Network inference and hypotheses-generation from single-cell transcriptomic data using multivariate information measures

Thalia E Chan[1,†], Michael PH Stumpf[1,†,*], Ann C Babtie[1,†,*]

1 Department of Life Sciences, Imperial College London, London, UK.
†All authors contribute equally.
∗ To whom correspondence should be addressed: a.babtie@imperial.ac.uk, m.stumpf@imperial.ac.uk

Developmental processes are carefully orchestrated. A multi-cellular organism can only emerge from a single fertilised egg cell because gene expression is robustly regulated in space and time by networks of transcriptional regulators. Single cell transcriptomic data allow us to probe and map these networks in unprecedented detail. Here we develop an information theoretical framework to infer candidate gene (co-)regulatory networks and distill mechanistic hypotheses from single cell data. Information theory offers clear advantages for such data, where cell-to-cell variability is all pervasive and sample sizes are large. Higher-order information theoretical functionals capture interactions and dependencies between genes reliably in both *in silico* and real data.

Background

Transcriptional regulation

Precisely controlled patterns of gene expression are essential for the survival and reproduction of all life-forms. Development provides the canonical example, where changes in gene regulation determine the path by which from a single fertilised egg cell emerges a complex multicellular organism. Intricate networks of transcriptional activators and repressors have evolved to regulate the spatial and temporal expression of genes, enabling organisms to adjust transcription levels in response to environmental, developmental and physiological cues [1–6]. Elucidating the structure of such gene-regulatory networks (GRNs) has been a central goal of much recent systems biology research [7–14].

The structure of GRNs alone does not fully constrain their function, but it serves as an important starting point for further analysis. The simplest mathematical representations of GRNs are static graphs, where each node represents a gene, and edges depict either (i) relationships between transcription factors and their targets; or (ii) connect pairs of genes which show coordinated changes in expression across conditions and/or time. The former *causal* relationships are of course more interesting from a mechanistic point of view, but distinguishing regulatory and co-regulatory relationships is often not feasible from expression data without making further assumptions, or using temporal or perturbation data. There are relatively few comprehensively characterized GRNs even at this coarse level of description; the most reliable have been generated over many years and in an often painstaking manner [5, 6, 15–17].

Network inference

The introduction of efficient high-throughput expression assays has driven interest in *network inference* methods, and a variety of machine learning and statistical approaches are now available to identify likely (co-)regulatory relationships be-

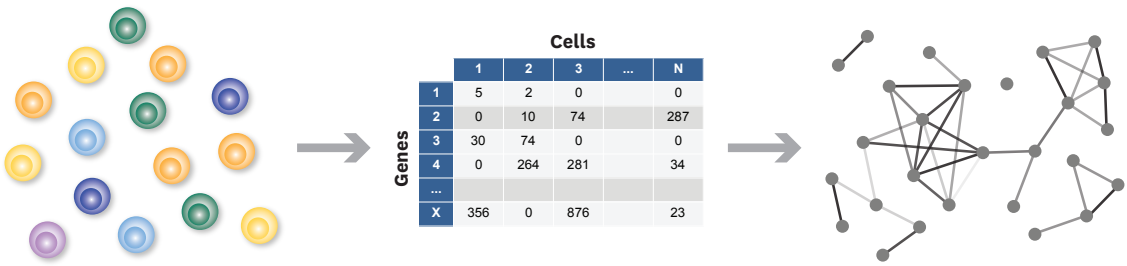

Figure 1: *Single cell transcriptomic data provides copy numbers (or suitable proxies thereof) of mRNA species present inside individual cells. By considering pairs (or triplets, quadruplets, etc.) of mRNA species we can test for statistical relationships among them. These dependencies may reflect coordinated gene expression of these pairs (or groups) of genes, resulting from gene regulatory interactions or co-regulation. Once such sets of genes that jointly change in expression are known, other statistical, bioinformatic, or text-mining analyses can be used to identify likely transcriptional regulators for these sets of genes. By iterating such in silico analyses with further, targeted experimental studies we can, in principle, build up a representation of the gene regulatory network.*

tween genes based on their gene expression patterns. In addition to correlation-based networks (perhaps the simplest way of identifying putative relationships), Gaussian graphical models, (Dynamical) Bayesian networks, regression analyses, and information theoretical approaches have been used for network inference [18–24]. Combining such approaches often confers slight but consistent improvements in the quality of the predicted networks [25–27].

A common challenge for all of these methods — though not always fully acknowledged — is the so-called *large–p, small–n* problem. This refers to the situation, common for bulk transcriptomic studies, where p, the number of hypotheses, is greater than n, the number of measurements; much innovative statistical research over the past two decades has targeted this problem [28]. The increasing availability of single cell data is, however, ushering in a new era, where the number of experiments — here, measurements obtained from single cells — is becoming comparable to the number of potential pairwise gene interactions (see Fig. 1); continuing advances in single cell technologies will enable substantial increases in sample size [29, 30].

Single cell experiments pose new statistical and analytical challenges due to high levels of technical noise, but importantly also provide information about the biological heterogeneity within cell populations [31–35]. Such cell–to–cell variability offers the potential to gain new insight into cell fate decisions and transitions between cell states [1, 3, 4, 36–38]. Identifying the (co-)regulatory interactions that contribute to controlling these processes is a key aim of many single-cell transcriptomic studies [1, 36–40].

Here we develop, evaluate and apply an information theoretical framework to detect statistical dependencies among single cell mRNA (scRNA) expression levels in order to infer candidate functional relationships between genes. Information theory provides a set of measures, chiefly among them *mutual information* (MI), that allow us to characterize statistical dependencies between pairs of random variables [41, 42]. It has considerable advantages over simpler measures such as (Pearson) correlation, as it is capable of capturing non-linear dependencies, and reflecting the dynamics between pairs or groups of genes [43, 44]. Calculating these measures involves estimating joint probability distributions, generally requiring computationally expensive density estimation [45, 46] or data discretization. These approaches are ideally suited to single cell data as the large sample sizes enable more accurate (and higher dimensional) empirical probability distribution estimates, while discretization limits the influence of noise and differing expression magnitudes of genes.

In the next section, we summarise key information theoretic measures [41, 42], including our preferred multivariate information (MVI) measure, *partial information decomposition* (PID) [47, 48],

before developing them into a framework for network inference.

Information theoretic measures

The entropy, $H(X)$, quantifies the uncertainty in the probability distribution, $p(x)$, of a random variable X. For a discrete random variable, the entropy is given by,

$$H(X) = -\sum_{x \in X} p(x) \log p(x), \tag{1}$$

which is maximal for a uniform distribution. When considering the relationship between X and a second random variable, Y, we quantify the information that one variable provides about the other using the mutual information,

$$\begin{aligned} I(X;Y) &= \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \\ &= H(X) + H(Y) - H(X,Y). \end{aligned} \tag{2}$$

This quantifies the difference between the joint entropy, and the joint entropy assuming independence of X and Y, and thus provides a non-negative, symmetric measure of the statistical dependency between the two variables. Given a third variable, Z, the conditional mutual information (CMI),

$$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z), \tag{3}$$

quantifies the information between X and Y given knowledge of Z.

A number of MVI measures have been defined that aim to quantify the statistical dependencies between three or more variables, but there is little consensus as to the most appropriate metric [47]. Some of these measures (such as three-variable MI, CMI and conditional entropies) are used in existing information theoretic based network inference algorithms [22, 23, 49, 50]. However, alternative MVI measures exist that offer a more detailed examination of the relationships between three variables, and in this work we make use of PID [48].

PID considers the information provided by a subset of *source* variables, e.g. $\mathcal{S} = \{X, Y\}$, about another *target* variable, e.g. Z. The variables in set $\mathcal{S}$ can provide information about the target uniquely, redundantly, or synergistically. In the case of three variables, the mutual information between the set $\mathcal{S}$ and the target variable is equal to the sum of four partial information terms,

$$\begin{aligned} I(X,Y;Z) = &\text{Synergy}(Z;X,Y) + \text{Unique}(Z;X) + \\ &\text{Unique}(Z;Y) + \text{Redundancy}(Z;X,Y) \end{aligned} \tag{4}$$

The MI between a single variable in $\mathcal{S}$ and the target comprises a unique and redundant contribution, e.g.,

$$I(X;Z) = \text{Unique}(Z;X) + \text{Redundancy}(Z;X,Y). \tag{5}$$

Conceptually, redundancy is the portion of information about the target that can be provided by either variable in $\mathcal{S}$ alone, the unique contributions are linked to a specific variable in $\mathcal{S}$, whereas the synergistic contribution requires knowledge of both variables in $\mathcal{S}$ (see *Methods* for more mathematical detail).

Existing network inference methods

Information theoretical methods of network inference are widely seen as powerful alternatives to other approaches, such as regression-based models [51], Bayesian and dynamical Bayesian networks [52–54] and Gaussian graphical models [55, 56]. Here we compare our new algorithm against several common information theoretic based network inference methods, and thus briefly summarise these approaches here.

Relevance networks [57] use the MI estimates (or, in some cases, correlation) in order to detect edges. As there is no reliable universal way of determining the statistical significance of MI values, a threshold is typically chosen to determine which edges are present and which ones are not. This fails to account for the fact that MI may be increased for nodes X and Z even though they only indirectly interact via an intermediate node Y. The *Data Processing Inequality* (DPI) allows us to sort out some of these cases by virtue of the relationship

$$I(X;Z) \leq \min\left(I(X;Y), I(Y;Z)\right), \tag{6}$$

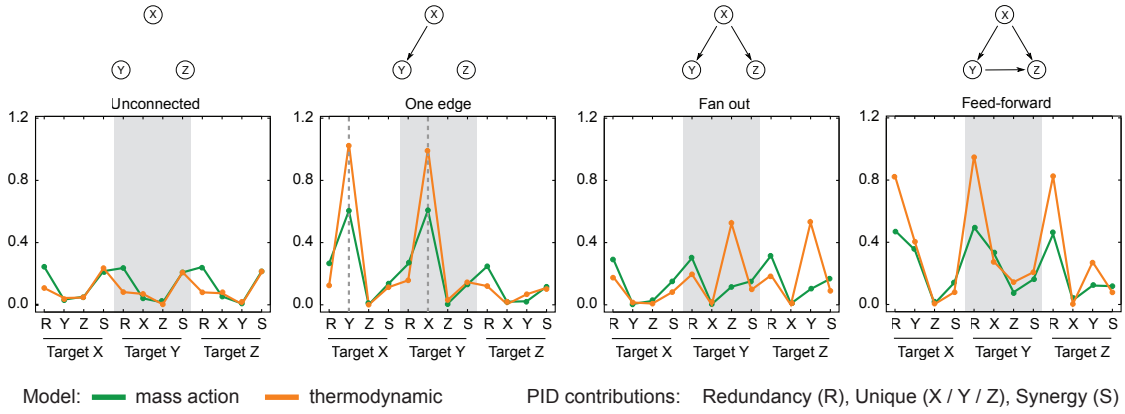

Figure 2: *Mean PID values for 3-gene networks with different topologies. PID values were calculated using data simulated from 3-gene networks with the topologies illustrated above each plot; the models used for simulation assumed mass action (green) or thermodynamic (orange) kinetics. Each line graph shows the mean PID values calculated, with the horizontal axis labels indicating the PID contribution, e.g. the first four values show the PID values with gene X as the target, consisting of the redundancy (R), unique contributions from gene Y (Y) and gene Z (Z), and the synergistic contribution (S). The vertical dashed grey lines in the 'One edge' plot indicate the unique PID values that are used as the basis for our inference algorithm. Here, all regulatory interactions were assumed to be activating, the additional stimulating ligand targeted gene X, and the values indicated are the mean PID values calculated from five sets of simulations (with different randomly sampled initial conditions); results obtained with models that include both activating and inhibitory regulation are shown in (see also Fig. S1).*

which holds whenever X, Y and Z form a Markov Chain. This post-processing of the MI values using the DPI is at the core of the popular *ARACNE* algorithm [20, 58]. Thresholds on the pairwise MIs are used to identify likely dependent pairs X, Z; MI values above the threshold are then considered with every possible other node Y in light of the DPI.

Given that MI values are affected by a number of factors, including especially the variability of each individual random variable, any global *a priori* threshold may be highly problematic: it will give rise to false positives as well as false-negatives. In the *context likelihood of relatedness* (CLR) algorithm [50, 59] the MI between X and Z is considered against all MI values for pairings of X and Z with all other variables Y. Thus the threshold for each pair will reflect the variabilities of both genes, as well as their relative levels of statistical dependence on other genes. *MRNET* [60] aims to identify a minimally redundant but maximally explanatory set of variables/predictors (containing up to f variables/genes) for each target gene X in a greedy manner.

All these methods start from the pairwise MI matrix, and then use it in different ways. Even compiling the MI matrix is, however, fraught with potential problems: the manner in which data are treated (e.g. discretization), and the estimator used for the entropy and MI both affect the performance of the algorithms [61, 62]. When comparing different approaches it is therefore important to ensure that discretization and estimation of MI are done identically. Without this it becomes impossible to disentangle the relative strengths and weaknesses of the different approaches that are based on and interpret MI values. Below, in the *Methods*, we discuss the different estimators and discretization approaches that we use and which are implemented in the `InformationMeasures.jl` package (see Results: Software) . Our comparisons with existing methods thus always start from the same MI matrix.

At every stage it has to be kept in mind, though, what network inference is capable of [12], which is not a reliable faultless reconstruction of the molecular networks underpinning cell physiology; instead it takes the observed variabilities in gene expression (or any other random variable) and looks for patterns of statistical dependence between pairs or groups of nodes (random variables) in the network, quantifies these, and allows us to rank these in an order that should mean

that higher-ranked relationships are more likely to represent true and meaningful functional relationships, e.g. among genes in a cellular process. It is therefore important to stress that "inferred" edges are best considered as *hypotheses* that can be distilled from the data.The aim here is to develop a better quantification and ranking of such dependencies/hypotheses and we will later revisit the interpretation and limitations of such approaches in more detail.

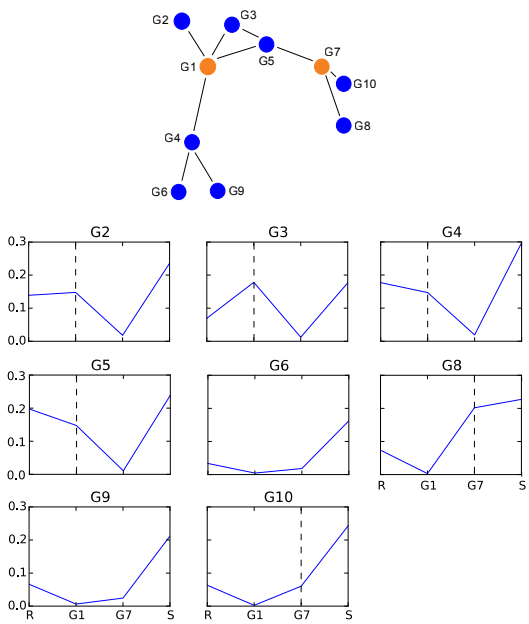

Figure 3: *PID values were estimated from data simulated from a 10-gene in silico network (top) using* `GeneNetWeaver` *[64]. Each line graph shows PID values estimated using genes 1 and 7 as the sources, and each of the remaining genes in turn as the target (graph titles, GX, indicate the target gene). Four PID values are given in each graph — the redundancy (R), the unique information between gene 1 and the target (G1), the unique information between gene 7 and the target (G7), and the synergy (S). The mutual information between two genes is the sum of their unique information and the redundancy (Eq. 5). The ratio of the unique information to the mutual information tends to be higher between pairs of connected genes (dashed vertical lines indicate the unique contributions for connected genes).*

Results

Single cell data are notoriously complex and their analysis fraught with a host of potential problems [31–35, 63]: technical problems — e.g. are zero mRNA expression measurements really absence of expression, or due to measurement sensitivity — and conceptual problems abound. Crucially, bulk or conventional mRNA expression measurements cannot be relied upon as guides or experimental validation of the much more detailed population measurements. So in order to validate and illustrate our approach we use *in silico* data sets where the true answer can be known by definition to compare our approach to other information theoretical methods. We then consider three real scRNA datasets; here the focus will be to show which interactions/relationships are suggested by our approach and to discuss the merit and relevance of these predictions or inferences. Along the way we will develop a set of guidelines as to how these methods are best applied in the analysis of new scRNA data.

Synthetic data generation and analysis

We investigate the usefulness of PID (including in comparison to MI) for inferring network edges using data generated from *in silico* models. Initially, we used stochastic simulations from simple directed 3-node networks of varying topologies, and estimated PID values from these simulated data (as described in *Methods*). A distinctive pattern is apparent in networks with a single directed edge between two variables ('one edge' topology, Fig. 2) — the unique information between the two connected genes is notably higher than both the unique information between unconnected genes and the redundancy values between all three genes. With increasing numbers of edges within the network this pattern is lost, as we would intuitively expect (Fig. 2 and S1). It is important to note that the pattern can only be observed under simulation conditions that generate variability in the observed variables (i.e. statistical relationships cannot be detected when the system is at steady state; Fig. S1).

To explore whether this pattern also occurs for triplets of nodes embedded in large networks, we consider time series expression data simulated from five different 50-gene networks using `GeneNetWeaver` [64]. This software generates dynamical models of networks inspired by known gene connectivity patterns in *E.coli* and *S.cerevisiae* (Fig. S2). PID values were esti-

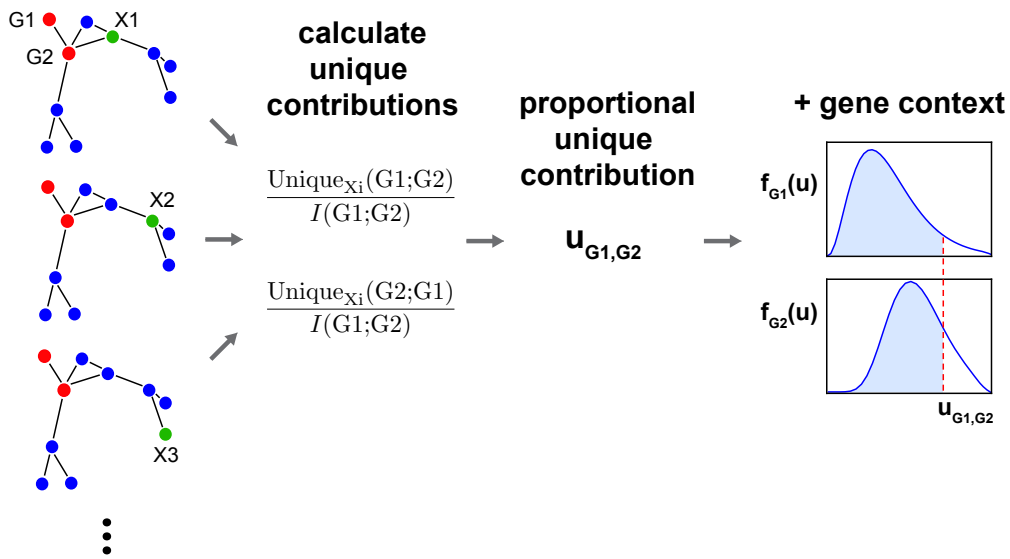

Figure 4: *PID values are estimated for every gene triplet (with each gene treated as the target gene in turn), and from these the PUC, $u_{X,Y}$, is estimated for every pair of genes. For each gene, X, a distribution, $f_X(u)$, is estimated from its PUC scores with all other genes. The confidence of an edge between a pair of genes depends on the cumulative distribution functions for each gene within the pair (i.e. the blue shaded areas); these confidence scores are used to rank all possible network edges.*

mated from these data for every triplet of genes within the networks, and each triplet was classified according to its topology — six topological arrangements were possible given the model assumptions (maximum of one edge between each pair of nodes, no self-regulation, and no feedback loops). Mean PID values were calculated for each group (Fig. S3), and the same distinctive pattern (high unique versus redundant contributions for connected genes) was observed for the triplets with a single directed edge. Examining *in silico* data from a 10-gene network suggests that the relative size of the unique information compared to the redundancy — i.e. the proportion of MI accounted for by the unique contribution (Eq. 5) — may be more informative than the absolute unique information (Fig. 3).

It should be noted that the unique information between any source gene X and target gene Y varies depending on the third gene Z considered in the triplet. The sum of the redundancy between the three genes and the unique information between X and Y is always the same, however: it is equal to their MI (Eq. 5). Hence for any gene pair, the ratio of unique information to MI varies depending on the identity of gene Z. We define the *proportional unique contribution* (PUC) between two genes X and Y as the sum of this ratio calculated using every other gene in a network (where S is the complete set of genes),

$$u_{X,Y} = \sum_{Z \in S \setminus \{X,Y\}} \frac{\text{Unique}_Z(X;Y)}{I(X;Y)} + \sum_{Z \in S \setminus \{X,Y\}} \frac{\text{Unique}_Z(Y;X)}{I(X;Y)}; \quad (7)$$

this measure may be thought of as the mean proportion of MI between two genes X and Y that is accounted for by the unique information.

Incorporating PID into an inference algorithm

A novel network inference algorithm was developed using the PUC measure (Fig. 4). The redundancy and unique information contributions are first estimated for every gene triplet, then the PUC is calculated for each pair of genes in the network (Eq. 7).

Finding a threshold for defining an edge at this stage is problematic, because the PUC scores are distributed differently for each gene (see Fig. S4), thus setting a global threshold for PUC scores across the whole network risks biasing the results by factors such as expression variability. This was previously observed with MI and led to the devel-

opment of measures that take into account the network context, central to the CLR algorithm [50, 59]. A similar solution is employed here: an empirical probability distribution is estimated from the PUC scores for each gene (paired to every other gene in the network), and the confidence of an edge between a pair of genes is given by,

$$c = F_X(u_{X,Y}) + F_Y(u_{X,Y}), \tag{8}$$

where $F_X(U)$ is the cumulative distribution function of all the PUC scores involving gene X (here, we assume either a Gamma or Gaussian empirical probability distribution).

Algorithm performance

Undirected networks were inferred from *in silico* datasets for five 50-gene networks (Fig. S2) using ARACNE, CLR, MI, MRNET [65] and the new PID-based algorithm. Accuracy of the inferred networks was evaluated using the area under the precision-recall curve (AUPR) and the area under the receiver operating characteristic curve (AUROC) [66].

The PID-based algorithm performs favourably compared to the other algorithms (Fig. 5). In agreement with the previous comparisons [27], CLR performs particularly well and rivals PID closely; it performs better than MI alone or the other MI-based network reconstruction algorithms, MRNET and ARACNE.

Application to single-cell data

Psaila *et al.* [67] used single-cell quantitative PCR (sc-qPCR) to map the behaviour of megakaryocyte-erythroid progenitors (MEP). Their analysis revealed the existence of heterogeneity and population structure in this class of cells, which had thus far been considered as a uniform cell population. The parallel existence of lineages that are primed preferentially for a particular cell fate — megakaryocytic (MK-MEP) or erythroid (E-MEP) — in addition to multi-potent progenitors (Pre-MEP) that have still some vestiges of myeloid differentiation capacity, is just one of many emerging examples for the need to revise our understanding of the haematopoietic hierarchy [68–70] (which may be flatter than had previously been thought). Here we apply our PID-based algorithm to the complete dataset of Psaila *et al.* [67] and infer a candidate network that depicts the statistical dependencies among the analysed genes. The resulting network (we restrict the figure to show the top 5% of edge confidence scores) is shown in Fig. 6 (All Data). Network structures based on pairwise MI scores from the same data showed skewed degree distributions with the bulk of the nodes unconnected (Fig. S5A).

Network inference requires variability in the data and the level of this variability is crucially linked to the ability to detect pairs of genes that vary in harmony. Here, we are interested in the genes involved in differentiation processes, and thus aim to detect statistical relationships associated with variability between different cell types, rather than analysing the heterogenous expression of genes within a single cell population. We therefore infer networks using subsets of the data comprising pairwise combinations of the three groups of cells identified in [67]; these networks differ from the network inferred using all the data, but much less than would be expected for any random pair of networks with the same number of nodes and degree sequence (Fig. S5B). Furthermore, the networks inferred from E-MEP + MK-MEP and Pre-MEP + MK-MEP data are more similar to each other than either is to the E-MEP + Pre-MEP data (Fig. 6). This could reflect that E-MEP cells are closer to Pre-MEP cells than they are to the MK-MEPs; this resonates well with the original authors' results.

We next wanted to consider the biological plausibility of specific interactions identified using our inference method (rather than considering the structural characteristics of the whole networks as above). To do so, we used two published mouse embryonic datasets [70, 71] where the roles of selected genes during the studied developmental processes are fairly well understood.

Fig. 7A shows the network inferred using data generated by Guo *et al.* comprising expression measurements of selected genes during early embryonic development (from oocyte to 64-cell blastocyst stages) [71]. We again inferred networks using subsets of these data — in this case, we considered overlapping subsets of 'early' and 'late' cells to reveal any temporal dependencies in the de-

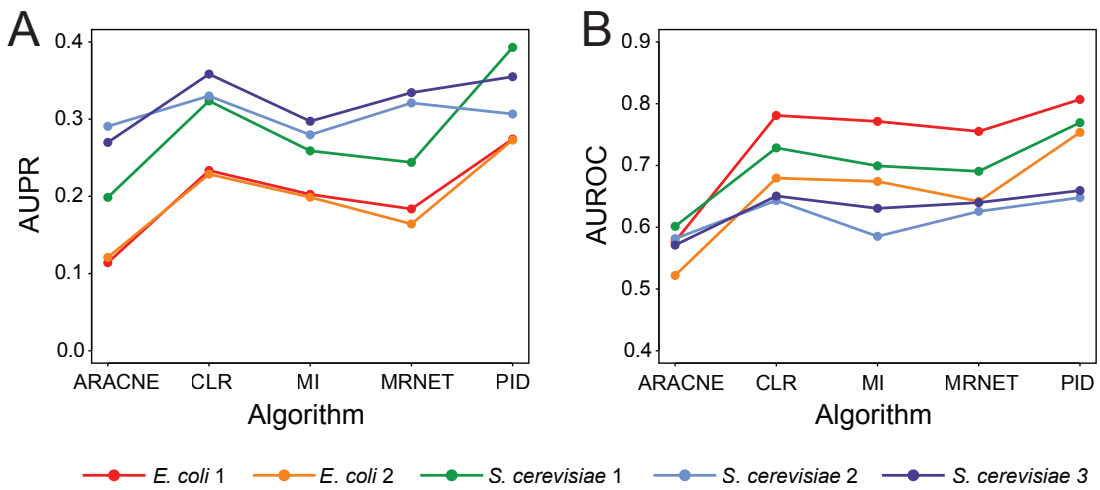

Figure 5: *AUPR* ***(A)*** *and AUROC (**B**) were calculated for several algorithms applied to five in silico datasets, generated from 50-node networks (Fig. S2). Using Bayesian Blocks discretization and the maximum likelihood estimator, the PID-based algorithm had the highest AUPR in three of the networks, and the highest AUROC in all five networks. The algorithms used are described in Methods and Background, with MI indicating the use of mutual information scores alone to rank edges (i.e. MI relevance network); the R package* `minet` *was used for the existing inference algorithms [65].*

tected interactions. A number of known relationships between genes are apparent in the network, e.g. upregulation of *Cdx2* and *Gata3* transcription factors (TFs) during the 8-cell to morula transition is identified as an edge in the 'early' network; while the co-expression of primitive endoderm specific TFs *Creb312* and *Sox17* is detected as an edge in the 'late' network (consistent with the appearance of distinct cell types, including primitive endoderm cells, in the blastocyst). A cluster of known pluripotency and reprogramming factors is also identified in the network (*Pou5f1*, *Nanog*, *Esrrb*, *Klf2* and *Klf4*) — importantly, *Sox2*, another key reprogramming factor is not connected with these genes, but is known to be upregulated later than the other factors [71].

Moignard *et al.* studied mouse embryonic haematopoietic development, and used sc-qPCR data to develop a Boolean network model of the GRN underlying blood development [70]. In Fig. 7B we compare the networks inferred from these same data using our PID-based algorithm and MI — we find the inferred PID network shares a higher number of edges with the Boolean network model, than the network constructed using MI values alone (see Fig. S6 for node labels). Although we of course do not know the 'true' structure of the GRN in this case (this is only feasible when using *in silico* data), the Boolean model was shown to capture key cell states observed experimentally and generated several experimentally-validated predictions [70], thus we use this as a benchmark to indicate the biological plausibility of our inferred networks.

Guidelines & limitations

Any comparative analysis of information-based GRN inference algorithms is influenced by a number of decisions, in particular: (i) how the data are discretized, (ii) the choice of MI estimator, and (iii) the metric used to evaluate performance. We discuss the impact of each of these decisions and offer guidelines for future analyses.

The information measures described here are all built upon estimates of discrete probability distributions, and yet mRNA expression data are continuous. Several algorithms and heuristics have been developed to discretize data and estimate empirical probability distributions from the resulting discrete frequencies. We investigated two methods for discretization, along with four MI estimators, as described in *Methods*.

All estimators were fairly accurate when used to calculate joint entropies for two uniformly distributed random variables, but for more variables their performance varies (Table 1). In such high-

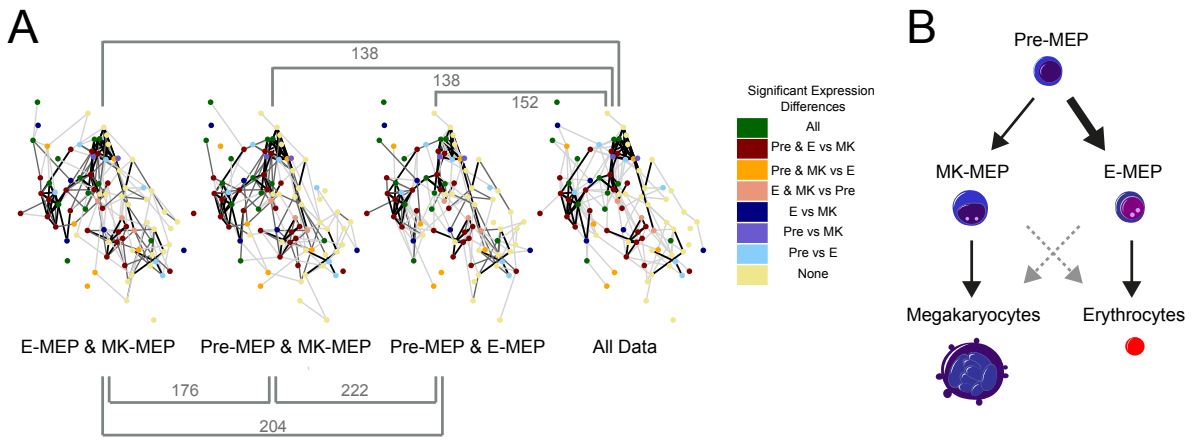

Figure 6: ***(A)*** *Networks inferred using the PID algorithm from subsets and the complete set of data (the top 5 % of edges are shown). Node colours reflect differences in expression profiles between the three cell populations and indicate which pairwise comparisons are statistically significant; we used Benjamini-Hochberg to correct for multiple testing by controlling the false discovery rate to be below 5 %. The numbers on the brackets above and below the networks indicate the distances (we used the network edit or Hamming distance) between pairs of networks (see Methods).* ***(B)*** *Illustration of the relationship between the three subpopulations of MEP cells — 'Pre-MEP' cells are enriched for erythroid/megakaryocyte progenitors but still retain some potential to differentiate into other cell types (myeloid cells); 'E-MEP' and 'MK-MEP' cells are strongly biased towards erythroid and megakaryocyte differentiation respectively (for details see [67]).*

Table 1: *Joint entropy estimates, in bits, for up to four independent random variables, calculated using different estimators. Each variable is uniformly distributed over $n = 64$ bins, with theoretical entropy, $\log_2(n)$. The theoretical joint entropy of independent variables is the sum of their entropies. The Dirichlet estimator is given priors of $\frac{1}{n}$ or 1, and the shrinkage estimator is given a uniform target.*

No. of variables	Theoretical	Maximum likelihood	Miller-Madow	Dirichlet ($\frac{1}{n}$)	Dirichlet (1)	Shrinkage (uniform)
1	6	5.9896	5.9973	5.9896	5.9911	6.0000
2	12	11.1629	11.4771	11.2134	11.8230	11.9998
3	18	11.9844	12.4803	15.9343	17.9915	18.0000
4	24	12.0000	12.4999	23.9283	23.9999	24.0000

dimension situations, the two estimators that take a prior, the Dirichlet and shrinkage estimators, retained their accuracy – at least when the correct priors were used. When given the wrong prior the Dirichlet and shrinkage estimators were less accurate (Table 2). Rank agreement between the estimators was increased by employing Bayesian Blocks discretization (Fig. S7); in light of these findings, use of the Bayesian Blocks algorithm is advised, and the maximum likelihood estimator is favoured due to its simplicity.

The choice of discretization method and estimator was found to influence the performance of the inference algorithms, with effects varying depending on the algorithm and on the true network (Fig. S8). Sampling frequency and dataset size also had an effect, with better performance in general with the availability of more data (Fig. S9). Due to the number of influential factors, we advise caution when interpreting the results of this or any such comparison as an exhaustive exploration of these factors is not feasible; however we note that the PID-based algorithm performed well in general across the tested combinations of discretization methods, estimators and datasets.

The metric used to evaluate algorithms also affects the apparent performance of the inference algorithms, evident here in the higher scores for AUROC than AUPR (Fig. 5 and S8). This is a well-documented phenomenon, caused by the true negatives in a GRN vastly outnumbering the true positives, meaning that AUPR is the more meaning-

Table 2: *Estimates of the difference between joint and marginal entropies of four independent random variables. The theoretical difference for independent variables is 0. Estimates are made for three sets of variables, drawn from three different distributions (uniform, normal or exponential). The Dirichlet estimator is given priors of $\frac{1}{n}$ and 1, and the shrinkage estimator is given a uniform target.*

	Maximum likelihood	Miller-Madow	Dirichlet ($\frac{1}{n}$)	Dirichlet (1)	Shrinkage (uniform)
Uniform	11.9523	11.4831	0.0240	0.0461	0.0009
Normal	8.7385	8.2674	3.1902	3.1170	3.1168
Exponential	4.5887	4.1413	7.2181	7.4598	6.5322

ful measure, despite AUROC being more widely-used [66].

Software

A new open-source package for estimating MVI measures was implemented in the Julia programming language [72]. The package, named `InformationMeasures.jl` supports information measures such as entropy, MI, conditional MI, interaction information and PID; the maximum likelihood, Miller-Madow, Dirichlet and shrinkage estimators; and the Bayesian Blocks, uniform width and uniform count discretization methods.

Julia was chosen for its speed (Fig. S10), clear mathematical syntax, growing availability of libraries and good integration with other languages. The existing `Discretizers.jl` package is used to implement the discretization methods (uniform width, uniform count, Bayesian Blocks); in some of our initial analyses we used the `AstroML` Python implementation of the Bayesian Blocks algorithm [73, 74].

In order to meet a wide range of requirements, the package can be used simply for discretizing data or for probability estimation. The formulae for the information measures can be accessed directly and applied to pre-discretized data or probability distributions that have already been estimated elsewhere.

A script that implements the PID-based network inference algorithm (using the Julia programming language [72]) is available from our website. The algorithm has complexity $O(n^3)$ meaning that it is best suited for small to medium-sized inference problems (Fig. S10). When analysing larger datasets (e.g. as obtained using scRNA sequencing where data is available for thousands of genes), a subset of (up to hundreds of) genes should first be selected, e.g. based on prior knowledge of the specific system being studied (or using functional annotations of genes), or those that show high variability between cell states or samples. This not only makes the subsequent network inference analysis computationally tractable, but will also aid interpreting the results. As is the case for other network inference algorithms, genes with no variability in mRNA expression are uninformative, and should always be removed prior to analysis.

Discussion

Numerous algorithms have been developed that aim to infer the structure of biological networks from observations of the constituent components. Many make use of information theoretic measures that, unlike correlation, enable detection of non-linear relationships between variables and, in some cases, allow us to explore statistical dependencies between three or more variables. Here, we demonstrate that decomposing the mutual information between three variables into different contributions using the multivariate information measure PID [47, 48] can help us to identify putative relationships between variables. We used *in silico* datasets generated from known networks to demonstrate that, as we might intuitively expect, the unique information terms (Eq. 4) are particularly informative about direct interactions.

Using these unique information terms, we developed an inference algorithm to score potential network edges in a manner which explicitly takes into account the local context for each gene rather than setting global thresholds for edge scores. Similar to CLR [50, 59], we use empirical distri-

butions to find the highest scoring edges for each gene in turn; both methods performed particularly well across our comparisons with different *in silico* datasets, indicating the value and relevance of this type of approach.

Considering gene context also impacts the types of network structures we infer: in many cases, we found that our approach results in more connected networks which capture the most important interactions for each node. By contrast, standard MI relevance networks tend to produce unbalanced graphs which leave many orphan nodes, and which have skewed degree distributions (e.g. Fig. S5) as a result of high-variability nodes predominating among the nodes involved in pairs with the highest MI values; such highly variable genes would act as false-positive generators. This should be borne in mind when selecting inference algorithms, as depending on our beliefs about the structure of the true network, and the aims of our analyses, we may want to emphasise (or elucidate) different network features. The choice of network inference algorithm thus implicitly biases the results that we get. In favour of information theoretical measures (including PID) is that this is an *equitable* measure of statistical dependency — by equitability we mean that we do not bias systematically in favour of certain types of statistical dependencies [75].

When comparing inference algorithms it is also crucial to consider — or at least be aware of — the many other factors (beyond the actual inference algorithm) that may influence the result. Choice of discretization and entropy estimation methods, and performance metrics, as well as the particular dataset can all affect final algorithm rankings. Although there are practical limits to how many factors we can consider, we emphasise the need to explore these issues to some extent to obtain reliable and fair comparisons. Here we demonstrate robustly good performance of our algorithm as we vary many of these influential factors.

Such comparisons will help us to understand the strengths and weaknesses of different approaches, thus allowing us to choose the most appropriate inference algorithms for our particular situation and data type. There is good reason to favour ensemble or committee-type predictors where more than one method is used to infer edges; their predictions are pooled; and a merged network is generated [13, 25–27, 76]. This simple description hides a great number of steps, few of which are statistically well worked out or understood. For example, it is unclear what the minimal requirements on an estimator are to be useful in an ensemble (except that this depends, of course, also on the whole panel of estimators included in the procedure). Neither do we understand how optimally to combine and weight different predictors. This is in principle straightforward in cases where we have training data (supervised learning), but this is rarely the case for the problems currently being addressed with single cell data. And while *in silico* performance lends credibility to an inference procedure (we should not trust a method that performs poorly under such a sanitized setting), this may inadequately reflect the performance in real settings, especially when our ambition is to model the dynamics (as well as the connectivity) of biological networks.

Algorithm selection may also be influenced by our prior knowledge about the underlying network structure when this is available [77]. Some inference algorithms incorporate these prior beliefs [78, 79], rather than a pure data-driven approach where the network solely depends on the current dataset. While in many cases this can enhance performance, it may not always be appropriate to rely on existing knowledge. For example, when analysing single cell data, we may want to avoid biasing our conclusions by relying on knowledge largely gained from population-averaged bulk data, in order to fully exploit the potential of single cell data to provide new insights.

In parallel to addressing these methodological challenges it is also important to develop better, more informative experiments. Single cell transcriptomic data often comprises cells harvested at different time-points (or under different conditions); in some cases, cells may be divided prior to subsequent transcriptomic analyses into distinct subpopulations / phenotypes using FACS or similar methods that rely on known cell-type markers. However, there is not necessarily a direct correspondence between sampling times and biologically interesting or relevant events *a priori* due to heterogeneity within the population (e.g. asynchronous development). Instead pseudo-temporal ordering is used to place cells into a sequence that may correspond to de-

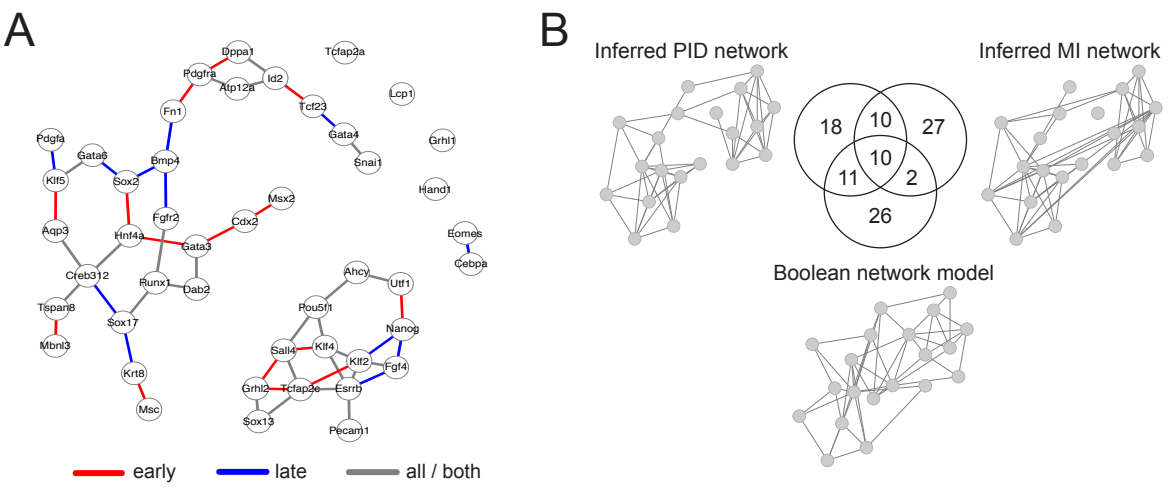

Figure 7: *(A) PID interaction network inferred using early embryonic development data (oocyte to E4.25 blastocyst stages) [71]. Graph edges indicate the top 5 % of putative interactions detected using the PID algorithm on the complete dataset. Networks were also inferred using two overlapping subsets of the data — an 'early' subset that includes all cells collected from oocyte up to 32-cell E3.5 blastocyst stages, and a 'late' subset including cells collected from 16-cell morula to 64-cell E4.25 blastocyst stages. Edge colours indicate temporal dependencies of the identified relationships; red indicates an edge ranks in the top 5 % of edges in the early network but not the late network, blue indicates the converse (late but not early), while grey indicates relationships without specific temporal dependencies (i.e. only present in the network constructed using the complete dataset, or present in both the early/late networks). (B) Comparison of networks inferred using haematopoietic development data in [70]. Moignard et al. used single-cell expression data for 20 transcription factors to infer a Boolean network model of blood development; we show a simplified representation of their model, where nodes (genes) are linked by an edge if those genes are either directly linked, or linked via one Boolean operation or set of update rules (genes linked in their model via a chain of multiple sets of update rules are not connected here). We used our PID algorithm or MI alone to infer networks of putative interactions between genes using these same data, and compare the edges identified in each of these three networks (numbers of shared edges are indicated by the Venn diagram). For node labels, see Fig. S6.*

velopmental change over time [1, 38, 80–82] (we note that recently these approaches have received some criticism as they are based on simplistic assumptions about the development process [4]). In situations where we believe these assumptions to be reasonable (and thus the inferred temporal orderings realistic), we could potentially make use of this temporal information to infer directionality of gene interactions within networks [22, 49, 83, 84]. As fine-tuning the temporal sampling scheme is less likely to be successful (except, perhaps, for the simple maxim "the more time points the better") than for bulk-data we need to identify other interventions that could lead to more informative data. When to collect data, and which external conditions should be considered, is still largely guided by domain expertise with all the advantages and problems that this entails. Simulation-based experimental design approaches [85–88] are required to ensure that we use single cell technologies to their best effect.

Methods that allow us to explore high-throughput datasets and identify putative functional relationships between genes are clearly needed to allow us to effectively analyse the large quantities of single cell transcriptomic data now available. It is crucial, however, to remember the limitations of such data-driven inference approaches when interpreting the results: we can only hope to detect relationships where there is sufficient variability in gene expression observed under the chosen experimental conditions (and we may thus miss some relevant functional relationships). The statistical relationships we do observe may be merely correlative (as without perturbation or temporal data we cannot judge causality), however, in many instances identifying genes which show coordinated changes in expression state will still be of biological interest (even if there is no direct regulatory relationship between the genes). The inferred relationships and network structure should be viewed as a way of identifying patterns and pu-

tative relationships from the data — which can be used as the basis for further targeted experiments or integration with other types of data — thus serving as a hypothesis-generation tool to guide promising areas for future research and investigation. Validating or invalidating any of these hypotheses experimentally may, of course, lead to a revision of the network model.

Conclusions

n this study, we describe the development of a network inference algorithm that is based on PID, an easily interpretable multivariate information measure. Using *in silico* single-cell gene expression datasets, we demonstrate better or comparable performance of this algorithm relative to several common inference algorithms that also rely on information-theoretic measures.

Our extensive comparisons highlight the importance of considering a multitude of influential factors when making such comparisons — the algorithms used to discretize continuous expression data and to estimate information theoretic measures, as well as the metrics used to evaluate performance can all influence the relative ranking of competing inference methods. Of course, the performance may also be affected by the nature of the dataset — including the sampling frequency, dataset size, and the structure of the true biological network. It is clearly infeasible to exhaustively consider all possible factors, but we emphasise the need to explore the influence of these variables in order to obtain robust and unbiased comparisons.

We apply our method to several published single-cell gene expression datasets to illustrate how we can explore statistical dependencies and identify putative biological relationships between the observed variables (genes). Single-cell gene expression data are increasingly common and, with advances in experimental technologies, dataset sizes (i.e. number of cells) will further increase enabling more accurate estimation of multivariate information measures. As with all network inference methods, we cannot expect to reconstruct the exact structure of the underlying biological networks, but instead view such methods as tools to explore the data; generate hypotheses; represent the current state of understanding; and guide further experiments, model development and analyses.

Methods

Calculating PID terms

The partial information terms are calculated using the specific information, I_{spec}, which quantifies the information provided by one variable about a specific state of another variable [89, 90]; e.g. if we consider the information provided by X about state z of variable Z,

$$I_{\text{spec}}(z;X) = \sum_{x \in X} p(x|z) \times \left(\log\left(\frac{1}{p(z)}\right) - \log\left(\frac{1}{p(z|x)}\right)\right). \quad (9)$$

If we consider $S = \{X_1, X_2\}$ and a target variable Z, the redundant contribution is calculated by comparing the amount of information provided by each variable within S about each state of the target Z,

$$\text{Redundancy}(Z; X_1, X_2) = \sum_{z \in Z} p(z) \min_{X_i} I_{\text{spec}}(z; X_i). \quad (10)$$

Unique and synergistic contributions can then be calculated using the redundancy and MI terms (Eq. 4, 5).

Discretization algorithms

In order to use the entropy estimators described here, continuous datasets must first be discretized. A number of algorithms exist to define the total number and boundaries of the resulting partitions (or bins). One common simple approach is to use bins of equal width, with the number of bins determined heuristically, e.g. here we use the nearest integer to the square root of the size of the dataset, $\sqrt{n}$ [91]. A more sophisticated approach is the Bayesian Blocks algorithm [74], in which the number and widths of bins are chosen by optimising a fitness function, without constraining the bins to be of equal width.

Estimators

When a dataset is large enough, the empirical frequencies can be considered to be an approximation of the true probabilities, referred to here as the maximum likelihood approach. For sparser datasets, a number of methods have been developed either for estimating the probability distribution from a set of frequencies, such as the Dirichlet estimator and the shrinkage estimator, or for estimating the entropy directly, as in the Miller-Madow estimator [92, 93].

The Dirichlet estimator refers to a group of Bayesian estimators that take a Dirichlet distribution as prior, but each with different parameters [92, 94]. There is no consensus on the best parameters to use, despite several proposed alternatives [92]; here, we use the same parameter, 1, for each bin unless otherwise stated.

The shrinkage estimator is also Bayesian, compromising between the observed frequencies, unbiased but with a high variance, and a prior (or target) distribution, biased but with low variance [92]. The estimate is affected by both the choice of target distribution and the weight given to the target (or shrinkage intensity). In the current analysis the optimal shrinkage intensity is calculated as described in [92], and the target distribution is the uniform distribution.

The Miller-Madow estimator is an entropy bias correction that does not estimate the probability distribution, and therefore cannot be meaningfully applied to higher order information measures. Despite this it has been applied for the comparison of different MI-based algorithms [65], and so it is included in this analysis, with the caveat that its meaning is unclear.

Simulation of *in silico* 3-gene network data

We considered six 3-gene topologies (Fig. S1), and used the Gillespie algorithm [95] to generate stochastic simulations of gene expression timecourse data using two alternative model definitions (based on thermodynamic or mass action kinetics). In both cases, we include an additional activating stimulus that is present from halfway through the simulation timecourse. This additional stimulus acts to perturb the system away from a steady state, driving changes in gene expression that are necessary for relationships between genes to be observable. For model and simulation details, see *SI Methods*.

Simulation of *in silico* GeneNetWeaver network data

Data were simulated using GeneNetWeaver [64], which was used in the DREAM5 (Dialogue on Reverse Engineering Assessment and Methods) network inference competition [27]. GeneNetWeaver uses dynamical models that consider mRNA transcription and translation processes and generates stochastic time series simulations. To mimic single cell data, we simulated 700 time series experiments for each network, with mRNA measurements generated at seven time points; we sampled a single time point from each time series, resulting in gene expression measurements for 700 'cells' (100 cells at each time point).

Methods for analysis of real datasets

Published datasets

We analysed three published qRT-PCR datasets to illustrate our network inference algorithm. Normalised Ct values from Psaila *et al.* [67] were subtracted from the assumed maximum, 40, and the resulting dCt values used in our analyses. Normalised dCt values from Moignard *et al.* [70] were used directly for our analyses. Raw Ct data from Guo *et al.* [71] were treated as described by the original authors (dCt values were calculated assuming a limit of detection of 28, and normalised on a cell-wise basis by subtracting the mean expression of housekeeping genes *Actb* and *Gapdh*; all values corresponding to expression below the limit of detection were set to -15).

Network Distances and Statistical Significance

The edit distance of a symmetric network [96] is defined here as

$$d(\mathcal{G}_1, \mathcal{G}_2) = \sum_{0 \leq i \leq j \leq N} |A^1_{ij} - A^2_{ij}|$$

where A^q_{ij} denotes the (i,j) element of the adjacency matrix A^q corresponding to network $\mathcal{G}_q$; for binary matrices such as here, this is simply a Hamming distance, only for undirected networks can we restrict the distance calculation to the upper or lower triangular matrix.

In order to test whether this distance is in line with expectations, we compare to the configuration graph ensemble corresponding to the networks compared. That is for both $\mathcal{G}_1$ and $\mathcal{G}_2$ we randomly rewire edges but keep the degree of each node fixed, and we repeat this procedure 10,000 times. This approach is appropriate as the most general null model to generate confidence intervals for network properties [97].

Availability of data and materials

The `InformationMeasures.jl` package is available from our website.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

TEC, MPHS, and ACB designed and performed research, analysed data, and wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by a Biotechnology and Biological Sciences Research Council (BBSRC) DTP Studentship and a BBSRC Future Leaders Fellowship.

SI Methods

Simulation of *in silico* 3-gene network data

The first *thermodynamic* model includes seven species: mRNA (x_i) and protein (y_i) corresponding to three genes ($i = 1, 2, 3$), and a stimulating ligand (s) which targets a selected gene. We define the following reaction types with associated propensities,

Reaction	Propensity
$x_i \rightarrow \emptyset$	λx_i
$y_i \rightarrow \emptyset$	λy_i
$x_i \rightarrow x_i + y_i$	$r_i(x_i)$
$\emptyset \rightarrow x_i$	$f_i(\mathbf{y}, s)$

to represent mRNA decay, protein decay, translation and transcription respectively, where λ is the protein / mRNA decay rate, translation is modelled according to saturation kinetics (with maximum rate $\alpha_{\text{translation}}$), i.e.,

$$r_i(x_i) = \alpha_{\text{translation}} \cdot \frac{1}{1 + (k_i/x_j)},$$

and transcription rates depend on the concentration of any regulating proteins (including the stimulating ligand s if present) according to the relationship,

$$f_i(\mathbf{y}, s) = \alpha_{\text{transcription}} \cdot \sum_{m=0}^{M} \alpha_m P\{S_m\},$$

where $\alpha_{\text{transcription}}$ is a constant transcription rate, M is the total number of possible states S_m for gene i (either unbound, or bound by one or two regulating proteins), and α_m is the relative activation rate for each state. The probability of each state, $P\{S_m\}$, depends on the concentrations of the regulating proteins, modelled according to standard thermodynamic principles (see e.g. [26] for details). For example, if a gene has two possible regulators (proteins y_j, y_k), we calculate the mean activation of transcription of the target gene i using the function,

$$f_i(\mathbf{y}) = \frac{\alpha_0 + \alpha_j \chi_j + \alpha_k \chi_k + \alpha_{jk} \chi_j \chi_k}{1 + \chi_j + \chi_k + \chi_j \chi_k},$$

where $\chi_j = (y_j/k_j)$, k_j is the dissociation constant, and the possible states of the gene are unbound, bound by y_j or y_k alone, or by both y_j and y_k. For our models, the maximum number of regulators for a given gene is three (proteins y_j and y_k plus the stimulating ligand s), but we assume a maximum of two regulators can bind the gene at any one time (so we consider the gene states with each possible pair of ligands bound, but not a state with all three bound).

For the thermodynamic model, we simulated timecourses from time 0 to 1000, with the stimulating ligand s present from time 500 at a constant level of 20 molecules, and recorded the system state at 40 equally spaced intervals. We repeated the simulations 25 times — and calculated PID scores using the resulting data — with the stimulus targeting each of the three genes in the network in turn (75 simulations in total, 3 sets of PID scores). We randomly sample initial mRNA (x_i) and protein (y_i) levels from a $U(0, 5)$ distribution. Model parameters were $\lambda = 0.02, \alpha_{\text{transcription}} = 2, \alpha_{\text{translation}} = 2$, and $k_i = 50$ (for all $i = 1, 2, 3$). Relative activation rates for transcription, α_m, depended on the number of activating and inhibiting regulators present in each possible state S_m: $\alpha_m = 0.1$ (for the unbound state, i.e. basal transcription), 0.001 if an inhibitor was bound (we assumed inhibition dominated activation), and 5 if only activating regulators were bound.

The second *mass action* model that we consider also includes seven species: genes (g_i) and

mRNA/protein (x_i) for $i = 1, 2, 3$, and the stimulating ligand s. We assume that protein and mRNA concentrations are equal (i.e. translation is instantaneous) and, unlike our first model, assume that a gene can only be bound by a single protein at any time. The possible reactions and associated propensities are,

Reaction			Propensity
g_i	$\rightarrow$	$g_i + x_i$	$k_{\text{txn}} \cdot x_i$
x_i	$\rightarrow$	$\emptyset$	$k_{\text{decay}} \cdot x_i$
$g_i + x_j$	$\rightarrow$	$g_i x_j$	$k_{\text{on}} \cdot g_i \cdot x_j$
$g_i x_j$	$\rightarrow$	$g_i + x_j$	$k_{\text{off}} \cdot g_i x_j$
$g_i x_j$	$\rightarrow$	$g_i x_j + x_i$	$k_{\text{regulated}} \cdot g_i x_j$

for basal transcription and protein decay, protein binding and unbinding from a target gene, and transcription from a gene bound to a regulating protein respectively (where $g_i x_j$ indicates gene i is bound by regulating protein j).

For the mass action model, we simulated timecourses from time 0 to 400, with the stimulating ligand present from time 200 at a constant level of 20 molecules, and recorded the system state at 20 equally spaced intervals. We repeated the simulations 50 times with the stimulating ligand targeting each gene. We initiated simulations with two copies of each gene ($g_i = 2$), no stimulating ligand, and initial mRNA/protein levels sampled from a uniform distribution ($x_i \sim U(0, 50)$). Model parameters were $k_{\text{txn}} = 1, k_{\text{decay}} = 0.05, k_{\text{on}} = 0.01, k_{\text{off}} = 0.25$, and $k_{\text{regulated}} = 10$ or 0.1 for activating and inhibiting regulation respectively. Information measures were calculated from these data using the Matlab package written by Timme *et al.* [47], following discretisation with the AstroML implementation of the Bayesian Blocks algorithm [73, 74].

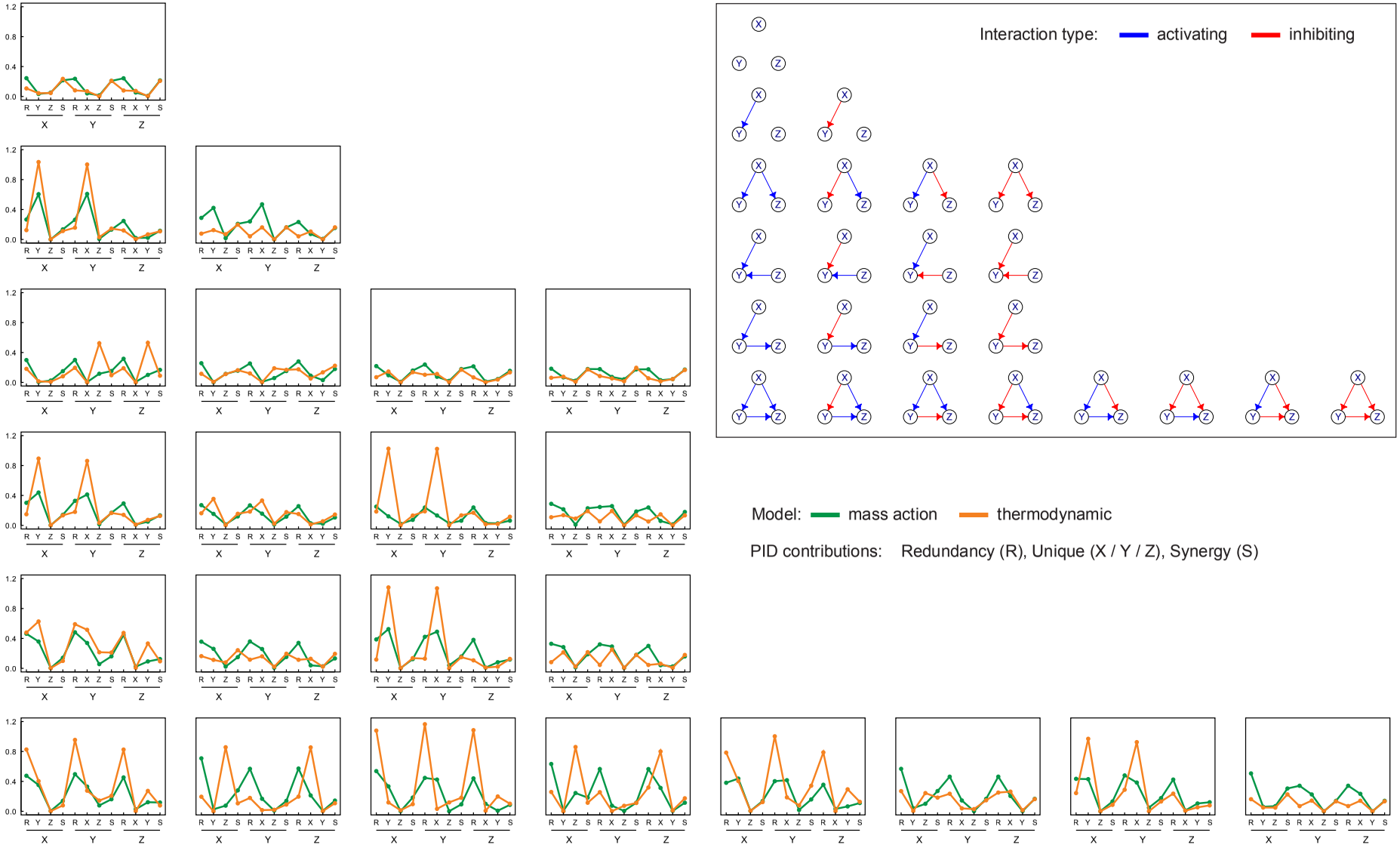

Figure S1A: *Mean PID values for 3-gene networks with different topologies (extension of results in Fig. 2). PID values were calculated using data simulated from 3-gene networks with the topologies illustrated in the top right inset (blue and red arrows indicate activating and inhibiting interactions respectively). Each line graph in the main figure shows the mean PID values calculated using models with the topology indicated in the equivalent grid position (i.e. the same row/column); the horizontal axis labels indicate the PID contribution, e.g. the first four values show the PID values with gene X as the target, consisting of the redundancy (R), unique contributions from gene Y (Y) and gene Z (Z), and the synergistic contribution (S), red. The models used for simulation assumed mass action (green) or thermodynamic (orange) kinetics, and the additional stimulation ligand targeted gene X from halfway through the simulation time (see SI Methods). The values plotted are the mean PID values calculated from five sets of simulations (with different randomly sampled initial conditions). Note that the same profiles are not necessarily seen for networks with equivalent connectivity, but different types of edges (results in the same row of this figure) — this is to be expected, as statistical relationships between genes can only be detected when there is sufficient variability in the observed data (which may not occur in some cases, e.g. if expression of a gene is inhibited throughout the simulation time).*

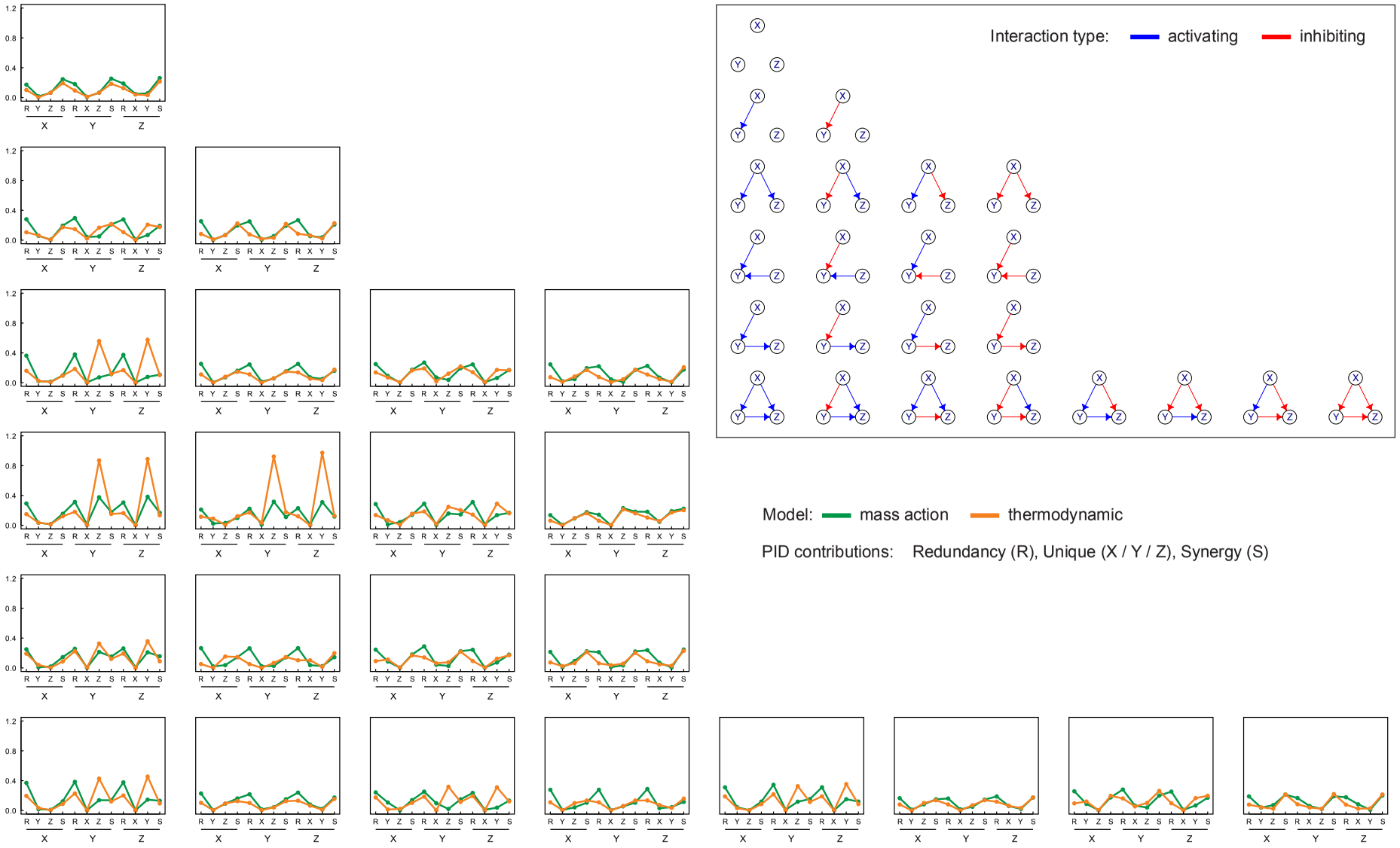

Figure S1B: *Mean PID values for 3-gene networks with different topologies, as in Fig. 2 and Fig. S1A, but in this case simulations were carried out with the additional stimulating ligand targeting gene Z. These results again demonstrate the need for variability of gene expression levels in the data in order to detect statistical relationships. In these simulations, variability is driven by introducing an additional stimulating ligand to perturb the system from a steady state during the simulation timecourse, but many of the topologies do not enable transmission of this signal to genes X or Y if this ligand targets gene Z. For example, for the one edge activating topology (second row, first column), we cannot observe the high unique scores indicative of an edge connecting genes X and Y under this scenario, but this edge is clearly detected in the simulated data analysed in Fig. S1A.*

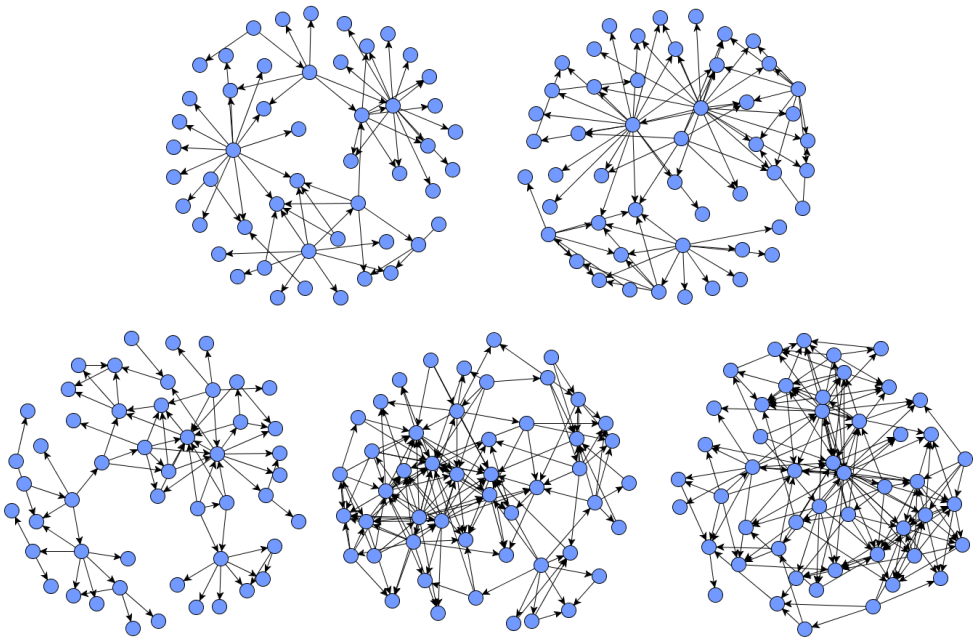

Figure S2: *In silico data were generated from models of five 50-gene networks with structures that mimic known transcriptional sub-networks from E. coli (top row) and S. cerevisiae (bottom row); simulated data and the network images here were generated using GeneNetWeaver [64].*

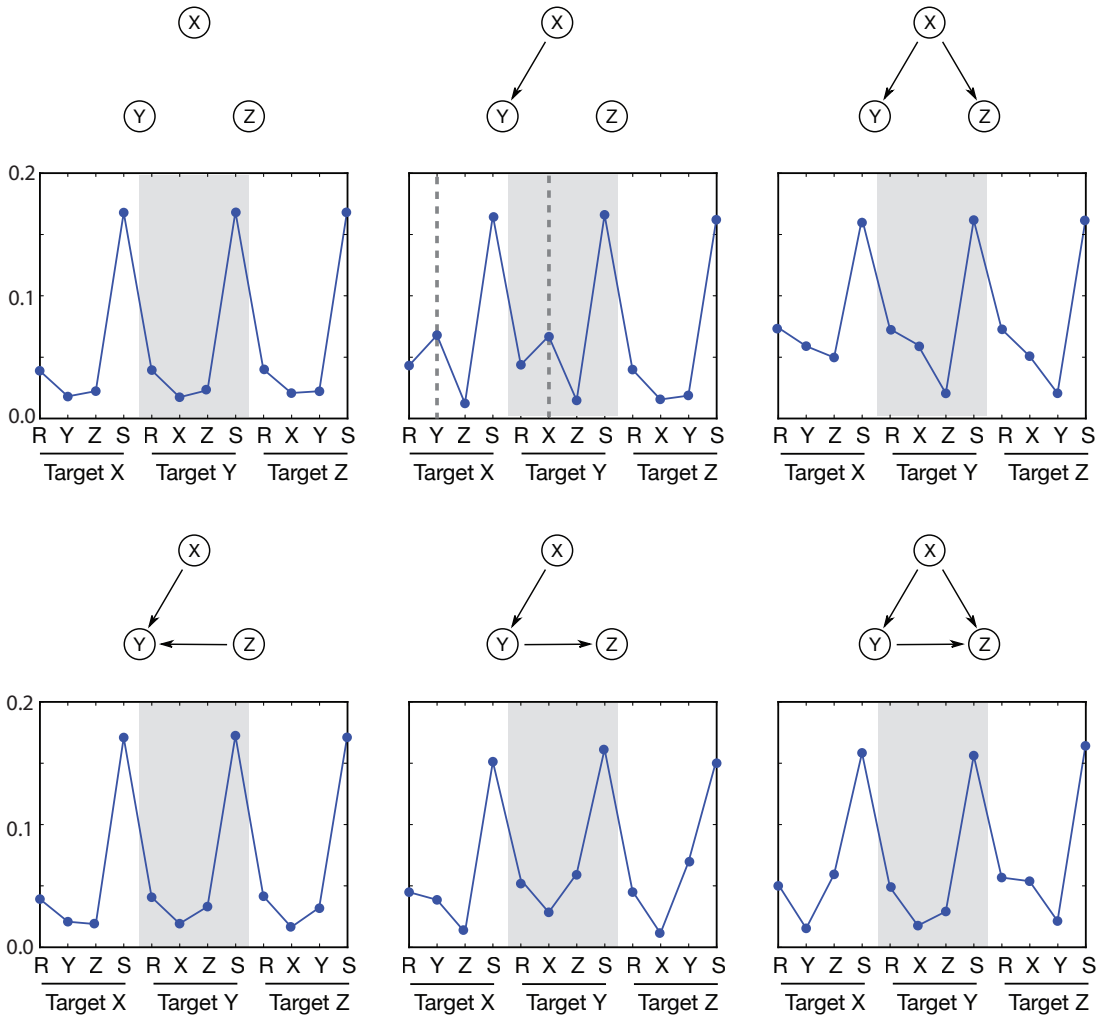

Figure S3: *Every triplet of nodes (genes) in the network was assigned to one of six possible classes (based on the known connectivity of genes, as indicated in network diagrams above each plot). For each triplet, twelve PID values were calculated from the simulated data — there are four PID contributions with each gene treated as the target gene in turn, consisting of a redundant, synergistic and two unique contributions (Eq. 4). Each line graph shows the mean PID values calculated across triplets with the same topology, with the horizontal axis labels indicating the PID contribution, e.g. the first four values show the PID values with gene X as the target, consisting of the redundancy (R), unique contributions from gene Y (Y) and gene Z (Z), and the synergistic contribution (S).*

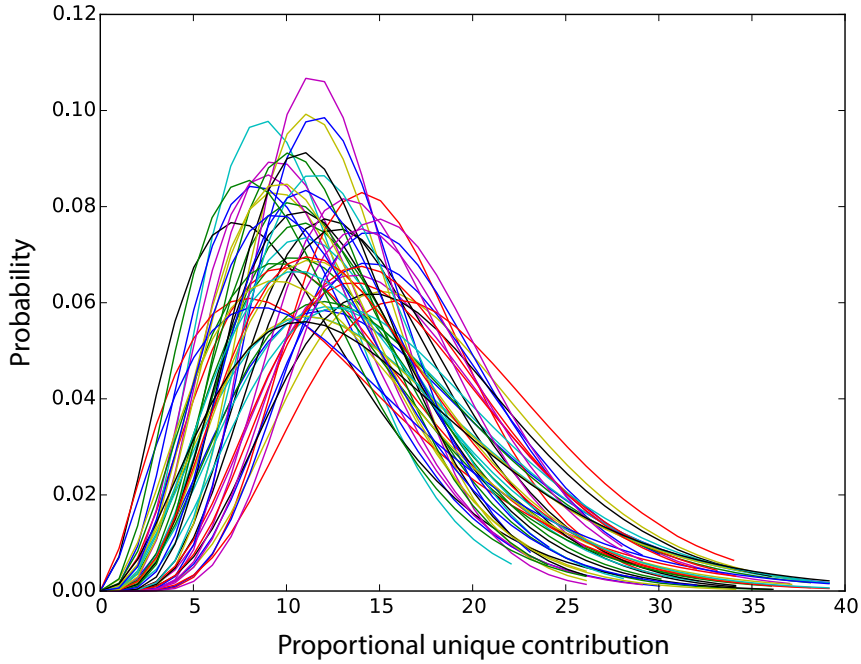

Figure S4: *Gamma distributions were fitted to the proportional unique contribution (PUC) scores (Eq. 7) for each gene in a 50-node in silico network (for each gene X, a PUC score, $u_{X,Y}$, is obtained for that gene paired with each other gene Y in the network). Due to the variability of these distributions, using a universal threshold for inferring edges is problematic, thus we use the cumulative probability distributions for each gene to obtain a final confidence score for network edges.*

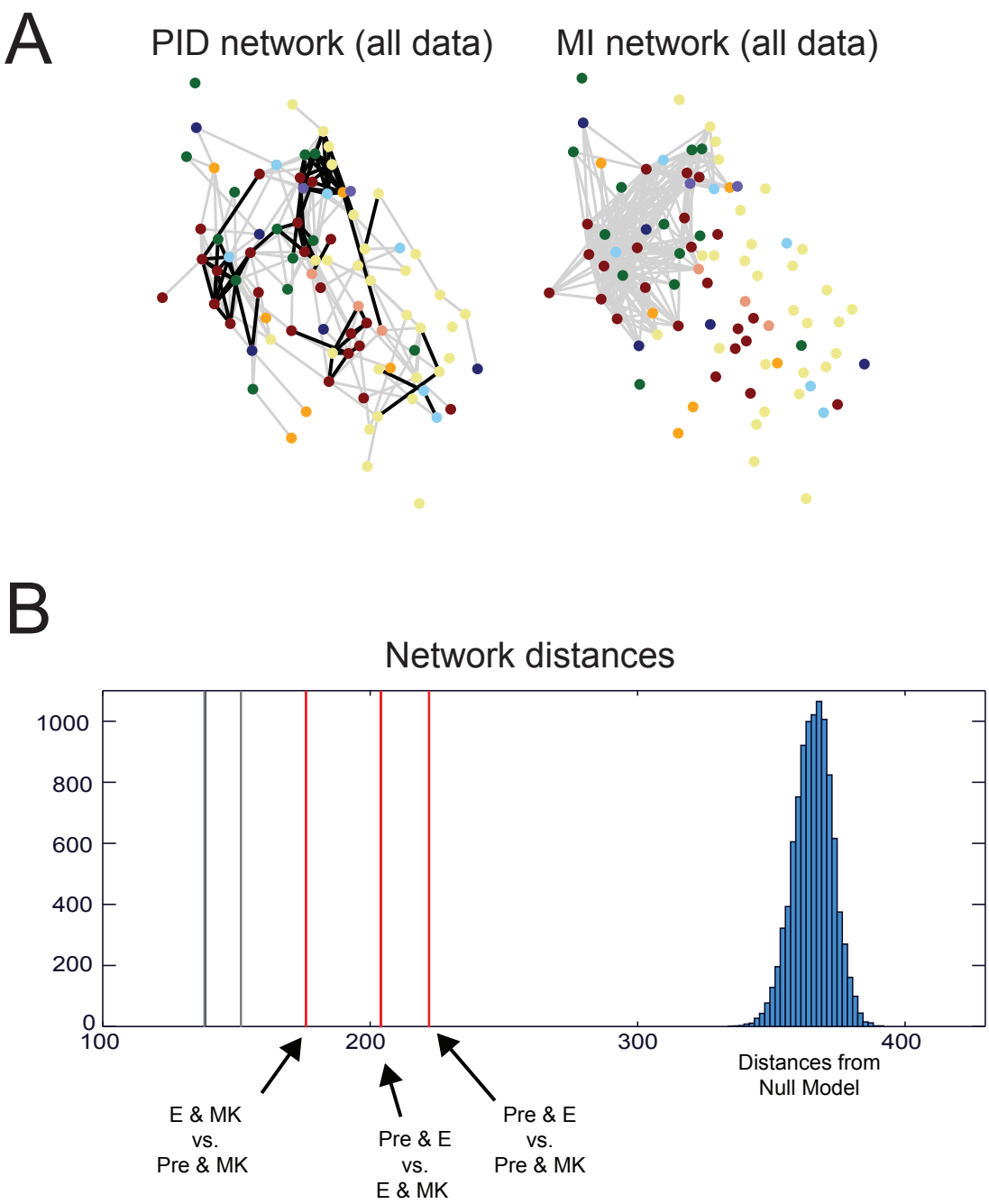

Figure S5: ***(A)*** *Comparison of the networks obtained using our PID algorithm (left) and MI values (right) to detect relationships between genes using data from Psaila et al. [67]; in both cases, the top 5 % of PID confidences scores or pairwise MI values are shown as edges.* ***(B)*** *Illustration of distances calculated between the networks inferred using the PID algorithm, and different subsets of the single-cell data (see Results and Methods for details). The blue histogram illustrates distances between pairs of random networks (with the same number of nodes and degree sequence as the inferred networks, but random connectivity); vertical grey lines indicate the distances calculated between the networks inferred using each subset of the data and the complete dataset; and vertical red lines indicated the distances between networks inferred using each subset of the data.*

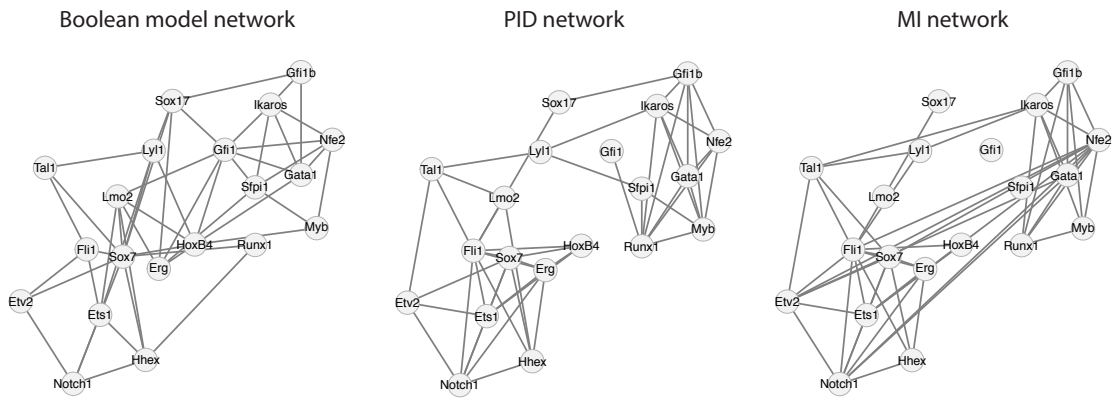

Figure S6: *Three networks inferred using sc-qPCR data from Moignard et al. [70]. A Boolean logic model was inferred by the original authors [70] — here, the left image shows a simplified version of their model where edges are only included between genes linked by one or a single set of Boolean update functions; the middle and right images show the networks we infer using our PID-based algorithm or MI from the same dataset (note that we restrict these networks to have the same number of edges in total as the Boolean model).*

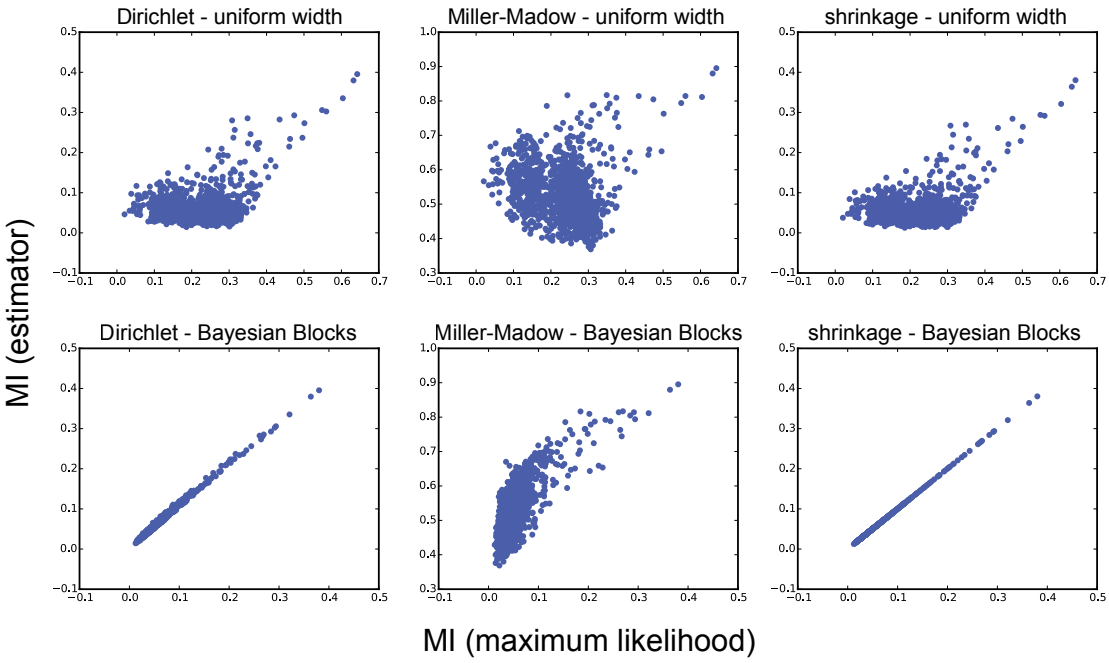

Figure S7: *MI was estimated for every pair of genes in a 50-gene in silico network, using different combinations of discretization algorithms (uniform width or Bayesian Blocks [73, 74]) and MI estimators (Dirichlet, Miller-Madow, shrinkage and maximum likelihood); see Methods for details. Each plot shows the relative ranks of MI scores obtained using the maximum likelihood estimator (horizontal axis) versus one of the other estimators (vertical axis); the top and bottom rows show results obtained using data discretized using a uniform width or Bayesian Blocks algorithm respectively. MI ranks were the most consistent when using Bayesian Blocks discretization.*

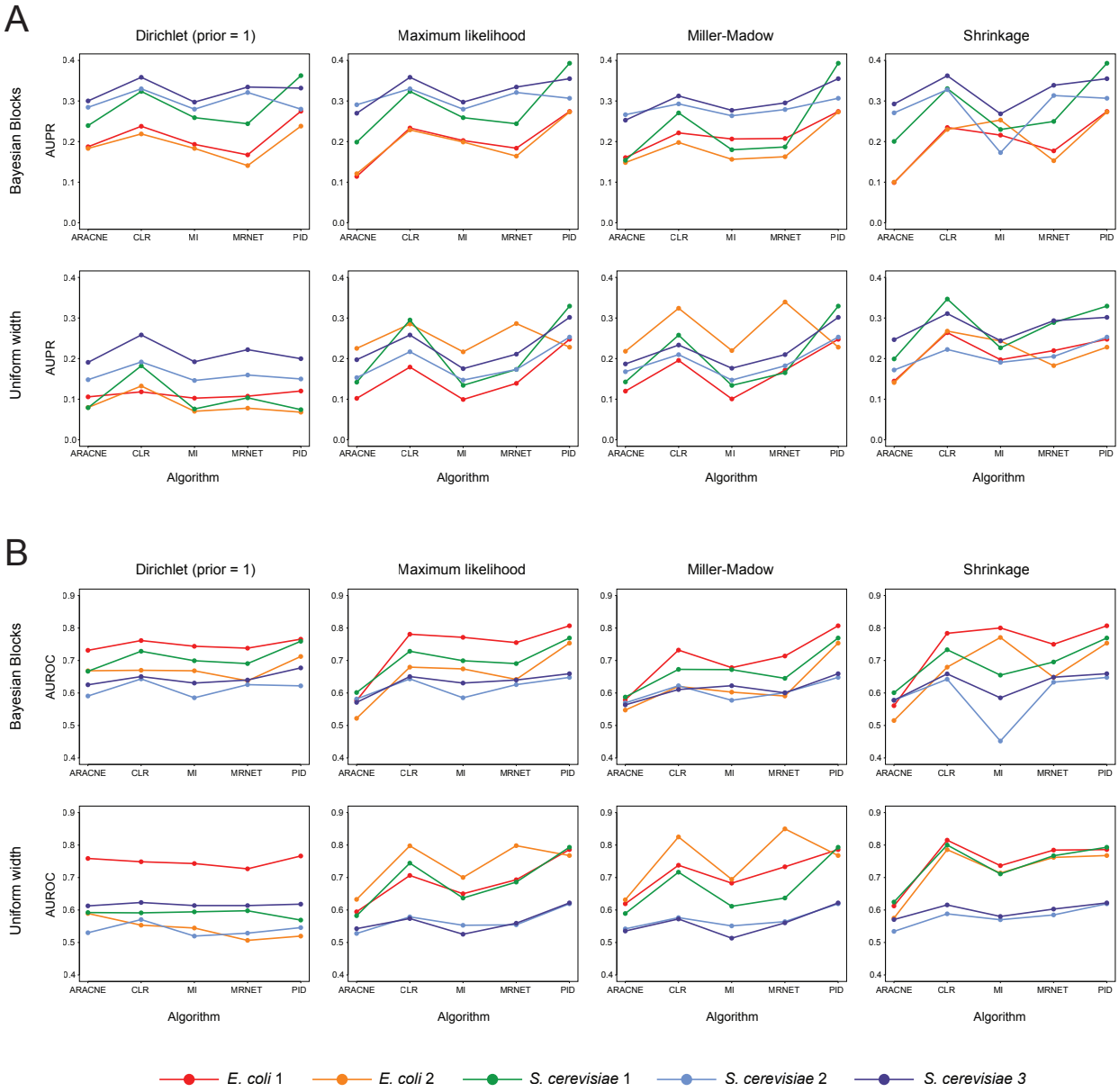

Figure S8: *AUPR* ***(A)*** *and AUROC* ***(B)*** *scores quantifying the accuracy of inferred networks calculated using in silico data simulated from five 50-gene networks (see Methods). Coloured lines indicate the results obtained with different datasets. Each plot shows the results obtained using a different combination of discretization algorithm (rows) and MI estimator (columns). Choice of algorithm and estimator clearly affects the relative scores of the network inference algorithms, suggesting that comparisons made using only one combination should be interpreted with caution; we find that the PID-based inference algorithm performs consistently well across the different combinations. The R package* `minet` *was used to implement the existing network inference algorithms [65].*

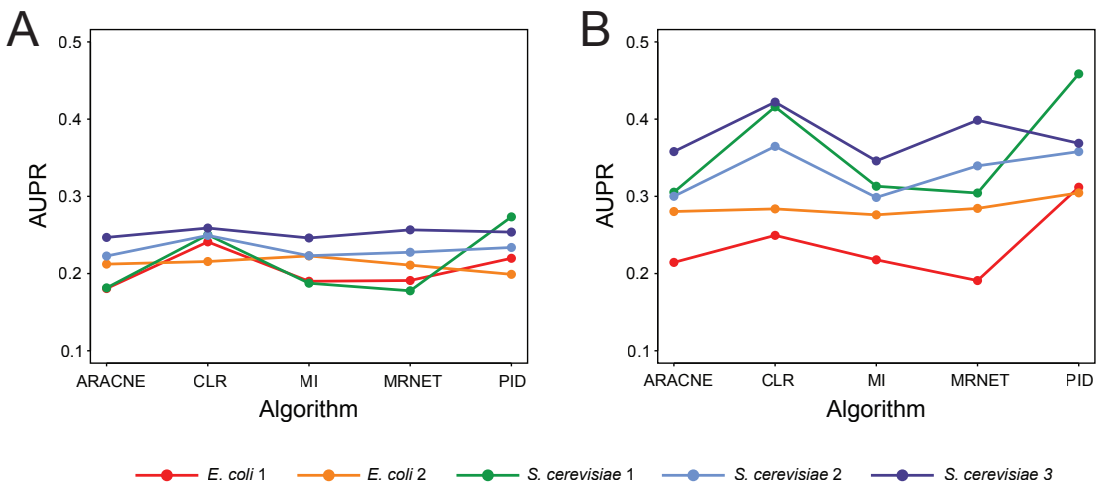

Figure S9: *Networks were inferred using alternative datasets generated from the 50-gene in silico networks, with a shorter time course and fewer cells (left) or a longer time course and more cells (right). AUPR scores indicate the performance of different methods — in most cases, as expected, more data led to better performance. The R package* `minet` *was used to implement the existing network inference algorithms [65].*

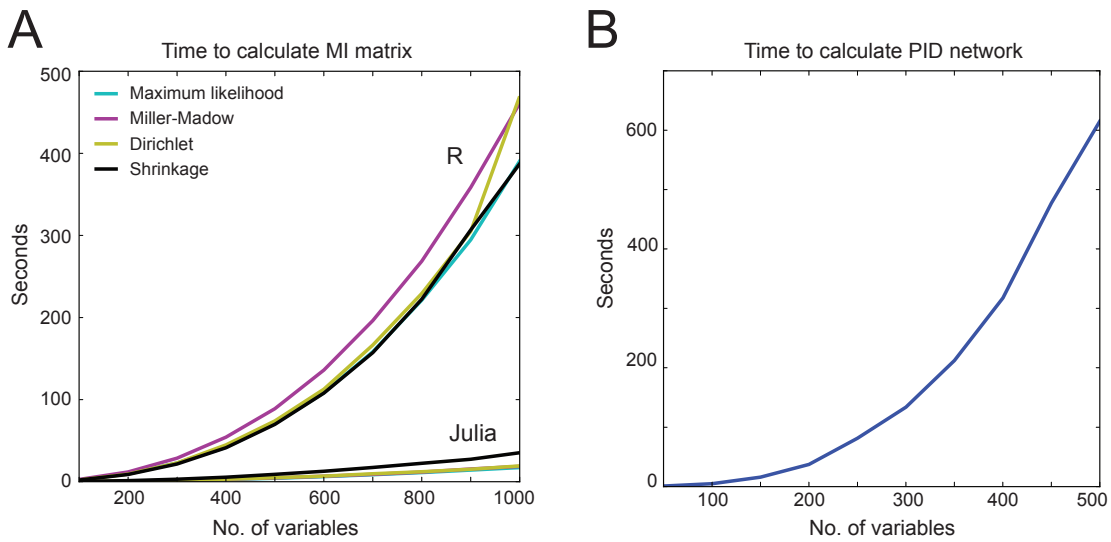

Figure S10: ***(A)*** *Comparison of the times taken to calculate a matrix of pairwise MI values using the R package* `minet` *[65] and our Julia package* `InformationMeasures.jl`*. Input data were simulated expression values for up to 1000 genes, with 700 values per gene; data were discretised using the uniform width algorithm and times were measured using inbuilt functions in R and Julia.* ***(B)*** *Time taken to infer networks of varying sizes using the PID-based inference algorithm implemented in Julia. Networks were inferred for simulated datasets of up to 500 genes, with 700 expression values per gene.*

References

[1] Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., Rinn, J.L.: The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature biotechnology **32**(4), 381–251 (2014). doi:10.1038/nbt.2859

[2] Harrington, H.A., Azogui, H.H., Yahalom-Ronen, Y., Plotnikov, A., Stumpf, M.P.H., Stumpf, M.P.H.: Nuclear to cytoplasmic shuttling of ERK promotes differentiation of muscle stem/progenitor cells. Development **141**(13), 2611–2620 (2014). doi:10.1242/dev.107078

[3] Rué, P., Martinez Arias, A.: Cell dynamics and gene expression control in tissue homeostasis and development. Molecular Systems Biology **11**(2), 792–792 (2015). doi:10.15252/msb.20145549

[4] Moris, N., Pina, C., Martinez Arias, A.: Transition states and cell fate decisions in epigenetic landscapes. Nature Reviews Genetics (2016). doi:10.1038/nrg.2016.98

[5] Gouti, M., Metzis, V., Briscoe, J.: The route to spinal cord cell types: a tale of signals and switches. Trends in Genetics **31**(6), 282–289 (2015)

[6] Göttgens, B.: Regulatory network control of blood stem cells. Blood **125**(17), 2614–2620 (2015)

[7] De Smet, R., Marchal, K.: Advantages and limitations of current network inference methods. Nature Reviews Microbiology **8**(10), 717–729 (2010). doi:10.1038/nrmicro2419

[8] Oates, C.J., Mukherjee, S.: Network Inference and Biological Dynamics. The annals of applied statistics **6**(3), 1209–1235 (2012). doi:10.1214/11-AOAS532

[9] Thorne, T.W., Stumpf, M.P.H.: Inference of temporally varying Bayesian networks. Bioinformatics **28**(24), 3298–3305 (2012). doi:10.1093/bioinformatics/bts614

[10] Thorne, T.W., Fratta, P., Hanna, M.G., Cortese, A., Plagnol, V., Fisher, E.M., Stumpf, M.P.H., Stumpf, M.P.H.: Graphical modelling of molecular networks underlying sporadic inclusion body myositis. Molecular Biosystems **9**(7), 1736–1742 (2013). doi:10.1039/c3mb25497f

[11] Siegenthaler, C., Gunawan, R.: Assessment of Network Inference Methods: How to Cope with an Underdetermined Problem. PLoS One **9**(3), 90481 (2014). doi:10.1371/journal.pone.0090481

[12] Oates, C.J., Amos, R., Spencer, S.E.F.: Quantifying the multi-scale performance of network inference algorithms. Statistical Applications in Genetics and Molecular Biology **13**(5), 611–631 (2014). doi:10.1515/sagmb-2014-0012

[13] Huang, X., Zi, Z.: Inferring cellular regulatory networks with Bayesian model averaging for linear regression (BMALR). Molecular BioSystems **10**(8), 2023–2030 (2014). doi:10.1039/c4mb00053f

[14] Young, W.C., Raftery, A.E., Yeung, K.Y.: Fast Bayesian inference for gene regulatory networks using ScanBMA. BMC Systems Biology **8**(1), 47 (2014). doi:10.1186/1752-0509-8-47

[15] Thomson, M., Liu, S.J., Zou, L.-N., Smith, Z., Meissner, A., Ramanathan, S.: Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. **145**(6), 875–889 (2011). doi:10.1016/j.cell.2011.05.017

[16] Peter, I., Davidson, E.H.: Genomic Control Process. Academic Press, London, UK (2015)

[17] Clark, E., Akam, M.: Odd-paired controls frequency doubling in *Drosophila* segmentation by altering the pair-rule gene regulatory network. eLife **5**, 18215 (2016). doi:10.7554/eLife.18215

[18] Penfold, C.A., Wild, D.L.: How to infer gene networks from expression profiles, revisited. Interface Focus **1**(6), 857–870 (2011)

[19] Bonneau, R., Reiss, D.J., Shannon, P., Facciotti, M., Hood, L., Baliga, N.S., Thorsson, V.: The Inferelator: an algorithm for learning parsimonious regulatory networks from

systems-biology data sets de novo. Genome ... (2006)

[20] Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R., Califano, A.: ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. BMC Bioinformatics **7**(Suppl 1), 7–15 (2006)

[21] Villaverde, A.F., Banga, J.R.: Reverse engineering and identification in systems biology: strategies, perspectives and challenges. Journal of The Royal Society Interface **11**(91), 20130505–20130505 (2013)

[22] Villaverde, A., Ross, J., Banga, J.: Reverse Engineering Cellular Networks with Information Theoretic Methods. Cells **2**(2), 306–329 (2013)

[23] Liang, K.-C., Wang, X.: Gene Regulatory Network Reconstruction Using Conditional Mutual Information. EURASIP journal on bioinformatics & systems biology **2008**(1), 1–14 (2008). doi:10.1155/2008/253894

[24] Madar, A., Greenfield, A., Vanden-Eijnden, E., Bonneau, R.: DREAM3: Network Inference Using Dynamic Context Likelihood of Relatedness and the Inferelator. PLoS ONE **5**(3), 9803–13 (2010)

[25] Hill, S.M., Heiser, L.M., Cokelaer, T., Unger, M., Nesser, N.K., Carlin, D.E., Zhang, Y., Sokolov, A., Paull, E.O., Wong, C.K., Graim, K., Bivol, A., Wang, H., Zhu, F., Afsari, B., Danilova, L.V., Favorov, A.V., Lee, W.S., Taylor, D., Hu, C.W., Long, B.L., Noren, D.P., Bisberg, A.J., Consortium, H.-D., Mills, G.B., Gray, J.W., Kellen, M., Norman, T., Friend, S., Qutub, A.A., Fertig, E.J., Guan, Y., Song, M., Stuart, J.M., Spellman, P.T., Koeppl, H., Stolovitzky, G., Saez-Rodriguez, J., Mukherjee, S.: Inferring causal molecular networks: empirical assessment through a community-based effort. Nature Methods **13**(4), 310–318 (2016)

[26] Marbach, D., Prill, R.J., Schaffter, T., Mattiussi, C., Floreano, D., Stolovitzky, G.: Revealing strengths and weaknesses of methods for gene network inference. Proceedings of the National Academy of Sciences **107**(14), 6286–6291 (2010)

[27] Costello, J.C., Vega, N.M., Camacho, D.M., Allison, K.R., Aderhold, A., Allison, K.R., Bonneau, R., Camacho, D.M., Chen, Y., Cordero, F., Costello, J.C., Crane, M., Dondelinger, F., Drton, M., Esposito, R., Foygel, R., de la Fuente, A., Gertheiss, J., Geurts, P., Greenfield, A., Grzegorczyk, M., Haury, A.-C., Holmes, B., Hothorn, T., Husmeier, D., Huynh-Thu, V.A., Irrthum, A., Karlebach, G., Küffner, R., Lèbre, S., De Leo, V., Madar, A., Mani, S., Marbach, D., Mordelet, F., Ostrer, H., Ouyang, Z., Pandya, R., Petri, T., Pinna, A., Poultney, C.S., Prill, R.J., Rezny, S., Ruskin, H.J., Saeys, Y., Shamir, R., Sîrbu, A., Song, M., Soranzo, N., Statnikov, A., Vega, N., Vera-Licona, P., Vert, J.-P., Visconti, A., Wang, H., Wehenkel, L., Windhager, L., Zhang, Y., Zimmer, R., Kellis, M., Collins, J.J., Stolovitzky, G.: Wisdom of crowds for robust gene network inference. Nature Methods **9**(8), 796–804 (2012)

[28] Efron, B.: Large-Scale Inference. Empirical Bayes Methods for Estimation, Testing, and Prediction. Cambridge University Press, Cambridge, UK (2010)

[29] Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., Trombetta, J.J., Weitz, D.A., Sanes, J.R., Shalek, A.K., Regev, A., McCarroll, S.A.: Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell **161**(5), 1202–1214 (2015)

[30] Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., Kirschner, M.W.: Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell **161**(5), 1187–1201 (2015)

[31] Grün, D., van Oudenaarden, A.: Design and Analysis of Single-Cell Sequencing Experiments. Cell **163**(4), 799–810 (2015)

[32] Stegle, O., Teichmann, S.A., Marioni, J.C.: Computational and analytical challenges in single-cell transcriptomics. Nature Reviews Genetics **16**(3), 133–145 (2015)

[33] Bacher, R., Kendziorski, C.: Design and computational analysis of single-cell RNA-sequencing experiments. Genome Biology **17**(1), 1–14 (2016)

[34] Liu, S., Trapnell, C.: Single-cell transcriptome sequencing: recent advances and remaining challenges. F1000Research, 1–9 (2016)

[35] Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., Stegle, O.: Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nature Biotechnology **33**(2), 155–160 (2015)

[36] Pina, C., Teles, J., Fugazza, C., May, G., Wang, D., Guo, Y., Soneji, S., Brown, J., Edén, P., Ohlsson, M., Peterson, C., Enver, T.: Single-Cell Network Analysis Identifies DDIT3 as a Nodal Lineage Regulator in Hematopoiesis. Cell Reports **11**(10), 1503–1510 (2015). doi:10.1016/j.celrep.2015.05.016

[37] Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A.J., Tanaka, Y., Wilkinson, A.C., Buettner, F., Macaulay, I.C., Jawaid, W., Diamanti, E., Nishikawa, S.-I., Piterman, N., Kouskoff, V., Theis, F.J., Fisher, J., Göttgens, B.: Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nature biotechnology **33**(3), 269–276 (2015). doi:10.1038/nbt.3154

[38] Bendall, S.C., Davis, K.L., Amir, E.-a.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., Pe'er, D.: Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development. Cell **157**(3), 714–725 (2014)

[39] Kharchenko, P.V., Silberstein, L., Scadden, D.T.: Bayesian approach to single-cell differential expression analysis. Nature methods **11**(7), 740–742 (2014). doi:10.1038/nmeth.2967

[40] Ocone, A., Haghverdi, L., Mueller, N.S., Theis, F.J.: Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. Bioinformatics **31**(12), 89–96 (2015). doi:10.1093/bioinformatics/btv257

[41] Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, Hoboken, New Jersey (2012)

[42] Mc Mahon, S.S., Sim, A., Johnson, R., Liepe, J., Stumpf, M.P.H.: Information theory and signal transduction systems: from molecular information processing to network inference. Seminars in Cell & Developmental Biology **35**, 98–108 (2014). doi:10.1016/j.semcdb.2014.06.011

[43] Mc Mahon, S.S., Lenive, O., Filippi, S., Stumpf, M.P.H.: Information processing by simple molecular motifs and susceptibility to noise. Journal of the Royal Society Interface **12**(110), 20150597 (2015). doi:10.1098/rsif.2015.0597

[44] Uda, S., Saito, T.H., Kudo, T., Kokaji, T., Tsuchiya, T., Kubota, H., Komori, Y., Ozaki, Y.-i., Kuroda, S.: Robustness and Compensation of Information Transmission of Signaling Pathways. Science **341**(6145), 558–561 (2013). doi:10.1126/science.1234511

[45] Kraskov, A., Stögbauer, H., Grassberger, P.: Estimating mutual information. Phys. Rev. E **69**, 066138 (2004). doi:10.1103/PhysRevE.69.066138

[46] Steuer, R., Kurths, J., Daub, C.O., Weise, J., Selbig, J.: The mutual information: Detecting and evaluating dependencies between variables. Bioinformatics **18**(suppl 2), 231–240 (2002)

[47] Timme, N., Alford, W., Flecker, B., Beggs, J.M.: Synergy, redundancy, and multivariate information measures: an experimentalist's perspective. Journal of Computational Neuroscience **36**(2), 119–140 (2014). doi:10.1007/s10827-013-0458-4

[48] Williams, P.L., Beer, R.D.: Nonnegative Decomposition of Multivariate Information. arXiv.org (2010). 1004.2515v1

[49] Villaverde, A.F., Ross, J., Morán, F., Banga, J.R.: MIDER: Network Inference with Mutual Information Distance and Entropy Reduction. PLoS ONE **9**(5), 96732–15 (2014)

[50] Watkinson, J., Liang, K.-C., Wang, X., Zheng, T., Anastassiou, D.: Inference of regulatory gene interactions from expression data using three-way mutual information. Annals of the New York Academy of Sciences **1158**(1), 302–313 (2009). doi:10.1111/j.1749-6632.2008.03757.x

[51] Penfold, C.A., Shifaz, A., Brown, P.E., Nicholson, A., Wild, D.L.: CSI: a nonparametric Bayesian approach to network inference from multiple perturbed time series gene expression data. Statistical Applications in Genetics and Molecular Biology **14**(3), 307–310 (2015). doi:10.1515/sagmb-2014-0082

[52] Hill, S.M., Lu, Y., Molina, J., Heiser, L.M., Spellman, P.T., Speed, T.P., Gray, J.W., Mills, G.B., Mukherjee, S.: Bayesian inference of signaling network topology in a cancer cell line. **28**(21), 2804–2810 (2012). doi:10.1093/bioinformatics/bts514

[53] Lèbre, S., Becq, J., Devaux, F., Stumpf, M.P.H., Stumpf, M.P.H., Lelandais, G.: Statistical inference of the time-varying structure of gene-regulation networks. BMC Systems Biology **4**(1), 130 (2010). doi:10.1186/1752-0509-4-130

[54] Beal, M.J., Falciani, F., Ghahramani, Z., Rangel, C., Wild, D.L.: A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. Bioinformatics **21**(3), 349–356 (2005). doi:10.1093/bioinformatics/bti014

[55] Schäfer, J., Strimmer, K.: An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics **21**(6), 754–764 (2005). doi:10.1093/bioinformatics/bti062

[56] Vinciotti, V., Augugliaro, L., Abbruzzo, A., Wit, E.C.: Model selection for factorial Gaussian graphical models with an application to dynamic regulatory networks. Statistical Applications in Genetics and Molecular Biology **15**(3), 193–212 (2016). doi:10.1515/sagmb-2014-0075

[57] Butte, A.J., Tamayo, P., Slonim, D., Golub, T.R., Kohane, I.S.: Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. Proceedings of the National Academy of Sciences **97**(22), 12182–12186 (2000). doi:10.1073/pnas.220392197

[58] Margolin, A.A., Wang, K., Lim, W.K., Kustagi, M., Nemenman, I., Califano, A.: Reverse engineering cellular networks. Nature Protocols **1**(2), 662–671 (2006). doi:10.1038/nprot.2006.106

[59] Faith, J.J., Hayete, B., Thaden, J.T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J., Gardner, T.S.: Large-Scale Mapping and Validation of Escherichia coli Transcriptional Regulation from a Compendium of Expression Profiles. PLoS Biology **5**(1), 8–13 (2007)

[60] Meyer, P.E., Kontos, K., Lafitte, F., Bontempi, G.: Information-Theoretic Inference of Large Transcriptional Regulatory Networks. EURASIP Journal on Bioinformatics and Systems Biology **2007**(1), 1–9 (2007)

[61] Simoes, R.d.M., Emmert-Streib, F.: Influence of Statistical Estimators of Mutual Information and Data Heterogeneity on the Inference of Gene Regulatory Networks. PLoS One **6**(12) (2011). doi:10.1371/journal.pone.0029279

[62] Zhang, Z., Zheng, L.: A mutual information estimator with exponentially decaying bias. Statistical Applications in Genetics and Molecular Biology **14**(3), 243–252 (2015). doi:10.1515/sagmb-2014-0047

[63] Pierson, E., Yau, C.: ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. Genome Biology **16**(1), 241 (2015). doi:10.1186/s13059-015-0805-z

[64] Schaffter, T., Marbach, D., Floreano, D.: GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. Bioinformatics **27**(16), 2263–2270 (2011)

[65] Meyer, P.E., Lafitte, F., Bontempi, G.: minet: A R/Bioconductor Package for Inferring Large Transcriptional Networks Using Mutual Information. BMC Bioinformatics **9**(1), 461–10 (2008)

[66] Murphy, K.P.: Machine Learning. A Probabilistic Perspective. MIT Press, Cambridge, MA (2012)

[67] Psaila, B., Barkas, N., Iskander, D., Roy, A., Anderson, S., Ashley, N., Caputo, V.S., Lichtenberg, J., Loaiza, S., Bodine, D.M., Karadimitris, A., Mead, A.J., Roberts, I.: Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. Genome Biology **17**(1), 1 (2016). doi:10.1186/s13059-016-0939-7

[68] Wilson, N.K., Kent, D.G., Buettner, F., Shehata, M., Macaulay, I.C., Calero-Nieto, F.J., Sánchez Castillo, M., Oedekoven, C.A., Diamanti, E., Schulte, R., Ponting, C.P., Voet, T., Caldas, C., Stingl, J., Green, A.R., Theis, F.J., Gottgens, B.: Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. Cell Stem Cell **16**(6), 712–724 (2015). doi:10.1016/j.stem.2015.04.004

[69] Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K.L., Ema, H., Nakauchi, H.: Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell **154**(5), 1112–1126 (2013). doi:10.1016/j.cell.2013.08.007

[70] Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A.J., Tanaka, Y., Wilkinson, A.C., Buettner, F., Macaulay, I.C., Jawaid, W., Diamanti, E., Nishikawa, S.-I., Piterman, N., Kouskoff, V., Theis, F.J., Fisher, J., Göttgens, B.: Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nature Biotechnology **33**(3), 269–276 (2015)

[71] Guo, G., Huss, M., Tong, G.Q., Wang, C., Sun, L.L., Clarke, N.D., Robson, P.: Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst. Developmental Cell **18**(4), 675–685 (2010)

[72] Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B.: Julia: A Fresh Approach to Numerical Computing. arXiv.org (2014). 1411.1607v2

[73] Vanderplas, J.T., Connolly, A.J., Ivezi , Ž., Gray, A.: Introduction to astroml: Machine learning for astrophysics. In: Conference on Intelligent Data Understanding (CIDU), pp. 47–54 (2012). doi:10.1109/CIDU.2012.6382200

[74] Scargle, J.D., Norris, J.P., Jackson, B., Chiang, J.: Studies in Astronomical Time Series Analysis. VI. Bayesian Block Representations. Astrophysical Journal **764**, 167 (2013). doi:10.1088/0004-637X/764/2/167. 1207.5578

[75] Kinney, J., Atwal, G.S.: Equitability, mutual information, and the maximal information coefficient. Proceedings of the National Academy of Sciences of the United States of America **111**(9), 3354–3359 (2014). doi:10.1073/pnas.1309933111

[76] Dietterich, T.G.: Ensemble Methods in Machine Learning. In: Multiple Classifier Systems, pp. 1–15. Springer, Berlin, Heidelberg (2000)

[77] Aijö, T., Bonneau, R.: Biophysically motivated regulatory network inference: progress and prospects. bioRxiv (2016). doi:10.1101/051847

[78] Mukherjee, S., Speed, T.P.: Network inference using informative priors. Proceedings of the National Academy of Sciences of the United States of America **105**(38), 14313–14318 (2008). doi:10.1073/pnas.0802272105

[79] Studham, M.E., Tjärnberg, A., Nordling, T.E.M., Nelander, S., Sonnhammer, E.L.L.: Functional association networks as priors for gene regulatory network inference. Bioinformatics **30**(12), 130–8 (2014)

[80] Haghverdi, L., ttner, M.B.u., Wolf, F.A., Buettner, F., Theis, F.J.: Diffusion pseudotime robustly reconstructs lineage branching. Nature Methods, 1–6 (2016)

[81] Reid, J.E., Wernisch, L.: Pseudotime estimation: deconfounding single cell time series. Bioinformatics **32**(19), 2973–2980 (2016)

[82] Setty, M., Tadmor, M.D., Reich-Zeliger, S., Angel, O., Salame, T.M., Kathail, P., Choi, K.,

Bendall, S., Friedman, N., Pe'er, D.: Wishbone identifies bifurcating developmental trajectories from single-cell data. Nature Publishing Group **34**(6), 637–14 (2016)

[83] Zoppoli, P., Morganella, S., Ceccarelli, M.: TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. BMC Bioinformatics **11**(1), 154 (2010)

[84] Opgen-Rhein, R., Strimmer, K.: From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. BMC Systems Biology **1**(1), 37 (2007). doi:10.1186/1752-0509-1-37

[85] Vanlier, J., Tiemann, C.A., Hilbers, P.A.J., van Riel, N.A.W.: A Bayesian approach to targeted experiment design. Bioinformatics **28**(8), 1136–1142 (2012)

[86] Liepe, J., Filippi, S., Komorowski, M., Stumpf, M.P.H.: Maximizing the information content of experiments in systems biology. PLoS Computational Biology **9**(1), 1002888 (2013). doi:10.1371/journal.pcbi.1002888

[87] Sunnåker, M., Zamora-Sillero, E., Dechant, R., Ludwig, C., Busetto, A.G., Wagner, A., Stelling, J.: Automatic Generation of Predictive Dynamic Models Reveals Nuclear Phosphorylation as the Key Msn2 Control Mechanism. Science Signaling **6**(277), 41–41 (2013)

[88] Silk, D., Kirk, P.D.W., Barnes, C.P., Toni, T., Stumpf, M.P.H., Stumpf, M.P.H.: Model Selection in Systems Biology Depends on Experimental Design. PLoS Computational Biology **10**(6), 1003650 (2014). doi:10.1371/journal.pcbi.1003650

[89] Deweese, M.R., Meister, M.: How to measure the information gained from one symbol. Network: Computation in Neural Systems **10**(4), 325–340 (1999)

[90] Timme, N., Alford, W., Flecker, B., Beggs, J.M.: Synergy, redundancy, and multivariate information measures: an experimentalist's perspective. Journal of Computational Neuroscience **36**(2), 119–140 (2013)

[91] Mosteller, F., Tukey, J.W.: Data Analysis and Regression: A Second Course in Statistics, p. 49. Addison-Wesley, Reading, MA (1977)

[92] Hausser, J., Strimmer, K.: Entropy Inference and the James-Stein Estimator, with Application to Nonlinear Gene Association Networks. The Journal of Machine Learning Research **10**, 1469–1484 (2009)

[93] Paninski, L.: Estimation of entropy and mutual information. Neural Computation **15**, 1191–1253 (2003)

[94] Agresti, A., Hitchcock, D.B.: Bayesian inference for categorical data analysis. Statistical Methods and Applications **14**(3), 297–330 (2005)

[95] Gillespie, D.: Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry **81**, 2340–2361 (1977)

[96] Thorne, T.W., Stumpf, M.P.H.: Graph spectral analysis of protein interaction network evolution. Journal of the Royal Society, Interface **9**(75), 2653–2666 (2012). doi:10.1098/rsif.2012.0220

[97] Thorne, T.W., Stumpf, M.P.H.: Generating confidence intervals on biological networks. BMC Bioinformatics **8**(1), 467 (2007). doi:10.1186/1471-2105-8-467