

GSA-SNP2 Users' manual

GSA-SNP2 basically supports command-line interface (CLI) for Windows, Linux and MAC OS. For Windows, it also supports graphic user interface (GUI) for those who are not familiar with CLI. Detailed instructions for GUI and CLI version of GSA-SNP2 is as follows:

■ GUI (Graphic User Interface) version

1. Installing GSA-SNP2

- A. Download the GUI version of GSA-SNP2 program ([gsasnp2-windows-gui.zip](#)) and related data files from the website (<https://sourceforge.net/projects/gsasnp2/files/>).

- B. Unzip the file and execute the program by double-clicking 'gsasnp2-gui.exe'.

* If it is your first time to use GSA-SNP2, we strongly recommend you to download [test-package.zip](#) first. It consists of GSA-SNP2 program (gsasnp2.exe) and 'data' folder that contains necessary files so that users can easily run the program with the example data.

* The 'data' folder must contain (1) [pathway file](#) (e.g., c2.cp.v5.2.symbols.gmt), (2) [SNP-gene mapping file](#) (e.g., db19_20k), (3) [Adjacent gene correlation file](#) (e.g., EUR_Adjacent_correlation), (4) [gene list files](#) (e.g., hg19GeneList), (5) [SNP position file](#) (e.g., rsloc_hg19) and (6) [network file](#) (e.g., STRING_NETWORK.txt). This folder must locate where the main program (gsasnp2.exe) exists.

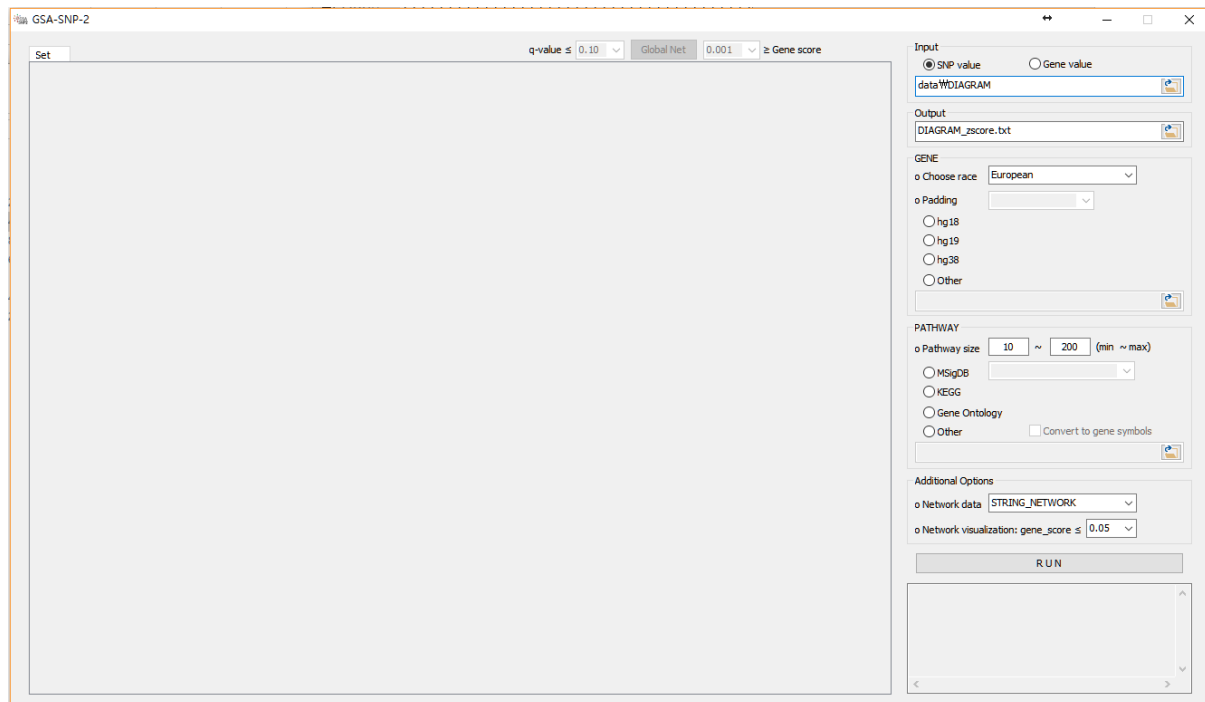
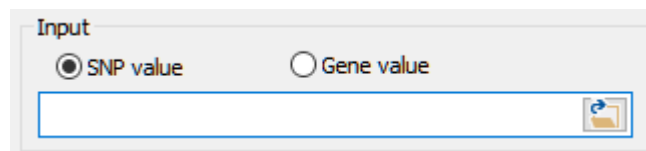


Figure 1. Initial interface of GSA-SNP2.

2. Input data and options

A. Input

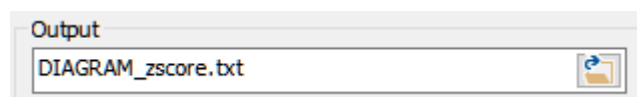


- i. SNP value: GSA-SNP2 accepts simple tab-delimited two column text file composed of SNP IDs (e.g. rs numbers) and corresponding p-values obtained from a GWA study (e.g. PLINK) (Fig.2). As an example, we provide DIAGRAM summary p-values for European population from our website ('DIAGRAM' file in the test-sample-data.zip).
- ii. Gene value: Alternatively, same format with gene ID (Ensembl, Entrez or gene symbol) and its p-value can be uploaded. We provide an example gene value data calculated from VEGAS2 program ('DIAGRAMgene' file in the test-sample-data.zip). In addition to the gene-based p-value for GWA studies, users also can upload gene p-values assessed from microarray or RNA-sequencing differential expression analysis.

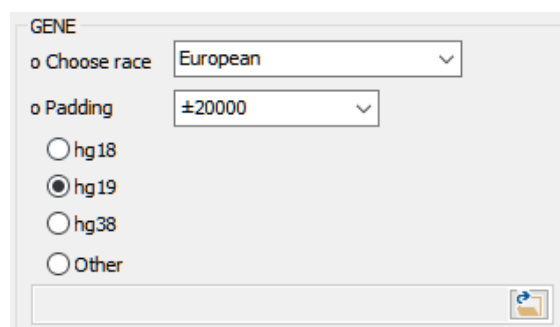
rs12565286	0.16	ZMIZ1	3e-05
rs3094315	0.82	THADA	3e-05
rs2905035	0.36	SSR1	3.5e-05
rs2980319	0.39	MIR486-2	3.7e-05
rs4040617	0.5	CISD2	3.9e-05
rs2977612	0.35	TIMP4	4.1e-05
rs2905062	0.33	MIR486	4.2e-05
rs2980300	0.61	FEZF1-AS1	4.7e-05
rs11240777	0.78	PLEKHA1	4.9e-05
rs17160906	0.34	MIR3941	5.3e-05
rs3121561	0.3	SLC9B1	5.8e-05
rs3813193	0.66	UBE2D3	6.1e-05
rs4075116	0.97	HSPA1B	6.6e-05
rs3934834	0.43	SLC9B2	7.2e-05
rs3766193	0.29	LOC646736	8.2e-05
		ZNF239	8.4e-05

Figure 2. Input file format for GSA-SNP2 (SNP p-value: left, Gene p-value: right).

- B. Output: The path for output file is designated here.



C. GENE



- i. Choose race: Choose the race of input data. This option removes highly correlated adjacent genes to reduce false positives. Inter-gene correlations of adjacent genes were pre-

calculated for five human races (African, American, East Asian, European and South Asian) using the 1000 Genomes genotype data: The SNP with the smallest p-value in each gene were mapped to the closest SNP among the three gene points (start, center, end) and corresponding genotype data in 1000 Genomes were used to calculate correlation. The European inter-gene correlation file ('EUR_Adjacent_correlation') is in the test-sample-data.zip, and the others can be downloaded from 'data/adjacent_gene_all_races-20170225T013103Z-001.zip'.

- ii. Padding: The extension of gene ranges where each SNP is assigned to the corresponding gene (Default: 20kb).
- iii. Reference genome: Choose the version of reference genome (hg18,19 or 38).
 - ✓ Note: You MUST download proper SNP-Gene mapping file from the GSA-SNP2 web site and put them in the 'data' folder for this mapping process. We provide multiple mapping files from *data/gene-map* folder in Data session of the web site.
 - ✓ There are 9 SNP-Gene map files (db18_0k, db18_10k, db18_20k, db18_exon, db19_0k, db19_10k, db19_20k, db19_exon and db38_0k) and the mapping file name indicates the genome version and the padding size. Its format is shown in Figure 3. Each row represents rs number of a SNP and matched genes (tab-delimited between SNP and gene ID, and comma-delimited between gene IDs). Another mapping file with same format can be generated and uploaded by user.

```
3  EEF1DP3,
4  EEF1DP3,
5  KRIT1,
6  AKAP9,CYP51A1,
7  CYP51A1,LRRD1,
8  CDK6,
9  CDK6,
10 CDK6,
15 THSD7A,
16 THSD7A,
17 THSD7A,
18 THSD7A,
19 THSD7A,
```

Figure 3. An example of SNP-Gene mapping file (db19_20k)

D. PATHWAY

PATHWAY

o Pathway size ~ (min ~ max)

☒ MSigDB

☐ KEGG

☐ Gene Ontology

☐ Other ☐ Convert to gene symbols

- i. Pathway size: Minimum and maximum number of gene-set size
- ii. Gene-sets: Choose a gene-set database. Some popularly used gene-set DBs (KEGG, Gene

Ontology, mSigDB C1, C2, C5 gene sets) are provided from 'data/popular_pathway_data-20170227T151601Z-001.zip' in the web-site. Among them, mSigDB sets are included in the test-sample-data.zip. These data must be in the 'data' folder.

The users can also upload their own gene-set DB using the format shown in Figure 4: Each line is tab-delimited and composed of gene set name, some description and the list of gene-set members. The user can download gene set data from mSigDB page (<http://software.broadinstitute.org/gsea/msigdb/collections.jsp>) and directly use them for GSA-SNP2.

* If the **gene IDs are represented in Ensembl or Entrez ID** in your gene-set data, please check the box 'Convert to gene symbols'.

KEGG GLYCOLYSIS GLUCONEOGENESIS	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_GLICOLYSIS_GLUCONEOGENESIS	AC92	GCK	PFK2	PFK1	PFKB	PFHA1	PFHA2	PFM2	PF11	ACSB1	FBP1	ADH1B	PF2	ADH1C
KEGG CITRATE CYCLE TCA CYCLE	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_CITRATE_CYCLE_TCA_CYCLE	IDH3B	DLST	PFK2	CS	PFHB	PFK1	PFHA1	LOC642502	PFHA2	LOC283398	FR	SDHD	OGDH	
KEGG PENTOSE PHOSPHATE PATHWAY	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_PENTOSE_PHOSPHATE_PATHWAY	RPE	PF1A	PFM2	PGL8	PFPS2	FBP2	PFPM	PFKL	TALDO1	TKT	FBP1	TKTL2	PGD	RBSK
KEGG PENTOSE AND GLUCURONATE INTERCONVERSIONS	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_PENTOSE_AND_GLUCURONATE_INTERCONVERSIONS	UGT1A10	UGT1A8	RPE	UGT1A7	UGT1A6	UGT2B28	UGT1A5	CRYL1	UGDH	UGT				
KEGG FRUCTOSE AND MANNOSE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_FRUCTOSE_AND_MANNOSE_METABOLISM	MTF	PFM2	PFM1	FBP2	PFPM	OMG5	PFPS4	PFKL	WTM65	PF11	PFH11	PF2		
KEGG GALACTOSE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_GALACTOSE_METABOLISM	GCK	GALK1	GLB1	GAL2	B4GALT1	PFM2	LALBA	PFPM	PFKL	MGAM	HK2	HK1	HK3	GALT
KEGG ASCORBATE AND ALDARATE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_ASCORBATE_AND_ALDARATE_METABOLISM	UGT1A10	UGT1A8	UGT1A7	UGT1A6	ALDH1B1	UGT2B28	ALDH2	UGT1A5	MIOX	UGDH	UGT2A1			
KEGG FATTY ACID METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_FATTY_ACID_METABOLISM	CPPIA	CPPI1	ACAD8	ALDH1B1	ACAD8	ACADL	ALDH2	ACADM	CYF4A11	ACR22	ACADVL	ACAT1	ACAA2	FAH
KEGG STEROID BIOSYNTHESIS	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_STEROID_BIOSYNTHESIS	SOAT1	LSS	SQLE	EBP	CYP51A1	DMCR7	CYP27B1	DMCR24	HSID17B7	MSMO1	FDFT1	SC5DL	LIPA	CEL
KEGG PRIMARY BILE ACID BIOSYNTHESIS	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_PRIMARY_BILE_ACID_BIOSYNTHESIS	CYP46A1	SLC27A5	BAAT	CYP7B1	AKR1C4	HSID17B4	SCF2	AKH1D1	ACOX2	HSID3B7	CYP27A1	AMACR		
KEGG STEROID HORMONE BIOSYNTHESIS	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_STEROID_HORMONE_BIOSYNTHESIS	SRD5A3	AMP1C4	CYP19A5	HSID3B2	UGT2B28	HSID3B1	COMT	SULT2B1	CYP24A	CYP17A1	CYP11A1	AKR1C3		
KEGG OXIDATIVE PHOSPHORYLATION	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_OXIDATIVE_PHOSPHORYLATION	ATP4V1G1	UQC10	NDUFA5	NDUFA4	COX6CP3	PF2A2	ATP5J2	NDUF57	CYTB	ATP6V0A1	ATP6V1G2	ATP		
KEGG PURINE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_PURINE_METABOLISM	PF1A2	PF1A1	PF1A3	PF1A4	PF1A5	PF1A6	PF1A7	PF1A8	PF1A9	PF1A10	PF1A11	PF1A12	PF1A13	PF1A14
KEGG PYRIMIDINE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_PYRIMIDINE_METABOLISM	MTF2	PF1A2	PF1A1	PF1A3	PF1A4	PF1A5	PF1A6	PF1A7	PF1A8	PF1A9	PF1A10	PF1A11	PF1A12	PF1A13
KEGG ALANINE ASPARTATE AND GLUTAMATE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_ALANINE_ASPARTATE_AND_GLUTAMATE_METABOLISM	GLUD2	GLUT2	GLUT1	GLUT3	GLUT4	GLUT5	GLUT6	GLUT7	GLUT8	GLUT9	GLUT10	GLUT11	GLUT12	GLUT13
KEGG GLYCINE SERINE AND THREONINE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM	ALAS1	ALAS2	GLYCTK	MAOB	AGXT2	MAOA	ACOC2	GATM	SDS	ACOC				
KEGG CYSTEINE AND METHIONINE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_CYSTEINE_AND_METHIONINE_METABOLISM	AMO1	SRM	ADH1	ABCY	DMNT1	SDS	TRMT1	DMNT3A	DMNT3B	LDHC	IL1411	SNS	LDH	
KEGG VALINE LEUCINE AND ISOLEUCINE BIOSYNTHESIS	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_BIOSYNTHESIS	ACAD8	ACAD9	ACAD10	ACAD11	ACAD12	ACAD13	ACAD14	ACAD15	ACAD16	ACAD17	ACAD18	ACAD19	ACAD20	ACAD21
KEGG LYSINE DEGRADATION	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_LYSINE_DEGRADATION	SUV39H2	ASORHPT	DLST	ALDH1B1	ALDH2	ALDH3	ALDH4	ALDH5	ALDH6	ALDH7	ALDH8	ALDH9	ALDH10	ALDH11
KEGG ARGININE AND PROLINE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_ARGININE_AND_PROLINE_METABOLISM	SRM	ACG	GLUD2	GLU2	ARG1	ARG2	GLUD1	GLU3	GLU4	GLU5	GLU6	GLU7	GLU8	GLU9
KEGG HISTIDINE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_HISTIDINE_METABOLISM	CNDP1	MAOB	MAOA	ALDH1B1	ALDH2	MTTTL6	ALDH3A1	ALDH3A2	ALDH3A3	ALDH3A4	ALDH3A5	ALDH3A6	ALDH3A7	ALDH3A8
KEGG TYROSINE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_TYROSINE_METABOLISM	MAOB	MAOA	HPD	AOX1	AOX2	TYR	HSD	MTTTL6	ALDH3A1	COMT	AOX3	WBSCR22	IL1411	GSTT1
KEGG PHENYLALANINE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_PHENYLALANINE_METABOLISM	ALDH1A3	ALDH1B1	ALDH2	MAOB	MAOA	HPD	GOT2	ACO2	ALDH3A1	AOX3	FAH	GOT1	PKRX6	TAT
KEGG TRYPTOPHAN METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_TRYPTOPHAN_METABOLISM	MAOA	IDO2	AOX1	ALDH1B1	ANAT	ALDH2	WARS	IDO1	CAT	ACAT2	ACAT1	IL1411	RAH6	OGH
KEGG BETA ALANINE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_BETA_ALANINE_METABOLISM	CNDP1	SRM	ECM1	ALDH9A1	ALDH3A2	GAD1	GAD2	ACO2	ALDH1B1	HIBCH	ABAT	ERHADH	ALDH2	ACA
KEGG TAURINE AND HYPOXANTHINE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_TAURINE_AND_HYPOXANTHINE_METABOLISM	GSTT1	CSAD	GOT5	GOT1	GAD1	BAAT	GAD2	GOT6	CDOL	ADO				
KEGG SILENIN METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_SILENIN_METABOLISM	MTTTL6	ARCT	WBSCR22	SEPR1	GOT1	SEPR2	MTTTL2	HUMK1	GOT6	LCMT2	FAPB51	FAPB52		
KEGG GLUTATHIONE METABOLISM	http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_GLUTATHIONE_METABOLISM	SRM	GSTT1	GSTT2	GSTT3	GSTT4	RRM2B	SMS	PGD	GSTO1	GSTA5	MGST2	LAP3	MGST1	MGST3

Figure 4. Gene-set list file format

E. Network visualization of gene-sets (Additional Options)

Additional Options

☐ Network data

STRING_NETWORK

☐ Network visualization: gene_score ≤

0.05

i. Network data: Type of network. Currently, STRING (<http://www.internationalgenome.org/>) and HIPPIE(<http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/download.php>) network data are provided. Users can also use their own network file. The file must contain three tab-delimited columns including two genes and their interaction score. The interaction scores must be scaled to 0~1.

ii. Network visualization: Threshold of gene-score (best SNP p-value) to be displayed in the network. (default: 0.05)

3. Result

A. Result table

After clicking 'RUN' button (

RUN

), the results are shown in the main panel and also saved as a text file in the local (Fig. 5) in only seconds to minutes depending on the input data size. The results consist of gene-set name, size of gene-set, count of detected genes, z-score, adjusted z-score, p-value, q-value and list of member genes. Significance of a gene-set is determined by q-value which is based on the trend curve adjusted gene scores. Users can sort the gene-sets based on score of your interest by clicking the header of each column.

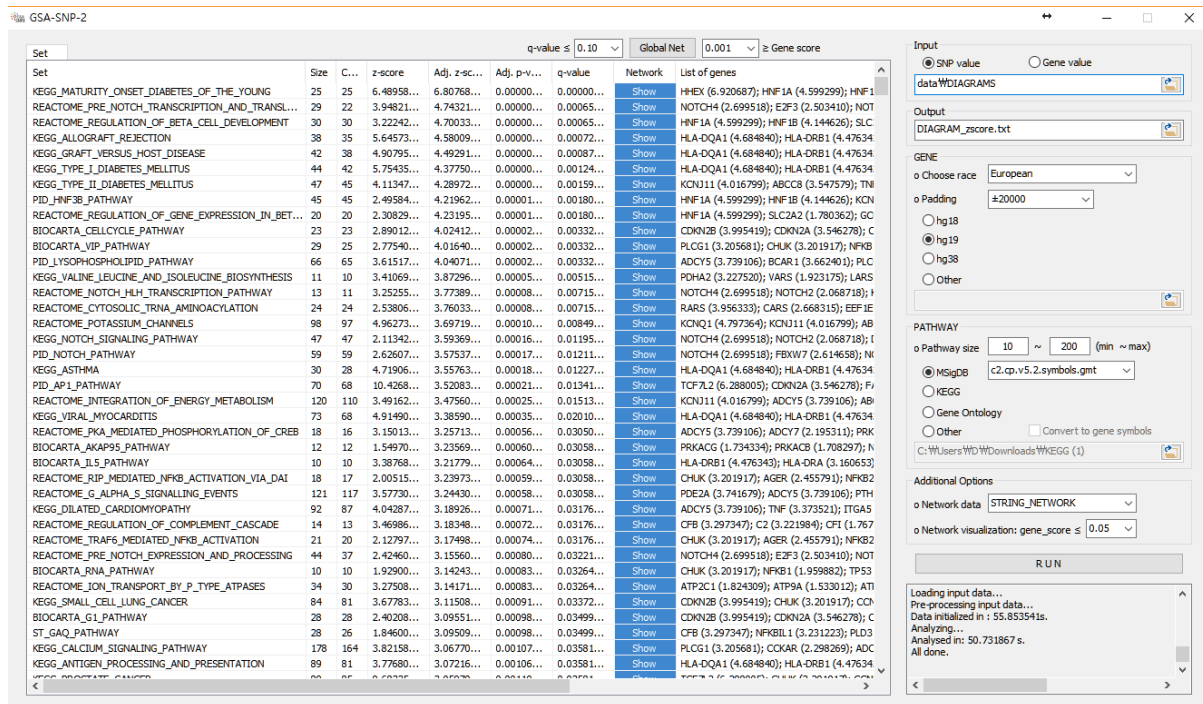
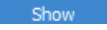


Figure 5. Result screen

B. Network Visualization of individual gene-set

If the users click on the 'show' button () of a gene-set of interest, protein interaction network among the member genes will be shown (Fig.6) and it is saved in the local directory. Large nodes represent more connections to other nodes and the node color represents the gene score (dark color represents high level of association to the phenotype). Network edges will be shown only for genes with p-values less than a user-defined threshold, but their interacting partner gene nodes will all be displayed.

Link to external page: If the users click on an interesting gene in the network, a box representing the gene name and the matched SNPs ($p\text{-value} \leq 0.05$) with their p-values will be shown (Fig. 8). By clicking on the gene name or SNP ID, the users can access the corresponding GeneCards (<http://www.genecards.org/>) or dbSNP (<https://www.ncbi.nlm.nih.gov/SNP>) pages for more information (Fig. 9).

*Note: Network graph is optimized for Microsoft Edge. For Internet Explorer (IE) users, if the IE restricts executing script or Active X on the network page, please allow blocked contents.

Network for KEGG_TYPE_II_DIABETES_MELLITUS

Legends

1. Gene score

Min  Max

2. Interaction score

— Highest(0.9)
— High(0.7)
— Medium(0.4)
— Lowest(0.15)

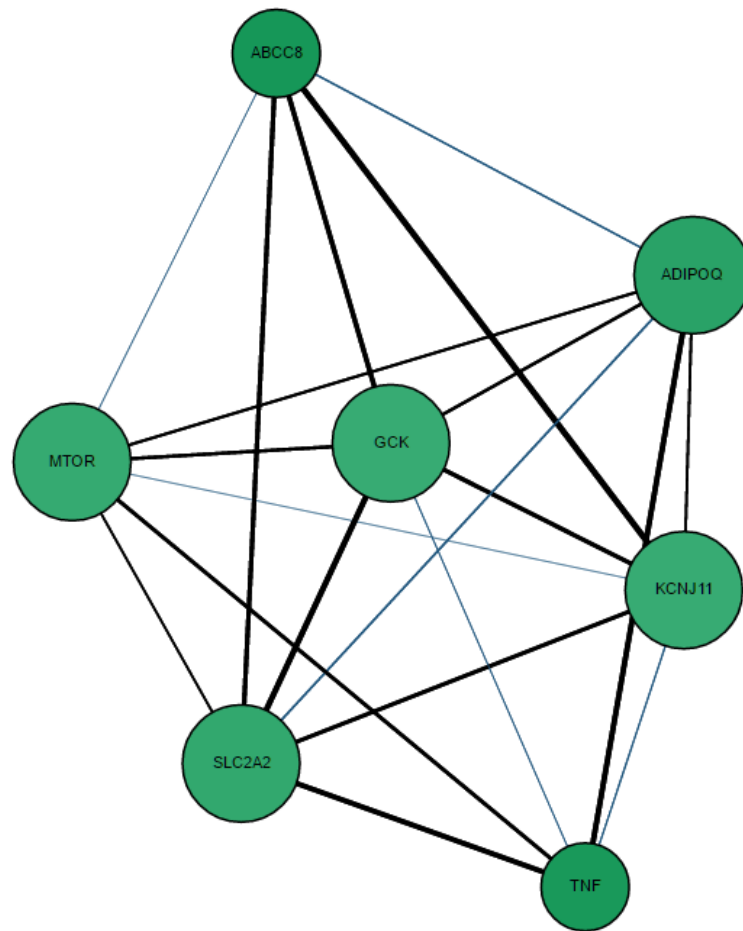


Figure 6. Network visualization of a single gene-set.

Gene-SNPs Reference	
Gene	GSK (0.036947)
SNPs ($p_value \leq 0.05$)	rs1303722 (0.001400) rs2041547 (0.002300) rs2268576 (0.002600) rs2268574 (0.002600) rs7793213 (0.002900) rs758989 (0.002900) rs2300584 (0.004800) rs2908292 (0.010000) rs2971685 (0.017000) rs10278336 (0.020000) rs2908282 (0.026000) rs882019 (0.032000) rs3824065 (0.035000) rs1799884 (0.042000)
Neighboring genes (interaction weight)	SLC2A2 (0.932000) ABCC8 (0.811000) KCNJ11 (0.755000) MTOR (0.374000) ADIPOQ (0.330000) TNF (0.191000)
KEGG_TYPE_II_DIABETES_MELLITUS	
ABCC8	

Figure 7. Box showing a gene and its matched SNPs

GeneCards | MalaCards | LifeMap Discovery | PathCards | TbxCards | Variants | GeneAnalytics | GeneRatCart | GeneLife

Free for academic non-profit institutions. Other users need a Commercial license.

WETZMANN INSTITUTE OF SCIENCE

GeneCards®
HUMAN GENE DATABASE

Keywords + Search Term

Home User Guide Analysis Tools News And Views About My Genes Log In

HNF4A Gene (Protein Coding) ★

Hepatocyte Nuclear Factor 4 Alpha

GCID: GC20P044355
GIFTS: 73

Jump to section

Aliases	Disorders	Domains	Drugs	Expression	Function	Genomics	Localization	Orthologs
Paralogs	Pathways	Products	Proteins	Publications	Sources	Summaries	Transcripts	Variants

Proteins & Enzymes
Antibodies Assays & Kits

Genes Peptides Proteins
CRISPR

Proteins Antibodies
Assays Genes shRNA
Primers CRISPR

Genes (ado)
Genes (gen)
shRNA (AA)

Aliases for HNF4A Gene

Aliases for HNF4A Gene

Hepatocyte Nuclear Factor 4 Alpha ^{1 2 3}	HNF-4-Alpha ⁴
Nuclear Receptor Subfamily 2 Group A Member 1 ⁴	HNF4alpha ³
Transcription Factor HNF-4 ^{1 4}	HNF4a7 ³
Transcription Factor 14 ^{1 4}	HNF4a8 ³
TCF14 ^{1 4}	HNF4a9 ³
TCF14 ^{1 4}	NR2A21 ³
NR2A1 ^{1 4}	FRTS3 ³
HNF4 ^{2 4}	MODY1 ³
Hepatocyte Nuclear Factor 4-Alpha ³	MODY ³
Hepatic Nuclear Factor 4 Alpha ³	TCF ²
HNF4alphaT011/T2 ³	

External IDs for HNF4A Gene

HGNc: 5024 Entrez Gene: 3172 Ensembl: ENSG00000101076 OMIM: 600281 UniProtKB: P41235

Previous HGNC Symbols for HNF4A Gene

TCF14, MODY, MODY1

Rapid Di-
Genetic
 GeneCards provides
 interpretation
FREE FOR A
START AT YOU

Figure8. External pages showing the information of a gene (GeneCards) and a SNP (dbSNP)

C. Global network visualization for top-ranked gene sets.

Adj. p-v...	q-value	Network	List of genes
0.00000...	0.00000...	Show	HHEX (6.920687); HNF1A (4.599299); HNF1
0.00000...	0.00065...	Show	NOTCH4 (2.699518); E2F3 (2.503410); NOT
0.00000...	0.00065...	Show	HNF1A (4.599299); HNF1B (4.144626); SLC
0.00000...	0.00072...	Show	HLA-DQA1 (4.684840); HLA-DRB1 (4.47634
0.00000...	0.00087...	Show	HLA-DQA1 (4.684840); HLA-DRB1 (4.47634
0.00000...	0.00124...	Show	HLA-DQA1 (4.684840); HLA-DRB1 (4.47634

After the computation is finished, user can view the global network of top-ranked gene-sets by

choosing cutoffs for gene-set q-value and gene-score (best SNP p-value), and then clicking on 'Global Net' button (Global Net). Figure 9 shows a core network of DIAGRAM data.

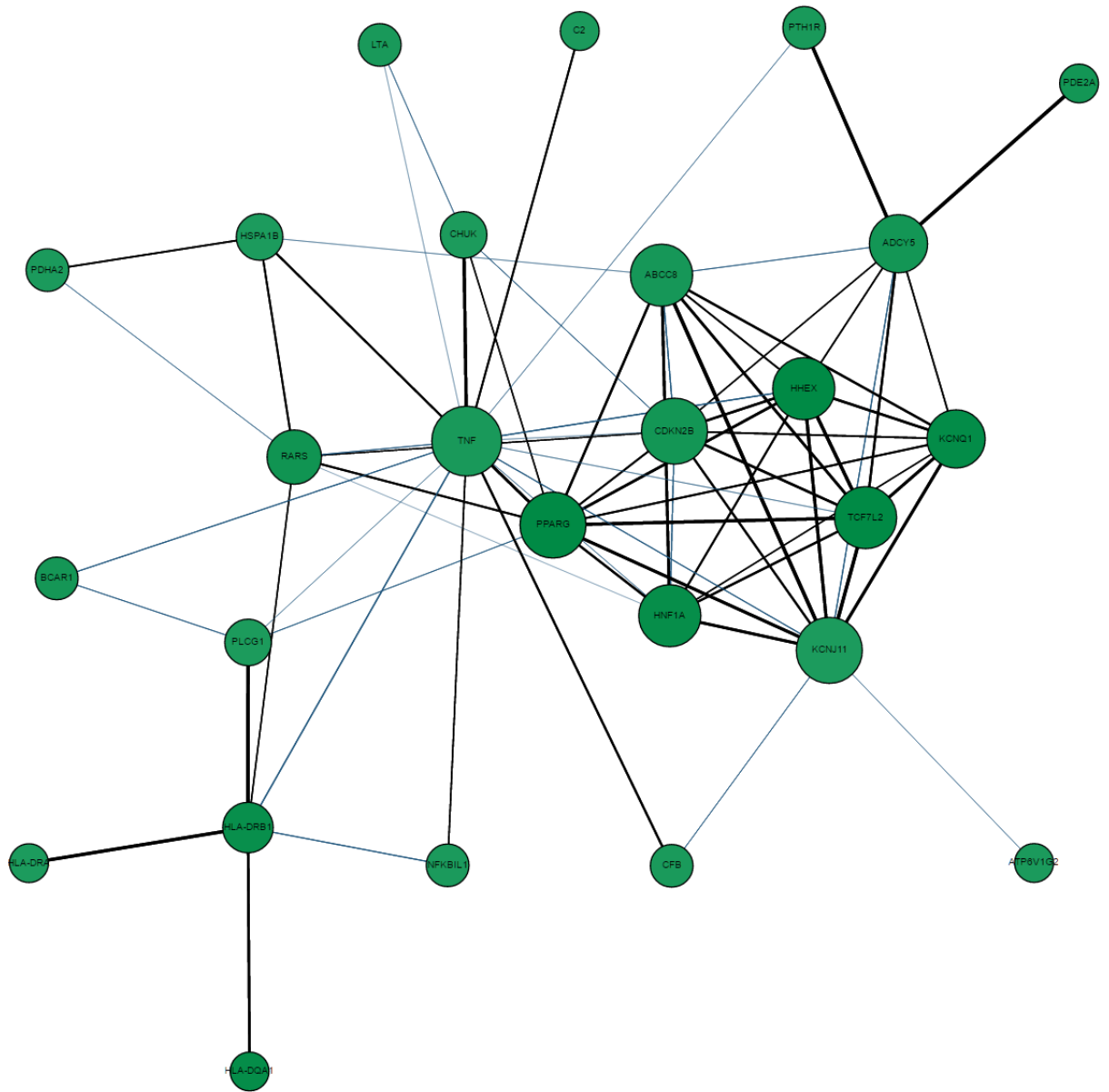


Figure 9. Global network visualization

■ CLI (Command-Line Interface) version

I. Installing GSA-SNP2

1. Download the GSA-SNP2 program and related data files from the website. (<https://sourceforge.net/projects/gsasnp2/files/>)

* GSA-SNP2 programs

- gsasnp2-window-cmd.zip for Windows. It contains main program (gsasnp2.exe), a batch file, and other required files.

- [gsasnp2_mac_cmd.zip](#) for MAC OS. It contains main program (gsasnp2), a library file (libboost_program_options.dylib) and a bash file (run_gsasnp2.sh).
- [gsasnp2-linux-cmd.zip](#) for Linux. It contains main program (gsasnp2), lib folder where required library files exist and a bash file (run_gsasnp2.sh).

* The [test-sample-data.zip](#) file in the web-site includes 'data' folder that contains example data for simple test. After downloading it, please locate this folder where the main program exists. The 'data' folder must contain (1) [pathway file](#) (e.g., c2.cp.v5.2.symbols.gmt), (2) [SNP-gene mapping file](#) (e.g. db19_20k), (3) [Adjacent gene correlation file](#) (e.g., EUR_Adjacent_correlation), (4) [gene list files](#) (e.g., hg19GeneList), (5) [SNP position file](#) (e.g., snplocmap) and (6) [network file](#) (e.g., STRING_NETWORK.txt).

All these required files can be easily downloaded from the web-site. For detailed explanation of each file type, please refer to the GUI part.

II. Usage

Basic usage:

```
gsasnp2 [-input/-i] <snp/gene_list_file> [--pathway/-p] <set/pathway_list_file> [--snpgene/-s]
<0/1> <additional parameters.....>
```

Example

✓ Linux User

```
cd /home/DIRECTORY_FOR_GSASNP2_PROGRAM # Move to the directory where gsasnp2
program is.

export LD_LIBRARY_PATH=./lib:$LD_LIBRARY_PATH # including required library files.

./gsasnp2 -p data/c2.cp.v5.2.symbols.gmt -i data/DIAGRAM -s 0 -o DIAGRAM_result.txt -g
data/db19_20k -a data/EUR_Adjacent_correlation --minset 10 --maxset 200 -b 0
```

✓ MAC OS User

```
cd /home/DIRECTORY_FOR_GSASNP2_PROGRAM # Move to the directory where gsasnp2
program is

./gsasnp2 -p data/c2.cp.v5.2.symbols.gmt -i data/DIAGRAM -s 0 -o DIAGRAM_result.txt -g
data/db19_20k -a data/EUR_Adjacent_correlation --minset 10 --maxset 200 -b 0
```

✓ Windows User

```
cd C:\Users\DIRECTORY_FOR_GSASNP2_PROGRAM # Move to the directory where gsasnp2
program is.

gsasnp2 -p data\c2.cp.v5.2.symbols.gmt -i data\DIAGRAM -s 0 -o DIAGRAM_result.txt -g
data\db19_20k -a data\EUR_Adjacent_correlation --minset 10 --maxset 200 -b 0
```

Option description (Some file formats are described in Windows GUI version part.)

- p [--pathway]: Path to input set/pathway list file. Necessary!
- i [--input]: Path to input SNP/GENE file. Necessary!

-s [--snpgene]: This parameter indicates that input file is SNP (0) or GENE (1) p-value.

-o [--output]: Path to output result. Default is 'adjusted_zscore_result.txt'.

-g [--genemap]: Gene map file. (e.g., data/db19_20k)

-a [--adj]: Inter-gene genotype correlation by race file. (e.g. data/EUR_Adjacent_correlation)

--minset: Minimum set size (number of genes). Default minset = 10

--maxset: Minimum set size (number of genes). Default maxset = 200

-n [--netfile]: Path to network file (e.g., data/STRING_NETWORK.txt)

-v [--gval]: Gene score cutoff for selecting reduced network (0~1). Default value = 0.05

-q [--qval]: q-value cutoff for selecting core network (0~0.25). Default value = 0.15. Note: high q-value may cause memory overloaded.

-b [--symbol] : 0 or 1. 1 to convert gene ID (Ensembl or Entrez) in gene set/pathway to gene symbol, 0 for otherwise. Default = 0.

III. Result

CLI version of GSA-SNP2 provides three result files as follows:

A. Gene-set analysis result

It gives gene-set analysis result. Each row consists of gene-set name, gene-set size, count (the number of gene-set elements in the gene list file), original z-score, adjusted z-score (z-score evaluated with adjusted gene scores), p-value, q-value and the list of pathway genes with the gene scores. Unless specified, the result file name is 'zscore_result.txt'.

Set	Size	Count	z-score	Adj. z-score	p-value	q-value	List of genes
KEGG_MATURITY_ONSET_DIABETES_OF_THE_YOUNG	25	25	5.87817	6.02094	8.67E-10	1.08E-06	HHEX(3.56277); HNF1A(3.27547); HNF1B(3.27547);
REACTOME_PRE_NOTCH_TRANSCRIPTION_AND_TRANS	29	22	3.57161	4.8595	5.88E-07	0.00036687	NOTCH4(2.50353); E2F3(2.45492); NOTCH1(3.27547);
REACTOME_REGULATION_OF_BETA_CELL_DEVELOPME	30	30	2.91042	4.49532	3.47E-06	0.0014437	HNF1A(3.27547); HNF1B(3.20346); SLC2A2(2.82564);
KEGG_TYPE_II_DIABETES_MELLITUS	47	45	3.71599	4.15556	1.62E-05	0.00505793	KCNJ11(3.3128); ABCB1(2.82564); TNF(2.73922);
KEGG_VIP_PATHWAY	29	25	2.50588	4.12927	1.82E-05	0.00505793	CHUK(3.17987); PLCG1(3.15853); NFkB1(1.64951);
REACTOME_REGULATION_OF_GENE_EXPRESSION_IN_BI	20	20	2.08325	4.10865	1.99E-05	0.00505793	HNF1A(3.27547); SLC2A2(1.74435); GSK3B(1.64951);
KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_BIOSYNTHES	11	10	3.08786	3.97989	3.45E-05	0.00547857	PDHA2(3.1915); VARS(1.85385); LARS2(1.64951);
KEGG_ALLOGRAFT_REJECTION	38	35	5.10942	3.93853	4.10E-05	0.00547857	HLA-DQA1(3.33243); TNF(2.73922); HLA-DQB1(3.33243);
KEGG_GRAFT_VERSUS_HOST_DISEASE	42	38	4.43886	3.88439	5.13E-05	0.00547857	HLA-DQA1(3.33243); TNF(2.73922); HLA-DQB1(3.33243);
KEGG_CELL_CYCLE_PATHWAY	23	23	2.61062	3.99463	3.24E-05	0.00547857	CDKN2B(3.2927); CDKN2A(3.29108); CCN1(3.29108);
PID_LYSOPHOSPHOLIPID_PATHWAY	66	65	3.25975	4.00695	3.08E-05	0.00547857	BCAR1(3.28671); PLCG1(3.15853); ADCY5(1.64951);
PID_HNF3B_PATHWAY	45	45	2.24727	3.92158	4.40E-05	0.00547857	KCNJ11(3.3128); HNF1A(3.27547); HNF1B(3.27547);
REACTOME_NOTCH_HLH_TRANSCRIPTION_PATHWAY	13	11	2.94384	3.82873	6.44E-05	0.00617781	NOTCH4(2.50353); NOTCH2(2.04808); KA1(2.04808);
REACTOME_CYTOSOLIC_TRNA_AMINOACYLATION	24	24	2.29067	3.78478	7.69E-05	0.00685166	RARS(3.17518); CARS(2.64627); EEF1E1(2.04808);
KEGG_NOTCH_SIGNALING_PATHWAY	47	47	1.89963	3.71227	0.0001027	0.00853813	NOTCH4(2.50353); NOTCH2(2.04808); DU(2.04808);
KEGG_TYPE_II_DIABETES_MELLITUS	44	42	5.20646	3.69224	0.0001115	0.00859921	HLA-DQA1(3.33243); LTA(2.74207); TNF(2.73922);
PID_NOTCH_PATHWAY	59	59	2.36277	3.6941	0.00011034	0.00859921	FBXW7(2.58397); NOTCH4(2.50353); NOTCH1(3.27547);
REACTOME_POTASSIUM_CHANNELS	98	97	4.47825	3.42028	0.00031279	0.0216692	KCNJ11(3.3128); ABCB1(2.82564); KCNQ1(3.3128);
REACTOME_INTEGRATION_OF_ENERGY_METABOLISM	120	110	3.14077	3.39046	0.00034888	0.0228972	KCNJ11(3.3128); ADCY5(3.01851); ABCB1(2.82564);
BIOCARTA_AKAP95_PATHWAY	12	12	1.39733	3.34373	0.0004133	0.0257695	PRKACG(1.68229); PRKACB(1.64951); NCA(1.64951);

Figure 10. An example of gene-set analysis result

B. List of significant genes in the core pathways (*.gnl)

It represents the list of significant genes (satisfying the gene p-value cutoff) contained in the core pathways (satisfying pathway q-value cutoff). For MAC OS, the file extension is *.gnl.txt

C. Gene network for core pathways (*.net)

In this file, each row represents two genes, their interaction score, and the core pathways where both genes are involved. If two genes, each of which belongs to at least one core pathway, are not included in same core pathway, the fourth column will be empty. For MAC OS, the file extension is *.net.txt