



International Journal of
Molecular Sciences



Article

Biomarker Prioritisation and Power Estimation Using Ensemble Gene Regulatory Network Inference

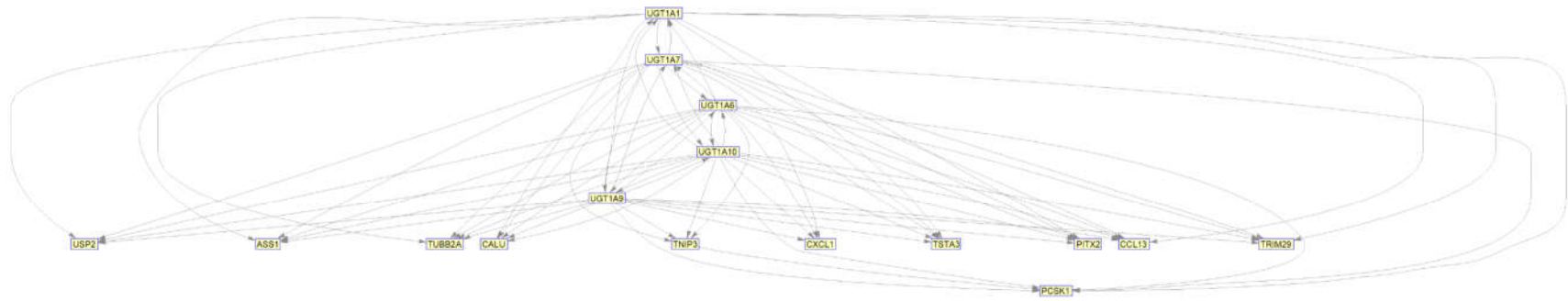


Figure S1. GRN generated from the application of PLSNET on the IBD dataset. The threshold value is selected in such a way that the top 30% edges are selected. UGT genes are delineating as regulators for other genes.

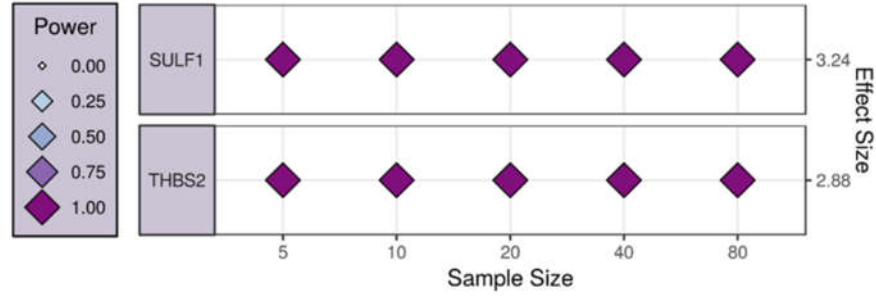


Figure S2. The two identified regulators are represented by the largest observed effect size. The effect size of each assessed variable is shown along the y axis and a series of sample sizes along the x axis.

Table S1. Distribution of the posteriors versus observed experimental states for GSE15471.

Genes	Predicted Marginals		Observed States	
	0	1	0	1
<i>SULF1</i>	0.6026	0.3974	0.6026	0.3974
<i>COL8A1</i>	0.4103	0.5897	0.4103	0.5897
<i>INHBA</i>	0.6026	0.3974	0.6026	0.3974
<i>FN1</i>	0.6026	0.3974	0.6026	0.3974
<i>COL10A1</i>	0.5637	0.4363	0.5641	0.4359
<i>THBS2</i>	0.5641	0.4359	0.5641	0.4359
<i>NTM</i>	0.6021	0.3979	0.6026	0.3974
<i>NOX4</i>	0.3846	0.6154	0.3846	0.6154
<i>RASAL2</i>	0.5385	0.4615	0.5385	0.4615
<i>ADAMTS12</i>	0.5769	0.4231	0.5769	0.4231
<i>CAPG</i>	0.6667	0.3333	0.6667	0.3333
<i>LTBP1</i>	0.6538	0.3462	0.6538	0.3462
<i>CTHRC1</i>	0.5128	0.4872	0.5128	0.4872
<i>FAP</i>	0.5769	0.4231	0.5769	0.4231
<i>WISP1</i>	0.6538	0.3462	0.6538	0.3462
<i>VCAN</i>	0.4235	0.5765	0.4231	0.5769
<i>TIMP1</i>	0.5641	0.4359	0.5641	0.4359
<i>MIR34AHG</i>	0.8462	0.1538	0.8462	0.1538
<i>SLPI</i>	0.5513	0.4487	0.5513	0.4487
<i>GPRC5A</i>	0.6406	0.3594	0.641	0.359

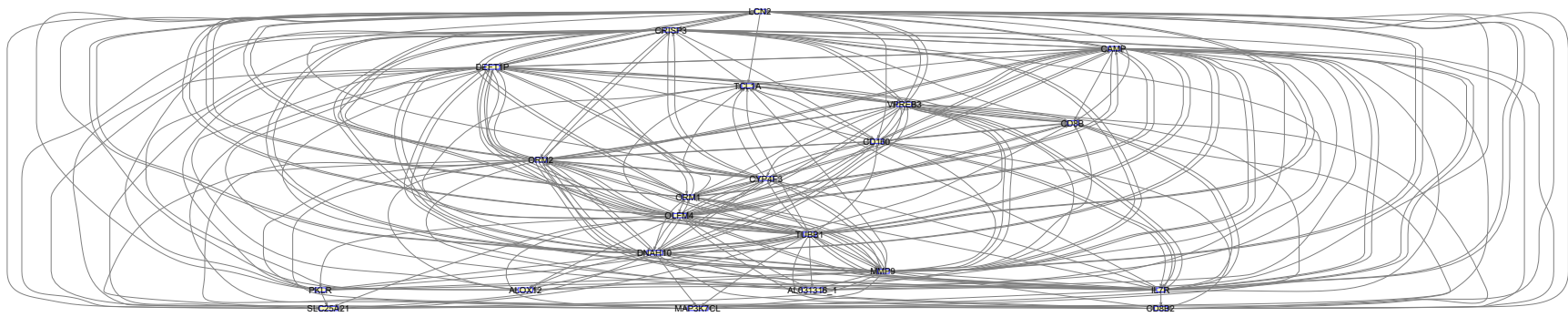


Figure S3. GRN generated using PLSNET for AML. The threshold value is selected in such a way that the top 10% of all edges are selected. This figure does not show Table 22. genes that do not have any interactions. These genes include *HHEX*, *KLF10*, *SOX4*, *MAPK12*, *TRIO*, *GAS5*, *SOCS5*, *MARK4*, *ZNF469*, *EEF1A1*, *C6orf48*, *ASPH*, *SLC17A9*, *RPL22*, *RAB37*, *FBXL19*, *RTN2*, *MSRB3*, *C5*, *SNX33*, *FLT3*, *MERTK*, *CD44*, *ITPR1*, *CRTC1*, *IGF2BP2*, *URB1_AS1*, and 'ARRDC4'.

<i>MSRB3</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
<i>C5</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
<i>CFD</i>	0	0	0	0	0	0	2	0	0	2	0	0	2	0	
<i>ORM2</i>	0	43	57	0	14	86	0	6	94	0	6	94	0	6	94
<i>SNX33</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>FLT3</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>AL022313_1</i>	0	0	0	0	1	0	0	4	0	0	4	0	0	4	0
<i>MERTK</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>CD160</i>	4	65	26	0	24	76	0	8	92	0	8	92	0	8	92
<i>ALOX12</i>	0	75	0	0	92	8	0	70	30	0	70	30	0	70	30
<i>PKLR</i>	2	78	8	0	46	54	0	19	81	0	19	81	0	19	81
<i>CD44</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>CD8B</i>	7	49	31	0	23	77	0	5	95	0	5	95	0	5	95
<i>TCL1A</i>	0	77	23	0	36	64	0	11	89	0	11	89	0	11	89
<i>ITPR1</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>AL031316_1</i>	0	97	3	0	72	28	0	43	57	0	43	57	0	43	57
<i>CRTC1</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>IGF2BP2</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>URB1_AS1</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>SLC25A21</i>	0	99	0	0	91	9	0	78	22	0	78	22	0	78	22
<i>ARRDC4</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

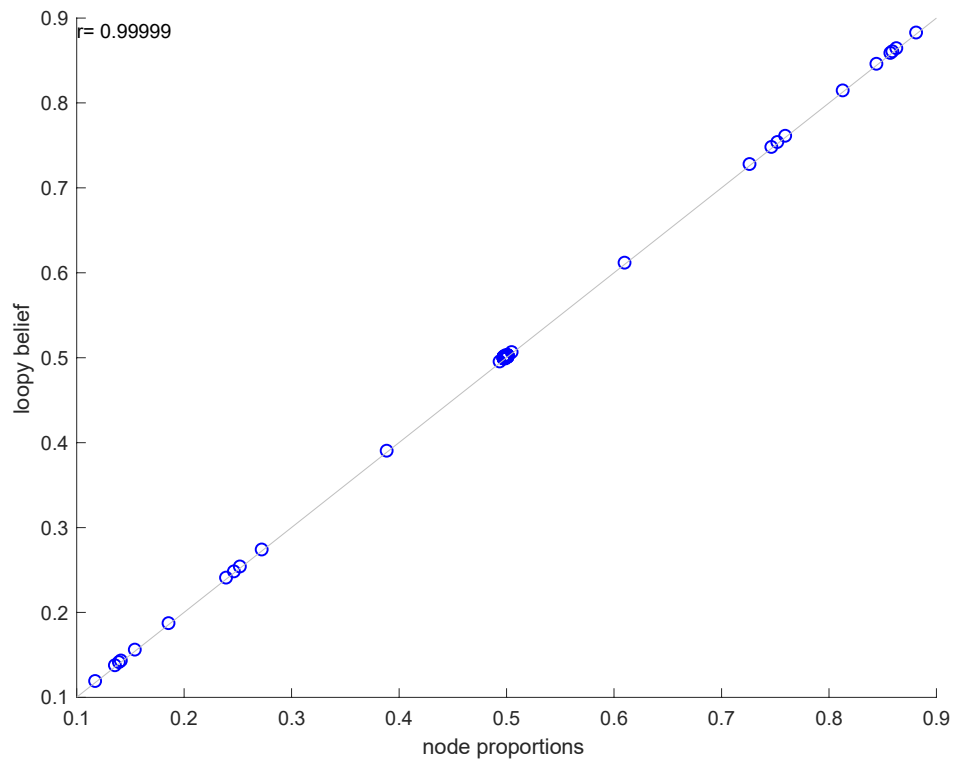


Figure S4. Pearson correlation plots for the AML dataset.

Table S3. Distribution of the posteriors versus observed experimental states in the AML dataset.

Genes	Predicted Marginals		Observed States	
	0	1	0	1
<i>LCN2</i>	0.5	0.5	0.5	0.5
<i>HHEX</i>	0.5	0.5	0.4982	0.5018
<i>CRISP3</i>	0.5	0.5	0.5	0.5
<i>CAMP</i>	0.5	0.5	0.5	0.5
<i>KLF10</i>	0.5	0.5	0.5	0.5
<i>SOX4</i>	0.5	0.5	0.5	0.5
<i>HOXA9</i>	0.5	0.5	0.5	0.5
<i>MAP3K7CL</i>	0.1863	0.8137	0.1863	0.8137
<i>MAPK12</i>	0.5	0.5	0.5	0.5
<i>TRIO</i>	0.5	0.5	0.5	0.5
<i>DEFT1P</i>	0.2731	0.7269	0.2731	0.7269
<i>GAS5</i>	0.5	0.5	0.5	0.5
<i>BMP1</i>	0.5	0.5	0.5	0.5
<i>CD8B2</i>	0.8596	0.1404	0.8598	0.1402
<i>SOCS5</i>	0.5	0.5	0.4982	0.5018
<i>PTH2R</i>	0.5	0.5	0.5	0.5
<i>MARK4</i>	0.5	0.5	0.5	0.5
<i>FSCN1</i>	0.5	0.5	0.5	0.5
<i>CYP4F3</i>	0.5	0.5	0.5	0.5
<i>ZNF469</i>	0.5	0.5	0.5	0.5
<i>DNAH10</i>	0.2472	0.7528	0.2472	0.7528
<i>HOXA10</i>	0.5	0.5	0.5	0.5
<i>VPREB3</i>	0.1551	0.8449	0.155	0.845
<i>EEF1A1</i>	0.5	0.5	0.5	0.5
<i>C6orf48</i>	0.5018	0.4982	0.5018	0.4982
<i>ASPH</i>	0.5	0.5	0.5	0.5
<i>MMP9</i>	0.5	0.5	0.5	0.5
<i>SLC17A9</i>	0.5	0.5	0.5	0.5
<i>UBE2E2</i>	0.5	0.5	0.5	0.5
<i>RPL22</i>	0.5	0.5	0.5	0.5
<i>NETO2</i>	0.5	0.5	0.5	0.5
<i>OLFM4</i>	0.6107	0.3893	0.6107	0.3893
<i>IL7R</i>	0.7471	0.2529	0.7472	0.2528
<i>PNKD</i>	0.5	0.5	0.5	0.5
<i>ORM1</i>	0.5	0.5	0.5	0.5
<i>RAB37</i>	0.5	0.5	0.5	0.5
<i>TUBB1</i>	0.5	0.5	0.5	0.5
<i>FBXL19</i>	0.5	0.5	0.5	0.5
<i>RTN2</i>	0.5	0.5	0.5	0.5
<i>MSRB3</i>	0.4945	0.5055	0.4945	0.5055

<i>C5</i>	0.4982	0.5018	0.4982	0.5018
<i>CFD</i>	0.5	0.5	0.5	0.5
<i>ORM2</i>	0.7601	0.2399	0.7601	0.2399
<i>SNX33</i>	0.5	0.5	0.5	0.5
<i>FLT3</i>	0.5	0.5	0.5	0.5
<i>AL022313_1</i>	0.5018	0.4982	0.5	0.5
<i>MERTK</i>	0.5018	0.4982	0.5018	0.4982
<i>CD160</i>	0.1367	0.8633	0.1365	0.8635
<i>ALOX12</i>	0.5	0.5	0.5	0.5
<i>PKLR</i>	0.5	0.5	0.5	0.5
<i>CD44</i>	0.5	0.5	0.5	0.5
<i>CD8B</i>	0.1422	0.8578	0.1421	0.8579
<i>TCL1A</i>	0.8818	0.1182	0.8819	0.1181
<i>ITPR1</i>	0.5	0.5	0.5	0.5
<i>AL031316_1</i>	0.5	0.5	0.5	0.5
<i>CRTC1</i>	0.5	0.5	0.5	0.5
<i>IGF2BP2</i>	0.4982	0.5018	0.4982	0.5018
<i>URB1_AS1</i>	0.5018	0.4982	0.5018	0.4982
<i>SLC25A21</i>	0.5	0.5	0.5	0.5
<i>ARRDC4</i>	0.5	0.5	0.5	0.5

Table S4. Frequencies of different genes appearing as Regulatory (R), Target (T), or Intermediate (I) genes for different threshold values in the DREAM4 dataset.

Genes	Top 2%			Top 5%			Top 10%			Top 15%			Top 20%		
	R	T	I	R	T	I	R	T	I	R	T	I	R	T	I
<i>G1</i>	0	0	0	8	21	1	1	0	99	0	0	100	0	0	100
<i>G2</i>	0	0	0	0	0	0	0	69	0	0	100	0	0	100	0
<i>G3</i>	0	0	0	0	0	0	0	88	0	0	99	0	0	100	0
<i>G4</i>	0	0	0	0	0	0	0	29	0	0	99	0	0	100	0
<i>G5</i>	0	0	0	21	8	1	0	1	99	0	0	100	0	0	100
<i>G6</i>	1	3	0	2	8	90	0	0	100	0	0	100	0	0	100
<i>G7</i>	0	0	0	0	79	0	0	100	0	0	100	0	0	92	8
<i>G8</i>	3	1	0	8	2	90	0	0	100	0	0	100	0	0	100
<i>G9</i>	98	0	0	100	0	0	100	0	0	100	0	0	82	0	18
<i>G10</i>	0	96	0	0	100	0	0	100	0	0	100	0	0	100	0

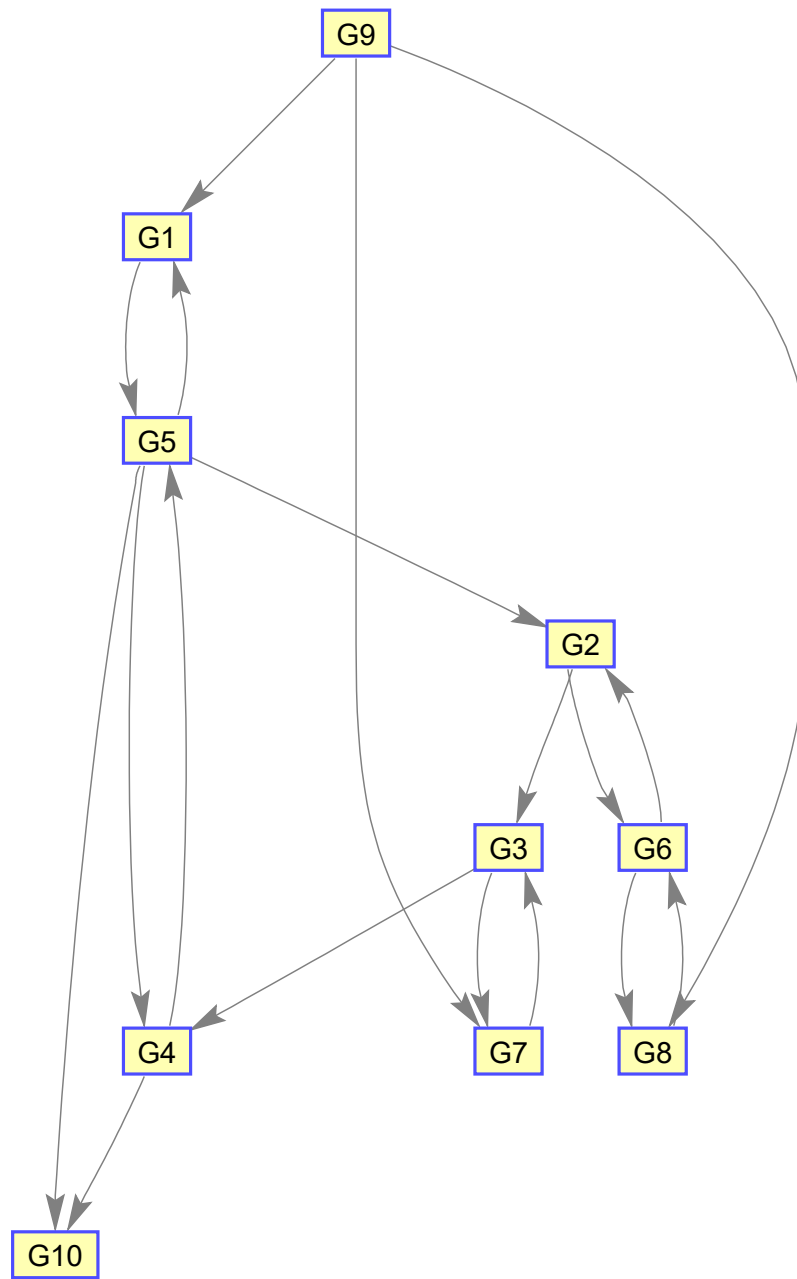


Figure S5. GRN generated from the application of MIDER on the DREAM dataset.

Table S5. Distribution of the posteriors versus observed experimental states for DREAM4.

Genes	Predicted Marginals		Observed States	
	0	1	0	1
<i>G1</i>	0.6026	0.3974	0.6026	0.3974
<i>G2</i>	0.4103	0.5897	0.4103	0.5897
<i>G3</i>	0.6026	0.3974	0.6026	0.3974
<i>G4</i>	0.6026	0.3974	0.6026	0.3974
<i>G5</i>	0.5637	0.4363	0.5641	0.4359
<i>G6</i>	0.5641	0.4359	0.5641	0.4359
<i>G7</i>	0.6021	0.3979	0.6026	0.3974
<i>G8</i>	0.3846	0.6154	0.3846	0.6154
<i>G9</i>	0.5385	0.4615	0.5385	0.4615
<i>G10</i>	0.5769	0.4231	0.5769	0.4231

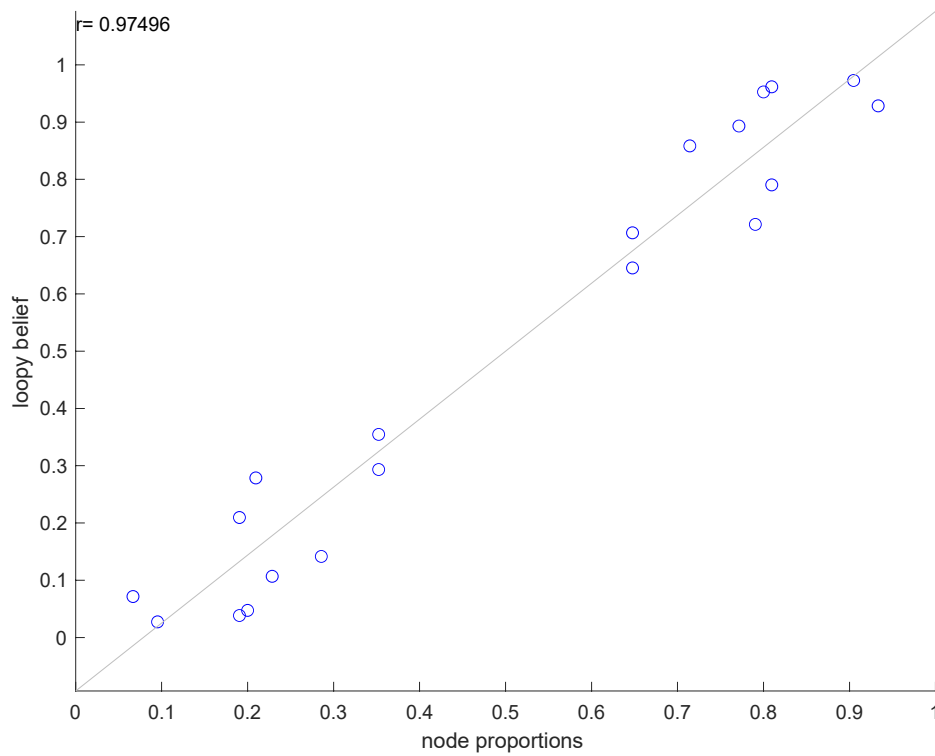
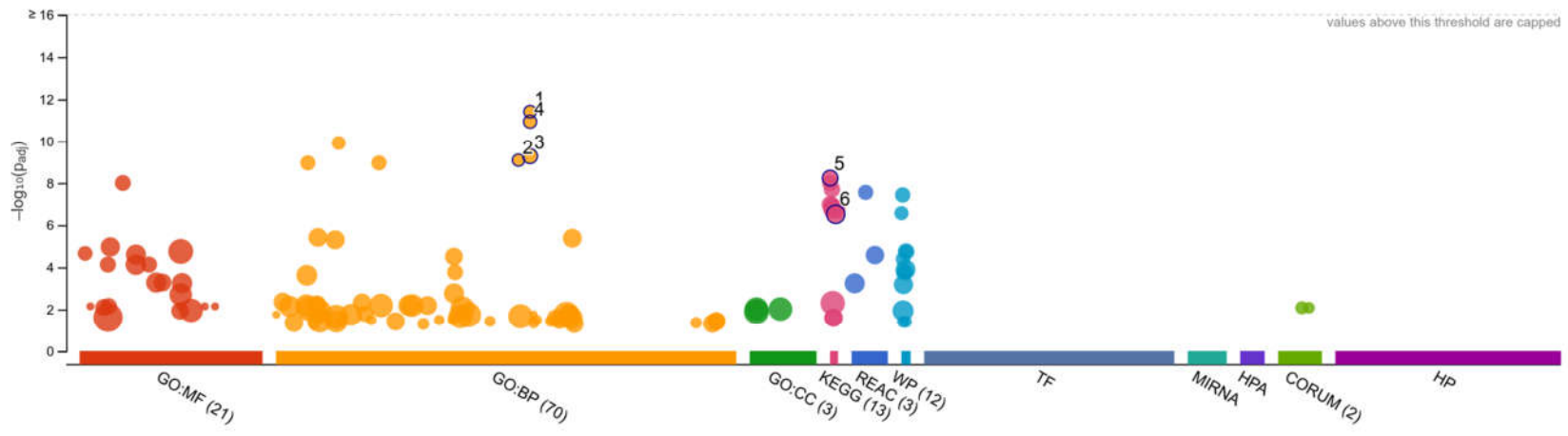


Figure S6. Pearson correlation plots for the DREAM4.

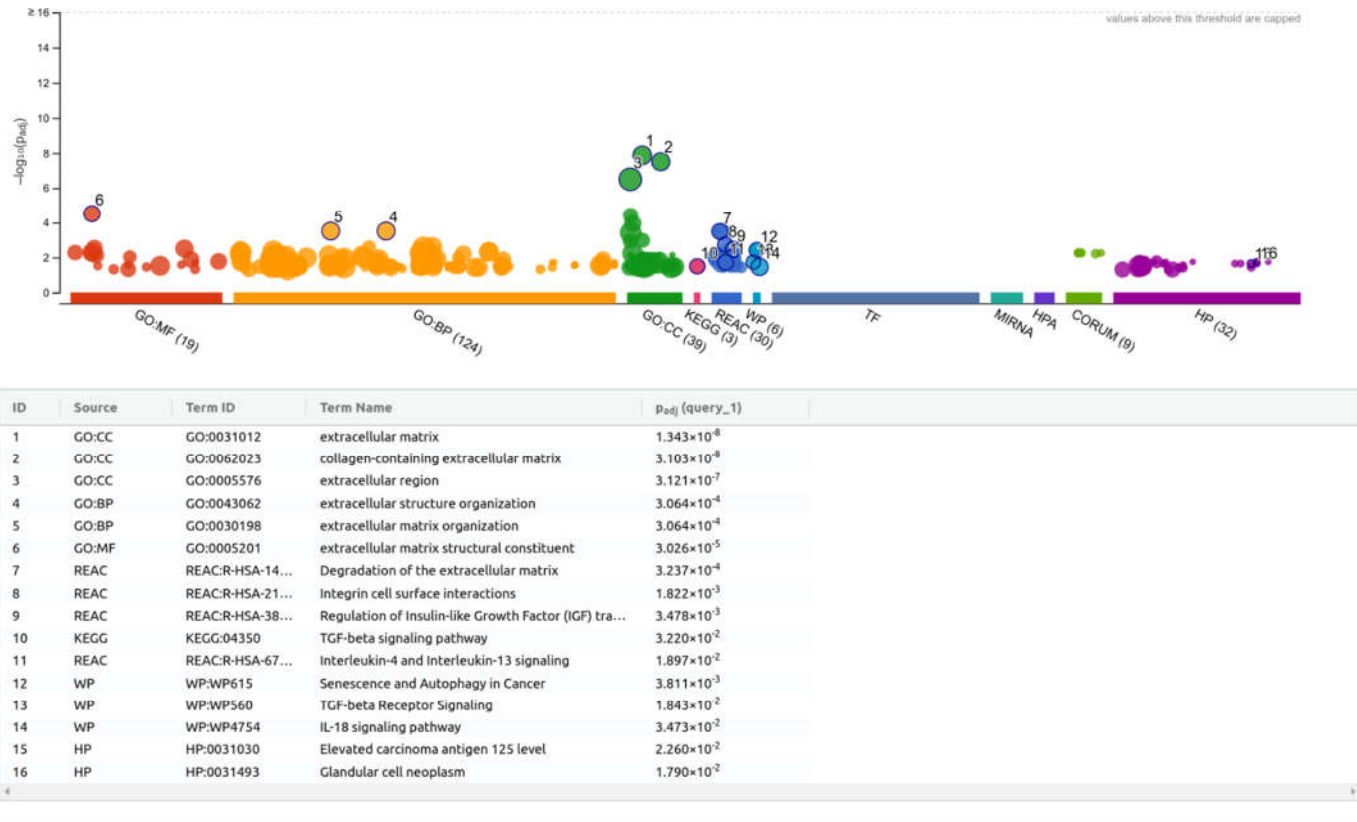


ID	Source	Term ID	Term Name	Padj (query_1)
1	GO:BP	GO:0052696	flavonoid glucuronidation	3.862×10^{-12}
2	GO:BP	GO:0051552	flavone metabolic process	8.236×10^{-10}
3	GO:BP	GO:0052695	cellular glucuronidation	5.314×10^{-10}
4	GO:BP	GO:0052697	xenobiotic glucuronidation	1.158×10^{-11}
5	KEGG	KEGG:00053	Ascorbate and aldarate metabolism	5.858×10^{-9}
6	KEGG	KEGG:04976	Bile secretion	3.061×10^{-7}

version e100_eg47_p14_7733820
date 24/09/2020, 12:48:08
organism hsapiens

g:Profiler

Figure S7. Gene set enrichment results from the IBD dataset. Genes were entered into the gProfiler database. P values are FDR adjusted. GO:BP = biological process, KEGG = Kyoto Encyclopedia of Genes and Genomes database.



version e100_eg47_p14_7733820
 date 23/09/2020, 16:24:50
 organism hsapiens

g:Profiler

Figure S8. Gene set enrichment results from the PDAC dataset. Genes were entered into the gProfiler database. *p* values are FDR adjusted. GO:BP = biological process, GO:CC = cellular component, REAC = Reactome database, WP = WikiPathways, HP = Human Phenotype Ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes database.

