Supplementary Material: Significance and Functional Similarity for Identification of Disease Genes

Pradipta Maji and Ekta Shah

Description of Data Sets

This section presents the brief descriptions of five microarray data sets, which are used to evaluate the performance of different criteria, namely, maximum significance-maximum functional similarity, t-test, maximum relevance (MR), minimum redundancy-maximum relevance (mRMR) [1], maximum relevance-maximum significance (MRMS) [2], maximum significance (MS) [2], maximum functional similarity (MFS), maximum relevance-maximum functional similarity (MRMFS) [3], and different integrated methods for disease gene selection, namely, RelSim [3], MR+PPIN [4], mRMR+PPIN [5], MRMS+PPIN [6], and CLAIM [7]. Majority of the experimental results on these five data sets are reported in the main paper, while a few of them are presented in this supplementary material. Each data set is preprocessed by standardizing each sample to zero mean and unit variance. The descriptions of these five microarray data sets used are as follows:

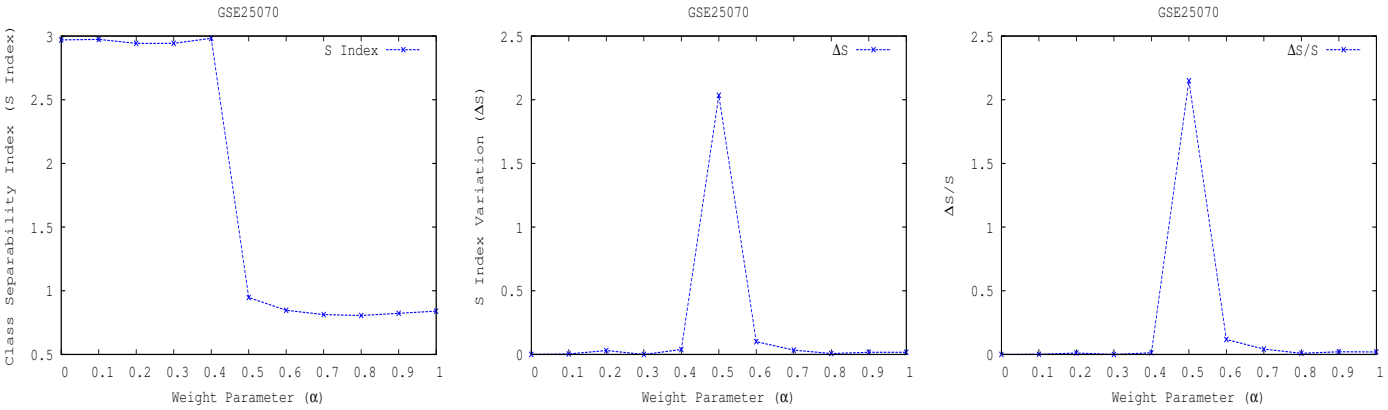
1) *GSE25070:* It is the gene expression data retrieved from study of Hinoue et al. [8]. The data set contains the expression profiles of 26 colorectal tumors matched histological to normal adjacent colonic tissue samples. Illumina Ref-8 whole-genome expression BeadChip with 24526 probes corresponding to 18491 genes was used to obtaing the gene expression profiles.
2) *GSE10950:* The gene expression data comapres 24 colon normal to tumor pairs, making use of the Illumina BeadChip Human Ref8-v2. The data had been put forward by Jiang et al. to study the role of DACT3 as an epigenetic regulator of Wnt/β-catenin signaling in colorectal cancer [9]. It contains the gene expression profiles of 22184 genes.
3) *GSE11223:* This data contains the transcriptional profiling of colon epithelial biopsies from ulcerative colitis patients and healthy control donors. It contains the expression profiles of 44290 probes corresponding to 40991 genes, for the 202 samples, of which 129 are diseased and 73 are non-diseased [10].
4) *Breast Cancer:* The breast cancer data set contains expression levels of 7129 genes in 49 breast tumor samples [11]. The samples are classified according to their estrogen receptor (ER) status: 25 samples are ER positive while other 24 samples are ER negative.
5) *Leukemia:* It is an Affymetrix high density oligonucleotide array that contains 7070 genes and 72 samples from two classes of leukemia [12]: 47 acute lymphoblastic leukemia and 25 acute myeloid leukemia.

Optimum Value of Weight Parameter

Fig. 1(a) plots the variation of class separability index $\mathcal{S}$ for different values of weight parameter α for GSE25070. Similarly, Fig. 1(b) and Fig. 1(c) plot the variation of the change in $\mathcal{S}$ index, that is $\Delta\mathcal{S}$, and $\mathcal{S}^{-1}\Delta\mathcal{S}$, respectively, for different values of weight parameter α. From all these figures, it can be seen that the optimum value of α is 0.5 for GSE25070. More results of the SiFS on five microarray data sets for different values of α are reported in Appendix of this material.

• *P. Maji and E. Shah are with the Biomedical Imaging and Bioinformatics Lab, Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India. E-mail: {pmaji,ekta_r}@isical.ac.in.*

This work is partially supported by the Department of Electronics and Information Technology, Government of India (PhD-MLA/4(90)/2015-16).
Digital Object Identifier no. 10.1109/TCBB.2016...



(a) (b) (c)

Fig. 1. Variation of $\mathcal{S}$ index, $\Delta\mathcal{S}$, and $\mathcal{S}^{-1}\Delta\mathcal{S}$ for different values of weight parameter α

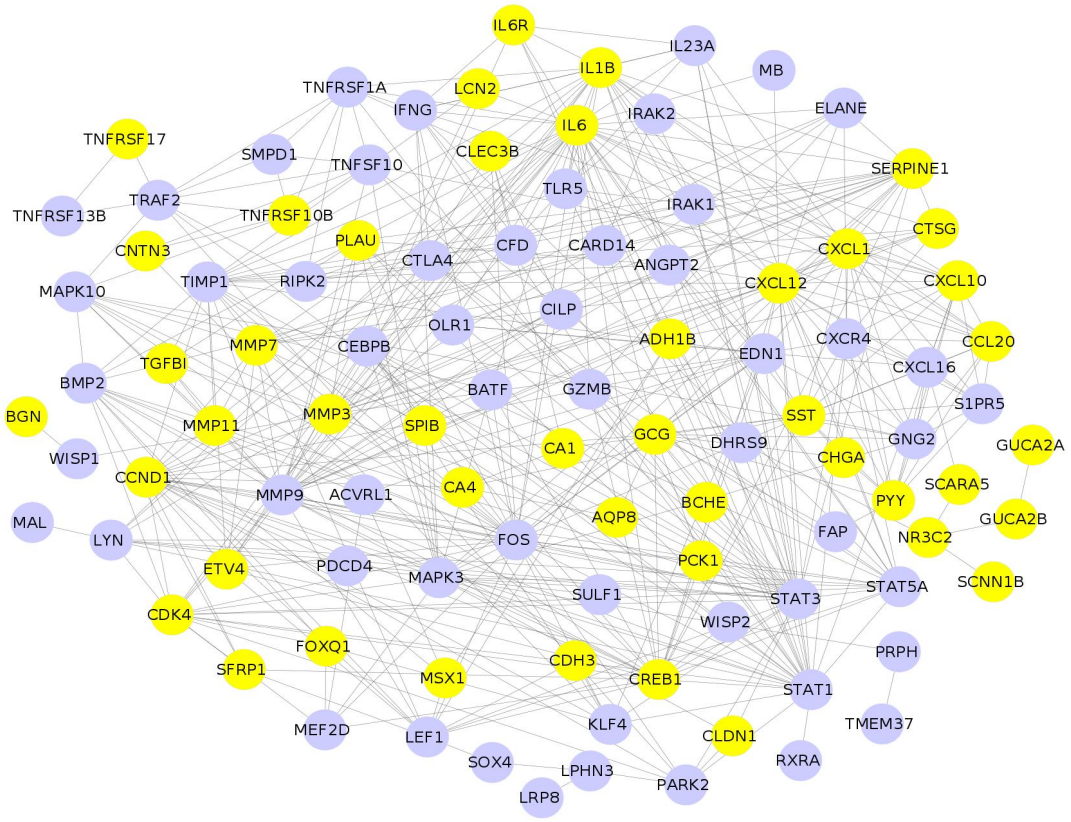


Fig. 2. PPI network for 100 genes obtained by the SiFS algorithm: genes marked in yellow are known disease genes

Example PPI Network

The PPI network is constructed for the set of genes selected by the MSMFS criterion based proposed SiFS algorithm. The network is generated using the STRING database. Fig. 2 shows the PPI network for the 100 genes selected using the proposed SiFS algorithm. The nodes in yellow represent the known disease genes, while those in violet mark the predicted disease genes. The constructed PPI network validates the concept of functional similarity introduced in the proposed work. Several small clusters of disease related genes can be found. For example, the genes **IL6**, **IL6R**, **IL1B**, and **LCN2** form a connected component. Similarly, **CXCL1**, **CXCL10**, **CXCL12**, and **CXCL20** represent another such component. Several other such associations can also be seen in the appended PPI network.

KEGG Enrichment Analysis on GSE25070 for SiFS

Table 1 reports the complete KEGG pathway analysis results on GSE25070 data for the 100 genes selected by the MSMFS criterion and weighted Cosine coefficient based proposed SiFS algorithm.

TABLE 1
KEGG Enrichment Analysis for SiFS

ID	Term	%	P-Value	Associated Genes
04668	TNF signaling pathway	13.64	1.57E-11	CCL20, CEBPB, CREB1, CXCL1, CXCL10, EDN1, FOS, IL1B, IL6, MAPK10, MAPK3, MMP3, MMP9, TNFRSF1A, TRAF2
04060	Cytokine-cytokine receptor interaction	6.42	7.40E-08	BMP2, CCL20, CXCL1, CXCL10, CXCL12, CXCL16, CXCR4, IFNG, IL1B,IL23A, IL6, IL6R, TNFRSF10B, TNFRSF13B, TNFRSF17, TNFRSF1A, TNFSF10
05152	Tuberculosis	7.34	1.75E-06	CEBPB, CREB1, IFNG, IL1B, IL23A, IL6, IRAK1, IRAK2, MAPK10, MAPK3, RIPK2, STAT1, TNFRSF1A
05323	Rheumatoid arthritis	10.99	1.78E-06	CCL20, CTLA4, CXCL1, CXCL12, FOS, IFNG, IL1B, IL23A, IL6, MMP3
05162	Measles	8.21	7.34E-06	CCND1, CDK4, IFNG, IL1B, IL6, IRAK1, STAT1, STAT3, STAT5A, TNFRSF10B, TNFSF10
05161	Hepatitis B	7.53	1.77E-05	CCND1, CDK4, CREB1, FOS, IL6, MAPK10, MAPK3, MMP9, STAT1, STAT3, STAT5A
05200	Pathways in cancer	4.28	3.00E-05	BMP2, CCND1, CDK4, CXCL12, CXCR4, FOS, GNG2, IL6, LEF1, MAPK10, MAPK3, MMP9, RXRA, STAT1, STAT3, STAT5A, TRAF2
04066	HIF-1 signaling pathway	8.74	6.77E-05	ANGPT2, EDN1, IFNG, IL6, IL6R, MAPK3, SERPINE1, STAT3, TIMP1
05142	Chagas disease (American trypanosomiasis)	8.65	7.35E-05	FOS, IFNG, IL1B, IL6, IRAK1, MAPK10, MAPK3, SERPINE1, TNFRSF1A
04620	Toll-like receptor signaling pathway	8.49	8.64E-05	CXCL10, FOS, IL1B, IL6, IRAK1, MAPK10, MAPK3, STAT1, TLR5
05132	Salmonella infection	9.30	1.75E-04	CXCL1, FOS, IFNG, IL1B, IL6, MAPK10, MAPK3, TLR5
04062	Chemokine signaling pathway	5.88	2.09E-04	CCL20, CXCL1, CXCL10, CXCL12, CXCL16, CXCR4, GNG2, LYN, MAPK3, STAT1, STAT3
05321	Inflammatory bowel disease (IBD)	10.45	3.51E-04	IFNG, IL1B, IL23A, IL6, STAT1, STAT3, TLR5
04380	Osteoclast differentiation	6.87	5.05E-04	CREB1, FOS, IFNG, IL1B, MAPK10, MAPK3, STAT1, TNFRSF1A, TRAF2
04917	Prolactin signaling pathway	9.72	5.70E-04	CCND1, FOS, MAPK10, MAPK3, STAT1, STAT3, STAT5A
05133	Pertussis	9.33	7.50E-04	FOS, IL1B, IL23A, IL6, IRAK1, MAPK10, MAPK3
05164	Influenza A	5.71	7.88E-04	CXCL10, IFNG, IL1B, IL6, MAPK10, MAPK3, STAT1, TNFRSF10B, TNFRSF1A, TNFSF10
04621	NOD-like receptor signaling pathway	10.53	1.64E-03	CXCL1, IL1B, IL6, MAPK10, MAPK3, RIPK2
04210	Apoptosis	8.24	1.71E-03	IL1B, IRAK1, IRAK2, TNFRSF10B, TNFRSF1A, TNFSF10, TRAF2
04064	NF-kappa B signaling pathway	7.69	2.68E-03	CXCL12, IL1B, IRAK1, LYN, PLAU, TNFRSF1A, TRAF2
05212	Pancreatic cancer	9.09	3.82E-03	CCND1, CDK4, MAPK10, MAPK3, STAT1, STAT3
05160	Hepatitis C	6.02	4.36E-03	CLDN1, MAPK10, MAPK3, RXRA, STAT1, STAT3, TNFRSF1A, TRAF2
04920	Adipocytokine signaling pathway	8.57	5.33E-03	MAPK10, PCK1, RXRA, STAT3, TNFRSF1A, TRAF2
05202	Transcriptional misregulation in cancer	5.03	6.06E-03	CEBPB, ELANE, ETV4, GZMB, IL6, MMP3, MMP9, PLAU, RXRA
05140	Leishmaniasis	8.33	6.25E-03	FOS, IFNG, IL1B, IRAK1, MAPK3, STAT1
04672	Intestinal immune network for IgA production	10.20	9.41E-03	CXCL12, CXCR4, IL6, TNFRSF13B, TNFRSF17
05216	Thyroid cancer	13.79	1.39E-02	CCND1, LEF1, MAPK3, RXRA
04630	Jak-STAT signaling pathway	5.06	1.45E-02	CCND1, IFNG, IL23A, IL6, IL6R, STAT1, STAT3, STAT5A
05145	Toxoplasmosis	5.83	1.54E-02	IFNG, IRAK1, MAPK10, MAPK3, STAT1, STAT3, TNFRSF1A
05221	Acute myeloid leukemia	8.77	1.94E-02	CCND1, LEF1, MAPK3, STAT3, STAT5A
05210	Colorectal cancer	8.06	2.87E-02	CCND1, FOS, LEF1, MAPK10, MAPK3
04068	FoxO signaling pathway	5.22	3.03E-02	CCND1, IL6, MAPK10, MAPK3, PCK1, STAT3, TNFSF10
05168	Herpes simplex infection	4.30	4.36E-02	FOS, IFNG, IL1B, IL6, MAPK10, STAT1, TNFRSF1A, TRAF2

This table shows that the gene set under consideration is associated to pathways like *"TNF-signaling pathway"*, *"cytokine-cytokine receptor interaction"*, *"JAK-STAT signaling pathway"*, *"Chemokine signaling pathway"*, etc., which are known to play an active role in colorectal cancer growth and metastasis. Several evidences exist to suggest the significance of cytokines in various biological responses, like hematopoiesis, oncogenesis, inflammation, cell growth monitoring, survival and differentiation. The *"cytokine- cytokine receptor interaction"* is known to induce a cascade of signaling pathways, for example, the interaction of TNFα and IL1β with their respective receptors activates the *"NF-kappa B signaling pathway"*, while IL-6 with its receptor gp130 activates STAT3, a major oncogenic transcription factor. Several other such interactions exist, which demonstrate the importance of cytokine interaction in colorectal cancer [13], [14]. The association of *"tuberculosis"* and cancer has ever since been controversial. There exist rare cases that mark the presence of both. Tanaka et al. [15] put forward a possibility that cancer in the colon may originate from a tuberculous lesion. Patients having a history of *"Rheumatoid arthritis"* are known to be at a lower risk of having colorectal cancer [16]. The selected set of genes also annotate to the *"Measles"* pathway. The measles virus, being an oncolytic virus, is known to have anti-tumor properties. Measles viruses are known to target $CD133^{+}$, an antigen found in abundance in colon cancer cell lines [17]. Patients having a history of thyroid cancer have also been known to develop malignancies in the gastrointestinal tract and colon regions [18], [19]. Abnormalities in regular *"apoptosis "* process is known to contribute to invasiveness of colorectal cancer and its resistance to chemotherapeutic drugs [20]. Hardwick et al. [21] showed that leptin, a member of the adipocytokine family, induces

cell proliferation and p42/44 MAPK phosphorylation. Thus, the *"adipocytokine signaling pathway"* is associated to the growth and spread of cancer. *"HIF-1 signaling pathway"*, mediated by hypoxia-inducible factor-1, functions to transactivate genes like, cathepsin D, MMP2, uPAR, FN1, etc., protein products of which play a pivotal role in tumor angiogenesis and glycolytic metabolism [22]. The *"Toll-like Receptor Signaling Pathway"* is important for microbiota-induced development of colitis-associated cancers, severity of which is known to be correlated to colorectal tumor development. This bacterial derived inflammation leads to progression from adenoma to invasive carcinoma [23]. *"Inflammatory bowel disease"* results in the release of cytokines and growth factors, expression of which is regulated by the *"NF-kappa B signaling pathway"*, which itself is involved in growth and development of cancer [24]. Prolactin, a cytokine secreted by many tissues, binds to its receptor to activate the JAK2-STAT3 and JAK2-ERK1/2 signaling within the cell, which result in an increased expression of Jagged1, a Notch-1 receptor ligand which regulates the colorectal cancer stem cell population. Thus, the *"Prolactin signaling pathway"* is involved in promoting the colorectal cancer growth [25]. The *"NOD-like receptor signaling pathway"* is another pathway that bears a significant association to the set of genes selected using the proposed method. The activation of this pathway initiates the recruitment of adaptor proteins, like RICK and CARD9, which leads to K-63 linked ubiquitylation and activation of the *"NF-kappa B signaling pathway"* and the MAP-kinase signaling cascade [26]. The *"JAK-STAT signaling pathway"* is known to be intensely involved in the growth and progression of colorectal cancer. They are known to be associated with cell growth, survival, invasion and motility through regulation of genes like BCL2, P16, VEGF, MMPs, etc. [27]. Thus, most of the pathways, annotated significantly by the selected set of genes, contribute in some form to the growth, proliferation, motility or survival of colorectal cancer cells.

Importance of Weighted Similarity Measure

The proposed similarity measure makes use of the fact that disease related genes always have a tendency to interact with each other. This association information has been captured using PPI networks and can be represented in weighted form in the PPI network database. As mentioned in the main paper, the proposed similarity measure reduces to Cosine coefficient if $\omega_{ij} \in \{0, 1\}$. However, a 0 or 1 association represents a uniform confidence in the interaction between the proteins. Hence, $\omega_{ij} \in [0, 1]$ may be a better source of information.

TABLE 2
Comparative Performance Between Weighted and Non-Weighted Similarity Measures

Data Sets	Similarity Measure	LIST 1		LIST 2		LIST 3		LIST 2-3	LIST 1-2-3
		Overlap	P-Value	Overlap	P-Value	Overlap	P-Value	Overlap	
GSE25070	Weighted (Wt)	12	6.59E-04	31	6.21E-28	17	3.62E-20	36	45
	Non-weighted (NWt)	2	9.14E-01	4	1.60E-01	4	3.34E-03	5	7
GSE10950	Weighted (Wt)	21	3.65E-10	3	3.52E-01	3	2.36E-02	5	26
	Non-weighted (NWt)	21	3.65E-10	2	6.25E-01	7	2.74E-06	7	28
GSE11223	Weighted (Wt)	21	3.65E-10	7	5.19E-03	7	2.74E-06	10	31
	Non-weighted (NWt)	27	1.48E-15	5	6.00E-02	5	3.77E-04	7	34

Data Sets	Mea-sure	Biological Process		KEGG Pathway		Disease Ontology	
		Term	P-Value	Term	P-Value	Term	P-Value
GSE 25070	Wt	response to lipid	2.0E-23	TNF signaling pathway	1.6E-11	colorectal cancer	8.8E-08
	NWt	nucleoside phosphate metabolic process	1.1E-23	Purine metabolism	2.1E-35	*	*
GSE 10950	Wt	response to cytokine	3.5E-36	Cytokine-cytokine receptor interaction	7.9E-23	colorectal cancer	3.4E-11
	NWt	regulation of immune response	1.4E-30	Cytokine-cytokine receptor interaction	1.1E-15	colorectal cancer	1.9E-12
GSE 11223	Wt	response to cytokine	2.8E-32	Jak-STAT signaling pathway	1.9E-19	colorectal cancer	6.4E-12
	NWt	positive regulation of immune system process	2.6E-27	Jak-STAT signaling pathway	1.1E-16	colorectal cancer	1.2E-13
Breast	Wt	positive regulation of immune system process	6.8E-33	Pathways in cancer	1.1E-21	breast cancer	3.6E-22
	NWt	regulation of cell proliferation	1.1E-38	Pathways in cancer	1.3E-22	breast cancer	4.0E-22
Leuk emia	Wt	leukocyte activation	5.9E-22	NF-kappa B signaling pathway	3.2E-12	leukemia	3.9E-25
	NWt	response to xenobiotic stimulus	8.7E-08	Metabolism of xenobiotics by cytochrome P450	1.0E-08	*	*

In order to establish the importance of weighted similarity measure over its non-weighted counterpart, extensive experimentation is carried out and the results are reported in Table 2. The comparative performance analysis between weighted and non-weighted similarity measures, with respect to the degree of overlapping with known gene lists, GO, and KEGG pathway based analysis, is provided in Table 2. From the results reported in the table, it can be seen that the numbers of overlapped genes for the proposed weighted similarity measure are better than its non-weighted counterpart in most of cases. Also, the proposed weighted measure attains significantly lower p-values than its non-weighted counterpart, irrespective of the gene lists used.

Table 2 also compares the performance of weighted and non-weighted similarity measures using the KEGG pathway enrichment analysis of the obtained gene sets. Analyzing Table 2, it is seen that the term annotated by the non-weighted measure for GSE25070 has the lowest p-value, but the term *"purine metabolism"* does not bear a significant association to colon cancer. On the other hand, although the terms annotated by both measures are

TABLE 3
Gene Ontology Based Analysis for 100 Genes Identified by Different Criteria

Different Data Sets	Different Criteria	Cellular Components		Molecular Functions	
		Term	P-Value	Term	P-Value
GSE25070	*t*-test	*	*	glycosaminoglycan binding	2.9E-05
	MR	*	*	carbonate dehydratase activity	3.1E-04
	MS	secretory granule lumen	6.7E-03	integrin binding	3.9E-05
	mRMR	histone acetyltransferase complex	2.3E-03	cargo receptor activity	1.4E-03
	MRMS	*	*	glycosaminoglycan binding	5.1E-06
	MFS	integrin complex	1.0E-03	cytokine activity	1.7E-07
	MRMFS	secretory granule lumen	3.5E-03	G-protein coupled receptor binding	1.1E-06
	MSMFS	secretory granule lumen	2.4E-04	cytokine receptor binding	1.6E-12
GSE10950	*t*-test	*	*	carboxy-lyase activity	2.0E-04
	MR	*	*	protein methyltransferase activity	1.0E-03
	MS	protein-DNA complex	2.4E-04	Wnt-activated receptor activity	2.3E-05
	mRMR	*	*	histone methyltransferase activity	2.9E-03
	MRMS	*	*	histone methyltransferase activity	2.5E-03
	MFS	membrane raft	7.0E-11	inorganic anion transmembrane transporter activity	3.6E-13
	MRMFS	external side of plasma membrane	3.8E-15	cytokine receptor binding	3.1E-09
	MSMFS	side of membrane	4.9E-23	cytokine receptor binding	9.6E-15
GSE11223	*t*-test	*	*	CCR chemokine receptor binding	1.9E-03
	MR	cytoplasmic mRNA processing body	6.0E-03	*	*
	MS	actin filament bundle	1.4E-02	protease binding	2.5E-03
	mRMR	cytoplasmic mRNA processing body	5.3E-03	snoRNA binding	1.1E-03
	MRMS	cytoplasmic mRNA processing body	5.5E-03	*	*
	MFS	external side of plasma membrane	1.3E-24	cytokine receptor binding	6.4E-30
	MRMFS	side of membrane	2.5E-14	cytokine receptor binding	2.6E-21
	MSMFS	external side of plasma membrane	3.2E-14	cytokine receptor binding	1.3E-22
Breast	*t*-test	extracellular matrix component	3.9E-05	core promoter binding	1.8E-05
	MR	MHC class II protein complex	5.8E-07	MHC class II receptor activity	2.1E-06
	MS	costamere	1.1E-03	antioxidant activity	3.2E-05
	mRMR	MHC class II protein complex	2.9E-05	MHC class II receptor activity	1.6E-04
	MRMS	MHC class II protein complex	6.4E-07	MHC class II receptor activity	2.0E-06
	MFS	side of membrane	1.3E-13	transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding	9.5E-18
	MRMFS	side of membrane	9.9E-14	cytokine receptor binding	6.9E-17
	MSMFS	side of membrane	1.1E-11	cytokine receptor binding	2.1E-14
Leukemia	*t*-test	nuclear chromatin	5.9E-07	enhancer binding	5.4E-04
	MR	mast cell granule	1.4E-06	protease binding	3.0E-03
	MS	mast cell granule	8.1E-07	peroxidase activity	5.3E-04
	mRMR	mast cell granule	1.4E-06	peroxidase activity	9.4E-03
	MRMS	mast cell granule	1.0E-06	peroxidase activity	7.5E-03
	MFS	I-kappaB/NF-kappaB complex	5.3E-12	oxidoreductase activity	1.3E-17
	MRMFS	I-kappaB/NF-kappaB complex	6.9E-15	antioxidant activity	2.1E-11
	MSMFS	I-kappaB/NF-kappaB complex	6.9E-15	antioxidant activity	6.5E-10

same for GSE10950, GSE11223, and breast cancer data, the p-values for weighted measure are significantly lower than non-weighted counterpart for two colon data sets and similar for breast cancer data. A similar analysis is reported in Table 2 based on GO, where the weighted measure is found to give better performance compared to non-weighted counterpart, in terms of biological process, and the annotated terms for the weighted measure are more relevant to the corresponding diseases compared to the non-weighted measure. Moreover, the sets of genes obtained using weighted measure annotate to DO terms for all data sets with significantly lower p-values, which bear close resemblance with related diseases, while non-weighted counterpart does not annotate to any term significantly for GSE25070 and leukemia.

Cellular Components and Molecular Functions

Table 3 compares the performance of different criteria with respect to cellular components and molecular functions. Similarly, Table 4 compares the performance of different weighted similarity measures, namely, Cosine, Dice, Geometric, Simpson, and Jaccard, with respect to cellular components and molecular functions. Finally, the performance of the proposed SiFS algorithm is compared with that of MR+PPIN [4], mRMR+PPIN [5], MRMS+PPIN [6], RelSim [3], and CLAIM [7] with respect to cellular components and molecular functions, and corresponding results are reported in Table 5.

REFERENCES

[1] C. Ding and H. Peng, "Minimum Redundancy Feature Selection from Microarray Gene Expression Data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185–205, 2005.

TABLE 4
Gene Ontology Based Analysis for 100 Genes Identified by Different Similarity Measures

Different Data Sets	Different Criteria	Cellular Components		Molecular Functions	
		Term	P-Value	Term	P-Value
GSE25070	Jaccard	secretory granule lumen	6.7E-03	G-protein coupled receptor binding	2.1E-22
	Simpson	pronucleus	4.7E-04	cytokine activity	1.6E-07
	Geometric	secretory granule lumen	6.7E-03	integrin binding	3.5E-05
	Dice	secretory granule lumen	3.5E-03	cytokine receptor binding	2.5E-08
	Cosine	secretory granule lumen	2.4E-04	cytokine receptor binding	1.6E-12
GSE10950	Jaccard	caveola	3.5E-03	Wnt-activated receptor activity	4.2E-12
	Simpson	actin cytoskeleton	2.6E-33	actin binding	2.3E-26
	Geometric	cytosolic ribosome	2.3E-35	structural constituent of ribosome	1.8E-30
	Dice	membrane microdomain	4.6E-08	frizzled binding	5.8E-17
	Cosine	side of membrane	4.9E-23	cytokine receptor binding	9.6E-15
GSE11223	Jaccard	melanosome	9.7E-04	cytokine receptor binding	1.2E-06
	Simpson	external side of plasma membrane	6.4E-18	cytokine receptor binding	4.8E-27
	Geometric	actin filament bundle	1.4E-02	CXCR chemokine receptor binding	8.9E-04
	Dice	external side of plasma membrane	4.9E-13	cytokine receptor binding	3.6E-24
	Cosine	external side of plasma membrane	3.2E-14	cytokine receptor binding	1.3E-22
Breast	Jaccard	catalytic step 2 spliceosome	1.3E-09	single-stranded DNA binding	2.3E-09
	Simpson	plasma membrane receptor complex	2.2E-12	cytokine receptor binding	3.2E-18
	Geometric	catalytic step 2 spliceosome	6.4E-10	RNA polymerase II activity	6.0E-06
	Dice	side of membrane	1.1E-12	transcription factor binding	4.2E-16
	Cosine	side of membrane	1.1E-11	cytokine receptor binding	2.1E-14
Leukemia	Jaccard	I-kappaB/NF-kappaB complex	7.7E-15	antioxidant activity	6.1E-10
	Simpson	I-kappaB/NF-kappaB complex	6.9E-15	antioxidant activity	4.8E-10
	Geometric	I-kappaB/NF-kappaB complex	1.2E-14	peroxidase activity	1.5E-07
	Dice	I-kappaB/NF-kappaB complex	3.5E-12	oxidoreductase activity	1.3E-17
	Cosine	I-kappaB/NF-kappaB complex	6.9E-15	antioxidant activity	6.5E-10

[2] P. Maji and S. Paul, "Rough Set Based Maximum Relevance-Maximum Significance Criterion and Gene Selection from Microarray Data," *International Journal of Approximate Reasoning*, vol. 52, no. 3, pp. 408–426, 2011.

[3] P. Maji, E. Shah, and S. Paul, "A New Similarity Measure for Identification of Disease Genes," in *Pattern Recognition and Machine Intelligence*, ser. Lecture Notes in Computer Science. Springer International Publishing, 2015, vol. 9124, pp. 451–461.

[4] P. Maji and S. Paul, *Scalable Pattern Recognition Algorithms: Applications in Computational Biology and Bioinformatics*. Springer-Verlag, London, April 2014, p. 304.

[5] B.-Q. Li, T. Huang, L. Liu, Y.-D. Cai, and K.-C. Chou, "Identification of Colorectal Cancer Related Genes with mRMR and Shortest Path in Protein-Protein Interaction Network," *PLoS ONE*, vol. 7, no. 4, p. e33393, 2012.

[6] S. Paul and P. Maji, "Gene Expression and Protein-Protein Interaction Data for Identification of Colon Cancer Related Genes Using f-Information Measures," *Natural Computing*, doi:10.1007/s11047-015-9485-6, 2016.

[7] S. Daniele, S. Aleksandra, Z. Agnieszka, K. Marta, B. Marek, B. Paola, and F. Giovanni, "An Integrated Approach (CLuster Analysis Integration Method) to Combine Expression Data and Protein-Protein Interaction Networks in Agrigenomics: Application on Arabidopsis thaliana," *OMICS: A Journal of Integrative Biology*, vol. 18, no. 2, pp. 155–165, 2014.

[8] T. Hinoue, D. J. Weisenberger, C. P. Lange, H. Shen, H. M. Byun, D. Van Den Berg, S. Malik, F. Pan, H. Noushmehr, C. M. van Dijk, R. A. E. M. Tollenaar, and P. W. Laird, "Genome-Scale Analysis of Aberrant DNA Methylation in Colorectal Cancer," *Genome Research*, vol. 22, no. 2, pp. 271–282, 2012.

[9] X. Jiang, J. Tan, J. Li, S. Kivime, X. Yang, L. Zhuang, P. L. Lee, M. T. Chan, L. W. Stanton, E. T. Liu, B. N. Cheyette, and Q. Yu, "{DACT3} Is an Epigenetic Regulator of Wnt/β-Catenin Signaling in Colorectal Cancer and Is a Therapeutic Target of Histone Modifications ," *Cancer Cell*, vol. 13, no. 6, pp. 529– 541, 2008.

[10] C. L. Noble, A. R. Abbas, J. Cornelius, C. W. Lees, G.-T. Ho, K. Toy, Z. Modrusan, N. Pal, F. Zhong, S. Chalasani, H. Clark, I. D. Arnott, I. D. Penman, J. Satsangi, and L. Diehl, "Regional Variation in Gene Expression in the Healthy Colon is Dysregulated in Ulcerative Colitis," *Gut*, vol. 57, no. 10, pp. 1398–1405, 2008.

[11] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J. A. Olson, J. R. Marks, and J. R. Nevins, "Predicting the Clinical Status of Human Breast Cancer by Using Gene Expression Profiles," *Proceedings of the National Academy of Sciences, USA*, vol. 98, no. 20, pp. 11 462–11 467, 2001.

[12] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring," *Science*, vol. 286, no. 5439, pp. 531–537, 1999.

[13] L. Klampfer, "Cytokines, Inflammation and Colon Cancer," *Current Cancer Drug Targets*, vol. 11, no. 4, pp. 451–464, 2011.

[14] T. Hirano, K. Ishihara, and M. Hibi, "Roles of STAT3 in Mediating the Cell Growth, Differentiation and Survival Signals Relayed Through the IL-6 Family of Cytokine Receptors Space," *Oncogene*, vol. 19, no. 21, pp. 2548–2556, 2000.

[15] Tanaka et al., "A Case of Colonic Carcinoma Associated with Intestinal Tuberculosis, and an Analysis of 26 cases Reported in Japan," *Japan Journal of Cancer Clinics*, vol. 33, no. 9, pp. 1117–1123, 1987.

[16] A. Rostom, C. Dube, G. Lewin, A. Tsertsvadze, N. Barrowman, C. Code, M. Sampson, and D. Moher, "Nonsteroidal Anti-inflammatory Drugs and Cyclooxygenase-2 Inhibitors for Primary Prevention of Colorectal Cancer: A Systematic Review Prepared for the U.S. Preventive Services Task Force," *Annals of Internal Medicine*, vol. 146, no. 5, pp. 376–389, 2007.

[17] P. Bach, T. Abel, C. Hoffmann, Z. Gal, G. Braun, I. Voelker, C. R. Ball, I. C. Johnston, U. M. Lauer, C. Herold-Mende, M. D. Muhlebach, H. Glimm, and C. J. Buchholz, "Specific Elimination of CD133+ Tumor Cells with Targeted Oncolytic Measles Virus," *Cancer Research*, vol. 73, no. 2, pp. 865–874, 2013.

[18] W. C. Hanna, T. A. Ponsky, G. D. Trachiotis, and S. M. Knoll, "Colon Cancer Metastatic to the Lung and the Thyroid Gland," *Archives of Surgery*, vol. 141, no. 1, pp. 93–96, 2006.

[19] C. Goatman, P. J. Goldsmith, V. Antonopoulos, and B. Ali, "Metastasis of Colorectal Adenocarcinoma to the Thyroid: A Case Report and Review of the Literature," *Case Reports in Surgery*, vol. 2012, 2012.

TABLE 5
Gene Ontology Based Analysis for Genes Identified by Different Integrated Methods

Different Data Sets	Different Methods	Cellular Components		Molecular Functions	
		Term	P-Value	Term	P-Value
GSE25070	MR+PPIN	histone deacetylase complex	8.5E-05	RNA polymerase II transcription factor binding	4.8E-09
	mRMR+PPIN	nuclear chromatin	1.4E-16	RNA polymerase II transcription factor binding	3.7E-14
	MRMS+PPIN	nuclear chromatin	3.9E-07	chromatin binding	1.3E-11
	CLAIM	platelet alpha granule	5.5E-03	carbohydrate transmembrane transporter activity	1.8E-03
	RelSim	apical plasma membrane	2.3E-07	G-protein coupled receptor binding	8.7E-06
	SiFS	secretory granule lumen	2.4E-04	cytokine receptor binding	1.7E-12
GSE10950	MR+PPIN	secretory granule membrane	6.9E-05	*	*
	mRMR+PPIN	secretory granule membrane	8.5E-04	*	*
	MRMS+PPIN	secretory granule membrane	6.7E-05	ubiquitin conjugating enzyme activity	3.3E-04
	CLAIM	dynein complex	2.9E-03	Wnt-activated receptor activity	5.0E-04
	RelSim	DNA-directed RNA polymerase II, core complex	4.6E-04	transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding	1.5E-03
	SiFS	side of membrane	4.9E-23	cytokine receptor binding	9.6E-15
GSE11223	MR+PPIN	cell surface furrow	5.4E-03	snRNA binding	7.2E-04
	mRMR+PPIN	cell surface furrow	2.8E-03	receptor signaling protein serine/threonine kinase activity	3.6E-04
	MRMS+PPIN	nuclear pore	4.9E-04	hormone activity	9.2E-05
	CLAIM	*	*	aminopeptidase activity	6.7E-03
	RelSim	external side of plasma membrane	3.4E-19	cytokine receptor binding	1.6E-25
	SiFS	external side of plasma membrane	3.2E-14	cytokine receptor binding	1.3E-22
Breast	MR+PPIN	platelet alpha granule	3.2E-04	integrin binding	4.4E-05
	mRMR+PPIN	*	*	integrin binding	2.2E-06
	MRMS+PPIN	cytoplasmic membrane-bounded vesicle lumen	1.6E-03	core promoter binding	1.1E-03
	CLAIM	cytosolic small ribosomal subunit	7.2E-15	structural constituent of ribosome	2.2E-11
	RelSim	sarcolemma	1.9E-03	enhancer sequence-specific DNA binding	1.9E-05
	SiFS	side of membrane	1.1E-11	cytokine receptor binding	2.1E-14
Leukemia	MR+PPIN	integrin complex	4.0E-04	inositol trisphosphate kinase activity	9.8E-05
	mRMR+PPIN	multivesicular body	1.0E-04	inositol trisphosphate kinase activity	2.0E-06
	MRMS+PPIN	integrin complex	4.0E-04	inositol trisphosphate kinase activity	8.8E-05
	CLAIM	I-kappaB/NF-kappaB complex	6.7E-05	transcriptional activator activity, RNA polymerase II distal enhancer sequence-specific binding	1.7E-03
	RelSim	mast cell granule	2.4E-06	peroxidase activity	5.5E-05
	SiFS	I-kappaB/NF-kappaB complex	6.9E-15	antioxidant activity	6.5E-10

[20] A. J. M. Watson, "Apoptosis and Colorectal Cancer," *Gut*, vol. 53, no. 11, pp. 1701–1709, 2004.

[21] J. C. Hardwick, L. L. Kodach, G. J. Offerhaus, and G. R. van den Brink, "Bone Morphogenetic Protein Signalling in Colorectal Cancer," *Nature Review Cancer*, vol. 8, no. 10, pp. 806–812, 2008.

[22] B. Krishnamachary, S. Berg-Dixon, B. Kelly, F. Agani, D. Feldser, G. Ferreira, N. Iyer, J. LaRusch, B. Pak, P. Taghavi, and G. L. Semenza, "Regulation of Colon Carcinoma Cell Invasion by Hypoxia-Inducible Factor 1," *Cancer Research*, vol. 63, no. 5, pp. 1138–1143, 2003.

[23] T.-T. Li, S. Ogino, and Z. R. Qian, "Toll-like Receptor Signaling in Colorectal Cancer: Carcinogenesis to Cancer Therapy," *World Journal of Gastroenterology*, vol. 20, no. 47, pp. 17 699–17 708, 2014.

[24] J. K. Triantafillidis, G. Nasioulas, and P. A. Kosmidis, "Colorectal Cancer and Inflammatory Bowel Disease: Epidemiology, Risk Factors, Mechanisms of Carcinogenesis and Prevention Strategies," *Anticancer Research*, vol. 29, no. 7, pp. 2727–2737, 2009.

[25] N. K. Neradugomma, D. Subramaniam, O. W. Tawfik, V. Goffin, T. R. Kumar, R. A. Jensen, and S. Anant, "Prolactin Signaling Enhances Colon Cancer Stemness by Modulating Notch Signaling In a Jak2-STAT3/ERK manner," *Carcinogenesis*, vol. 35, no. 4, pp. 795–806, 2014.

[26] H. Z. Md. Zaki and T.-D. Kanneganti, "NOD-Like Receptors in Colon Tumorigenesis," *Immunogastroenterology*, vol. 2, no. 2, pp. 90–95, 2013.

[27] H. Xiong, Z.-G. Zhang, X.-Q. Tian, D.-F. Sun, Q.-C. Liang, Y.-J. Zhang, R. Lu, Y.-X. Chen, and J.-Y. Fang, "Inhibition of JAK1, 2/STAT3 Signaling Induces Apoptosis, Cell Cycle Arrest, and Reduces Tumor Cell Invasion in Colorectal Cancer Cells ," *Neoplasia*, vol. 10, no. 3, pp. 287 – 297, 2008.

Appendix (Performance Analysis of SiFS)

Performance for Different Values of Weight Parameter

GSE25070	Biological Process	
Alpha	**Term**	**P-value**
0	response to lipopolysaccharide	1.10E-11
0.1	response to lipopolysaccharide	1.10E-11
0.2	response to lipopolysaccharide	1.15E-10
0.3	response to lipopolysaccharide	1.15E-10
0.4	response to lipopolysaccharide	1.62E-09
0.5	response to lipid	2.00E-23
0.6	extracellular matrix organization	3.09E-09
0.7	negative regulation of cell adhesion	5.72E-05
0.8	negative regulation of cell adhesion	5.37E-05
0.9	negative regulation of locomotion	2.10E-05
1	negative regulation of locomotion	2.00E-05

GSE10950	Biological Process	
Alpha	**Term**	**P-value**
0	positive regulation of cell proliferation	1.16E-23
0.1	inorganic anion transport	1.58E-14
0.2	regulation of immune response	3.67E-27
0.3	regulation of immune response	8.87E-32
0.4	response to cytokine	3.54E-36
0.5	Wnt signaling pathway	1.09E-20
0.6	Wnt signaling pathway	3.07E-15
0.7	SRP-dependent cotranslational protein targeting to membrane	2.10E-31
0.8	cellular response to nutrient	6.70E-05
0.9	cellular response to nutrient	6.76E-05
1	cellular extravasation	7.61E-05

GSE11223	Biological Process	
Alpha	**Term**	**P-value**
0	immune response	1.14E-54
0.1	immune response	1.14E-54
0.2	immune response	3.61E-53
0.3	response to cytokine	1.25E-41
0.4	positive regulation of immune system process	2.16E-42
0.5	regulation of immune system process	8.76E-44
0.6	positive regulation of immune system process	2.36E-42
0.7	response to cytokine	2.83E-32
0.8	regulation of lymphocyte activation	4.78E-08
0.9	intrinsic apoptotic signaling pathway	3.94E-06
1	intrinsic apoptotic signaling pathway in response to oxidative stress	5.58E-04

Performance for Different Values of Weight Parameter

Breast	Biological Process	
Alpha	**Term**	**P-value**
0	positive regulation of immune system process	2.37E-35
0.1	positive regulation of immune system process	6.48E-38
0.2	positive regulation of immune system process	3.36E-39
0.3	regulation of immune system process	3.58E-43
0.4	positive regulation of immune system process	6.84E-33
0.5	mRNA splicing, via spliceosome	7.70E-25
0.6	mRNA splicing, via spliceosome	1.56E-23
0.7	mRNA splicing, via spliceosome	3.39E-18
0.8	response to vitamin	1.10E-04
0.9	response to vitamin	1.05E-04
1	response to vitamin	2.94E-07

Leukemia	Biological Process	
Alpha	**Term**	**P-value**
0	response to molecule of bacterial origin	1.12E-20
0.1	response to xenobiotic stimulus	2.65E-19
0.2	leukocyte activation	5.93E-22
0.3	leukocyte activation	9.01E-20
0.4	leukocyte activation	1.21E-17
0.5	inflammatory response	5.75E-15
0.6	response to oxidative stress	6.32E-10
0.7	myeloid leukocyte activation	3.40E-08
0.8	myeloid leukocyte activation	3.95E-07
0.9	myeloid leukocyte activation	3.00E-05
1	myeloid leukocyte activation	3.00E-05

Performance for Different Values of Weight Parameter

GSE25070	Cellular Components	
Alpha	**Term**	**P-value**
0	integrin complex	1.04E-03
0.1	integrin complex	5.19E-03
0.2	integrin complex	5.04E-03
0.3	integrin complex	5.04E-03
0.4	pronucleus	4.87E-04
0.5	secretory granule lumen	2.44E-04
0.6	basal part of cell	6.88E-03
0.7	secretory granule lumen	6.69E-03
0.8	secretory granule lumen	6.69E-03
0.9	secretory granule lumen	6.69E-03
1	secretory granule lumen	6.69E-03

GSE10950	Cellular Components	
Alpha	**Term**	**P-value**
0	membrane raft	7.02E-11
0.1	apical part of cell	2.51E-10
0.2	side of membrane	1.12E-13
0.3	side of membrane	9.20E-16
0.4	side of membrane	4.88E-23
0.5	membrane microdomain	3.86E-07
0.6	caveola	2.90E-04
0.7	ribosome	3.04E-29
0.8	***	***
0.9	***	***
1	protein-DNA complex	2.37E-04

GSE11223	Cellular Components	
Alpha	**Term**	**P-value**
0	external side of plasma membrane	1.33E-24
0.1	external side of plasma membrane	1.33E-24
0.2	external side of plasma membrane	2.91E-23
0.3	external side of plasma membrane	7.62E-25
0.4	external side of plasma membrane	2.91E-23
0.5	external side of plasma membrane	2.91E-23
0.6	external side of plasma membrane	2.27E-19
0.7	external side of plasma membrane	3.19E-14
0.8	melanosome	1.62E-03
0.9	actin filament bundle	1.44E-02
1	actin filament bundle	1.44E-02

Performance for Different Values of Weight Parameter

Breast	Cellular Components	
Alpha	Term	P-value
0	side of membrane	1.30E-13
0.1	side of membrane	8.36E-16
0.2	side of membrane	1.30E-13
0.3	side of membrane	9.44E-15
0.4	side of membrane	1.09E-11
0.5	catalytic step 2 spliceosome	1.35E-09
0.6	catalytic step 2 spliceosome	6.77E-10
0.7	catalytic step 2 spliceosome	9.77E-09
0.8	basement membrane	1.67E-03
0.9	costamere	1.51E-03
1	costamere	1.13E-03

Leukemia	Cellular Components	
Alpha	Term	P-value
0	I-kappaB/NF-kappaB complex	5.31E-12
0.1	I-kappaB/NF-kappaB complex	4.72E-12
0.2	I-kappaB/NF-kappaB complex	6.87E-15
0.3	I-kappaB/NF-kappaB complex	8.59E-15
0.4	I-kappaB/NF-kappaB complex	1.12E-14
0.5	I-kappaB/NF-kappaB complex	1.12E-14
0.6	I-kappaB/NF-kappaB complex	7.08E-12
0.7	mast cell granule	2.42E-06
0.8	mast cell granule	2.02E-06
0.9	mast cell granule	8.08E-07
1	mast cell granule	8.08E-07

Performance for Different Values of Weight Parameter

GSE25070	Molecular Function	
Alpha	Term	P-value
0	cytokine activity	1.67E-07
0.1	cytokine activity	1.67E-07
0.2	cytokine activity	1.49E-07
0.3	cytokine activity	1.49E-07
0.4	cytokine activity	1.67E-07
0.5	cytokine receptor binding	1.65E-12
0.6	G-protein coupled receptor binding	5.98E-06
0.7	metalloendopeptidase activity	3.03E-05
0.8	metalloendopeptidase activity	2.78E-05
0.9	integrin binding	3.22E-05
1	integrin binding	3.86E-05

GSE10950	Molecular Function	
Alpha	Term	P-value
0	inorganic anion transmembrane transporter activity	3.59E-13
0.1	inorganic anion transmembrane transporter activity	8.73E-18
0.2	cytokine receptor binding	2.94E-11
0.3	cytokine receptor binding	1.18E-14
0.4	cytokine receptor binding	9.58E-15
0.5	Wnt-activated receptor activity	1.04E-11
0.6	Wnt-activated receptor activity	5.21E-12
0.7	structural constituent of ribosome	5.95E-26
0.8	Wnt-activated receptor activity	2.32E-05
0.9	Wnt-activated receptor activity	3.09E-05
1	Wnt-activated receptor activity	2.32E-05

GSE11223	Molecular Function	
Alpha	Term	P-value
0	cytokine receptor binding	6.38E-30
0.1	cytokine receptor binding	6.38E-30
0.2	cytokine receptor binding	1.87E-31
0.3	cytokine receptor binding	4.21E-30
0.4	cytokine receptor binding	6.98E-30
0.5	cytokine receptor binding	1.38E-25
0.6	cytokine receptor binding	6.68E-30
0.7	cytokine receptor binding	1.26E-22
0.8	chemokine receptor binding	3.80E-04
0.9	cription factor activity, RNA polymerase II distal enhancer sequence-specific bir	2.43E-03
1	protease binding	2.51E-03

Performance for Different Values of Weight Parameter

Breast	Molecular Function	
Alpha	Term	P-value
0	activator activity, RNA polymerase II transcription regulatory region sequence-s	9.47E-18
0.1	cytokine receptor binding	3.30E-18
0.2	cytokine receptor binding	4.40E-18
0.3	cytokine receptor binding	4.28E-18
0.4	cytokine receptor binding	2.07E-14
0.5	RNA polymerase II activity	3.46E-08
0.6	RNA polymerase II activity	6.01E-06
0.7	enhancer sequence-specific DNA binding	1.48E-05
0.8	enhancer sequence-specific DNA binding	1.64E-04
0.9	enhancer sequence-specific DNA binding	1.49E-04
1	antioxidant activity	3.18E-05

Leukemia	Molecular Function	
Alpha	Term	P-value
0	oxidoreductase activity	1.28E-17
0.1	antioxidant activity	6.99E-10
0.2	antioxidant activity	6.53E-10
0.3	antioxidant activity	7.29E-10
0.4	antioxidant activity	5.62E-10
0.5	peroxidase activity	3.65E-09
0.6	peroxidase activity	1.19E-07
0.7	peroxidase activity	8.56E-04
0.8	peroxidase activity	6.42E-04
0.9	peroxidase activity	5.35E-04
1	peroxidase activity	5.35E-04

Performance for Different Values of Weight Parameter

GSE25070	KEGG	
Alpha	**Term**	**P-value**
0	Signaling pathways regulating pluripotency of stem cells	7.08E-06
0.1	Signaling pathways regulating pluripotency of stem cells	8.30E-06
0.2	Signaling pathways regulating pluripotency of stem cells	7.08E-06
0.3	Signaling pathways regulating pluripotency of stem cells	7.08E-06
0.4	Signaling pathways regulating pluripotency of stem cells	7.08E-06
0.5	TNF signaling pathway	1.57E-11
0.6	TNF signaling pathway	1.57E-06
0.7	Pentose and glucuronate interconversions	3.26E-03
0.8	Pentose and glucuronate interconversions	2.99E-03
0.9	Pentose and glucuronate interconversions	1.78E-03
1	Pentose and glucuronate interconversions	2.55E-04

GSE10950	KEGG	
Alpha	**Term**	**P-value**
0	Hepatitis B	2.02E-13
0.1	Hepatitis B	1.76E-10
0.2	Cytokine-cytokine receptor interaction	2.29E-12
0.3	Cytokine-cytokine receptor interaction	4.00E-16
0.4	Cytokine-cytokine receptor interaction	7.88E-23
0.5	Pathways in cancer	4.48E-20
0.6	Wnt signaling pathway	8.77E-15
0.7	Ribosome	1.21E-21
0.8	African trypanosomiasis	8.20E-03
0.9	African trypanosomiasis	7.05E-03
1	Nucleotide excision repair	1.12E-02

GSE11223	KEGG	
Alpha	**Term**	**P-value**
0	Cytokine-cytokine receptor interaction	9.20E-29
0.1	Cytokine-cytokine receptor interaction	9.20E-29
0.2	Cytokine-cytokine receptor interaction	1.41E-30
0.3	Cytokine-cytokine receptor interaction	3.41E-29
0.4	Cytokine-cytokine receptor interaction	3.09E-30
0.5	Cytokine-cytokine receptor interaction	1.49E-27
0.6	Cytokine-cytokine receptor interaction	6.59E-29
0.7	Jak-STAT signaling pathway	1.93E-19
0.8	Pathways in cancer	1.45E-08
0.9	Pathways in cancer	4.24E-06
1	Prostate cancer	5.15E-04

Performance for Different Values of Weight Parameter

Breast	KEGG	
Alpha	**Term**	**P-value**
0	Pathways in cancer	2.89E-23
0.1	Pathways in cancer	2.92E-23
0.2	Pathways in cancer	3.49E-24
0.3	Pathways in cancer	2.15E-22
0.4	Pathways in cancer	1.06E-21
0.5	Epstein-Barr virus infection	1.17E-07
0.6	Spliceosome	2.73E-04
0.7	Pertussis	3.33E-04
0.8	Pertussis	2.58E-03
0.9	Leishmaniasis	1.70E-02
1	Leishmaniasis	1.68E-04

Leukemia	KEGG	
Alpha	**Term**	**P-value**
0	Drug metabolism	2.89E-10
0.1	Drug metabolism	3.30E-10
0.2	NF-kappa B signaling pathway	3.19E-12
0.3	Pathways in cancer	8.92E-12
0.4	Osteoclast differentiation	8.24E-11
0.5	Osteoclast differentiation	4.69E-12
0.6	Osteoclast differentiation	4.11E-09
0.7	Osteoclast differentiation	9.27E-09
0.8	Osteoclast differentiation	6.08E-06
0.9	Epstein-Barr virus infection	3.42E-03
1	Epstein-Barr virus infection	6.98E-04

Performance for Different Values of Weight Parameter

GSE25070	Disease Ontology		Area Under Curve (n =100)	
Alpha	Term	P-value	ROC	PR
0	colorectal cancer	2.57E-02	0.991	0.953
0.1	colorectal cancer	2.55E-02	0.991	0.953
0.2	colorectal cancer	2.46E-02	0.991	0.953
0.3	colorectal cancer	2.46E-02	0.991	0.953
0.4	colorectal cancer	2.59E-02	0.991	0.953
0.5	colorectal cancer	8.81E-08	0.999	0.960
0.6	colorectal cancer	4.48E-04	0.999	0.960
0.7	colorectal cancer	1.32E-02	0.999	0.960
0.8	colorectal cancer	3.53E-02	0.999	0.960
0.9	gastrointestinal system disease	6.61E-05	0.999	0.960
1	gastrointestinal system disease	2.62E-04	0.999	0.960

GSE10950	Disease Ontology		Area Under Curve (n =100)	
Alpha	Term	P-value	ROC	PR
0	colorectal cancer	2.40E-12	1	0.958
0.1	colorectal cancer	3.37E-09	1	0.958
0.2	colorectal cancer	9.68E-14	1	0.958
0.3	colorectal cancer	1.61E-12	1	0.958
0.4	colorectal cancer	3.36E-11	1	0.958
0.5	colorectal cancer	1.12E-07	1	0.958
0.6	colorectal cancer	3.44E-03	1	0.958
0.7	***	***	1	0.958
0.8	***	***	1	0.958
0.9	***	***	1	0.958
1	***	***	1	0.958

GSE11223	Disease Ontology		Area Under Curve (n =100)	
Alpha	Term	P-value	ROC	PR
0	colorectal cancer	5.70E-08	0.987	0.970
0.1	colorectal cancer	5.70E-08	0.987	0.970
0.2	colorectal cancer	4.80E-08	0.987	0.971
0.3	colorectal cancer	6.47E-09	0.986	0.971
0.4	colorectal cancer	4.95E-08	0.987	0.972
0.5	colorectal cancer	4.78E-08	0.987	0.974
0.6	colorectal cancer	1.60E-10	0.987	0.975
0.7	colorectal cancer	6.40E-12	0.988	0.976
0.8	colorectal cancer	1.08E-05	0.987	0.977
0.9	colorectal cancer	2.68E-02	0.987	0.978
1	colorectal cancer	5.70E-08	0.987	0.977

Performance for Different Values of Weight Parameter

Breast	Disease Ontology		Area Under Curve (n =100)	
Alpha	Term	P-value	ROC	PR
0	breast cancer	4.75E-21	0.817	0.809
0.1	breast cancer	4.23E-21	0.780	0.791
0.2	breast cancer	4.00E-22	0.817	0.809
0.3	breast cancer	3.84E-22	0.907	0.879
0.4	breast cancer	3.61E-22	0.967	0.931
0.5	breast cancer	5.25E-03	0.683	0.704
0.6	breast cancer	1.11E-03	0.880	0.858
0.7	breast cancer	4.57E-03	0.867	0.844
0.8	breast cancer	8.36E-04	0.960	0.925
0.9	breast cancer	4.63E-03	0.953	0.918
1	breast cancer	1.94E-02	0.945	0.910

Leukemia	Disease Ontology		Area Under Curve (n =100)	
Alpha	Term	P-value	ROC	PR
0	leukemia	1.87E-19	0.954	0.955
0.1	leukemia	6.52E-19	0.952	0.955
0.2	leukemia	3.90E-25	0.964	0.961
0.3	leukemia	3.38E-22	0.982	0.970
0.4	leukemia	4.08E-17	0.969	0.963
0.5	leukemia	5.96E-16	0.975	0.966
0.6	leukemia	3.60E-11	0.984	0.970
0.7	leukemia	1.06E-07	0.987	0.972
0.8	leukemia	2.15E-06	0.991	0.974
0.9	leukemia	6.95E-04	0.992	0.975
1	leukemia	5.30E-04	0.992	0.975

Performance for Different Values of Weight Parameter

GSE25070	List-1		List-2		List-3		List-2-3	List-1-2-3
Alpha	Overlap	P-value	Overlap	P-value	Overlap	P-value	Overlap	Overlap
0	9	1.90E-02	5	6.00E-02	6	3.50E-05	10	19
0.1	9	1.90E-02	5	6.00E-02	6	3.50E-05	10	19
0.2	9	1.79E-02	5	5.79E-02	6	3.30E-05	10	19
0.3	9	1.79E-02	5	5.79E-02	6	3.30E-05	10	19
0.4	10	6.83E-03	5	6.00E-02	6	3.50E-05	10	20
0.5	12	6.59E-04	31	6.21E-28	17	3.62E-20	36	45
0.6	12	6.59E-04	32	2.68E-29	12	1.11E-12	34	42
0.7	10	6.83E-03	31	6.21E-28	11	2.66E-11	32	38
0.8	9	1.90E-02	29	2.89E-25	11	2.66E-11	30	37
0.9	8	4.78E-02	30	1.37E-26	11	2.66E-11	31	37
1	8	4.78E-02	31	6.21E-28	11	2.66E-11	32	38

GSE10950	List-1		List-2		List-3		List-2-3	List-1-2-3
Alpha	Overlap	P-value	Overlap	P-value	Overlap	P-value	Overlap	Overlap
0	19	1.45E-08	5	6.00E-02	5	3.77E-04	9	28
0.1	15	1.04E-05	5	6.00E-02	5	3.77E-04	9	24
0.2	15	1.04E-05	3	3.52E-01	7	2.74E-06	8	23
0.3	17	4.44E-07	3	3.52E-01	7	2.74E-06	8	25
0.4	19	1.45E-08	6	1.90E-02	8	1.85E-07	11	30
0.5	22	5.30E-11	3	3.52E-01	2	1.25E-01	5	27
0.6	19	1.45E-08	4	1.60E-01	2	1.25E-01	5	24
0.7	6	2.13E-01	1	8.81E-01	0	1.00E+00	1	7
0.8	10	6.83E-03	4	1.60E-01	3	2.36E-02	5	15
0.9	12	6.59E-04	4	1.60E-01	3	2.36E-02	5	17
1	12	6.59E-04	4	1.60E-01	3	2.36E-02	5	17

GSE11223	List-1		List-2		List-3		List-2-3	List-1-2-3
Alpha	Overlap	P-value	Overlap	P-value	Overlap	P-value	Overlap	Overlap
0	15	1.04E-05	6	1.90E-02	7	2.74E-06	10	25
0.1	15	1.04E-05	6	1.90E-02	7	2.74E-06	10	25
0.2	15	1.04E-05	6	1.90E-02	7	2.74E-06	10	25
0.3	17	4.44E-07	6	1.90E-02	6	3.50E-05	9	26
0.4	19	1.45E-08	6	1.90E-02	5	3.77E-04	8	27
0.5	22	5.30E-11	5	6.00E-02	4	3.34E-03	7	29
0.6	21	3.65E-10	5	6.00E-02	5	3.77E-04	8	29
0.7	21	3.65E-10	7	5.19E-03	7	2.74E-06	10	31
0.8	21	3.65E-10	6	1.90E-02	6	3.50E-05	8	29
0.9	18	8.28E-08	3	3.52E-01	4	3.34E-03	5	23
1	15	1.04E-05	4	1.60E-01	4	3.34E-03	6	21