

Applying causal discovery to single-cell analyses using CausalCell

Yujian Wen^{1†}, Jielong Huang^{1†}, Shuhui Guo¹, Yehezqel Elyahu², Alon Monsonego², Hai Zhang^{3*}, Yanqing Ding^{4*}, Hao Zhu^{1,5,6*}

¹Bioinformatics Section, School of Basic Medical Sciences, Southern Medical University, Guangzhou, China; ²The Shraga Segal Department of Microbiology, Immunology and Genetics, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel; ³Network Center, Southern Medical University, Guangzhou, China; ⁴Department of Pathology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, China; ⁵Guangdong-Hong Kong-Macao Greater Bay Area Center for Brain Science and Brain-Inspired Intelligence, Southern Medical University, Guangzhou, China; ⁶Guangdong Provincial Key Lab of Single Cell Technology and Application, Southern Medical University, Guangzhou, China

Abstract Correlation between objects is prone to occur coincidentally, and exploring correlation or association in most situations does not answer scientific questions rich in causality. Causal discovery (also called causal inference) infers causal interactions between objects from observational data. Reported causal discovery methods and single-cell datasets make applying causal discovery to single cells a promising direction. However, evaluating and choosing causal discovery methods and developing and performing proper workflow remain challenges. We report the workflow and platform CausalCell (<http://www.gaemons.net/causalcell/causalDiscovery/>) for performing single-cell causal discovery. The workflow/platform is developed upon benchmarking four kinds of causal discovery methods and is examined by analyzing multiple single-cell RNA-sequencing (scRNA-seq) datasets. Our results suggest that different situations need different methods and the constraint-based PC algorithm with kernel-based conditional independence tests work best in most situations. Related issues are discussed and tips for best practices are given. Inferred causal interactions in single cells provide valuable clues for investigating molecular interactions and gene regulations, identifying critical diagnostic and therapeutic targets, and designing experimental and clinical interventions.

***For correspondence:**
zhangh@smu.edu.cn (HZ);
dyqgz@126.com (YD);
zhuhao@smu.edu.cn (HZ)

[†]These authors contributed equally to this work

Competing interest: The authors declare that no competing interests exist.

Funding: See page 13

Received: 28 June 2022

Preprinted: 19 August 2022

Accepted: 01 May 2023

Published: 02 May 2023

Reviewing Editor: Babak Momeni, Boston College, United States

© Copyright Wen, Huang et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editor's evaluation

This manuscript presents an important tool for causal inference intended for the analysis of single cell datasets but possibly with broader applications. It compares several algorithms and incorporates a number of them in the platform and offers convincing evidence of its usefulness. With the rapid expansion of large datasets, this tool is beneficial in offering several causal inference analysis options and expediting the interpretation of data.

Introduction

RNA-sequencing (RNA-seq) has been used to detect gene expression in a lump of cells for years. Many statistical methods have been developed to explore correlation/association between transcripts in RNA-seq data, including the 'weighted gene co-expression network analysis' that infers networks of correlated genes (Joehanes, 2018). Since a piece of tissue may contain many different cells and the

sample sizes of most RNA-seq data are <100, causal interactions in single cells, which to a great extent are emergent events (Bhalla and Iyengar, 1999), cannot be revealed by these statistical methods. Averaged gene expression in heterogeneous cells also makes causal interactions blurred or undetectable. Except for some annotated interactions in signaling pathways, most causal interactions in specific cells remain unknown (e.g. in developing cells undergoing rapid fate determination and in diseased cells expressing genes aberrantly).

Single-cell RNA-sequencing (scRNA-seq) has been widely used to detect gene expression in single cells, providing large samples for analyzing cell-specific gene expression and regulation. On statistical data analysis, it is argued that 'statistics alone cannot tell which is the cause and which is the effect' (Pearl and Mackenzie, 2019). Corresponding to this, causal discovery is a science that distinguishes between causes and effects and infers causal interactions from observational data. Many methods have been designed to infer causal interactions from observational data. For single-cell analysis, any method faces the three challenges – high-dimensional data, data with missing values, and inferring with incomplete model (with missing variables). The constraint-based methods are a class of causal discovery methods (Glymour et al., 2019; Yuan and Shou, 2022), and the PC algorithm is a classic constraint-based method. Testing conditional independence (CI, CI≠unconditional independence [UI]≠uncorrelation) between variables is at the heart of constraint-based methods. Many CI tests have been developed (Verbyla, 2018; Zhang and Peters, 2011), from the fast GaussCItest to the time-consuming kernel-based CI tests. GaussCItest is based upon partial correlations between variables. Kernel-based CI tests estimate the dependence between variables upon their observations without assuming any relationship between variables or distribution of data. These features of kernel-based CI tests enable relationships between any genes and molecules, not just transcription factors (TFs) and their targets, to be inferred. Thus, CI tests critically characterize constraint-based causal discovery and distinguish causal discovery from other network inferences, including 'regulatory network inference' (Nguyen et al., 2021; Pratapa et al., 2020), 'causal network inference' (Lu et al., 2021), 'network inference' (Deshpande et al., 2019), and 'gene network inference' (Marbach et al., 2012).

Kernel-based CI tests are highly time-consuming and thus infeasible for transcriptome-wide causal discovery. Recently other causal discovery methods are reported, especially continuous optimization-based methods (Bello et al., 2022a; Zheng et al., 2018). Thus, identifying the best methods and CI tests, developing reasonable workflows, developing measures for quality control, and making trade-offs between time consumption, network size, and network accuracy are important. This *Tools and Resources* article addresses the above issues by benchmarking multiple causal discovery methods and CI tests, applying causal discovery to multiple scRNA-seq datasets, developing a causal discovery workflow/platform (called CausalCell), and summarizing tips for best practices. Specifically, the workflow combines feature selection and causal discovery. The benchmarking includes 11 causal discovery methods, 10 CI tests, and 9 feature selection algorithms. In addition, measures for estimating and ensuring the reliability of causal discovery are developed. Our results indicate that when relationships between variables are free of missing variables and missing values, continuous optimization-based methods perform well. Otherwise, the PC algorithm with kernel-based CI tests can better tolerate incomplete models and missing values. Inferred relationships between gene products help researchers draw causal hypotheses and design experimental studies. The remaining sections describe the workflow/platform and data analysis examples, discuss specific issues, and present tips for best practices. The details of methods and algorithms, benchmarking results, and data analysis results are described in appendix files.

Materials and methods

Features of different algorithms

Causal discovery cannot be performed transcriptome-wide due to time consumption and the power of methods. A way to choose a subset of genes based on one or several genes of interest is feature selection. A feature selection algorithm combines a search technique and an evaluation measure and works upon one or several response variables (i.e. genes of interest). After obtaining a measure between the response variable(s) and each feature (i.e. variable, gene), a subset of features most related to the response variable(s) are extracted from the whole dataset. Using simulated data and real scRNA-seq data (Appendix 1—table 1), we benchmarked nine feature selection algorithms. The

Table 1. Performance of feature selection methods.

Algorithm	Category	Time consumption	Accuracy	Scalability	Advantage/disadvantage
RandomForest	Ensemble learning-based methods use many trees of a random forest to calculate the importance of features, then perform regression based on the response variable(s) to identify the most relevant features.	+	++	++	These algorithms are indeterministic (the same input may generate slightly different outputs). ExtraTrees and RandomForest perform better than XGBoost.
ExtraTrees		+	++	++	
XGBoost		++	+	+	
BAHSIC	The three are Hilbert-Schmidt independence criterion (HSIC)-based algorithms. HSIC is used as the measure of dependency between the response variable and features.	+	+++	+	BAHSIC and SHS are the best and second best.
SHS		+	+++	+	
HSIC Lasso		++	++	++	Inferior to BAHSIC and SHS.
Lasso	Lasso is a regression analysis method that performs both variable selection and regularization (which adds additional constraints or penalties to a regression model). Lasso, RidgeRegression, and ElasticNet are three regulation terms.	+++	+	+++	Inferior to BAHSIC and SHS. Accuracy is not high and scalability is poor.
RidgeRegression		+++	+	+++	
ElasticNet		+++	+	+++	

Time consumption is estimated upon simulated data (Appendix 2—figure 1). Accuracy is estimated upon simulated and real data (Appendix 2—figures 2–7). Scalability is estimated upon simulated data (Appendix 2—figure 2). Advantage/disadvantage is made upon accuracy together with algorithms’ other properties.

properties and advantages/disadvantages of these algorithms are summarized, with ‘+++’ and ‘+’ indicating the most and least recommended ones (Table 1; Appendix 2—figures 1–7).

Many causal discovery methods have been proposed. Constraint-based causal discovery identifies causal relationships between a set of variables in two steps: skeleton estimation (determining the skeleton of the causal network) and orientation (determining the direction of edges in the causal network). The PC algorithm is a classic and widely recognized algorithm (Glymour et al., 2019). Causal discovery using the PC algorithm is different in that PC can work with different CI tests to perform the first step. We combined the PC algorithm with 10 CI tests to form 10 constraint-based causal discovery algorithms. The properties and advantages/disadvantages of the 10 algorithms are summarized, with ‘+++’ and ‘+’ indicating the most and least recommended ones (Table 2; Appendix 3—figures 1 and 2). In addition to constraint-based methods, there are other kinds of methods, including score-based methods that assign a score function to each directed acyclic graph (DAG) and optimize the score via

Table 2. Performance of causal discovery methods.

Methods	CI tests	Category	Time consumption	Accuracy	Stability	Features
	GaussCitest	GaussCitest assumes all variables are multivariate Gaussian, which impairs GussCitest’s performance when data are complex.	+++	+	+++	Fast and inaccurate
	CMlknn	Conditional mutual information (CMI) is based on mutual information.	+++	++	+	Fast and inaccurate
	RCIT		++	++	++	
	RCoT	Two approximation methods of KCIT (the Kernel conditional independence test).	++	++	++	Fast and moderately accurate
	HSIC.clust		+	++	+	
	HSIC.gamma	Extra transformations make HSIC determine if X and Y are conditionally independent given a conditioning set. HSIC.gamma and HSIC.perm employ gamma test and permutation test to estimate a p-value.	+	+++	++	
	HSIC.perm		+	+++	+	Slow and accurate
	DCC.gamma	Distance covariance is an alternative to HSIC for measuring independence. DCC.gamma and DCC.perm employ gamma test and permutation test to estimate a p-value.	+	+++	++	
	DCC.perm		+	+++	+	Slow and accurate
PC GSP	GCM	The generalized covariance measure-based (also classified as regression-based).	+	++	+++	Slower than DCC.gamma
GES		Score-based causal.	++	++	++	Fast and moderately accurate
DAGMA-nonlinear		Continuous optimization-based.	+	++	+++	Performs well with complete models

Time consumption is estimated upon simulated data (Appendix 3—figure 1). Accuracy is estimated upon the lung cancer cell lines (Figure 2; Appendix 3—figure 2). Stability is estimated upon the relative structural Hamming distance (SHD, a standard distance to compare graphs by their adjacency matrix), which is used to measure the extent an algorithm produces the same results when running multiple times (Appendix 3—table 1). Advantage/disadvantage is made upon accuracy.

The screenshot displays the CausalCell web application interface, organized into three main steps:

- Step 1: Data Upload**
 - Choose the dataset of a cell type to analyze: Not selected yet
 - Or: Upload the dataset of a cell type to analyze (.csv or .tsv): H838_10x_norm.csv (Upload complete)
 - Choose the dataset of a control for the analysis: Not selected yet
 - Or: Upload the dataset of a control for the analysis (.csv or .tsv): Alveolar_Epithelial_10x_norm.csv (Upload complete)
 - What is the data format: log2-transformed, z-scores, Raw values
 - Select a TF list to show TF genes: Human TF list (Bahrami et al. BMC Res Notes 2015)
- Step 2: Feature Selection Parameters**
 - gene expression threshold: Slider from -0.5 to 0.5 (set at 0)
 - cellsInCase: Slider from 0.003 to 1 (set at 0.5)
 - varianceInCase: Slider from 0 to 1.618 (set at 1.618)
 - foldChange: Slider from 0 to 3 (set at 0.25)
 - TF: No, Yes
 - max candidate gene number: Slider from 0 to 2,288 (set at 2,288)
 - Select fold change computation of case-specific genes: Set FC of case-specific genes to 999, Assuming the value of case-specific genes' counterparts in control is 0.01
 - Select feature selection algorithms (one or multiple): Random Forest *, Extra Trees *, BAHSIC **, SHS *, HSIClasso, Lasso, Ridge, Elastic Net, XGBoost
 - Number of selected genes for each algorithm: Slider from 10 to 100 (set at 50)
 - Input one or multiple response variables (target genes): gene1, gene2
 - Run Feature Selection.
- Step 3: Causal Discovery Parameters**
 - Accept the just finished feature selection result: Yes, No
 - If No is chosen, input feature genes: [Empty text area]
 - Select a causal discovery framework: PC, GSP, GES, DAGMA_nonlinear
 - Select causal discovery conditional independence test(s) (one or multiple): gaussCTest, dcc.gamma ***, dcc.perm ***, hsc.gamma **, hsc.perm **, hsc.clust, RCIT *, RCoT, cmIKnn, GCM
 - Use spike-in data or not: Yes, No
 - Available spike-in data in this server: Not selected yet
 - Or upload a spike-in data (optional): Browse... No file selected
 - Set maximal size of the conditioning sets that are considered in the conditional independence tests: Slider from -1 to 10 (set at -1)
 - Set the alpha level for conditional independence test: Slider from 0 to 1 (set at 0.1)
 - Select the number of cells for causal discovery: Slider from 100 to 1,000 (set at 300)

Figure 1. The user interface of CausalCell. Multiple algorithms and functions are integrated and implemented to facilitate and compare feature selection and causal discovery.

greedy searches (*Chickering, 2003*), hybrid methods that combine score-based and constraint-based methods (*Solus et al., 2021*), and continuous optimization-based methods that convert the traditional combinatorial optimization problem into a continuous program (*Bello et al., 2022a; Zheng et al., 2018*). When benchmarking the four classes of methods, multiple simulated data, real scRNA-seq data, and signaling pathways were used to evaluate their performance (*Appendix 1—table 1*).

The results of benchmarking the 11 causal discovery methods and 10 CI tests show that when causal discovery is without the problems of incomplete models (i.e. ones that miss nodes or edges from the data-generating model) and missing values, nonlinear versions of continuous optimization-based methods (especially DAGMA-nonlinear) perform better than others (*Bello et al., 2022a*). When causal discovery is applied to a set of highly expressed or differentially expressed genes in an scRNA-seq dataset (which has both missing variables and missing values), the PC algorithm with kernel-based CI tests (especially DCC.gamma) performs well. Therefore, the CausalCell platform includes 4 causal discovery methods (PC, GES, GSP, and DAGMA-nonlinear) to suit different data, together with 10 CI tests and 9 feature selection algorithms.

Developing the workflow/platform for causal discovery

The CausalCell workflow/platform is implemented using the Docker technique and Shiny language and consists of feature selection, causal discovery, and several auxiliary functions (**Figure 1**). A parallel version of the PC algorithm is used to realize parallel multi-task causal discovery (**Le et al., 2019**). In addition, the platform also includes the GES, GSP, and DAGMA-nonlinear methods. PC and GSP can work with 10 CI tests. Annotations of functions and parameters and the detailed description of a causal discovery process are available online.

Data input and pre-processing

scRNA-seq and proteomics data generated by different protocols or methods (e.g. 10x Genomics, Smart-seq2, and flow cytometry) can be analyzed. CausalCell accepts log₂-transformed data and z-score data and can turn raw data into either of the two forms. A dataset (i.e. the 'case') can be analyzed with or without a control dataset (i.e. the 'control'). Researchers often identify and analyze special genes, such as highly expressed or differentially expressed genes. For each gene in a case and control, three attributes (the averaged expression value, percentage of expressed cells, and variance) are computed. Fold changes of gene expression are also computed (using the *FindMarkers* function in the *Seurat* package) if a control is uploaded. Genes can be ordered upon any attribute and filtered upon a combination of five conditions (i.e. expression value, percentage of expressed cells, variance, fold change, and being a TF or not). Since performing feature selection transcriptome-wide is unreliable due to too many genes, filtering genes before feature selection is necessary, and different filtering conditions generate different candidates for feature selection.

Batch effects may influence identifying differentially expressed genes. Since removing batch effects should be performed with raw data before integrating batches and there are varied batch effect removal methods (**Tran et al., 2020**), it should be performed by the user if necessary.

Feature selection

Feature selection selects a set of genes (i.e. features) from the candidate genes upon one or multiple genes of interest (i.e. response variables). As above-mentioned, candidate genes are extracted from the whole dataset upon specific conditions because performing feature selection transcriptome-wide is unreliable. Based on the accuracy, time consumption, and scalability of the nine feature selection algorithms (**Table 1**), BAHSIC is the most recommended algorithm. The joint use of two kinds of algorithms (e.g. Random Forest+BAHSIC) is also recommended to ensure reliability. Feature genes are usually 50–70, but the number also depends on the causal discovery algorithms. Genes can be manually added to or removed from the result of feature selection (i.e., the feature gene list) to address a biological question specifically. The input for causal discovery can also be manually selected without performing feature selection; for example, the user can examine a specific Gene ontology (GO) term.

Causal discovery

The PC and GSP algorithms can work with the 10 CI tests to provide varied options for causal discovery. In the inferred causal networks, direction of edges is determined by the meek rules (**Meek, 1997**), and each edge has a sign indicating activation or repression and a thickness indicating CI test's statistical significance. The sign of an edge from A to B is determined by computing a Pearson correlation coefficient between A and B, which is 'repression' if the coefficient is negative or 'activation' if the coefficient is positive. In most situations, 'A activating B' and 'A repressing B' correspond to up-regulated A in the case dataset, with up- and down-regulated B in the case dataset compared with in the control dataset.

There are two ways to construct a consensus network that is statistically more reliable. One way is to run multiple algorithms (i.e. multiple CI tests) and take the intersection of some or all inferred networks as the consensus network (**Figure 2**). The other is to run an algorithm multiple times and take the intersection of all inferred networks as the consensus network (**Figure 3**).

If a scRNA-seq dataset is large, a subset of cells should be sampled to avoid excessive time consumption. We suggest that 300 and 600 cells are suitable for reliable inference if the input is Smart-seq2 and 10x Genomics data, respectively, the input contains about 50 genes, and genes are expressed in >50% cells. Here, reliable inference means that key interactions (those with high CI test significance) are inferred (Appendices 3, 4). More cells are needed if the input genes are expressed in

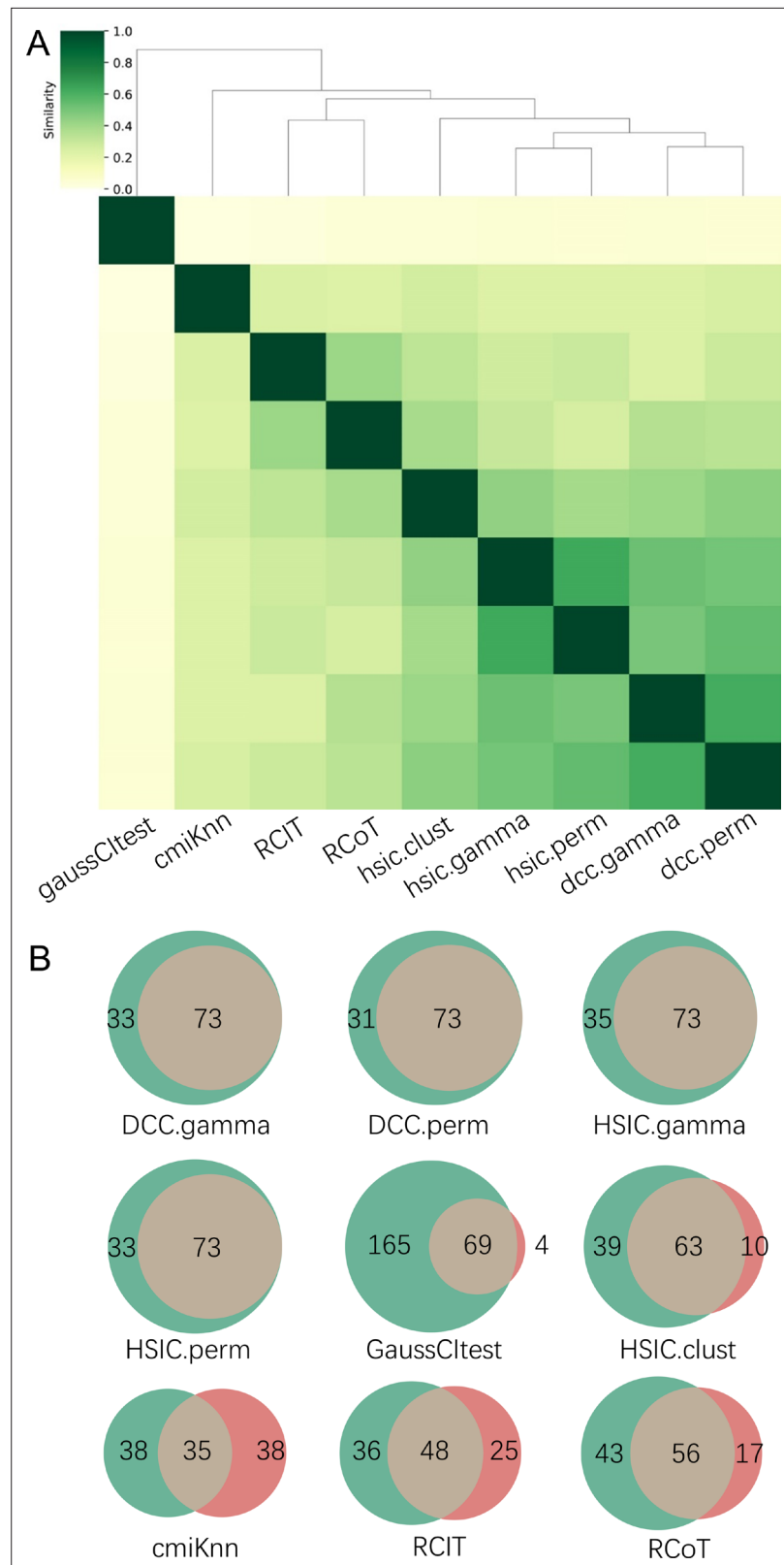


Figure 2. The accuracy of PC+nine CI tests was evaluated with four steps. First, nine causal networks were inferred using the nine CI tests. Second, pairwise structural Hamming distances (SHD) between these networks were computed, and the matrix of SHD values was transformed into a matrix of similarity values (using the equation $Similarity = \exp(-Distance/2\sigma^2)$, where $\sigma=5$). The networks of DCC.gamma, DCC.perm, HSIC.gamma, and HSIC.

Figure 2 continued on next page

Figure 2 continued

perm share the highest similarity. Third, a consensus network was built using the networks of the above four CI tests, which was assumed to be closer to the ground truth than the network inferred by any single algorithm. Fourth, each of the nine networks was compared with the consensus network. **(A)** The cluster map shows the similarity values (darker colors indicating higher similarity). **(B)** Shared and specific interactions in each algorithm's network and the consensus network. In each panel, the gray-, green-, and pink-circled areas and numbers indicate the overlapping interactions, interactions identified specifically by the algorithm, and interactions specifically in the consensus network. There are 73 overlapping interactions between DCC.gamma's network and the consensus network, and 33 interactions were identified specifically by DCC.gamma. Thus, the true positive rate (TPR) of DCC.gamma is $73 / (73 + 33) = 68.9\%$. The TPRs of DCC.perm, HSIC.gamma, HSIC.perm, GaussCtest, HSIC.clust, cmknn, RCIT, and RCoT are 70.2%, 67.6%, 68.9%, 29.5%, 61.8%, 47.9%, 57.1%, and 56.6%, indicating that the two distance covariance criteria (DCC) CI tests perform better than others.

fewer cells and if the input contains >50 genes. Larger sample sizes (more cells) may make more interactions be inferred, but the key interactions are stable (**Appendix 3—figure 3**). As HSIC.perm and DCC.perm employ permutation to perform CI test, the networks inferred each time may be somewhat different. Our data analyses suggest that interactions inferred by running distance covariance criteria (DCC) algorithms multiple times are quite stable (**Figure 3**).

Four parameters influence causal discovery. First, 'set the alpha level' determines the statistical significance cut-off of the CI test, and large and small values make more and fewer interactions be inferred. Second, 'select the number of cells' controls sample size, and selecting more cells makes the inference more reliable but also more time-consuming. Third, 'select how a subset of cells is sampled' determines how a subset of cells is sampled. If a subset is sampled randomly, the inferred network is not exactly reproducible (but by running multiple times, the inferred edges may show high consistency, see **Figure 3**). Fourth, 'set the size of conditional set' controls the size of conditional set when performing CI tests; it influences both network topology and time consumption and should be set with care. Since some CI tests are time-consuming and running causal discovery with multiple algorithms are especially time-consuming, providing an email address is necessary to make the result sent to the user automatically.

The performance of different PC+CI tests was intensively evaluated. First, we evaluated the accuracy, time consumption, sample requirement, and stability of PC+nine CI tests using simulated data and the non-small cell lung cancer (NSCLC) cell line H2228 and the normal lung alveolar cells (as the case and control) (**Tian et al., 2019; Travaglini et al., 2020**). Comparing inferred networks with the

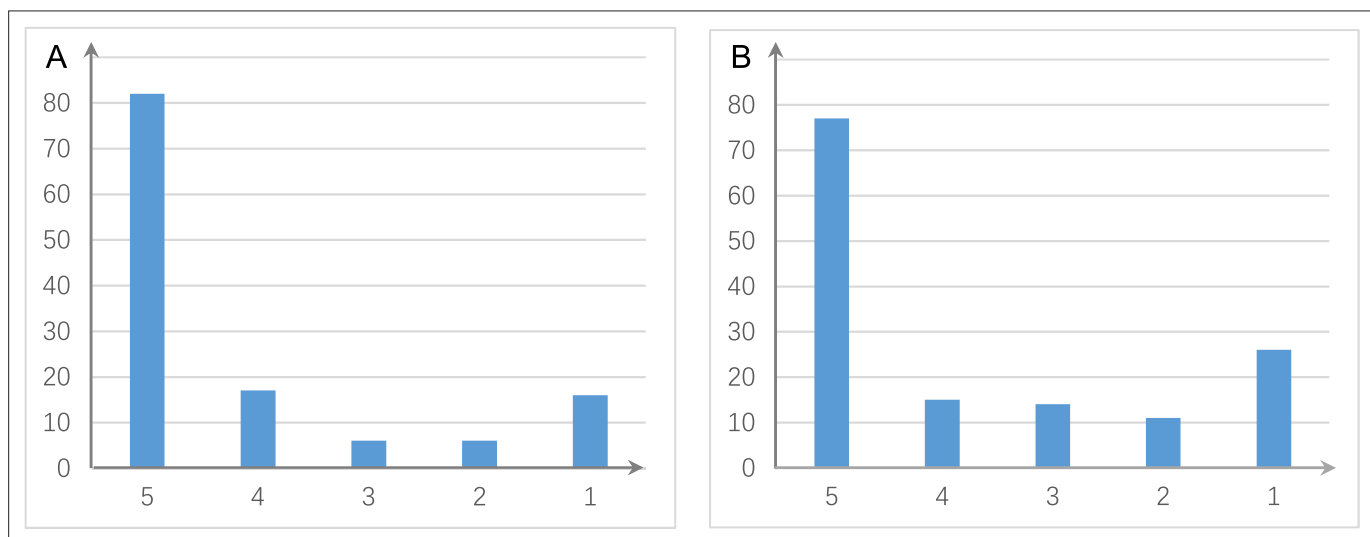


Figure 3. The shared and distinct interactions inferred by running causal discovery five times using the H2228 cell line dataset. Numbers on the vertical and horizontal axes represent the percentages of interactions in 1, 2, 3, 4, and 5 networks, respectively. **(A)** The results of PC+DCC.gamma. **(B)** The results of PC+DCC.perm. These results indicate that 78% and 64.3% of interactions occurred stably in ≥ 4 networks, suggesting that the inferred networks are quite stable.

consensus network suggests that the two DCC CI tests are most accurate and most time-consuming, suitable for small-scale network inference. RCIT and RCoT, two approximated versions of the KCIT, are moderately accurate and relatively fast, suitable for large-scale network inference. GaussCitest is the fastest and suitable for data with Gaussian distribution (**Figure 2; Appendix 3—figure 2**). Second, we compared the performance of PC+DCC.gamma, GSP+DCC.gamma, and GES. The former two have comparable performance, and both are more accurate and time-consuming than GES (**Appendix 3—figures 4 and 5**).

Verification of causal discovery

We used the five NSCLC cell lines (A549, H1975, H2228, H838, and HCC827), the normal alveolar cells, and genes in specific pathways to validate network inference by PC+DCC.gamma (**Tian et al., 2019; Travaglini et al., 2020**). First, upon the combined conditions of (a) gene expression value >0.1 , (b) gene expression in $>50\%$ cells, and (c) fold change >0.3 , we identified highly and differentially expressed genes in each cell line against the alveolar cells. Second, we applied gene set enrichment analysis to differentially expressed genes in each cell line using the g:Profiler and GSEA programs. g:Profiler identified 'Metabolic reprogramming in colon cancer' (WP4290), 'Pyrimidine metabolism' (WP4022), and 'Nucleotide metabolism' (hsa01232) as enriched pathways in all cancer cell lines, and GSEA identified 'Non-small cell lung cancer' (hsa05223) as an enriched pathway in cancer cell lines ('WP' and 'hsa' indicate WikiPathways and KEGG pathways). Many studies reveal that glucose metabolism is reprogrammed and nucleotide synthesis is increased in cancer cells. Key features of reprogrammed glucose metabolism in cancer cells include increased glucose intake, increased lactate generation, and using the glycolysis/TCA cycle intermediates to synthesize nucleotides. The networks inferred by PC+DCC.gamma capture these features despite of the absence of metabolites in these datasets. The networks of WP4022 also capture the key features of pyrimidine metabolism. In the networks of hsa05223, over 50% inferred interactions agree with pathway annotations. These results support network inference (**Appendix 4**).

Evaluating and ensuring the reliability

Single-cell data vary in quality and sample sizes; thus, it is important to effectively evaluate and ensure the reliability of network inference. Inspired by using RNA spike-in to measure RNA-seq quality (**Jiang et al., 2011**), we developed a method to evaluate and ensure the reliability of causal discovery. This method includes three steps: extracting the data of several well-known genes and their interactions from certain dataset as the 'spike-in' data, integrating the spike-in data into the case dataset, and applying causal discovery to the integrated dataset (the latter two steps are performed automatically when a spike-in dataset is chosen or uploaded). The user can choose a spike-in dataset in the platform or design and upload a spike-in dataset. In the inferred network, a clear separation of genes and their interactions in the spike-in dataset from genes and interactions in the case dataset is an indicator of reliable inference (**Appendix 4—figure 1**). Some public databases (e.g. the STRING database, <https://string-db.org/>) can also be used to evaluate inferred interactions (**Appendix 4—figures 2 and 3**).

Results

The analysis of lung cancer cell lines and alveolar epithelial cells

Down-regulated MHC-II genes help cancer cells avoid being recognized by immune cells (**Rooney et al., 2015**); thus, identifying genes and interactions involved in MHC-II gene down-regulation is important. To assess if causal discovery helps identify the related interactions, we examined the five NSCLC cancer cell lines (A549, H1975, H2228, H838, and HCC827) and the normal alveolar epithelial cells (**Tian et al., 2019; Travaglini et al., 2020**). For each of the six datasets, we took the five MHC-II genes (*HLA-DPA1*, *HLA-DPB1*, *HLA-DRA*, *HLA-DRB1*, *HLA-DRB5*) as the response variables (genes of interest, hereafter also called target genes) and selected 50 feature genes (using BAHSIC, unless otherwise stated) from all genes expressed in $>50\%$ cells. Then, we applied the nine causal discovery algorithms to the 50 genes in 300 cells sampled from each of the datasets. The two DCC algorithms performed the best when processing the H2228 cells and lung alveolar epithelial cells (**Appendix 5—figures 1 and 2**).

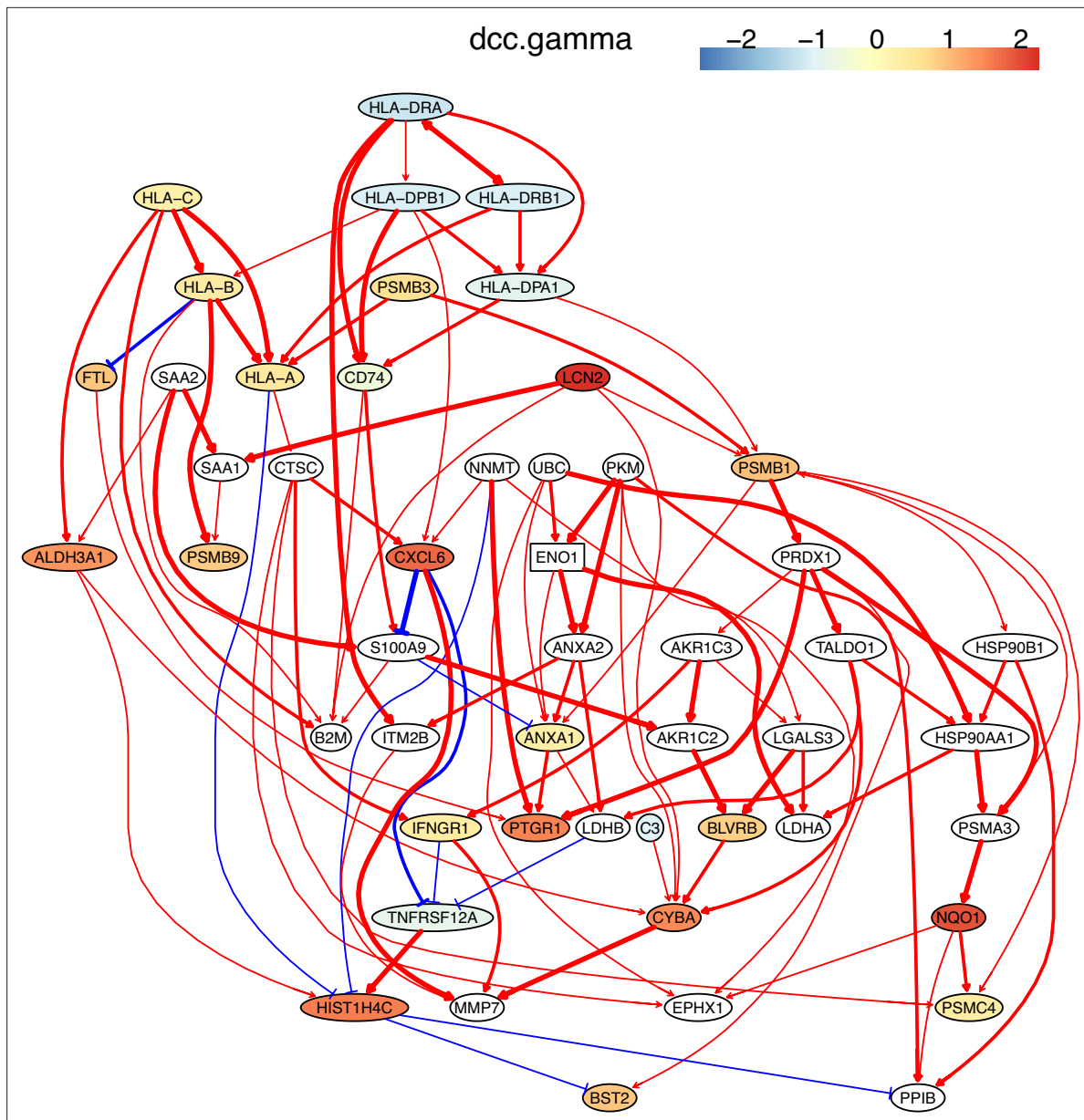


Figure 4. The network of the 50 genes inferred by DCC.gamma from the H2228 dataset (the alpha level for CI test was 0.1). Red → and blue -| arrows indicate activation and repression, and colors indicate fold changes of gene expression compared with genes in the alveolar epithelial cells.

The inferred networks also show that down-regulated genes weakly but up-regulated genes strongly regulate downstream targets and that activation and repression lead to up- and down-regulation of target genes. These features are biologically reasonable. Many inferred interactions, including those between MHC-II genes and CD74, between CXCL genes, and between MHC-I genes and B2M, are supported by the STRING database (<http://string-db.org>) and experimental findings (Figure 4; Appendix 4—figure 2; Castro et al., 2019; Karakikes et al., 2012; Szklarczyk et al., 2021). An interesting finding is the PRDX1→TALDO1→HSP90AA1→NQO1→PSMC4 cascade in H2228 cells. Interactions between PRDX1/TALDO1/HSP90AA1 and NQO1 were reported (Mathew et al., 2013; Yin et al., 2021), but the interaction between NQO1 and PSMC4 was not. Previous findings on NQO1 include that it determines cellular sensitivity to the antitumor agent napabucasin in many cancer cell lines (Guo et al., 2020), is a potential poor prognostic biomarker, and is a promising therapeutic target for patients with lung cancers (Cheng et al., 2018; Siegel et al., 2012), and that mutations in NQO1 are associated with susceptibility to various forms of cancer. Previous findings on

PSMC4 include that high levels of PSMC4 (and other PSMC) transcripts were positively correlated with poor breast cancer survival (Kao et al., 2021). Thus, the inferred NQO1→PSMC4 probably somewhat explains the mechanism behind these experimental findings.

The analysis of macrophages isolated from glioblastoma

Macrophages critically influence glioma formation, maintenance, and progression (Gutmann, 2020), and CD74 is the master regulator of macrophage functions in glioblastoma (Alban et al., 2020; Quail and Joyce, 2017; Zeiner et al., 2015). To examine the function of CD74 in macrophages in gliomas, we used CD74 as the target gene and selected 50 genes from genes expressed in >50% of macrophages isolated from glioblastoma patients (Nefitel et al., 2019). In the networks of DCC algorithms (Appendix 5—figure 3), CD74 regulates MHC-II genes, agreeing with the finding that CD74 is an MHC-II chaperone and plays a role in the intracellular sorting of MHC class II molecules. The network includes interactions between C1QA/B/C, agreeing that they form the complement C1q complex. The identified TYROBP→TREM2→A2M→APOE→APOC1 cascade is supported by the reports that TREM2 is expressed in tumor macrophages in over 200 human cancer cases (Molgora et al., 2020) and that there are interactions between TREM2/A2M, TREM2/APOE, A2M/APOE, and APOE/APOC1 (Krasemann et al., 2017).

The analysis of tumor-infiltrating exhausted CD8 T cells

Tumor-infiltrating exhausted CD8 T cells are highly heterogeneous yet share common differentially expressed genes (McLane et al., 2019; Zhang et al., 2018), suggesting that CD8 T cells undergo different processes to reach exhaustion. We analyzed three exhausted CD8 T datasets isolated from human liver, colorectal, and lung cancers (Appendix 5—figure 4; Guo et al., 2018; Zhang et al., 2018; Zheng et al., 2017). A key feature of CD8 T cell exhaustion identified in mice is PDCD1 up-regulation by TOX (Khan et al., 2019; Scott et al., 2019; Seo et al., 2019). Using TOX and PDCD1 as the target gene, we selected 50 genes expressed in >50% exhausted CD8 T cells and 50 genes expressed in >50% non-exhausted CD8 T cells, respectively. Transcriptional regulation of PDCD1 by TOX was observed in LCMV-infected mice without mentioning any role of CXCL13 (Khan et al., 2019). Here, indirect TOX→PDCD1 (via genes such as CXCL13) was inferred in exhausted CD8 cells, and direct TOX→PDCD1 was inferred in non-exhausted CD8 T cells (although the expression of TOX and PDCD1 is low in these cells) (Appendix 5—figure 4). Recently, CXCL13 was found to play a critical role in T cells for effective responses to anti-PD-L1 therapies (Zhang et al., 2021b). The causal discovery results help reveal differences in CD8 T cell exhaustion between humans and mice and under different pathological conditions. The PDCD1→TOX inferred in exhausted and non-exhausted CD8 T cells may indicate some feedback between TOX and PDCD1, as on the proteome level, a study reported that the binding of PD1 to TOX in the cytoplasm facilitates the endocytic recycling of PD1 (Wang et al., 2019).

Identifying genes and inferring interactions that signify CD4 T cell aging

How immune cells age and whether some senescence signatures reflect the aging of all cell types draw wide attention (Gorgoulis et al., 2019). We analyzed gene expression in naive, TEM, rTreg, naive_lsg15, cytotoxic, and exhausted CD4 T cells from young (2–3 months, n=4) and old (22–24 months, n=4) mice (Appendix 5—figures 5; Elyahu et al., 2019). For each cell type, we compared the combined data from all four young mice with the data from each old mouse to identify differentially expressed genes. If genes were expressed in >25% cells and consistently up/down-regulated ($|\text{fold change}| > 0$) in most of the 24 comparisons, we assumed them as aging-related (Appendix 5—table 1). Some of these identified genes play important roles in the aging of T cells or other cells, such as the mitochondrial genes encoding cytochrome c oxidases and the gene *Sub1* in the mTOR pathway (Bektas et al., 2019; Gorgoulis et al., 2019; Goronzy and Weyand, 2019; Walters and Cox, 2021). We directly used these genes, plus one CD4-specific biomarker (Cd28) and two reported aging biomarkers (Cdkn1b, Cdkn2d) (Gorgoulis et al., 2019; Larbi and Fulop, 2014), as feature genes to infer their interactions in different CD4 T cells in young and old mice. The inferred causal networks unveil multiple findings (Appendix 5—figure 5). First, B2m→H2-Q7 (a mouse MHC class I gene), Gm9843→Rps27rt (Gm9846), and the interactions between the five mitochondrial genes

(MT-ATP6, MT-CO1/2/3, and MT-Nd1) were inferred in nearly all CD4 T cells. Second, many interactions are supported by the STRING database (**Appendix 4—figure 3**). Third, some interactions agree with experimental findings, including Sub1-|Lamtor2 (**Chen et al., 2021**) and the regulations of these mitochondrial genes by Lamtor2 (**Morita et al., 2017**). Fourth, Gm9843→Rps27rt→Junb were inferred in multiple CD4 T cells (both Gm9843 and Rps27rt are mouse-specific). Since JUNB belongs to the AP-1 family TFs that are increased in all immune cells during human aging (**Zheng et al., 2020**), Gm9843→Rps27rt→Junb could highlight a counterpart regulation of JUNB in human immune cells.

Discussion

Single-cell causal discovery

Various methods have been proposed to infer interactions between variables from observational data. As surveyed recently (**Nguyen et al., 2021; Pratapa et al., 2020**), many methods assume linear relationships between variables and the Gaussian distribution of data. These assumptions enable these methods to run fast, handle many genes and even perform transcriptome-wide prediction. However, our algorithm benchmarking results suggest that networks inferred by fast methods with these assumptions should be concerned.

Causal discovery infers causal interactions directly upon observations of variables without assuming relationships between variables and the distribution of data. Because genes and molecules have varied relationships in different cells, causal discovery better satisfies inferring their interactions than other methods. Causal discovery methods have reviewed recently (**Glymour et al., 2019; Yuan and Shou, 2022**), but workflows and platforms integrating multiple methods for analyzing scRNA-seq data remain rare.

Our integration and benchmarking of multiple methods (note that these methods are not for inferring causal relationships from temporal data) and analysis of multiple datasets generate several conclusions. First, although kernel-based CI tests are time-consuming (**Shah and Peters, 2020**), applying them to a set of genes is feasible. A set of genes can be generated by feature selection, by gene set enrichment analysis, or by manual selection. Second, the cost of time consumption pays off in network accuracy, as the most time-consuming CI tests generate the most reliable results. Thus, trade-offs between time consumption, network size, and network accuracy should be made. Third, causal discovery can infer signaling networks or gene regulatory networks, depending on the input. If genes encoding TFs and their targets are the input, gene regulatory networks are inferred. Fourth, dropouts and noises in scRNA-seq data concern researchers and trouble correlation analysis (**Hou et al., 2020; Mohan and Pearl, 2018; Tu et al., 2019**), but can be well tolerated by PC+kernel-based CI tests if samples are sufficiently large. Finally, using 'spike-in' data can effectively evaluate the reliability of causal discovery.

Challenges of data analysis

Single-cell causal discovery also faces several challenges. First, causal discovery assumes there are no unmeasured common causes (the causal sufficiency assumption), but in real data latent and unobserved variables are common and hard to identify. Specifically, inferring interactions between highly expressed or differentially expressed genes is a case of causal discovery with incomplete models (i.e. models with missing variables from the data-generating model). In this situation, what are inferred are indirect relationships instead of direct interactions between gene products. Second, constraint-based methods cannot differentiate networks belonging to a Markov equivalent class (the causal Markov assumption). This can be solved partly by combined use of PC and DAGMA-nonlinear (which can better determine the direction of edges). Third, the following examples indicate that the lack of relevant information makes judging inferred interactions and relationships difficult. (a) TOX is reported to activate PDCD1 in exhausted CD8 T cells in mice (**Khan et al., 2019**), but whether CXCL13 is involved in (or required for) the TOX-PDCD1 interaction in exhausted CD8 T cells in humans is unclear, until recently CXCL13 is reported to play critical roles in T cells for effective responses to anti-PD-L1 therapies (**Zhang et al., 2021b**). (b) The differences in inferred networks in exhausted CD8 T cells from different cancers are puzzling, until a recent study reports that exhausted CD8 T cells show high heterogeneity and exhaustion can follow different paths (**Zheng et al., 2021**). (c) It is difficult to explain multiple genes encoding ribosomal proteins in the inferred networks in CD4 cells from old

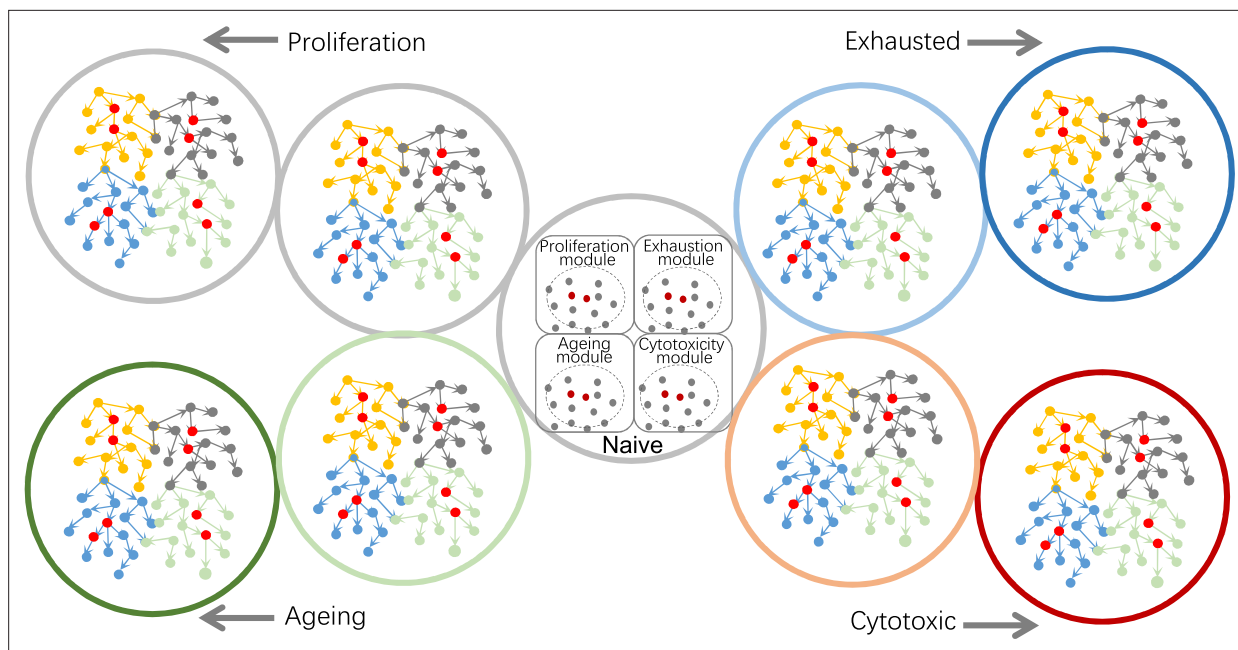


Figure 5. Using causal discovery to analyze different cells, cells at different stages, or different biological processes in cells. The red and gray dots within the four circles in the central cell indicate the four modules' core genes and related genes. Genes in different modules should be chosen as target genes when exploring different biological processes.

mice, until a recent study reports that aging impairs ribosomes' ability to synthesize proteins efficiently (Stein *et al.*, 2022).

Limitations of the study

The time consumption of kernel-based CI tests disallows inferring large networks, and how this challenge can be solved remains unsolved. C codes may be developed to replace the most time-consuming parts of the R functions, but this has not been done.

Tips for best practices

First, exploring different biological modules or processes needs careful selection of genes (Figure 5). When it is unclear what genes are most relevant to one or several target genes, it is advisable to run multiple rounds of feature selection using different combination of target genes as response variable(s). Second, when feature genes are identified by gene set enrichment analysis or upon highly expressed genes, PC+kernel-based CI tests perform better than continuous optimization-based methods, and the inferred networks consist more likely of indirect causal relationships instead of direct causal interactions. Third, BAHSIC and SHS are the best feature selection algorithms. Since selecting feature genes from too many candidates is unreliable, filtering genes upon specific conditions (e.g. expression values, expressed cells, fold changes) is necessary. Fourth, DCC.gamma and DCC.perm are the best CI tests working with PC. When building consensus networks, it is advisable to use the results of just DCC CI tests. Fifth, trade-offs between scale, reliability, and accuracy are inevitable. When examining many genes, RCIT/RCoT may be proper, and when examining large datasets, sub-sampling is necessary. For Smart-seq2 and 10x Genomics datasets, 300 and 600 cells are recommended for analyzing 50–60 genes expressed in >50% of cells. More cells are needed if more genes are selected and/or selected genes are expressed in fewer cells (e.g. 25%). Sixth, when it is unclear if a sub-sampled dataset is large enough, repeat causal discovery several times using different sizes of sub-samples. If the inferred networks are similar, the sub-samples should be sufficient. Seventh, using "spike-in" datasets helps measure and ensure reliability. Eighth, carefully inspect the potential influence of cell heterogeneity on causal discovery, and caution is needed when interpreting the results of heterogeneous cells.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31771456) and the Department of Science and Technology of Guangdong Province (2020A1515010803). We appreciate the help from Prof. Ruichu Cai at the Guangdong University of Technology.

Additional information

Funding

Funder	Grant reference number	Author
National Natural Science Foundation of China	31771456	Hao Zhu
Department of Science and Technology of Guangdong Province	2020A1515010803	Hao Zhu

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Yujian Wen, Jielong Huang, Software, Formal analysis, Methodology; Shuhui Guo, Software; Yehezqel Elyahu, Resources, Data curation; Alon Monsonogo, Resources, Data curation, Writing - review and editing; Hai Zhang, Software, Visualization; Yanqing Ding, Data curation, Supervision; Hao Zhu, Conceptualization, Formal analysis, Supervision, Funding acquisition, Writing - original draft, Project administration, Writing - review and editing

Author ORCIDs

Hao Zhu  <http://orcid.org/0000-0001-7384-3840>

Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.81464.sa1>
 Author response <https://doi.org/10.7554/eLife.81464.sa2>

Additional files

Supplementary files

- MDAR checklist

Data availability

Only public data were used. Links to all data are provided in the manuscript.

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Geirsdottir L	2019	Cross-species analysis across 450 million years of evolution reveals conservation and divergence of the microglia program (scRNA-seq)	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134705	NCBI Gene Expression Omnibus, GSE134705
Tian L	2019	Designing a single cell RNA sequencing benchmark dataset to compare protocols and analysis methods [5 Cell Lines 10X]	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126906	NCBI Gene Expression Omnibus, GSE126906

Continued on next page

Continued

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Travaglini KJ	2020	Human Lung Cell Atlas	https://www.synapse.org/#!Synapse:syn21041850	Synapse, syn21041850
Elyahu Y	2019	Study: Aging promotes reorganization of the CD4 T cell landscape toward extreme regulatory and effector phenotypes	https://singlecell.broadinstitute.org/single_cell/study/SCP490/aging-promotes-reorganization-of-the-cd4-t-cell-landscape-toward-extreme-regulatory-and-effector-phenotypes	Single Cell Portal, SCP490
Neftel C	2019	Single cell RNA-seq analysis of adult and paediatric IDH-wildtype Glioblastomas	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131928	NCBI Gene Expression Omnibus, GSE131928
Guo X	2018	T cell landscape of non-small cell lung cancer revealed by deep single-cell RNA sequencing	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99254	NCBI Gene Expression Omnibus, GSE99254
Zhang L	2018	Lineage tracking reveals dynamic relationships of T cells in colorectal cancer	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108989	NCBI Gene Expression Omnibus, GSE108989
Zheng C	2018	Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98638	NCBI Gene Expression Omnibus, GSE98638

References

- Alban TJ**, Bayik D, Otvos B, Rabljenovic A, Leng L, Jia-Shiun L, Roversi G, Lauko A, Momin AA, Mohammadi AM, Peereboom DM, Ahluwalia MS, Matsuda K, Yun K, Bucala R, Vogelbaum MA, Lathia JD. 2020. Glioblastoma myeloid-derived suppressor cell subsets express differential macrophage migration inhibitory factor receptor profiles that can be targeted to reduce immune suppression. *Frontiers in Immunology* **11**:1191. DOI: <https://doi.org/10.3389/fimmu.2020.01191>, PMID: 32625208
- Amoah A**, Keller A, Emini R, Hoenicka M, Liebold A, Vollmer A, Eiwien K, Soller K, Sakk V, Zheng Y, Florian MC, Geiger H. 2021. Aging of human hematopoietic stem cells is linked to changes in Cdc42 activity. *Haematologica* **107**:393–402. DOI: <https://doi.org/10.3324/haematol.2020.269670>
- Angelidis I**, Simon LM, Fernandez IE, Strunz M, Mayr CH, Greiffo FR, Tsitsiridis G, Ansari M, Graf E, Strom T-M, Nagendran M, Desai T, Eickelberg O, Mann M, Theis FJ, Schiller HB. 2019. An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nature Communications* **10**:963. DOI: <https://doi.org/10.1038/s41467-019-08831-9>, PMID: 30814501
- Avelar RA**, Ortega JG, Tacutu R, Tyler EJ, Bennett D, Binetti P, Budovsky A, Chatsirisupachai K, Johnson E, Murray A, Shields S, Tejada-Martinez D, Thornton D, Fraifeld VE, Bishop CL, de Magalhães JP. 2020. A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biology* **21**:91. DOI: <https://doi.org/10.1186/s13059-020-01990-9>, PMID: 32264951
- Balcazar N**, Sathyamurthy A, Elghazi L, Gould A, Weiss A, Shiojima I, Walsh K, Bernal-Mizrachi E. 2009. Mtorc1 activation regulates beta-cell mass and proliferation by modulation of cyclin D2 synthesis and stability. *The Journal of Biological Chemistry* **284**:7832–7842. DOI: <https://doi.org/10.1074/jbc.M807458200>, PMID: 19144649
- Bektas A**, Schurman SH, Gonzalez-Freire M, Dunn CA, Singh AK, Macian F, Cuervo AM, Sen R, Ferrucci L. 2019. Age-Associated changes in human CD4+ T cells point to mitochondrial dysfunction consequent to impaired autophagy. *Aging* **11**:9234–9263. DOI: <https://doi.org/10.18632/aging.102438>, PMID: 31707363
- Bello K**, Aragam B, Ravikumar P. 2022a. DAGMA: Learning DAGs via M-Matrices and a Log-Determinant Acyclicity Characterization NeurIPS.
- Bello K**, Aragam B, Ravikumar P. 2022b. Dagma. 7c6ba5f. Github. <https://github.com/kevinsbello/dagma>
- Bhalla US**, Iyengar R. 1999. Emergent properties of networks of biological signaling pathways. *Science* **283**:381–387. DOI: <https://doi.org/10.1126/science.283.5400.381>, PMID: 9888852

- Breiman L.** 2001. Random forests. *Machine Learning* **45**:5–32. DOI: <https://doi.org/10.1023/A:1010933404324>
- Butovsky O**, Jedrychowski MP, Moore CS, Cialic R, Lanser AJ, Gabriely G, Koeglsperger T, Dake B, Wu PM, Doykan CE, Fanek Z, Liu L, Chen Z, Rothstein JD, Ransohoff RM, Gygi SP, Antel JP, Weiner HL. 2014. Identification of a unique TGF- β -dependent molecular and functional signature in microglia. *Nature Neuroscience* **17**:131–143. DOI: <https://doi.org/10.1038/nn.3599>, PMID: 24316888
- Castro A**, Ozturk K, Pyke RM, Xian S, Zanetti M, Carter H. 2019. Elevated neoantigen levels in tumors with somatic mutations in the HLA-A, HLA-B, HLA-C and B2M genes. *BMC Medical Genomics* **12**:107. DOI: <https://doi.org/10.1186/s12920-019-0544-1>, PMID: 31345234
- Chen T**, Guestrin C. 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (San Francisco, California, USA: Association for Computing Machinery). , 785, 794. DOI: <https://doi.org/10.1145/2939672.2939785>
- Chen L**, Liao F, Wu J, Wang Z, Jiang Z, Zhang C, Luo P, Ma L, Gong Q, Wang Y, Wang Q, Luo M, Yang Z, Han S, Shi C. 2021. Acceleration of ageing via disturbing mtor-regulated proteostasis by a new ageing-associated gene PC4. *Aging Cell* **20**:e13370. DOI: <https://doi.org/10.1111/accel.13370>
- Cheng X**, Liu F, Liu H, Wang G, Hao H. 2018. Enhanced glycometabolism as a mechanism of NQO1 potentiated growth of NSCLC revealed by metabolomic profiling. *Biochemical and Biophysical Research Communications* **496**:31–36. DOI: <https://doi.org/10.1016/j.bbrc.2017.12.160>
- Chickering DM.** 2003. Optimal structure identification with greedy search. *Journal of Machine Learning Research* **3**:507–554.
- Climente-González H**, Azencott C-A, Kaski S, Yamada M. 2019. Block HSIC LASSO: model-free biomarker detection for ultra-high dimensional data. *Bioinformatics* **35**:i427–i435. DOI: <https://doi.org/10.1093/bioinformatics/btz333>, PMID: 31510671
- Deming Y**, Filipello F, Cignarella F, Cantoni C, Hsu S, Mikesell R, Li Z, Del-Aguila JL, Dube U, Farias FG, Bradley J, Budde J, Ibanez L, Fernandez MV, Blennow K, Zetterberg H, Heslegrave A, Johansson PM, Svensson J, Nellgård B, et al. 2019. The MS4A gene cluster is a key modulator of soluble TREM2 and Alzheimer's disease risk'. *Science Translational Medicine* **11**:eaau2291. DOI: <https://doi.org/10.1126/scitranslmed.aau2291>
- Deshpande A**, Chu LF, Stewart R, Gitter A. 2019. Network Inference with Granger Causality Ensembles on Single-Cell Transcriptomic Data. *bioRxiv* . . DOI: <https://doi.org/10.1101/534834>
- Dong Q**, Du Y, Li H, Liu C, Wei Y, Chen M-K, Zhao X, Chu Y-Y, Qiu Y, Qin L, Yamaguchi H, Hung M-C. 2019. Egfr and c-Met cooperate to enhance resistance to PARP inhibitors in hepatocellular carcinoma. *Cancer Research* **79**:819–829. DOI: <https://doi.org/10.1158/0008-5472.CAN-18-1273>, PMID: 30573522
- Dupont-Versteegden EE**, Nagarajan R, Beggs ML, Bearden ED, Simpson PM, Peterson CA. 2008. Identification of cold-shock protein RBM3 as a possible regulator of skeletal muscle size through expression profiling. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **295**:R1263–R1273. DOI: <https://doi.org/10.1152/ajpregu.90455.2008>
- Elyahu Y**, Hekselman I, Eizenberg-Magar I, Berner O, Strominger I, Schiller M, Mittal K, Nemirovsky A, Eremenko E, Vital A, Simonovsky E, Chalifa-Caspi V, Friedman N, Yeger-Lotem E, Monsonogo A. 2019. Aging promotes reorganization of the CD4 T cell landscape toward extreme regulatory and effector phenotypes. *Science Advances* **5**:eaaw8330. DOI: <https://doi.org/10.1126/sciadv.aaw8330>, PMID: 31457092
- Gangeh MJ**, Zarkoob H, Ghodsi A. 2017. Fast and scalable feature selection for gene expression data using hilbert-schmidt independence criterion. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **14**:167–181. DOI: <https://doi.org/10.1109/TCBB.2016.2631164>, PMID: 28182548
- Garcia GG**, Miller RA. 2009. Age-Related changes in lck-vav signaling pathways in mouse CD4 T cells. *Cellular Immunology* **259**:100–104. DOI: <https://doi.org/10.1016/j.cellimm.2009.06.001>, PMID: 19577230
- Geiger H**, Zheng Y. 2013. Cdc42 and aging of hematopoietic stem cells. *Current Opinion in Hematology* **20**:295–300. DOI: <https://doi.org/10.1097/MOH.0b013e3283615aba>, PMID: 23615056
- Geirsdottir L**, David E, Keren-Shaul H, Weiner A, Bohlen SC, Neuber J, Balic A, Giladi A, Sheban F, Dutertre C-A, Pfeifle C, Peri F, Raffo-Romero A, Vizioli J, Matiassek K, Scheiwe C, Meckel S, Mätz-Rensing K, van der Meer F, Thormodsson FR, et al. 2019. Cross-Species single-cell analysis reveals divergence of the primate microglia program. *Cell* **179**:1609–1622.. DOI: <https://doi.org/10.1016/j.cell.2019.11.010>, PMID: 31835035
- Geurts P**, Ernst D, Wehenkel L. 2006. Extremely randomized trees. *Machine Learning* **63**:3–42. DOI: <https://doi.org/10.1007/s10994-006-6226-1>
- Glymour C**, Zhang K, Spirtes P. 2019. Review of causal discovery methods based on graphical models. *Frontiers in Genetics* **10**:524. DOI: <https://doi.org/10.3389/fgene.2019.00524>, PMID: 31214249
- Gorgoulis V**, Adams PD, Alimonti A, Bennett DC, Bischof O, Bishop C, Campisi J, Collado M, Evangelou K, Ferbeyre G, Gil J, Hara E, Krizhanovsky V, Jurk D, Maier AB, Narita M, Niedernhofer L, Passos JF, Robbins PD, Schmitt CA, et al. 2019. Cellular senescence: defining a path forward. *Cell* **179**:813–827. DOI: <https://doi.org/10.1016/j.cell.2019.10.005>, PMID: 31675495
- Goronzy JJ**, Weyand CM. 2019. Mechanisms underlying T cell ageing. *Nature Reviews. Immunology* **19**:573–583. DOI: <https://doi.org/10.1038/s41577-019-0180-1>, PMID: 31186548
- Gretton A**, Bousquet O, Smola A. 2005. Measuring statistical dependence with Hilbert-Schmidt norms. [International Conference on Algorithmic Learning Theory].63–77. DOI: <https://doi.org/10.1007/11564089>
- Guo X**, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, Kang B, Liu Z, Jin L, Xing R, Gao R, Zhang L, Dong M, Hu X, Ren X, Kirchhoff D, Roeder HG, Yan T, Zhang Z. 2018. Publisher correction: global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nature Medicine* **24**:978–985. DOI: <https://doi.org/10.1038/s41591-018-0167-7>, PMID: 30093729

- Guo G**, Gao Z, Tong M, Zhan D, Wang G, Wang Y, Qin J. 2020. Nqo1 is a determinant for cellular sensitivity to anti-tumor agent napabucasin. *American Journal of Cancer Research* **10**:1442–1454 PMID: 32509390.
- Gutmann DH**. 2020. The sociobiology of brain tumors. *Advances in Experimental Medicine and Biology* **1225**:115–125. DOI: https://doi.org/10.1007/978-3-030-35727-6_8, PMID: 32030651
- Haas RH**. 2019. Mitochondrial dysfunction in aging and diseases of aging. *Biology* **8**:48. DOI: <https://doi.org/10.3390/biology8020048>
- He Y**, Li D, Cook SL, Yoon M-S, Kapoor A, Rao CV, Kenis PJA, Chen J, Wang F. 2013. Mammalian target of rapamycin and Rictor control neutrophil chemotaxis by regulating Rac/Cdc42 activity and the actin cytoskeleton. *Molecular Biology of the Cell* **24**:3369–3380. DOI: <https://doi.org/10.1091/mbc.E13-07-0405>, PMID: 24006489
- Hoerl AE**, Kennard RW. 2000. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **42**:80–86. DOI: <https://doi.org/10.1080/00401706.2000.10485983>
- Hou W**, Ji Z, Ji H, Hicks SC. 2020. A systematic evaluation of single-cell RNA-sequencing imputation methods. *Genome Biology* **21**:218. DOI: <https://doi.org/10.1186/s13059-020-02132-x>
- Hwang JS**, Hwang JS, Chang I, Kim S. 2007. Age-Associated decrease in proteasome content and activities in human dermal fibroblasts: restoration of normal level of proteasome subunits reduces aging markers in fibroblasts from elderly persons. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* **62**:490–499. DOI: <https://doi.org/10.1093/gerona/62.5.490>, PMID: 17522352
- Jiang L**, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, Gingeras TR, Oliver B. 2011. Synthetic spike-in standards for RNA-seq experiments. *Genome Research* **21**:1543–1551. DOI: <https://doi.org/10.1101/gr.121095.111>
- Joehanes R**. 2018. Network analysis of gene expression. *Methods in Molecular Biology* **1783**:325–341. DOI: https://doi.org/10.1007/978-1-4939-7834-2_16, PMID: 29767370
- Kao TJ**, Wu CC, Phan NN, Liu YH, Ta HDK, Anuraga G, Wu YF, Lee KH, Chuang JY, Wang CY. 2021. Prognoses and genomic analyses of proteasome 26S subunit, ATPase (PSMC) family genes in clinical breast cancer. *Aging* **13**:17970. DOI: <https://doi.org/10.18632/aging.203345>, PMID: 34329194
- Karakikes I**, Morrison IEG, O'Toole P, Metodieva G, Navarrete CV, Gomez J, Miranda-Sayago JM, Cherry RJ, Metodiev M, Fernandez N. 2012. Interaction of HLA-DR and CD74 at the cell surface of antigen-presenting cells by single particle image analysis. *FASEB Journal* **26**:4886–4896. DOI: <https://doi.org/10.1096/fj.12-211466>, PMID: 22889831
- Kerber RA**, O'Brien E, Cawthon RM. 2009. Gene expression profiles associated with aging and mortality in humans. *Aging Cell* **8**:239–250. DOI: <https://doi.org/10.1111/j.1474-9726.2009.00467.x>, PMID: 19245677
- Khan O**, Giles JR, McDonald S, Manne S, Ngiow SF, Patel KP, Werner MT, Huang AC, Alexander KA, Wu JE, Attanasio J, Yan P, George SM, Bengsch B, Staupel RP, Donahue G, Xu W, Amaravadi RK, Xu X, Karakousis GC, et al. 2019. Tox transcriptionally and epigenetically programs CD8+ T cell exhaustion. *Nature* **571**:211–218. DOI: <https://doi.org/10.1038/s41586-019-1325-x>
- Kim DH**, Park JW, Jeong HO, Lee B, Chung KW, Lee Y, Jung HJ, Hyun MK, Lee AK, Kim BM, Yu BP, Chung HY. 2019. Novel role of Lck in leptin-induced inflammation and implications for renal aging. *Aging and Disease* **10**:1174. DOI: <https://doi.org/10.14336/AD.2019.0218>
- Kirkland JL**, Hollenberg CH, Gillon WS. 1993. Effects of aging on ribosomal protein L7 messenger RNA levels in cultured rat preadipocytes. *Experimental Gerontology* **28**:557–563. DOI: [https://doi.org/10.1016/0531-5565\(93\)90044-e](https://doi.org/10.1016/0531-5565(93)90044-e), PMID: 8137892
- Krasemann S**, Madore C, Cialic R, Baufeld C, Calcagno N, El Fatimy R, Beckers L, O'Loughlin E, Xu Y, Fanek Z, Greco DJ, Smith ST, Tweet G, Humulock Z, Zrzavy T, Conde-Sanroman P, Gacias M, Weng Z, Chen H, Tjon E, et al. 2017. The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity* **47**:566–581. DOI: <https://doi.org/10.1016/j.immuni.2017.08.008>, PMID: 28930663
- Kunihiro B**, Ritei S, Masaaki S. 2004. Partial correlation and conditional correlation as measures of conditional independence. *Australian New Zealand Journal of Statistics* **46**:657–664. DOI: <https://doi.org/10.1111/j.1467-842X.2004.00360.x>
- Larbi A**, Fulop T. 2014. From "truly naive" to "exhausted senescent" T cells: when markers predict functionality. *Cytometry. Part A* **85**:25–35. DOI: <https://doi.org/10.1002/cyto.a.22351>, PMID: 24124072
- Le TD**, Hoang T, Li J, Liu L, Liu H, Hu S. 2019. A fast PC algorithm for high dimensional causal discovery with multi-core pcs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **16**:1483–1495. DOI: <https://doi.org/10.1109/TCBB.2016.2591526>, PMID: 27429444
- Le Page A**, Dupuis G, Larbi A, Witkowski JM, Fülöp T. 2018. Signal transduction changes in CD4 + and CD8 + T cell subpopulations with aging. *Experimental Gerontology* **105**:128–139. DOI: <https://doi.org/10.1016/j.exger.2018.01.005>
- Lee SE**, Sun SC, Choi HY, Uhm SJ, Kim NH. 2012. Mtor is required for asymmetric division through small GTPases in mouse oocytes. *Molecular Reproduction and Development* **79**:356–366. DOI: <https://doi.org/10.1002/mrd.22035>, PMID: 22407942
- Li H**, Liu P, Xu S, Li Y, Dekker JD, Li B, Fan Y, Zhang Z, Hong Y, Yang G, Tang T, Ren Y, Tucker HO, Yao Z, Guo X. 2017. Foxp1 controls mesenchymal stem cell commitment and senescence during skeletal aging. *Journal of Clinical Investigation* **127**:1241–1253. DOI: <https://doi.org/10.1172/JCI89511>
- Li X**, Zhang B, Wang H, Zhao X, Zhang Z, Ding G, Wei F. 2020. The effect of aging on the biological and immunological characteristics of periodontal ligament stem cells. *Stem Cell Research & Therapy* **11**:326. DOI: <https://doi.org/10.1186/s13287-020-01846-w>

- Li Z, Wang S, Liu S, Xu Z, Yi X, Wang H, Dang J, Wei X, Feng B, Liu Z, Zhao M, Wu Q, Hu D. 2022. New insights into aging-associated characteristics of female subcutaneous adipose tissue through integrative analysis of multi-omics data. *Bioengineered* **13**:2044–2057. DOI: <https://doi.org/10.1080/21655979.2021.2020467>
- Lu J, Dumitrascu B, McDowell IC, Jo B, Barrera A, Hong LK, Leichter SM, Reddy TE, Engelhardt BE. 2021. Causal network inference from gene transcriptional time-series response to glucocorticoids. *PLOS Computational Biology* **17**:e1008223. DOI: <https://doi.org/10.1371/journal.pcbi.1008223>, PMID: 33513136
- Maity P, Singh K, Koroma A, Hainzl A, Bloch W, Kochanek S, Wlaschek M, Schorpp-Kistner M, Angel P, Ignatius A, Geiger H, Scharfetter-Kochanek K. 2021. Persistent junB activation in fibroblasts disrupts stem cell niche interactions enforcing skin aging. *Cell Reports* **36**:109634. DOI: <https://doi.org/10.1016/j.celrep.2021.109634>, PMID: 34469740
- Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, DREAM5 Consortium, Kellis M, Collins JJ, Stolovitzky G. 2012. Wisdom of crowds for robust gene network inference. *Nature Methods* **9**:796–804. DOI: <https://doi.org/10.1038/nmeth.2016>, PMID: 22796662
- Mathew B, Jacobson JR, Siegler JH, Moitra J, Blasco M, Xie L, Unzueta C, Zhou T, Evenoski C, Al-Sakka M, Sharma R, Huey B, Bulent A, Smith B, Jayaraman S, Reddy NM, Reddy SP, Fingerle-Rowson G, Bucala R, Dudek SM, et al. 2013. Role of migratory inhibition factor in age-related susceptibility to radiation lung injury via NF-E2-related factor-2 and antioxidant regulation. *American Journal of Respiratory Cell and Molecular Biology* **49**:269–278. DOI: <https://doi.org/10.1165/rcmb.2012-0291OC>, PMID: 23526214
- McCartney DL, Min JL, Richmond RC, Lu AT, Sobczyk MK, Davies G, Broer L, Guo X, Jeong A, Jung J, Kasela S, Katrinli S, Kuo P-L, Matias-Garcia PR, Mishra PP, Nygaard M, Palviainen T, Patki A, Raffield LM, Ratliff SM, et al. 2021. Genome-Wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging. *Genome Biology* **22**:194. DOI: <https://doi.org/10.1186/s13059-021-02398-9>, PMID: 34187551
- McLane LM, Abdel-Hakeem MS, Wherry EJ. 2019. Cd8 T cell exhaustion during chronic viral infection and cancer. *Annual Review of Immunology* **37**:457–495. DOI: <https://doi.org/10.1146/annurev-immunol-041015-055318>, PMID: 30676822
- Meek C. 1997. *Graphical Models: Selecting Causal and Statistical Models* Carnegie Mellon University Diss.
- Micheli L, Creanza TM, Ceccarelli M, D'Andrea G, Giacobazzo G, Ancona N, Coccorello R, Scardigli R, Tirone F. 2021. Transcriptome analysis in a mouse model of premature aging of dentate gyrus: rescue of alpha-synuclein deficit by virus-driven expression or by running restores the defective neurogenesis. *Frontiers in Cell and Developmental Biology* **9**:696684. DOI: <https://doi.org/10.3389/fcell.2021.696684>, PMID: 34485283
- Mohan K, Pearl J. 2018. Graphical Models for Processing Missing Data. *arXiv*. <https://arxiv.org/abs/1801.03583>
- Molgora M, Esaulova E, Vermi W, Hou J, Chen Y, Luo J, Brioschi S, Bugatti M, Omodei AS, Ricci B, Fronick C, Panda SK, Takeuchi Y, Gubin MM, Faccio R, Cella M, Gilfillan S, Unanue ER, Artyomov MN, Schreiber RD, et al. 2020. Trem2 modulation remodels the tumor myeloid landscape enhancing anti-PD-1 immunotherapy. *Cell* **182**:886–900. DOI: <https://doi.org/10.1016/j.cell.2020.07.013>, PMID: 32783918
- Morita M, Prudent J, Basu K, Goyon V, Katsumura S, Hulea L, Pearl D, Siddiqui N, Strack S, McGuirk S, St-Pierre J, Larsson O, Topisirovic I, Vali H, McBride HM, Bergeron JJ, Sonenberg N. 2017. Mtor controls mitochondrial dynamics and cell survival via mtfp1. *Molecular Cell* **67**:922–935. DOI: <https://doi.org/10.1016/j.molcel.2017.08.013>, PMID: 28918902
- Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, Rahme GJ, Richman AR, Silverbush D, Shaw ML, Hebert CM, Dewitt J, Gritsch S, Perez EM, Gonzalez Castro LN, Lan X, Druck N, Rodman C, Dionne D, Kaplan A, Bertalan MS, et al. 2019. An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* **178**:835–849. DOI: <https://doi.org/10.1016/j.cell.2019.06.024>, PMID: 31327527
- Nguyen H, Tran D, Tran B, Pehlivan B, Nguyen T. 2021. A comprehensive survey of regulatory network inference methods using single cell RNA sequencing data. *Briefings in Bioinformatics* **22**:bbaa190. DOI: <https://doi.org/10.1093/bib/bbaa190>, PMID: 34020546
- Patir A, Shih B, McColl BW, Freeman TC. 2019. A core transcriptional signature of human microglia: derivation and utility in describing region-dependent alterations associated with Alzheimer's disease. *Glia* **67**:1240–1253. DOI: <https://doi.org/10.1002/glia.23572>, PMID: 30758077
- Pearl J, Mackenzie D. 2019. *The Book of Why - The New Science of Cause and Effect* Penguin.
- Peters J, Shah RD. 2022. Generalisedcovariancemeasure: test for conditional independence based on the generalized covariance measure (GCM). CRAN. <https://cran.r-project.org/web/packages/GeneralisedCovarianceMeasure/index.html>
- Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. 2020. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature Methods* **17**:147–154. DOI: <https://doi.org/10.1038/s41592-019-0690-6>, PMID: 31907445
- Quail DF, Joyce JA. 2017. The microenvironmental landscape of brain tumors. *Cancer Cell* **31**:326–341. DOI: <https://doi.org/10.1016/j.ccell.2017.02.009>, PMID: 28292436
- Rahimi A, Recht B. 2007. Random features for large-scale kernel machines. Proceedings of the 20th International Conference on Neural Information Processing Systems. 1177–1184.
- Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. 2015. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**:48–61. DOI: <https://doi.org/10.1016/j.cell.2014.12.033>, PMID: 25594174
- Rudd CE. 2021. How the discovery of the CD4/CD8-p56lck complexes changed immunology and immunotherapy. *Frontiers in Cell and Developmental Biology* **9**:626095. DOI: <https://doi.org/10.3389/fcell.2021.626095>, PMID: 33791292

- Runge J.** 2018. Conditional Independence Testing Based on a Nearest-Neighbour Estimator of Conditional Mutual Information. Proceedings of the 21st International Conference on Artificial Intelligence and Statistics. .
- Runge J.** 2020. Tigramite. V4.2.2.1. Github. <https://github.com/jakobrunge/tigramite>
- Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP.** 2005. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308**:523–529. DOI: <https://doi.org/10.1126/science.1105809>, PMID: 15845847
- Scott AC, Dündar F, Zumbo P, Chandran SS, Klebanoff CA, Shakiba M, Trivedi P, Menocal L, Appleby H, Camara S, Zamarin D, Walther T, Snyder A, Femia MR, Comen EA, Wen HY, Hellmann MD, Anandasabapathy N, Liu Y, Altorki NK, et al.** 2019. Tox is a critical regulator of tumour-specific T cell differentiation. *Nature* **571**:270–274. DOI: <https://doi.org/10.1038/s41586-019-1324-y>, PMID: 31207604
- Seo H, Chen J, González-Avalos E, Samaniego-Castruita D, Das A, Wang YH, López-Moyado IF, Georges RO, Zhang W, Onodera A, Wu C-J, Lu L-F, Hogan PG, Bhandoola A, Rao A.** 2019. Tox and TOX2 transcription factors cooperate with NR4A transcription factors to impose CD8 + T cell exhaustion . *PNAS* **116**:12410–12415. DOI: <https://doi.org/10.1073/pnas.1905675116>
- Shah RD, Peters J.** 2020. The hardness of conditional independence testing and the generalised covariance measure. *The Annals of Statistics* **48**:1514–1538. DOI: <https://doi.org/10.1214/19-AOS1857>
- Shiow LR, Roadcap DW, Paris K, Watson SR, GrigoroVA IL, Lebet T, An J, Xu Y, Jenne CN, Föger N, Sorensen RU, Goodnow CC, Bear JE, Puck JM, Cyster JG.** 2008. The actin regulator coronin 1A is mutant in a thymic egress-deficient mouse strain and in a patient with severe combined immunodeficiency. *Nature Immunology* **9**:1307–1315. DOI: <https://doi.org/10.1038/ni.1662>, PMID: 18836449
- Siegel D, Yan C, Ross D.** 2012. Nad (P) H: quinone oxidoreductase 1 (NQO1) in the sensitivity and resistance to antitumor quinones. *Biochemical Pharmacology* **83**:1033–1040. DOI: <https://doi.org/10.1016/j.bcp.2011.12.017>, PMID: 22209713
- Smith LK, He Y, Park J-S, Bieri G, Snethlage CE, Lin K, Gontier G, Wabl R, Plambeck KE, Udeochu J, Wheatley EG, Bouchard J, Eggel A, Narasimha R, Grant JL, Luo J, Wyss-Coray T, Villeda SA.** 2015. B2-Microglobulin is a systemic pro-aging factor that impairs cognitive function and neurogenesis. *Nature Medicine* **21**:932–937. DOI: <https://doi.org/10.1038/nm.3898>, PMID: 26147761
- Solus L, Wang Y, Uhler C.** 2021. Consistency guarantees for greedy permutation-based causal inference algorithms. *Biometrika* **108**:795–814. DOI: <https://doi.org/10.1093/biomet/asaa104>
- Song L, Bedo J, Borgwardt KM, Gretton A, Smola A.** 2007. Gene selection via the bahsic family of algorithms. *Bioinformatics* **23**:i490–i498. DOI: <https://doi.org/10.1093/bioinformatics/btm216>, PMID: 17646335
- Squires C.** 2018. Causaldag. V1. Github. <https://github.com/uhlerlab/causaldag>
- Steffen KK, Dillin A.** 2016. A ribosomal perspective on proteostasis and aging. *Cell Metabolism* **23**:1004–1012. DOI: <https://doi.org/10.1016/j.cmet.2016.05.013>, PMID: 27304502
- Stein KC, Morales-Polanco F, van der Lienden J, Rainbolt TK, Frydman J.** 2022. Ageing exacerbates ribosome pausing to disrupt cotranslational proteostasis. *Nature* **601**:637–642. DOI: <https://doi.org/10.1038/s41586-021-04295-4>, PMID: 35046576
- Stelekati E, Chen Z, Manne S, Kurachi M, Ali M-A, Lewy K, Cai Z, Nzingha K, McLane LM, Hope JL, Fike AJ, Katsikis PD, Wherry EJ.** 2018. Long-Term persistence of exhausted CD8 T cells in chronic infection is regulated by microRNA-155. *Cell Reports* **23**:2142–2156. DOI: <https://doi.org/10.1016/j.celrep.2018.04.038>
- Strobl EV.** 2019. Rcit. V0.1.0. Github. <https://github.com/ericstrobl/RCIT>
- Strobl EV, Zhang K, Visweswaran S.** 2019. Approximate kernel-based conditional independence tests for fast non-parametric causal discovery. *Journal of Causal Inference* **7**:20180017. DOI: <https://doi.org/10.1515/jci-2018-0017>
- Székely GJ, Rizzo ML, Bakirov NK.** 2007. Measuring and testing dependence by correlation of distances. *The Annals of Statistics* **35**:2769–2794. DOI: <https://doi.org/10.1214/009053607000000505>
- Székely GJ, Rizzo ML.** 2009. Brownian distance covariance. *The Annals of Applied Statistics* **3**:1236–1265. DOI: <https://doi.org/10.1214/09-AOAS312>
- Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C.** 2021. The string database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research* **49**:D605–D612. DOI: <https://doi.org/10.1093/nar/gkaa1074>
- Tian L, Dong X, Freytag S, Lê Cao K-A, Su S, JalalAbadi A, Amann-Zalcenstein D, Weber TS, Seidi A, Jabbari JS, Naik SH, Ritchie ME.** 2019. Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments. *Nature Methods* **16**:479–487. DOI: <https://doi.org/10.1038/s41592-019-0425-8>, PMID: 31133762
- Tibshirani R.** 1997. The LASSO method for variable selection in the COX model. *Statistics in Medicine* **16**:385–395. DOI: [https://doi.org/10.1002/\(sici\)1097-0258\(19970228\)16:4<385::aid-sim380>3.0.co;2-3](https://doi.org/10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3), PMID: 9044528
- Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, Goh M, Chen J.** 2020. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biology* **21**:12. DOI: <https://doi.org/10.1186/s13059-019-1850-9>, PMID: 31948481
- Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, Chang S, Conley SD, Mori Y, Seita J, Berry GJ, Shrager JB, Metzger RJ, Kuo CS, Neff N, Weissman IL, Quake SR, Krasnow MA.** 2020. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**:619–625. DOI: <https://doi.org/10.1038/s41586-020-2922-4>, PMID: 33208946
- Tu R, Zhang C, Ackermann P, Mohan K, Kjellstrom H, Zhang K.** 2019. Causal discovery in the presence of missing data. Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS) 2019.

- Van Pelt DW**, Confides AL, Abshire SM, Hunt ER, Dupont-Versteegden EE, Butterfield TA. 2019. Age-Related responses to a bout of mechanotherapy in skeletal muscle of rats. *Journal of Applied Physiology* **127**:1782–1791. DOI: <https://doi.org/10.1152/jappphysiol.00641.2019>, PMID: 31670600
- Verbyla P**, Desgranges N, Richardson S, Wernisch L. 2017 Exploiting General Independence Criteria for Network Inference. *bioRxiv*. DOI: <https://doi.org/10.1101/138669>
- Verbyla P**. 2018. *Network Inference Using Independence Criteria* Cambridge University.
- Walters HE**, Cox LS. 2021. Intercellular transfer of mitochondria between senescent cells through cytoskeleton-supported intercellular bridges requires mTOR and Cdc42 signalling. *Oxidative Medicine and Cellular Longevity* **2021**:6697861. DOI: <https://doi.org/10.1155/2021/6697861>, PMID: 34373767
- Wang X**, Lin Y, Lan F, Yu Y, Ouyang X, Liu W, Xie F, Wang X, Huang Q. 2014. Bax and CDKN1A polymorphisms correlated with clinical outcomes of gastric cancer patients treated with postoperative chemotherapy. *Medical Oncology* **31**:249. DOI: <https://doi.org/10.1007/s12032-014-0249-4>
- Wang X**, He Q, Shen H, Xia A, Tian W, Yu W, Sun B. 2019. Tox promotes the exhaustion of antitumor CD8+ T cells by preventing PD1 degradation in hepatocellular carcinoma. *Journal of Hepatology* **71**:731–741. DOI: <https://doi.org/10.1016/j.jhep.2019.05.015>
- Wherry EJ**, Ha SJ, Kaech SM, Haining WN, Sarkar S, Kalia V, Subramaniam S, Blattman JN, Barber DL, Ahmed R. 2007. Molecular signature of CD8+ T cell exhaustion during chronic viral infection. *Immunity* **27**:670–684. DOI: <https://doi.org/10.1016/j.immuni.2007.09.006>
- Xiong D**, Wang Y, You M. 2020. A gene expression signature of trem2hi macrophages and $\gamma\delta$ T cells predicts immunotherapy response. *Nature Communications* **11**:5084. DOI: <https://doi.org/10.1038/s41467-020-18546-x>
- Xu S**, Li X, Tang L, Liu Z, Yang K, Cheng Q. 2021. Cd74 correlated with malignancies and immune microenvironment in gliomas. *Frontiers in Molecular Biosciences* **8**:706949. DOI: <https://doi.org/10.3389/fmolb.2021.706949>, PMID: 34540893
- Yamada M**, Jitkrittum W, Sigal L, Xing EP, Sugiyama M. 2014. Riken-Aip/pyHSICLasso. 1.4.2. Github. <https://github.com/riken-aip/pyHSICLasso>
- Yang J**, Lin P, Yang M, Liu W, Fu X, Liu D, Tao L, Huo Y, Zhang J, Hua R, Zhang Z, Li Y, Wang L, Xue J, Li H, Sun Y. 2021. Integrated genomic and transcriptomic analysis reveals unique characteristics of hepatic metastases and pro-metastatic role of complement C1q in pancreatic ductal adenocarcinoma. *Genome Biology* **22**:4. DOI: <https://doi.org/10.1186/s13059-020-02222-w>
- Yin H**, Huang YH, Best SA, Sutherland KD, Craik DJ, Wang CK. 2021. An integrated molecular grafting approach for the design of keap1-targeted peptide inhibitors. *ACS Chemical Biology* **16**:1276–1287. DOI: <https://doi.org/10.1021/acscchembio.1c00388>, PMID: 34152716
- Yuan AE**, Shou W. 2022. Data-Driven causal analysis of observational biological time series. *eLife* **11**:e72518. DOI: <https://doi.org/10.7554/eLife.72518>, PMID: 35983746
- Yun UJ**, Park SE, Shin DY. 2011. P41-arc, a regulatory subunit of Arp2/3 complex, can induce premature senescence in the absence of p53 and Rb. *Experimental & Molecular Medicine* **43**:389–392. DOI: <https://doi.org/10.3858/emmm.2011.43.7.042>, PMID: 21628992
- Zeiner PS**, Preusse C, Blank A-E, Zachskorn C, Baumgarten P, Caspary L, Braczynski AK, Weissenberger J, Bratzke H, Reiß S, Pennartz S, Winkelmann R, Senft C, Plate KH, Wischhusen J, Stenzel W, Harter PN, Mittelbronn M. 2015. Mif receptor CD74 is restricted to microglia/macrophages, associated with a M1-polarized immune milieu and prolonged patient survival in gliomas. *Brain Pathology* **25**:491–504. DOI: <https://doi.org/10.1111/bpa.12194>, PMID: 25175718
- Zhang K**, Peters J. 2011. Kernel-based conditional independence test and application in causal discovery. *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*. 804–813.
- Zhang L**, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, Gao R, Kang B, Zhang Q, Huang JY, Konno H, Guo X, Ye Y, Gao S, Wang S, Hu X, Ren X, Shen Z, Ouyang W, Zhang Z. 2018. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**:268–272. DOI: <https://doi.org/10.1038/s41586-018-0694-x>, PMID: 30479382
- Zhang Y**, Chen H, Mo H, Hu X, Gao R, Zhao Y, Liu B, Niu L, Sun X, Yu X, Wang Y, Chang Q, Gong T, Guan X, Hu T, Qian T, Xu B, Ma F, Zhang Z, Liu Z. 2021a. Single-Cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. *Cancer Cell* **39**:1578–1593. DOI: <https://doi.org/10.1016/j.ccell.2021.09.010>
- Zhang H**, Weyand CM, Goronzy JJ. 2021b. Hallmarks of the aging T-cell system. *The FEBS Journal* **288**:7123–7142. DOI: <https://doi.org/10.1111/febs.15770>, PMID: 33590946
- Zheng C**, Zheng L, Yoo J-K, Guo H, Zhang Y, Guo X, Kang B, Hu R, Huang JY, Zhang Q, Liu Z, Dong M, Hu X, Ouyang W, Peng J, Zhang Z. 2017. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* **169**:1342–1356. DOI: <https://doi.org/10.1016/j.cell.2017.05.035>, PMID: 28622514
- Zheng X**, Aragam B, Ravikumar P, Xing EP. 2018. DAGs with NO TEARS: Continuous Optimization for Structure Learning NeurIPS.
- Zheng Y**, Liu X, Le W, Xie L, Li H, Wen W, Wang S, Ma S, Huang Z, Ye J, Shi W, Ye Y, Liu Z, Song M, Zhang W, Han J-DJ, Belmonte JCI, Xiao C, Qu J, Wang H, et al. 2020. A human circulating immune cell landscape in aging and COVID-19. *Protein & Cell* **11**:740–770. DOI: <https://doi.org/10.1007/s13238-020-00762-2>, PMID: 32780218
- Zheng L**, Qin S, Si W, Wang A, Xing B, Gao R, Ren X, Wang L, Wu X, Zhang J, Wu N, Zhang N, Zheng H, Ouyang H, Chen K, Bu Z, Hu X, Ji J, Zhang Z. 2021. Pan-Cancer single-cell landscape of tumor-infiltrating T cells. *Science* **374**:abe6474. DOI: <https://doi.org/10.1126/science.abe6474>, PMID: 34914499

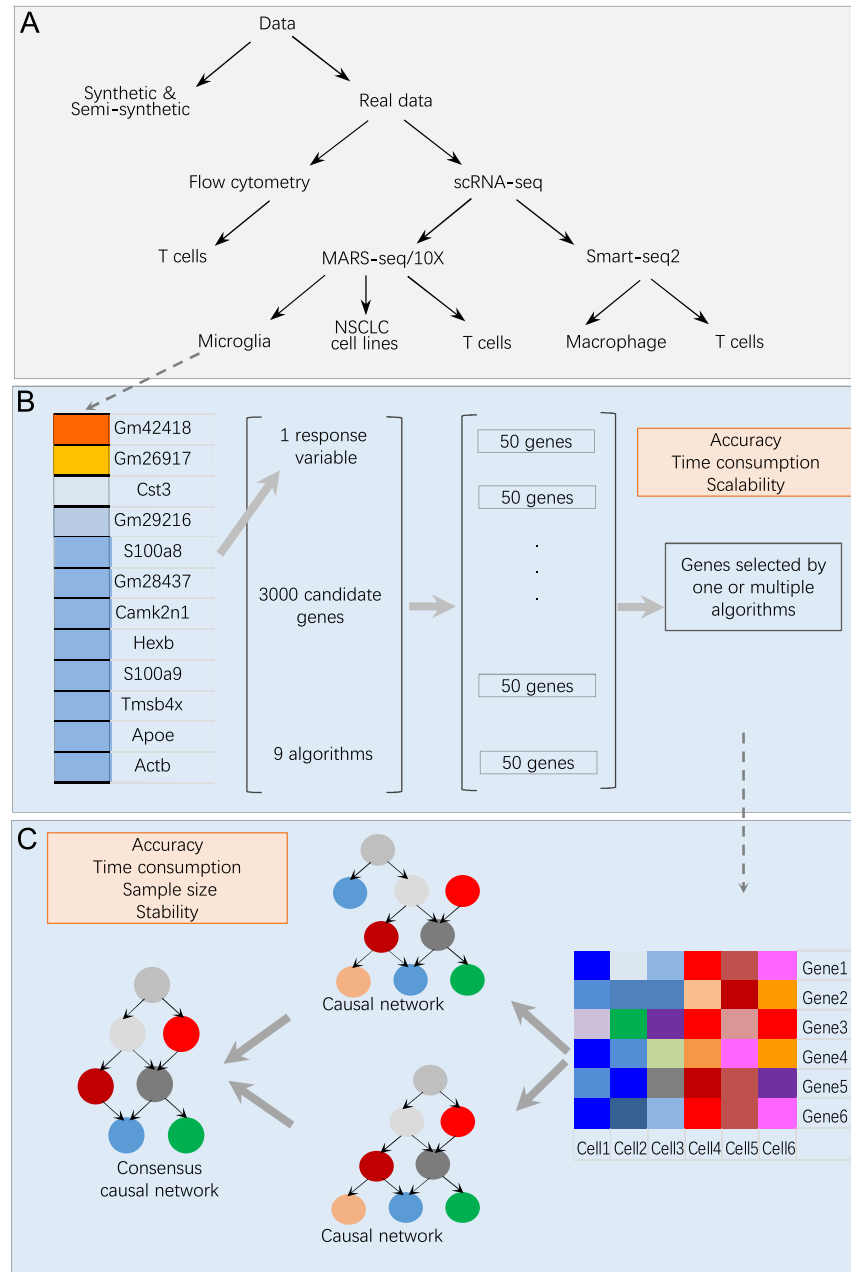
Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B* **67**:301–320. DOI: <https://doi.org/10.1111/j.1467-9868.2005.00503.x>

Appendix 1

Overview of data and algorithms

1. Algorithms and datasets

We combined feature selection and causal discovery to infer causal interactions among a set of gene products in single cells. We used synthetic data, semi-synthetic data, real scRNA-seq data, and flow cytometry data to benchmark nine feature selection algorithms and nine causal discovery algorithms (**Appendix 1—tables 1 and 2; Appendix 1—figure 1**).



Appendix 1—figure 1. Overview of data and benchmarking. **(A)** The single-cell RNA-sequencing (scRNA-seq) data were generated by different protocols and from different cell types (**Appendix 1—table 1**). **(B)** Illustration of feature selection benchmarking using data of microglia from humans and mice. Steps: (i) choose a target gene from a list of microglia biomarkers; (ii) let each algorithm select 50 genes from 3000 candidates expressed in most cells; (iii) merge the nine sets of feature genes into a superset; (iv) compare each selected feature gene set with the superset. **(C)** Illustration of causal discovery benchmarking using a set of feature genes. Steps: (i) use nine *Appendix 1—figure 1 continued on next page*

Appendix 1—figure 1 continued

algorithms (PC+CI tests) to generate nine causal networks; (ii) generate a type 1 consensus network upon the networks of multiple algorithms (a type 2 consensus network is generated upon running an algorithm multiple times); (iii) compare each causal network with the consensus network.

Appendix 1—table 1. Real single-cell RNA-sequencing (scRNA-seq) data.

Dataset	Cell type	Species	Protocols	Dataset URL	Database and Identifier	References
1	Microglia from humans and mice	Human and mouse	MARS-seq	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134705	NCBI Gene Expression Omnibus, GSE134705	Geirsdottir et al., 2019
2	Five lung cancer cell lines (A549, H1975, H2228, H838, HCC827) from the CellBench benchmarking dataset	Human	10x Genomics	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126906	NCBI Gene Expression Omnibus, GSE126906	Tian et al., 2019
3	Lung alveolar epithelial cells	Human	10x Genomics	https://www.synapse.org/#!Synapse:syn21041850	Synapase, syn21041850	Travaglini et al., 2020
4	Six types of CD4 T cells (naïve, TEM, rTregs, naïve_lsg15, cytotoxic, exhausted) from young and old mice	Mouse	10x Genomics	https://singlecell.broadinstitute.org/single_cell/study/SCP490/aging-promotes-reorganization-of-the-cd4-t-cell-landscape-toward-extreme-regulatory-and-effector-phenotypes	Single Cell Portal, SCP490	Elyahu et al., 2019
5	Macrophages isolated from glioblastomas	Human	Smart-seq2	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131928	NCBI Gene Expression Omnibus, GSE131928	Neftel et al., 2019
6	Exhausted CD8 T cells isolated from liver cancer, lung cancer, and CRC	Human	Smart-seq2	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99254 , https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108989 , https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98638 .	NCBI Gene Expression Omnibus, GSE99254. NCBI Gene Expression Omnibus, GSE108989. NCBI Gene Expression Omnibus, GSE98638.	Guo et al., 2018; Zhang et al., 2018; Zheng et al., 2017
7	Non-exhausted CD8 T cells isolated from the normal liver, lung, and colorectal tissues	Human	Smart-seq2	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99254 , https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108989 , https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98638 .	NCBI Gene Expression Omnibus, GSE99254. NCBI Gene Expression Omnibus, GSE108989. NCBI Gene Expression Omnibus, GSE98638.	Guo et al., 2018; Zhang et al., 2018; Zheng et al., 2017
8	CD4 T cell	Human	Flow cytometry	https://www.science.org/doi/10.1126/science.1105809	Science Supplementary Materials, doi: 10.1126/science.1105809	Sachs et al., 2005

Appendix 1—table 2. Feature selection and causal discovery algorithms.

Feature selection	Category	Causal discovery	Category
Random forests		GaussCItest	Test for CI between Gaussian random variables upon partial correlation
Extremely randomized trees		DCC.perm	
XGBoost	Ensemble learning-based	DCC.gamma	Test for CI using a distance covariance-based kernel
SHS		HSIC.perm	
BAHSIC		HSIC.gamma	
Block HSIC Lasso	HSIC-based	HSIC.clust	Test for CI using a HSIC-based kernel

Appendix 1—table 2 Continued on next page

Appendix 1—table 2 Continued

Feature selection	Category	Causal discovery	Category
Lasso		RCIT	Test for CI using an approximate KCIT kernel
Ridge regression		RCoT	
Elastic net	Regularization-based	CMIknn	Test for CI based on conditional mutual information

2. Synthetic data for feature selection

Fully synthetic dataset

N variables (indicating candidate genes) without a specific pattern were generated randomly from a $(0, 2)$ uniform distribution, from which n variables (indicating feature genes) were selected randomly to synthesize a response variable. Each feature influences the response variable depending randomly on one of the five functions: $y=x^2$, $y=\sin(x)$, $y=\cos(x)$, $y=\tanh(x)$, and $y=e^x$.

By combining different numbers of feature genes and candidate genes (4–50, 8–100, 8–200, 20–500, 20–1000, and 50–2000) and generating samples of different sizes (100, 200, 500, 1000, and 2000), we generated 30 schemes. For each algorithm, we ran each scheme 10 times, and each time the true positive rate (TPR) was calculated by:

$$TPR = \frac{\text{selected features} \cap \text{true features}}{\text{true features}}.$$

A TPR of 1.0 means that the feature selection algorithm completely correctly selects the features; a small TPR indicates poor performance.

Semi-synthetic dataset

First, we extracted genes from a benchmark scRNA-seq dataset (https://support.10xgenomics.com/single-cell-gene-expression/datasets/4.0.0/Parent_NGSC3_DI_PBMC), sorted these genes based on the cells in which they were expressed (expression level >0), and obtained the top 5000 genes expressed in most cells. Then, we obtained the top 5000 cells that contained the most expressed genes. These genes and cells formed a 5000*5000 matrix. Different candidate gene sets were sampled from this matrix, and different feature gene sets were selected randomly from each candidate gene set. Next, the response variables (target genes) were synthesized using feature genes.

3. Synthetic data for causal discovery

We used the *randomDAG* function in the *pcalg* package (<https://cran.r-project.org/web/packages/pcalg/index.html>) to generate DAGs with random topologies. Values of nodes (i.e. genes) in these DAGs were randomly generated using the following 10 functions that determined relationships between nodes:

$$y = x^2, y = \sqrt{\text{abs}(x)}, y = \sin(x) * \sin(x), y = \sin(x), y = \cos(x),$$

$$y = \cos(x) * \cos(x), y = \tanh(x), y = e^{\log_2(\text{abs}(x))}, y = \log(|x| + 1), y = \tanh(x) * \tanh(x)$$

With the variable number ranging from 20 to 80 (step = 20), and the sample size ranging from 500 to 1000 (step = 500), we generated eight datasets with known networks.

4. Real single-cell data for feature selection and causal discovery

Single-cell datasets in **Appendix 1—table 1** were used for benchmarking.

Appendix 2

Feature selection algorithms and benchmarking

1. Ensemble learning-based algorithms

Random forests

We used the *RandomForestRegressor* function (with default parameters) in the *sklearn* package (<https://scikit-learn.org/stable/>) to build random forest models (**Breiman, 2001**). Each model contained 200 decision trees. After regression based on the response variable(s), genes were sorted based on *Gini importance*, and the top genes were selected as feature genes.

Extremely randomized trees

We used the *ExtraTreesRegressor* function (with default parameters) in the *sklearn* package (<https://scikit-learn.org/stable/>) to generate extremely randomized trees (**Geurts et al., 2006**). Each tree model contained 200 decision trees. After regression based on the response variable(s), genes were sorted based on *Gini importance*, and the top genes were selected as feature genes.

XGBoost

We used the *XGBRegressor* function (with default parameters) in the *Scikit-Learn API* (https://xgboost.readthedocs.io/en/latest/python/python_api.html) to build the XGBoost models (**Chen and Guestrin, 2016**). Each XGBoost model contained 200 decision trees. After regression based on the response variable(s), genes were sorted based on *Gini importance*, and the top genes were selected as feature genes.

2. Regularization-based algorithms

Lasso

We used the *Lasso* function (with default parameters) in the *sklearn* package (<https://scikit-learn.org/stable/>) to produce the regression models. In the Lasso (least absolute shrinkage and selection operator) regression equation (**Tibshirani, 1997**):

$$\beta^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2N} \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\},$$

N is the number of samples, p is the number of features, β_j is the coefficient of the j th feature, and λ (by default $\lambda=0.5$) is a penalty coefficient controlling the shrinkage. Feature genes were selected based on the value of $|\beta_j|$, which indicates the importance of the j th feature for the response variable(s).

Ridge regression

We used the *Ridge* function (with default parameters) in the *sklearn* package (<https://scikit-learn.org/stable/>) to build Ridge regression models. The equation of Ridge regression is similar to that of Lasso (**Hoerl and Kennard, 2000**):

$$\beta^{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2N} \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\},$$

but the L2 penalty term is $\sum_{j=1}^p \beta_j^2$. Feature genes were selected based on the value of $|\beta_j|$, which indicates the importance of the j th feature for the response variable(s).

Elastic net

We used the *ElasticNet* function (with default parameters) in the *sklearn* package (<https://scikit-learn.org/stable/>) to build elastic net models. Elastic net linearly combines the L1 and L2 penalties of the Lasso and Ridge methods using the following equation (by default $\lambda=1$ and $\alpha=0.5$) (**Zou and Hastie, 2005**). In the equation:

$$\beta^{elastic\ net} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2N} \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \left(\frac{1-\alpha}{2} \sum_{j=1}^p \beta_j^2 + \alpha \sum_{j=1}^p |\beta_j| \right) \right\},$$

feature genes are selected upon the value of $|\beta_j|$, which indicates the importance of the j th feature for the response variable(s).

3. HSIC-based algorithms

BAHSIC

Hilbert-Schmidt independence criterion (HSIC) is a measure of dependency between two variables (**Gretton et al., 2005**). After obtaining a measure between a response variable and a feature, a backward elimination process is used to extract a subset of features that are most relevant to the response variable (**Song et al., 2007**). We used the BAHSIC program (<https://www.cc.gatech.edu/~lsong/code.html>), together with the nonlinear radial basis function kernel, to evaluate the dependency between feature genes and response variable(s), and set the parameter $flg3 = 1 - \text{float} \left(\frac{\text{desired feature number}}{\text{total feature number}} \right)$ to accelerate computation.

SHS

Sparse HSIC (SHS), which combines HSIC with fast sparse decomposition of matrices, is an HSIC-based feature selection algorithm without the backward elimination process to identify a sparse projection of all features (**Gangeh et al., 2017**). We translated the SHS program encoded in MATLAB (<https://uwaterloo.ca/data-science/sites/ca.data-science/files/uploads/files/shs.zip>) into a Python program and used eigenvalue decomposition as the matrix halving procedure. The parameters $\gamma=1.1$ (a penalty parameter that controls the sparsity of the solution) and $\rho=0.1$.

Block HSIC Lasso

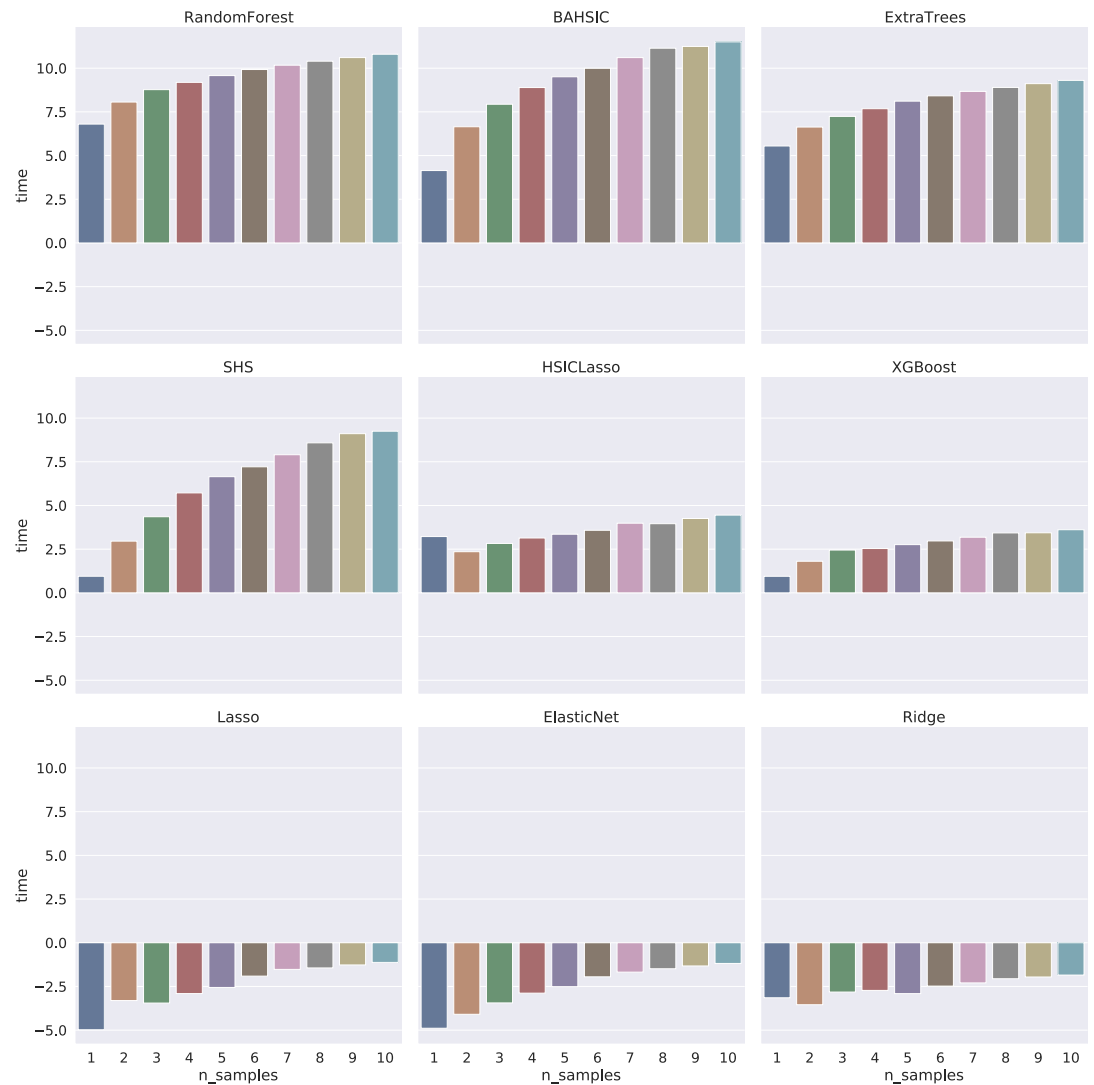
HSIC Lasso is a variant of the minimum redundancy maximum relevance feature selection algorithm and is suitable for high-dimensional small sample data. We used the pyHSICLasso program (<https://github.com/riken-aijp/pyHSICLasso>; **Yamada et al., 2014**), which is an approximation of HSIC Lasso but reduces memory usage dramatically while retaining the properties of HSIC Lasso (**Climente-González et al., 2019**). We used the function `get_index_score()` to compute feature importance and the function `get_features()` to return top feature genes.

4. Benchmarking results

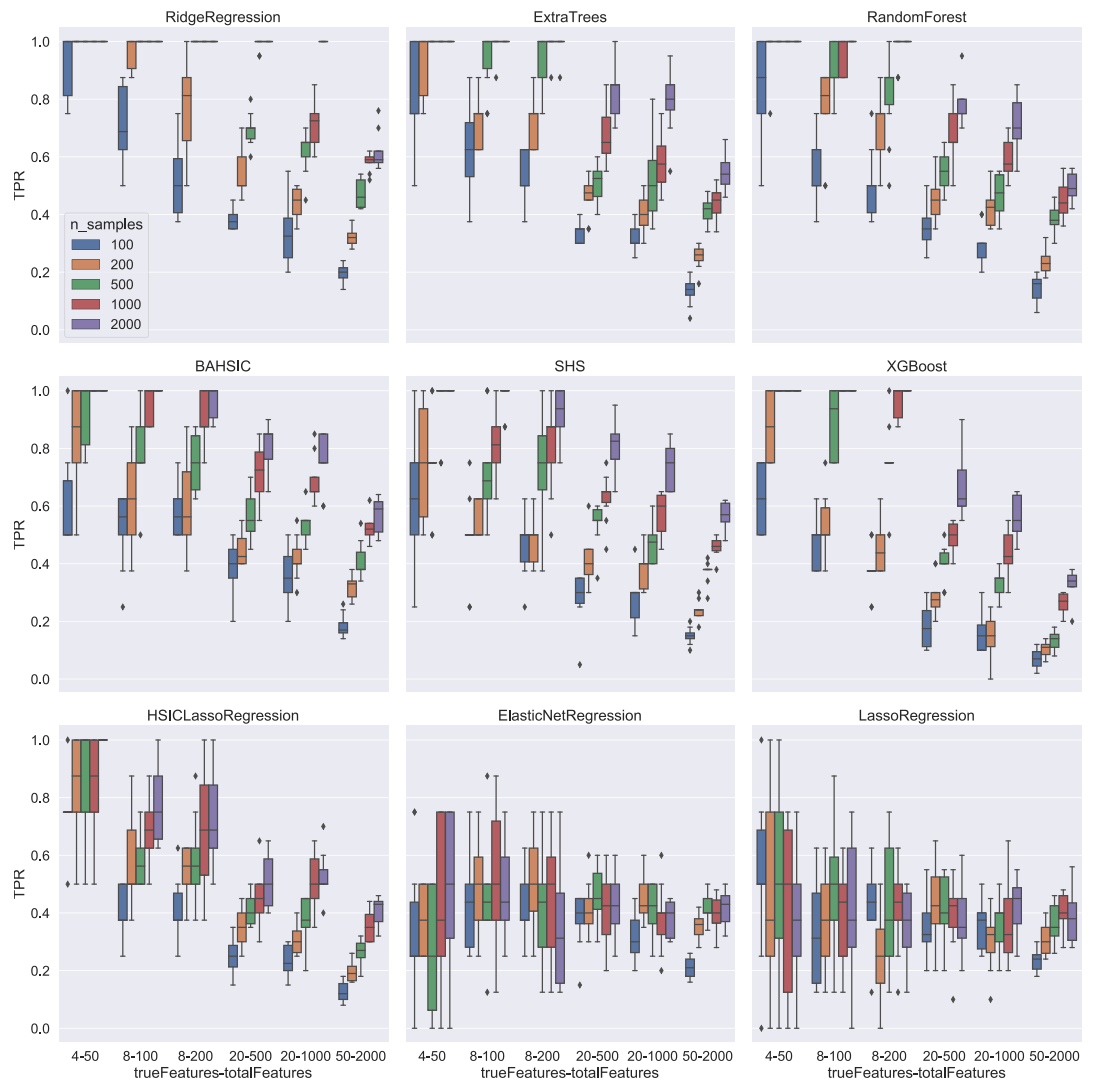
We evaluated the time consumption, accuracy, and scalability of nine feature selection algorithms in three categories (**Appendix 1—tables 1 and 2**). First, tested using synthesized data, all algorithms showed moderate time consumption, which increased insignificantly when the sample size increased (**Appendix 2—figure 1**). Second, using synthesized data, multiple algorithms selected all features correctly if schemes were simple (e.g. selecting 4 features from 50 candidates). If features and/or candidates increased (e.g. selecting 50 features from 2000 candidates), BAHSIC showed the best performance, with accuracy decreasing more slowly than others (**Appendix 2—figure 2**). Third, using well-known microglial biomarkers in humans and mice (**Butovsky et al., 2014; Patir et al., 2019**) and using scRNA-seq data of microglia from the human and mouse brain (**Geirsdottir et al., 2019**), we further evaluated feature selection algorithms' accuracy. We merged feature genes generated by the nine algorithms into a superset (**Appendix 1—figure 1B**), identified a subset generated by the majority of algorithms, and examined how many feature genes of each algorithm overlap with the subset. When selecting 50 genes from 3000 candidates upon a target gene (e.g. Hexb), BAHSIC and SHS were the best and second-best algorithms, and they also selected most microglia biomarkers (**Appendix 2—figures 3 and 4**). Fourth, we used real scRNA-seq data in applications to evaluate algorithms' accuracy and found that BAHSIC also performs well. Finally, to evaluate algorithms' scalability, we let algorithms select feature genes from different numbers of candidate genes. When the number of candidate genes is large (>10,000), the accuracy of feature selection is somewhat decreased.

BAHSIC's performance was examined further using macrophages isolated from human glioblastoma by checking whether the selected feature genes accurately characterize macrophages (**Neftel et al., 2019**). We used six macrophage biomarkers (CD14, AIF1, FCER1G, FCGR3A, TYROBP, and CSF1) exclusively expressed in these macrophages as the target genes (response variables) and used BAHSIC to select 50 feature genes from 3000 candidate genes expressed in >50% macrophages

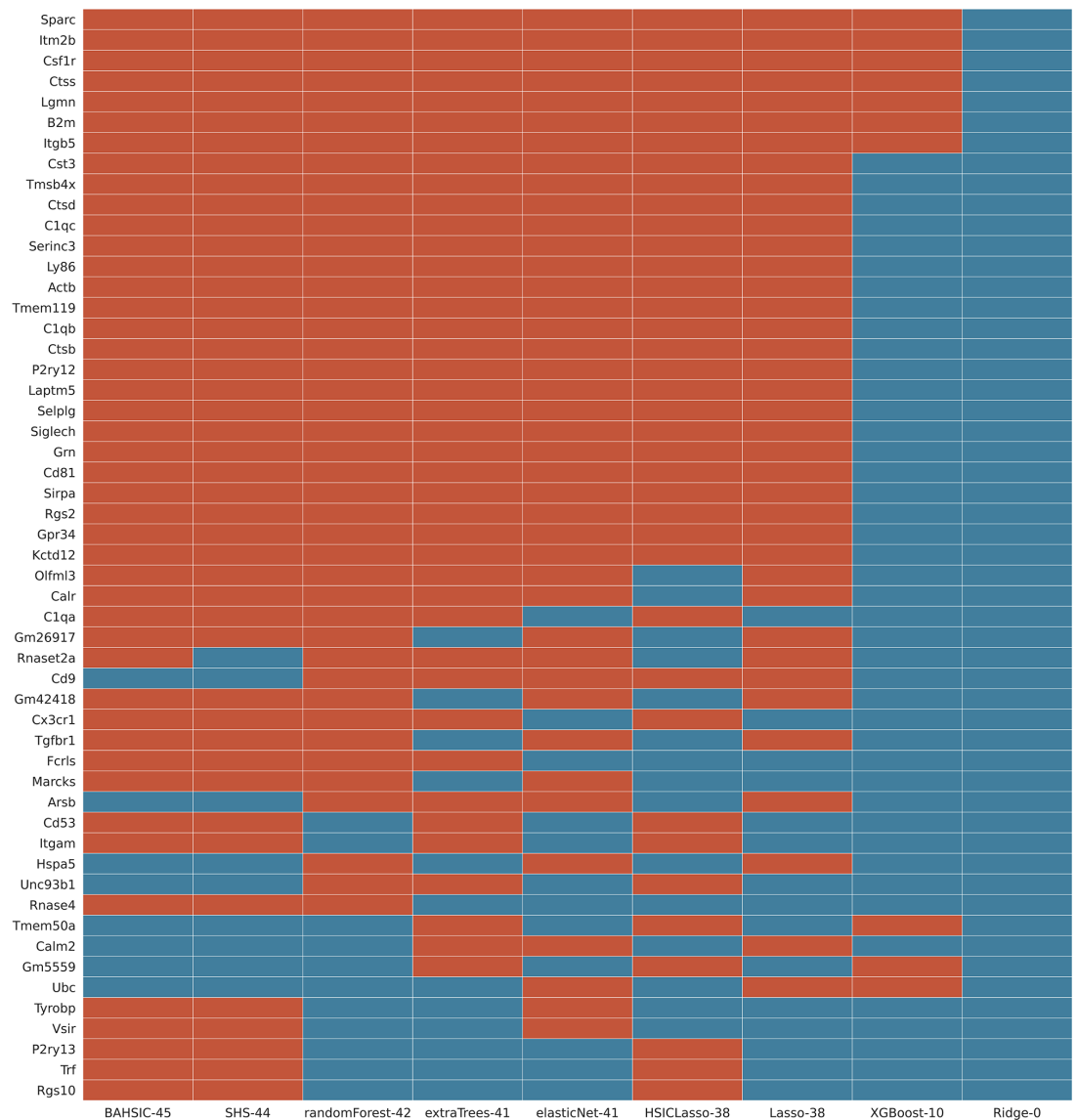
(Appendix 2—figure 5). Nearly all feature genes were expressed exclusively in the macrophages (Appendix 2—figure 6). Experimental findings support many feature genes. C1QA/B/C and C3 are supported by the finding that C1Q is produced and the complement cascade is up-regulated in cancer-infiltrated macrophages (Yang et al., 2021). CD74 and MHC-II genes are supported by the finding that CD74 is correlated with malignancies and the immune microenvironment in gliomas (Xu et al., 2021). TREM2 and APOE are supported by the finding that highly expressed TREM2 and APOC2 in macrophages contribute to immune checkpoint therapy resistance (Xiong et al., 2020). MS4A4A and MS4A6A are supported by the finding that APOE and TREM2 are up-regulated by MS4A (Deming et al., 2019). In contrast, feature genes selected by RidgeRegression upon the same six biomarkers were expressed in diverse cells (Appendix 2—figure 7). These confirm that BAHSIC can quite reliably select feature genes upon target genes.



Appendix 2—figure 1. Time consumption of feature selection algorithms increased mildly when the sample size increased from 1000 to 10,000 (the X-axis indicates the sample size; 1–10 indicate 1000–10,000, respectively). The Y-axis indicates time (s) in the log2 form. The log2 form makes the time consumption of Lasso, ElasticNet, and Ridge have negative values (<1 s).



Appendix 2—figure 2. Feature selection results using synthetic data. Colors indicate different sample sizes (see the inset in the top-left panel). The X-axis (depicted under the bottom panels) indicates different schemes. For example, 4–50 means that there are 50 candidates (i.e. total features), 4 of which are chosen randomly to generate the response variable, and feature selection should select the 4 features (i.e. true features) from the 50 candidates upon the response variable. The Y-axis indicates true positive rate (TPR). For simple schemes (e.g. 4–50), some algorithms reached a TPR of 1.0. For complex schemes (e.g. 50–2000, selecting 50 features from 2000 candidates upon the response variable), some algorithms (especially BAHSIC) reached a TPR of 0.6.



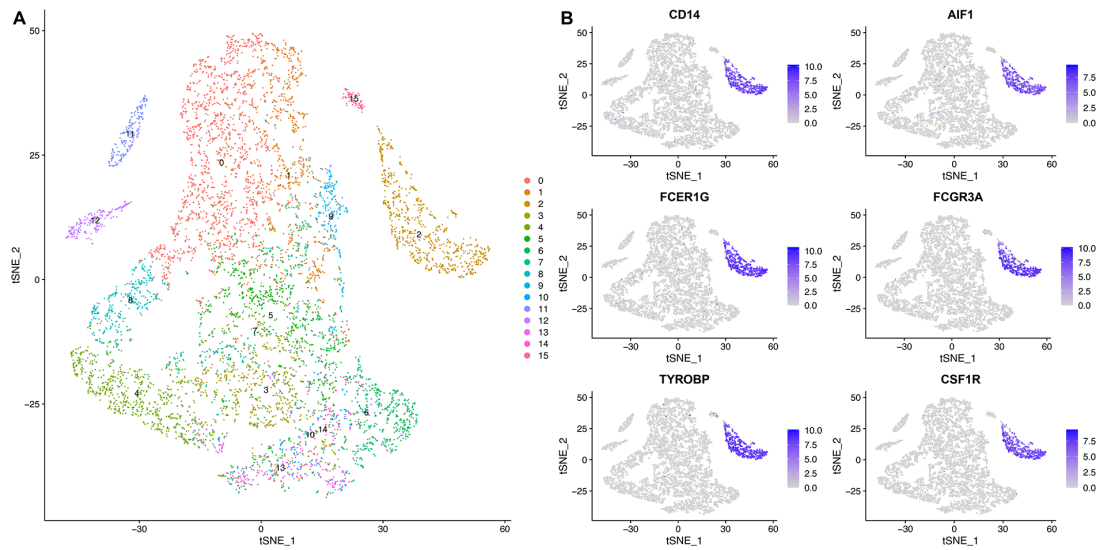
Appendix 2—figure 3. This screenshot showed a feature selection result (the superset of feature genes selected by ≥ 3 algorithms from the mouse microglia dataset) when all 9 algorithms were used. The *Hexb* gene was the target gene, and each algorithm selected 50 feature genes from the 3000 candidate genes expressed in most cells. BAHSIC, SHS, RandomForest, ExtraTrees, ElasticNet selected highly overlapping feature genes, many of which are microglia biomarkers in mice (**Appendix 2—figure 4**). The numbers right side of algorithm names indicate genes overlapping with the superset.

Human (CD74) (1)	Human (ACTB) (33)	Mouse (ACTB) (7)	Mouse (Gm42418) (1)	Mouse (Hexb) (4)
A2M	A2M	C1qa	Abce1	Actb
AC011481.3	ACTBP11	C1qb	AC124744.2	Arsb
AC025857.2	ACTG1P21	C1qc	Actb	B2m
ACTB	ATP5J2	Calm1	Aldoa	C1qa
ACTG1	B2M	Calr	Apoe	C1qb
APOE	BECN1	Cd81	Cnpy2	C1qc
B2M	C1QB	Cf1	Coro1a	Calm2
BOD1L1	C3	Csf1r	Cryl1	Calr
C1QA	CALM1	Cst3	Csf1r	Cd53
C1QB	CALM2	Ctsb	Ctsd	Cd81
C3	CCL3	Ctsd	Cx3cr1	Cd9
C3AR1	CD74	Ctss	Cyth4	Csf1r
CCL2	CD81	Cx3cr1	Fcgr3	Cst3
CCL4	CSF1R	Eef1a1	Gm13339	Ctsb
CD14	CST3	Epb41l2	Gm15725	Ctsd
CD81	CX3CR1	Fam102b	Gm21198	Ctss
CH25H	CYBB	Fcer1g	Gm22213	Cx3cr1
CSF1R	CYFIP1	Glul	Gm26917	Fcer1g
CSGALNACT1	DUSP7	Gm11942	Gm29047	Fcrls
CST3	FBXL18	Gm15725	Hbb-bs	Gm15344
DENND3	FCER1G	Gm26917	Hexb	Gm26917
FTL	FCGR3A	Gm42418	Kpnb1	Gm29216
FYB1	FTH1	Golm1	Laptm5	Gm42418
GLIPR1	FTL	Hexb	Lpin2	Gpr34
HECTD1	HLA-DRA	Itgb5	Lrrc3	Grn
HLA-B	IFNAR1	Itm2b	Malat1	Hspa5
HLA-C	ITM2B	Laptm5	Marcks	Itgam
HLA-DMB	MARCKS	Lcp1	Mcrip1	Itgb5
HLA-DPA1	MIR3064	Ly86	Mir6236	Itm2b
HLA-DPB1	PFN1	Man2b1	Nek7	Kctd12
HLA-DQA1	PRDX4	Marcks	Olfml3	Laptm5
HLA-DQB1	SNORA70	Olfml3	P2ry12	Lgmn
HLA-DRA	SPP1	P2ry12	Pmepa1	Ly86
HLA-DRB1	SRGN	Rassf2	Reep5	Marcks
HLA-DRB6	SSB	Rgs10	Selenop	Olfml3
HLA-E	TMSB4X	Rnase4	Serinc3	P2ry12
IL1B	TMSB4XP8	Selenop	Sipa1	P2ry13
ITM2B	TUBA1B	Selplg	Ski	Rgs10
LAPTM5	TYROBP	Serinc3	Sparc	Rgs2
MALAT1	VSIG4	Siglech	Stat6	Rnase4
NDUFS5	ZFP36L1	Slco2b1	Syng2	Rnaset2a
NEAT1	ZSWIM6	Sparc	Tmem119	Selplg
PLD4		Tgfbr1	Tmem19	Serinc3
PSAP		Tmem119	Ubc	Siglech
PTGS1		Tmsb4x	Vsir	Sirpa
RNASE6		Tyrobp	Wdr1	Sparc
RNASET2		Ubl3	Zfhx3	Tgfbr1
SLC15A2				Tmem119
SNORA31				Tmsb4x
SORL1				Tyrobp
SPP1				Unc93b1
TBC1D12				Vsir
TMSB4X				
TREM2				
VSIR				

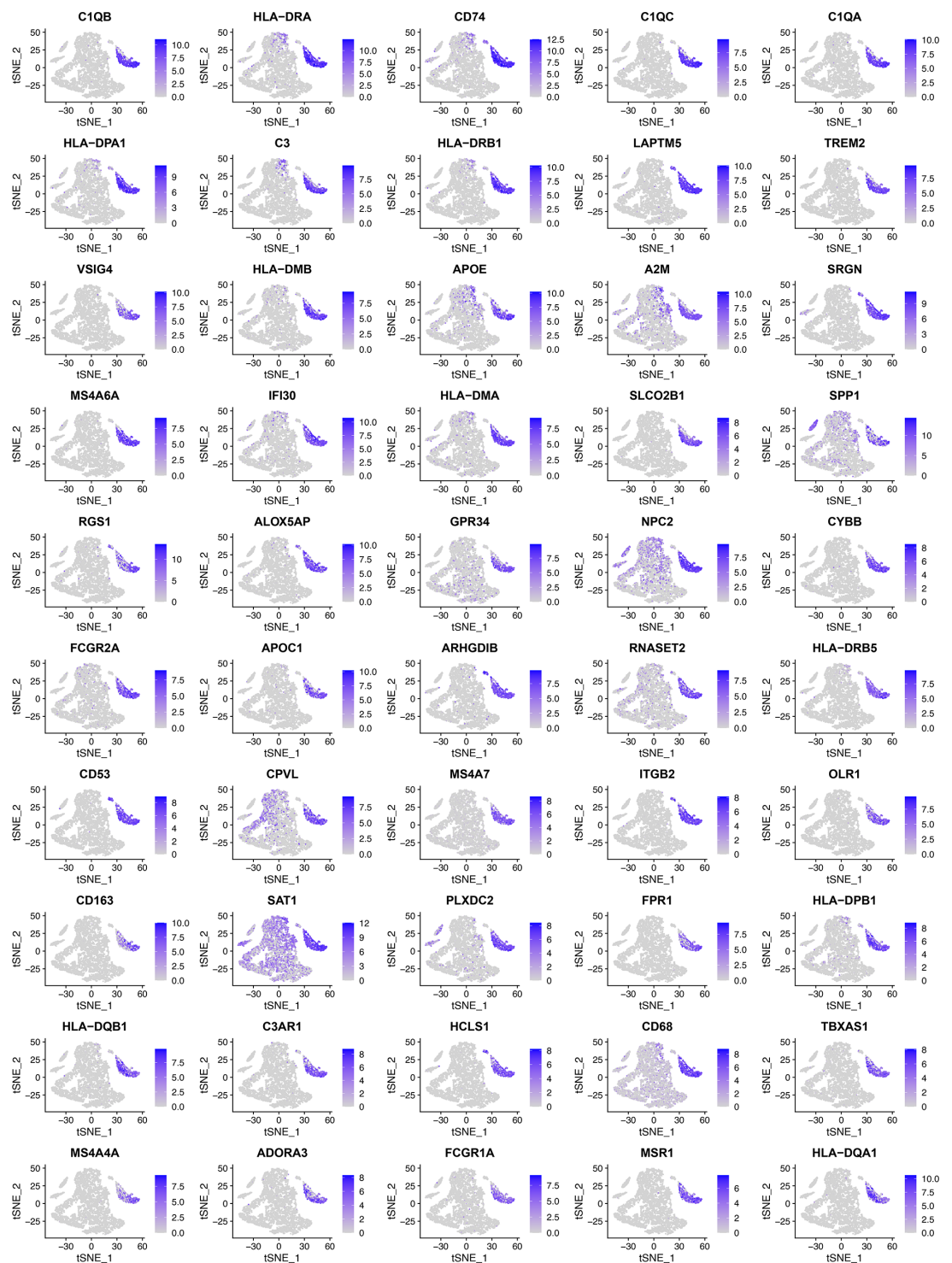
Appendix 2—figure 4. Feature genes selected by ≥ 3 algorithms in microglia from humans and mice using *CD74*, *ACTB*, *ACTB*, *Gm42418*, with *Hexb* as the target gene. The number right side of each target gene indicates the rank of its transcript's variance in the 3000 candidate genes. A large variance indicates that the gene may be important in the examined cells. Many feature genes are microglia biomarkers, indicating that feature genes
Appendix 2—figure 4 continued on next page

Appendix 2—figure 4 continued

selected by ≥ 3 algorithms are biologically rational. Annotated microglia biomarkers in humans and mice are marked in red (Patir et al., 2019).



Appendix 2—figure 5. Cell types and cells that express macrophage biomarker genes. (A) The tSNE plot shows all cells isolated from the human glioblastoma (Nefel et al., 2019). Region 2 (the blue area in (B)) is macrophages. (B) The six macrophage biomarker genes were exclusively expressed in macrophages (the blue area).



Appendix 2—figure 6. These tSNE plots show that nearly all of the 50 feature genes were exclusively expressed in macrophages. These feature genes were selected by BAHSIC using six macrophage biomarkers (CD14, AIF1, FCER1G, FCGR3A, TYROBP, and CSF1R) as the target genes. They include genes involved in macrophage activation (e.g. C1QA, CD74, TREM2) and multiple class II major histocompatibility complex (MHC) genes (e.g. HLA-DMA, HLA-DPA1). The interactions between CD74 and MHC-II genes (CD74 is an MHC class II chaperone) probably contribute to the co-selection of these genes.



Appendix 2—figure 7. When RidgeRegression was used to select 50 feature genes using the same six macrophage markers (CD14, AIF1, FCER1G, FCGR3A, TYROBP, and CSF1R) as target genes, as these tSNE plots show, many feature genes were expressed in diverse cells instead of macrophages.

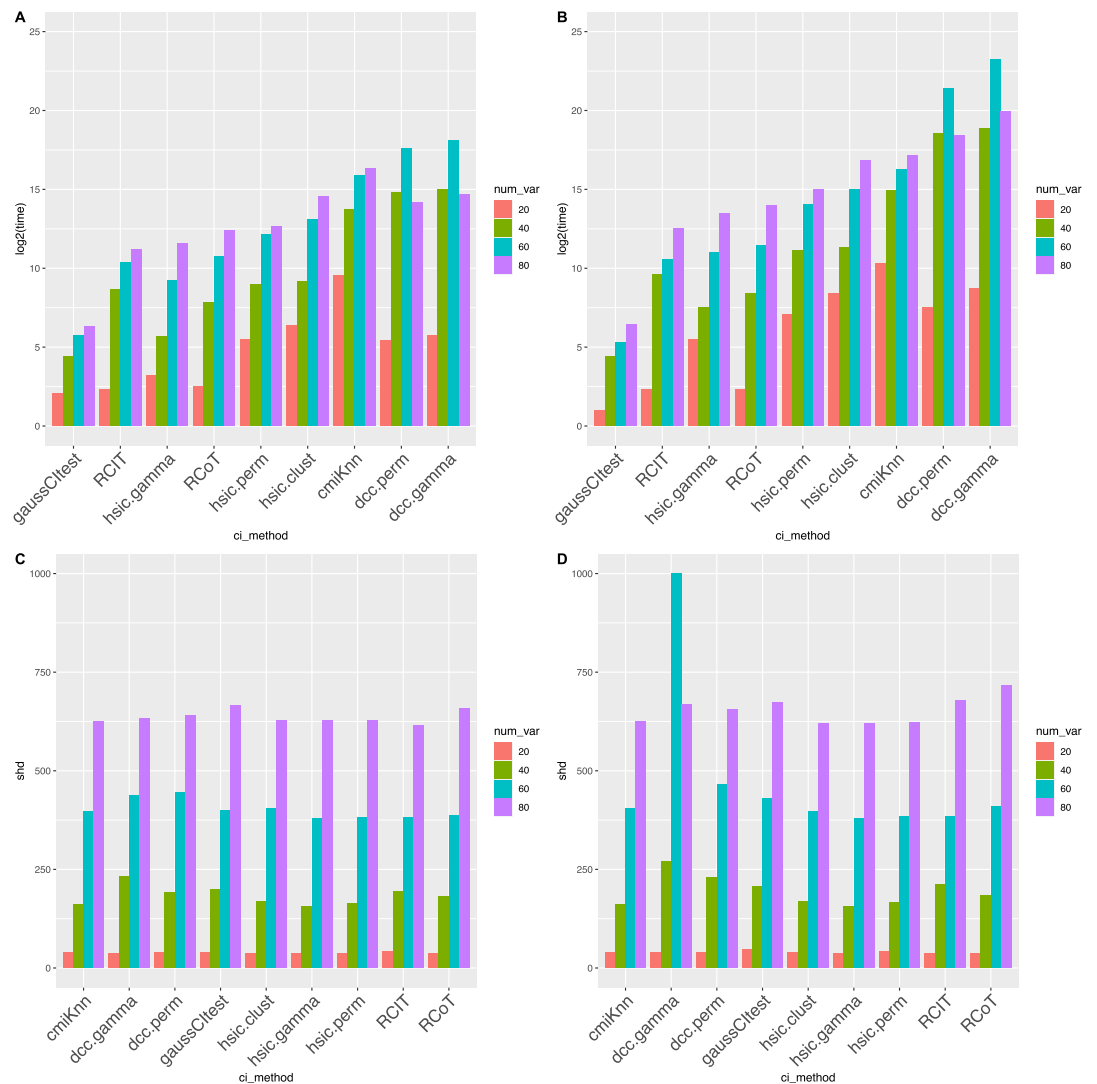
Appendix 3

Causal discovery algorithms and benchmarking

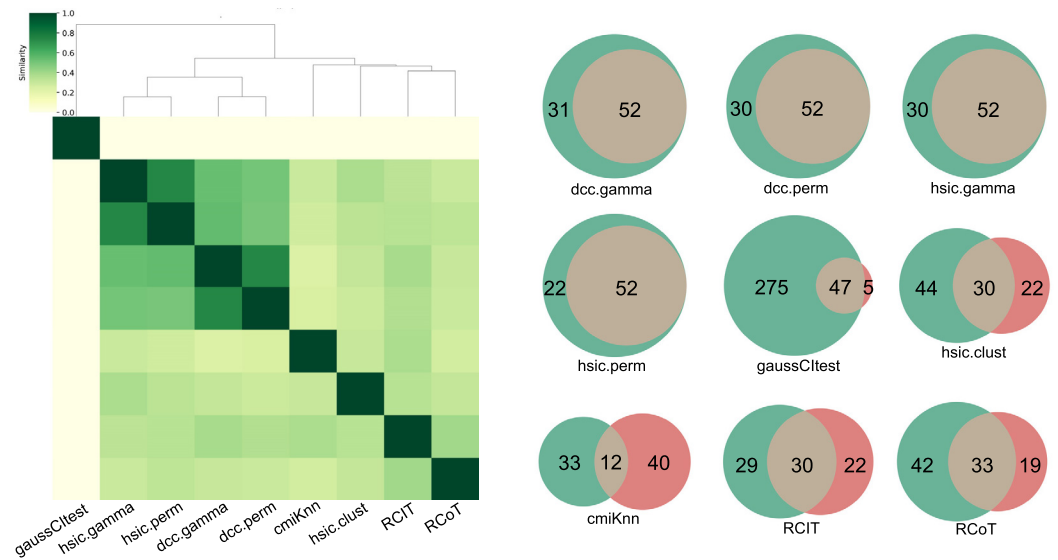
1. Causal discovery methods

CausalCell integrates four causal discovery methods – PC, GES, GSP, and DAGMA-nonlinear – which are representative constraint-based, score-based, hybrid, and continuous optimization-based methods. Constraint-based methods identify causal interactions in a set of variables in two steps: skeleton estimation and orientation. Score-based methods assign a score function (e.g. the Bayesian information criterion) to each potential causal network and optimize the score via greedy approaches. Hybrid methods combine score-based methods and CI tests. Continuous optimization-based methods recast the combinatoric graph search problem as a continuous optimization problem. The PC and GSP algorithms can be combined with different CI tests.

To benchmark the performance of different CI tests, we combined 10 CI tests with the parallel version of the PC algorithm (i.e. the *pc* function in the R package *pcalg*, with the default setting *skel.method*="stable") (Le et al., 2019). The results show that kernel-based CI tests (especially the two DCC CI tests) outperform other CI tests (Appendix 3—table 1; Appendix 3—figures 1 and 2). To evaluate the score-based and hybrid methods GES (<https://cran.r-project.org/web/packages/pcalg/index.html>) and GSP (<https://github.com/uhlerlab/causal DAG>; Chickering, 2003; Solus et al., 2021; Squires, 2018), we compared PC+DCC.gamma, GES, and GSP+DCC.gamma. The results show that PC+DCC.gamma and GSP+DCC.gamma have comparable network accuracy and time consumption, and both are more accurate but more time-consuming than GES (Appendix 3—figures 3–6).

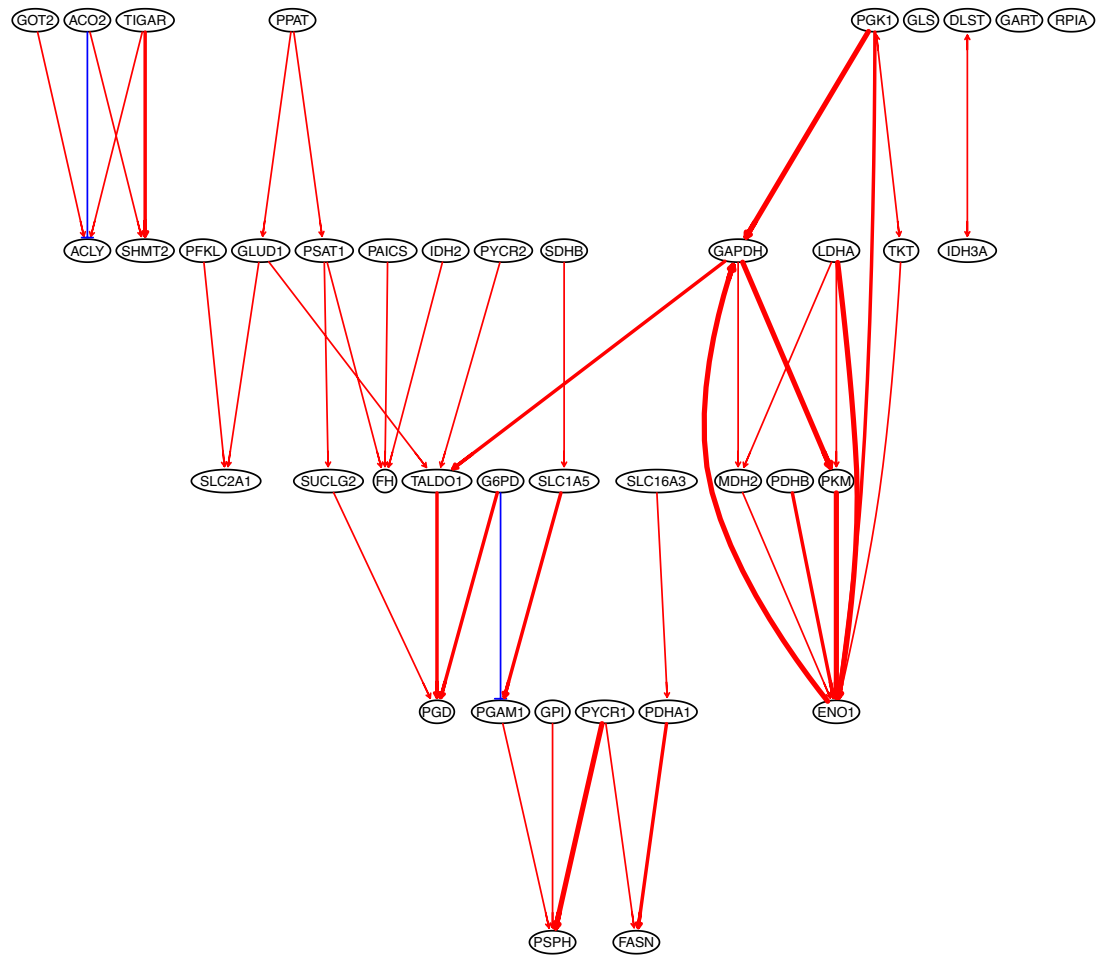


Appendix 3—figure 1. Causal discovery performance that was tested using synthetic data. DCC.perm and HSIC.gamma are the most time-consuming algorithms, but all algorithms have similar accuracy. The sample size is 500 (AB) or 1000 (CD), and the variable number ranges from 20 to 80. (AC) show the time consumption (in second) of different algorithms, and (BD) show structural Hamming distance (SHD) values. '*' in (CD) indicates that the algorithm did not finish running in 6 weeks. These two cases, and the two cases in (A) where DCC.perm and DCC.gamma took more time when there were 60 variables than when there were 80 variables, were anomalies caused by synthetic data. When testing using real scRNA-seq data, no such anomalies occurred.

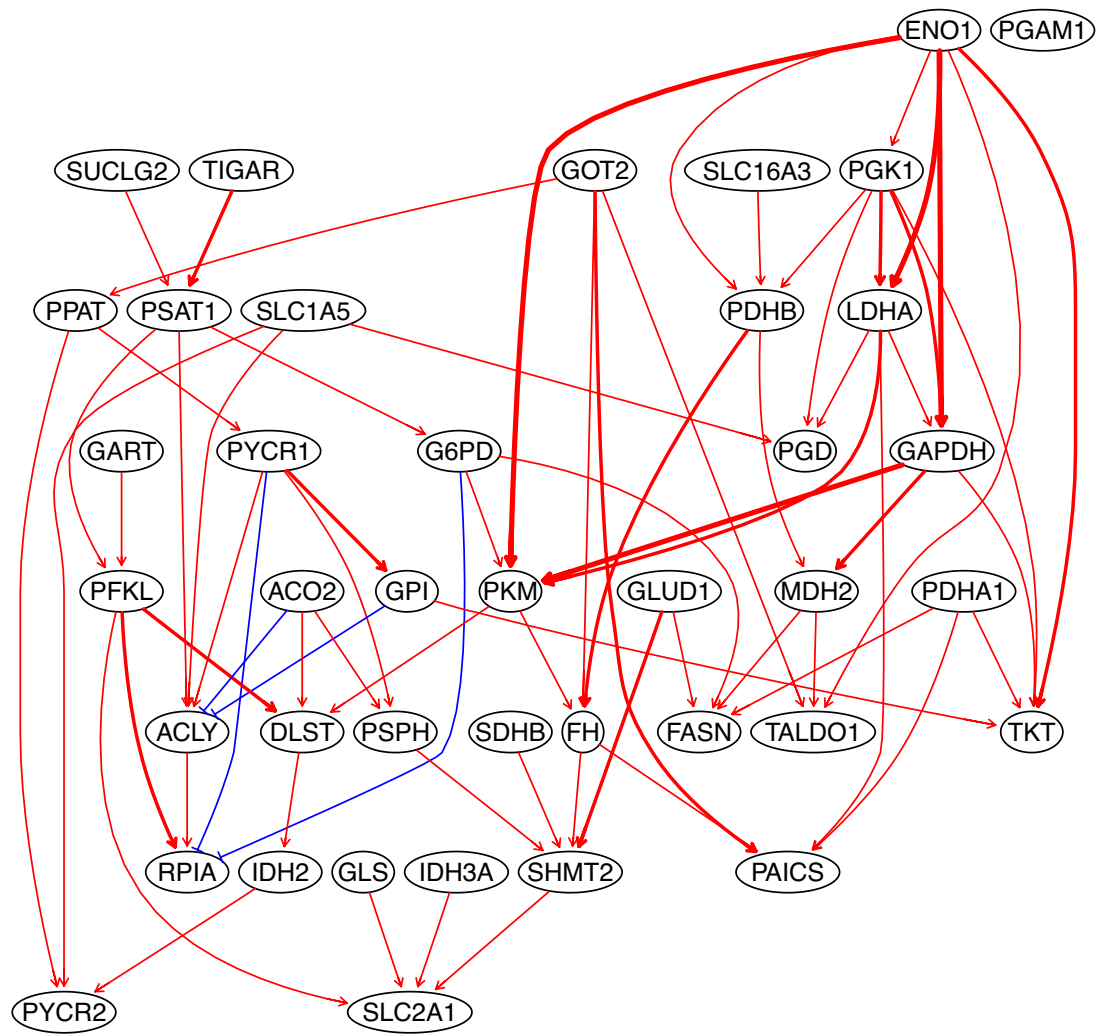
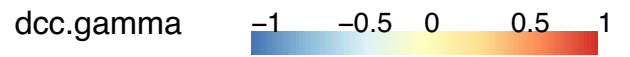


Appendix 3—figure 2. Consistency between each causal network and the consensus network. **(A)** In the cluster map of different CI tests, darker colors indicate a higher similarity of networks. The networks of HSIC.gamma, HSIC.perm, HSIC.gamma, and HSIC.perm have the highest similarity values, thus sharing the most similar structures. We used the four networks to build a consensus network, which was assumed most close to the ground truth. **(B)** For interactions inferred by each algorithm (green circle), we checked how many interactions overlap the interactions in the consensus network (pink circle). The true positive rate (TPR) of DCC.gamma, DCC.perm, HSIC.gamma, HSIC.perm, GaussCltest, HSIC.clust, cmiKnn, RCIT, and RCoT were 62.7%, 63.4%, 63.4%, 70.3%, 14.6%, 40.5%, 26.7%, 50.8%, and 44.0%, respectively, confirming that Hilbert-Schmidt independence criterion (HSIC) and distance covariance criteria (DCC) are better than others and that it is reasonable to use the consensus network generated upon the four algorithms' networks to evaluate algorithms' performance.

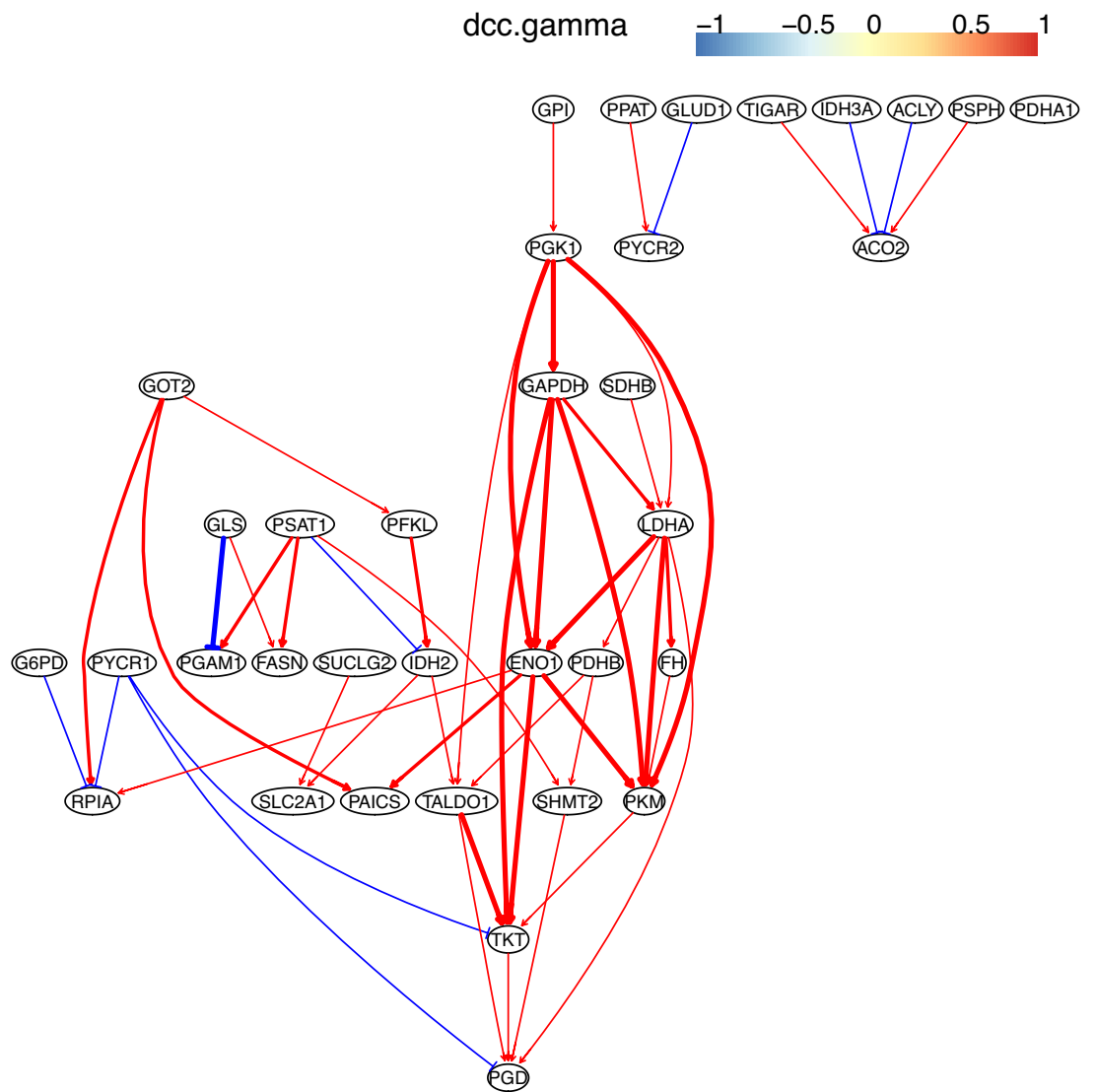
dcc.gamma



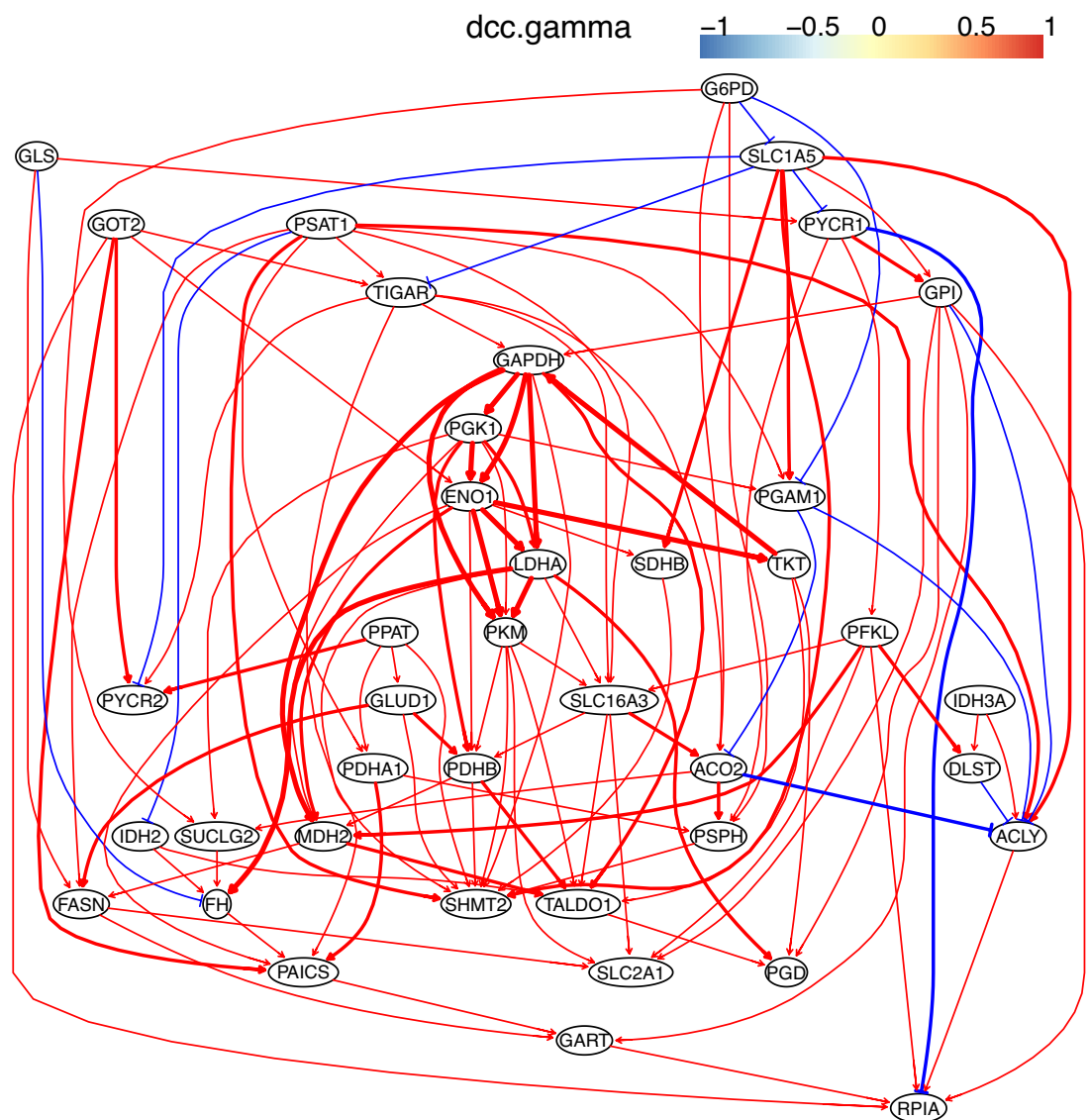
Appendix 3—figure 3. The causal relationships between genes in the WP4290 pathway in the cell line H838 inferred using 200 cells ($\alpha = 0.1$). The color bar indicates fold changes of genes in the case dataset compared with in the control dataset.



Appendix 3—figure 4. The causal relationships between genes in the WP4290 pathway in the cell line H838 inferred using 400 cells ($\alpha = 0.1$).



Appendix 3—figure 5. The causal relationships between genes in the WP4290 pathway in the cell line H838 inferred using 600 cells ($\alpha = 0.1$).



Appendix 3—figure 6. The causal relationships between genes in the WP4290 pathway in the cell line H838 inferred using 800 cells ($\alpha = 0.1$). More cells make more relationships be inferred, but relationships with high significance (with thick arrows) are stable.

Further, we benchmarked six continuous optimization-based methods (NOTEARS-linear, NOTEARS-nonlinear, DAGMA-linear, DAGMA-nonlinear, GOLEM, and DAG_GNN) (Bello *et al.*, 2022a; Zheng *et al.*, 2018), and two linear non-Gaussian acyclic model methods (ICLiGNAM and DirectLiGNAM). We compared the performance of these methods with PC+DCC.gamma and PC+GaussCltest. Continuous optimization-based methods, especially DAGMA-nonlinear (<https://github.com/kevinsbello/dagma>; Bello *et al.*, 2022b), perform well when relationships between variables are free of missing variables and missing values, otherwise they perform poorly and underperform PC+DCC.gamma. All benchmarking used both simulated data and multiple scRNA-seq datasets, especially the five lung cancer cell lines (A549, H1975, H2228, H838, HCC827) from the CellBench benchmarking dataset (Tian *et al.*, 2019). Genes differentially expressed in these cell lines were determined upon gene expression in the lung alveolar cells (Travaglini *et al.*, 2020).

2. Partial correlation-based CI test

GaussCltest

Gauss CI test examines CI using partial correlation, assuming that all variables are multivariate Gaussian. The partial correlation coefficient ρ_{XYZ} is zero if and only if X is conditionally independent

of Y given Z (Kunihiro et al., 2004). H_0 is $\hat{\rho}_{XYZ} = 0$, H_1 is $\hat{\rho}_{XYZ} \neq 0$, and a hypothesis test ($p < 0.05$) decides whether two variables are conditionally independent given Z . We used the `gaussCltest` function in the R package `pcalg` with default parameters (<https://cran.r-project.org/web/packages/pcalg/index.html>).

3. HSIC-based CI test

HSIC is a measure of dependency between two variables; $\text{HSIC}(X, Y) = 0$ if X and Y are unconditionally independent. Performing two extra transformations can determine if X and Y are conditionally independent given the conditioning set Z : first, performing nonlinear regressions for X and Z and for Y and Z , respectively, to generate the residuals X_{resid} and Y_{resid} based on Z ; then, calculating $\text{HSIC}(X_{resid}, Y_{resid})$ that indicates whether X and Y are conditionally independent given the conditioning set Z ($X \perp\!\!\!\perp Y | Z$) (Verbyla et al., 2017). We used the `gam()` function in the R package `mgcv` to build the nonlinear regression model and used the three HSIC-based functions (with default parameters unless otherwise specified) in the R package `kpcalg` (<https://cran.r-project.org/web/packages/kpcalg/index.html>) to perform CI test.

hsic.perm

In practice, $\text{HSIC}(X, Y)$ may be slightly larger than 0.0 when X and Y are independent, making it hard to judge whether X and Y are independent. `hsic.perm` uses a permutation test to solve this problem by assuming that permuting Y removes any dependency between X and Y . We used the `hsic.perm` function to permute Y 100 times to calculate $\text{HSIC}(X, Y_{perm})$, then we compared them with $\text{HSIC}(X, Y)$. The p-value was the fraction of times $\text{HSIC}(X, Y_{perm})$ was smaller than the $\text{HSIC}(X, Y)$.

hsic.gamma

We used the `hsic.gamma` function to fit a gamma distribution: $\text{Gamma}(\alpha, \theta)$ of the HSIC under the null hypothesis. The shape parameter α and the scale parameter θ were calculated using the equation:

$$\alpha = \frac{E[\hat{H}_{X,Y}]^2}{\text{Var}(\hat{H}_{X,Y})}, \theta = \frac{\text{Var}(\hat{H}_{X,Y})}{E[\hat{H}_{X,Y}]}$$

A p-value was obtained as an upper-tail quantile of $\text{HSIC}(X, Y)$.

hsic.clust

First, samples were clustered using the R function `kmeans()` by calculating the Euclidean distance between the Z coordinates of samples; then, Y was permuted based on the clustered Z . Within each Z , cluster Y_{perm} was generated, ensuring that the permuted samples break the dependency between X and Y but retain the dependency between X and Y on Z . After permutation, a p-value was calculated to make a statistical decision.

4. Distance covariance-based CI test

Distance covariance is an alternative to HSIC for measuring independence (Székely and Rizzo, 2009; Székely et al., 2007). We used two DCC-based functions `dcc.perm` and `dcc.gamma` (with default parameters) in the R package `kpcalg` (<https://cran.r-project.org/web/packages/kpcalg/index.html>) to perform CI test. Similar to HSIC-based algorithms, the two functions directly calculate DCC(X, Y) for an UI test, then, the nonlinear regression is performed, next, DCC(X_{resid}, Y_{resid}) is calculated for a CI test and a statistical decision (Verbyla et al., 2017).

dcc.perm

This program is similar to `hsic.perm` and uses a permutation test to estimate a p-value. The DCC statistic is calculated in each permutation, and finally, a statistical decision is made based on the p-value. We used the `dcov.test` function (with default parameters) in the R package `energy` to calculate the statistic DCC in the permutation test. The p-value was the fraction of times that DCC(X, Y_{perm}) was smaller than DCC(X, Y).

dcc.gamma

Similar to `hsic.gamma`, `dcc.gamma` uses the gamma distribution $\text{Gamma}(\alpha, \theta)$ of the DCC under the null hypothesis. The two parameters were estimated by

$$\alpha = \frac{E[\hat{D}_{X,Y}]^2}{\text{Var}(\hat{D}_{X,Y})}, \theta = \frac{\text{Var}(\hat{D}_{X,Y})}{E[\hat{D}_{X,Y}]}$$

We used the `dcov.gamma` function (with default parameters) in the R package `kpcalg` to calculate the p-value. The p-value was obtained as an upper-tail quantile of $DCC(X, Y)$.

5. Approximation of KCIT

The KCIT is another powerful CI test ([Zhang and Peters, 2011](#)), but it is time-consuming for large datasets. Based on random Fourier features ([Rahimi and Recht, 2007](#)), two approximation methods (randomized conditional independence test, RCIT, and randomized conditional correlation test, RCoT) were proposed ([Strobl et al., 2019](#)). RCIT and RCoT approximate KCIT by sampling Fourier features, return p-values orders of magnitude faster than KCIT when the sample size is large, and may also estimate the null distribution more accurately than KCIT.

RCIT

We used the `RCIT` function in the R package `RCIT` (with default parameters) (<https://github.com/ericstrobl/RCIT>; [Strobl, 2019](#)) to implement the randomized CI test.

RCoT

RCoT often outperforms RCIT, especially when the size of the conditioning set is greater than or equal to 4. We used the `RCoT` function in the R package `RCIT` (with default parameters) (<https://github.com/ericstrobl/RCIT>; [Strobl, 2019](#)) to implement the RCoT.

6. Conditional mutual information-based CI test

Mutual information is used to measure mutual dependence between two variables. Conditional mutual information (CMI) is a measure based on mutual information, which is zero if and only if $X \perp\!\!\!\perp Y|Z$.

CMIknn

`CMIknn` is a program that combines CMI with a local permutation scheme determined by the nearest-neighbor approach ([Runge, 2018](#)). We used the Python package `tigramite` (with default parameters) (<http://github.com/jakobrunge/tigramite>; [Runge, 2020](#)) to perform the CI test.

7. CI test based on generalized covariance measure

GCM

GCM (<https://cran.r-project.org/web/packages/GeneralisedCovarianceMeasure/index.html>; [Peters and Shah, 2022](#)) is a CI test based on generalized covariance measure. It is also classified as a regression-based CI test because it is based on a suitably normalized version of the empirical covariance between the residual vectors from the regressions ([Shah and Peters, 2020](#)).

8. Benchmarking results

The time consumption, accuracy, sample requirement, and stability of the PC+ nine CI tests were evaluated ([Appendix 3—table 1](#)). First, we simulated eight datasets with known causal networks, whose variable numbers and sample sizes ranged from 20 to 80 (step = 20) and 500 to 1000 (step = 500), respectively, to evaluate causal discovery algorithms' time consumption, scalability, and accuracy. Algorithms based on the DCC kernel were more time-consuming than others ([Appendix 3—figure 1A,C](#)). Algorithms' accuracy was assessed based on the structural Hamming distance (SHD) between the inferred and the true networks (SHD = 0 indicates no difference). The networks of all algorithms showed similar SHD when the sample size was 500 ([Appendix 3—figure 1B](#)); the close performance was probably because synthetic data were generated using a few simple functions. When the sample size was increased from 500 to 1000, time consumption increased (but was not doubled), but SHD did not decrease (i.e. algorithms' performance did not increase) significantly, indicating that 500 cells may be adequate for causal discovery ([Appendix 3—figure 1D](#)).

Second, to further evaluate algorithms' accuracy, for each feature gene set, we merged causal networks generated by multiple good algorithms into a consensus network (multi-algorithm-based

consensus network), then compared the network of each algorithm with the consensus network (Main text-**Figure 2**; **Appendix 3—figure 2**). We used the SHD to define the difference between two networks, and the network with the shortest SHD with the consensus network is assumed to be the most accurate.

Third, to evaluate the impact of sample size on algorithms' performance, we ran the nine algorithms using 200 (instead of 300) H2228 cells. The results of 200 cells were poorer than the results of 300 cells (compared with the consensus network in Main text-**Figure 2** and **Appendix 3—figure 2**). Still, the two DCC algorithms performed the best and were less sensitive to the decreased sample size than the two HSIC algorithms. We also inferred interactions between genes in the 'Metabolic reprogramming in colon cancer' (WP4290) pathway using 200, 400, 600, and 800 cells in the H838 (**Appendix 3—figures 3–6**). We find that more cells make more interactions be inferred, but the interactions with high significance are quite stable.

Fourth, to evaluate algorithms' stability, we used the H2228 dataset to run the nine algorithms five times and estimated each algorithm's stability by computing the mean relative SHD of the five networks. The networks of gaussCltest have the smallest mean relative SHD and the networks of HSIC.perm, HSIC.clust, and DCC.perm have the largest mean relative SHD (**Appendix 3—table 1**). As DCC.perm and DCC.gamma are the most accurate algorithms, we examined whether their stability impairs their accuracy by checking the distribution of interactions in the five networks. DCC.gamma and DCC.perm inferred 127 and 143 interactions, 78% and 64.3% occurred stably in ≥ 4 networks, and many inconsistent interactions occurred in just one network (Main text-**Figure 3**), indicating that most interactions were stably inferred in multiple running. The networks of multiple running can be merged into a consensus network (multi-running-based consensus network), which can be used to examine which algorithm generated the most consistent networks.

Fifth, we compared the accuracy of PC+DCC.gamma, GES, and GSP+DCC.gamma using genes in the WikiPathways 'Metabolic reprogramming in colon cancer' (WP4290) and 600 cells in the A549, H2228, and H838 datasets. GSP+DCC.gamma (the significance level $\alpha = 0.01$) inferred much more interactions than PC+DCC.gamma ($\alpha = 0.1$) and GES ($\alpha = 0.1$). The results indicate that PC+DCC.gamma ($\alpha = 0.1$) and GSP+DCC.gamma ($\alpha = 0.05$) have comparable accuracy and time consumption, and both are more accurate but time-consuming than GES ($\alpha = 0.1$) (**Appendix 3—figures 7–12**).

Appendix 3—table 1. Performance of the nine causal discovery algorithms ('+++′ and '+′ indicate the best and worst, respectively).

Algorithm	Time complexity*	Time consumption†	Accuracy ‡	Sample size	Stability (mean of rSHD) §
GaussCltest	$O(q^3)$	+++	+	+++	0.075 (+++)
CMlkn	$O(n^2)$	+	++	+	0.25 (+)
RCIT	$O(d^2 \times n)$	++	++	++	0.12 (++)
RCoT	$O(d^2 \times n)$	++	++	++	0.12 (++)
HSIC.clust	$O\left(\sum_{k=1}^K m_k^3\right)$	+	+++	++	0.26 (+)
HSIC.gamma	$O(n^3)$	+	+++	++	0.12 (++)
HSIC.perm	$O(r \times n^3)$	+	+++	++	0.28 (+)
DCC.gamma	$O(n^3)$	+	+++	+++	0.14 (++)
DCC.perm	$O(r \times n^3)$	+	+++	+++	0.24 (+)
GCM	Depending on regression methods	+	++	++	

Appendix 3—table 1 Continued on next page

Appendix 3—table 1 Continued

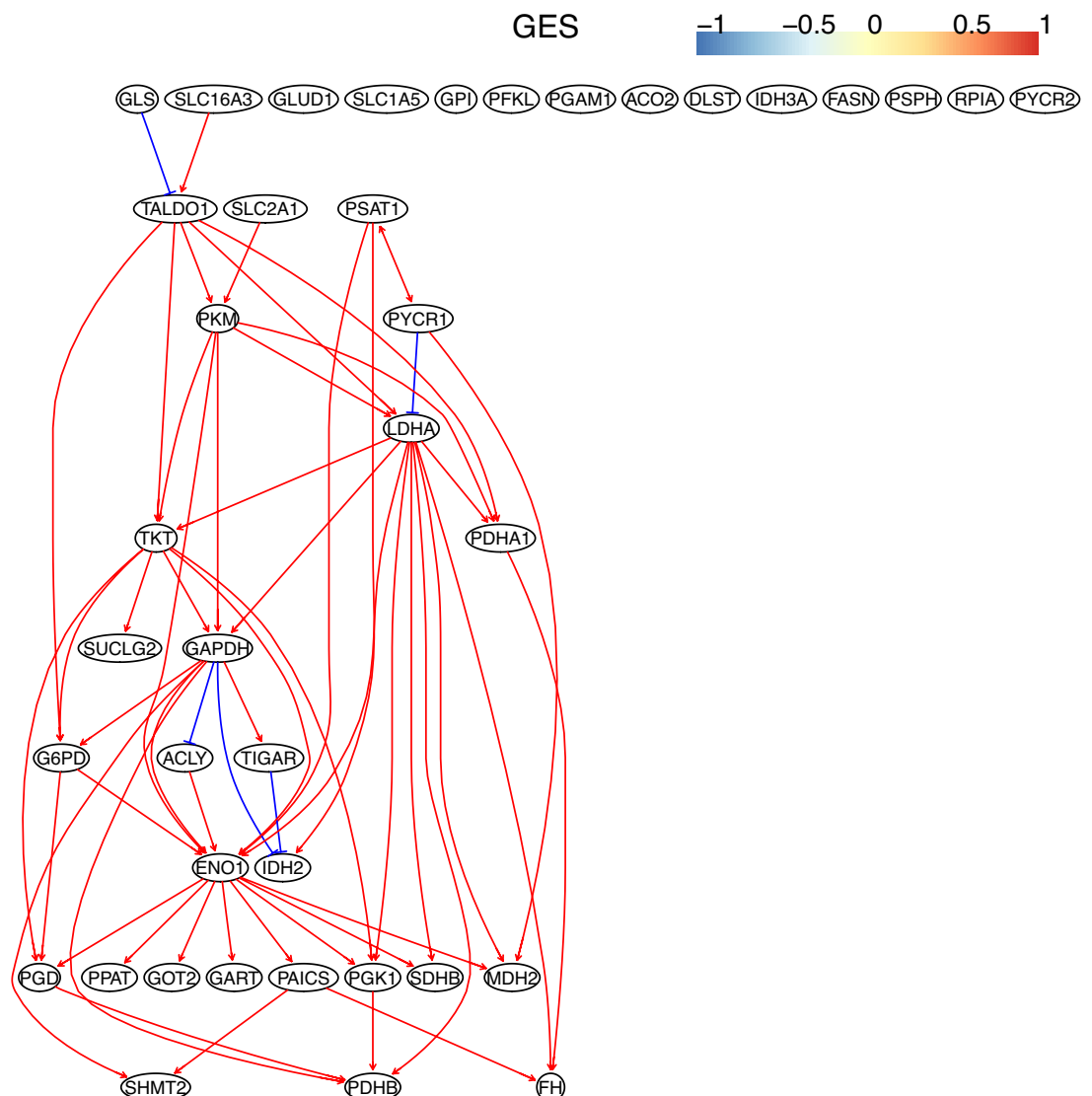
Algorithm	Time complexity*	Time consumption†	Accuracy ‡	Sample size	Stability (mean of rSHD) §
-----------	------------------	-------------------	------------	-------------	----------------------------

*(a) Assuming a dataset has n samples and the total dimension of X, Y, Z is q . Generally, $q \ll n$. (b) r is the time of permutation. (c) d is the number of random Fourier features, generally $d < n$. (d) The time complexity of the PC algorithm is $\frac{N^2(N-1)^{deg-1}}{(deg-1)!}$ where N is the number of nodes and deg is the maximal degree.

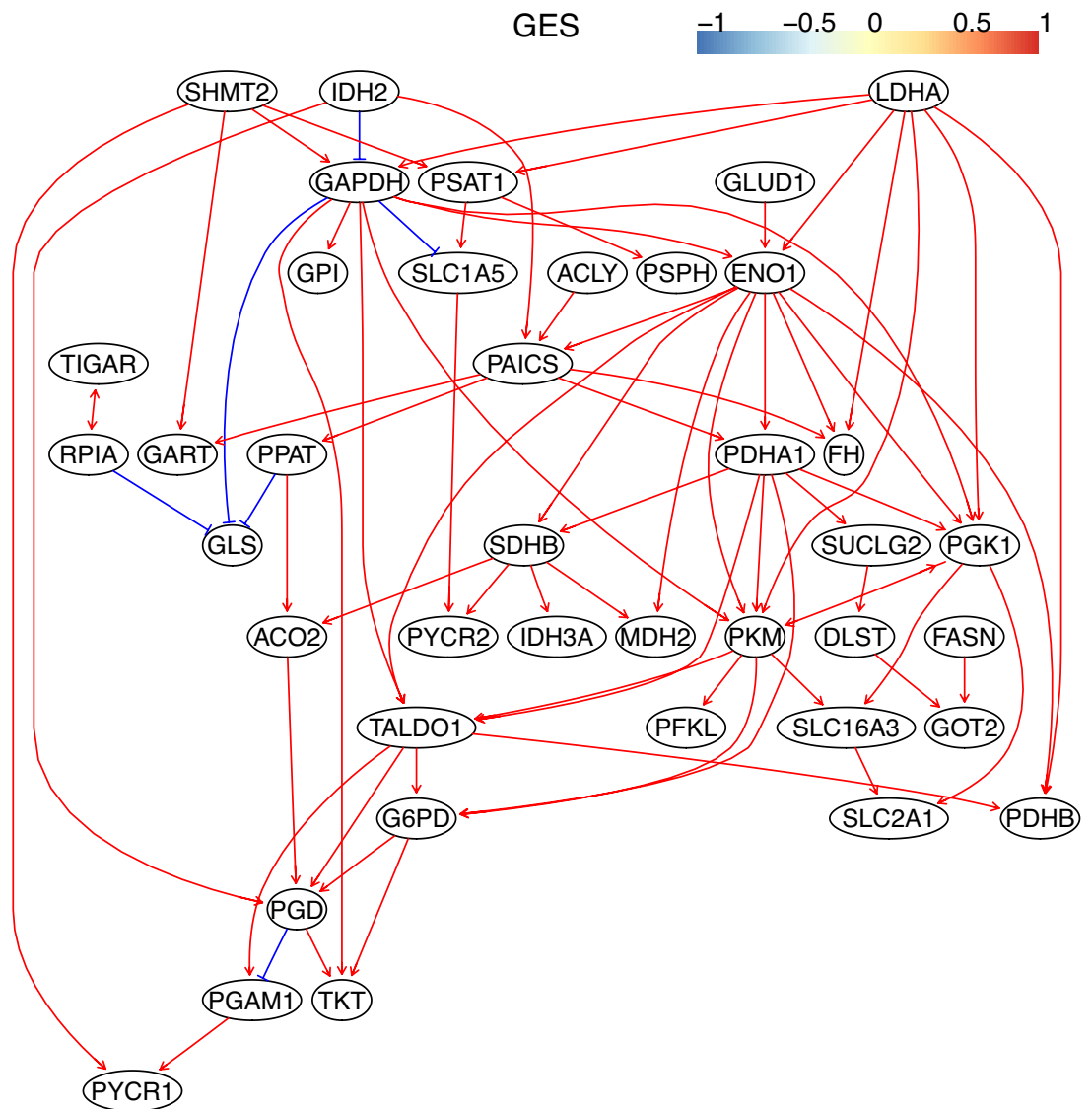
†Time-consuming levels are estimated upon simulated data (Appendix 3—figure 1).

‡Accuracy is estimated upon the lung cancer cell lines (Main text: Figure 2; Appendix 3—figure 2). We performed causal discovery using the nine algorithms five times for the H2228 cell line and obtained $9 \times 5 = 45$ causal networks.

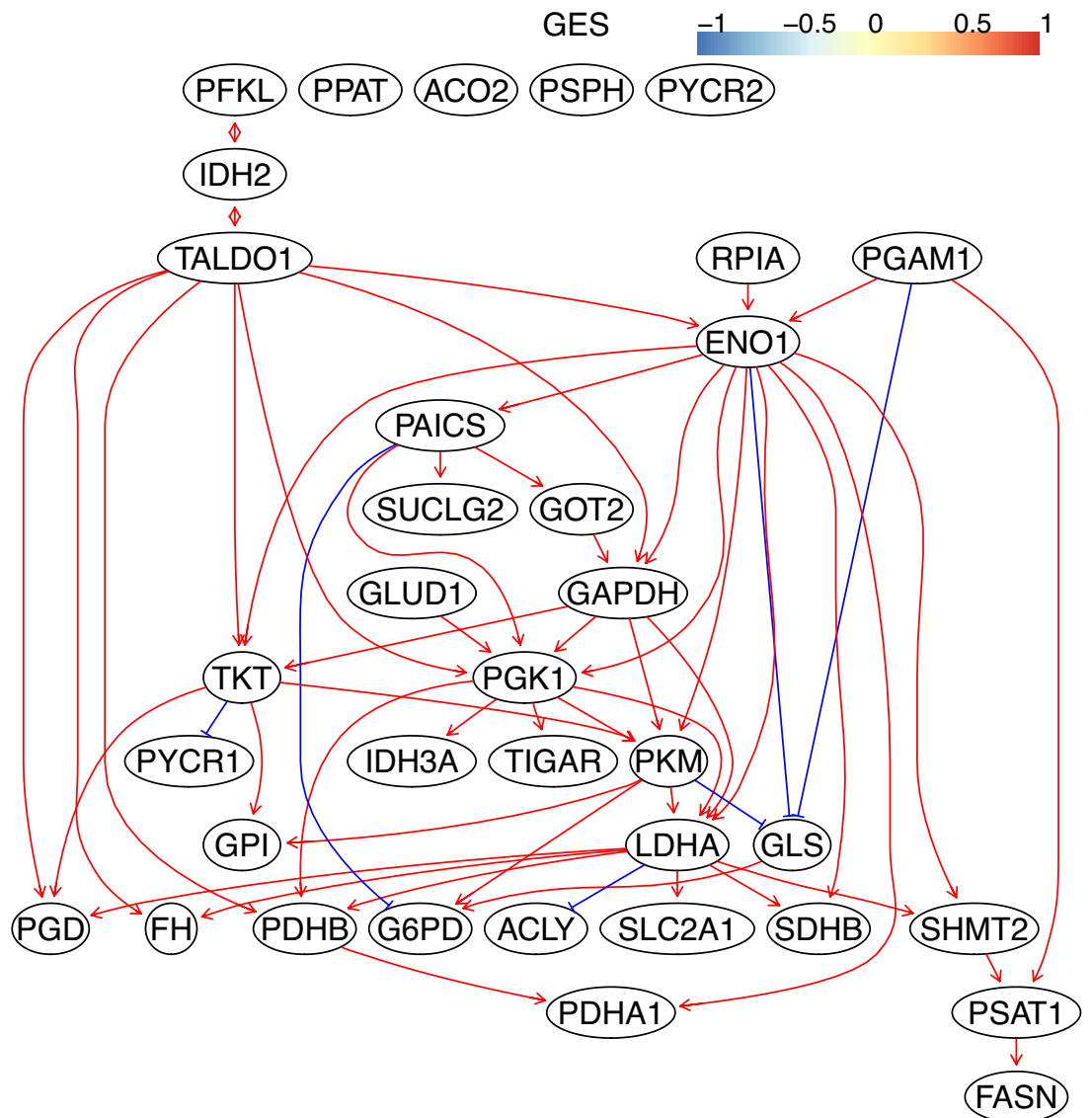
§We estimated the stability of each algorithm's performance by computing the mean relative SHD for the five causal networks the algorithm generated using the equation: $\frac{2}{NG(NG-1)} \sum_{i=1}^{NG-1} \sum_{j=i+1}^{NG} \frac{SHD(G_i, G_j)}{\#edges(G_i)}$. In this equation, $SHD(G_i, G_j)$ is the structural Hamming distance between causal network G_i and G_j , $\#edges(G_i)$ is the number of edges in G_i , and $NG = 5$ because each algorithm generates five causal networks.



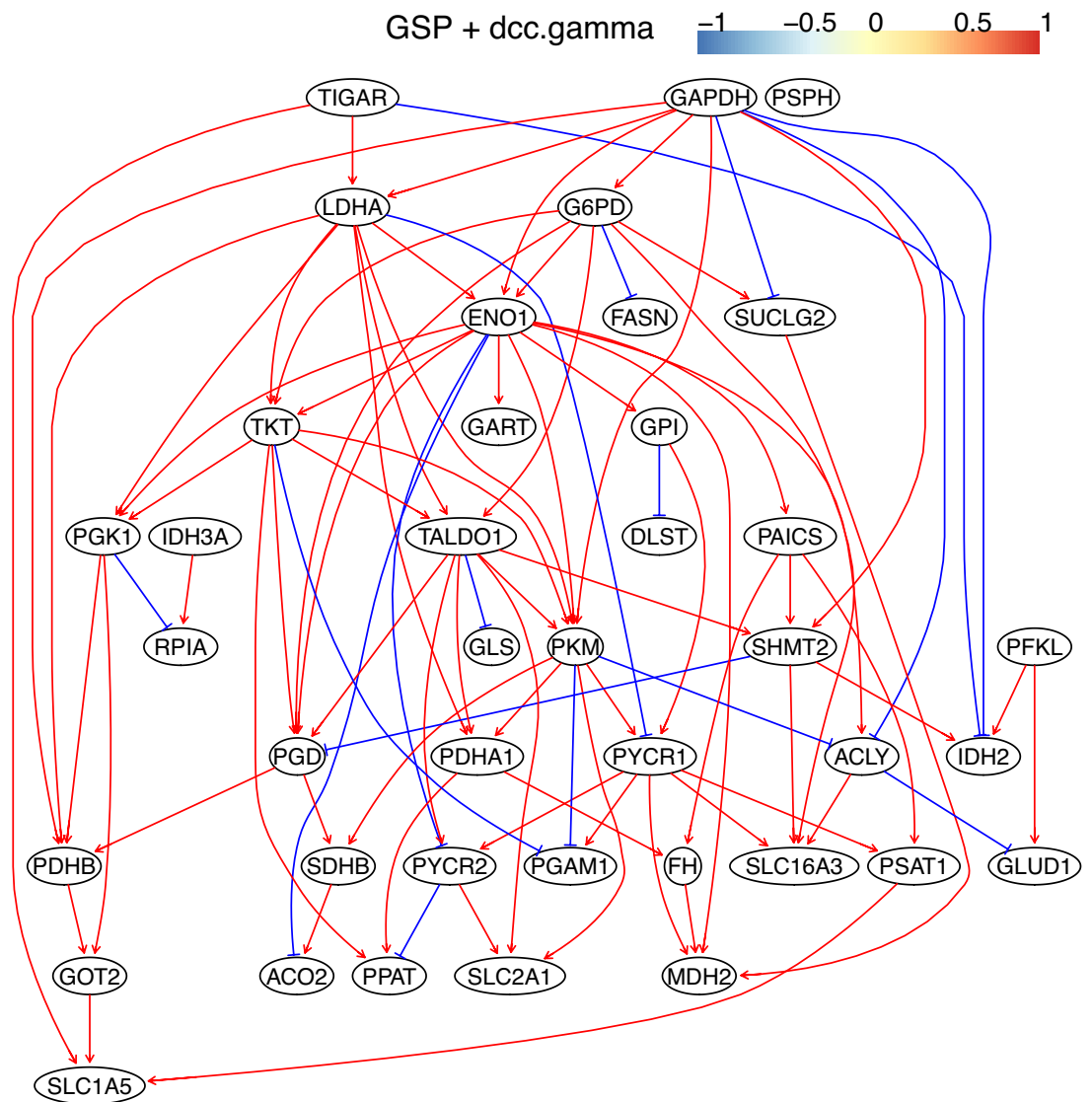
Appendix 3—figure 7. The causal relationships between genes in the WP4290 pathway in the cell line A549 inferred using the GES method and 600 cells ($\alpha = 0.1$). Compared with the networks in these cells inferred using PC+DCC.gamma, here there are more isolated nodes.



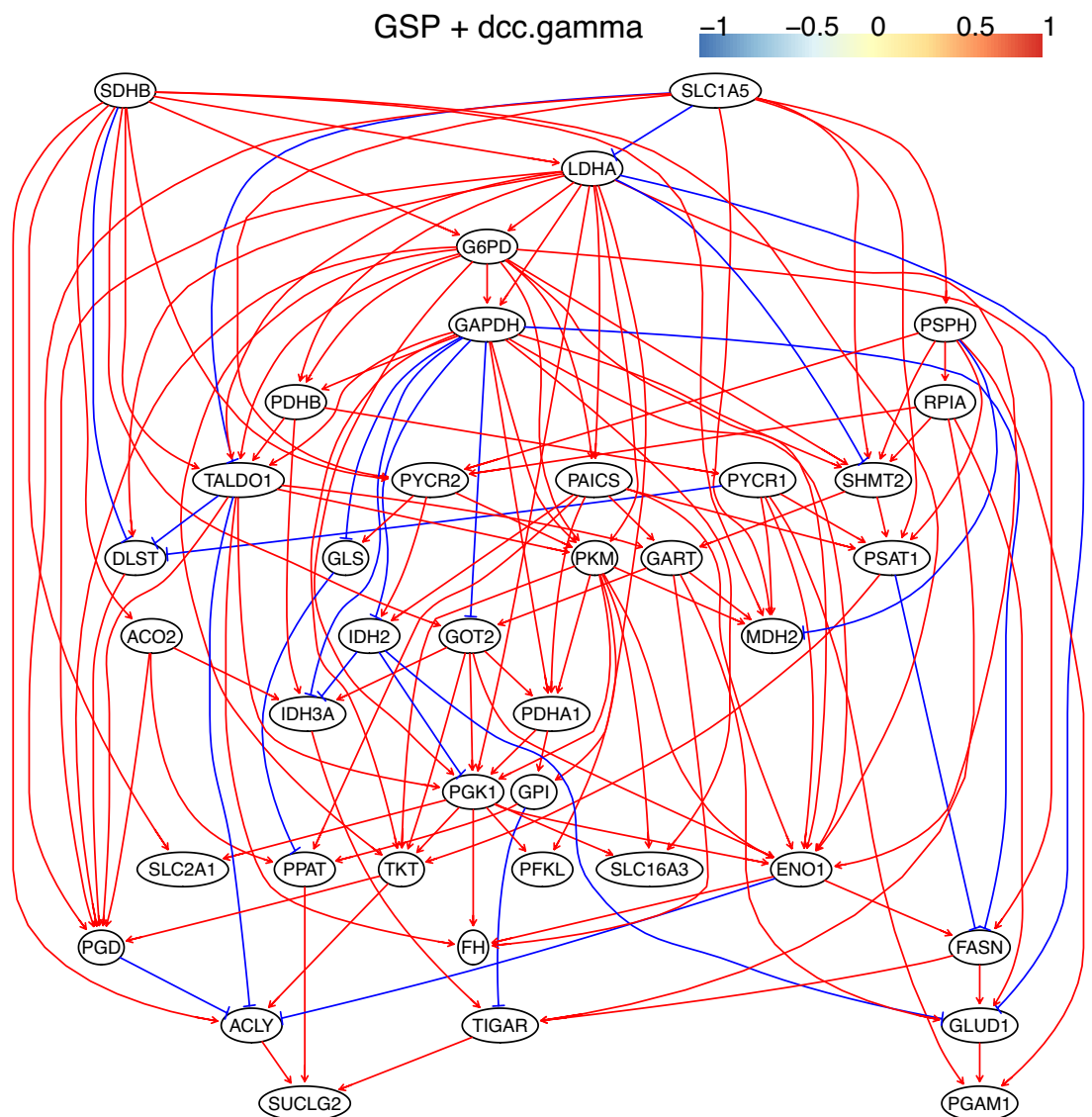
Appendix 3—figure 8. The causal relationships between genes in the WP4290 pathway in the cell line H2228 inferred using the GES method and 600 cells ($\alpha = 0.1$).



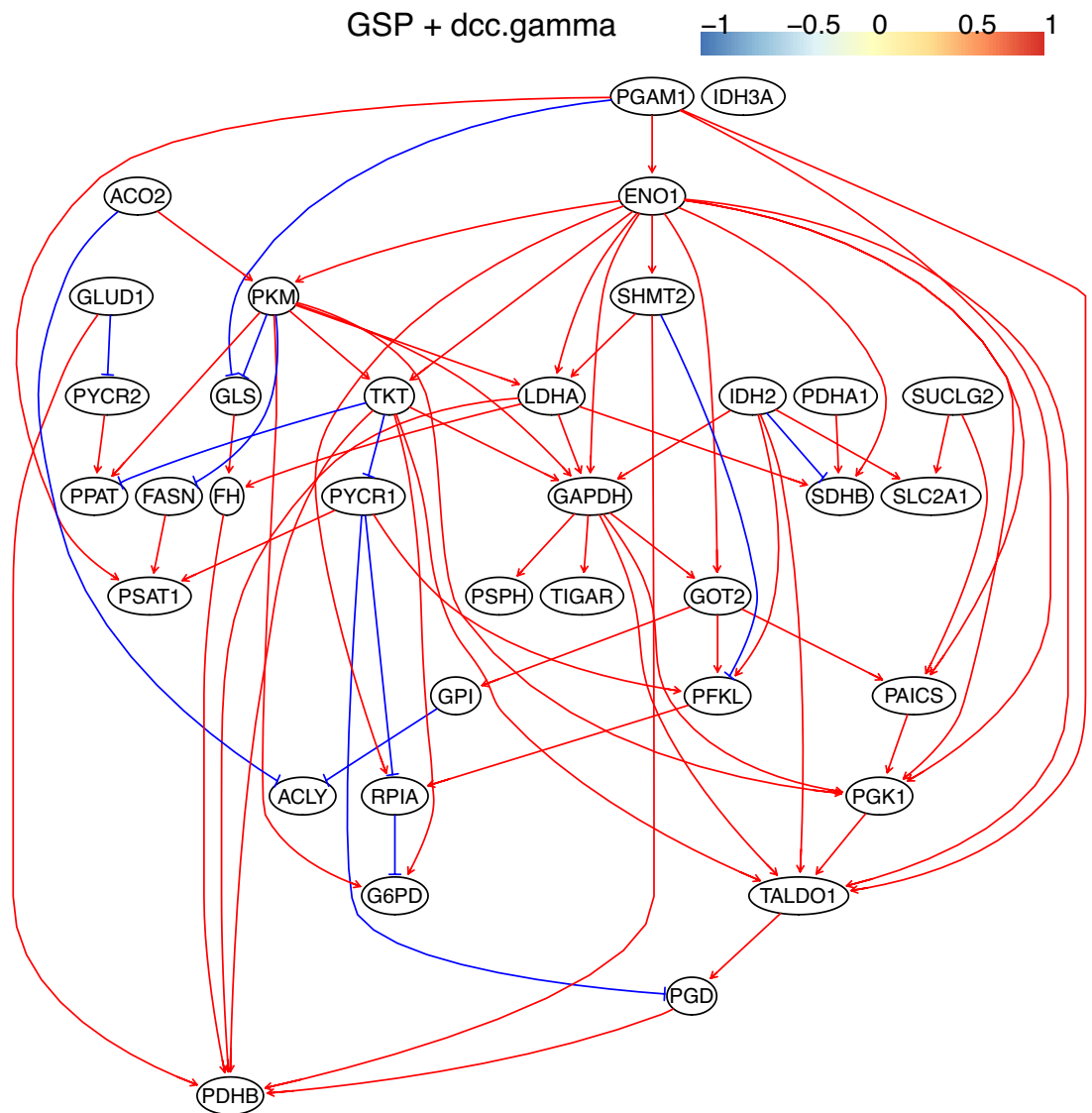
Appendix 3—figure 9. The causal relationships between genes in the WP4290 pathway in the cell line H838 inferred using the GES method and 600 cells ($\alpha = 0.1$).



Appendix 3—figure 10. The causal relationships between genes in the WP4290 pathway in the cell line A549 inferred using the GSP+DCC.gamma and 600 cells ($\alpha = 0.05$). Compared with the networks in these cells inferred using PC+DCC.gamma (**Appendix 4—figures 6–8**), more relationships are inferred even if the significance level is 0.05. The key features of the reprogrammed glucose metabolism (as indicated in the inferred networks of PC+DCC.gamma, see Appendix 4) also occur in the network.



Appendix 3—figure 11. The causal relationships between genes in the WP4290 pathway in the cell line H2228 inferred using the GSP+DCC.gamma and 600 cells ($\alpha = 0.05$). Compared with the networks in these cells inferred using PC+DCC.gamma (**Appendix 4—figures 6–8**), more relationships are inferred even if the significance level is 0.05. The key features of the reprogrammed glucose metabolism (as indicated in the inferred networks of PC+DCC.gamma, see Appendix 4) also occur in the network.

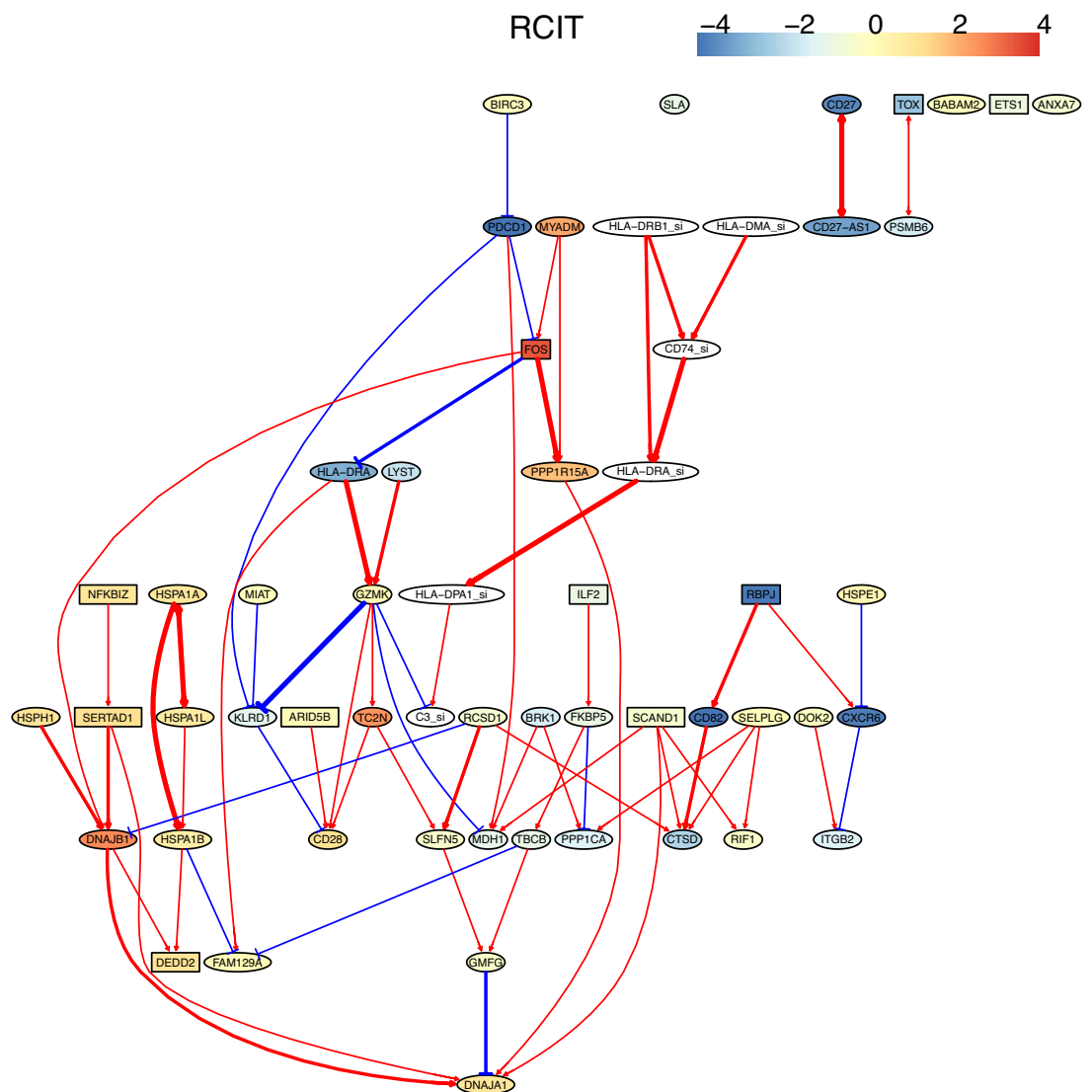


Appendix 3—figure 12. The causal relationships between genes in the WP4290 pathway in the cell line H838 inferred using the GSP+DCC.gamma and 600 cells ($\alpha = 0.05$). Compared with the networks in these cells inferred using PC+DCC.gamma (**Appendix 4—figures 6–8**), more relationships are inferred even if the significance level is 0.05. The key features of the reprogrammed glucose metabolism (as indicated in the inferred networks of PC+DCC.gamma, see Appendix 4) also occur in the network.

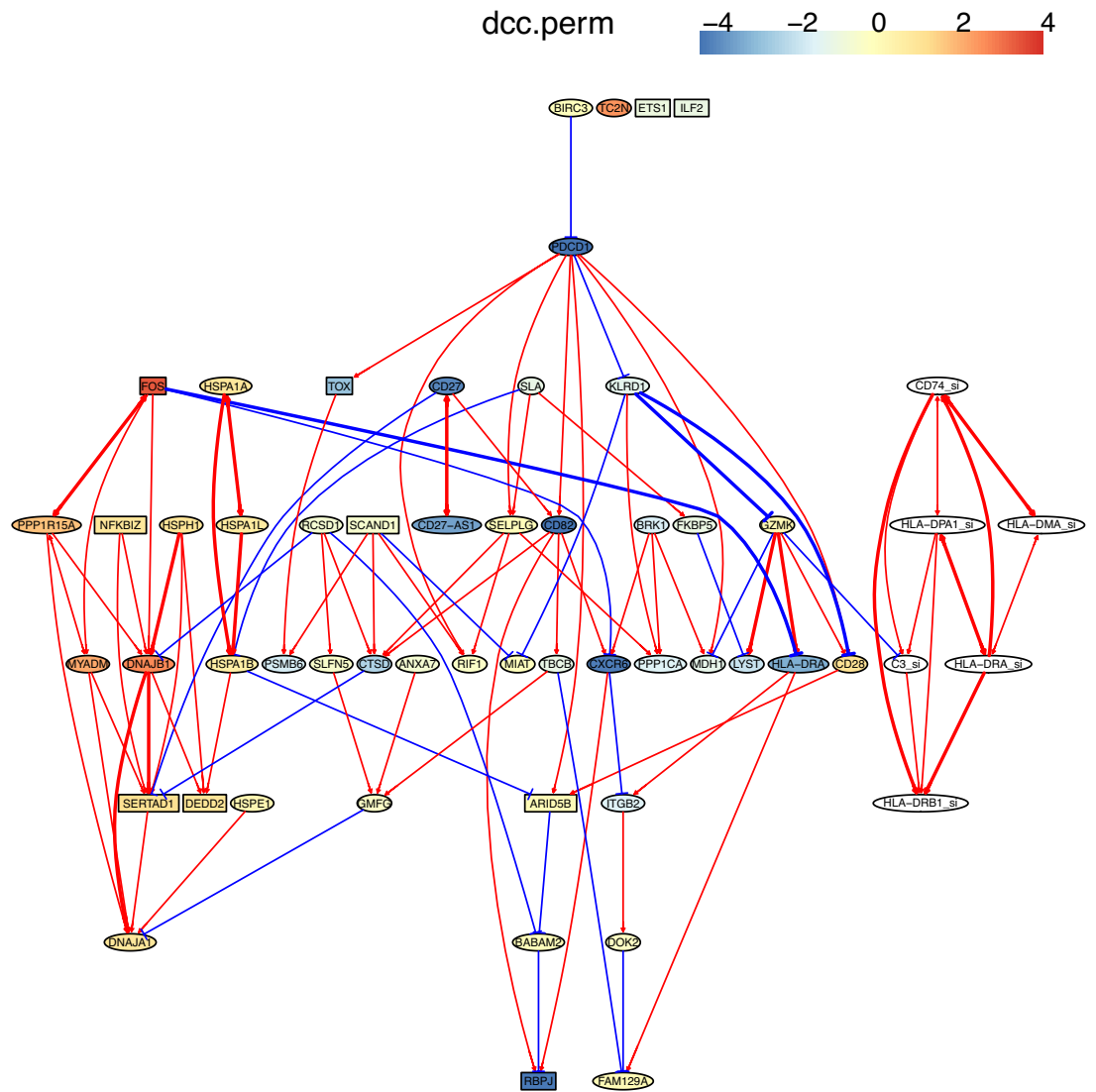
Appendix 4

Evaluating the reliability and verifying causal discovery results

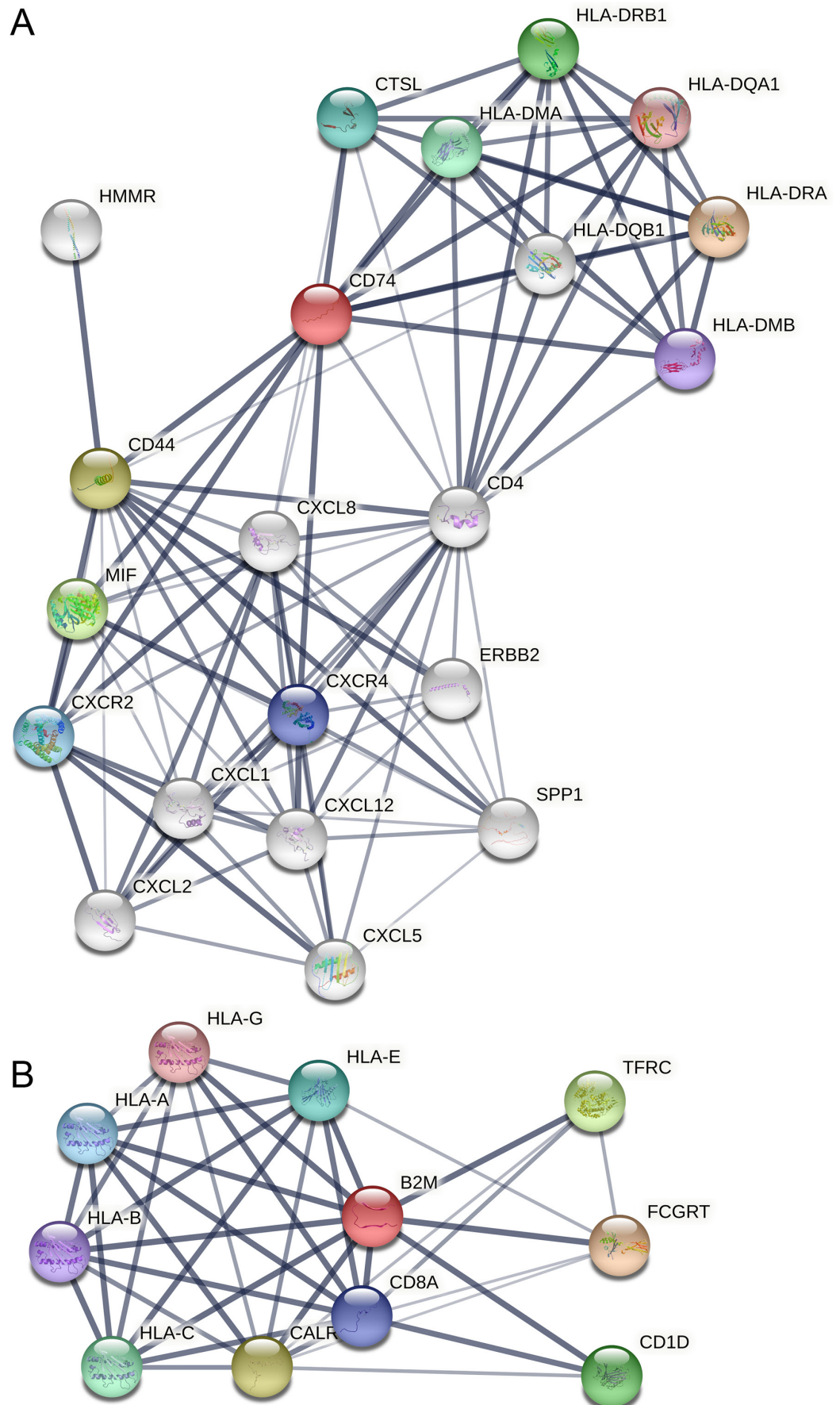
We evaluated the reliability of causal discovery by examining whether algorithms can differentiate interactions between genes in different cells. Inspired by using RNA spike-in to measure RNA-seq quality, we extracted the data of six MHC-II-related genes (HLA-DRB1, HLA-DMA, HLA-DRA, HLA-DPA, CD74, C3, which have the suffix *_si* to mark them) from the macrophage dataset (generated by Smart-seq2 sequencing) and the alveolar epithelial cell dataset (generated by 10x Genomics) to form two spike-in datasets. We mixed the spike-in dataset with the dataset of exhausted CD8 T cells and examined if the causal discovery was able to separate MHC-II genes and their interactions in the spike-in dataset from feature genes and their interactions in the exhausted CD8 T dataset. When the datasets contain sufficient cells (usually >300), the two DCC algorithms can discriminate genes and interactions in the two datasets quite well (**Appendix 4—figures 1 and 2**), indicating the power of causal discovery based on kernel-based CI tests. The inferred causal interactions can be verified using annotated protein interactions in the STRING database (<https://string-db.org/>). The results of our application cases indicate that many inferred interactions are supported by annotated protein interactions in the STRING database (**Appendix 4—figures 3 and 4**).



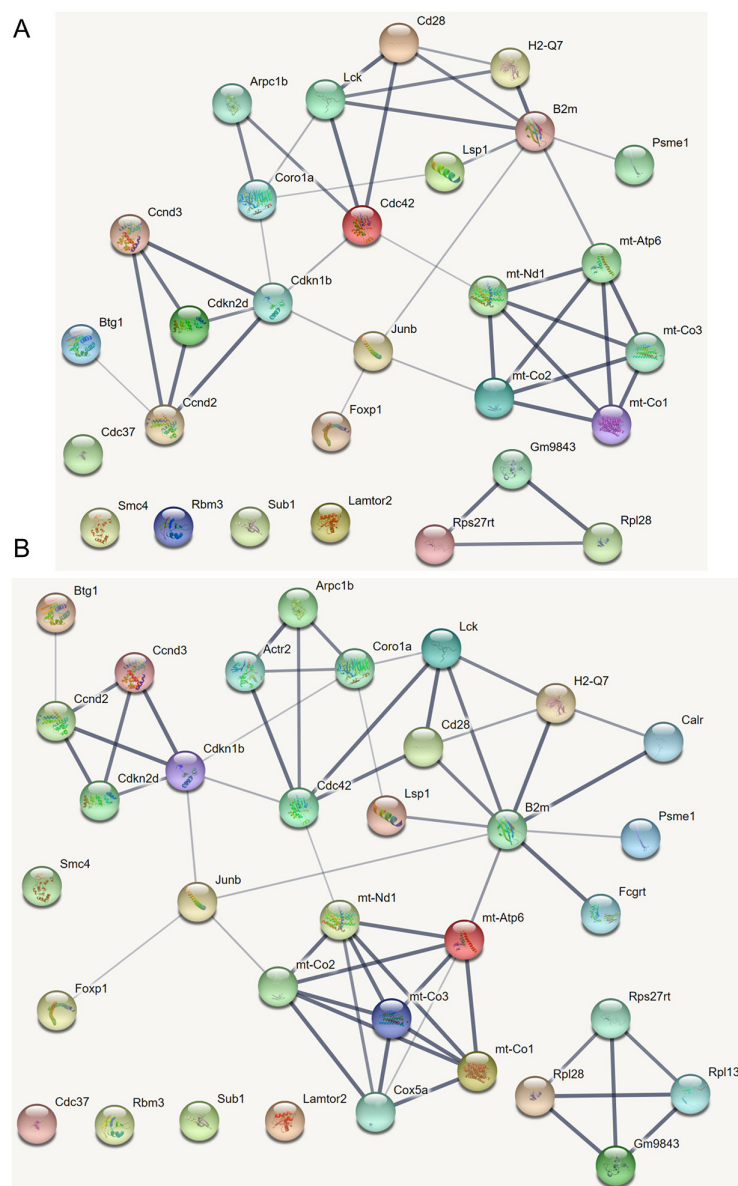
Appendix 4—figure 1. Multiple algorithms can identify genes and their relationships in the spike-in dataset from genes and their relationships in the primary dataset. The six genes and their relationships were identified in the network of RCIT, but the network has multiple orphan nodes.



Appendix 4—figure 2. Multiple algorithms can identify genes and their relationships in the spike-in dataset from genes and their relationships in the primary dataset. The six genes and their relationships were identified in the network of RCIT, but the network has multiple orphan nodes. GZMK-|C3_si is a wrong interaction in the networks of RCIT and DCC.perm.



Appendix 4—figure 3. Protein interactions in the STRING database (parameter settings: *network type* = full STRING network, *meaning of network edges* = confidence, *active interaction sources* = all, *minimum required interaction score* = medium confidence, 0.4). **(A)** Interactions among MHC-II genes and CD74 and among CXCL and CXCR genes. **(B)** Interactions among MHC-I genes and B2M.



Appendix 4—figure 4. Interactions among the differentially expressed genes in CD4 T cells in the STRING database (<https://string-db.org/>) (parameter settings: *network type* = full STRING network, *meaning of network edges* = confidence, *active interaction sources* = all, *minimum required interaction score* = medium confidence, 0.4). **(A)** The interactions. **(B)** The extended interactions ('add more nodes to current network' is chosen).

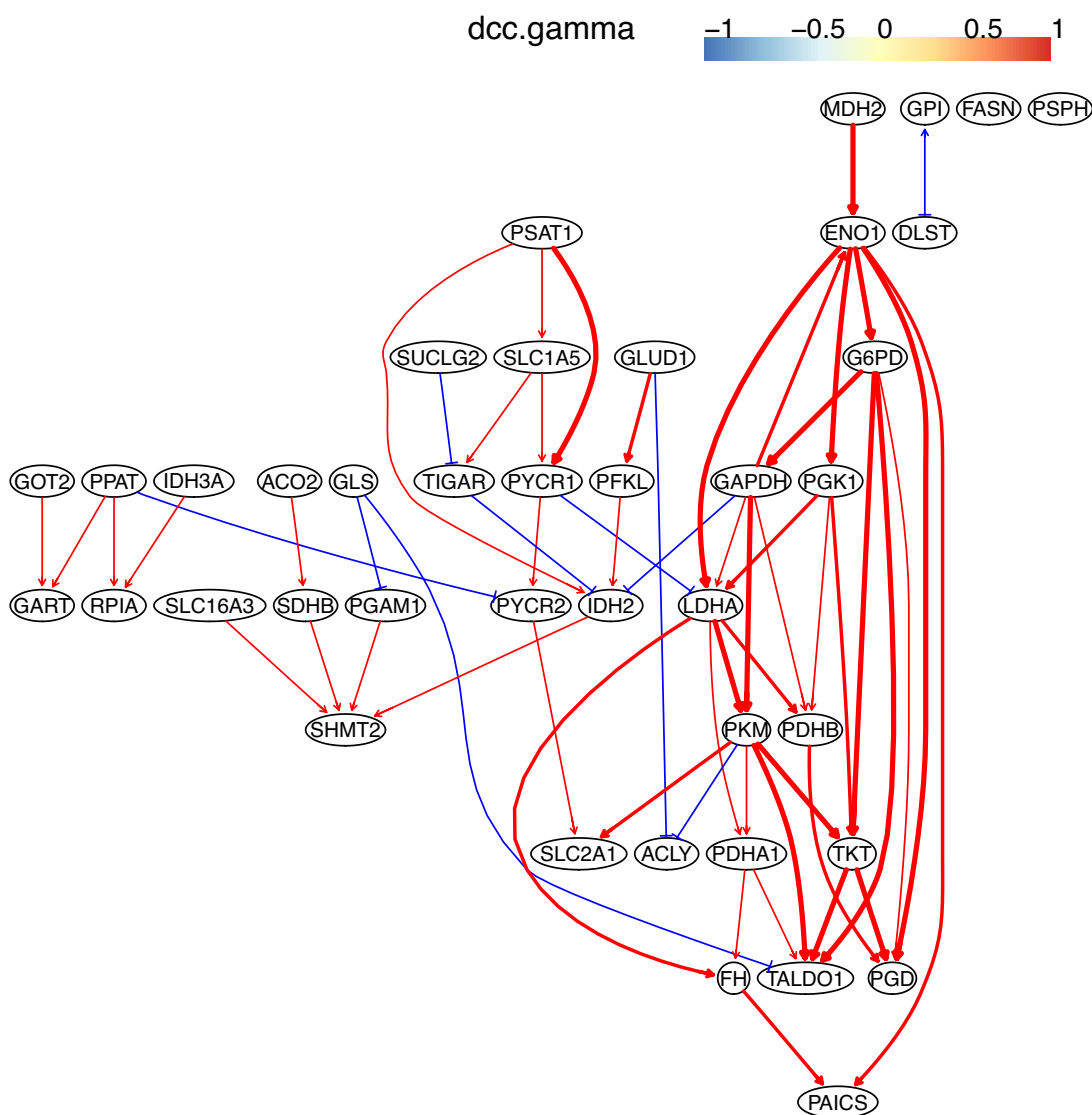
We have taken a systematic approach to validate causal discovery using the five lung cancer cell lines and lung alveolar cells. First, upon (a) gene expression value >0.1, (b) gene expression >50% cells, (c) fold change >0.3, we identified differentially expressed genes in each cell line against the alveolar cells. Second, we applied GO analysis to the differentially expressed genes in each cancer dataset using g:Profiler (<https://biit.cs.ut.ee/gprofiler/gost>) (parameters: Significance threshold = Benjamini-Hochberg FDR, User threshold = 0.05, Data sources = KEGG and WikiPathways). The WikiPathways and KEGG pathways 'Metabolic reprogramming in colon cancer' (WP4290), 'Pyrimidine metabolism' (WP4022), and 'Nucleotide metabolism' (hsa01232) are commonly enriched in all cancer cell lines

(**Appendix 4—figure 5**). We also performed GO analysis using the GSEA package, which identified the KEGG pathway ‘Non-small cell lung cancer’ (hsa05223) as an enriched pathway in cancer cell lines (note that these lung cancer cell lines were derived from NSCLC). We used the PC+DCC. gamma to infer interactions among genes in the three pathways in the five cancer cell lines and the alveolar cells.

Source	Term_name	Term_id	Adjusted p_value	Term_size	Intersect ion_size
KEGG	Ribosome	KEGG:03010	2.25E-34	127	72
KEGG	Parkinson disease	KEGG:05012	4.45E-33	238	99
KEGG	Huntington disease	KEGG:05016	5.62E-29	272	101
KEGG	Prion disease	KEGG:05020	2.25E-26	243	91
KEGG	Amyotrophic lateral sclerosis	KEGG:05014	6.00E-25	327	106
KEGG	Alzheimer disease	KEGG:05010	5.01E-22	353	106
KEGG	Oxidative phosphorylation	KEGG:00190	5.01E-22	118	56
KEGG	Chemical carcinogenesis - reactive oxygen species	KEGG:05208	1.54E-19	206	73
KEGG	Pathways of neurodegeneration - multiple diseases	KEGG:05022	1.76E-19	442	117
KEGG	Diabetic cardiomyopathy	KEGG:05415	5.83E-19	187	68
KEGG	Thermogenesis	KEGG:04714	5.02E-15	207	66
KEGG	Proteasome	KEGG:03050	7.21E-14	39	25
KEGG	Non-alcoholic fatty liver disease	KEGG:04932	5.62E-12	142	48
KEGG	Spliceosome	KEGG:03040	2.14E-09	123	40
KEGG	Coronavirus disease - COVID-19	KEGG:05171	4.04E-08	207	53
KEGG	Metabolic pathways	KEGG:01100	4.27E-06	1457	217
KEGG	Carbon metabolism	KEGG:01200	1.42E-05	112	31
KEGG	Spinocerebellar ataxia	KEGG:05017	7.13E-05	132	33
KEGG	Nucleotide metabolism	KEGG:01232	9.90E-05	83	24
KEGG	Protein processing in endoplasmic reticulum	KEGG:04141	3.19E-04	154	35
KEGG	Citrate cycle (TCA cycle)	KEGG:00020	4.02E-04	29	12
KEGG	Cardiac muscle contraction	KEGG:04260	8.53E-04	77	21
KEGG	Glutathione metabolism	KEGG:00480	9.14E-04	56	17
KEGG	Biosynthesis of unsaturated fatty acids	KEGG:01040	9.14E-04	27	11
KEGG	Protein export	KEGG:03060	1.63E-03	20	9
KEGG	Glycolysis / Gluconeogenesis	KEGG:00010	4.90E-03	64	17
KEGG	Base excision repair	KEGG:03410	4.90E-03	32	11
KEGG	Retrograde endocannabinoid signaling	KEGG:04723	7.56E-03	135	28
KEGG	DNA replication	KEGG:03030	1.06E-02	35	11
KEGG	Pyruvate metabolism	KEGG:00620	1.11E-02	46	13
KEGG	Fatty acid elongation	KEGG:00062	1.81E-02	27	9
KEGG	Salmonella infection	KEGG:05132	1.95E-02	236	41
KEGG	Pyrimidine metabolism	KEGG:00240	2.39E-02	56	14
KEGG	Cell cycle	KEGG:04110	2.47E-02	120	24
KEGG	Folate biosynthesis	KEGG:00790	2.88E-02	24	8
KEGG	Shigellosis	KEGG:05131	3.80E-02	239	40
KEGG	Bacterial invasion of epithelial cells	KEGG:05100	3.80E-02	72	16
KEGG	p53 signaling pathway	KEGG:04115	3.80E-02	72	16
KEGG	Pathogenic Escherichia coli infection	KEGG:05130	4.36E-02	183	32
KEGG	Cysteine and methionine metabolism	KEGG:00270	4.72E-02	49	12
WP	Cytoplasmic ribosomal proteins	WP:WP477	1.60E-25	69	47
WP	Electron transport chain: OXPHOS system in mitochondria	WP:WP111	5.24E-24	92	53
WP	Oxidative phosphorylation	WP:WP623	2.28E-11	55	29
WP	Metabolic reprogramming in colon cancer	WP:WP4290	2.23E-10	42	24
WP	Nonalcoholic fatty liver disease	WP:WP4396	2.25E-10	146	49
WP	Mitochondrial complex I assembly model OXPHOS system	WP:WP4324	2.69E-09	50	25
WP	Proteasome degradation	WP:WP183	3.15E-08	55	25
WP	mRNA processing	WP:WP411	1.01E-05	114	34
WP	VEGFA-VEGFR2 signaling pathway	WP:WP3888	2.34E-05	415	83
WP	Alzheimer's disease	WP:WP5124	2.34E-05	250	57
WP	Mitochondrial complex IV assembly	WP:WP4922	6.41E-05	29	14
WP	Cori cycle	WP:WP1946	2.06E-04	17	10
WP	Parkin-ubiquitin proteasomal system pathway	WP:WP2359	3.76E-04	60	20
WP	Aerobic glycolysis	WP:WP4629	4.24E-04	12	8
WP	Retinoblastoma gene in cancer	WP:WP2446	4.24E-04	86	25
WP	Pathogenic Escherichia coli infection	WP:WP2272	4.24E-04	47	17
WP	Glycolysis and gluconeogenesis	WP:WP534	2.88E-03	44	15
WP	Pentose phosphate metabolism	WP:WP134	3.89E-03	6	5
WP	TCA cycle (aka Krebs or citric acid cycle)	WP:WP78	9.88E-03	17	8
WP	Base excision repair	WP:WP4752	1.04E-02	30	11
WP	Alzheimer's disease and miRNA effects	WP:WP2059	1.04E-02	319	58
WP	nsp1 from SARS-CoV-2 inhibits translation initiation in the host	WP:WP5027	1.28E-02	14	7
WP	Pyrimidine metabolism	WP:WP4022	1.28E-02	78	20
WP	Eukaryotic transcription initiation	WP:WP405	1.32E-02	36	12

Appendix 4—figure 5. The enriched KEGG and WikiPathways pathways of differentially expressed genes in the A549 cell line. KEGG:01232, WP4290, and WP4022 are enriched in all of the five lung cancer cell lines.

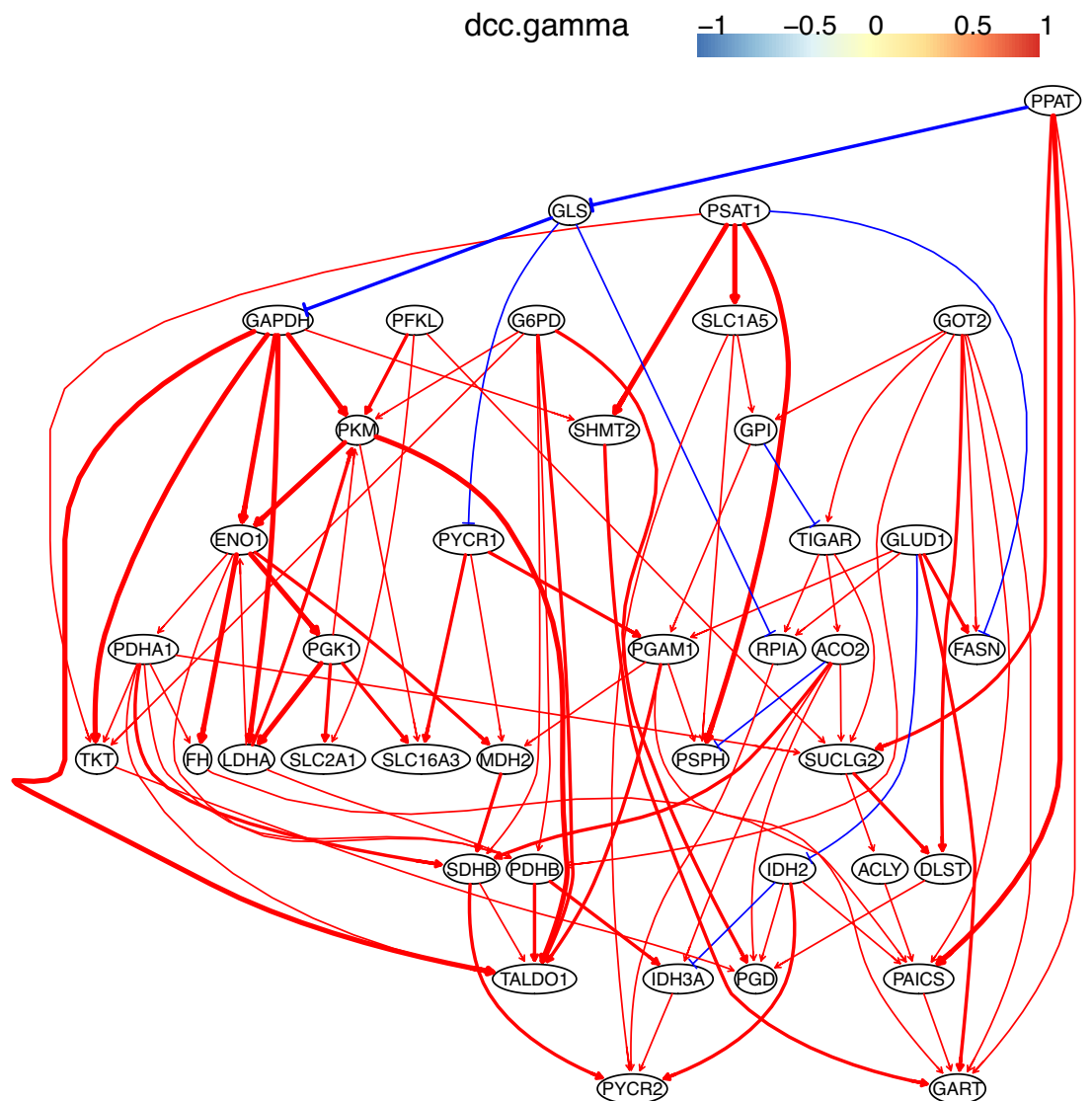
First, we examined the ‘Metabolic reprogramming in colon cancer’ (WP4290) pathway (Appendix 4—figures 6–9). Numerous studies report that glucose metabolism is reprogrammed and nucleotides synthesis is increased in cancer cells. Thus, we first examined and compared the WP4290 pathway in the five lung cancer cell lines and lung alveolar cells. The key features of the reprogrammed glucose metabolism are that (a) glucose intake is increased, (b) the glycolysis/TCA cycle intermediates are used for synthesizing nucleotide, (c) lactate generation is increased. The inferred networks capture these features. (a) multiple activations of SLC2A1 (which encodes a major glucose transporter and controls glucose intake), PGD (which promotes glucose metabolism into the pentose phosphate shunt), PSAT1 (which encodes a phosphoserine aminotransferase that catalyzes the reversible conversion of 3-phosphohydroxypyruvate to phosphoserine), and LDHA (whose protein catalyzes the conversion of pyruvate to lactate) are inferred in all cancer cell lines but not in alveolar cells. (b) Many activations of genes by downstream genes are inferred, and this sort of feedback regulations is an intrinsic feature of metabolism. Especially, the controlling factor SLC2A1 is activated by multiple genes. (c) In contrast, none of these features occur in the alveolar cells (partly due to key genes such as SLC2A1 is not expressed). These inferred results are literature-supported and biologically reasonable, despite that the causal inference is flawed by the absence of metabolites in the data.



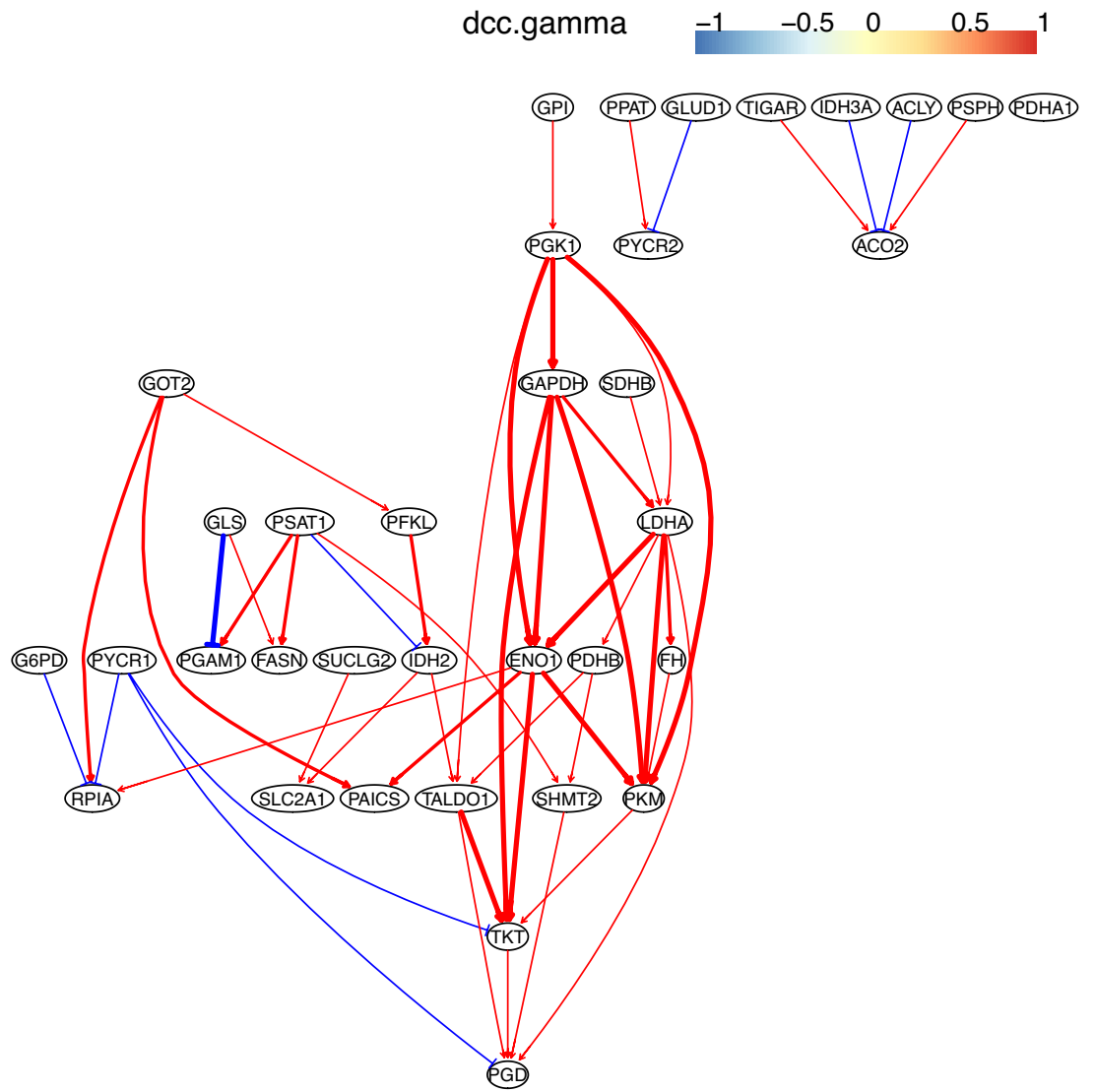
Appendix 4—figure 6. The causal relationships between genes in the WP4290 pathway in the cell line A549 inferred using PC+DCC.gamma. The inference is flawed because the true network contains both gene Appendix 4—figure 6 continued on next page

Appendix 4—figure 6 continued

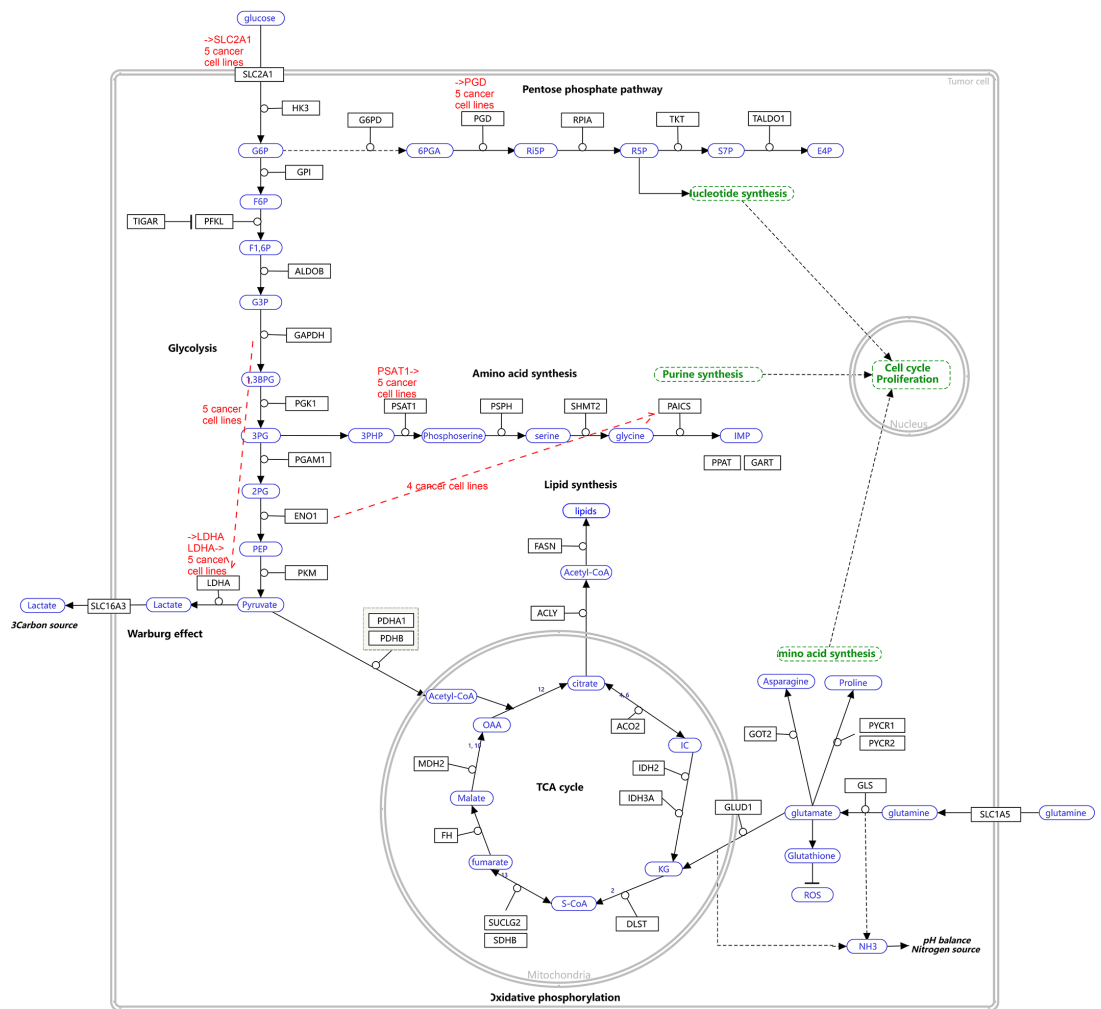
products and metabolites but single-cell RNA-sequencing (scRNA-seq) data do not contain metabolites. Nevertheless, **Appendix 4—figures 6–8** show that multiple inferred interactions reasonably reveal key features of reprogrammed glucose metabolism. First, shared interactions in ≥ 2 datasets are 21.12%, 58.82%, 50.51%, 40.82%, 53.09%, and 50.0% in alveolar cells, H838 cells, H2228 cells, HCC827 cells, H1975 cells, and A549 cells, indicating that causal inference differentiates glucose metabolism in cancer cells from in alveolar cells. Second, the inferred networks reflect key features of reprogrammed glucose metabolism, especially the activation of SLC2A1 (which encodes a major glucose transporter and controls glucose intake), PGD (which promotes glucose metabolism toward nucleotide synthesis), PSAT1 (which promotes glucose metabolism toward nucleotide synthesis), and LDHA (which promotes glucose metabolism toward lactate generation) in cancer cell lines.



Appendix 4—figure 7. The causal relationships between genes in the WP4290 pathway in the cell line H2228 inferred using PC+DCC.gamma.



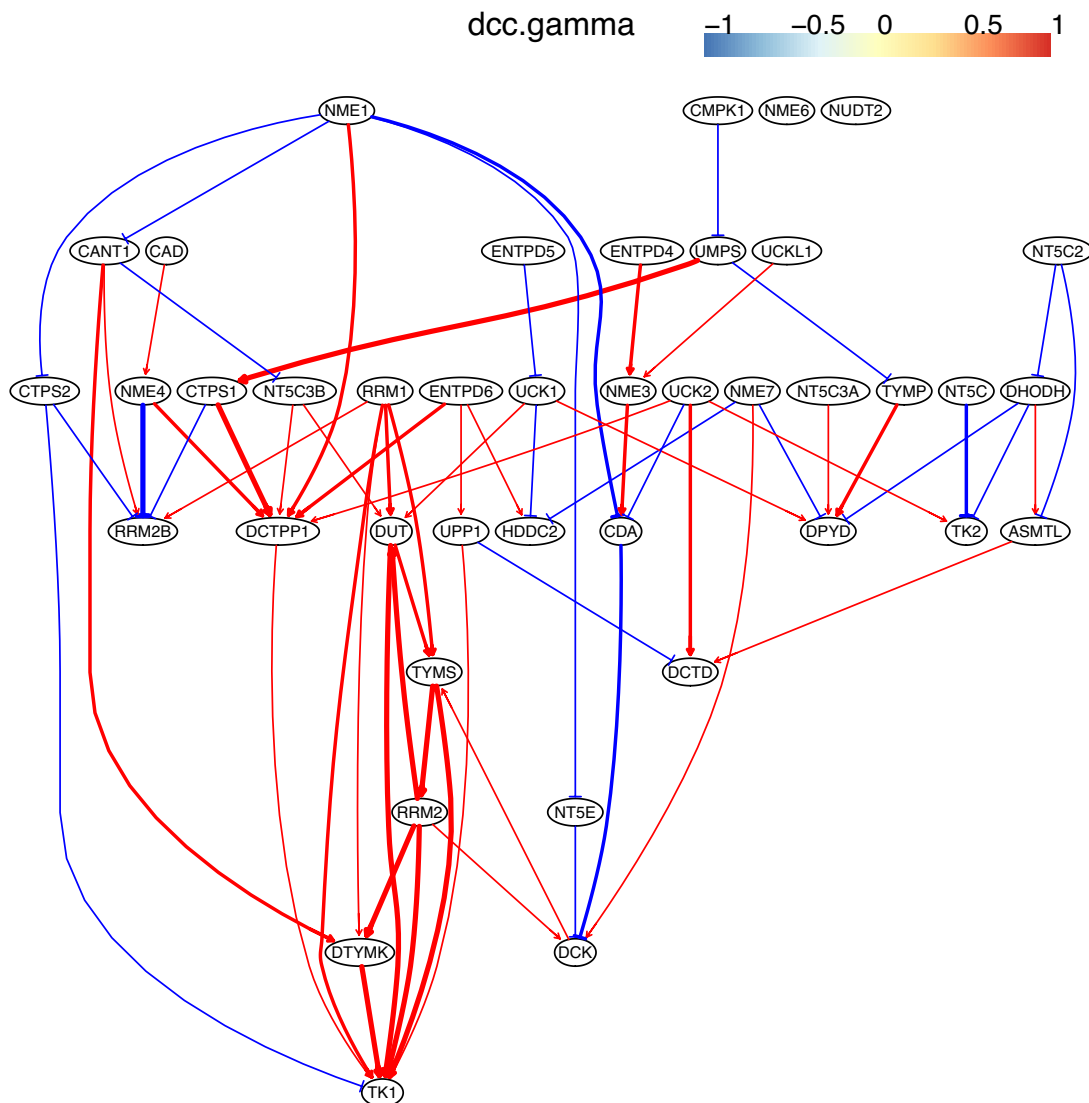
Appendix 4—figure 8. The causal relationships between genes in the WP4290 pathway in the cell line H838 inferred using PC+DCC.gamma.



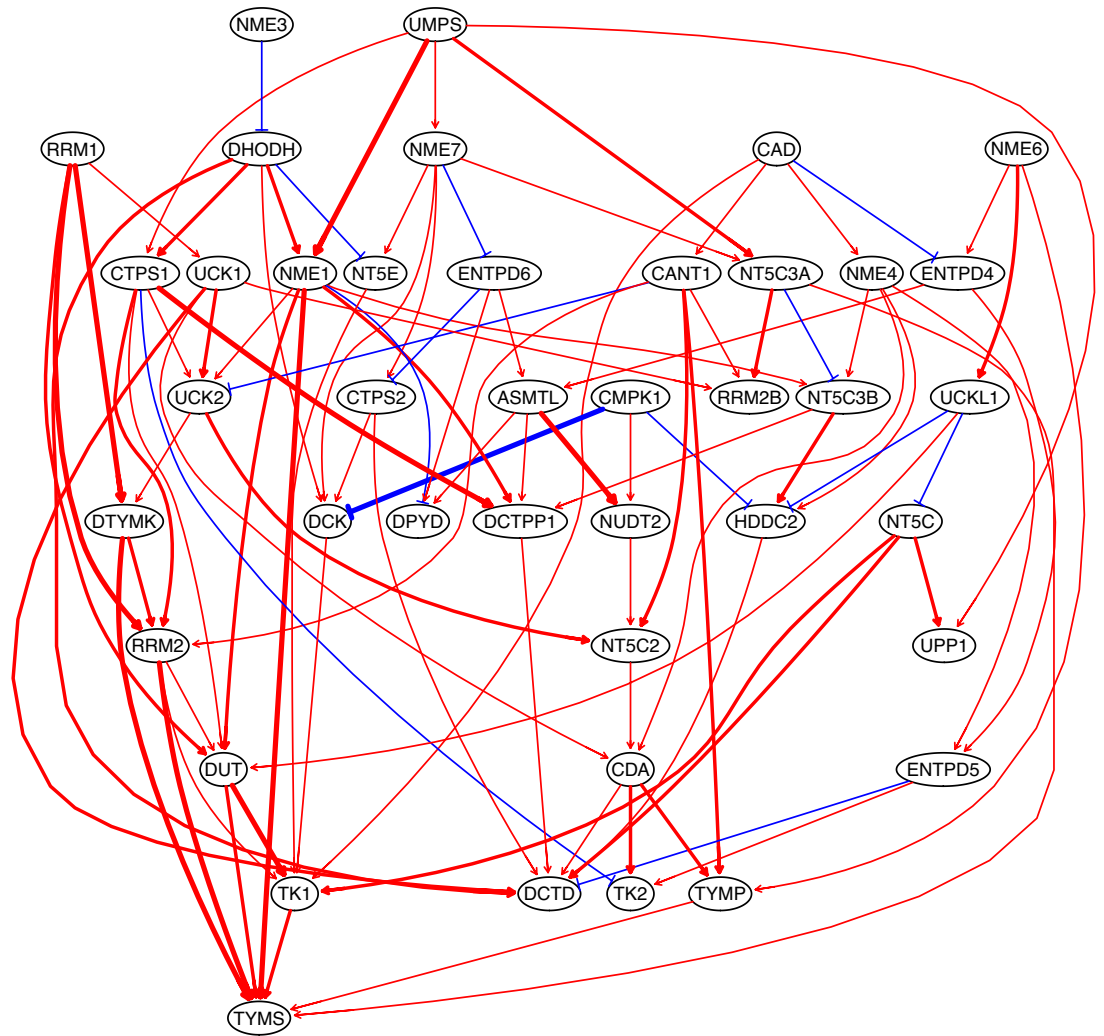
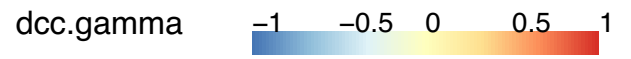
Appendix 4—figure 9. The WP4290 pathway and the key features of inferred interactions in lung cancer cell lines. Glucose intake is greatly increased in cancer cells. The increased glucose consumption is used as a carbon source for anabolic processes and this excess carbon is also used for the de novo generation of nucleotides, lipids, and proteins. This so-called Warburg effect is proposed to be an adaptation mechanism to support these biosynthesis processes for uncontrolled proliferation of cancer cells. SLC2A1, PGD, PSAT1, and LDHA are critical genes controlling glucose intake and the generation of nucleotides and lactate. The added red notes indicate key inferred interactions. ‘→XXX’ and ‘XXX→’ indicate the activation of the gene XXX by others and the activation of others by the gene XXX, respectively. In the alveolar cells, SLC2A1, PGD, PSAT1, ENO1, and LDHA are not expressed and none of these interactions are inferred.

Second, we examined the ‘Pyrimidine metabolism’ (WP4022) pathway (**Appendix 4—figures 10–13**). We used genes in the ‘Pyrimidine metabolism’ (hsa00240) to perform the inference (because WP4022 contains too many POLR gene families) and used the more readable WP4022 pathway to illustrate the results. Compared with glucose metabolism, pyrimidine metabolism has many reversible reactions, making interactions vary greatly in cells and the differences between cancer and alveolar cells opaque. The following genes and reactions are notable. (a) TYMS catalyzes dUMP→dTMP unidirectionally toward DNA synthesis. (b) Tk1/2 catalyze thymidine→dTMP and deoxyuridine→dUMP toward DNA synthesis (while NT5C/E/M do the opposite). (c) DUT catalyzes dUTP→dUMP (and dUMP is the substrate for TYMS). (d) TYMP catalyzes thymidine→thymine unidirectionally away from DNA synthesis. (e) ENTPD1/3 catalyze dTTP→dTDP→dTMP, UTP→UDP→UMP, and CTP→CDP→CMP away from DNA and RNA synthesis (but AK9/NME reverse these reactions). (f) NT5C/E/M catalyze dCMP→deoxycytidine, dUMP→deoxyuridine, and dTMP→thymidine away from DNA synthesis. Accordingly, the following interactions were inferred from cancer cell lines. (a) TYMS (the most critical gene promoting DNA synthesis) is activated in all cancer cell lines but not in alveolar

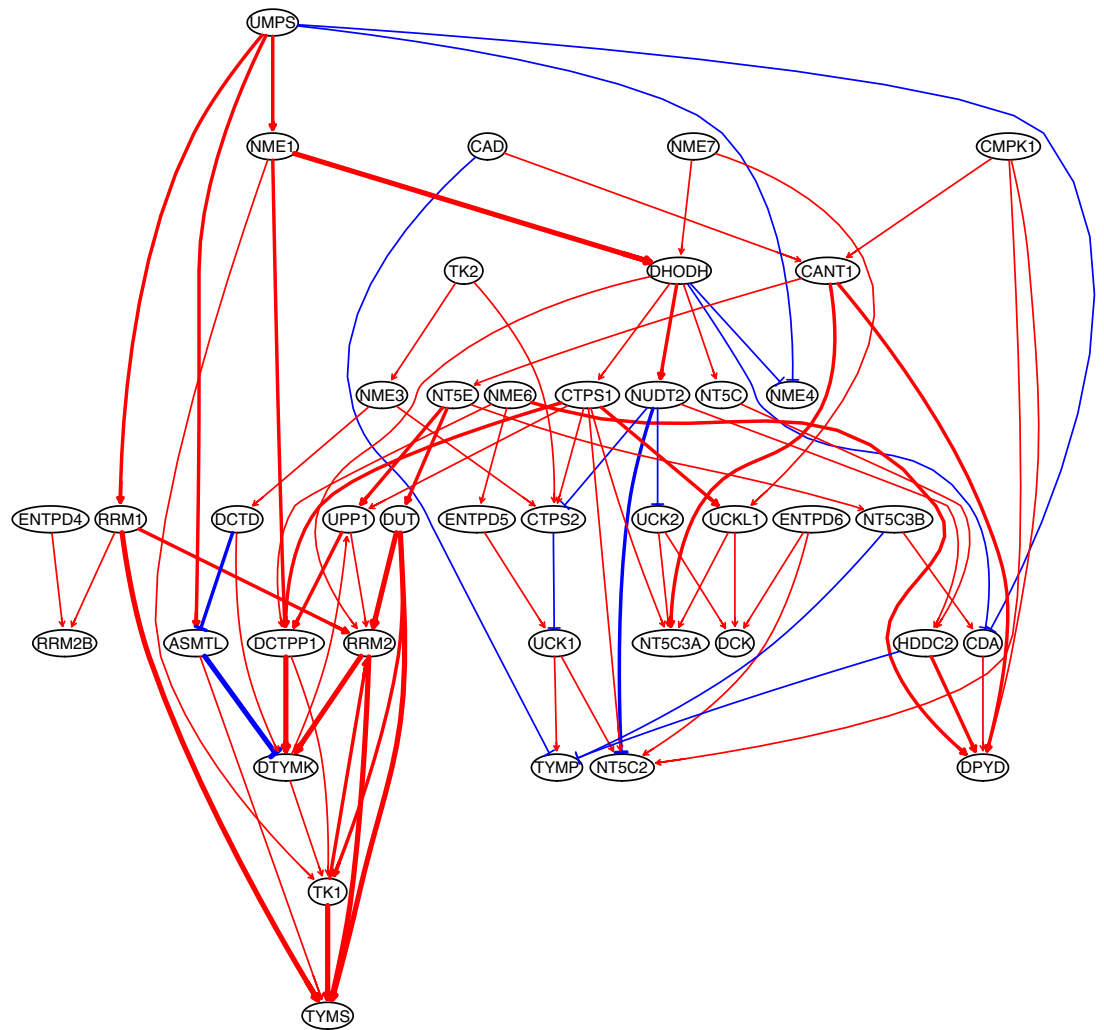
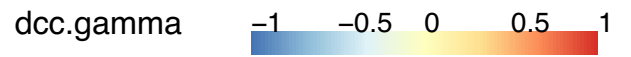
cells, and it is not repressed by any gene in cancer cell lines. (b) Tk1/2 are activated in cancer cells and alveolar cells. (c) DUT is activated in all cancer cell lines but is not expressed in alveolar cells. (d) activations of TYMP (the critical gene making reactions away from DNA synthesis) by multiple others are inferred in alveolar cells. (e) ENTPD1/3 (genes making reactions away from DNA synthesis) are activated only in alveolar cells. (f) NT5C/E/M are repressed in all cancer cell lines but are not expressed in alveolar cells. The most notable may be DUT->Tk1 and DUT->TYMS in all cancer cell lines, indicating feedforward or coordinated regulations that promote DNA synthesis. These features are literature-supported and biologically reasonable, despite that the causal inference is flawed by the absence of metabolites in the data.



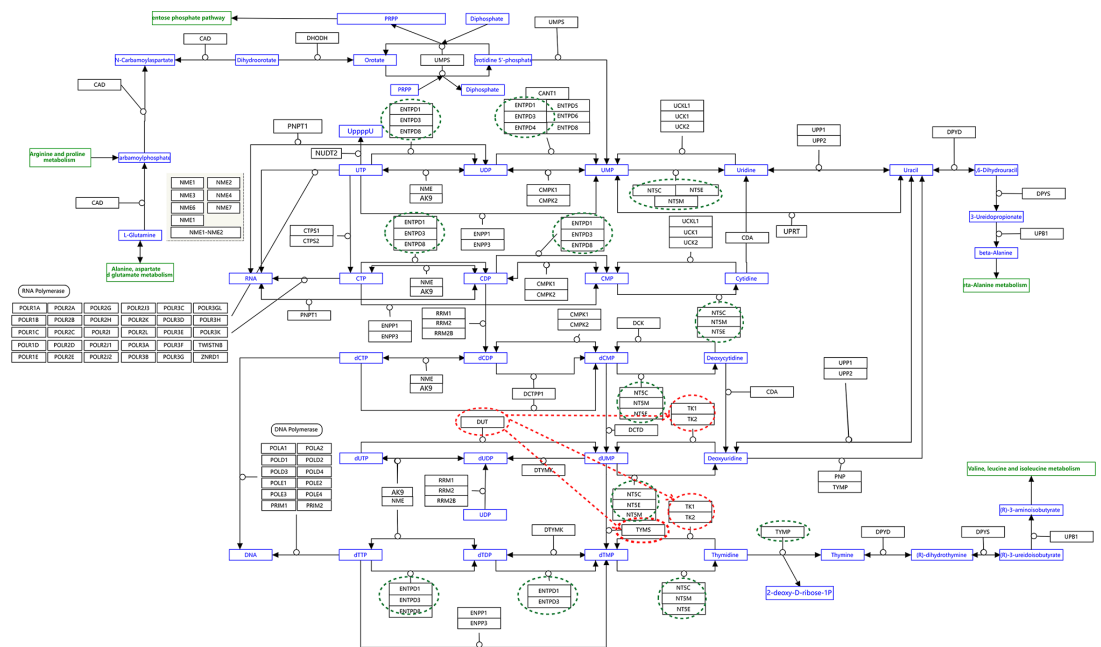
Appendix 4—figure 10. The causal interactions between genes in the hsa00240 pathway in the A549 cells. Since pyrimidine metabolism consists of many reversible reactions (Appendix 4—figure 13), inferred interactions are more varied than those of glucose metabolism. Appendix 4—figures 10–12 show that causal inference reasonably reveals the critical differences between cancer cells and alveolar cells, which include that percentages of interactions shared by ≥ 2 cell lines are 19.74%, 41.38%, 42.42%, 30.86%, 36.67%, and 32.99% in alveolar, H838, H2228, HCC827, H1975, and A549 cells. The regulations of important genes are notable (Appendix 4—figure 13). (1) TYMS is activated in five cancer cell lines but not in alveolar cells, and is not repressed in cancer cell lines. (2) Tk1/2 are activated in five cancer cells and alveolar cells. (3) DUT is not expressed in alveolar cells and is activated in the five cancer cell lines. (4) Multiple TYMP activations are inferred in alveolar cells. (5) ENTPD1/3 are activated only in alveolar. (6) NT5C/E/M are repressed in five cancer cell lines but are not expressed in alveolar cells. (7) There are many cases where downstream enzymes activate upstream enzymes, such as ENTPD3->CTPS2. Of note, DUT->Tk1 and DUT->TYMS in all five cancer cell lines indicate well-coordinated causal interactions for DNA synthesis in cancer cells.



Appendix 4—figure 11. The causal interactions between genes in the hsa00240 pathway in the HCC827 cells.



Appendix 4—figure 12. The causal interactions between genes in the hsa00240 pathway in the H1975 cells.



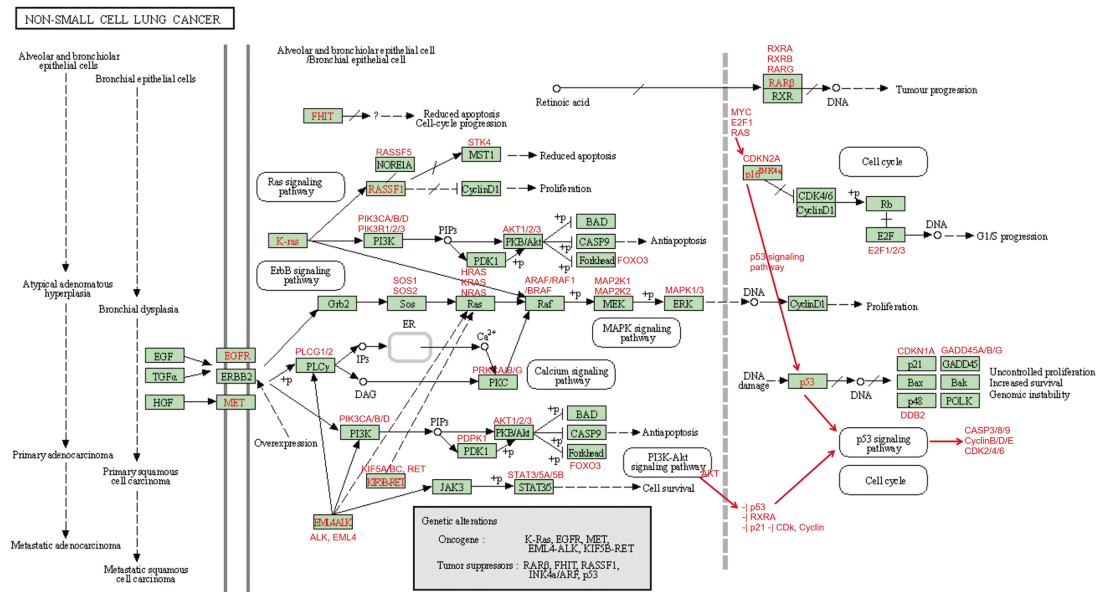
Appendix 4—figure 13. The pyrimidine metabolism pathway and key features of inferred causal networks in lung cancer cell lines and alveolar cells. Genes in the KEGG ‘Pyrimidine metabolism’ (hsa00240) pathway were used to perform causal inference (because WP4022 contains too many POLR gene families) and the figure of WP4022 was used to illustrate the results (this figure is more readable). This figure indicates that pyrimidine metabolism has many reversible reactions, and these reactions somewhat blur the key features in cancer cell lines and alveolar cells. The following genes and reactions are notable. (1) TYMS turns dUMP->dTMP unidirectionally toward DNA synthesis. (2) Tk1/2 turn thymidine->dTMP and deoxyuridine->dUMP toward DNA synthesis (while NT5C/E/M do the opposite). (3) DUT turns dUTP->dUMP, and dUMP is the substrate for TYMS. (4) TYMP turns thymidine->thymine unidirectionally away from DNA synthesis. (5) ENTPD1/3 turn dTTP->dTDP->dTMP, UTP->UDP->UMP, and CTP->CDP->CMP away from DNA synthesis and RNA synthesis (but these reactions can be reversed by AK9/NME). (6) NT5C/E/M turn dCMP->deoxycytidine, dUMP->deoxyuridine, and dTMP->thymidine away from DNA synthesis. Red and green ellipses mark genes that promote DNA synthesis and genes that do not promote DNA synthesis. In the inferred causal networks, accordingly, there are following interactions. (1) TYMS is activated in five cancer cell lines but not in alveolar cells, and is not repressed in cancer cell lines. (2) Tk1/2 are activated in five cancer cells and alveolar cells. (3) DUT is not expressed in alveolar cells and is activated in the five cancer cell lines. (4) Multiple TYMP activations are inferred in alveolar cells. (5) ENTPD1/3 are activated only in alveolar. (6) NT5C/E/M are repressed in five cancer cell lines but are not expressed in alveolar cells. (7) There are many cases where downstream enzymes activate upstream enzymes, such as ENTPD3->CTPS2. Of note, there are DUT->Tk1 and DUT->TYMS in all five cancer cell lines, indicating coordinated molecular interaction and gene regulation for DNA synthesis in cancer cells.

Third, we examined the ‘Non-small cell lung cancer’ (hsa05223) pathway (**Appendix 4—table 1; Appendix 4—figure 14**). We used the ‘graphite’ R package to turn hsa05223 into an adjacency matrix and mapped inferred interactions to the matrix. If an interaction can be mapped to an edge or a path with any directions (forward, inverse, or undirected) in hsa05223, it was assumed mapped to the pathway. hsa05223 contains sub-pathways such as p53 signaling pathway and PI3K-AKT pathway, therefore there are considerable epistatic interactions that are not annotated in hsa05223. Also, synergistic interactions (e.g. CDKN1A->BAX and EGFR->MET, see *Dong et al., 2019; Wang et al., 2014*), and many of which are literature-supported but not annotated. We additionally examined hsa05223 and sub-pathways wherein manually and found that many inferred interactions can be mapped to epistatic and synergistic interactions. Taken together, in each cell line, about 50% of inferred interactions can be mapped to the pathway. Note that this is the result without considering feedback regulations by TFs. For example, many EGF1-related interactions were inferred (e.g. E2F1->EGFR and RB1->ERBB2), but these interactions were not accounted because they are not annotated in the KEGG database. Two extra notes here. First, unlike reprogrammed glucose metabolism, common interactions between genes in different cell lines are not impressive, probably because these cell lines are generated with different genetic basis despite being derived

from NSCLC. Second, the annotation of hsa05223 has defects, because it is not in the list of enriched pathways identified by g:Profiler.

Appendix 4—table 1. The percentages of mapped edges between inferred networks and the hsa05223 pathway.

Cell lines	Inferred interactions	Num of 'forward'	Num of 'reverse'	Num of 'undirected'	Num of 'epistatic' and 'synergistic'	All
A549	82	10 (12.2%)	12 (14.63%)	2 (2.44%)	17 (20.73%)	50%
H838	77	9 (11.69%)	10 (12.99%)	1 (1.3%)	21 (27.27%)	53.25%
H1975	102	13 (12.75%)	16 (15.69%)	1 (0.98%)	17 (16.67%)	46.09%
H2228	106	13 (12.26%)	20 (18.87%)	2 (1.89%)	25 (23.58%)	56.60%
HCC827	122	13 (10.66%)	26 (21.31%)	NA (NA%)	28 (22.95%)	54.92%



Appendix 4—figure 14. The 'Non-small cell lung cancer' (hsa05223) pathway. Annotated sub-pathways, genes, and interactions are marked in red.

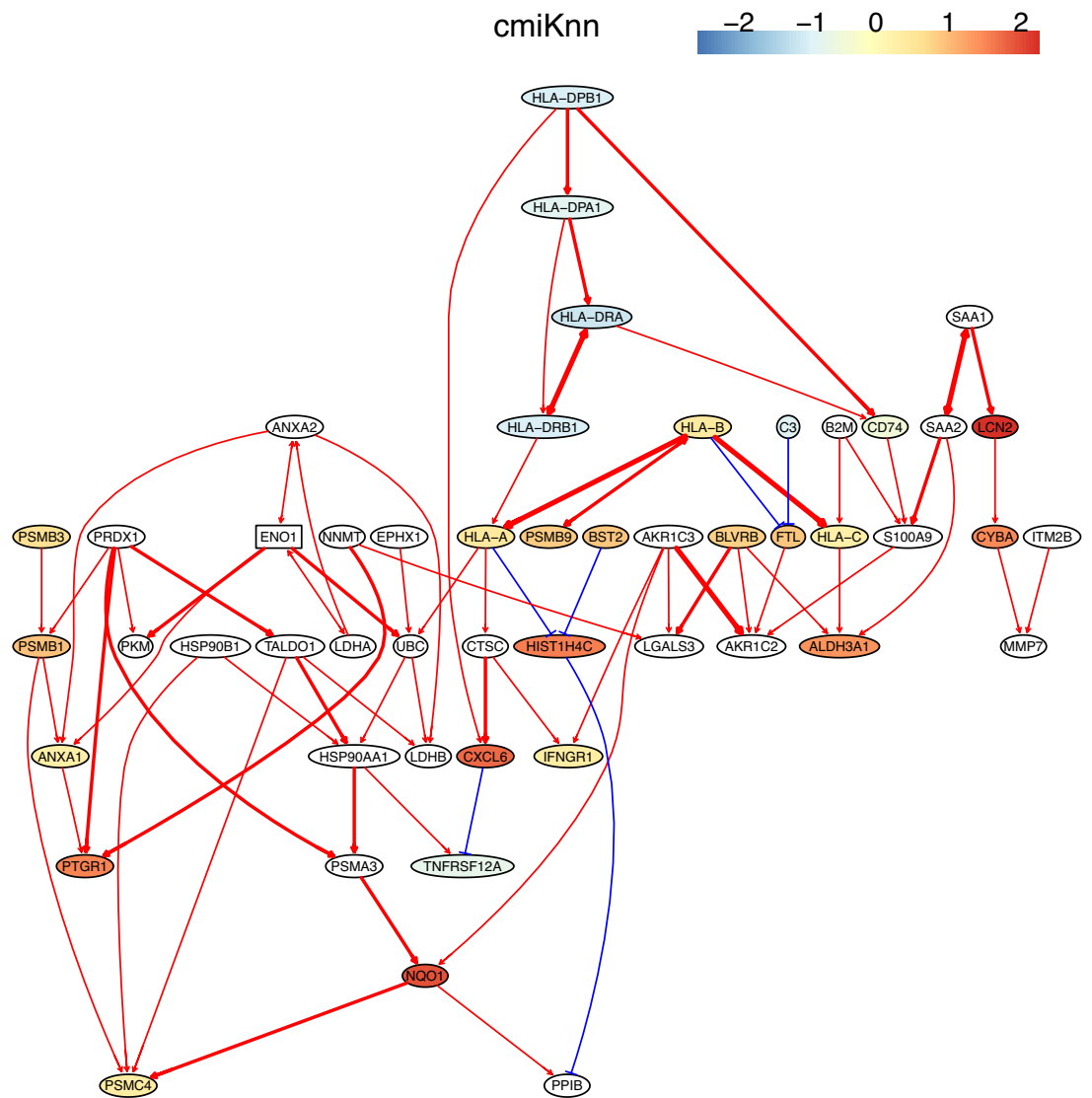
Appendix 5

Additional results of applications

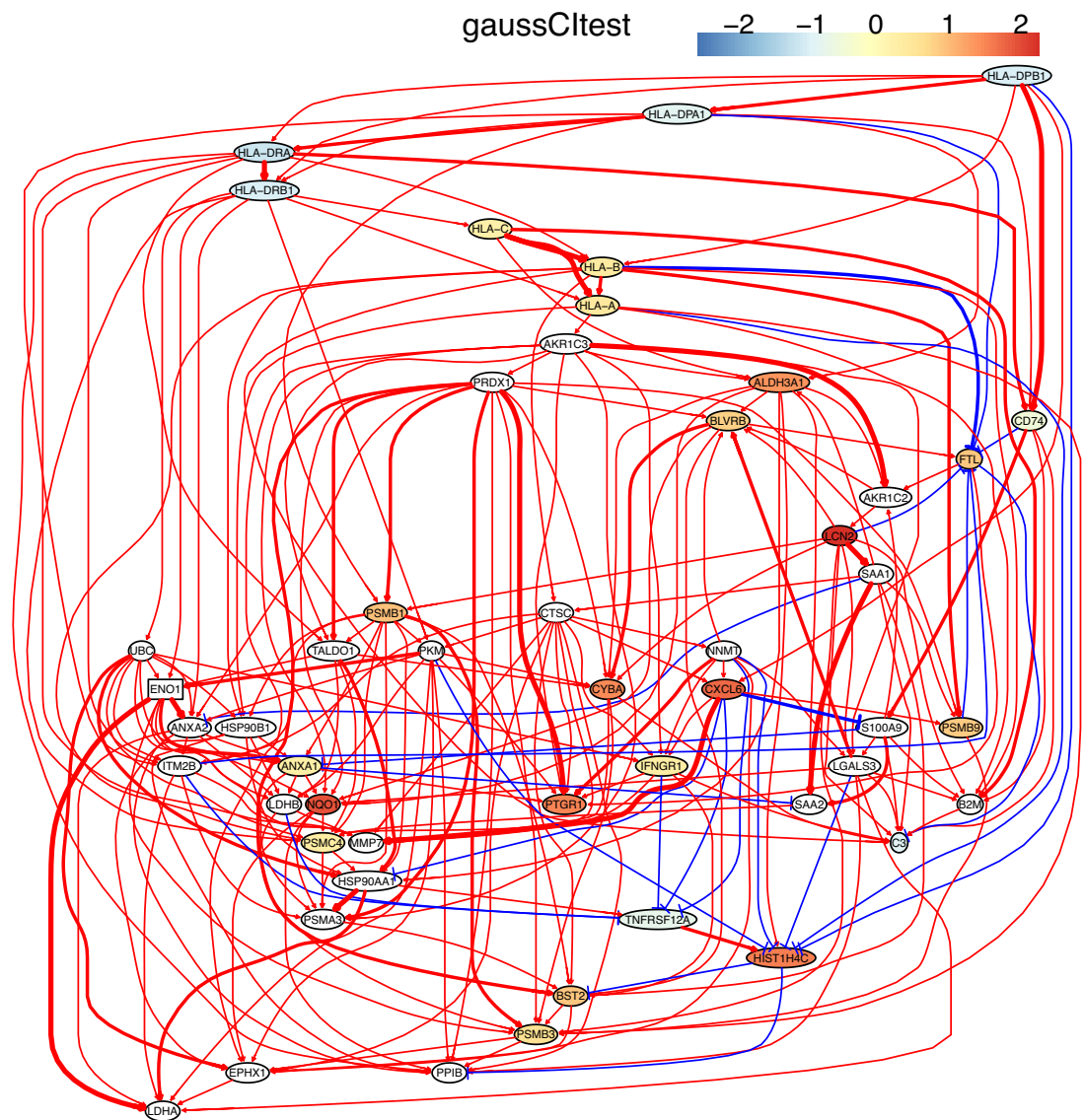
This appendix file describes the additional results of five applications, including the analysis of lung cancer cell lines and alveolar epithelial cells, the analysis of macrophages isolated from glioblastoma, the analysis of tumor-infiltrating exhausted CD8 T cells, identifying genes and inferring interactions that signify CD4 T cell aging, and the analysis of a flow cytometry dataset. These examples were used to examine the applicability of causal discovery to varied cell types and sequencing protocols. To same running time and also examine algorithms' power, varied sample sizes were used. All of these data were analyzed using the PC+CI method. The results indicate that causal discovery can be applied flexibly to varied cells. The appendix text (including appendix tables and figures) is brief and divided into five subsections, with the first four corresponding to the four subsections in the Results section in the main text, following appendix figures that are ordered accordingly.

1. The analysis of lung cancer cell lines and lung alveolar epithelial cells

As expected, feature genes and causal networks in H2228 and lung alveolar epithelial cells are distinctly different (Main text-**Figure 4**; **Appendix 5—figures 1–8**). (a) HLA Class II genes and CD74 are down-regulated in H2228 cells but up-regulated in lung alveolar epithelial cells. (b) LCN2 is up-regulated in H2228 cells but down-regulated in lung alveolar epithelial cells. (c) Algorithms inferred multiple interactions between PRDX1, TALDO1, HSP90AA1, NQO1, and PSMC4 in H2228 cells, but none of them were inferred in lung alveolar epithelial cells. (d) HLA Class I genes are feature genes in H2228 cells but not in the lung alveolar epithelial cells. HLA genes make proteins called human leukocyte antigens (HLA), which take bits and pieces of proteins from inside the cell and display them on the cell's surface. If the cell is cancerous or infected, the HLA proteins display abnormal fragments that trigger immune cells to destroy that cell. Down-regulated HLA genes may help cancer cells escape from immune cells. Annotating the networks upon related experimental findings suggest that DCC algorithms are the best and cmiKnn and GaussCltest are the poorest.

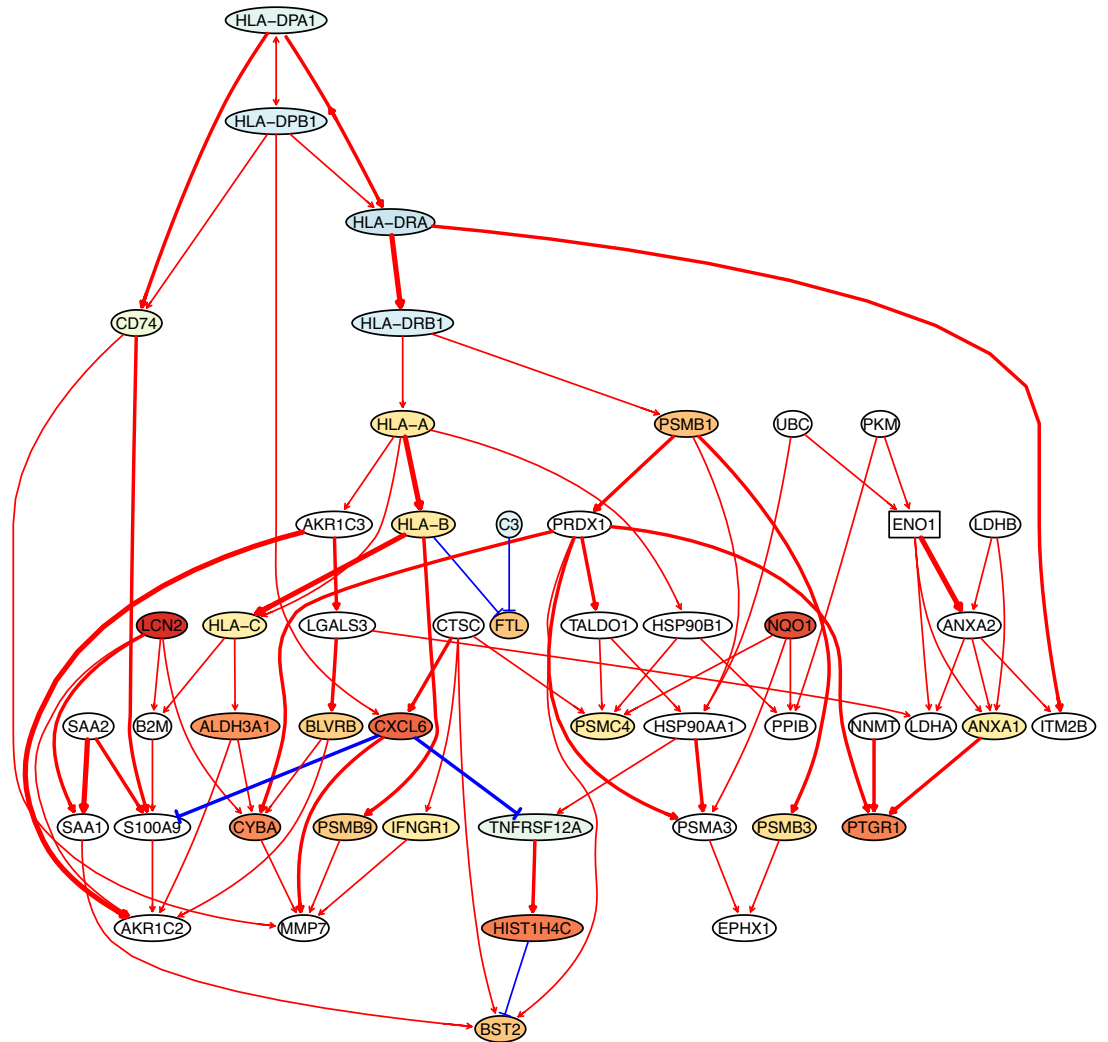


Appendix 5—figure 1. The causal network of the 50 feature genes inferred by PC+cmiKnn from the H2228 dataset (settings: feature genes were expressed in >50% cells, the alpha level for CI test was 0.1, and the 300 cells with more feature genes expressed in cells were used). In these and the following figures, red and blue arrows indicate activation and inhibition, double arrows indicate undermined direction, arrows' thickness indicates the statistical significance of CI test, and node colors indicate fold changes of gene expression. MHC-II genes were significantly down-regulated compared with the control (the lung alveolar epithelial cells).



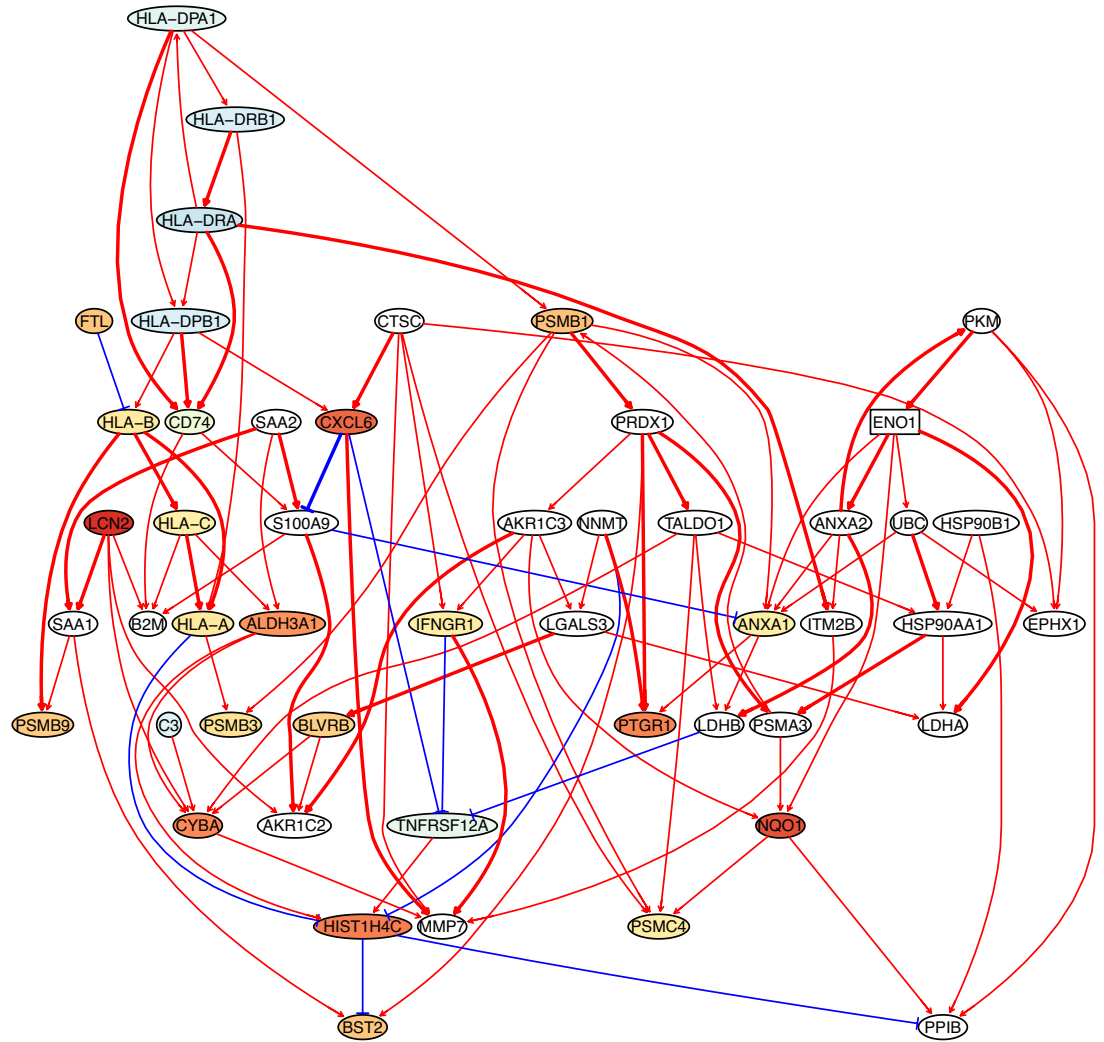
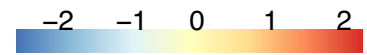
Appendix 5—figure 2. The causal network of the 50 feature genes inferred by PC+GaussCltest from the H2228 dataset (settings: feature genes were expressed in >50% cells, the alpha level for CI test was 0.1, and the 300 cells with more feature genes expressed in cells were used). MHC-II genes were significantly down-regulated compared with the control (the lung alveolar epithelial cells).

RCIT

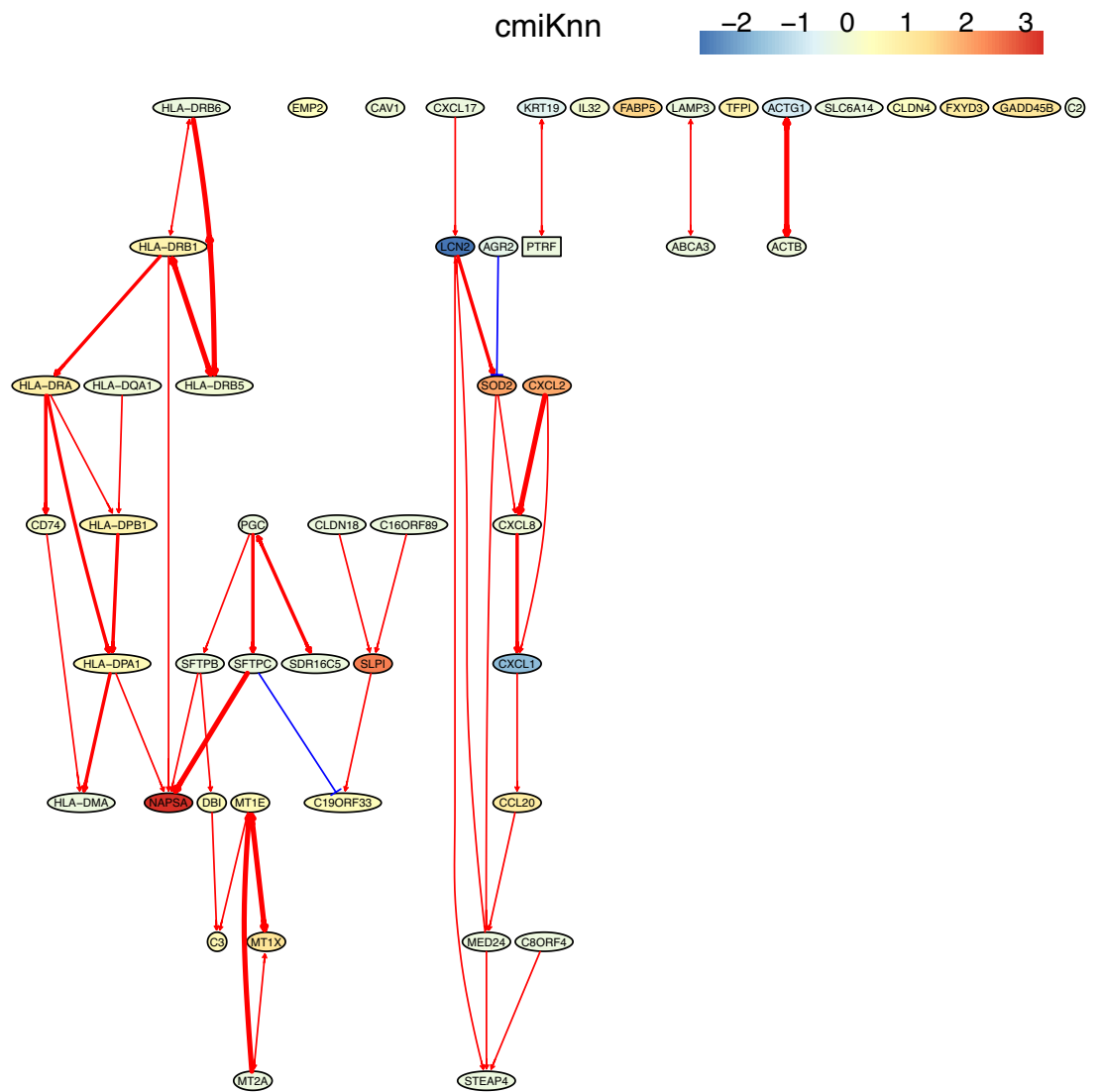


Appendix 5—figure 3. The causal network of the 50 feature genes inferred by PC+RCIT from the H2228 dataset (settings: feature genes were expressed in >50% cells, the alpha level for CI test was 0.1, and the 300 cells with more feature genes expressed in cells were used). MHC-II genes were significantly down-regulated compared with the control (the lung alveolar epithelial cells).

dcc.perm

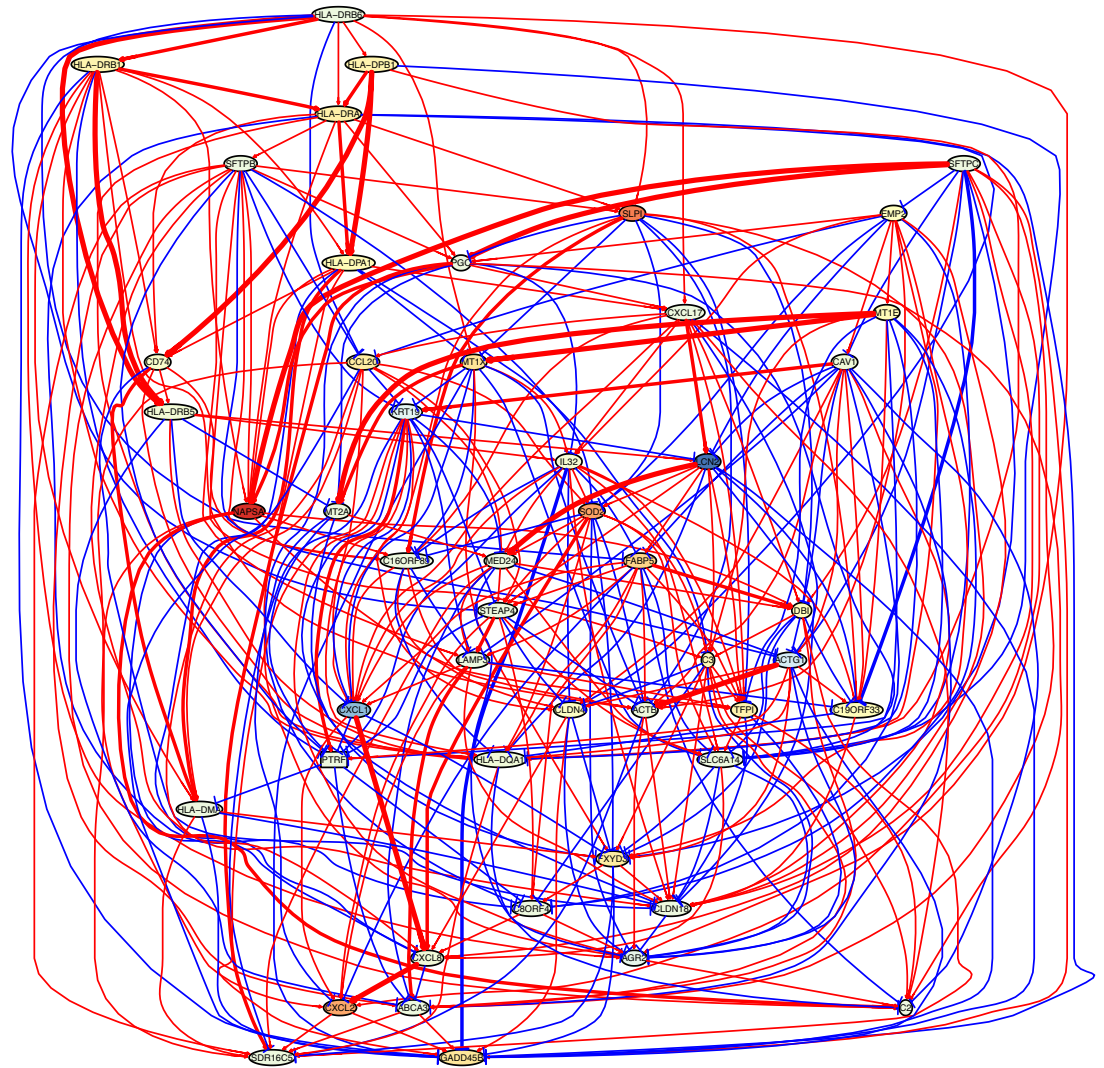
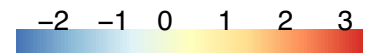


Appendix 5—figure 4. The causal network of the 50 feature genes inferred by PC+DCC.perm from the H2228 dataset (settings: feature genes were expressed in >50% cells, the alpha level for CI test was 0.1, and the 300 cells with more feature genes expressed in cells were used). MHC-II genes were significantly down-regulated compared with the control (the lung alveolar epithelial cells).

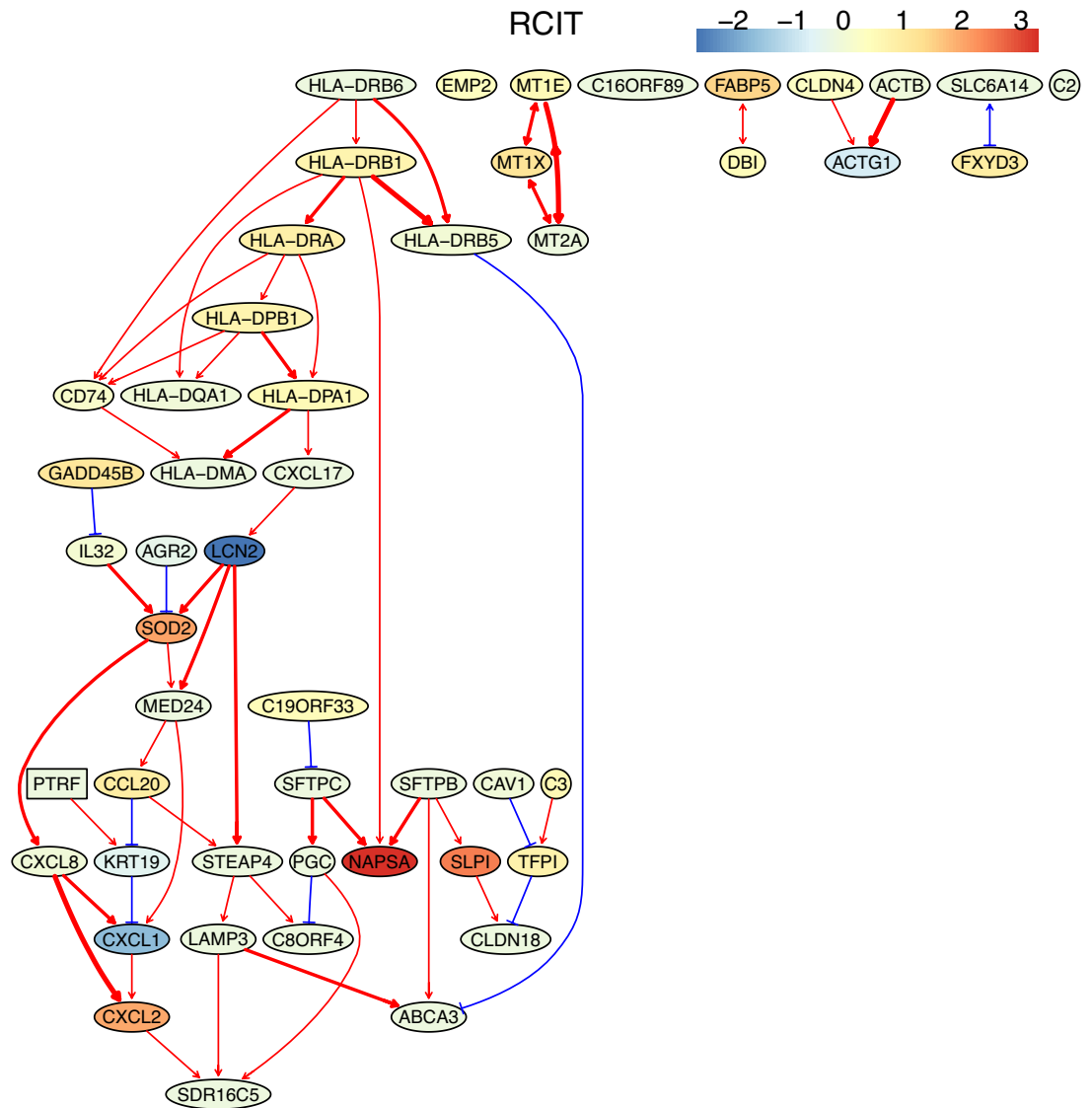


Appendix 5—figure 5. The causal network of the 50 feature genes inferred by PC+cmiKnn from the alveolar epithelial cell dataset (settings: feature genes were expressed in >50% cells, the alpha level for CI test was 0.1, and the 300 cells with more feature genes expressed the cells were used). The control of the case was the H2228 cells. Compared with H2228 cells, MHC-II genes in alveolar epithelial cells were highly expressed. The relationships between MHC-II genes and the relationships between MHC-II genes and CD74 (which is a key regulator of MHC-II proteins) are supported by annotated interactions in the STRING database.

gaussCltest

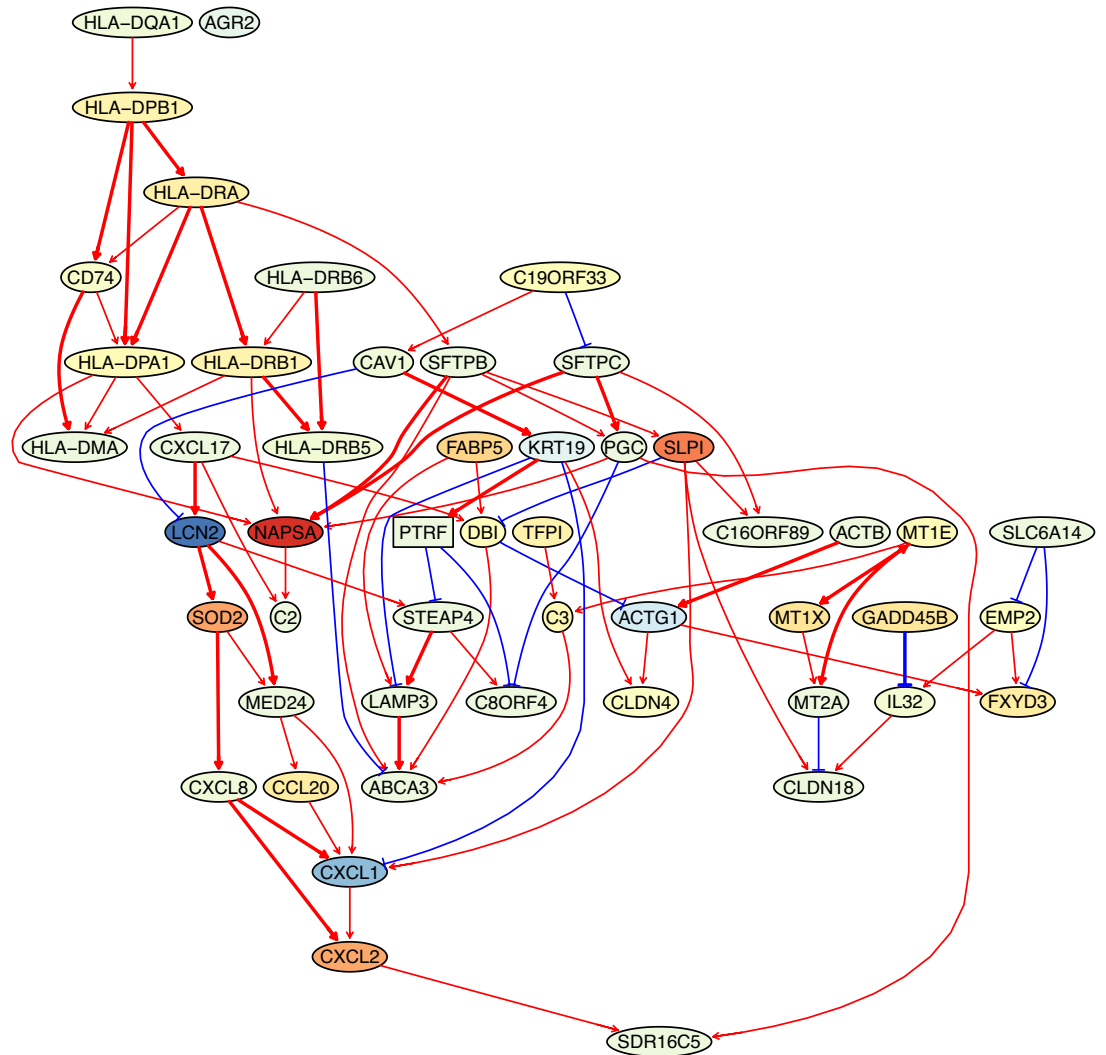


Appendix 5—figure 6. The causal network of the 50 feature genes inferred by PC+GaussCltest from the alveolar epithelial cell dataset (settings: feature genes were expressed in >50% cells, the alpha level for CI test was 0.1, and the 300 cells with more feature genes expressed the cells were used). The control of the case was the H2228 cells. Compared with H2228 cells, MHC-II genes in alveolar epithelial cells were highly expressed. The relationships between MHC-II genes and between MHC-II genes and CD74 are much more dense than those inferred by PC+cmiknn.



Appendix 5—figure 7. The causal network of the 50 feature genes inferred by PC+RCIT from the alveolar epithelial cell dataset (settings: feature genes were expressed in >50% cells, the alpha level for CI test was 0.1, and the 300 cells with more feature genes expressed the cells were used). The control of the case was the H2228 cells. Compared with H2228 cells, MHC-II genes in alveolar epithelial cells were highly expressed.

dcc.perm

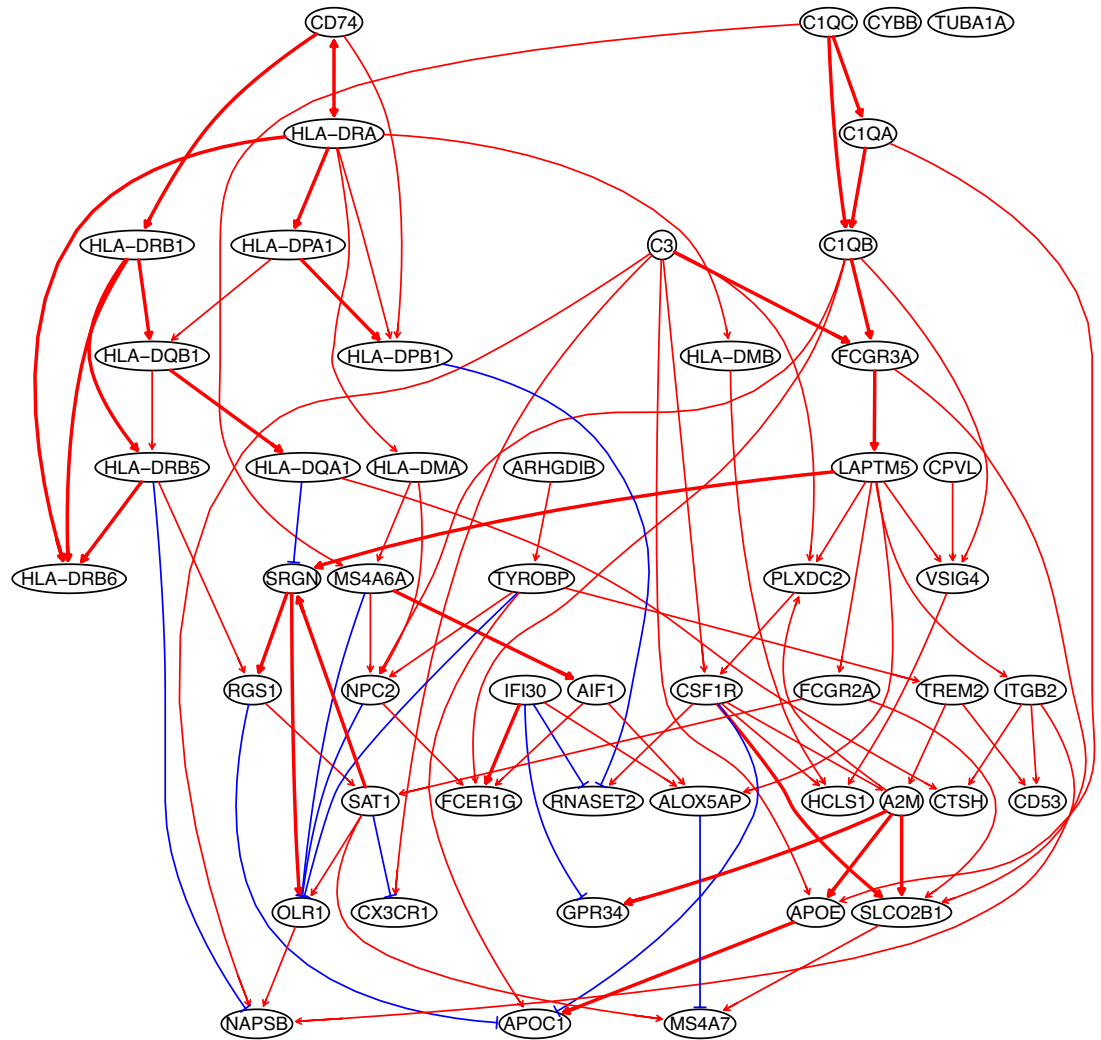
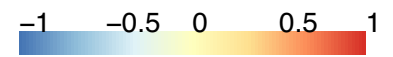


Appendix 5—figure 8. The causal network of the 50 feature genes inferred by PC+DCC.perm from the alveolar epithelial cell dataset (settings: feature genes were expressed in >50% cells, the alpha level for CI test was 0.1, and the 300 cells with more feature genes expressed the cells were used). The control of the case was the H2228 cells. Compared with H2228 cells, MHC-II genes in alveolar epithelial cells were highly expressed.

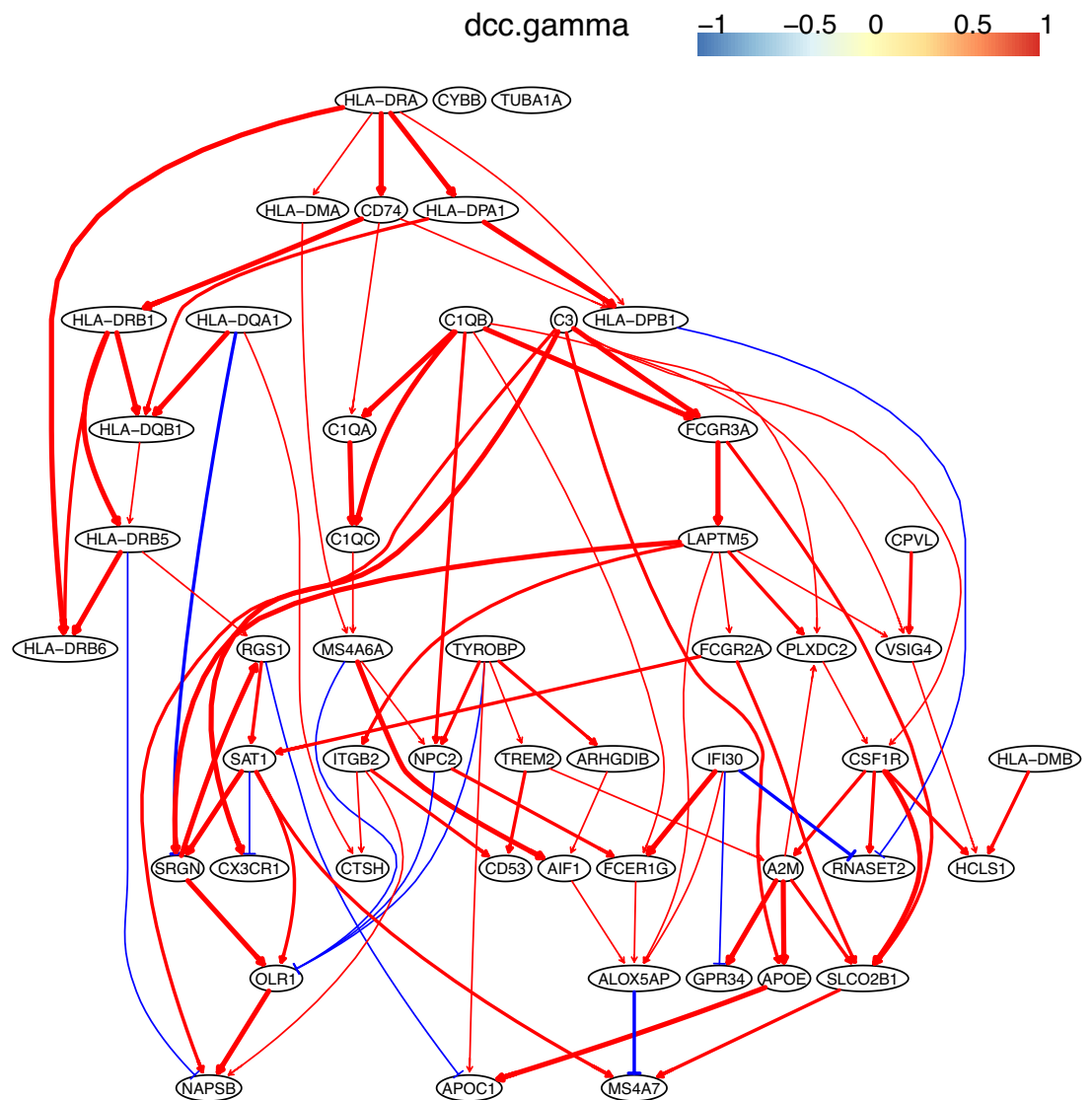
2. The analysis of the macrophages from glioblastoma

After using the dataset of macrophage isolated from glioblastoma to examine feature selection algorithms, we also used it to examine causal discovery algorithms. Again, feature genes include HLA genes to examine whether reported interactions are inferred (**Appendix 5—figures 9 and 10**).

dcc.perm



Appendix 5—figure 9. The causal network of the 50 feature genes inferred by PC+DCC.perm from the macrophage dataset (macrophages isolated from human glioblastoma) (setting: feature genes were expressed in >50% cells, the alpha level for CI test was 0.1, and the 300 cells with more feature genes expressed were used). Because no control data was used, the differential expression of genes was not computed. Note the interactions between MHC-II genes and CD74, between C1QA/B/C, and the TYROBP→TREM2→A2M→APOE→APOC1 cascade.

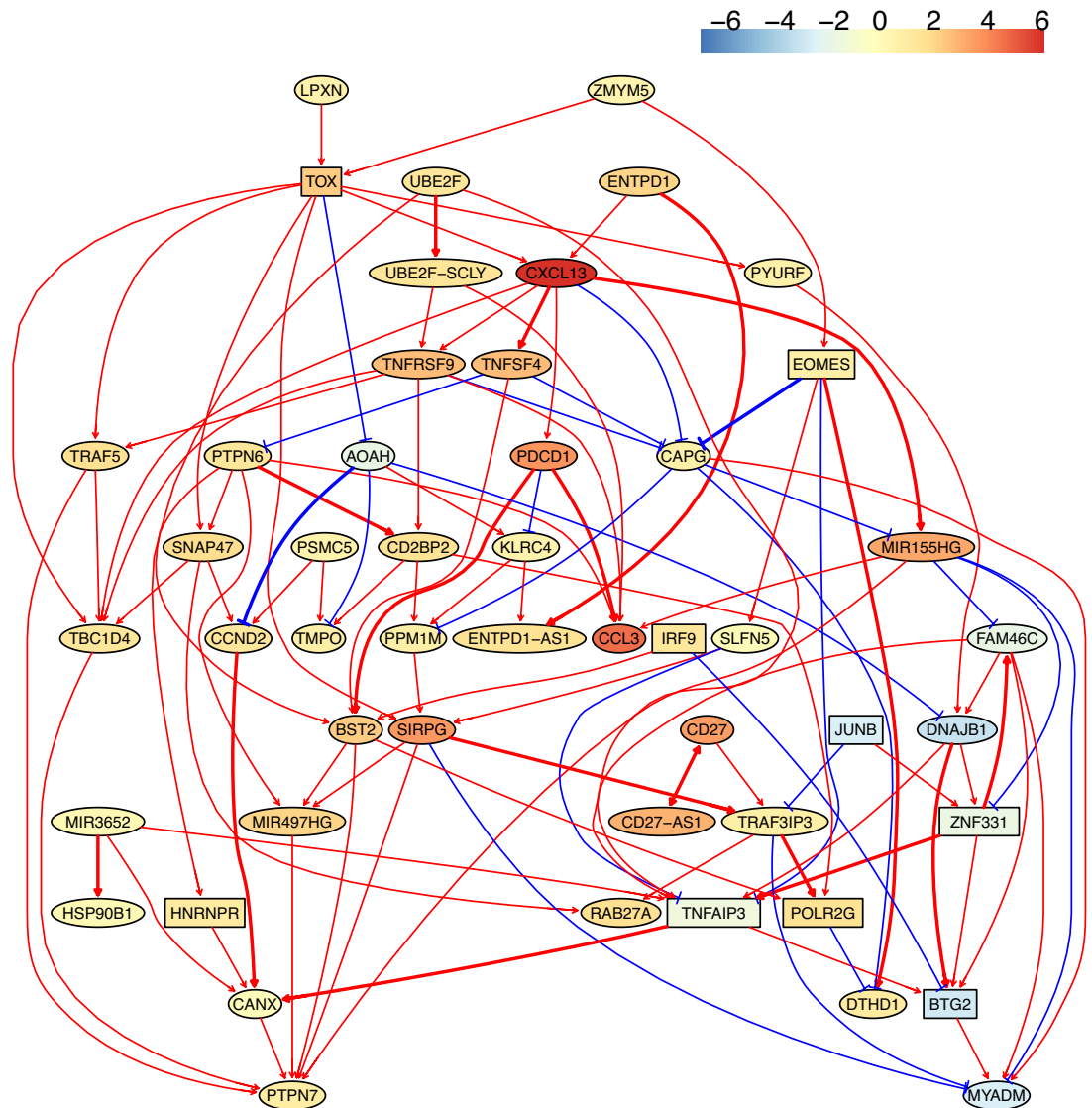


Appendix 5—figure 10. The causal network of the 50 feature genes inferred by PC+DCC.gamma from the macrophage dataset (macrophages isolated from human glioblastoma) (setting: feature genes were expressed in >50% cells, the alpha level for CI test was 0.1, and the 300 cells with more feature genes expressed were used). Because no control data was used, the differential expression of genes was not computed. Note the interactions between MHC-II genes and CD74, between C1QA/B/C, and the TYROBP→TREM2→A2M→APOE→APOC1 cascade.

3. The analysis of exhausted CD8 T cells from multiple cancers

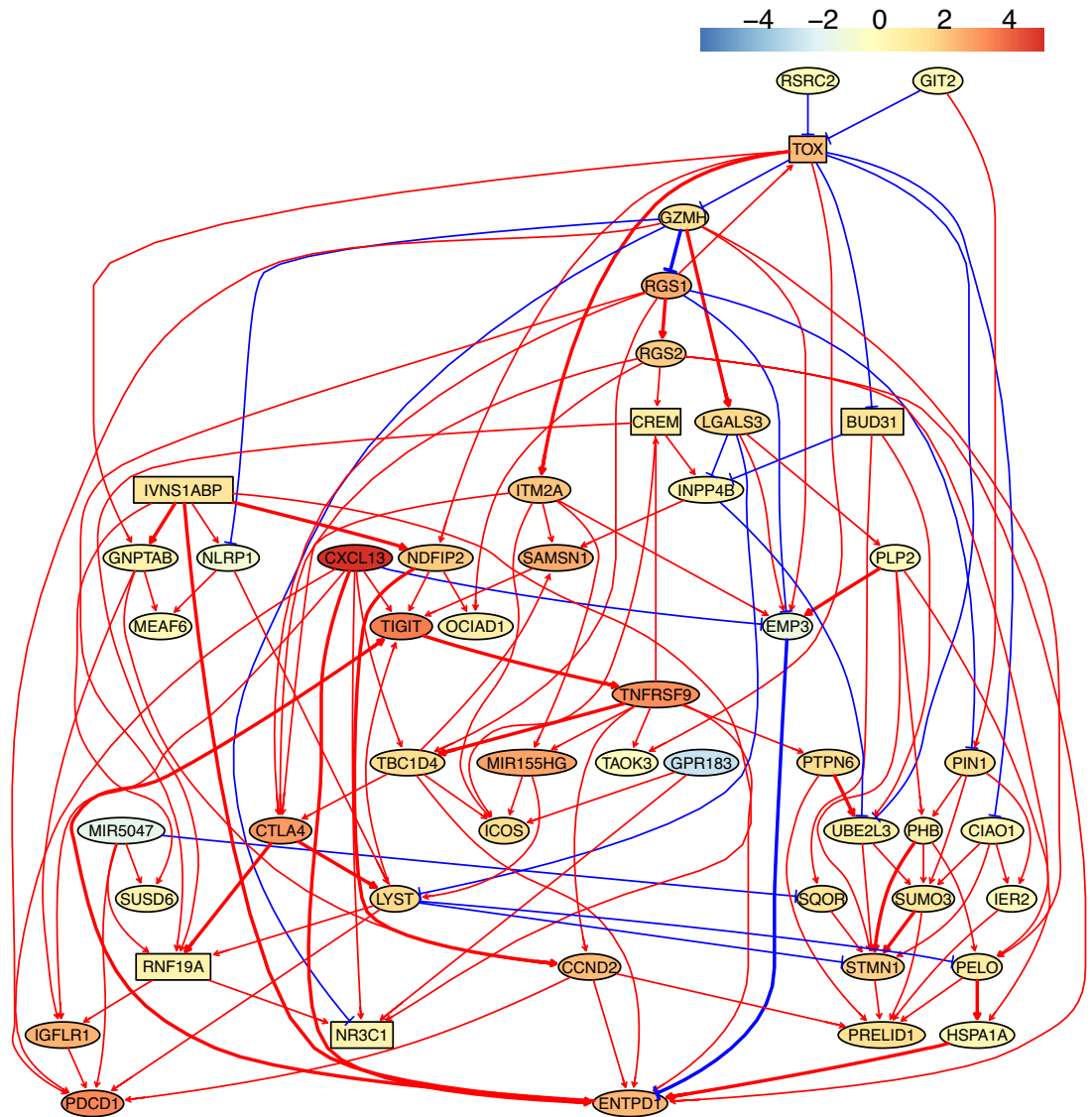
We used TOX and PDCD1 as the target gene, respectively, to select 50 genes from genes expressed in >50% exhausted CD8 T cells (from liver, colorectal, and lung cancers) and in >50% non-exhausted CD8 T cells (from the normal tissues neighboring these cancers). Networks with TOX and PDCD1 as the target gene are called TOX-network and PDCD1-network, respectively. In this application case, we demonstrate consensus networks; unless otherwise specified, all panels are consensus networks of the two DCC algorithms. Therefore, we use letters but not algorithms to label panels. Networks were inferred from 500 cells (the case of colorectal cancer) and 463 cells (the case of lung cancer). Exhausted and non-exhausted were mutually used as case and control. In panels, → and -| represent indirect activation and inhibition (**Appendix 5—figures 11–17**).

consensusNetwork (dcc.gamma + dcc.perm)



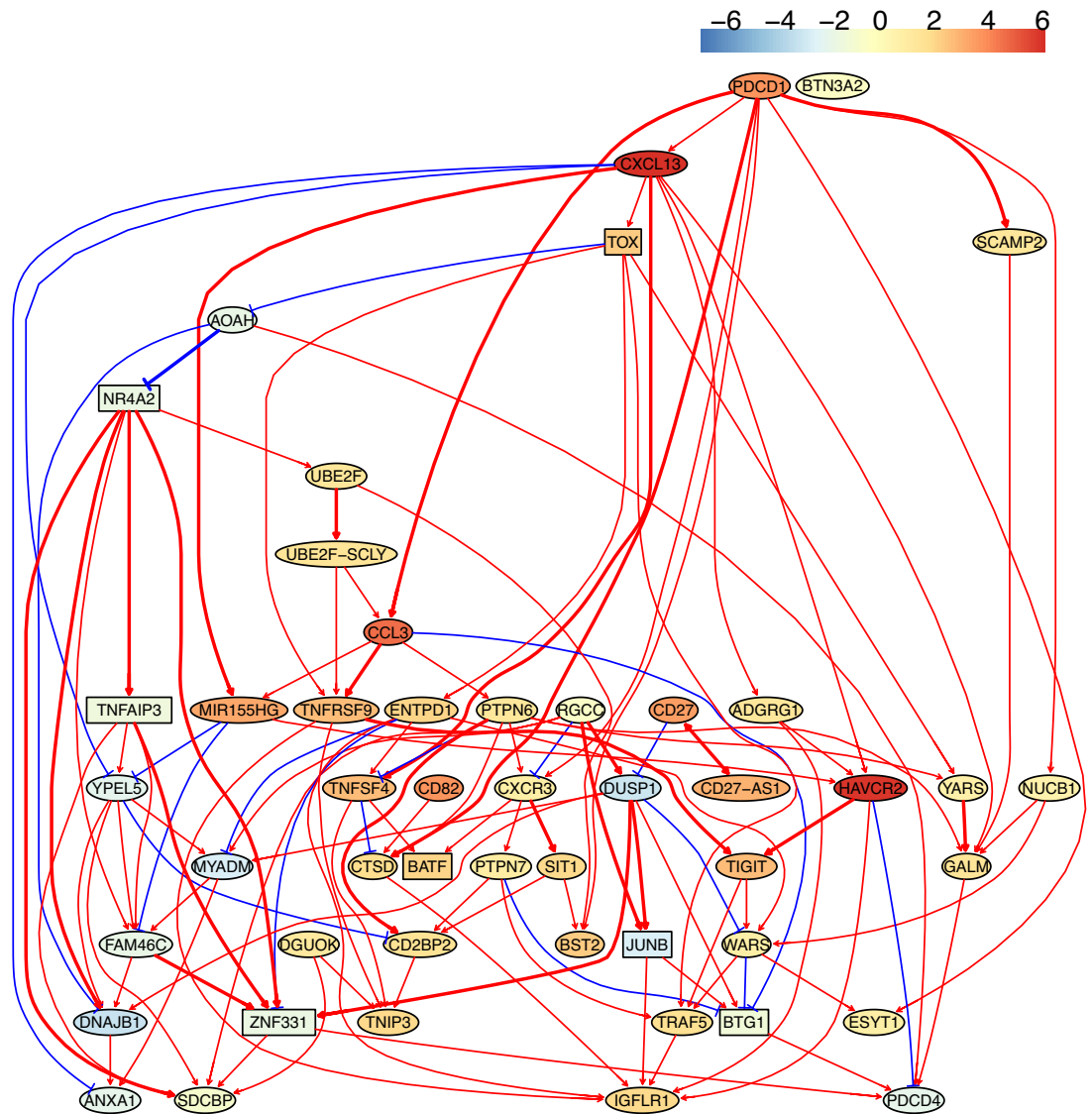
Appendix 5—figure 11. The networks of 50 genes in exhausted CD8 T cells and non-exhausted CD8 T cells from colorectal and lung cancers and the normal tissues neighboring the cancers. **(A)** The TOX-network inferred from exhausted CD8 T cells from colorectal cancer. TOX→PDCD1 (TOX→CXCL13→PDCD1), TNFRSF9→TRAF5, and TOX→MIR155HG have related reports including (1) TOX up-regulates PDCD1 expression (*Khan et al., 2019*), (2) TNF receptors bind to TRAF2/5 to activate NF-κB signaling, (3) in mice up-regulated miR-155 represses Fosl2 by inhibiting Fosb and causes long-term persistence of exhausted CD8 T cells during chronic infection (*Stelekati et al., 2018*). We found that if more feature genes were selected (to include FOSB), the MIR155HG-YPPEL5→DNAJB1→FOSB were inferred, agreeing with inhibited FOXB by up-regulated MIR155.

consensusNetwork (dcc.gamma + dcc.perm)



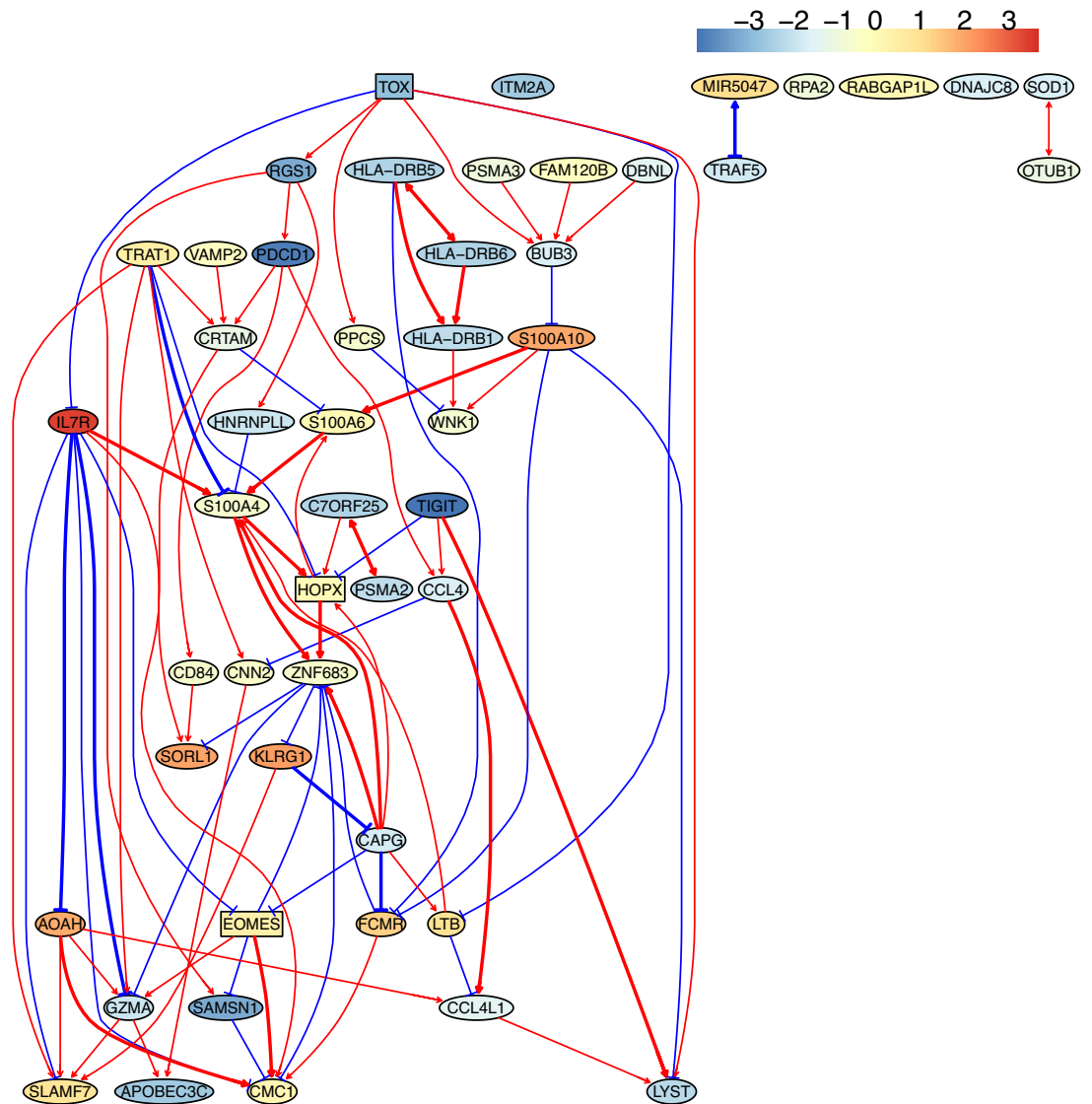
Appendix 5—figure 12. The networks of 50 genes in exhausted CD8 T cells and non-exhausted CD8 T cells from colorectal and lung cancers and the normal tissues neighboring the cancers. **(B)** The TOX-network inferred from exhausted CD8 T cells from lung cancer. Different route of TOX→PDCD1 and TOX→MIR155HG were inferred (i.e. TOX→GNPTAB→IGFLR1→PDCD1, TOX→ITM2A→MIR155HG).

consensusNetwork (dcc.gamma + dcc.perm)



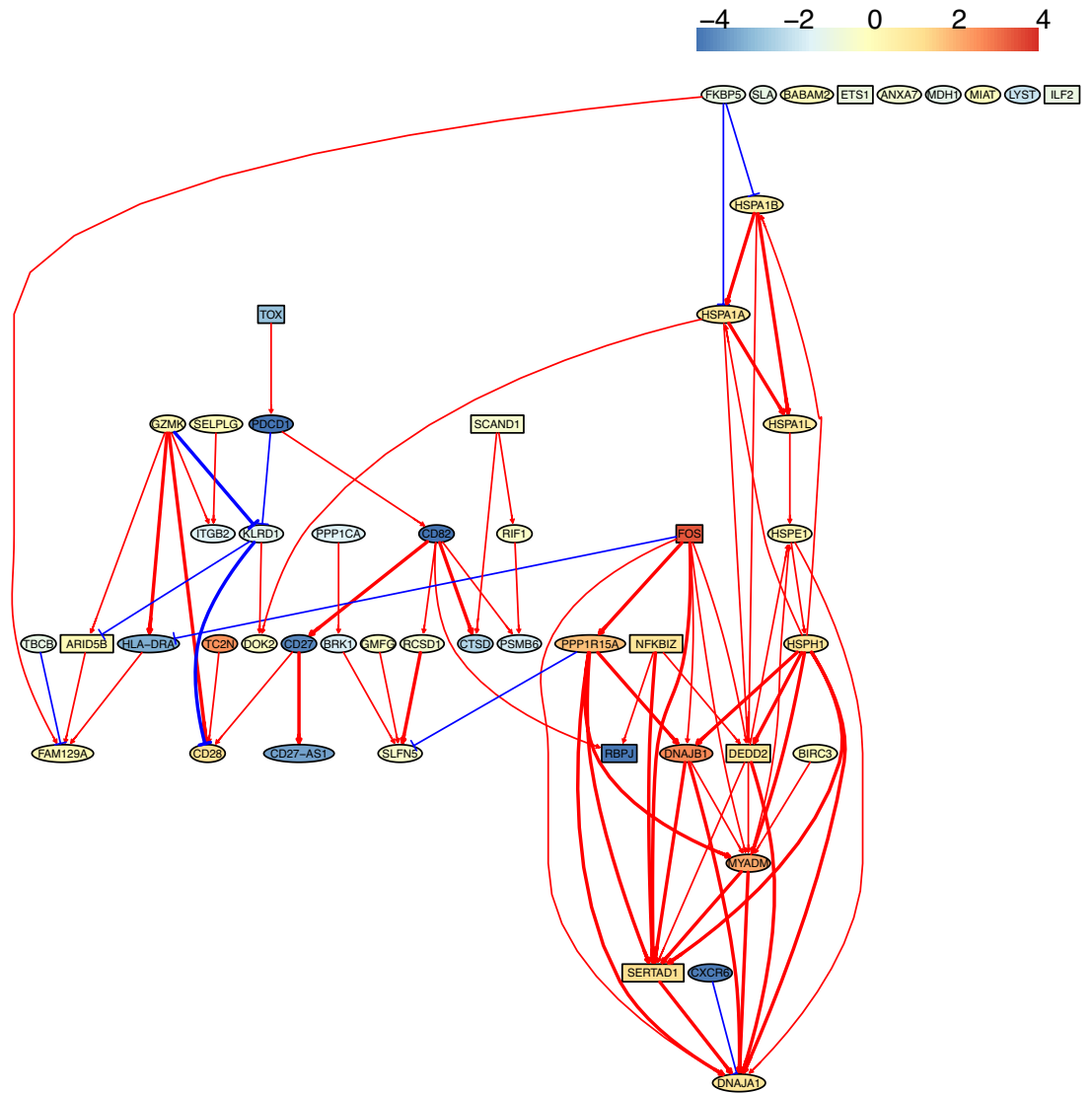
Appendix 5—figure 13. The networks of 50 genes in exhausted CD8 T cells and non-exhausted CD8 T cells from colorectal and lung cancers and the normal tissues neighboring the cancers. (C) The PDCD1-network inferred from exhausted CD8 T cells from colorectal cancer. PDCD1→TOX (PDCD1→CXCL13→TOX), PDCD1→MIR155HG (there were two routes: PDCD1→CXCL13→MIR155HG, PDCD1→CCL3→MIR155HG), MIR155HG→YPEL5→DNAJB1, and HAVCR2→PDCD4 were inferred. Related reports of these interactions include (1) TOX transcription factors cooperate with NR4A transcription factors to impose CD8+ T cell exhaustion (Seo et al., 2019), (2) CCL3 is one of the up-regulated chemokine genes in exhausted CD8 T cells (Wherry et al., 2007).

consensusNetwork (dcc.gamma + dcc.perm)



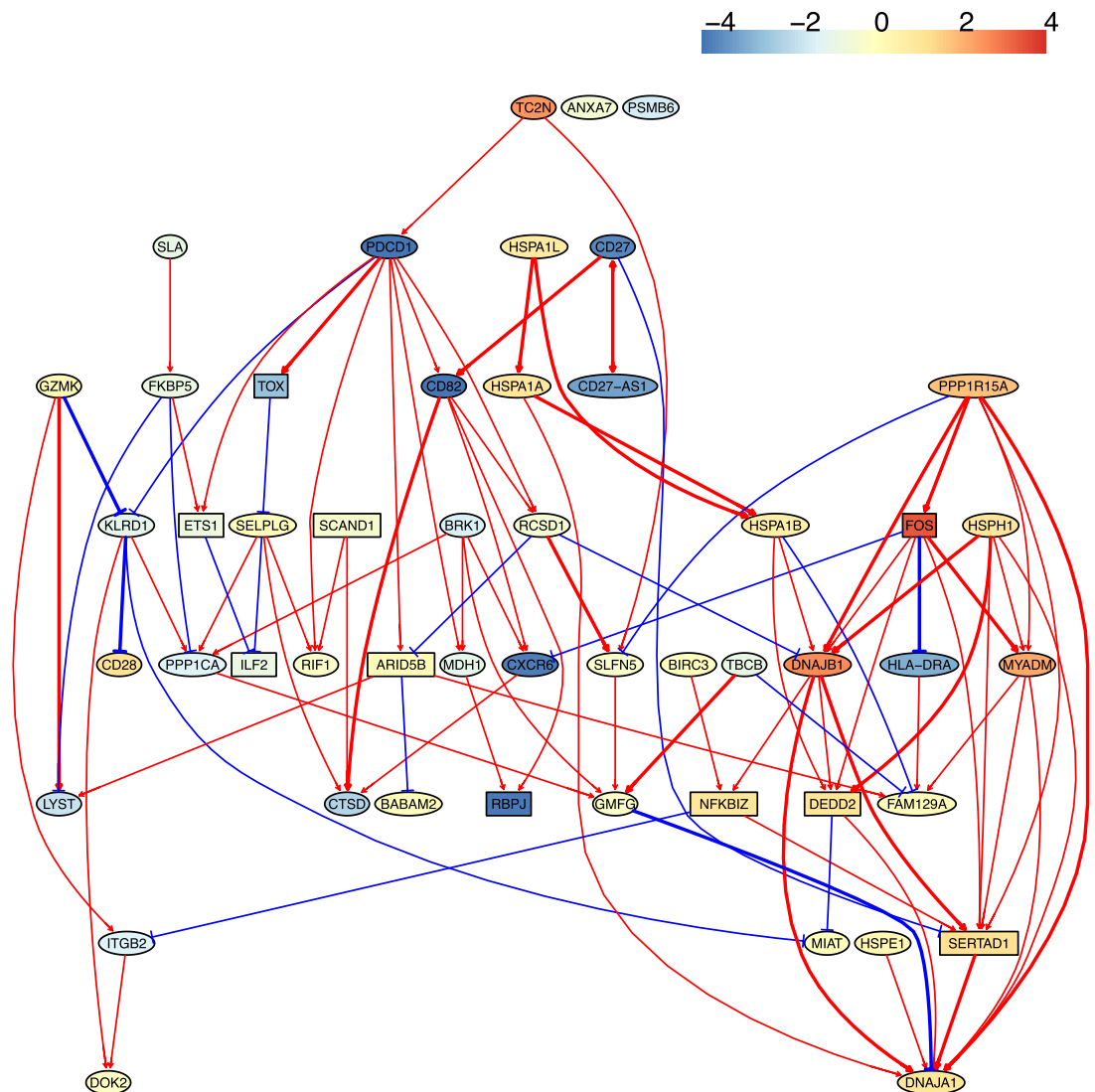
Appendix 5—figure 15. The networks of 50 genes in exhausted CD8 T cells and non-exhausted CD8 T cells from colorectal and lung cancers and the normal tissues neighboring the cancers. **(E)** The TOX-network inferred from non-exhausted CD8 T cells from the normal tissue neighboring lung cancer. TOX→RGS1→PDCD1 and interactions between HLA-DRB2, HLA-DRB6, and HLA-DRB1 were inferred.

consensusNetwork (hsic.gamma + hsic.perm)



Appendix 5—figure 16. The networks of 50 genes in exhausted CD8 T cells and non-exhausted CD8 T cells from colorectal and lung cancers and the normal tissues neighboring the cancers. (F) The PDCD1-network of HSIC algorithms inferred from non-exhausted CD8 T cells from the normal tissue neighboring colorectal cancer.

consensusNetwork (dcc.gamma + dcc.perm)



Appendix 5—figure 17. The networks of 50 genes in exhausted CD8 T cells and non-exhausted CD8 T cells from colorectal and lung cancers and the normal tissues neighboring the cancers. (G) The PDCD1-network inferred from non-exhausted CD8 T cells from the normal tissue neighboring colorectal cancer.

4. The analysis of CD4 T cells from young and old mice

Since aging occurs gradually and ubiquitously in almost all cells, we assumed that consistent up- or down-regulation in all CD4 T cell types better defines CD4 aging-related genes than large fold changes. Upon this, we obtained the presumably CD4 aging-related genes (**Appendix 5—table 1**). Many of these genes are not the senescence signatures (*Gorgoulis et al., 2019*), indicating that different genes may be involved in the aging of different cells, but the mitochondrial genes have been well recognized as being important for aging in many cells.

Data in the STRING database support many inferred interactions, especially interactions between the mitochondrial genes, between *Ccnd2*, *Ccnd3*, *Cdkn1b*, and *Cdkn2d*, between *B2m* and *H2-Q7*, between *Lck* and *Cd28*, and between *Gm9843* and *Rps27rt* (**Appendix 4—figure 3**). Interactions supported by experimental findings include *Cdc42*→*Coro1a* (*CDC42* and *CORO1A* exhibit strong associations both with age) (*Kerber et al., 2009*), *Arpc1b*→*Coro1a* (in mouse T cells *Coro1a* is involved in Arp2/3 regulation) (*Shiow et al., 2008*), *B2m*→*H2-Q7* (*B2m* is associated with the MHC class I heavy chain) (*Smith et al., 2015*), *Lck*→*Cd28* (*Lck* is found to associate with *CD28* by using its

SH2 domain to bind to a phospho-specific site) (Rudd, 2021), Cdc42-|Lamtor2 (mTOR is required for asymmetric division through small GTPases in mouse oocytes) (He et al., 2013; Lee et al., 2012), Ccnd2-|Lamtor2 (mTORC1 activation regulates beta-cell mass and proliferation by modulation of Ccnd2 synthesis and stability) (Balcazar et al., 2009), and Sub1-|Ccnd2-|Lamtor2 (Sub1 can accelerate aging via disturbing mTOR-regulated proteostasis) (Chen et al., 2021).

Several inferred results are noticeable. First, interactions among the mitochondrial genes were inferred in all cases, whose expression levels were low in cells from young mice but high in cells from old mice. These indicate that these genes may be common biomarkers of aging for CD4 T cells. Second, in the inferred networks, these mitochondrial genes do not have consistent inputs and outputs, which can probably be explained by the finding that the metabolic system undergoes extensive rewiring upon normal T-cell activation and differentiation (Zhang et al., 2021b). With the report of increasing experimental findings, mitochondrial dysfunction in aging and diseases of aging has drawn increasing attention (Haas, 2019). Third, Junb is activated. Persistent JUNB activation in human fibroblasts enforces skin aging and the AP-1 family TFs (including FOSL2 and JUNB) are increased in all immune cells during aging (Maity et al., 2021; Zheng et al., 2020). The findings of Junb/JUNB indicate that JUNB/Junb plays a critical role in aging. Fourth, the Gm9843→Rps27rt→Junb cascade (Rps27rt is also called Gm9846, and both Gm9843 and Rps27rt are mouse-specific genes) was inferred in many cases; it is interesting whether these interactions' counterparts exist in humans (Appendix 5—figures 18–21).

Appendix 5—table 1. Up- and down-regulated genes in CD4 T cells from young and old mice.

Gene	FC >0 cases	Annotation and evidence	References
Rpl28	24	Ribosomal proteins influence aging.	Kirkland et al., 1993; Steffen and Dillin, 2016
Arpc1b	24	Arpc1b may induce senescence in a p53-independent manner.	Li et al., 2022; Yun et al., 2011
Smc4	24		Goronzy and Weyand, 2019; McCartney et al., 2021
Sub1	23	Sub1 is increased and becomes activated with age, and transgenic expression of PC4 disturbs mTOR-regulated proteostasis and causes global accelerated aging.	Chen et al., 2021
Cdc37	23		
Lck	23	Lck is a positive regulator of inflammatory signaling and a potential treatment target for age-related diseases.	Garcia and Miller, 2009; Kim et al., 2019
Cdc42	23	Mouse model studies have found that aging is associated with elevated activity of the Rho GTPase Cdc42 in hematopoietic stem cells. In humans, CDC42 and CORO1A exhibited strong associations with age.	Amoah et al., 2021; Geiger and Zheng, 2013; Kerber et al., 2009
Ccnd2	22	Ccnd2 is an aging marker	Goronzy and Weyand, 2019; McCartney et al., 2021
Ccnd3	22	Ccnd2 is an aging marker	Goronzy and Weyand, 2019; Li et al., 2020; McCartney et al., 2021
Foxp1	22	FOXP1 controls mesenchymal stem cell commitment and senescence during skeletal aging.	Li et al., 2017
Coro1a	21	CORO1A is a senescence-related gene.	Avelar et al., 2020; Kerber et al., 2009
Gm26740	21	A mouse-specific gene without annotation.	
Lamtor2	20	MAPK and MTOR activator 2. It is involved in the activation of mTORC1.	Morita et al., 2017; Walters and Cox, 2021
Lsp1	20	Lymphocyte-specific protein 1; may play a role in mediating neutrophil activation and chemotaxis.	
Gene	FC <0 cases	Annotation and evidence	References

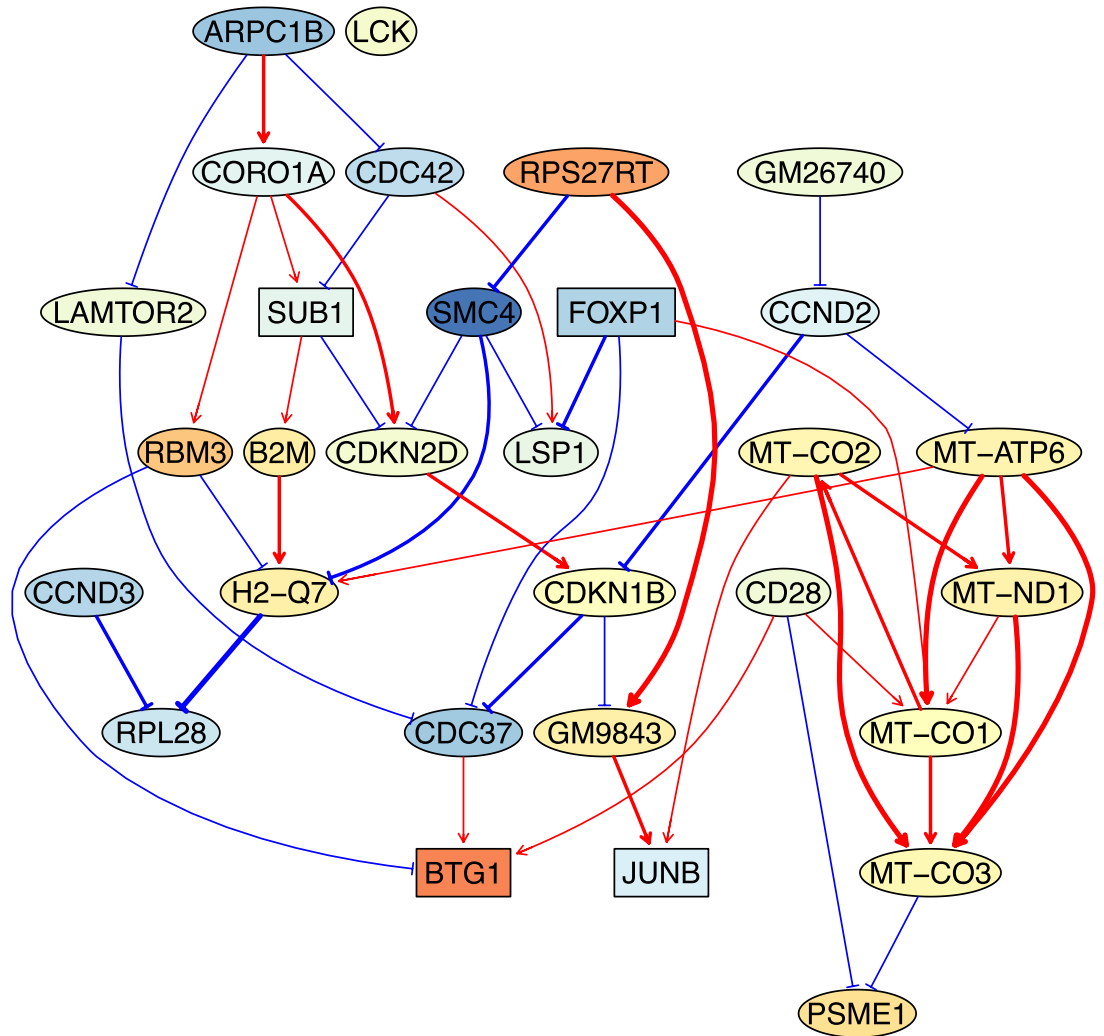
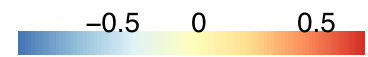
Appendix 5—table 1 Continued on next page

Appendix 5—table 1 Continued

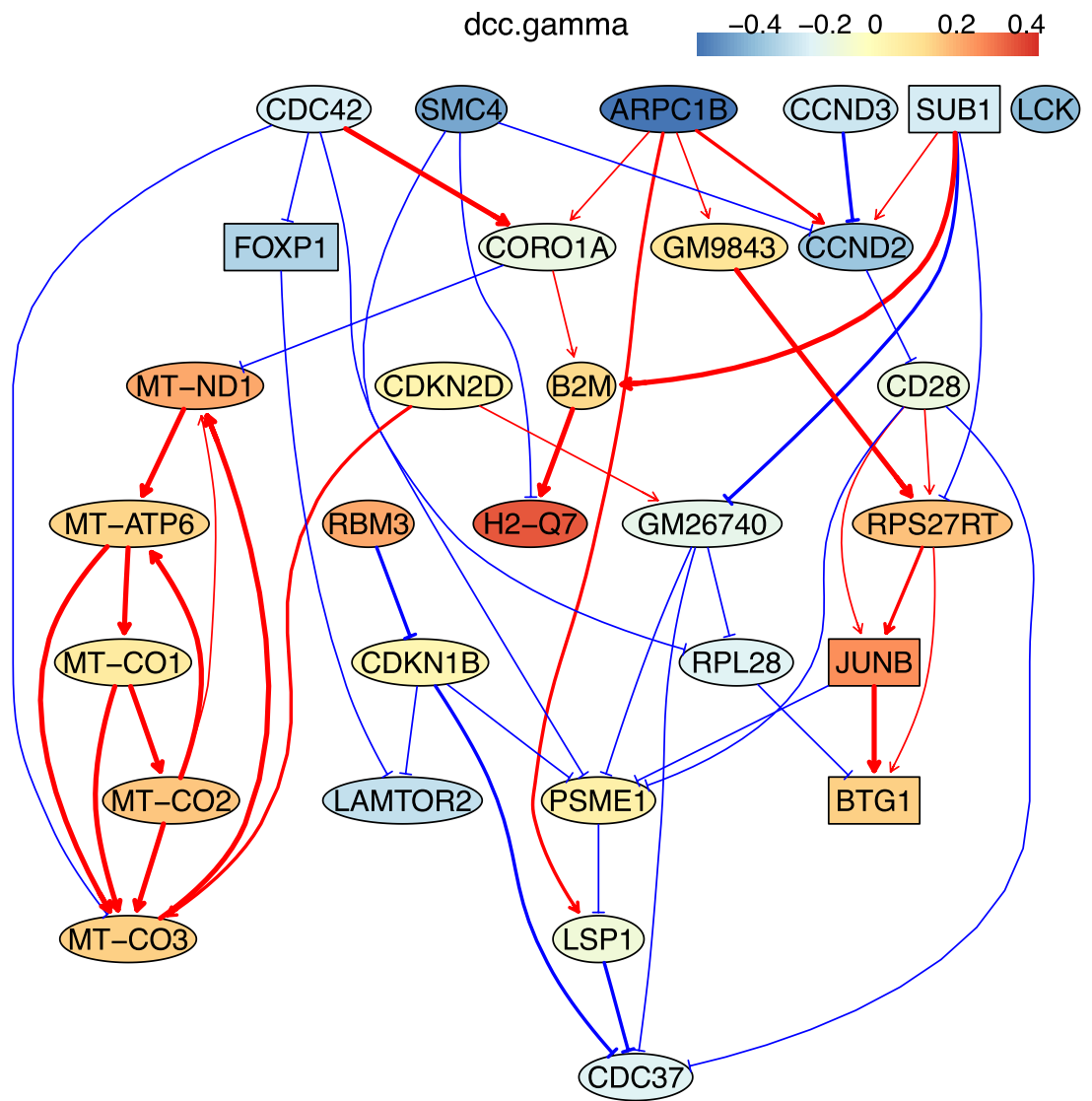
Gene	FC >0 cases	Annotation and evidence	References
Rbm3	24	Muscle from aged rats exhibited an increase in heat shock protein (HSP) 25 and HSP70 and in the cold shock protein RNA-binding motif 3 (RBM3).	Dupont-Versteegden et al., 2008; Van Pelt et al., 2019
H2-Q7	23	A strong increase of the MHC class I genes (including H2-Q7) and B2m is observed in the aging lung.	Angelidis et al., 2019
Btg1	23	Btg1 is involved in neural aging.	Micheli et al., 2021
Gm9843	23	A mouse-specific gene without annotation.	
Rps27rt (Gm9846)	23	Ribosomal protein S27 retrogene, mouse-specific.	
mt-Atp6	22		
mt-Co1	22		
mt-Co2	22		
mt-Co3	22	Mitochondrial proteins involved in the electron transport chain are overrepresented in cells from older participants, with prevalent dysregulation of oxidative phosphorylation and energy metabolism molecular pathways.	Bektas et al., 2019; Goronzy and Weyand, 2019; Haas, 2019
mt-Nd1	21		
Junb	21	JUNB is increased in all human immune cells during aging.	Maity et al., 2021; Zheng et al., 2020
Psmc1	21	Proteasome activator subunit 1. It is implicated in immuno-proteasome assembly and required for efficient antigen processing.	Hwang et al., 2007
B2m	20	B2m is in GO:0007568, a mouse aging GO term. B2M is elevated in the blood of aging humans and mice.	Smith et al., 2015
Gene	Other cases	Annotation and evidence	References
Cd28	12	Cd28 is an aging biomarker of T cells.	Le Page et al., 2018; Zhang et al., 2021a
Cdkn2d	6	Cdkn2d is an aging biomarker.	Goronzy and Weyand, 2019
Cdkn1b	5	Cdkn1b is an aging biomarker.	Goronzy and Weyand, 2019

#1Genes and numbers in red indicate fold change (FC)>0, genes and numbers in blue indicate FC <0, genes and numbers in black do not show clear differential expression in a majority of cell groups.

dcc.gamma

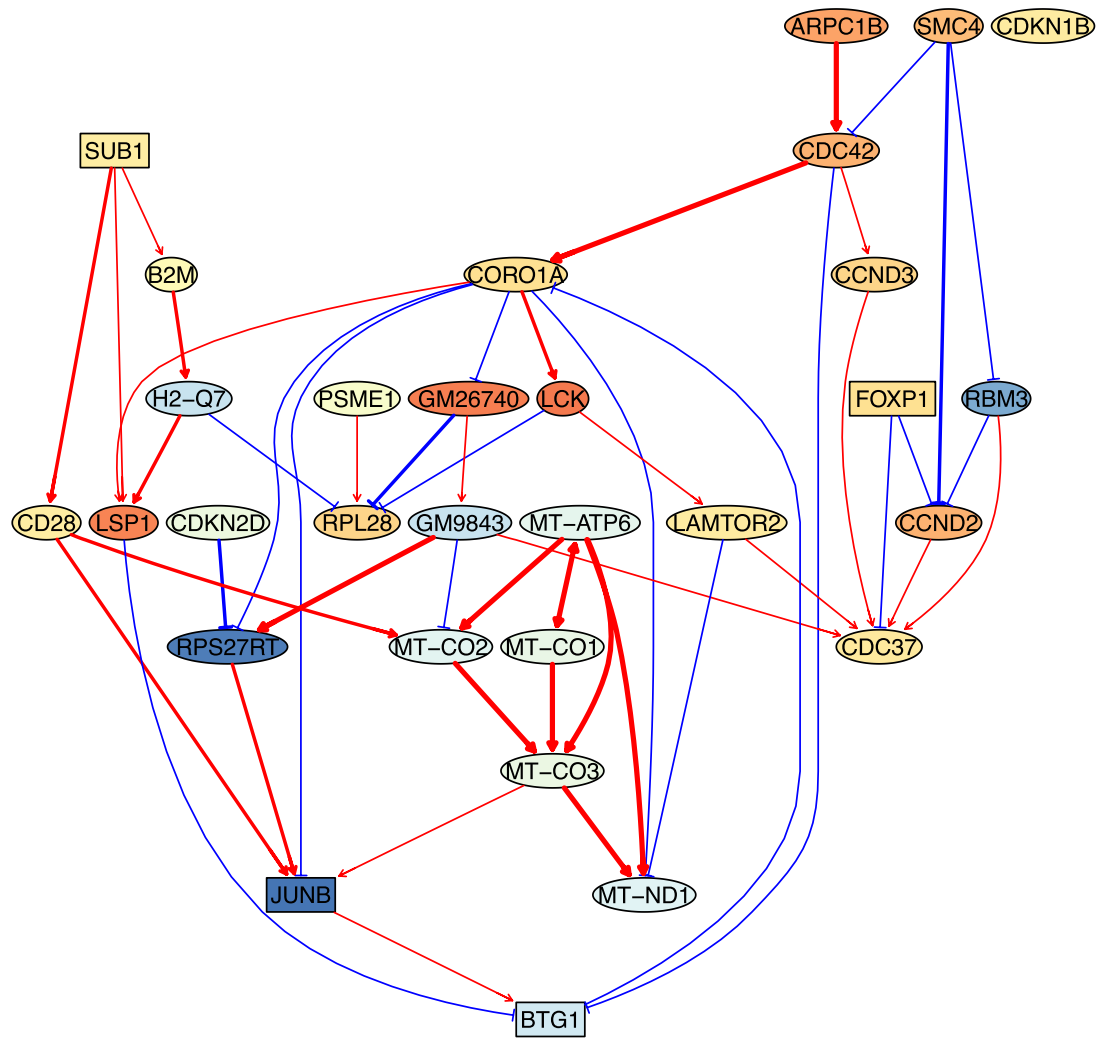


Appendix 5—figure 18. The causal relationships between genes that were differentially expressed in old cytotoxic CD4 T cells. The network was inferred from 600 cells.



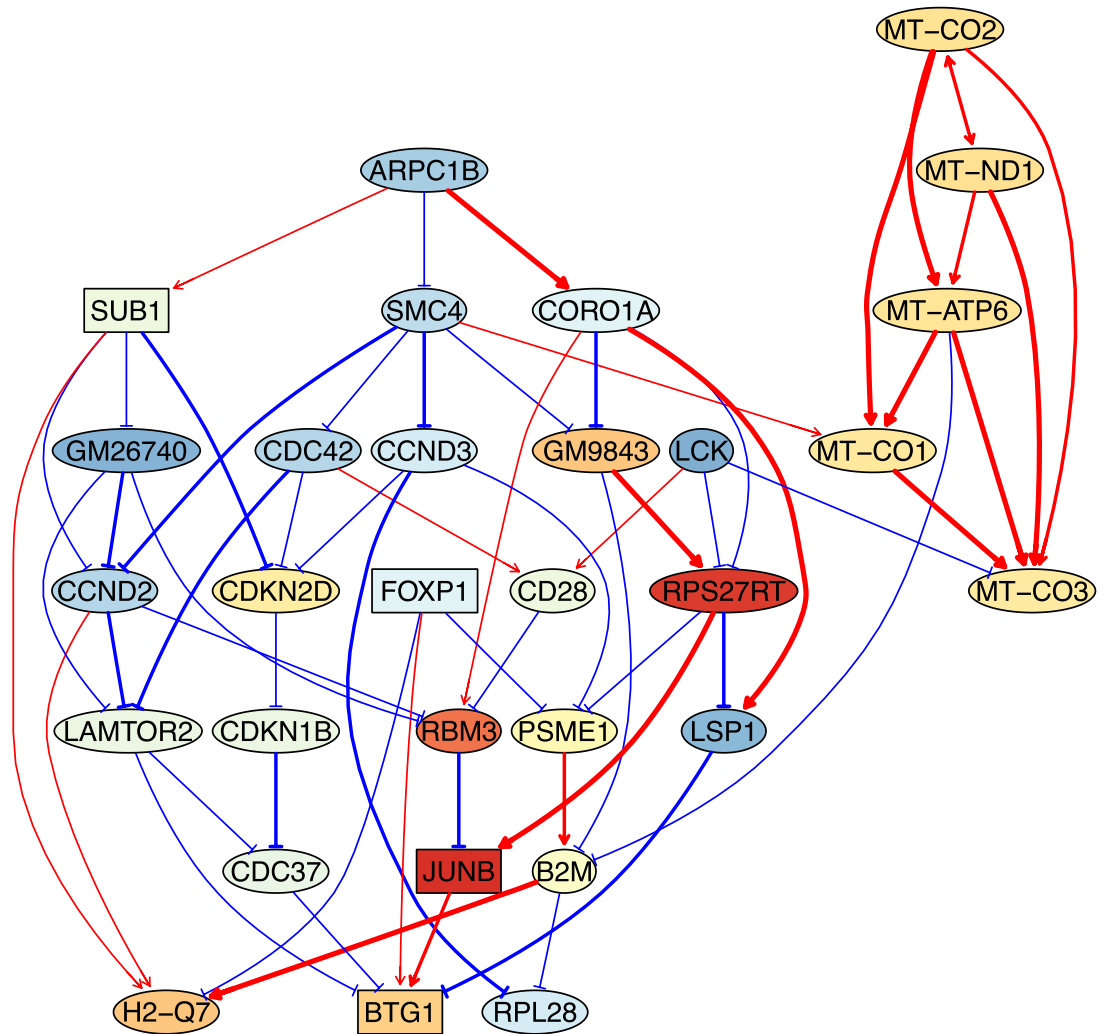
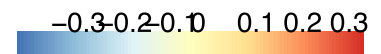
Appendix 5—figure 19. The causal relationships between genes that were differentially expressed in old exhausted CD4 T cells. The network was inferred from 600 cells.

dcc.gamma



Appendix 5—figure 20. The causal relationships between genes that were differentially expressed in young TEM CD4 T cells. The network was inferred from 600 cells.

dcc.gamma



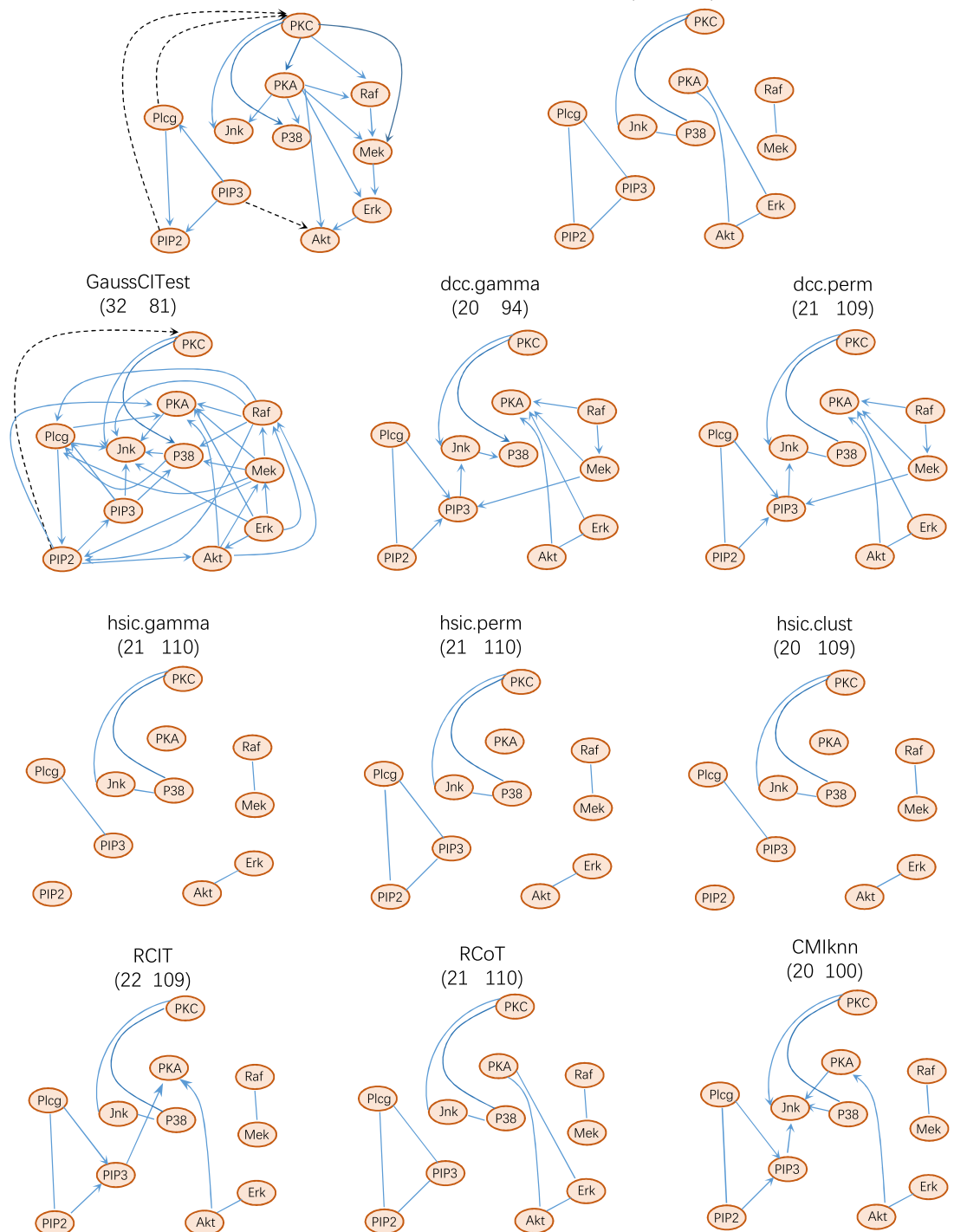
Appendix 5—figure 21. The causal relationships between genes that were differentially expressed in old TEM CD4 T cells. The network was inferred from 600 cells.

5. The analysis of a flow cytometry dataset

Finally, we analyzed the flow cytometry data reported by Sachs et al. This dataset, due to the ground truth given by the authors, has been used to test other algorithms. The computed structural intervention distance (SID) and SHD between networks inferred by different algorithms and the ground truth network also suggest that the DCC CI tests outperform others. See **Appendix 5—figure 22**.

The real network (solid-line edges indicate the Bayesian network inferred using both observational and interventional data)

The Bayesian network inferred using only observational data (21 110)



Appendix 5—figure 22. The performance of causal discovery algorithms upon the Sachs dataset (Sachs et al., 2005). Structural intervention distance (SID) is another important measure for evaluating causal graphs. The numbers in the bracket are structural Hamming distance (SHD) and SID values (the smaller, the better). These values indicate that DCC.gamma and DCC.perm outperform others. The network inferred by Bayesian inference contains only undirected edges.