Electronic Supplementary Material (ESI) for Molecular Omics.

Comparison between instrumental variable and mediation-based methods for reconstructing causal gene networks in yeast

(Supplementary Information)

Adriaan-Alexander Ludl and Tom Michoel*

December 9, 2020

Computational Biology Unit, Department of Informatics, University of Bergen, PO Box 7803, 5020 Bergen, Norway

* Corresponding author, email: tom.michoel@uib.no

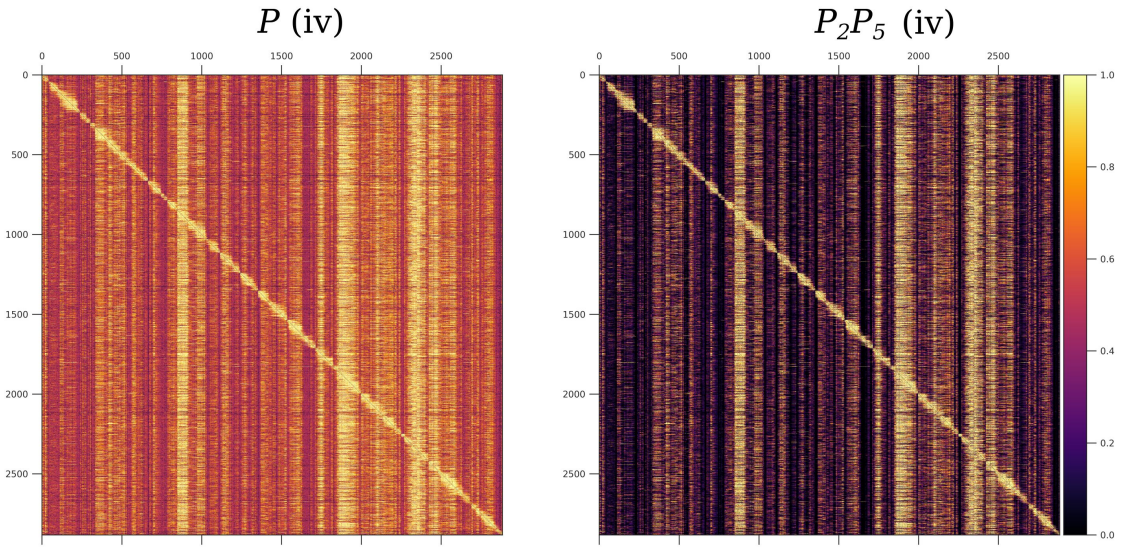

Figure S1: **Matrices of predicted gene interactions.** These square matrices represent the interactions between 2884 genes with causal anchors (eQTLs), probability values are color coded. Vertical bands correspond to hotspots. **Left**: The instrumental variable test with partial pleiotropy P. **Right**: The instrumental variable test with perfect pleiotropy P_2P_5. The genes are ordered according to the position of their causal anchor in the full yeast genome. Definitions of the tests are given in the Methods section. This figure complements Fig. 2 in the main text.

Method	p_{th}	FDR
P_2P_3	0.8175	0.09953
P_2	0.825	0.04974
P_2P_5	0.8375	0.04994
P	0.8575	0.04982
P_0	0.86	0.00986

Table S1: **FDR thresholds.** The thresholds (p_{th}) reported here were used to select significant interactions for the methods shown in figure 5 and S3.

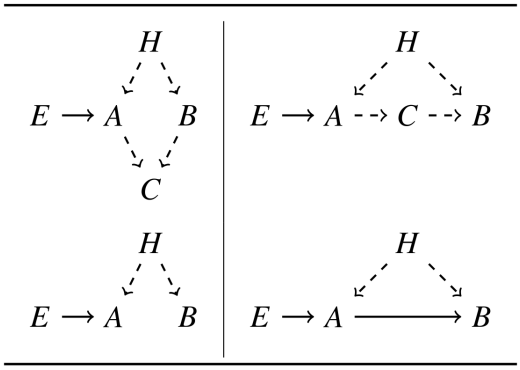

Figure S2: **Comparison of causal models.** These graphical models illustrate possible interactions between genes. The hidden confounder H would be a common parent of A and B as shown also in Fig. 1[B]. The upper row shows a possible ground truth, the lower row shows how this scenario would be classified by *Findr*. **Left**: A common child C of A and B does not affect the predictions of *Findr* regardless of the presence of a common parent H. **Right**: An intermediary node C could mediate the interaction between A and B, regardless of a common parent H.

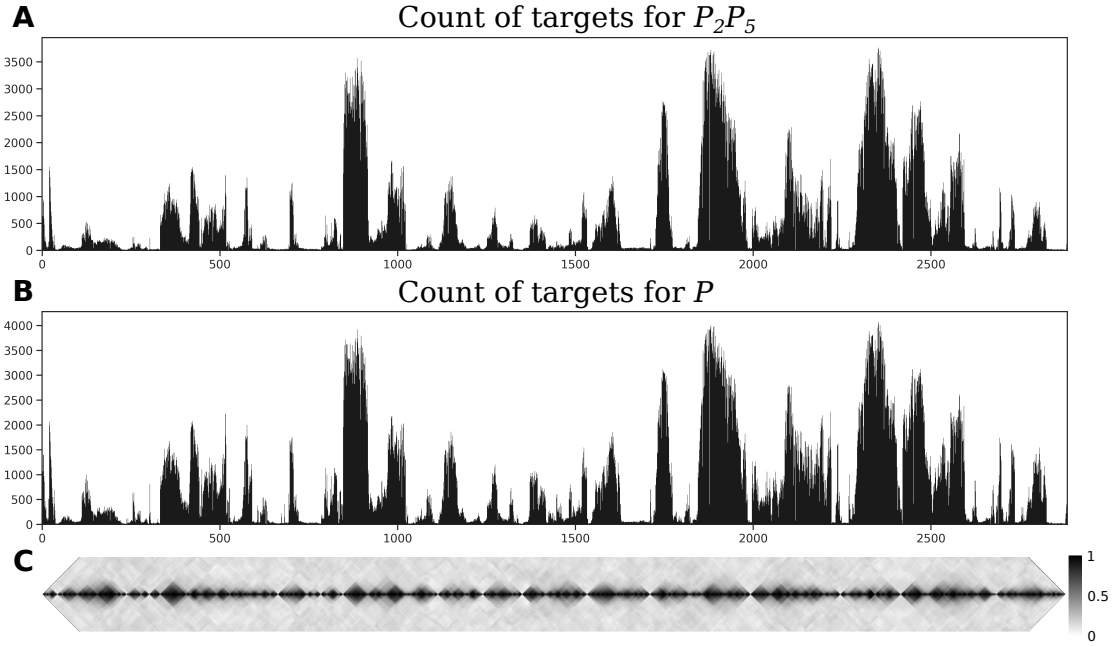

Figure S3: **Hotspots and genotype covariance.** **A** and **B** show the counts of significant interactions for two inference methods. Genes are ordered along the horizontal axis according to the position of their causal anchor in the full yeast genome. **A**: instrumental variables with perfect pleiotropy (P_2P_5) at FDR 5%. **B**: instrumental variables with partial pleiotropy (P) at FDR 5%. The thresholds used are reported in Tab. S1. **C**: The diagonal of the genotype covariance matrix for the 2884 eQTLs.

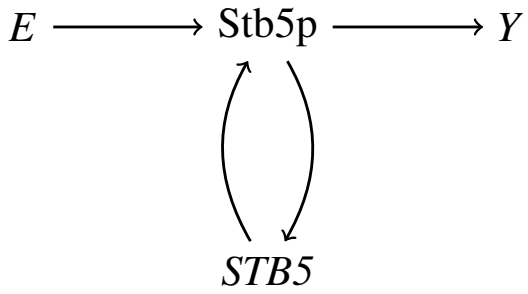

Figure S4: **Hypothetical model for the *STB5* hotspot.** Stb5p protein level is determined by *STB5* transcription level and the genotype of one or more protein-altering variants E, and in turn affects *STB5* transcription level by an auto-regulatory loop. Expression of *STB5* target genes Y is determined by *STB5* transcription only through Stb5p level. Even in the absence of any hidden confounders, *STB5* transcription does not block the path between E and Y, and unless the correlation between *STB5* transcription and Stb5p level is perfect (no biological or experiment noise), conditioning on *STB5* transcription level will not remove the statistical association between E and Y. This model is consistent with the observed lack of allele-specific expression of *STB5* [20], and with the fact that the instrumental variable method P_2 correctly identifies target genes with Stb5p binding sites, but the mediation-based method P_2P_3 does not.