SUPPLEMENT

A Proof of Theorem 1

Throughout the proof, we denote the covariance between variables $\tilde{X}_i$ and $\tilde{X}_j$ as $\tilde{\Sigma}_{ij}$ for $i, j \in \{1, 2, 3, 4\}$. First we prove the direction ' $\Longleftarrow$ ':

$$\begin{aligned}
&\begin{cases} \tilde{\rho}^u_{13|2} = 0 \iff \tilde{\Sigma}_{12}\tilde{\Sigma}_{23} - \tilde{\Sigma}_{13}(\tilde{\Sigma}_{22} - u) = 0 \\ \tilde{\rho}^u_{14|2} = 0 \iff \tilde{\Sigma}_{12}\tilde{\Sigma}_{24} - \tilde{\Sigma}_{14}(\tilde{\Sigma}_{22} - u) = 0 \\ \tilde{\rho}^u_{34|2} = 0 \iff \tilde{\Sigma}_{32}\tilde{\Sigma}_{24} - \tilde{\Sigma}_{34}(\tilde{\Sigma}_{22} - u) = 0 \end{cases} \\
&\iff (\tilde{\Sigma}_{22} - u) = \frac{\tilde{\Sigma}_{12}\tilde{\Sigma}_{23}}{\tilde{\Sigma}_{13}} = \frac{\tilde{\Sigma}_{12}\tilde{\Sigma}_{24}}{\tilde{\Sigma}_{14}} = \frac{\tilde{\Sigma}_{23}\tilde{\Sigma}_{24}}{\tilde{\Sigma}_{34}} \\
&\iff \frac{\tilde{\Sigma}_{23}}{\tilde{\Sigma}_{13}} = \frac{\tilde{\Sigma}_{24}}{\tilde{\Sigma}_{14}}, \qquad \frac{\tilde{\Sigma}_{12}}{\tilde{\Sigma}_{14}} = \frac{\tilde{\Sigma}_{23}}{\tilde{\Sigma}_{34}} \\
&\iff \tilde{\Sigma}_{12}\tilde{\Sigma}_{34} = \tilde{\Sigma}_{13}\tilde{\Sigma}_{24} = \tilde{\Sigma}_{14}\tilde{\Sigma}_{23}.
\end{aligned}$$

The result follows by applying Lemma 1 in the main paper, and observing that these are the only structures for a random measurement model where all d-separations hold.

Now we prove ' $\Longrightarrow$ '. Since the true underlying causal graph is as in Figure 1 in the main paper, we have that there is an α such that $\tilde{X}_2 = \alpha L + E_2 + M_2$ for some $\alpha \neq 0$, where E_2 is an independent noise variable with variance τ for X_2 and M_2 is an independent random measurement error for $\tilde{X}_2$. We let Σ_{ij} denote the covariance between variables X_i and X_j for $i, j \in \{1, 2, 3, 4\}$. Covariances between L and variables (X_1, X_2, X_3, X_4) are denoted as Σ_{Li} and Σ_{iL} for $i \in \{1, 2, 3, 4\}$. Hence

$$\text{Cov}(\tilde{X}_1, \tilde{X}_2, \tilde{X}_3) = \begin{pmatrix} \Sigma_{11} + m_1 & \alpha\sigma_{L1} & \Sigma_{13} \\ \alpha\Sigma_{L1} & \alpha^2\Sigma_{LL} + \tau + m_2 & \alpha\Sigma_{L3} \\ \Sigma_{13} & \alpha\Sigma_{L3} & \Sigma_{33} + m_3, \end{pmatrix}.$$

where m_1, m_2, m_3 are the variances of M_1, M_2, M_3 respectively and Σ_{LL} denotes the variance of the latent variable L. From this we obtain the relation for the adjusted partial correlation:

$$\tilde{\rho}^u_{13|2} = 0 \qquad \iff \qquad \alpha^2(\Sigma_{L1}\Sigma_{L3} - \Sigma_{13}\Sigma_{LL}) - \Sigma_{13}(\tau + m_2 - u) = 0.$$

Since L d-separates X_1 and X_3, the partial correlation $\tilde{\rho}_{13|L} = 0$ by the Markov assumption. Therefore

$$\Sigma_{L1}\Sigma_{L3} - \Sigma_{13}\Sigma_{LL} = 0.$$

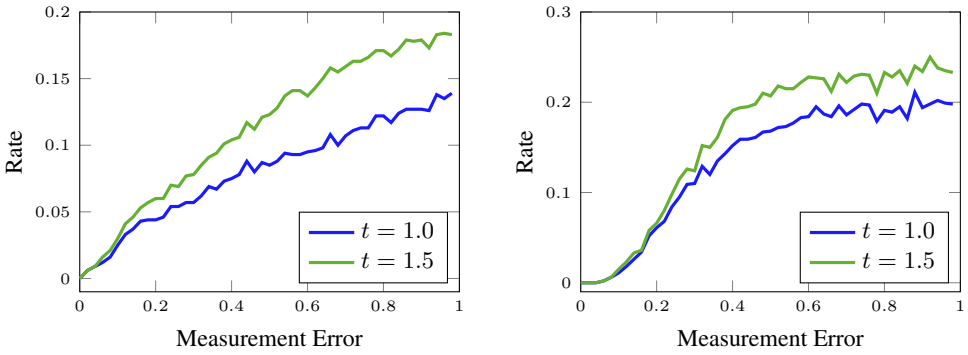
Because $\Sigma_{13} \neq 0$ by assumption, we find that $\tilde{\rho}^u_{13|2} = 0$ if and only if $u = \tau + m_2$. Via a similar argument we can show that for this u we also have that $\tilde{\rho}^u_{14|2} = 0$ and $\tilde{\rho}^u_{34|2} = 0$.

B Data simulations

In this section we give some additional details about the simulations that we used for the experiments in Section 7 of the main paper.

We obtained the results in Figure 7a of the main paper, by generating random DAGs for 6 variables with a connection probability of 0.7, parameters chosen uniformly at random from the interval $[-1.0, 1.0]$, and error variances chosen uniformly from the interval $[0.5, 1.0]$. Three out of the 6 variables were observed variables, and the remaining three were latent. We then used rejection sampling to select models for which the observed variables (X_1, X_2, X_3) satisfied: the λ-strong faithfulness assumption for $\lambda = 0.1$, $X_1 \not\perp\!\!\!\perp X_2$, $X_2 \not\perp\!\!\!\perp X_3$, and $X_1 \not\perp\!\!\!\perp X_3 \,|\, X_2$. For each model, we generated 10000 datapoints and added measurement error with varying variances.

For the experiment in Figure 7b in the main paper, we generated models for three variables (X_1, X_2, X_3) that satisfy: $X_1 \not\perp\!\!\!\perp X_2$, $X_2 \not\perp\!\!\!\perp X_3$, and $X_1 \perp\!\!\!\perp X_3 \,|\, X_2$. To do this, we considered all causal structures for (X_1, X_2, X_3) that



(a) Conditional dependences

(b) Conditional independences.

Figure 1: Rate of conditional dependences and independences that were not detected using a measurement error correction in the experiment in Figures 7a and 7b in the main paper.

satisfied these conditions. For each causal structure, parameters were chosen uniformly at random from the interval $[-1.0, 1.0]$ and error variances were chosen uniformly at random from the interval $[0.5, 1.0]$. We then used rejection sampling to select model that satisfied the λ-strong faithfulness for $\lambda = 0.1$. For each model, we generated 10000 datapoints and added measurement error with varying variances. We made sure that an equal amount of models was selected from each causal structure.

For the experiments in Figures 7a and 7b in the main paper, the measurement-error corrected method also gives the output 'unknown'. The rate of conditional dependences and independences that could not be detected in the experiment are shown in Figures 1a and 1b respectively.

To obtain the results in Figure 7c, we generated causal structures for triples of variables (X_1, X_2, X_3). To generate data for triples that did not have the structure of an LCD triple, we used the simulations for Figure 7a, but only selected causal structures where X_1 was not caused by X_2 and X_3. Similarly, to generate data for triples that did have the structure of LCD triples, we used the simulations for Figure 7b, but only selected causal structures where X_1 can be treated as an intervention variable. We added measurement error with a fixed variance of 0.8.

Finally, in order to obtain the results in Figure 7d, we generated random DAGs with 15 nodes and a connection probability of 0.15. In this case edge weights were chosen uniformly at random from the interval $[0.8, 1.2]$ and error variance were chosen uniformly at random from the interval $[0.5, 1.0]$. After generating 10000 datapoints for each model, we added measurement error with a fixed variance of 0.8.

C Protein Signaling Data

The raw data was pre-processed by transforming each datapoint x by

$$\hat{x} = \text{arcsinh}(x/5).$$

As a preprocessing step we filtered out cells that are in the M cell cycle phase according to the gating procedure described in [Behbehani et al., 2012]. We motivate this filtering step by several reasons. First, cells in the M phase form a distinct cluster and strongly violate our assumption of a linear-Gaussian model since they represent a separate cluster. Second, in the M phase the cells already have doubled nuclei and some other organelle, so we cannot safely assume that the causal mechanisms of signalling are the same anymore. Therefore, the removal of these cells should be seen as removal of a contaminating population. Practically this came down to selecting only single cell measurements for which the abundance of the phosphorylated protein pHH3 was smaller than 3.0.

We then only included data in our (conditional) independence tests, when all measurements that were needed to conduct the test exceeded a lower threshold of 0.5, to account for the detection limit in mass cytometry.

For the analysis in the main paper, we only considered the proteins that were over-expressed and whose phosphorylated abundances were also measured. We considered the measurements 5 minutes after stimulation, because at this time-

point the signaling responses were generally strong, see also Figure 3 in [Lun et al., 2017]. For our analysis we only used the first replica of the experiment, which had the most measurements for each condition.

We analyzed a subset of the available proteins, based on the recommendations in Lun et al. [2017], and excluded proteins from the cause variables when spill-over effects were reported under the condition that they were over-expressed, see also Table 1. We also excluded pS6 because over-expressing it induced no strong signaling responses. Finally, we also discarded SHP2. Although the condition where SHP2 was over-expressed was not affected by spill-over effects, the measured *phosphorylated* abundances of pSHP2 were affected by spill-over affects under multiple conditions.

Table 1: Proteins that are both over-expressed in one of the conditions and whose phosphorylated abundance is measured under all conditions, with an indication whether spillover effects are present.

Over-expressed protein	Measured protein	Spill-over
JNK1	pJNK	no
MKK6	pMKK3/6	no
PDPK1	pPDPK1	yes
P38	pP38	no
AKT1	pAKT	no
ERK2	pERK	no
SHP2	pSHP2	no
GSK3B	pGSK3B	yes
S6	pS6	no*
P90RSK	pP90RSK	yes
MEK1	pMEK1/2	no
P70S6K	pS6K	no

References

G.K. Behbehani, S.C. Bendall, M.R. Clutter, W.J. Fantl, and G.P. Nolan. Single-cell mass cytometry adapted to measurements of the cell cycle. *Cytometry A*, 81(7):552–566, 2012.

X. Lun, V.R.T. Zanotelli, J.D. Wade, D. Schapiro, M. Tognetti, N. Dobberstein, and B. Bodenmiller. Influence of node abundance on signaling network state and dynamics analyzed by mass cytometry. *Nature Biotechnology*, 35(2): 164–172, 2017.