Bi-I models. Surprisingly, we also found a positive covariance in the No-I model time augmented data (Sup. Fig. A f-g). The most drastic differences were seen in the skewness, where for all three models, the pseudo-time augmented data showed predominantly negative skewness, sharply contrasting against positive skewness seen in the original data.

In summary, pseudo-time augmented data can capture trends in the first two central moments, however, it could underestimate the skewness, hindering accurate GRN inference.

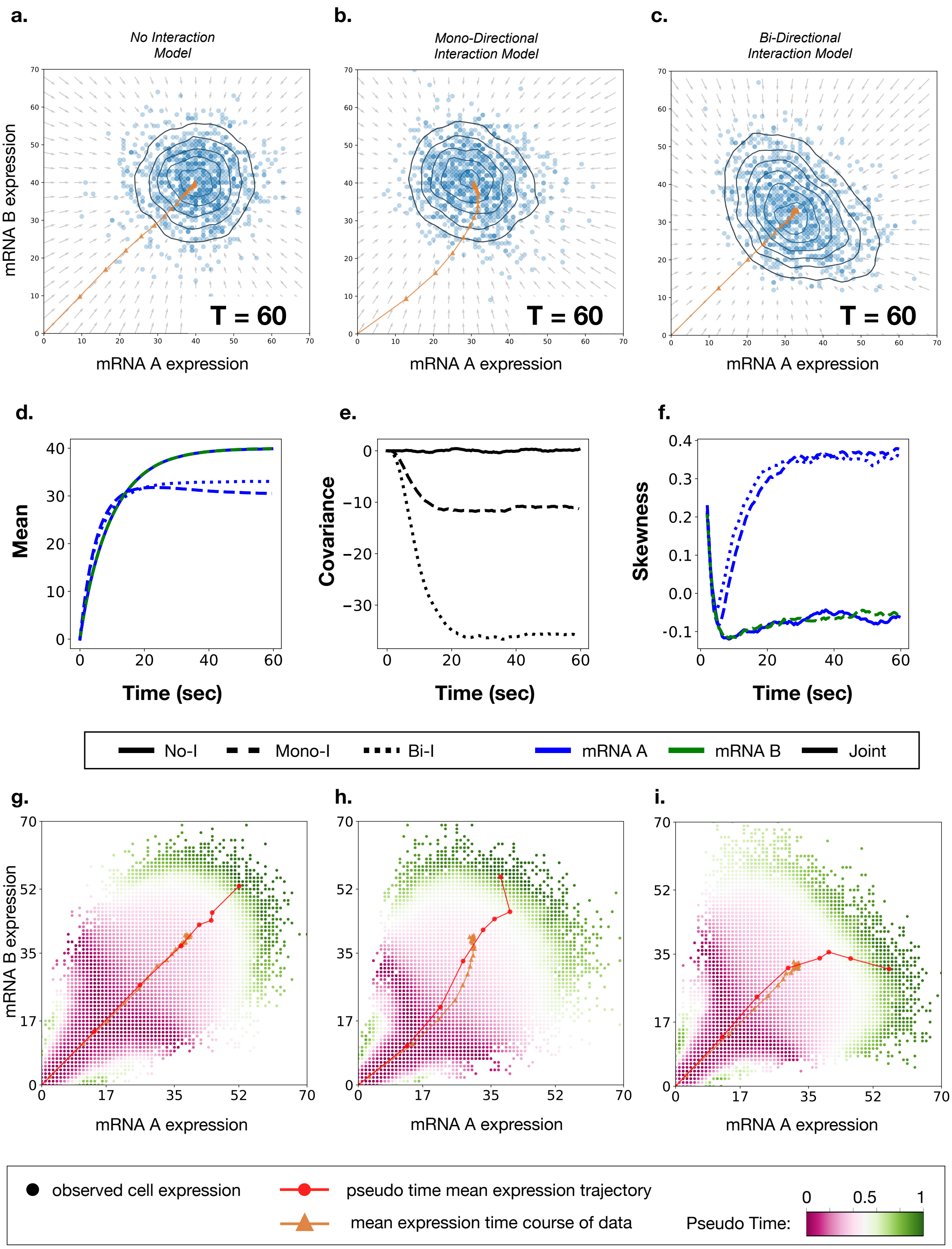


Figure 2: Snapshot Gene expression data at time $T = 60$ showing 1000 sample mRNA population counts. Top row figures are for **a.** No-I model, **b.** Mono-I model, and **c.** Bi-I model. Arrow represent the vector field composed of the derivative of the first order moment and orange line is the mean expression time course from an initial expression value (mRNA A, mRNA B)=(0, 0) to a similar steady state value of around (40, 40) for the three model. Bottom row figures are the same gene expression data reordered through Pseudo-time trajectory **d.** No-I model, **e.** Mono-I model, and **f.** Bi-I model.

2.1.3 No- and Mono-directional interactions are harder to infer than Bi-directional interactions

Four inference methods were applied to infer GRNs from the synthetic scRNAseq data of the three interaction models: the linear moment based method (Linear MBI, see Methods 4.3), the nonlinear moment based method

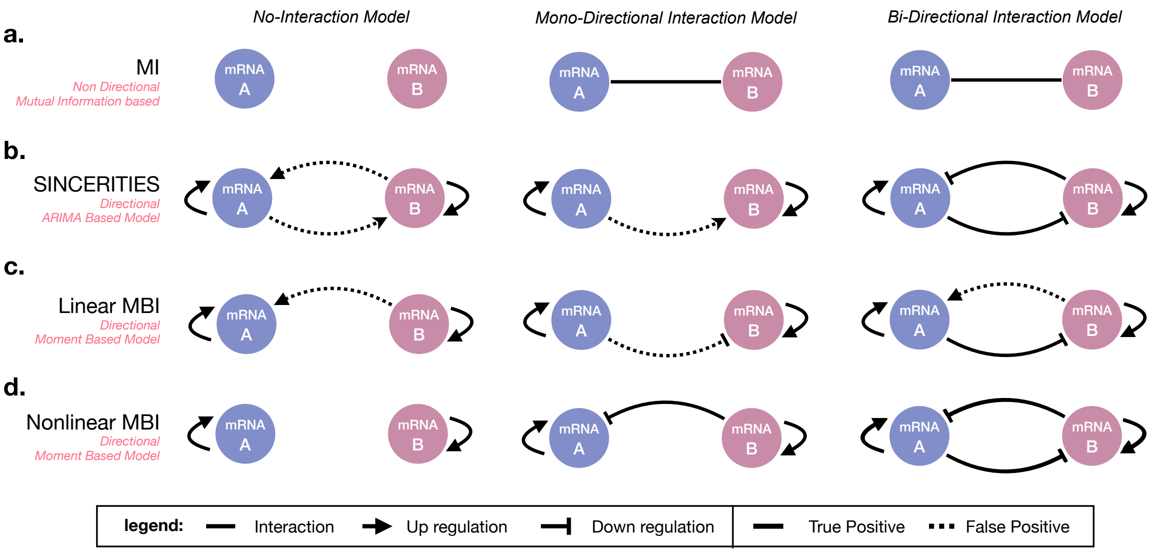


Figure 3: Comparison of GRN obtained from four network inference methods for the three models No-I, Mono-I, and Bi-I. The mRNA are shown as the nodes of the graphs (filled circles). The edges represent the regulatory relationships between the two mRNa. An edge looping over the same mRNA indicate self-regulation. Some algorithm infers non-directed edges: **a.** MI method (first row); and some algorithm infers directed edges: **b.** SINCERTIES, **c.** Linear MBI, and **d.** Nonlinear MBI. Non-directed edges indicate the presence of an undetermined regulatory relationship. An edge with an arrow end (→) toward an mRNA indicates the upregulation of the associated gene, whereas an edge with a flat end (⊣) toward an mRNA indicates the downregulation of the associated gene.

(Nonlinear MBI, see Methods 4.4), the Mutual Information (MI) method (see Methods 4.10), and the SINCERITIES method (see Methods 4.11). We repeated the inference 400 times, each time generating from new synthetic data.

The MI method inferred non zero MI scores for all three models (see Method 4.10). We observed a more than five fold increase in the mean edge score for the Bi-I model with respect to the No-I model, and furthermore, the mean edge score for the Mono-I model was found in between (Supp. Fig. B a). A one-way ANOVA analysis showed that the differences in the means of the edge scores of the three models were statistically significant (an F-value of 49418 and a p-value of strictly less than 0.001). Furthermore, a pairwise comparison with Tukey HSD (with a p-value of 0.001) also showed a significant difference between each two models. Using the mean MI-score of the No-I model as the minimum score edge cutoff, we concluded that the MI based approach was effective in detecting that the three models had a different magnitude of interactions between the genes (Fig. 3 a).

In SINCERITIES, interaction strength score is estimated by regularised regression of a system of distributional distances, while the sign of interaction (activation vs. repression) is determined by the sign of the partial correlation coefficient (see Method 4.10). In the No-I data, the SINCERITIES method inferred all possible activations between and within genes with a weak consistency in interaction scores (Fig. 3 b Left, Supp. Fig. B b). The interaction scores for the Mono-I model gave a clearer result, where a true positive self-activation of mRNA A and mRNA B were observed, however, the repression of mRNA A by mRNA B was missing. Instead, SINCERITIES inferred a false positive interaction of mRNA A activated by mRNA B (Fig. 3 b Middle). The inference of the Bi-I model was done correctly by SINCERITIES with high interaction scores (Fig. 3 b Right, Supp. Fig. B c).

We also investigated the ODE based inference using SCODE (32), which uses pseudo-time augmented data. However, we were not able to derive a sensible meaning of the parameters inferred by SCODE (Data not shown). The reason for this could be that SCODE was designed for inferring hundreds of genes and it was not possible to scale down the method to our simple two gene model.

The moment based inference (MBI) methods used the time course of up to degree three moments to infer their GRNs (see Methods 4.3-4.4). Given we knew *a priori* that the information was in the first three moments, to avoid over fitting, we used 15 times fewer snapshots in the moment based inference methods than in MI and SINCERITIES.

We saw that the Linear MBI performed slightly worse than SINCERITIES in inferring the underlying GRNs of the three models (Fig. 3 c). In particular, the Linear MBI method predicted a false positive activation between mRNA B and mRNA A in the Bi-I Model. Looking at individual GRNs inferred in the 400 replicates, we found that the Linear MBI method at best inferred the correct GRN 2.5 % of the time (Supp. Fig. C b).

Lastly, the Nonlinear MBI method performed the best out of the four methods. It predicted all true positive interactions and no false positive interactions (Fig. 3 d). Furthermore, looking to the individual GRNs in the replicates, we found that it correctly predicted the No-I model 54 % of the time, the Mono-I model 78 % of the time, and the Bi-I model 95 % of the time (Supp. Fig. C c).

In summary, in the four inference methods that we compared, the Bi-I model was the easiest to capture (Fig. 2 c). We suspect that this results from the strong double correlation signal present in the data, which results from the nonlinear interaction between gene A and gene B. The fact that the Mono-I model only had one interaction was detected by all methods, however, the directionality and regulatory mechanism could not be correctly detected. Lastly, the No-I model showed that not all methods are specific enough to correctly detect no interaction.

2.1.4 GRN inference is sensitive to starting populations

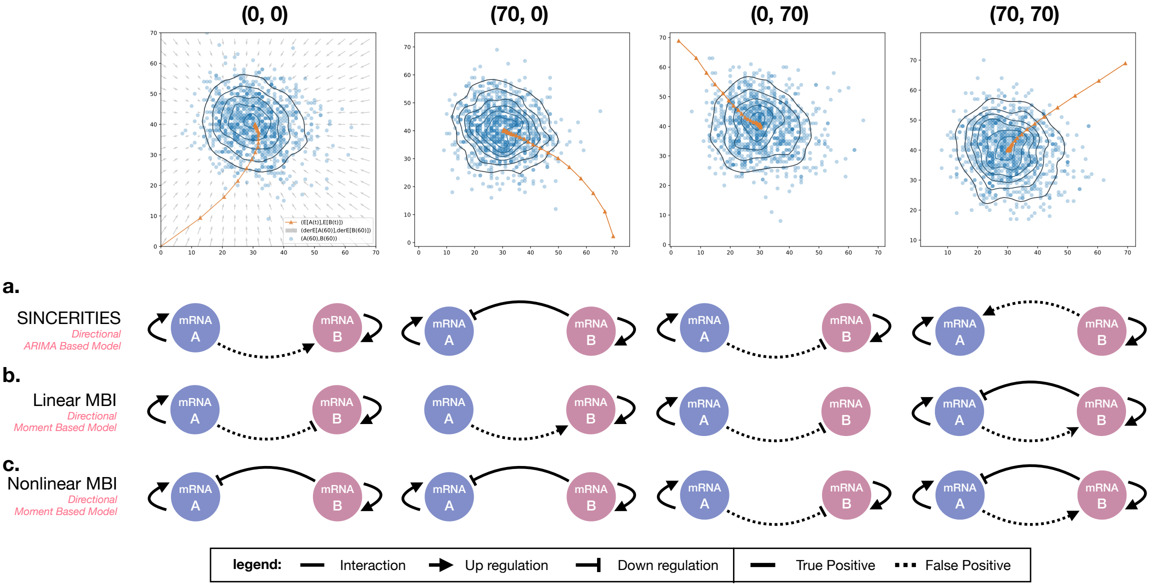


Figure 4: Illustration of GRNs obtained for the Mono-I model from mRNA counts simulations with different initial count configurations for (mRNA A, mRNA B): (0, 0), (70, 0), (0, 70), and (70, 70). Top row: 1000 sample mRNA population counts at time $T = 60$, the vector field indicating the direction of change in the model, i.e., derivatives of the first moments at a given point in the plane, and the orange line is the mean expression trajectory in the data. The three rows **a.**, **b.**, and **c.** show the inferred GRNs aligned with the corresponding to each initial conditions in the top row (refer to Fig. 3 for the meaning of the graphs).

The under-performance of the inference methods on the Mono-I model data was surprising. To investigate if the direction from which the target state is approached could influence the inference of the correct regulatory direction and mechanism, we simulated the Mono-I model with different starting population counts of (mRNA A, mRNA B): $(70, 0)$, $(0, 70)$ and $(70, 70)$ (Fig. 4 Top).

Firstly, we found that all methods struggled in accurately capturing the right GRN, each method failing at different starting points. SINCERITIES was able to detect that there was only one interaction in all scenarios, however, it only captured the right GRN for starting population $(70, 0)$. That is, when the starting population had an abundance of mRNA A, it could detect that A was being repressed (Fig. 4 a). Interestingly, SINCERI-

TIES predicted the right GRN in all the replicates for the case $(70, 0)$ (Supp. Fig. C a) The Linear MBI method did not predict the right GRN for any of the starting populations. At best, for population $(70, 70)$, Linear MBI predicted the right GRN 26 % of the time (Supp. Fig. C b). Surprisingly, in non-symmetric starting population cases, it did not capture some of the self-regulation edges (Fig. 4 b). Lastly, the Nonlinear MBI also under-performed when we started with an abundance of mRNA B. Looking into the replicates, we found that it predicted the right GRN 20 % of the time in the $(0, 70)$ case and 40 % of the time in the $(70, 70)$ case (Fig. 4 c, Supp. Fig. C c).

In summary, we saw that the direction from which we approach the target state influences the outcome of the inference of current methods, and this "directional bias" needs to be considered.

2.2 Linear MBI methods are sensitive to interval lengths between snapshots

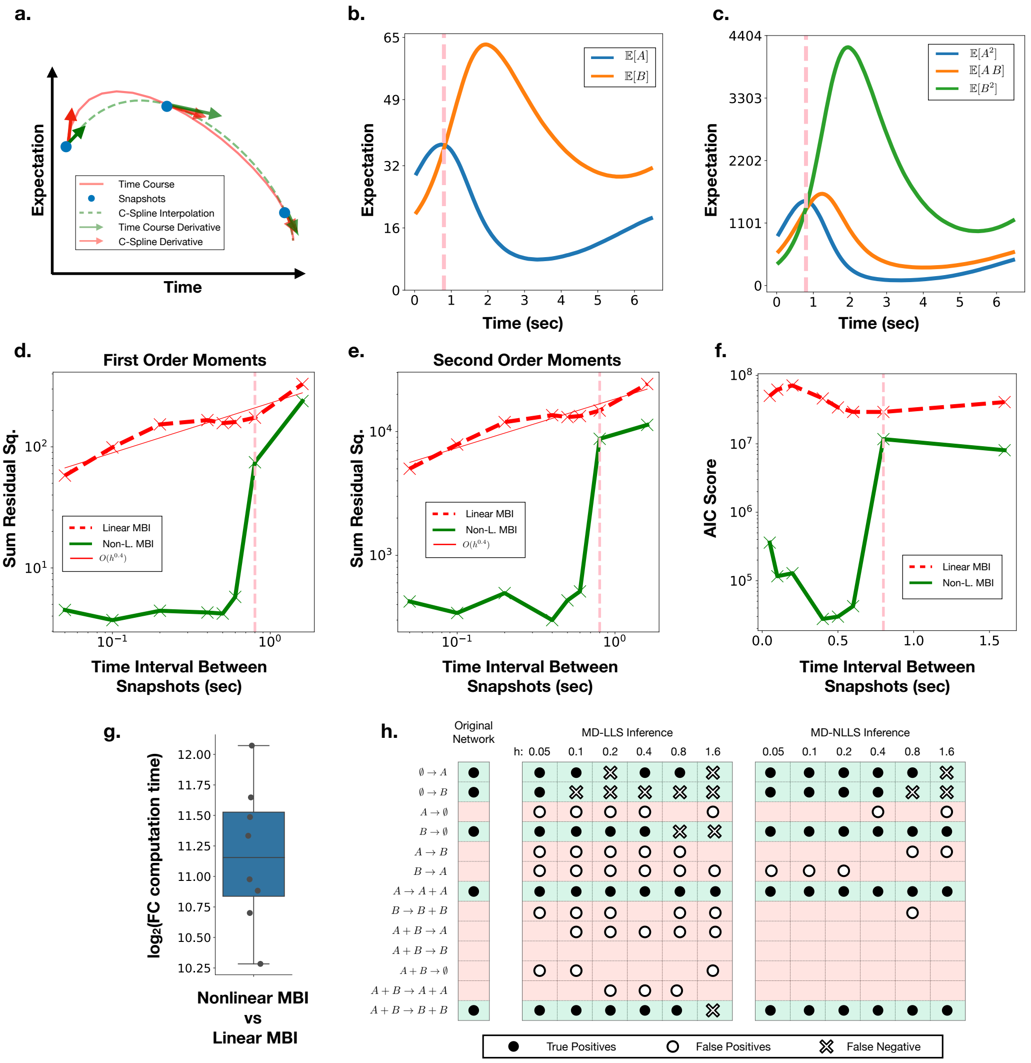
The linear least-squares method is well established and can be used to solve high-dimensional inference problems. To harness its scalability for inferring GRNs using moments (Linear MBI), good approximations of the derivatives of the moments' time courses are essential (see cartoon in Fig. 5 a). However, due to snapshot intervals generally being large in sequencing, good derivative approximations are seldom possible. We investigated the effect of interval lengths between snapshots on the accuracy of the inference by constructing a simple stochastic damped oscillator model (see Methods 4.8, Fig. 5 b-c). Snapshots of different time interval lengths were taken and their underlying network was inferred using the Linear MBI method and the Nonlinear MBI method.

2.3 The Linear MBI method struggles even at small snapshot intervals.

We observed that the residual sum of squares of the Linear MBI method increased with order $\mathcal{O}(h^{0.4})$ with respect to interval length h between snapshots (Fig. 5 d-e). Upon inspecting the inferred reactions, we found that for time interval of $h = 0.05$ sec, the Linear MBI method inferred the five true reactions and a further five false positive reactions (Fig. 5 h). For the subsequent interval lengths, we found that the Linear MBI method continued inferring five to six false positive reactions and the number of true positive reactions was decreasing. The mean time course of the SSA simulations with the inferred parameters showed that the Linear MBI method performed poorly in fitting data, even for the smallest interval length of $h = 0.05$ sec (Supp. Fig. D). In this case, 124 snapshots (1116 moments) were used to infer 13 reactions and surprisingly, we did not observe a close reconstruction to the real data. This suggests that the errors made in estimating the derivative could not be remedied by the large amount of snapshot data.

2.3.1 Nonlinear MBI method circumvents the derivative estimation step at the cost of a significant increase in computational time

The Nonlinear MBI method circumvents the derivative estimation by minimising the distance of the inferred model to the data. This results in a non-linear least squares problem, which does not need the time-course derivative of the moments. In comparison to the Linear MBI method, for time intervals less than $h = 0.6$ sec, we found that the Nonlinear MBI method had at least one order of magnitude lower residual sum of squares in all moments (Fig. 5 d-e). Furthermore, we found that the residual sum of squares did not increase linearly for small time interval lengths, showing a near flat trend between residual and interval length. Looking at the inferred reaction network, we observed that the Nonlinear MBI method captured all of the true reactions, and only inferred one false positive reaction for time intervals less than $h = 0.6$ sec (Fig. 5 h). Interestingly, we observed that for intervals larger than $h = 0.8$ sec, the Nonlinear MBI method starts to perform as poorly as the Linear MBI method. Upon closer inspection, we found that $h = 0.8$ sec is roughly where the first peak in the time course of population A occurs (Fig. 5 b-c). Comparing the AIC scores of the two approaches, we saw that the Nonlinear MBI's minimum AIC score was at least two orders of magnitude smaller than that of the



	Original Network	MD-LLS Inference						MD-NLLS Inference					
		h: 0.05	0.1	0.2	0.4	0.8	1.6	0.05	0.1	0.2	0.4	0.8	1.6
$\emptyset \to A$	●	●	●	☒	●	●	☒	●	●	●	●	●	☒
$\emptyset \to B$	●	●	☒	☒	☒	☒	☒	●	●	●	●	☒	☒
$A \to \emptyset$		○	○	○	○		○				○		○
$B \to \emptyset$	●	●	●	●	●	☒	☒	●	●	●	●	●	●
$A \to B$		○	○	○	○	○						○	○
$B \to A$		○	○	○	○	○	○	○	○	○			
$A \to A+A$	●	●	●	●	●	●	●	●	●	●	●	●	●
$B \to B+B$		○	○	○		○	○					○	
$A+B \to A$			○	○	○	○	○						
$A+B \to B$													
$A+B \to \emptyset$		○	○				○						
$A+B \to A+A$				○	○	○							
$A+B \to B+B$	●	●	●	●	●	●	☒	●	●	●	●	●	●

● True Positives ○ False Positives ☒ False Negative

Figure 5: Illustration of the performance of Linear MBI and Nonlinear MBI methods for the stochastic damped oscillator model. **a.** Estimation of time-course derivatives by computing a C-Spline Derivative from a C-Spline interpolation of the time-course of the snapshot data points, **b.** time-course of first order moments and **c.** second order moments. Pink dashed line indicates the highest time intervals between snapshots producing optimal results (also indicated in **d.**, **e.**, and **f.**). Second row shows the log of sum residuals of SSA produced by the inferred network parameters in the **a.** first, **b.** second, and **c.** third order moments as a function of the time-interval between snapshot data (the parameters found with the Linear MBI was used as initial condition in the Nonlinear MBI for the time intervals bellow 0.8 sec delimited with the pink dashed line. For time intervals above 0.8 sec, this methods was not computationally feasible thus the we used the parameters inferred from Nonlinear MBI with the time interval 0.4 sec as initial condition for Nonlinear MBI). **g.** The computational time of Linear MBI vs. Nonlinear MBI. **h.** The parameters inferred with Linear MBI and Nonlinear MBI as compared with the original network.

Linear MBI (Fig. 5 f). Even though the Nonlinear MBI method performed better, it must be noted that it took on average nearly 2000 times longer to compute than the Linear MBI method (Fig. 5 g).

In summary, the simple stochastic damped oscillator model highlighted the major challenges of using moments based methods for inference. In particular, we observed that the log of the residual scaled sublinearly with the interval length for the Linear MBI method. In order to achieve a similar accuracy as the Nonlinear MBI method at interval length $h = 0.05$ sec, the interval length of the data for the Linear MBI would have to be smaller than 10^{-3} sec. Furthermore, we saw that the snapshot interval has to be small enough to observe the turning points of the system for accurate inference.

3 Discussion

Due to the increase in accessibility and robustness of sequencing technology, single cell RNA-seq data is becoming more abundant. The technology has made significant contributions to discovering novel phenotypes and heterogeneities of cells. Recently, there has been a push for using single- or multiple scRNA-seq snapshots to infer the underlying gene regulatory networks steering the cells' biological functions. To date, this aspiration remains unrealised: a recent review by Pratapa et al. (15), who benchmarked twelve publicly available methods, demonstrated a high heterogeneity and overall under-performance of most current GRN inference methods. Even though we are convinced that their conclusions are true, we disagree with their interpretation.

There were clear indications from visual inspection alone of the simulated snapshots, that the regulatory information was captured in the first, second and third order moments of the cells' expression distribution. That is, we could see the downregulation of the genes resulting in a negative covariance structure in their mRNA expression distribution, as expected (25; 34; 35). However, surprisingly, the negative covariance structure was not shaped like a Gaussian, but rather was shaped like a waning crescent. We believe this is a consequence of having the regulation reactions placed down/up-stream of the mRNA, and not directly at the mRNAs. The shape is induced through the inhibition acting on the gene, in turn, altering the rate production of the mRNA. As only the production rate is altered and the degradation rate is the same, a shift in the skewness emerges.

The moments are simply monomial functions based summary statistics of the data. All current GRN inference methods transform the data into their respective summary statistics to derive their results (12; 15; 21; 26; 28; 32). Different summary statistics highlight different characteristics of the data, hence, the interpretability of different summary statistics might present a bigger dilemma and choosing the wrong statistics might lead to erroneous or misleading results. In light of our observations on the importance of up to third order moments on reconstructing regulatory interactions, we can speculate the source of the difficulties in the current approaches. For example, we know that ODE based methods, like SCODE (32), use means as their summary statistics. Means are a strong summary statistic when it comes to detecting reactions which conserve some measurable quantity, e.g. population, probability, etc. They have been very useful in reconstructing metabolic and protein cascading networks capturing educt–product reactions (36; 37). These assumptions are not coherent with the fact that regulation happens up/down stream from the mRNA. We suggest that ODEs are not suitable for inferring GRNs from scRNA data. A similar line of argument follow for inferring GRNs from pseudo-time augmented scRNA-seq data. Pseudo-time augmentation is a powerful tool for unravelling cell development, however, to date, the methods were not developed with the thought of preserving regulatory structures in the data. Hence, there is scope for new pseudo-time augmentation methods which can conserve certain summary statistics which capture the regulatory information.

Another aspect to summary statistics is the shape of their time course . For example, Linear Moment Based Inference (Linear MBI) uses the derivative of the time course of the moments as their summary statistics. Since we know that the derivatives of moments evolve linearly through time , fitting the derivatives reduces to solving the well established linear least-squares problem. A similar construction is also used in the SINCERITIES method, where instead of moments, their summary statistics are Kolmogorov-Smirnov distributional distances. These methods can handle inferring large GRNs harnessing the scalability of the linear least-squares problem they are based on.

We demonstrated that moments are good summary statistics for inferring underlying GRNs from data. However, the fact that moments evolve through time non-linearly requires the use of nonlinear least squares solvers, which currently are not optimised for inferring GRNs. New numerical schemes to solve high dimensional non-linear least squares problems would aid strongly in furthering the field of inferring GRNs.

To recapitulate, there is no effective method to date for inferring GRN networks from scRNA-seq data. It has become apparent that, in choosing summary statistics, we need to find a good trade-off between the compression of the data and the loss of regulatory information. Collating our insights, we propose two constraints which are significant in designing new GRN inference methods for scRNA-seq data: firstly, the proposed summary statistics must capture the information from moments up to order three of the data, secondly, the summary statistic should evolve linearly in time, or at worst, near linear in between snapshots. The former aspect is for capturing the correct regulatory dynamics and the later is for computational feasibility and scalability.

4 Method

4.1 Jump Markov Process

To derive an expression of the moments, we begin by modeling the interactions within a system of N_s species as a stochastic process representing the number of species undergoing N_r reactions. This process is well described as a jump Markov process known as *Kurtz process* which describes the population count configuration $Z(t) \in \mathbb{N}_0^{N_s}$ of the species at time t as

$$Z(t) = Z(t_0) + \sum_{j=1}^{N_r} \mathcal{P}\left(\int_{t_0}^{t} \theta_j f_j(Z(s))\, ds\right) v_j, \tag{4.1}$$

where $\mathcal{P}$ is an inhomogeneous Poisson process. t_0 is initial time, $\theta_j f_j$ is the propensity function representing the rate at which the j-th reaction fires. And $v_j \in \mathbb{N}_0^{N_s}$ is the stoichiometry vector representing the change in species count through the j-th reaction.

It was shown (34) that Equation 4.1 leads to the well known *Chemical Master Equation*, which represents the time evolution of the probability distribution p of the species count configuration $Z(t)$ as follows

$$\frac{\partial p(Z(t) = z)}{\partial t} = \sum_{j=1}^{N_r} \theta_j f_j(z - v_j) p(Z(t) = z - v_j) - \theta_j f_j(z) p(Z(t) = z), \quad z \in \mathbb{N}_0^{N_s} \tag{4.2}$$

4.2 GRN inference by means of parameter inference

Then from Equation 4.2, it can be shown (38) that for any monomial function ϕ, the derivatives of the expectation of $\phi(Z(t))$ is given by

$$\frac{d\mathbb{E}[\phi(Z(t))]}{dt} = \sum_{j=1}^{N_r} \mathbb{E}\Big[\Big(\phi(Z(t) + v_j) - \phi(Z(t))\Big)\theta_j f_j(Z(t))\Big]. \tag{4.3}$$

Using Equation 4.3, we can write down the raw moments, $m(t)$, of the process, $Z(t)$, as a linear system of ODEs.

$$\dot{m}(t) = \mathbf{A}(\theta) m(t). \tag{4.4}$$

$\mathbf{A}$ is known as the *design matrix* and is a linear matrix in $\theta = (\theta_j)_{j=0,\dots,N_r}$, the propensities coefficients, and depends on the stoichiometry $(v_j)_{j=0,\dots,N_r}$ we allow the GRN to undergo.

Using Equation 4.4, we can infer a causal relationship between interacting species by finding the parameter θ that best represents the data. The challenging aspect is that Equation 4.4 implies that when reactions involving two or more products are added, $\mathbf{A}$ is infinite dimensional. This means that truncation lead to instability, since the higher-order moments act as a damping for the lower order moments. In the next sections we present inference methods that are designed that circumvent this issue.

4.3 Linear Moment Based Inference (MBI)

Linear Moment Based Inference (Linear MBI) refers to a group of GRN inference methods that borrows the tool of Sparse Identification of Nonlinear Dynamics (SINDY)(24; 39). This method provides a powerful framework to infer parameters from large data because it approximates the system in Equation 4.4 with a linear system of ODE.

Let $\hat{m}$ be the vector containing the moments up to order l and $\bar{m}$ containing up to order $l-1$ (which is the set of moments that can be written as a function of all the moments in $\hat{m}$ and the propensity coefficients θ following Equation 4.4). We then have a closed system by reformulating Equation 4.4 as follows

$$\frac{d\bar{m}}{dt} = \hat{\mathbf{A}}(\theta)\hat{m}, \tag{4.5}$$

where $\hat{\mathbf{A}}$ is a rectangular matrix with dimension $(\dim \bar{m}, \dim \hat{m})$.

Now we can use numerical methods such as spline derivatives or finite difference to find an approximation $\mathbf{b}$ the moment derivatives from the data

$$\frac{d\bar{m}}{dt} \approx \frac{d\bar{m}_{data}}{dt} \approx \mathbf{b}, \tag{4.6}$$

where $\hat{m}_{data}$ is the moments data up to order l. This reduces the problem (Section 4.2) to a non-negative linear least square problem.

$$\hat{\theta} = \text{argmin}_{\theta \succeq 0} \left\| \mathbf{b} - \hat{\mathbf{A}}(\theta)\hat{m}_{data} \right\|_2^2 \tag{4.7}$$

In the simulations for the Linear MBI method we used the *reactionet lasso* (24) which solves Equation 4.7 by implementing an L1 norm regularization on the parameter θ. *reactionet lasso*, thus, prioritizes the inference small number of parameters, i.e., sparsely connected network. Sparsely connected networks reflect the minimal set of reactions that are involved in the network (39) and it has been suggested that robust networks are parsimonious (40).

4.4 Nonlinear Moment Based Inference (MBI)

We deal with truncation problem of Equation 4.4 by introducing the interpolations of the higher-order moments from the data source function into a truncated system. We refer to this method as Nonlinear Moment based inference (Nonlinear MBI)

Let $\hat{m}$ be the vector containing moments up to order l. We then approximate Equation 4.4 with

$$\frac{d\hat{m}(t)}{dt} = \bar{\mathbf{A}}(\theta)\hat{m}(t) + \mathbf{B}(\theta)u(t), \tag{4.8}$$

where $\bar{\mathbf{A}}(\theta)$ is a square matrix of dimensions (dim $\bar{m}$, dim $\bar{m}$) linear in the parameter θ, truncation of the matrix $\mathbf{A}$ for the moments $\bar{m}$. $\mathbf{B}$ is a rectangular matrix of dimensions (dim $\bar{m}$, dim $u(t)$). And $u(t)$ is the vector interpolation of the moments of order $l+1$.

The central limit theorem states that for a large number of measurements, the errors in the data follows are normally distributed. Thus, the parameter $\hat{\theta}$ that best represent the data can be obtained via maximum likelihood estimation. In this case, $\hat{\theta}$ can be obtained by minimising the negative log likelihood function

$$\|\hat{m}(t) - \hat{m}_{data}(t)\|_2^2 \tag{4.9}$$

where $\hat{m}_{data}$ is the moments data.

The problem of finding the parameter $\hat{\theta}$ characterizing the GRN network (Section 4.2) is then reduced to a non-linear least square minimization problem

$$\hat{\theta} = \text{argmin}_{\theta \succeq 0} \|\hat{m}(t) - \hat{m}_{data}(t)\|_2^2 \tag{4.10}$$

where $\hat{m}$ is the moments data up to order l.

We implemented the Nonlinear MBI method with python using Scipy (41). Numerical approximated the moments model $\hat{m}(t)$ were generated as solution of Equation 4.8 by solving the ODE between every two time-point data. For higher accuracy, we computed the splines of order $l+1$ moments ($u(t)$) by specifying their derivatives (requiring the order $l+2$ moments according to Equation 4.4) into the spline algorithm at the end-points of the time series. We then used a standard least square minimization routine to compute a solution of Equation 4.10. In the case of large difference between the magnitude of order l and order $l+1$ moments, the residuals of the order l moments were multiplied by a constant weight to avoid their underestimation. All codes are available upon request.

4.5 Two Gene Interaction Model Reaction Scheme

To illustrate our method, we consider three simple models of two-gene interaction (Fig. 1). Each models considers three class of variables: the Genes and their corresponding mRNA and Protein. We distinguish the mRNA and protein by appending the alphabet label of their respective origin gene. The first model considers no interaction between the two genes (No-Interaction model or No-I), the second model considers a single regulation pattern (Mono-Directional interaction model or Mono-I), and the third model includes a reciprocal regulation pattern (Bi-Directional interaction model or Bi-I). Additionally, we omitted the presence of technical errors, such as drop-outs, to the synthetic data, since technical- and technological challenges are beyond the scope of this work. For biological data, these technical errors can be dealt with a pre-processing of data using other algorithms such as MAST, scDoc (42; 43).

The three models No-I, Mono-I, Bi-I, share the same backbone. At any time, a gene has a binary state space { **on**, **off**}, mRNA and Proteins are described by their counts so they have a positive integer state space. In a two-gene interaction model for Gene A and Gene B, the species that are involve in the reactions are Gene A, Gene B, protein A, protein B, mRNA A, and mRNA B.

Table 1 and 2 describes each components of the No-I model for Gene A and Gene B, respectively. The Mono-I interaction model contains all reactions in Table 1 and 2, with the exception that reaction (1a) is modified to involves one of the protein product, here Protein B, which actively upregulates the switching off of Gene A in the reaction

$$\text{Gene A } \textbf{on} + \text{ Protein B} \longrightarrow \text{Gene A } \textbf{off},$$

with corresponding propensity σ_1[Protein B][Gene A **on**], where $\sigma_1 = 0.01875$, and stoichiometry vector (-1, 0, 0, -1, 0, 0).

The Bi-I interaction model also contains all reactions in Mono-I model, with the exception that reaction (1b) is also modified to include the upregulation of the switching off of Gene B by Gen A, i.e., reaction (1b) is replaced with

$$\text{Gene B } \textbf{on} + \text{ Protein A} \longrightarrow \text{Gene B } \textbf{off},$$

with corresponding propensity σ_1[Protein A][Gene B **on**], where $\sigma_1 = 0.01875$, and stoichiometry vector (0, -1, -1, 0, 0, 0)

All propensitey coefficients where chosen so that the three models reach the same steady state toward the time horizon.

Table 1: Components of two-gene No-Interaction model of Gene A. The positions in the stoichiometry vector corresponds to (Gene A, Gene B, protein A, protein B, mRNA A, mRNA B)

#	Reactions	Coefficients	Propensities	Stoichiometry	Description
1a	Gene A **on** ⟶ Gene A **off**	$\sigma_1 = 0.125$	σ_1 [Gene A **on**]	$(-1, 0, 0, 0, 0, 0)$	Inactivation
2a	Gene A **off** ⟶ Gene A **on**	$\sigma_2 = 0.5$	σ_2 [Gene A **off**]	$(1, 0, 0, 0, 0, 0)$	Activation
3a	Gene A **on** ⟶ Gene A **on**+ mRNA A	$\rho_1 = 4.75$	ρ_1 [Gene A **on**]	$(0, 0, 0, 0, 1, 0)$	Transcription
4a	Gene A **off** ⟶ Gene A **off**+ mRNA A	$\rho_2 = 1.0$	ρ_2 [Gene A **off**]	$(0, 0, 0, 0, 1, 0)$	Transcription
5a	mRNA A ⟶ mRNA A + Protein A	$\theta = 5.0$	θ [mRNA A]	$(0, 0, 1, 0, 0, 0)$	Translation
6a	Protein A ⟶ $\emptyset$	$\kappa = 0.1$	θ [Protein]	$(0, 0, -1, 0, 0, 0)$	Degradation
7a	mRNA A ⟶ $\emptyset$	$\delta = 0.1$	θ [mRNA A]	$(0, 0, 0, 0, -1, 0)$	Degradation

Table 2: Components of two-gene No-Interaction model of Gene B. The positions in the stoichiometry vector corresponds to (Gene A, Gene B, protein A, protein B, mRNA A, mRNA B)

#	Reactions	Coefficients	Propensities	Stoichiometry	Description
1b	Gene B **on** $\longrightarrow$ Gene B **off**	$\sigma_1 = 0.125$	σ_1 [Gene B **on**]	$(0, -1, 0, 0, 0, 0)$	Inactivation
2b	Gene B **off** $\longrightarrow$ Gene B **on**	$\sigma_2 = 0.5$	σ_2 [Gene B **off**]	$(0, 1, 0, 0, 0, 0)$	Activation
3b	Gene B **on** $\longrightarrow$ Gene B **on**+ mRNA B	$\rho_1 = 4.75$	ρ_1 [Gene B **on**]	$(0, 0, 0, 0, 0, 1)$	Transcription
4b	Gene B **off** $\longrightarrow$ Gene B **off**+ mRNA B	$\rho_2 = 1.0$	ρ_2 [Gene B **off**]	$(0, 0, 0, 0, 0, 1)$	Transcription
5b	mRNA B $\longrightarrow$ mRNA B + Protein B	$\theta = 5.0$	θ [mRNA B]	$(0, 0, 0, 1, 0, 0)$	Translation
6b	Protein B $\longrightarrow \emptyset$	$\kappa = 0.1$	θ [Protein]	$(0, 0, 0, -1, 0, 0)$	Degradation
7b	mRNA B $\longrightarrow \emptyset$	$\delta = 0.1$	θ [mRNA B]	$(0, 0, 0, 0, 0, -1)$	Degradation

4.6 Synthetic scRNA-seq data

For each of the two-gene interaction models: No-I, Mono-I, and Bi-I (see Section 4.5), species count of (Gene A, Gene B, Protein A, Protein B, mRNA A, mRNA B) were generated using stochastic simulation algorithm (SSA) (44). Initial population configuration was set to (0, 0, 0, 0) and the initial simulation time set to 0. All trajectories of populations counts in the simulations were sampled at each time intervals of 0.5 sec and up to a time horizon of 60 sec. We generated a total of 100000 time series trajectories of species counts. Only the mRNA time series, which reflects scRNA.seq data, were retained for GRNs inference.

The moments of the scRNA-seq data are required in GRNs inference methods . The moment model we used in the inference contains up to order three moments, which depend on moments up to order four Equation 4.4, thus we collected moments data for up to order four moments.

In the inference methods, we used the whole dataset for the MI (28) and SINCIRETIES (26) methods. For the MBI methods (Linear MBI, section 4.3, and Nonlinear MBI, section 4.4), we removed the first 30 data points for a more accurate representation of the moments and reduced the dataset by 15 times.

4.7 Two mRNA Reaction Library for GRNs inference

We infer GRNs from synthetic scRNA-seq data representing temporal snapshot of mRNA counts extracted from each two-gene interaction model described in Section 4.5. Since we only have information from mRNA counts, we are not aiming to reconstruct the two-gene interaction models. Instead, we aim to capture the regulatory relationships between the genes from interaction between mRNAs.

We set up the Linear MBI and NonLinear MBI methods to infer GRNs from the network depicted in Table 3. Reactions 6 and 8 represent the up-regulation of A by B and the vis-versa, while reactions 9 and 10 represent the down regulation of A by B and vis-versa. To draw the GRNs from the inferred parameter, we computed the adjacency matrix defining the regulatory (reg) rules as

$$\begin{pmatrix} \text{Auto-reg of mRNA A} & \text{B reg A} \\ \text{A reg B} & \text{Auto-reg of mRNA B} \end{pmatrix} = \begin{pmatrix} \bar{\theta}_5 - \bar{\theta}_3 & \bar{\theta}_6 - \bar{\theta}_9 \\ \bar{\theta}_8 - \bar{\theta}_{10} & \bar{\theta}_7 - \bar{\theta}_4 \end{pmatrix}, \tag{4.11}$$

where $\bar{\theta}_i$ are the propensity coefficients Table 3 averaged over 400 batch inferences, in each of which the snapshot dataset of moments were generated from 10000 subsampled mRNA count time series (see Section 4.6).

The diagonals of the adjacency matrix (Equation 4.11) represent the net rate of firing of the reactions self-production vs. death for mRNA A and mRNA B, whereas the remaining entries represent up-regulation vs down-regulation for mRNA A and mRNA B. Thus, the sign of each entry determines the direction of the regulation, whether it is a net self-production/death or a net up/down-regulation. We only considered the reactions that were firing at least 10 times within the time-window.

Table 3: Reaction Library for Two mRNA species Interaction. For simplicity we refere to mRNA A as A and mRNA B as B.

#	Reactions	Propensity	Stoichiometry	Description
1	$\emptyset \longrightarrow A$	θ_1	$(1, 0)$	birth of A
2	$\emptyset \longrightarrow B$	θ_2	$(0, 1)$	birth of B
3	$A \longrightarrow \emptyset$	$\theta_3\, A$	$(-1, 0)$	death of A
4	$B \longrightarrow \emptyset$	$\theta_4\, B$	$(0, -1)$	death of B
5	$A \longrightarrow A + A$	$\theta_5\, A$	$(1, 0)$	self production of A
6	$B \longrightarrow A + B$	$\theta_6\, B$	$(1, 0)$	production of A by B
7	$B \longrightarrow B + B$	$\theta_7\, B$	$(0, 1)$	self production of B
8	$A \longrightarrow A + B$	$\theta_8\, A$	$(0, 1)$	production of B by A
9	$A + B \longrightarrow B$	$\theta_9\, A\, B$	$(-1, 0)$	annihilation of A from encounter with B
10	$A + B \longrightarrow A$	$\theta_{10}\, A\, B$	$(0, -1)$	annihilation of B from encounter with A

4.8 Evaluation of MBI model Accuracy

We investigated the accuracy of the MBI methods by using the Stochastic Damped Oscillator model (Table 4). Using stochastic simulation algorithm (SSA), we generated synthetic population counts starting from an initial population count of (30, 20). The simulation were performed from an initial time of 0.0510 and a final time of 25.001 by taking sample population count at every 0.05 time intervals. The moments data were computed from 10000 SSA time series. For the purpose of this analysis, we only used the moments data up to time 6.4510 in the network inference method.

To investigate the effect of the time interval between data sampling, we inferred reaction networks from MBI methods for each dataset sub-sampled with the corresponding time interval separations. We then used the inferred parameters network to generate 1000 sample SSA trajectories from which we computed the new moments time series along the same time-interval as original dataset. We then computed the errors of the first order moments, second order moments at all time-interval rather than just on the sub-sampled time-intervals. This allows us to observe how well the approaches estimates the unseen data in between the fitted data (see Fig 5).

We compare the network inferred from the Linear MBI and the Nonlinear MBI by using the Akaike Information Criterion (AIC). The AIC is used to rank inference models by considering a trade-off between goodness of fit and over-fitting (45), it is given in the formula

$$\mathrm{AIC}_c = 2k - ln(\mathcal{L}), \tag{4.12}$$

where k in the number of inferred parameters and $\mathcal{L}$ is the likelihood function

To account for the small number of snapshot data points fitted in MBI methods, we used the AIC_c which corrects the original AIC formula for small sample size (46).

$$\mathrm{AIC}_c = \mathrm{AIC} + \frac{2k^2 + 2k}{N - k - 1}, \tag{4.13}$$

where N is the number of snapshot data points used in the inference method.

The case of MBI method, AIC follows is given by as follows

$$\mathrm{AIC}_c = 2k + \|\hat{m}(t) - \hat{m}_{data}(t)\|_2^2 + \frac{2k^2 + 2k}{N - k - 1}, \tag{4.14}$$

Table 4: Stochastic Damped Oscillator Reaction Library

#	Reactions	Coefficients	Stoichiometry	Description
1	$\emptyset \longrightarrow A$	$\theta_1 = 4.0$	$(1, 0)$	birth of A
2	$\emptyset \longrightarrow B$	$\theta_2 = 3.0$	$(0, 1)$	birth of B
3	$A \longrightarrow \emptyset$	$\theta_3 = 0.0$	$(-1, 0)$	death of A
4	$B \longrightarrow \emptyset$	$\theta_4 = 0.7$	$(0, -1)$	death of B
5	$A \longrightarrow B$	$\theta_5 = 0.0$	$(-1, 1)$	transition of A to B
6	$B \longrightarrow A$	$\theta_6 = 0.0$	$(1, -1)$	transition of B to A
7	$A \longrightarrow A + A$	$\theta_7 = 1.25$	$(1, 0)$	self production of A
8	$B \longrightarrow B + B$	$\theta_8 = 0.0$	$(0, 1)$	self production of B
9	$A + B \longrightarrow A$	$\theta_9 = 0.0$	$(0, -1)$	annihilation of B from encounter with A
10	$A + B \longrightarrow B$	$\theta_{10} = 0.0$	$(-1, 0)$	annihilation of A from encounter with B
11	$A + B \longrightarrow \emptyset$	$\theta_{11} = 0.0$	$(-1, -1)$	annihilation of A and B from encounter
12	$A + B \longrightarrow A + A$	$\theta_{12} = 0.04$	$(1, -1)$	birth of A from encounter with B
13	$A + B \longrightarrow B + B$	$\theta_{13} = 0.04$	$(-1, 1)$	birth of B from encounter with A

4.9 Pseudo-Time Approach

Pseudo-time analysis aims to achieve a temporal ordering of the unordered observations from an RNA-seq gene expression snapshot (30). The resulting “pseudo-time course” can be used to infer GRNs (13; 14; 32).

We performed pseudo-time analysis using diffusion maps as described by (31) and implemented in SCANPY (47). A diffusion map is a non-linear dimension reduction method which yields a lower-dimensional, de-noised representation of the high dimensional gene expression data. Haghverdi et al. derived a measure in the diffusion map space which recovers the dynamics of gene expression and is hence suited for inferring pseudo-time and GRNs. They define the pseudo-time of a cell as the distance from some root cell, which get assigned pseudo-time 0 a priori.

We recreated a typical pseudo-time analysis on simulated and sub-sampled gene expression snapshots generated according to the gene interaction networks described in section 4.5. We slightly over-sampled cells from earlier time points, as our simulation of gene expression converges to equilibrium in later time points and implicitly biases the pseudo-time approach. Furthermore we excluded the first snapshot at time 0. This resulted in three datasets of 50000 cells and two genes. We assigned an arbitrary cell from the first time point as root cell for analysis. Using the diffusion pseudo-time analysis pipeline implemented in SCANPY, we computed a sparse nearest neighbour graph (of 50 neighbours), embedded the observations in diffusion map space using 3 dimensions and computed diffusion pseudo-time for each cell with the first 2 dimensions.

4.10 Mutual Information

The mutual information (MI) measure quantifies the amount of information shared between two interacting species. The MI is a symmetric measure, therefore it has been used to infer non-directed GRN by using it as a score for the confidence of an edge between the genes (28). The MI can take any positive value and therefore there is no general way of interpreting it’s magnitude. A threshold of top scoring network is usually used to draw possible networks (28). In the case of our two gene interaction model (sec. 4.5), it is not possible choose a number of top scoring networks because there is only one possible edge given by the MI method. Therefore, we need to use a different way of interpreting the magnitude of the MI scores. We computed the MI score using the implementation by (28)). We generated MI scores for a batch of 400 time series of moments of gene expression data calculated from 10000 mRNA count time series (section 4.6). We infer the GRN from the resulting distribution of MI scores by conducting ANOVA analysis assessing the statistical difference between the three two gene interaction models No-I, Mono-I, and Bi-I.

4.11 SINCERITIES

SINCERITIES, or SINgle CEll Regularized Inference using TIme-stamped Expression profileS, has recently been proposed to infer a directed GRN by using time series of gene expression data (26). SINCERITIES algorithm performs a regularised regression of a system of Kolmogorov–Smirnov distributional distances to infer a set of scores $\boldsymbol{\alpha_j}$ representing the influence of gene j on all the other genes. Large score indicates a higher confidence that the corresponding edges exists. Similarly to the MI method (see section 4.10), there is not general way of interpreting the magnitude of these parameters. To infer a GRN and benchmark the results as compared to other methods, Gao et al. (26) used the top scoring edges. In addition, SINCERITIES algorithm infers the direction of the edges, i.e., the nature of the interaction, from the sing of the partial correlation coefficients between the each two genes.

We used SINCERITIES to compute interaction scores (i.e., $\boldsymbol{\alpha_j}$) and directions (sign of partial correlation coefficients) for 400 time series of moments of gene expression data calculated from 10000 trajectories from the SSA simulations of mRNA counts. The distributions of the resulting interaction scores are show in Supp. Fig. B. As the authors of SINCERITIES leave it to the user to choose a suitable interaction score threshold for deciding detection of an edge, we chose a generously low threshold of 0.05. The frequencies of the resulting interaction networks across the 400 runs are depicted in Supp. Fig. 1 a and the most frequent network drawn in Fig. 2.

References

[1] Kitano, H. Systems biology: A brief overview. *Science* **295**, 1662–1664 (2002).

[2] Altschuler, S. J. & Wu, L. F. Cellular heterogeneity: Do differences make a difference? *Cell* **141**, 559–563 (2010).

[3] Li, B. & You, L. Predictive power of cell-to-cell variability. *Quantitiative Biology* **1**, 131–139 (2013).

[4] Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews Genetics* **16**, 133–145 (2015).

[5] Dueck, H., Eberwine, J. & Kim, J. Variation is function: Are single cell differences functionally important?: Testing the hypothesis that single cell variation is required for aggregate function. *BioEssays: news and reviews in molecular, cellular and developmental biology* **38**, 172–180 (2016).

[6] Burdziak, C., Azizi, E., Prabhakaran, S. & Pe'er, D. A nonparametric multi-view model for estimating cell type-specific gene regulatory networks (2019). 1902.08138.

[7] Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Molecular Cell* **58**, 610–620 (2015).

[8] Fiers, M. W. E. J. *et al.* Mapping gene regulatory networks from single-cell omics data. *Briefings in Functional Genomics* **17**, 246–254 (2018).

[9] Ghanbari, M., Lasserre, J. & Vingron, M. The distance precision matrix: computing networks from non-linear relationships. *Bioinformatics* **35**, 1009–1017 (2018).

[10] Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine* **50** (2018).

[11] Giovanni, I., Ramon, M.-B. & Holger, H. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biology* **20** (2019).

[12] Delgado, F. M. & Gómez-Vela, F. Computational methods for gene regulatory networks reconstruction and analysis: A review. *Artificial Intelligence in Medicine* **95**, 133–145 (2019).

[13] Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology* **32**, 381–386 (2014).

[14] Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods* **14**, 979–982 (2017).

[15] Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A. & Murali, T. M. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature Methods* **17**, 147–154 (2020).

[16] Enze, L., Lang, L. & Lijun, C. Gene regulatory network review. In *Encyclopedia of Bioinformatics and Computational Biology*, 155–164 (Elsevier, 2019).

[17] Holehouse, J., Cao, Z. & Grima, R. Stochastic modeling of autoregulatory genetic feedback loops: A review and comparative study. *Biophysical Journal* **118**, 1517 – 1525 (2020).

[18] Davidson, E. H. *et al.* A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Developmental Biology* **246**, 162–190 (2002).

[19] Streit, A. *et al.* Experimental approaches for gene regulatory network construction: The chick as a model system. *Genesis* **51**, 296–310 (2013).

[20] Zheng, G. & Huang, T. The reconstruction and analysis of gene regulatory networks. In *Methods in Molecular Biology*, 137–154 (Springer New York, 2018).

[21] Barbuti, R., Gori, R., Milazzo, P. & Nasti, L. A survey of gene regulatory networks modelling methods: from differential equations, to boolean and qualitative bioinspired models. *Journal of Membrane Computing* (2020).

[22] Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* **13**, 227–232 (2012).

[23] Fortelny, N., Overall, C. M., Pavlidis, P. & Freue, G. V. C. Can we predict protein from mRNA levels? *Nature* **547**, E19–E20 (2017).

[24] Klimovskaia, A., Ganscha, S. & Claassen, M. Sparse regression based structure learning of stochastic reaction networks from single cell snapshot time series. *PLOS Computational Biology* **12**, 1–20 (2016).

[25] Stumpf, P. S. *et al.* Stem cell differentiation as a non-markov stochastic process. *Cell Systems* **5**, 268–282.e7 (2017).

[26] Papili Gao, N., Ud-Dean, S. M. M., Gandrillon, O. & Gunawan, R. SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. *Bioinformatics* **34**, 258–266 (2017).

[27] Yang, B. *et al.* MICRAT: a novel algorithm for inferring gene regulatory networks using time series gene expression data. *BMC Systems Biology* **12** (2018).

[28] Chan, T. E., Stumpf, M. P. H. & Babtie, A. C. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Systems* **5**, 251–267.e3 (2017).

[29] ching Liang, K. & Wang, X. Gene regulatory network reconstruction using conditional mutual information. *EURASIP Journal on Bioinformatics and Systems Biology* **2008**, 253894 – 253894 (2008).

[30] Magwene, P. M., Lizardi, P. & Kim, J. Reconstructing the temporal ordering of biological samples using microarray data. *Bioinformatics* **19**, 842–850 (2003).

[31] Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods* **13**, 845–848 (2016).

[32] Matsumoto, H. *et al.* SCODE: an efficient regulatory network inference algorithm from single-cell RNA-seq during differentiation. *Bioinformatics* **33**, 2314–2321 (2017).

[33] Cao, Z. & Grima, R. Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. *Proceedings of the National Academy of Sciences* **117**, 4682–4692 (2020).

[34] Sunkara, V. Algebraic expressions of conditional expectations in gene regulatory networks. *Journal of Mathematical Biology* **79**, 1779–1829 (2019).

[35] Vallejos, C. A., Richardson, S. & Marioni, J. C. Beyond comparisons of means: understanding changes in gene expression at the single-cell level. *Genome Biology* **17** (2016).

[36] Klarner, H., Streck, A., Šafránek, D., Kolčák, J. & Siebert, H. Parameter identification and model ranking of thomas networks. In *Computational Methods in Systems Biology*, 207–226 (Springer Berlin Heidelberg, 2012).

[37] Müller, A. C. & Bockmayr, A. Fast thermodynamically constrained flux variability analysis. *Bioinformatics* **29**, 903–909 (2013).

[38] Engblom, S. Computing the moments of high dimensional solutions of the master equation. *Applied Mathematics and Computation* **180**, 498 – 515 (2006).

[39] Hoffmann, M., Fröhner, C. & Noé, F. Reactive SINDy: Discovering governing reactions from concentration data. *The Journal of Chemical Physics* **150**, 025101 (2019).

[40] Leclerc, R. D. Survival of the sparsest: robust gene networks are parsimonious. *Molecular Systems Biology* **4** (2008).

[41] Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods* **17**, 261–272 (2020).

[42] Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biology* **16** (2015).

[43] Ran, D., Zhang, S., Lytal, N. & An, L. scDoc: correcting drop-out events in single-cell RNA-seq data. *Bioinformatics* (2020). Btaa283.

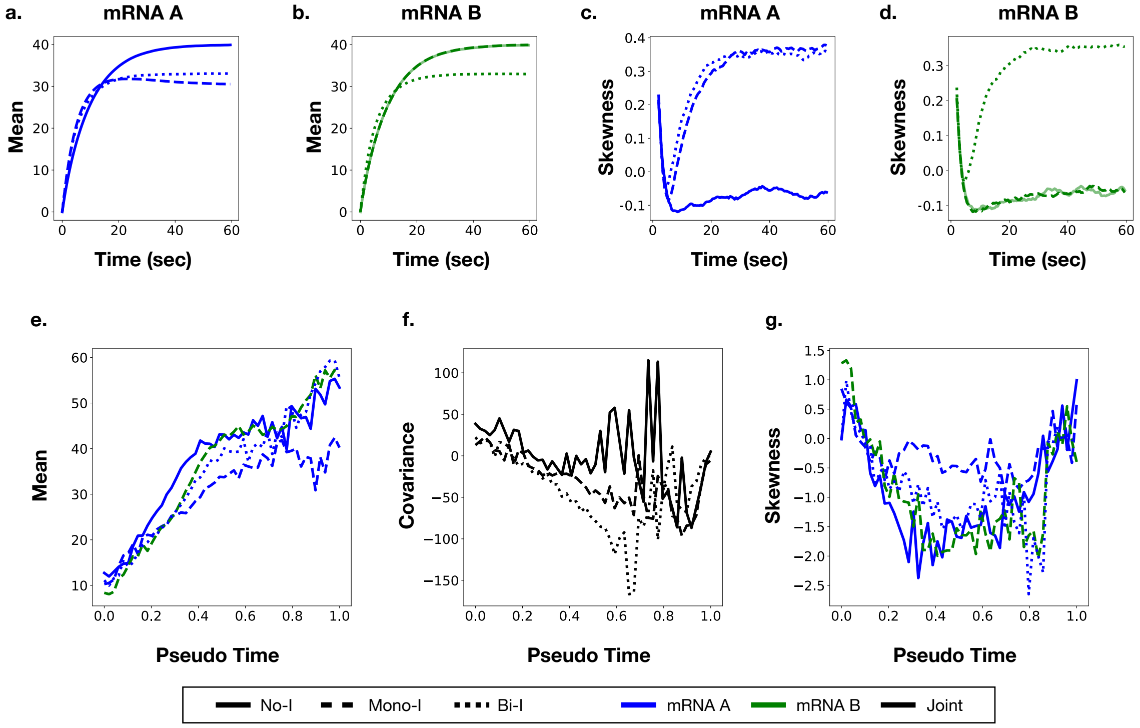
[44] Gillespie, D. T. A General method for numerically simulating the stochastic time evolution of coupled chemical reactions. *IFAC-PapersOnLine* **22**, 403–434 (1976).

[45] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723 (1974).

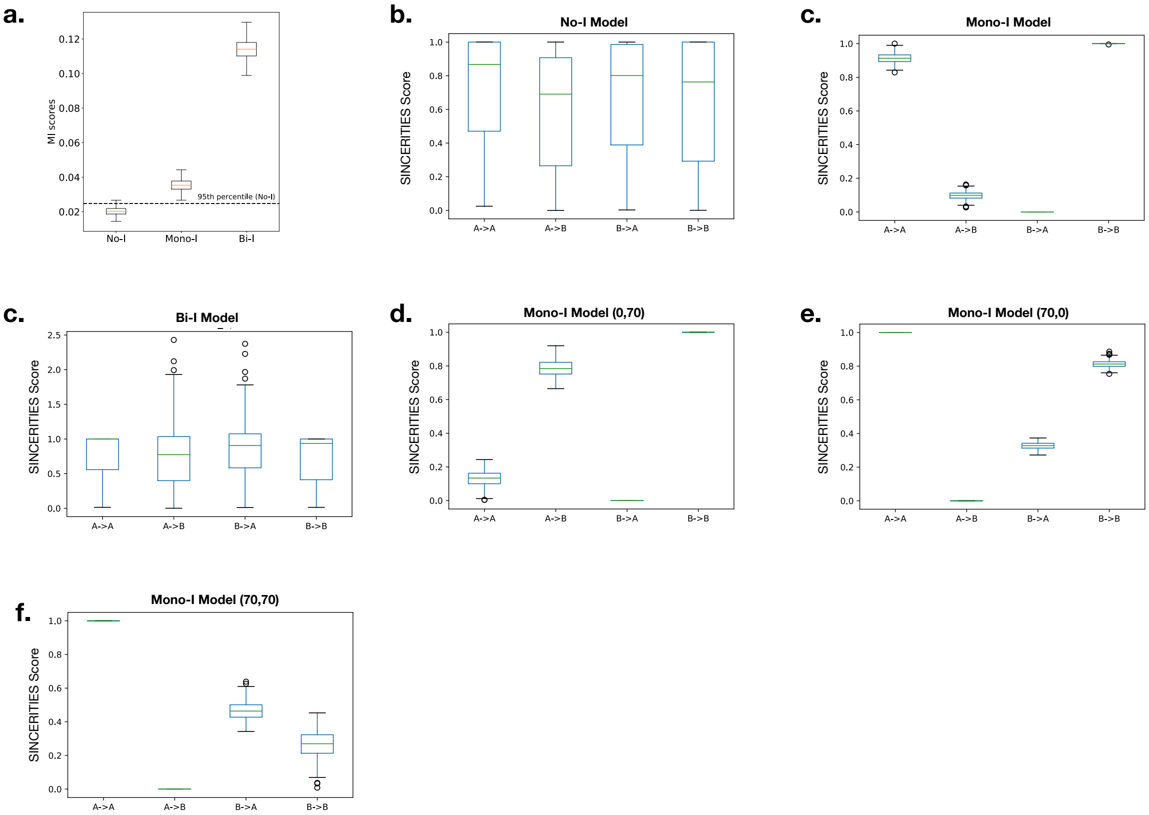
[46] Burnham, K. P. & Anderson, D. R. Multimodel inference: Understanding aic and bic in model selection. *Sociological Methods & Research* **33**, 261–304 (2004).

[47] Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19** (2018).

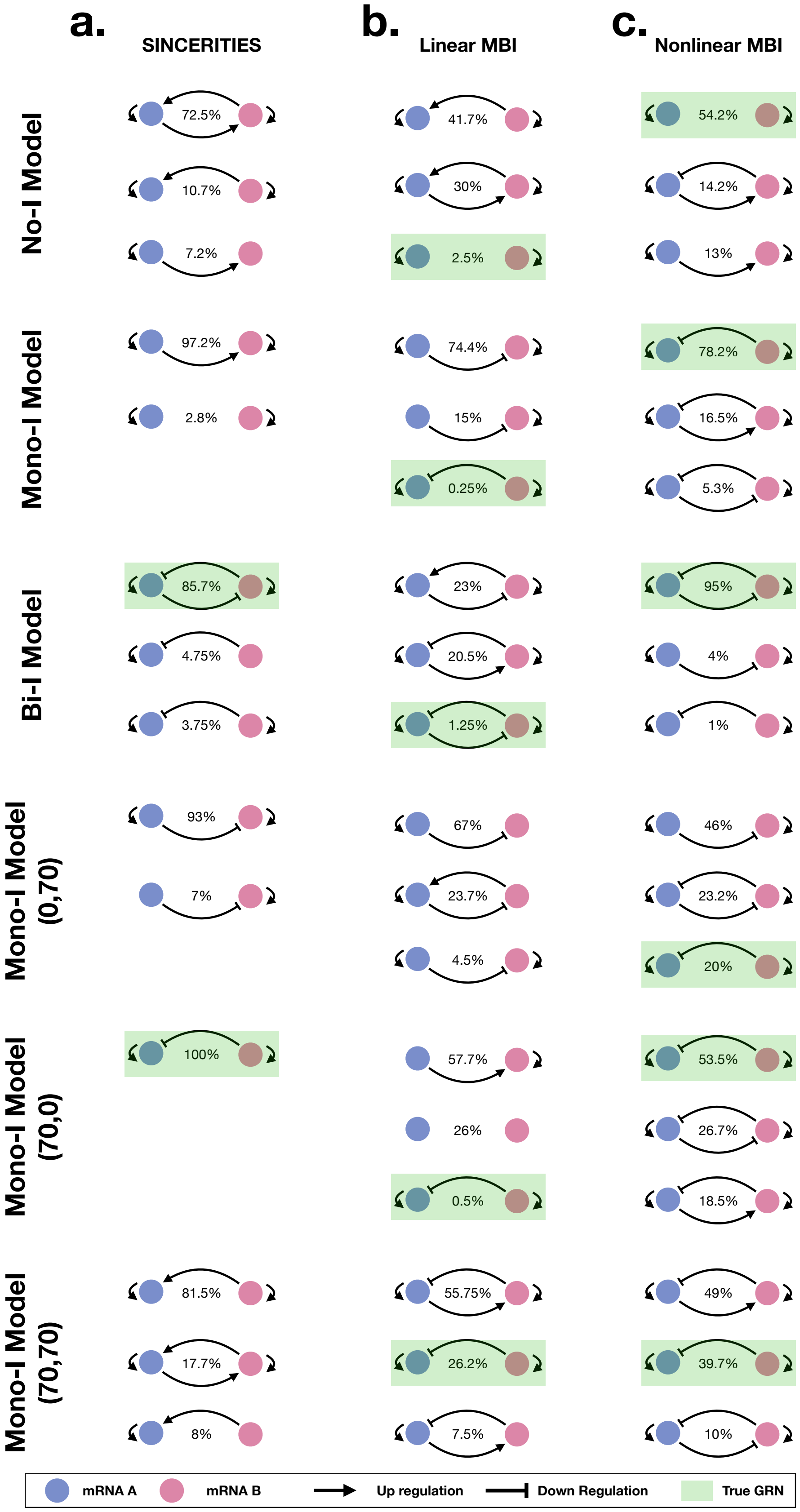
A Supplementary Figures



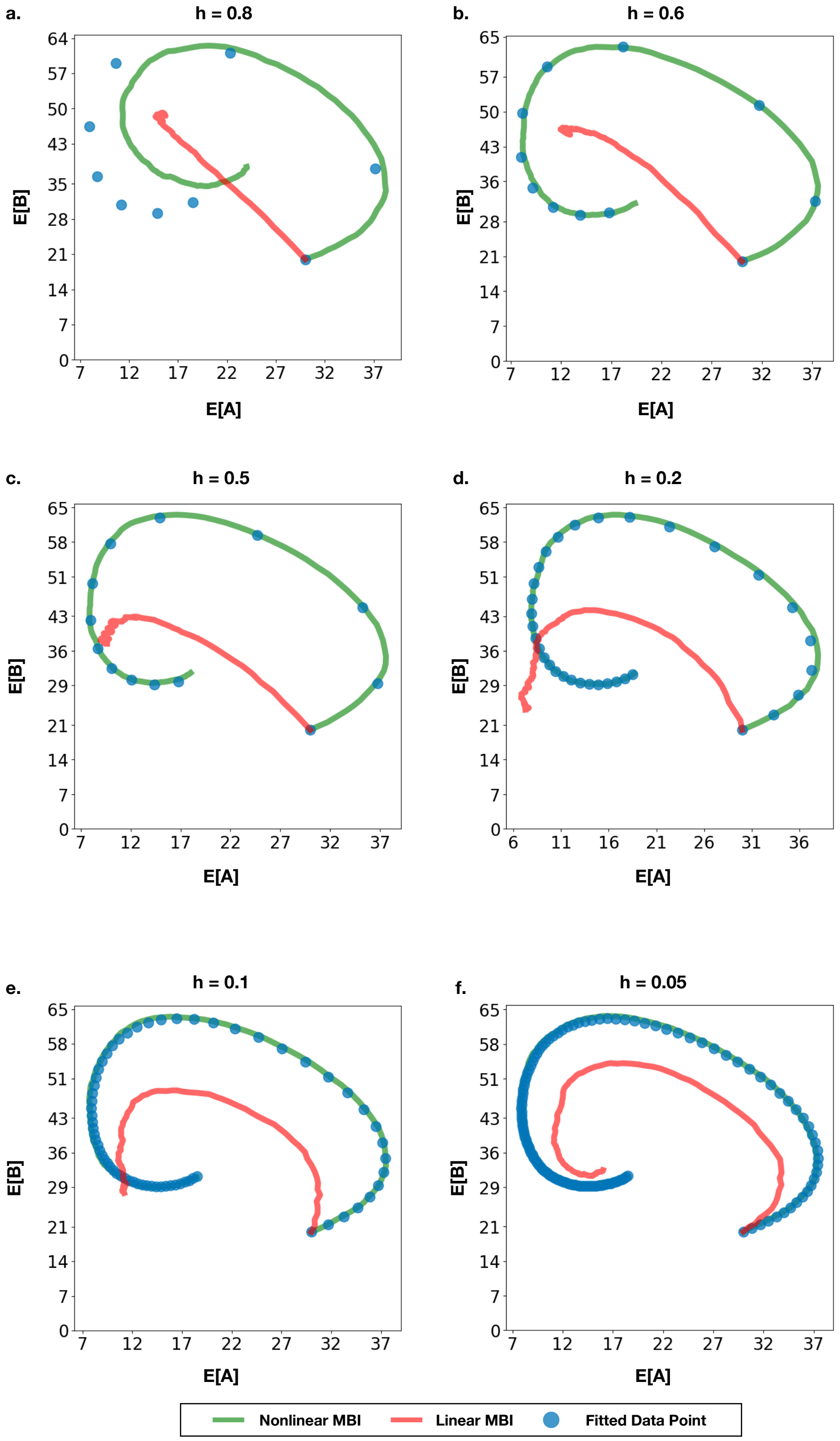
Supplementary Figure A: Comparison of the moments/statistics between simulated data (top row) and Pseudo Time augmented data (bottom row) for the three two-gene interaction models No-I, Mono-I, and Bi-I.



Supplementary Figure B: Comparison of interaction scores for the three GRN models: No-I, Mono-I, and Bi-I: (**a.**) MI scores with initial population counts (0, 0), (**b., c., d.**) SINCERITIES scores from data simulated with initial mRNA population counts (0, 0), (**e., f., g.**) SINCERITIES scores for the data simulated with initial mRNA population counts (0,70), (70, 0), and (70, 70).



Supplementary Figure C: Predicted GRNs and the percentage of times they were inferred by the methods within a batch of 400 snapshot time series of moments: **a.** SINCERITIES, **b.** Linear MBI, **c.** Nonlinear MBI.



Supplementary Figure D: Comparison of the first moments time course reconstruction of the stochastic damped oscillator model for different snapshots time-series data subsampled at time ntervals h (sec) between snapshots. Time course reconstruction were generated from SSA of the model using the parameters inferred by each MBI methods.