

NGS : DOWNSTREAM PROCESSING

The plan...

- Downstream analysis
 - Gene sets
 - Transcriptome studies
 - Cistrome/Epigenome studies
 - Function enrichment
 - GSEA, DAVID, IPA
 - STRING, Cytoscape
- Downstream processing
 - Genome coordinate sets
 - NGS alignment files (reads/tags)
 - ChIP-Seq peaks
 - BEDTools
 - Galaxy
 - UCSC Table browser

GSEA | MSigDB | Investigate Gene Sets

<http://www.broadinstitute.org/gsea/msigdb/annotate.jsp>
Google

logged in as nagarajanv@mail.nih.gov
[logout](#)

Gene Set Enrichment Analysis

[GSEA Home](#)
[Downloads](#)
[Molecular Signatures Database](#)
[Documentation](#)
[Contact](#)

[MSigDB Home](#)
[About Collections](#)
[Browse Gene Sets](#)
[Search Gene Sets](#)
[Investigate Gene Sets](#)
[View Gene Families](#)
[Help](#)

Investigate Gene Sets

Gain further insight into the biology behind a gene set by using the following tools:

- compute overlaps** with other gene sets in MSigDB ([more...](#))
- display the gene set expression profile** based on a selected compendium of expression data ([more...](#))
- categorize** members of the gene set by gene families ([more...](#))

Gene Identifiers

AIM1
BEX4
BTBD11
C12ORF29
C2ORF24
C2ORF54
C6ORF192
CCNYL1
CD24
CDC2L6
CITED2
CROT
CUL4B
CYP3A5
DHRS1
ERO1L
FBXO34
FTL
HBP1
HIBCH
IKZF2
IL13RA1
IL8RB
INPP1
LMBRD1
LTA4H
MOSPD1

Compute Overlaps

☐ C1: positional gene sets [?](#)
☐ C2: curated gene sets [?](#)
☐ CGP: chemical and genetic perturbations [?](#)
☐ CP: canonical pathways [?](#)
☐ CP:BIOCARTA: BioCarta gene sets [?](#)
☐ CP:KEGG: KEGG gene sets [?](#)
☐ CP:REACTOME: Reactome gene sets [?](#)
☐ C3: motif gene sets [?](#)
☐ MIR: microRNA targets [?](#)
☐ TFT: transcription factor targets [?](#)
☐ C4: computational gene sets [?](#)
☐ CGN: cancer gene neighborhoods [?](#)
☐ CM: cancer modules [?](#)
☐ C5: GO gene sets [?](#)
☐ BP: GO biological process [?](#)
☐ CC: GO cellular component [?](#)
☐ MF: GO molecular function [?](#)

show genesets
[compute overlaps](#)

Compendia expression profiles

☒ Human tissue compendium (Novartis)
☐ Global Cancer Map (Broad Institute)
☐ NCI-60 cell lines (National Cancer Institute)

[display expression profile](#)

Gene families

[show gene families](#)

Gene Identifier Platform

[Broad Home](#) | [Cancer Genomics](#)

MSigDB database v3.0 updated Sep 9, 2010
GSEA/MSigDB web site v3.82 released Oct 7, 2011

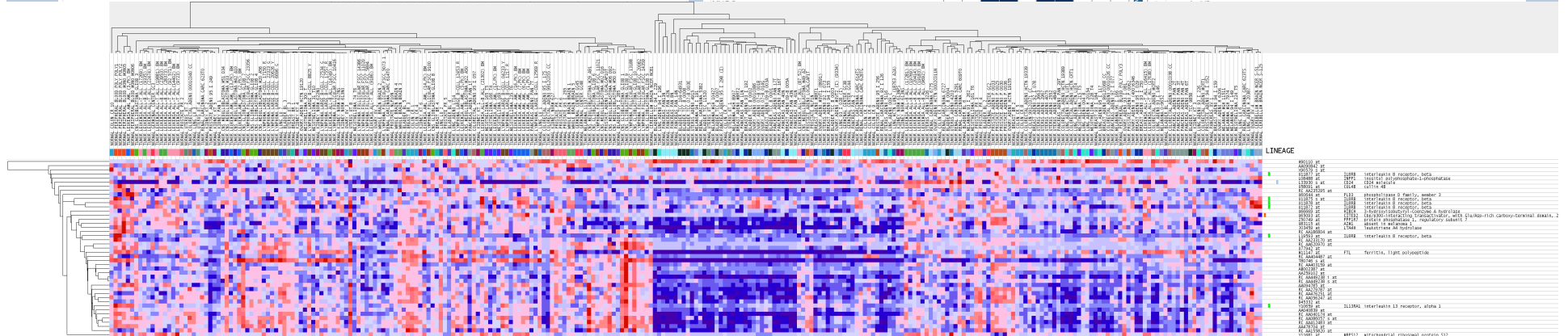
[Display a menu](#)

GSEA

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p value
CREIGHTON_ENDOCRINE_THERAPY_RESISTANCE_NCE_3 [723]	The 'group 3 set' of genes associated with acquired endocrine therapy resistance in breast tumors expressing ESR1 and ERBB2 [Gene ID=2099, 2064].	10		2.12 e ⁻⁸
RODRIGUES_THYROID_CARCINOMA_POORLY_DIF_DIFFERENTIATED_DN [796]	Genes down-regulated in poorly differentiated thyroid carcinoma (PDTc) compared to normal thyroid tissue.	9		6.49 e ⁻⁷
chr6q13 [23]	Genes in cytogenetic band chr6q13	3		3.52 e ⁻⁶
COLDREN_GEFITINIB_RESISTANCE_DN [228]	Genes down-regulated in NSCLC (non-small cell lung carcinoma) cell lines resistant to gefitinib [PubChem=123631] compared to the sensitive ones.	5		1.13 e ⁻⁵
NAKAMURA_TUMOR_ZONE_PERIPHERAL_VS_CENTRAL_DN [639]	Down-regulated genes in peripheral zone of human pancreatic cancer growing in the pancreas of nude mice compared to that of the tumor from the central zone.	7		1.63 e ⁻⁵
MEMBRANE_ORGANIZATION_AND_BIOGENESIS [133]	Genes annotated by the GO term GO:0016044. A process that is carried out at the cellular level which results in the formation, arrangement of constituent parts, or disassembly of membranes inside and surrounding the cell.	4		2.68 e ⁻⁵
WAMUNYOKOLI_OVARIAN_CANCER_GRADES_1_2_2_UP [138]	Genes up-regulated in mucinous ovarian carcinoma tumors of grades 1 and 2 compared to the normal ovarian surface epithelium tissue.	4		3.1 e ⁻⁵
BECKER_TAMOXIFEN_RESISTANCE_DN [51]	Genes down-regulated in a breast cancer cell line resistant to tamoxifen	3		4.03 e ⁻⁵

Gene/geneset overlap matrix

overlap matrix by gene and geneset	CREIGHTON_ENDOCRINE_THERAPY_RESISTANCE_3	RODRIGUES_THYROID_CARCINOMA_POORLY_DIF_DIFFERENTIATED_DN	chr6q13	COLDREN_GEFITINIB_RESISTANCE_DN	NAKAMURA_TUMOR_ZONE_PERIPHERAL_VS_CENTRAL_DN	MEMBRANE_ORGANIZATION_AND_BIOGENESIS	WAMUNYOKOLI_OVARIAN_CANCER_GRADES_1_2_UP	BECKER_TAMOXIFEN_RESISTANCE_DN	RODRIGUES_THYROID_CARCINOMA_ANAPLASTIC_DN	PARNER_BREAST_CANCER_APOCRINE_VS_LUMINAL	Entrez	Source	description
IKZF2												S	IKAROS family zinc finger 2 (Helios)
CITED2												S	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2
SASH1												S	SAM and SH3 domain containing 1
SH3BGRL2												S	SH3 domain binding glutamic acid-rich protein like 2
LMBRD1												S	LMBR1 domain containing 1
STK39												S	serine threonine kinase 39 (STE20/SPS1 homolog, yeast)
IL13RA1												S	interleukin 13 receptor, alpha 1
CYP3A5												S	cytochrome P450, family 3, subfamily A, polypeptide 5
INPP1												S	inositol polyphosphate-1-phosphatase
MOSPD1												S	motile sperm domain containing 1
CROT												S	carntine O-octanoyltransferase
CD24												S	CD24 molecule
TSPAN6												S	tetraspanin 6
C6ORF192												S	chromosome 6 open reading frame 192
DHRS1												S	dehydrogenase/reductase (SDR family) member 1
FBXO34												S	F-box protein 34



DAVID

DAVID: Functional Annotation Result Summary

http://david.abcc.ncifcrf.gov/summary.jsp

Functional Annotation Result Sum... Database for Annotation, Visualiz...

Functional Annotation Tool
DAVID Bioinformatics Resources 6.7, NIAID/NIH

Home Start Analysis Shortcut to DAVID Tools Technical Center Downloads & APIs Term of Service Why DAVID? About Us

Upload List Background

Upload Gene List

Demolist 1 Demolist 2 Upload Help

Step 1: Enter Gene List

A: Paste a list

AIM1
BEX4
BTBD11
C12ORF29
C2ORF24
C2ORF54
C6ORF192
CCNYL1
CD24
CDC2L6

Clear

Or
B: Choose From a File

Choose File no file selected

Multi-List File ?

Step 2: Select Identifier

Not Sure

Annotation Summary Results

Current Gene List: new_converted_list
Current Background: Homo sapiens

42 DAVID IDs
Check Defaults

Clear All

Help and Tool Manual

☒ Disease (1 selected)
☒ Functional_Categories (3 selected)
☒ Gene_Ontology (3 selected)
☒ General_Annotations (0 selected)
☒ Literature (0 selected)
☒ Main_Accessions (0 selected)
☒ Pathways (2 selected)
☒ Protein_Domains (3 selected)
☒ Protein_Interactions (0 selected)
☒ Tissue_Expression (0 selected)

☐ CGAP_EST_QUARTILE 71.4% 30 Chart
☐ CGAP_SAGE_QUARTILE 66.7% 28 Chart
☐ GNF_U133A_QUARTILE 88.1% 37 Chart
☐ PIR_TISSUE_SPECIFICITY 38.1% 16 Chart
☐ UNIGENE_EST_QUARTILE 85.7% 36 Chart
☐ UP_TISSUE 95.2% 40 Chart

Red annotation categories denote DAVID defined defaults

DAVID: Database for Annotation, Visualization, and Integrated Discovery (Laboratory of...and Infectious Diseases (NIAID); Science Applications International Corporation (SAIC);

http://david.abcc.ncifcrf.gov/chartReport.jsp?d=16544-s=5&d=16544-o=1&d=16544-p=1&annot=86

Functional Annotation Result Sum... Database for Annotation, Visualiz... Database for Annotation, Visualiz...

DAVID Bioinformatics Resources 6.7
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Functional Annotation Chart

Current Gene List: new_converted_list
Current Background: Homo sapiens
42 DAVID IDs

Options

Rerun Using Options Create Sublist

4 chart records

Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	UP_TISSUE	Liver	RT		10	23.8	2.5E-2	3.5E-1
<input type="checkbox"/>	UP_TISSUE	Colon	RT		8	19.0	1.0E-2	4.1E-1
<input type="checkbox"/>	UP_TISSUE	Pancreas	RT		6	14.3	5.3E-2	5.0E-1
<input type="checkbox"/>	UP_TISSUE	Aorta endothelial cell	RT		3	7.1	1.3E-2	2.8E-1

22 gene(s) from your list are not in the output.

Please cite [Nature Protocols 2009; 4\(1\):44](#) & [Genome Biology 2003; 4\(5\):P3](#) within any publication that makes use of any methods inspired by DAVID.



[Term of Service](#) | [Contact Us](#) | [Site Map](#)

IPA

IPA

File Edit View Window Help [Provide Feedback / Live Support](#) Vijayaraj Nagarajan [Close IPA](#)

[Genes and Chemicals](#) [Functions and Diseases](#) [Pathways and Tox Lists](#)

[NEW](#) [SEARCH](#) [Advanced Search](#)

Dataset Upload - testgeneset.txt

1. Select File Format: [More Info](#)
2. Contains Column Header: ☐ Yes ☒ No
3. Select Identifier Type: Specify the identifier type found in the dataset.
4. Array platform used for experiments: Select relevant array platform as a reference set for dat
5. Use the dropdown menus to specify the columns that contain identifiers and observations. For observations, select the appropriate expression va

Raw Data (44) \ Dataset Summary (40) \

	ID
1	AIM1
2	BEX4
3	BTBD11
4	C12ORF29
5	C2ORF24
6	C2ORF54
7	C6ORF192
8	CNLY1
9	CD24
10	CDC2L6
11	CITED2
12	CROT
13	CUL4B
14	CYP3A5
15	DHRS1
16	ERO1L
17	FBXO34
18	FTL
19	HBP1
20	HIBCH
21	IKZF2
22	IL13RA1
23	IL8RB
24	INPP1
25	LMBRD1
26	LTA4H
27	MOSPD1
28	MREG
29	MROCL3

IPA

File Edit View Window Help [Provide Feedback / Live Support](#) Vijayaraj Nagarajan [Close IPA](#)

[Genes and Chemicals](#) [Functions and Diseases](#) [Pathways and Tox Lists](#)

[NEW](#) [SEARCH](#) [Advanced Search](#)

testgeneset - 2011-10-31 12:14 PM

[Summary](#) \ [Networks](#) \ [Functions](#) \ [Canonical Pathways](#) \ [Lists](#) \ [My Pathways](#) \ [Molecules](#) \ [Network Explorer](#) \ [Overlapping Networks](#) \

[Download Summary \(PDF\)](#)

Analysis settings

Top Networks

ID	Associated Network Functions	Score
1	View Cellular Assembly and Organization, Drug Metabolism, Endocrine System Development and Function	30
2	View Cellular Development, Gene Expression, Infectious Disease	24
3	View Neurological Disease, Cellular Development, Hair and Skin Development and Function	21
4	View Gene Expression, Cardiovascular Disease, Cellular Function and Maintenance	2

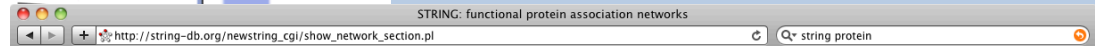
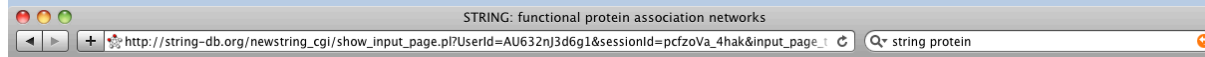
Top Bio Functions

Diseases and Disorders		
Name	p-value	# Molecules
Cancer	6.46E-05 - 4.34E-02	7
Infectious Disease	6.46E-05 - 4.75E-02	8
Neurological Disease	6.46E-05 - 4.75E-02	18
Respiratory Disease	2.04E-03 - 1.88E-02	4
Developmental Disorder	2.11E-03 - 4.55E-02	4
Molecular and Cellular Functions		
Name	p-value	# Molecules
Cellular Assembly and Organization	2.11E-03 - 4.88E-02	7
Cellular Function and Maintenance	2.11E-03 - 4.34E-02	6
Cellular Movement	2.11E-03 - 4.34E-02	3
Cell-To-Cell Signaling and Interaction	4.05E-03 - 4.34E-02	4
Carbohydrate Metabolism	4.22E-03 - 8.42E-03	1
Physiological System Development and Function		
Name	p-value	# Molecules
Nervous System Development and Function	2.11E-03 - 4.34E-02	3
Reproductive System Development and Function	4.05E-03 - 3.12E-02	3
Cardiovascular System Development and Function	4.22E-03 - 4.14E-02	2
Hematological System Development and Function	4.22E-03 - 4.54E-02	5
Immune Cell Trafficking	4.22E-03 - 1.47E-02	1

Top Canonical Pathways

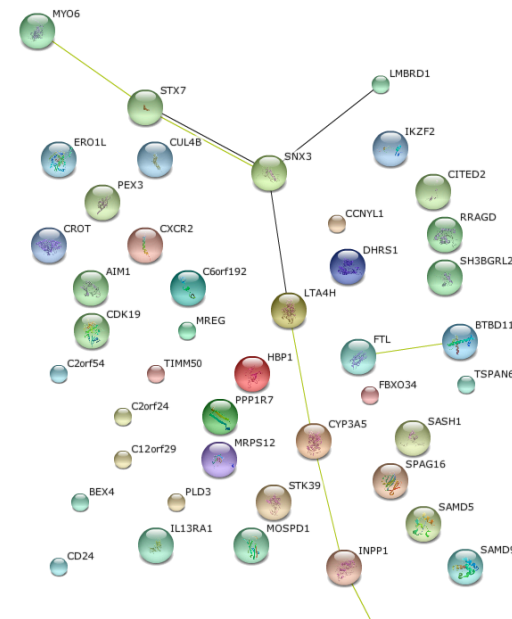
Name	p-value	Ratio
Xenobiotic Metabolism Signaling	2.29E-02	3/303 (0.01)

STRING



Home · Download · Help/Info

STRING 9.0



Display a menu

Display a menu for "http://string-db.org/newstring.cgi/show_input_page.pl"

Cytoscape

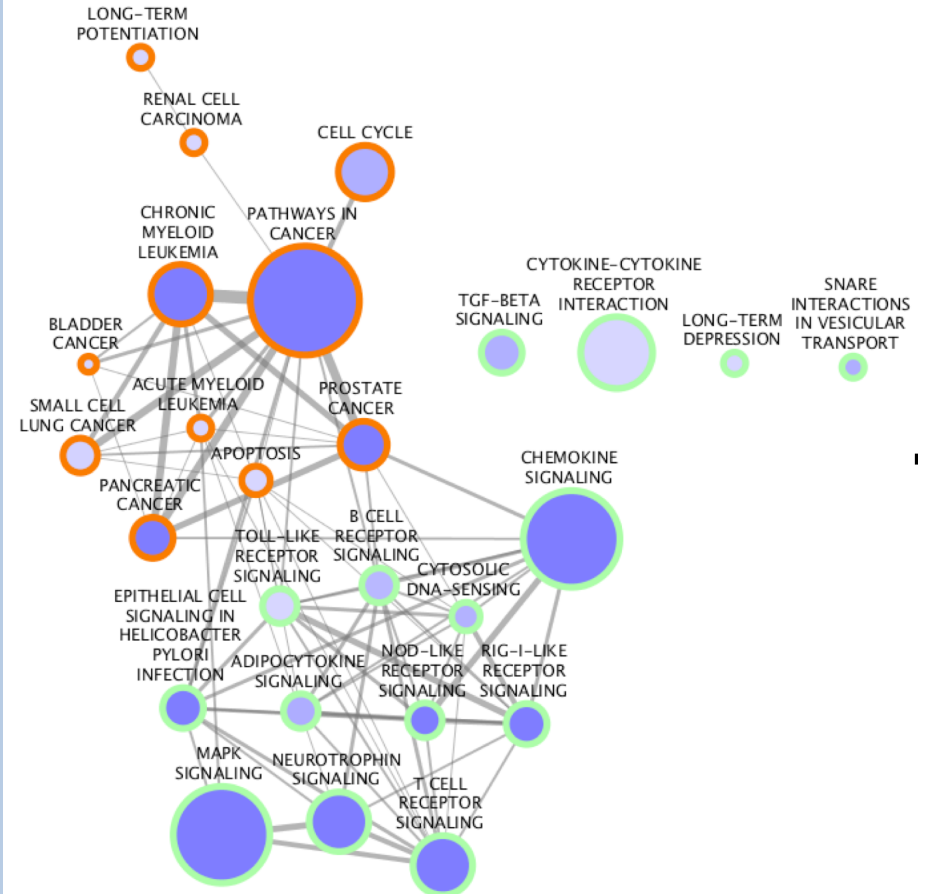
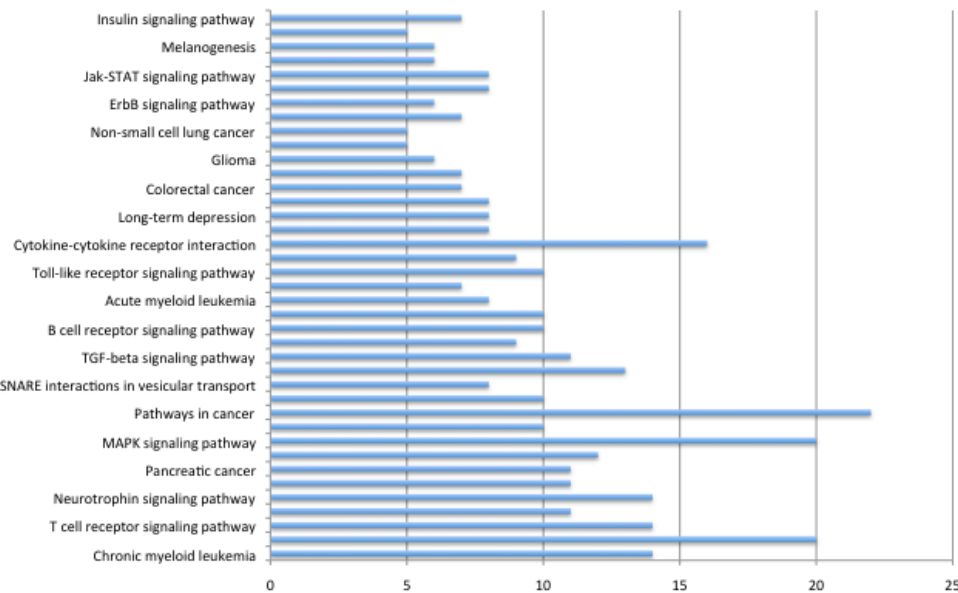
File Edit View Options Window Help

David Chart Kegg Pathways.txt

New Open Save Print Import Copy Paste Format Undo Redo AutoSum Sort A-Z Sort Z-A Gallery Toolbox Zoom Help

Verdana 10

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
1	KEGG_PATHW	Chronic myeloid leukemia	14	6.42201835	7.16E-09	EP2F1, BRAF, RELA, TGFBR1, ...	115	75	5085	8.25391304	7.30E-07	7.30E-07	7.97E-06
2	KEGG_PATHW	Chemokine signaling pathway	20	9.17431193	2.13E-08	CXCL1, GNAI3, GNAI2, LYN, ...	115	187	5085	4.72913276	2.17E-06	1.09E-06	2.37E-05
3	KEGG_PATHW	T cell receptor signaling pathway	14	6.42201835	6.26E-07	NFKB1E, RELA, NFKBIB, RAF1, ...	115	108	5085	5.73188406	6.39E-05	2.13E-05	6.98E-04
4	KEGG_PATHW	Epithelial cell signaling in Helicobacter pylori	11	5.04587156	2.10E-06	CXCL1, IL8, LYN, RELA, MAPK14, ...	115	68	5085	7.1528133	2.15E-04	5.37E-05	0.0023451
5	KEGG_PATHW	Neurotrophin signaling pathway	14	6.42201835	3.09E-06	BRAF, NFKB1E, RELA, NFKBIB, ...	115	124	5085	4.99228812	3.15E-04	6.30E-05	0.00344087
6	KEGG_PATHW	RIG-I-like receptor signaling pathway	11	5.04587156	3.16E-06	IKBKKE, IL8, MAP3K3, RELA, ...	115	71	5085	6.85058175	3.22E-04	5.37E-05	0.00351982
7	KEGG_PATHW	Pancreatic cancer	11	5.04587156	3.60E-06	EP2F1, BRAF, RELA, TGFBR1, ...	115	72	5085	6.75543478	3.67E-04	5.25E-05	0.00401135
8	KEGG_PATHW	Prostate cancer	12	5.04587156	3.72E-06	EP2F1, EP300, BRAF, RELA, ...	115	89	5085	5.96189546	3.79E-04	4.74E-05	0.00414368
9	KEGG_PATHW	MAPK signaling pathway	20	9.17431193	5.64E-06	BRAF, RELA, TGFBR1, ...	115	267	5085	3.11216414	5.75E-04	6.39E-05	0.00628337
10	KEGG_PATHW	NOD-like receptor signaling pathway	10	4.58715596	7.81E-06	CXCL1, IL8, RELA, MAPK14, ...	115	62	5085	7.13183731	7.97E-04	7.97E-05	0.00870448
11	KEGG_PATHW	Pathways in cancer	22	10.0917431	9.06E-06	EP2F1, TRAF1, BRAF, IL8, TGFBR1, ...	115	328	5085	2.96580064	9.24E-04	8.40E-05	0.01009719
12	KEGG_PATHW	Adipocytokine signaling pathway	10	4.58715596	1.50E-05	TNFRSF1A, PRKQC, TNFRSF1, ...	115	67	5085	6.59961064	0.00152823	1.27E-04	0.01670314
13	KEGG_PATHW	SNARE interactions in vesicular transport	8	3.66972477	1.68E-05	SNAP29, STX6, STX8, STX7, ...	115	38	5085	9.30892449	0.00171315	1.32E-04	0.01872583
14	KEGG_PATHW	Cell cycle	13	5.96330275	1.89E-05	EP2F1, RBL2, CREBBP, SMAD4, ...	115	125	5085	4.5986087	0.00152483	1.38E-04	0.02104151
15	KEGG_PATHW	TGF-beta signaling pathway	11	5.04587156	2.03E-05	SNAP29, EP300, RBL2, PPP2C, ...	115	87	5085	5.59070465	0.0020845	1.38E-04	0.02256983
16	KEGG_PATHW	Cytosolic DNA-sensing pathway	9	4.12844037	2.52E-05	IKBKKE, RELA, ARAF, IKKKG, ...	115	55	5085	7.23557312	0.00257058	1.61E-04	0.02810876
17	KEGG_PATHW	B cell receptor signaling pathway	10	4.58715596	3.79E-05	LYN, NFKB1E, RELA, JUN, NF, ...	115	75	5085	5.89565217	0.00385919	2.27E-04	0.04222377
18	KEGG_PATHW	Small cell lung cancer	10	4.58715596	9.39E-05	TRAF1, MAX, RELA, ...	115	84	5085	5.26397516	0.00952906	5.32E-04	0.10452342
19	KEGG_PATHW	Acute myeloid leukemia	8	3.66972477	2.77E-04	BRAF, RELA, ARAF, IKKKG, ...	115	58	5085	6.09859503	0.02782799	0.00148429	0.3077966
20	KEGG_PATHW	Bladder cancer	7	3.21100917	3.02E-04	EP2F1, IL8, BRAF, ARAF, RAF1, ...	115	42	5085	7.36956522	0.03035829	0.00154025	0.33615273
21	KEGG_PATHW	Toll-like receptor signaling pathway	10	4.58715596	3.87E-04	IKBKKE, IL8, RELA, MAPK14, ...	115	101	5085	4.37795944	0.03674268	0.00187981	0.43064356
22	KEGG_PATHW	Apoptosis	9	4.12844037	6.67E-04	TNFRSF1A, TNFRSF10B, REL, ...	115	87	5085	4.57421289	0.06581609	0.00308985	0.74085202
23	KEGG_PATHW	Cytokine-cytokine receptor interaction	16	7.33944954	6.69E-04	CXCL1, IL8, CXCL5, CXCL3, ...	115	262	5085	2.70029871	0.05950574	0.00296196	0.74241477
24	KEGG_PATHW	Long-term potentiation	8	3.66972477	7.43E-04	EP300, BRAF, GRIA3, ARA, ...	115	68	5085	5.20204604	0.0728939	0.00314866	0.82326609
25	KEGG_PATHW	Long-term depression	8	3.66972477	8.10E-04	GNAI3, GNAI2, BRAF, LYN, ...	115	69	5085	5.12665406	0.07937032	0.00330243	0.89917702
26	KEGG_PATHW	Renal cell carcinoma	8	3.66972477	8.84E-04	EP300, BRAF, PAK3, JUN, AR, ...	115	70	5085	5.05341615	0.08626282	0.00346369	0.98048529
27	KEGG_PATHW	Colorectal cancer	7	3.21100917	0.01094184	BRAF, JUN, TGFBR1, ARAF, ...	115	84	5085	3.68478261	0.67444313	0.04071171	11.5357593
28	KEGG_PATHW	Progesterone-mediated oocyte maturation	7	3.21100917	0.01220883	GNAI3, GNAI2, BRAF, MAPK1, ...	115	86	5085	3.59908999	0.7143433	0.04376228	12.7900939
29	KEGG_PATHW	Glioma	6	2.75229358	0.0129694	EP2F1, BRAF, ARAF, RAF1, ...	115	63	5085	4.21118012	0.73592752	0.04487675	13.5352682
30	KEGG_PATHW	Nucleotide excision repair	5	2.29357798	0.01612219	DNB1, LIG1, CDK7, CUL4B, ...	115	44	5085	5.02470356	0.81364488	0.0541673	16.7516039
31	KEGG_PATHW	Non-small cell lung cancer	6	2.29357798	0.0209175	EP2F1, BRAF, ARAF, RAF1, ...	115	54	5085	4.0942029	0.96410059	0.10176512	30.4683341
32	KEGG_PATHW	Vascular smooth muscle contraction	7	3.21100917	0.03902802	PRKQC, BRAF, ARAF, RAF1, ...	115	112	5085	2.76358696	0.98276172	0.11917337	35.8221766
33	KEGG_PATHW	ErbB signaling pathway	6	2.75229358	0.04409972	BRAF, PAK3, JUN, ARAF, RAF, ...	115	87	5085	3.04947326	0.9906853	0.13212002	39.9950582
34	KEGG_PATHW	Wnt signaling pathway	8	3.66972477	0.05196203	EP300, JUN, PPP2CA, CREBBP, ...	115	151	5085	2.34264325	0.9567263	0.14792624	44.8148929
35	KEGG_PATHW	Jak-STAT signaling pathway	8	3.66972477	0.05828133	TYK2, GHI, EP300, IL4R, ...	115	155	5085	2.28218794	0.99781244	0.16054219	48.7772222
36	KEGG_PATHW	GHRH signaling pathway	6	2.75229358	0.0607655	MAP3K3, MAP3K1, MAPK14, ...	115	98	5085	2.70718722	0.99524702	0.18107521	54.4096526
37	KEGG_PATHW	Melanogenesis	6	2.75229358	0.07047166	GNAI3, EP300, GNAI2, CREB, ...	115	99	5085	2.67984919	0.99942085	0.18246227	55.698052
38	KEGG_PATHW	Melanoma	5	2.29357798	0.07412135	EP2F1, BRAF, ARAF, RAF1, ...	115	71	5085	3.1139008	0.99961228	0.18674998	57.5978226
39	KEGG_PATHW	Insulin signaling pathway	7	3.21100917	0.08158513	BRAF, ARAF, TSC2, RAF1, ...	115	135	5085	2.29275362	0.99983019	0.19955311	61.2540165



BED-Tools

- Command line
- Open source
- A variety of input formats (BED, BEDPE, SAM/BAM, GFF, VCF)
- Things we could do...
 - Intersect, Union
 - Merge
 - Coverage
 - Subtract
 - Convert
 - Closest
 - Shuffle
 - Group, Sort