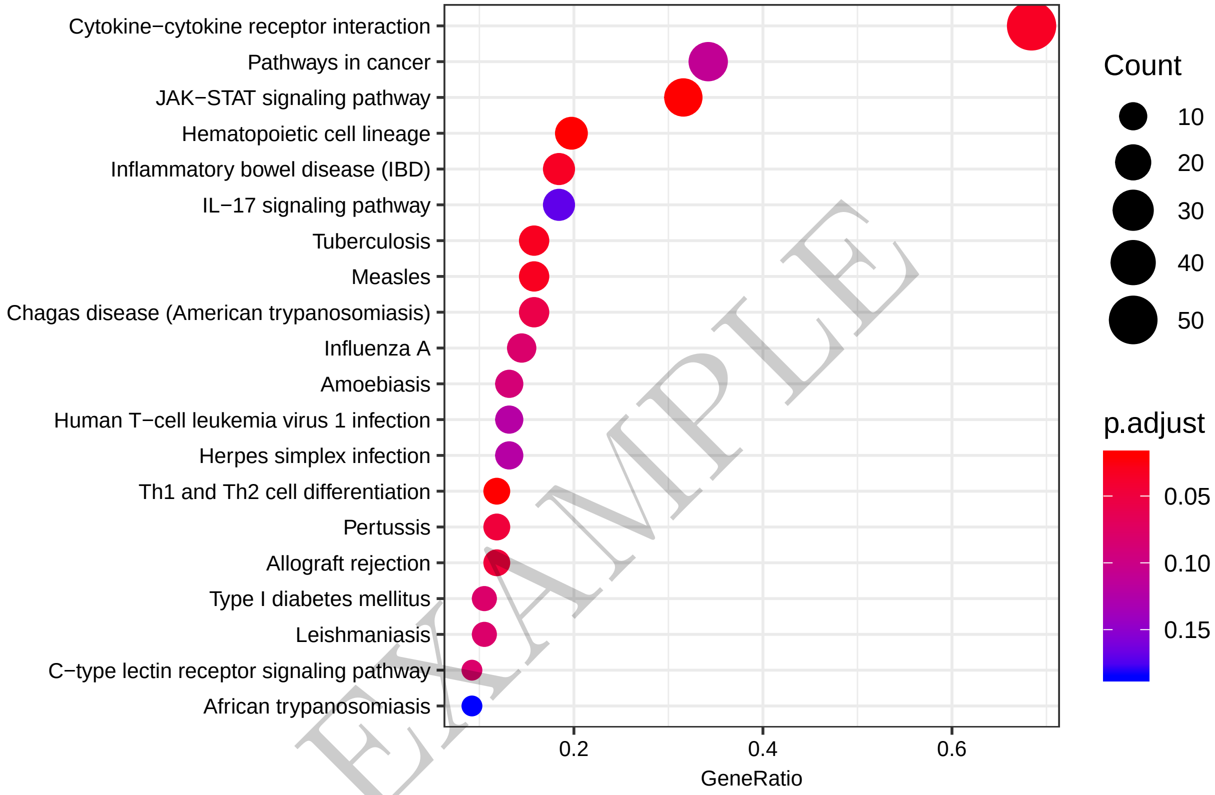
EXAMPLE REPORT

Biostatistics & Bioinformatics Services

"Pathway Analysis" Service



Bioinformatics Team, RayBiotech
December 04, 2018

Contents

1 Introduction

The “Pathway analysis” service assesses the enrichment of signaling pathways, biological functions, and gene sets based on the expression of biomarkers across the different groups. The analysis is conducted based on information from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2016) and Gene Ontology (GO) databases (Ashburner 2000). This service also determines whether differentially-expressed biomarkers have known protein-protein interactions (PPIs) curated by the STRING database (Szklarczyk et al. 2017).

> *Need help understanding how the statistical analyses were performed in layman’s terms? Please visit our website.*

Table 1 summarizes the samples and groups for this project.

Table 1: Samples in Control and Patient groups

sample	group	sample	group	sample	group	sample	group
SC1	Patient	SC11	Patient	SN1	Control	SN11	Control
SC2	Patient	SC12	Patient	SN2	Control	SN12	Control
SC3	Patient	SC13	Patient	SN3	Control	SN13	Control
SC4	Patient	SC14	Patient	SN4	Control	SN14	Control
SC5	Patient	SC15	Patient	SN5	Control	SN15	Control
SC6	Patient	SC16	Patient	SN6	Control		
SC7	Patient	SC17	Patient	SN7	Control		
SC8	Patient	SC18	Patient	SN8	Control		
SC9	Patient	SC19	Patient	SN9	Control		
SC10	Patient	SC20	Patient	SN10	Control		

2 Methods

2.1 Differential Expression

To determine whether biomarker was differentially-expressed across groups, the mean and median of each biomarker value per group were first calculated. The fold change is equal to the ratio of the mean or median of two groups (e.g., patient group to control group). The significance of the differential expression for each biomarker was evaluated using the Student’s t-test, Wilcoxon Rank-Sum, ANOVA, or Kruskal-Wallis test depending on the number of groups and distribution type (i.e., normal, non-normal) of the data. Differentially-expressed biomarkers were defined as biomarkers with p-values less than 0.05.

2.2 Enriched KEGG pathway and GO terms

The KEGG pathway enrichment and GO term enrichment were conducted in two approaches. The first one, “Over-representation analysis,” used hypergeometric distribution in the subset of differentially-expressed biomarkers. The second one, “Gene set enrichment analysis (GSEA)”, considered fold-changes between groups from all the biomarkers in a pathway/GO term, whether the biomarkers were differentially-expressed or not, and then calculated the enrichment score based on the magnitude of the fold changes.

2.3 Protein-protein interactions

The protein-protein interactions of differentially-expressed biomarkers were retrieved from the STRING database (Szklarczyk et al. 2017).

2.4 Software

All the analyses were conducted in the R programming language V3.5.1 (R Core Team 2017). The Pathway/GO over-representation and GSEA analyses were implemented with the R package $clusterProfiler$ (Yu et al. 2012).

3 Results

3.1 Differentially-expressed biomarkers

The 86 differentially-expressed biomarkers (p-value<0.05) are listed in Table 2.

Table 2: Differentially-expressed biomarkers between Control and Patient groups (p<0.05)

biomarker	Control	Patient	FoldChange	statistic	p.value	FDR	entrez_id
IL-10	15.81(3.7, 36.65)	2.66(0, 15.84)	0.1680	Wilcoxon W = 282	0.0000	0.0001	3586
IL-13 R2	34.04(6.56, 89.7)	4.43(0, 34.28)	0.1303	Wilcoxon W = 282	0.0000	0.0001	3598
TNFa	2483.15(1357.5, 4344.26)	792.32(171.54, 2988.36)	0.3191	Wilcoxon W = 277	0.0000	0.0002	7124
IL-1ra	26.62±7.88	11.97±6.77	0.4494	t = 5.7817	0.0000	0.0002	3557
IL-13	706.08(424.97, 1607.99)	316.53(61.56, 814.36)	0.4483	Wilcoxon W = 275	0.0000	0.0003	3596
IL-23	25939.37(14208.07, 68712.35)	10086.88(1439.62, 30838.83)	0.3889	Wilcoxon W = 275	0.0000	0.0003	51561
Cripto-1	2600.42±831.13	1017.1±926.62	0.3911	t = 5.3078	0.0000	0.0003	6997
TRAIL R4	9718.77(5181.03, 17139.64)	3186.7(676.92, 11833.9)	0.3279	Wilcoxon W = 273	0.0000	0.0003	8793
IL-5	58.97(27.56, 116.04)	23.38(7.87, 62.98)	0.3964	Wilcoxon W = 272	0.0000	0.0003	3567
Follistatin	1531.15(651.8, 3004.18)	486.78(159.18, 1918.07)	0.3179	Wilcoxon W = 270	0.0000	0.0004	10468
IL-6	153.42(78.74, 259.92)	62.05(26.09, 182.48)	0.4044	Wilcoxon W = 270	0.0000	0.0004	3569
GM-CSF	45.55(16, 96.15)	14.48(3.55, 52.62)	0.3180	Wilcoxon W = 269	0.0000	0.0004	1437
IL-2 Rb	1689.73(837.55, 5182.08)	686.31(344.7, 1932.85)	0.4062	Wilcoxon W = 269	0.0000	0.0004	3560
Angiostatin	41312.16(14358.44, 89185.45)	12967.54(2980.2, 48135.76)	0.3139	Wilcoxon W = 268	0.0000	0.0005	5340
BDNF	1120.92±484.24	2097.75±716.43	1.8715	t = -4.8069	0.0000	0.0006	627
Fcg RIIBC	231.24±71.93	118.49±60.05	0.5124	t = 4.9195	0.0000	0.0006	2213
IL-6R	8105.91±1761.84	5449.47±1159.76	0.6723	t = 5.0731	0.0000	0.0006	3570
E-Cadherin	844.76±318.56	366.96±210.47	0.4344	t = 5.042	0.0000	0.0006	999
SDF-1a	78.01(37.32, 133.73)	39.75(4.18, 67.73)	0.5095	Wilcoxon W = 265	0.0000	0.0006	6387
Galectin-7	78884.84±18279.3	51708.95±11723.07	0.6555	t = 5.0337	0.0000	0.0006	3963
IL-4	197.37(138.92, 602.18)	91.98(24.07, 254.02)	0.4660	Wilcoxon W = 264	0.0001	0.0007	3565
FAS L	170.97(117.81, 557.09)	76.93(18.95, 224.17)	0.4500	Wilcoxon W = 263	0.0001	0.0008	356
IL-16	95.48(60.19, 173.82)	43.2(26.38, 168.27)	0.4525	Wilcoxon W = 261	0.0001	0.0011	3603
NrCAM	170.22(98.29, 350.28)	63.45(36.58, 339.58)	0.3728	Wilcoxon W = 260	0.0001	0.0013	4897
EpCAM	1546.11(730.37, 3875.03)	662.47(242.84, 2121.75)	0.4285	Wilcoxon W = 257	0.0002	0.0020	4072
TGFb2	15175.75±2554.91	11916.25±1671.59	0.7852	t = 4.2989	0.0003	0.0029	7042
TNFb	105.41(0, 487.36)	0(0, 62.03)	0.0000	Wilcoxon W = 237	0.0003	0.0030	4049
TREM-1	34.14(0, 223.1)	0(0, 18.35)	0.0000	Wilcoxon W = 237	0.0003	0.0030	54210
CD14	23786.61(7429.2, 91955.19)	66283.41(16064.78, 206121.1)	2.7866	Wilcoxon W = 47	0.0003	0.0033	929
IFNg	337.01(181.36, 769.6)	68.28(0, 691.36)	0.2026	Wilcoxon W = 257	0.0004	0.0035	3458
IL-12p70	10.54(5.64, 29.6)	4.57(2.41, 13.49)	0.4333	Wilcoxon W = 252	0.0004	0.0035	3592
IL-8	20.83(9.27, 54.31)	6.66(0.23, 59.63)	0.3196	Wilcoxon W = 252	0.0004	0.0035	3576
CCL28	19.21(12.27, 29.31)	13.69(4.72, 53.14)	0.7125	Wilcoxon W = 251	0.0005	0.0038	56477
IL-13 R1	403.67(153.59, 1263.23)	106.66(2.13, 1427.68)	0.2642	Wilcoxon W = 251	0.0005	0.0038	3597
CTACK	633.07±267.47	1037.22±357.76	1.6384	t = -3.8242	0.0006	0.0044	10850
RANTES	9797.93±1397.52	8090.4±1164.91	0.8257	t = 3.8369	0.0007	0.0052	6352
ICAM-2	1363.17(558.03, 2736.41)	554.52(145.87, 3586.85)	0.4068	Wilcoxon W = 248	0.0007	0.0052	3384
IL-2	203.65(110.41, 431.41)	126.48(56.11, 263.07)	0.6211	Wilcoxon W = 248	0.0007	0.0052	3558
Nidogen-1	34826.75±4361.84	28831.06±5555.94	0.8278	t = 3.5756	0.0011	0.0079	4811
IL-17	27.12(15.17, 69.68)	12.36(4.87, 35.03)	0.4555	Wilcoxon W = 243	0.0014	0.0097	3605
OPG	379.6(241.24, 769.73)	575.74(289.44, 2171.87)	1.5167	Wilcoxon W = 57	0.0014	0.0097	4982
BTC	195.95(121.79, 317.64)	135.18(82.64, 522.39)	0.6898	Wilcoxon W = 242	0.0016	0.0105	685
PAI-1	158.52(81.85, 463.1)	65.39(22.57, 214.15)	0.4125	Wilcoxon W = 242	0.0016	0.0105	5054
ANG-1	1441.44(379.4, 5384.61)	360.46(0, 2445.32)	0.2501	Wilcoxon W = 243	0.0020	0.0130	284
IL-15	279.45(122.96, 875.47)	154.61(52.67, 646.48)	0.5533	Wilcoxon W = 240	0.0021	0.0130	3600
IGFBP-1	7997.67±4096.23	12842.87±4473.34	1.6058	t = -3.3284	0.0022	0.0136	3484
MCP-3	5.74±1.97	3.65±1.72	0.6351	t = 3.2839	0.0028	0.0165	6354
LAP(TGFb1)	180.16(73.9, 364.32)	95.67(29.82, 465.6)	0.5310	Wilcoxon W = 236	0.0034	0.0200	
IL-11	920.91(504.8, 2286.36)	575.22(237.08, 1382.91)	0.6246	Wilcoxon W = 235	0.0039	0.0220	3589
IL-2 Ra	374.69(72.82, 1298.19)	101.66(0, 795.13)	0.2713	Wilcoxon W = 237	0.0039	0.0220	3559
IL-7	160.12±43.57	108.15±56.67	0.6755	t = 3.0666	0.0043	0.0236	3574
Axl	428.58(210.7, 705.34)	762.12(166.78, 2094.89)	1.7783	Wilcoxon W = 67	0.0049	0.0258	558
LIGHT	39.55(30.42, 55.63)	30.61(9.81, 83.2)	0.7739	Wilcoxon W = 233	0.0049	0.0258	8740
DR6	6219.86(1856.71, 9302.98)	10145(3836.21, 30558.17)	1.6311	Wilcoxon W = 70	0.0068	0.0349	27242
GDF-15	641.59(442.17, 1073.45)	894.9(423.67, 2252.09)	1.3948	Wilcoxon W = 70	0.0068	0.0349	9518
BMP-4	10.61(0, 549.23)	36.79(1.54, 114.68)	3.4665	Wilcoxon W = 70	0.0080	0.0402	652
B2M	4153.15(699.88, 7752.17)	1407.77(230.24, 9341.52)	0.3390	Wilcoxon W = 228	0.0085	0.0418	567
IL-12p40	415.3(80.59, 1512.43)	222.69(40.97, 863.87)	0.5362	Wilcoxon W = 227	0.0095	0.0442	3593
IL-17B	908.59(455.17, 2835.52)	538.16(183.05, 1547.98)	0.5923	Wilcoxon W = 227	0.0095	0.0442	27190

Table 2: Differentially-expressed biomarkers between Control and Patient groups (p<0.05) *(continued)*

biomarker	Control	Patient	FoldChange	statistic	p.value	FDR	entrez_id
IL-18 BPa	57.7(6.68, 141.88)	28.02(6.66, 79.83)	0.4855	Wilcoxon W = 227	0.0095	0.0442	10068
DKK-1	244.58(71.4, 1082.42)	105.38(13.51, 3092.39)	0.4309	Wilcoxon W = 223	0.0143	0.0644	22943
SCF R	14200.8(11738.61, 19374.48)	16415.05(12967.4, 30392.35)	1.1559	Wilcoxon W = 77	0.0143	0.0644	3815
MIP-1d	3177.94(1325.69, 5611.54)	2477.68(1123.52, 5590.08)	0.7796	Wilcoxon W = 221	0.0173	0.0771	6359
HGF	102.36(59.08, 459.15)	165.07(52.13, 320.05)	1.6127	Wilcoxon W = 80	0.0191	0.0789	3082
IL-3	328.88(23.94, 6377.33)	124.76(0, 2152.33)	0.3794	Wilcoxon W = 220	0.0191	0.0789	3562
IP-10	139.86(103.11, 175.59)	174.54(101.05, 456.76)	1.2480	Wilcoxon W = 80	0.0191	0.0789	3627
Siglec-5	2490.11(872.42, 4925.89)	1856.25(720.78, 2846.56)	0.7455	Wilcoxon W = 220	0.0191	0.0789	8778
ErbB2	1.9(0, 115.2)	0(0, 48.32)	0.0005	Wilcoxon W = 218.5	0.0192	0.0789	2064
Eotaxin-2	233.62(91.2, 563.82)	156.85(48.6, 377.16)	0.6714	Wilcoxon W = 219	0.0210	0.0852	6369
IL-28A	2.92(0.93, 4.91)	1.05(0, 10.27)	0.3605	Wilcoxon W = 217	0.0266	0.1061	282616
MMP-3	12088.56(4152.25, 40637.18)	8262.79(3231.06, 20062.97)	0.6835	Wilcoxon W = 216	0.0277	0.1061	4314
TRAIL R3	1341.54(284.31, 2627.32)	1686.51(877.12, 21622.32)	1.2571	Wilcoxon W = 84	0.0277	0.1061	8794
MMP-2	1036.14(0, 15909.24)	91.02(0, 12689.66)	0.0878	Wilcoxon W = 216	0.0277	0.1061	4313
I-309	15.94(4.97, 56.64)	9.19(0, 34.68)	0.5766	Wilcoxon W = 215	0.0302	0.1144	6346
MBL	4869.27±1103.7	4079.38±875.74	0.8378	t = 2.2845	0.0307	0.1147	4153
ALCAM	284.41(84.72, 3622.15)	623.4(117.03, 2641.52)	2.1919	Wilcoxon W = 86	0.0330	0.1170	214
Angiogenin	401.64(147.61, 1207.63)	277.44(71.58, 761.25)	0.6908	Wilcoxon W = 214	0.0330	0.1170	283
CXCL16	2428.43(1445.92, 2841.72)	2640.95(1646.57, 3216.85)	1.0875	Wilcoxon W = 86	0.0330	0.1170	58191
G-CSF	409.57(84.7, 2171.77)	199.99(0, 1264.79)	0.4883	Wilcoxon W = 214	0.0330	0.1170	1440
IL-1a	79.16(0, 306.99)	0(0, 216.87)	0.0000	Wilcoxon W = 211	0.0340	0.1190	3552
Cathepsin S	17.98(4.57, 30.54)	9.49(0, 44.29)	0.5278	Wilcoxon W = 210	0.0463	0.1545	1520
ErbB3	646.88(140.86, 7446.32)	1450.08(371.71, 8729.04)	2.2416	Wilcoxon W = 90	0.0463	0.1545	2065
Leptin	8415.66(1402.18, 82488.18)	2704.97(737.1, 23761.42)	0.3214	Wilcoxon W = 210	0.0463	0.1545	3952
TIMP-1	30287.52(24332.14, 49866.78)	29087.94(18043.77, 36785.39)	0.9604	Wilcoxon W = 210	0.0463	0.1545	7076
TPO	176641.89±23630.93	159538.78±25072.51	0.9032	t = 2.0641	0.0474	0.1562	7066
Lymphotactin	10.91(0, 361.7)	143.02(0, 3393.29)	13.1061	Wilcoxon W = 91	0.0492	0.1603	6375

3.2 Over-representation of KEGG pathway/Gene Ontology(GO) terms

3.2.1 KEGG over-representation analysis

Table 3: 10 KEGG pathway based on 85 differentially-expressed biomarkers between Control and Patient groups

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
hsa04630	JAK-STAT signaling pathway	24/76	43/230	0.0005507	0.0205379	0.0154421
hsa04640	Hematopoietic cell lineage	15/76	23/230	0.0008925	0.0205379	0.0154421
hsa04658	Th1 and Th2 cell differentiation	9/76	11/230	0.0009780	0.0205379	0.0154421
hsa05152	Tuberculosis	12/76	18/230	0.0025052	0.0315654	0.0237334
hsa05162	Measles	12/76	18/230	0.0025052	0.0315654	0.0237334
hsa04060	Cytokine-cytokine receptor interaction	52/76	127/230	0.0033790	0.0333081	0.0250437
hsa05321	Inflammatory bowel disease (IBD)	14/76	23/230	0.0037009	0.0333081	0.0250437
hsa05133	Pertussis	9/76	13/230	0.0067232	0.0470621	0.0353850
hsa05330	Allograft rejection	9/76	13/230	0.0067232	0.0470621	0.0353850
hsa05142	Chagas disease (American trypanosomiasis)	12/76	20/230	0.0089680	0.0564983	0.0424799

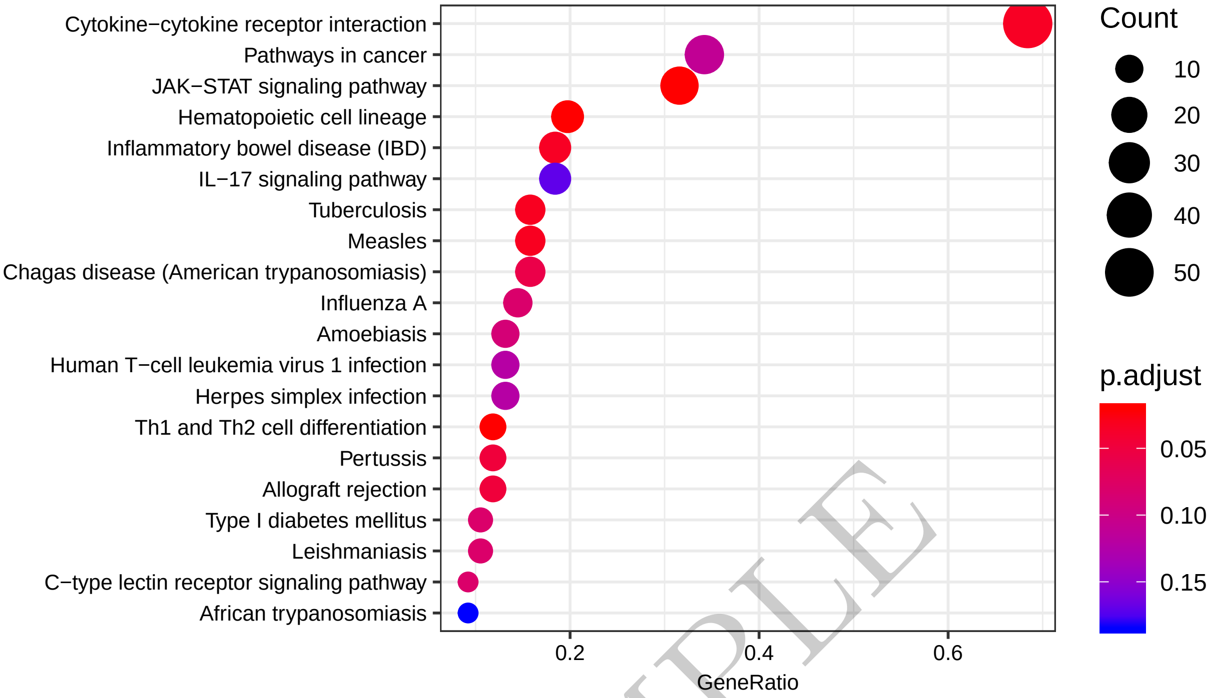


Figure 1: KEGG pathway over-representation analysis with 85 biomarkers between Control and Patient groups

3.2.2 GO biological process over-representation analysis

Table 4: 10 GO biological process based on 85 differentially-expressed biomarkers between Control and Patient groups

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
GO:0048534	hematopoietic or lymphoid organ development	30/84	60/276	0.0002550	0.1228285	0.1111921
GO:0030155	regulation of cell adhesion	33/84	69/276	0.0003345	0.1228285	0.1111921
GO:0030217	T cell differentiation	15/84	23/276	0.0003420	0.1228285	0.1111921
GO:0030098	lymphocyte differentiation	19/84	33/276	0.0005124	0.1228285	0.1111921
GO:0030097	hemopoiesis	29/84	59/276	0.0005144	0.1228285	0.1111921
GO:0022407	regulation of cell-cell adhesion	27/84	54/276	0.0006141	0.1228285	0.1111921
GO:0002520	immune system development	30/84	63/276	0.0008218	0.1286298	0.1164439
GO:0046651	lymphocyte proliferation	21/84	39/276	0.0008575	0.1286298	0.1164439
GO:0032943	mononuclear cell proliferation	21/84	40/276	0.0013570	0.1570732	0.1421926
GO:0050670	regulation of lymphocyte proliferation	20/84	38/276	0.0017586	0.1570732	0.1421926

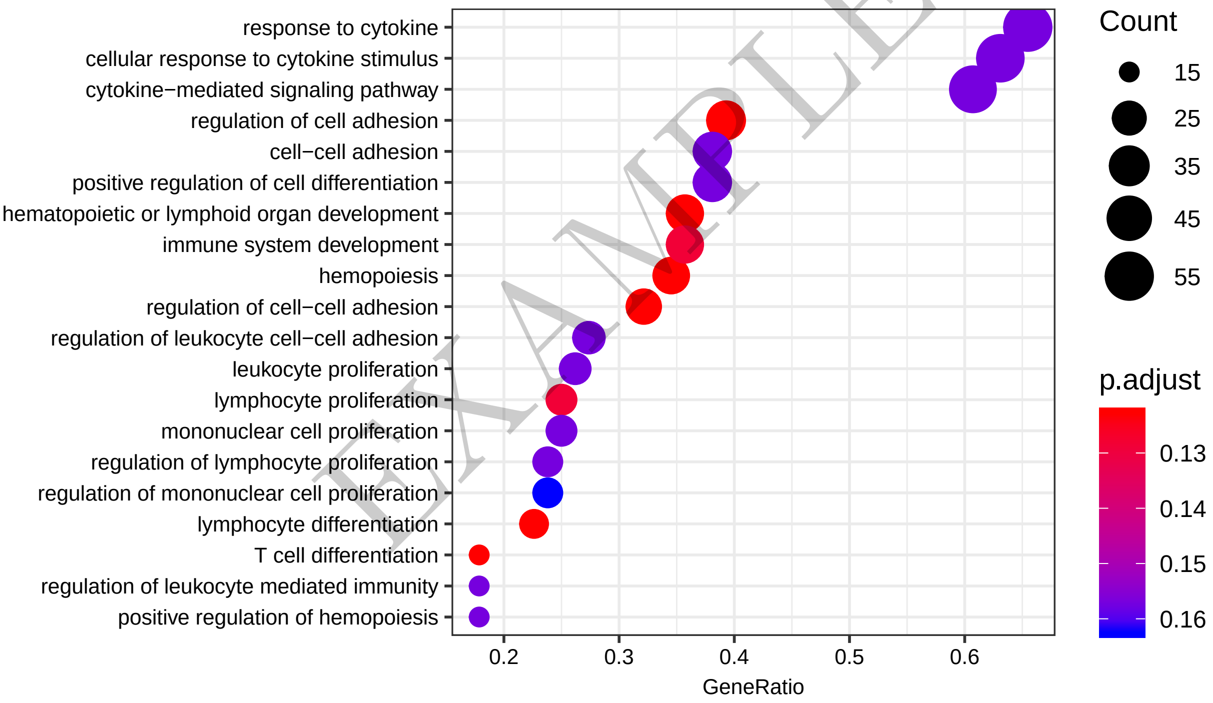


Figure 2: GO biological process over-representation analysis with 85 biomarkers between Control and Patient groups

3.2.3 GO cellular component over-representation analysis

Table 5: 10 GO cellular component basing on 85 differentially-expressed biomarkers between Control and Patient groups

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
GO:0005615	extracellular space	71/84	207/276	0.0101648	0.8777493	0.8777493
GO:0098552	side of membrane	19/84	42/276	0.0206123	0.8777493	0.8777493
GO:0044421	extracellular region part	71/84	211/276	0.0241582	0.8777493	0.8777493
GO:0009897	external side of plasma membrane	18/84	41/276	0.0346738	0.9448603	0.9448603
GO:0005768	endosome	11/84	27/276	0.1572636	1.0000000	1.0000000
GO:0044440	endosomal part	7/84	16/276	0.1791564	1.0000000	1.0000000
GO:0005788	endoplasmic reticulum lumen	8/84	19/276	0.1857301	1.0000000	1.0000000
GO:0005576	extracellular region	74/84	236/276	0.2707795	1.0000000	1.0000000
GO:0062023	collagen-containing extracellular matrix	9/84	24/276	0.2838064	1.0000000	1.0000000
GO:0031012	extracellular matrix	12/84	34/276	0.3179305	1.0000000	1.0000000

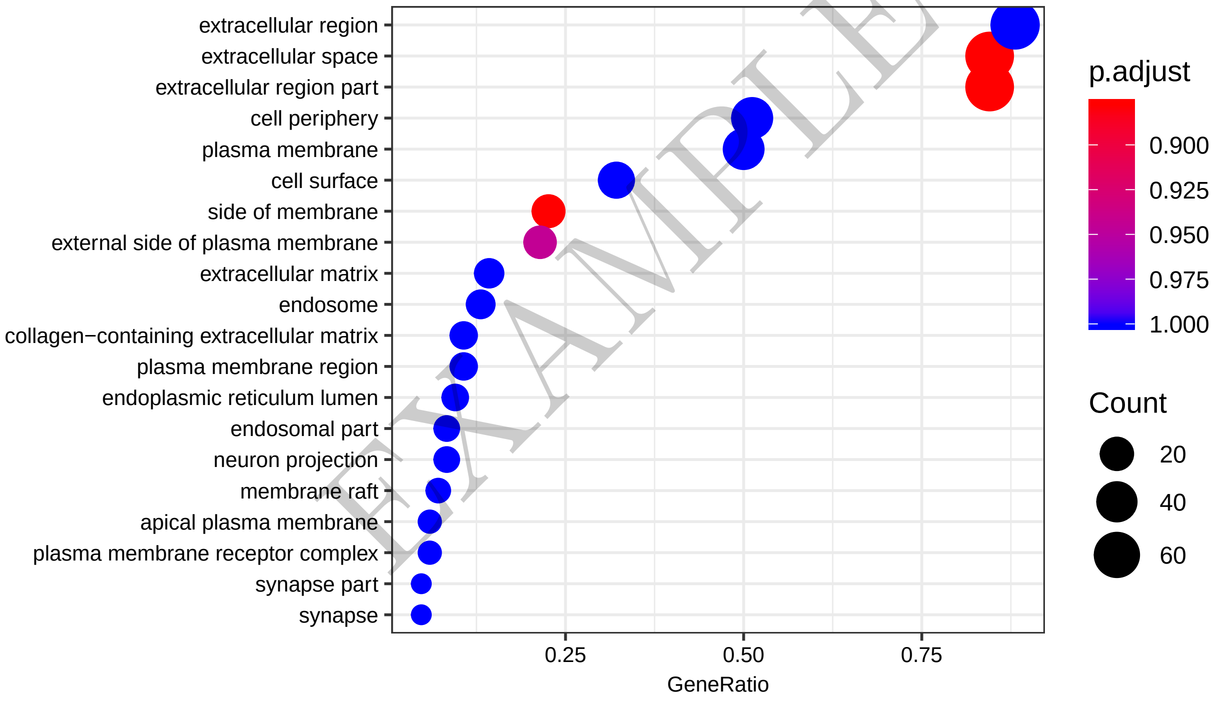


Figure 3: GO cellular component over-representation analysis with 85 biomarkers between Control and Patient groups

3.2.4 GO molecular function over-representation analysis

Table 6: 10 GO molecular function basing on 85 differentially-expressed biomarkers between Control and Patient groups

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
GO:0005125	cytokine activity	44/83	101/275	0.0002187	0.0196813	0.0186455
GO:0005515	protein binding	83/83	256/275	0.0008163	0.0367353	0.0348018
GO:0098772	molecular function regulator	60/83	161/275	0.0016136	0.0484068	0.0458591
GO:0030545	receptor regulator activity	53/83	141/275	0.0043425	0.0977073	0.0925648
GO:0005488	binding	83/83	262/275	0.0082469	0.1478903	0.1401066
GO:0005126	cytokine receptor binding	38/83	96/275	0.0098594	0.1478903	0.1401066
GO:0048018	receptor ligand activity	51/83	139/275	0.0121808	0.1566108	0.1483682
GO:0005102	signaling receptor binding	62/83	180/275	0.0225961	0.2542066	0.2408273
GO:0019955	cytokine binding	10/83	20/275	0.0435168	0.4351677	0.4122641
GO:0070851	growth factor receptor binding	16/83	37/275	0.0502414	0.4521726	0.4283740

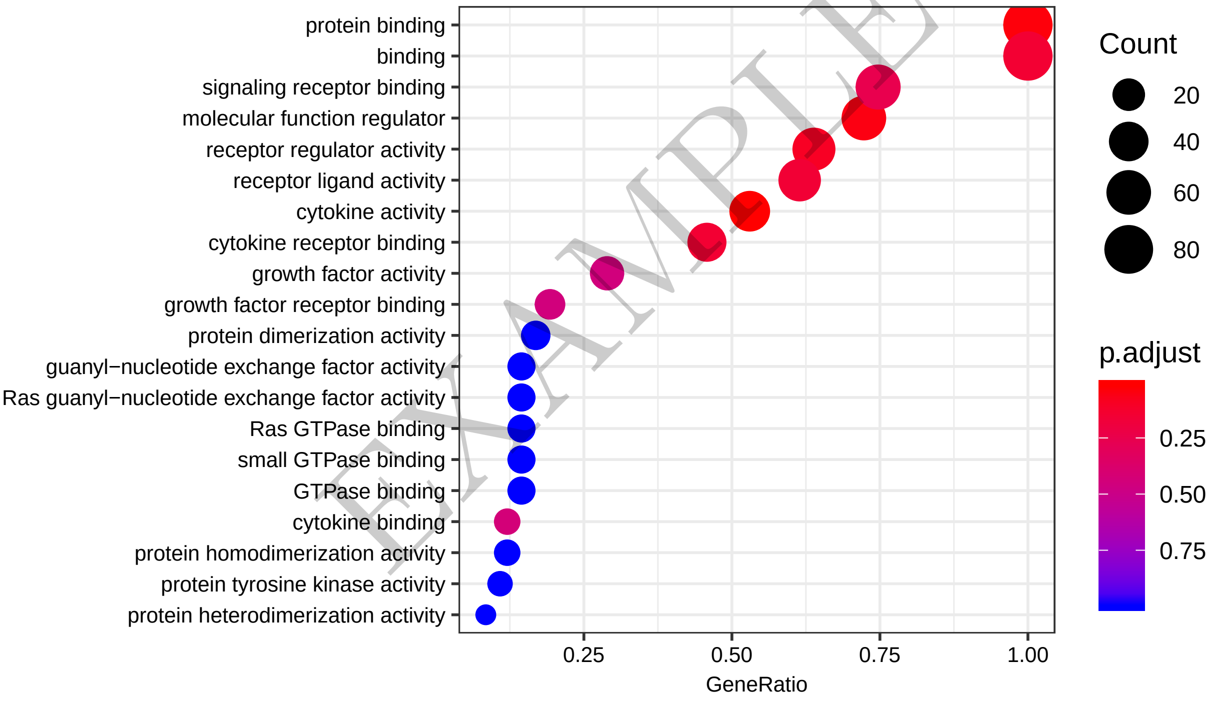


Figure 4: GO molecular function over-representation analysis with 85 biomarkers between Control and Patient groups

3.3 Gene Set Enrichment Analysis (GSEA)

3.3.1 KEGG Pathway Gene Set Enrichment Analysis

Table 7: 10 enriched KEGG pathway by GSEA with fold change of 277 biomarkers between Control and Patient groups

ID	Description	setSize	enrichmentScore	NES	pvalue	p.adjust
hsa04014	Ras signaling pathway	36	0.3617033	1.331017	0.0181818	0.3608247
hsa04940	Type I diabetes mellitus	12	-0.8251615	-1.595019	0.0282209	0.3608247
hsa04380	Osteoclast differentiation	13	-0.7907490	-1.552991	0.0425273	0.3608247
hsa04066	HIF-1 signaling pathway	18	-0.7597204	-1.527543	0.0492554	0.3608247
hsa05132	Salmonella infection	11	-0.7978870	-1.512208	0.0575000	0.3608247
hsa05166	Human T-cell leukemia virus 1 infection	18	-0.7437235	-1.495378	0.0652921	0.3608247
hsa04932	Non-alcoholic fatty liver disease (NAFLD)	13	-0.7694473	-1.511156	0.0680437	0.3608247
hsa05321	Inflammatory bowel disease (IBD)	23	-0.7140804	-1.450604	0.0691193	0.3608247
hsa05418	Fluid shear stress and atherosclerosis	21	-0.7179299	-1.456962	0.0725446	0.3608247
hsa04630	JAK-STAT signaling pathway	43	-0.6500179	-1.375273	0.0736515	0.3608247

3.3.2 GO biological process Gene Set Enrichment Analysis

Table 8: 10 enriched GO biological process by GSEA with fold change of 277 biomarkers between Control and Patient groups

ID	Description	setSize	enrichmentScore	NES	pvalue	p.adjust
GO:0043900	regulation of multi-organism process	22	-0.7912784	-1.614695	0.0055249	0.8134438
GO:0031330	negative regulation of cellular catabolic process	11	-0.8748342	-1.638318	0.0087173	0.8134438
GO:0043903	regulation of symbiosis, encompassing mutualism through parasitism	11	-0.8691080	-1.627595	0.0087173	0.8134438
GO:0009895	negative regulation of catabolic process	15	-0.8329189	-1.650569	0.0092700	0.8134438
GO:0010862	positive regulation of pathway-restricted SMAD protein phosphorylation	11	0.7160474	1.915406	0.0100503	0.8134438
GO:0007264	small GTPase mediated signal transduction	13	0.6858999	1.880901	0.0114286	0.8134438
GO:0007265	Ras protein signal transduction	13	0.6858999	1.880901	0.0114286	0.8134438
GO:0006959	humoral immune response	32	-0.7468140	-1.540305	0.0158061	0.8134438
GO:0043901	negative regulation of multi-organism process	12	-0.8433935	-1.605514	0.0158924	0.8134438
GO:0060395	SMAD protein signal transduction	12	0.6535576	1.731625	0.0163043	0.8134438

3.3.3 GO Cellular Component Gene Set Enrichment Analysis

Table 9: 10 enriched GO cellular component by GSEA with fold change of 277 biomarkers between Control and Patient groups

ID	Description	setSize	enrichmentScore	NES	pvalue	p.adjust
GO:0005829	cytosol	25	-0.7914919	-1.586257	0.0131004	0.8175000
GO:0044464	cell part	225	-0.6029880	-1.236879	0.0269730	0.8175000
GO:0005886	plasma membrane	133	-0.5967949	-1.264448	0.0270000	0.8175000
GO:0071944	cell periphery	134	-0.5944506	-1.258803	0.0300000	0.8175000
GO:0005634	nucleus	39	-0.6624543	-1.371132	0.0682057	0.9976152
GO:0043228	non-membrane-bounded organelle	14	-0.7807977	-1.500113	0.0739884	0.9976152
GO:0043232	intracellular non-membrane-bounded organelle	14	-0.7807977	-1.500113	0.0739884	0.9976152
GO:0036477	somatodendritic compartment	21	0.4247891	1.347954	0.0865385	0.9976152
GO:0016020	membrane	144	-0.5597711	-1.180944	0.0970000	0.9976152
GO:0031981	nuclear lumen	16	-0.7233025	-1.395023	0.1318807	0.9976152

3.3.4 GO Molecular Function Gene Set Enrichment Analysis

Table 10: 10 enriched GO molecular function by GSEA with fold change of 277 biomarkers between Control and Patient groups

ID	Description	setSize	enrichmentScore	NES	pvalue	p.adjust
GO:0043169	cation binding	38	-0.7655628	-1.575897	0.0041494	0.1867220
GO:0046872	metal ion binding	38	-0.7655628	-1.575897	0.0041494	0.1867220
GO:0005160	transforming growth factor beta receptor binding	11	0.7160474	1.891434	0.0151515	0.3718593
GO:0019900	kinase binding	13	-0.8479968	-1.620663	0.0181378	0.3718593
GO:0019901	protein kinase binding	11	-0.8654375	-1.632683	0.0223881	0.3718593
GO:0097159	organic cyclic compound binding	21	-0.7603426	-1.513556	0.0291153	0.3718593
GO:1901363	heterocyclic compound binding	21	-0.7603426	-1.513556	0.0291153	0.3718593
GO:0048020	CCR chemokine receptor binding	25	0.4579299	1.531686	0.0361446	0.3718593
GO:0043167	ion binding	75	-0.6346121	-1.341540	0.0371859	0.3718593
GO:0005515	protein binding	256	-0.6411434	-1.259755	0.0429570	0.3866134

3.4 Protein-protein interactions

The summary of protein-protein interactions of differentially-expressed biomarkers are shown in Figure 5 & 6.

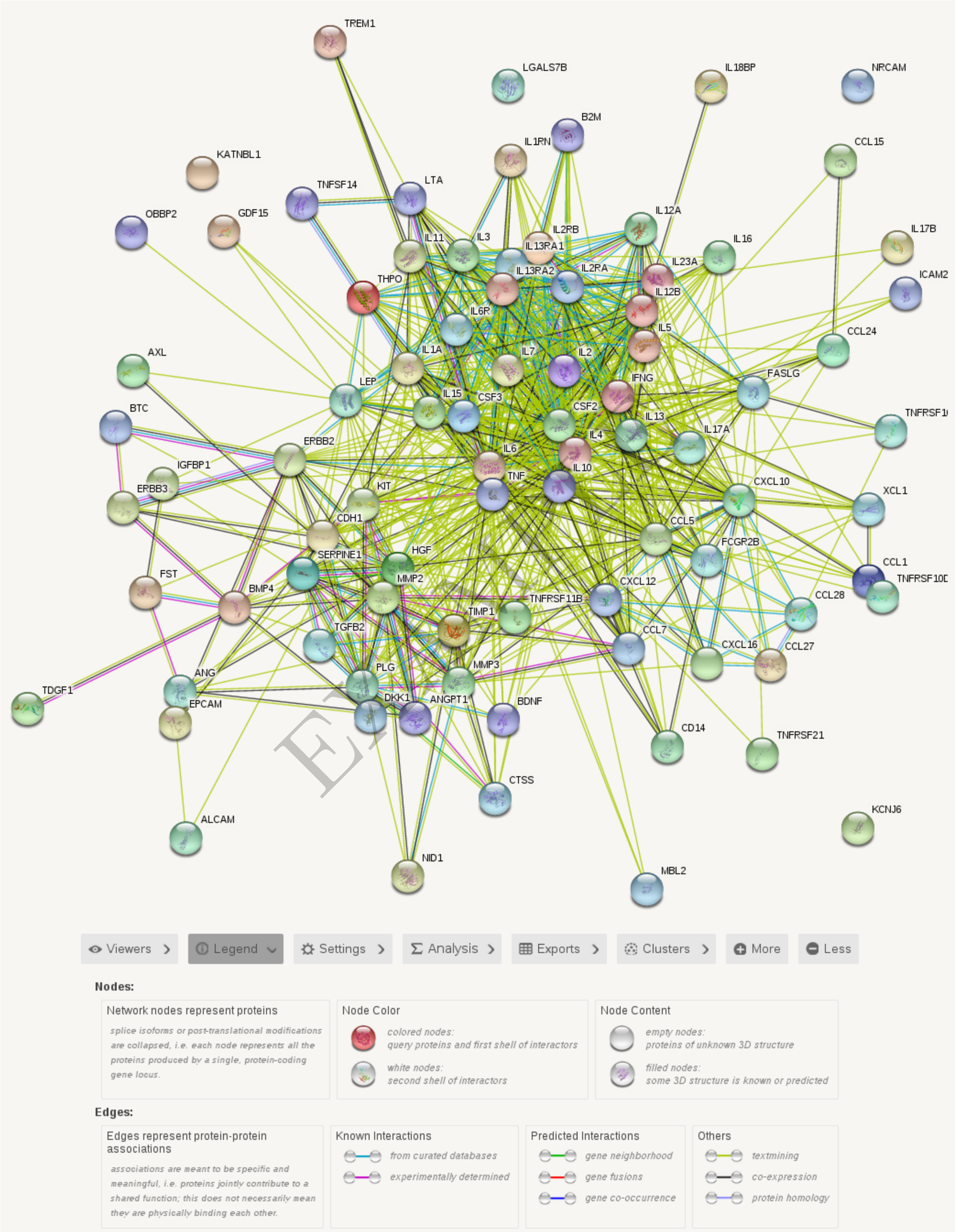
Geneset Analysis

Interaction Stats at Default Settings

number of nodes: 85
number of edges: 741
average node degree: 17.4
avg. local clustering coefficient: 0.692

expected number of edges: 108
PPI enrichment p-value: < 1.0e-16
your network has significantly more interactions than expected (what does that mean?)

Figure 5: Summary of protein-protein interactions of differentially-expressed biomarkers based on STRING database



Nodes:

Network nodes represent proteins

splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.

Node Color

colored nodes:
query proteins and first shell of interactors

white nodes:
second shell of interactors

Node Content

empty nodes:
proteins of unknown 3D structure

filled nodes:
some 3D structure is known or predicted

Edges:

Edges represent protein-protein associations

associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding each other.

Known Interactions

from curated databases
experimentally determined

Predicted Interactions

gene neighborhood
gene fusions
gene co-occurrence

Others

textmining
co-expression
protein homology

Figure 6: Protein-protein interactions of differentially-expressed biomarkers based on STRING database

References

Ashburner, M. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25: 25–29.

Kanehisa, Minoru, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. 2016. "KEGG as a Reference Resource for Gene and Protein Annotation." *Nucleic Acids Research* 44 (D1): D457–D462. https://doi.org/10.1093/nar/gkv1070.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Szklarczyk, D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, et al. 2017. "The String Database in 2017: Quality-Controlled Protein-Protein Association Networks, Made Broadly Accessible." *Nucleic Acids Res* 45 (January): D362–368.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. "ClusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters." *OMICS: A Journal of Integrative Biology* 16 (5): 284–87. https://doi.org/10.1089/omi.2011.0118.