

# GWAB tutorial

## 1. DATA SUBMISSION

Click 'GWAB RUN' menu.

### GWAB RUN

Submit files and run GWAB.

(The test example may take up to 10 - 30 minutes depending on the status of the server.)

FILL TEST EXAMPLE SHOW EXPECTED RESULT

**E-mail (Optional)**  
If you want to receive job status information and query results via email, please submit your email address

**GWAS data set**  
[sample file](#)  
 찾아보기...

**Disease name**

**Reference disease gene set**  
FILE UPLOADING MANUAL INPUT  
**Gene set**  
Pre-compiled gene sets from [DISEASES databases \(Pletscher-Frankild et al\)](#)

**Genome Build** hg18 ▼ **SNP-gene dist. range** 10000

**Test p-value threshold range: log(P)** -6 - -2 (1.00e-6 - 1.00e-2) **Interval of log(P) threshold** 0.3

SUBMIT RESET

- ① (Optional) If you want to receive job status information and query results via email, please submit your email address here.

② Upload GWAS data set file with the following format.

- a. Input file name should have \*.txt extension.
- b. Input data format: [chromosome number] [tab] [basepair] [tab] [p-value] [new line]
  - 1-22, X, and Y are acceptable for chromosome number.
  - We support various formats of p-value (such as 0.00001, .00001 and 1e-05).

CHR	BP	PVALUE
Chromosome number (1-22, X, Y)	Chromosomal position of each SNP (bp)	p-value of each SNP
1	1051029	.0089792
22	3312	0.0001251
Y	99102	4.213000e-05

c. [download example](#)

③ Enter a user-defined disease name. (ex. Coronary Artery Disease)

④ To find the optimal p-value threshold for boosting, you need to upload a set of reference disease gene set (only Entrez geneID accepted). We provide two ways to enter a reference disease gene set (File uploading and Manual input).

A. File uploading

- a. Input file name should have \*.txt extension.
- b. Input data format: [entrez geneID] [new line]

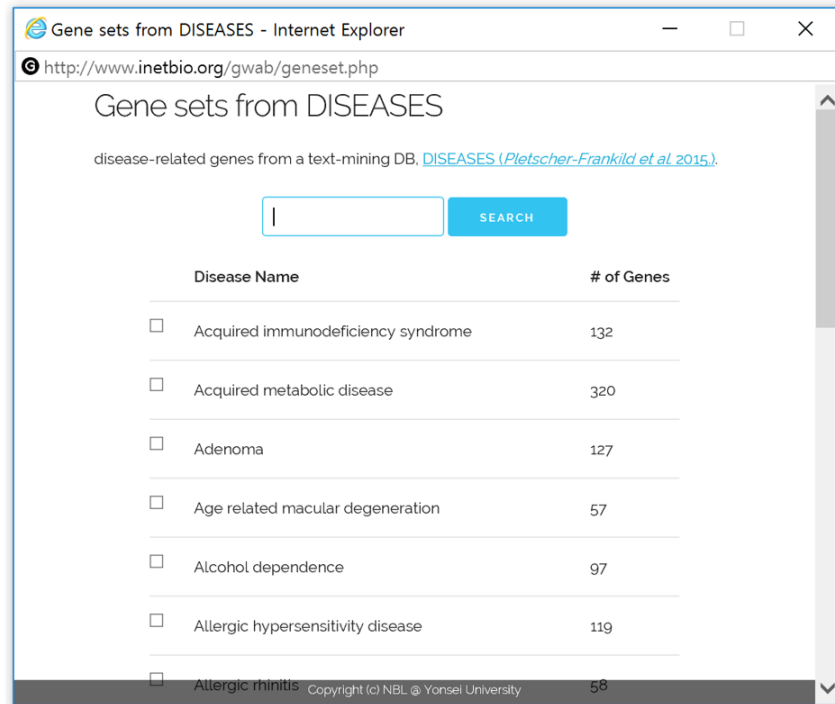
c. [download example](#)

B. Manual input

- a. If you have a list of disease genes, you can put them in a text box.
- b. If you do not have a disease gene set, you can select a disease gene list from the

pre-compiled DISEASES database as shown below.

- c. Click 'DISEASES database' link. You can search for disease name and then select multiple diseases, through the pop-up window as shown below.



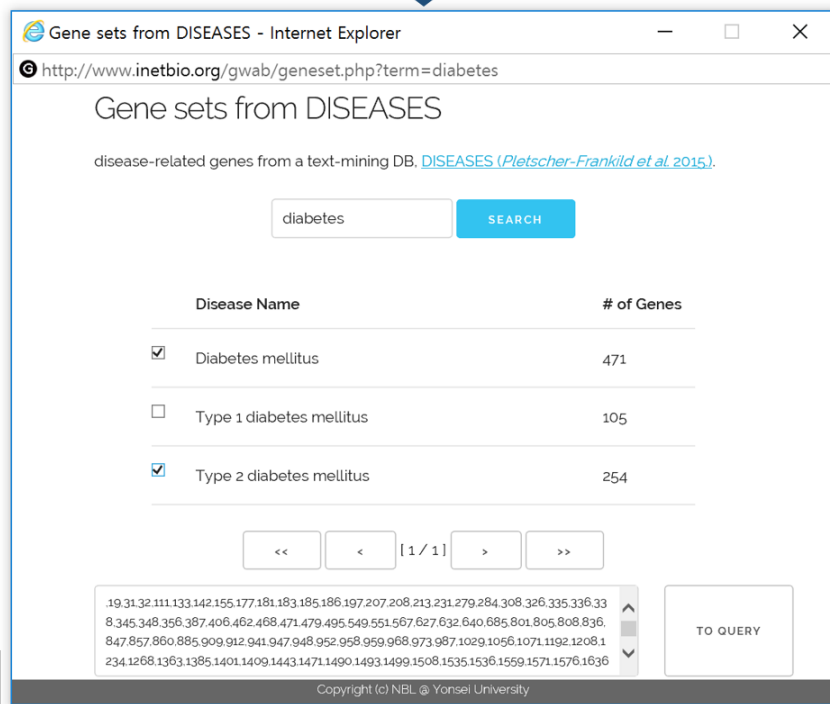
Gene sets from DISEASES - Internet Explorer  
http://www.inetbio.org/gwab/geneset.php

### Gene sets from DISEASES

disease-related genes from a text-mining DB, [DISEASES \(Pletscher-Frankild et al 2015\)](#).

Disease Name	# of Genes
<input type="checkbox"/> Acquired immunodeficiency syndrome	132
<input type="checkbox"/> Acquired metabolic disease	320
<input type="checkbox"/> Adenoma	127
<input type="checkbox"/> Age related macular degeneration	57
<input type="checkbox"/> Alcohol dependence	97
<input type="checkbox"/> Allergic hypersensitivity disease	119
<input type="checkbox"/> Allergic rhinitis	58

Copyright (c) NBL @ Yonsei University



Gene sets from DISEASES - Internet Explorer  
http://www.inetbio.org/gwab/geneset.php?term=diabetes

### Gene sets from DISEASES

disease-related genes from a text-mining DB, [DISEASES \(Pletscher-Frankild et al 2015\)](#).

Disease Name	# of Genes
<input checked="" type="checkbox"/> Diabetes mellitus	471
<input type="checkbox"/> Type 1 diabetes mellitus	105
<input checked="" type="checkbox"/> Type 2 diabetes mellitus	254

<< < [ 1 / 1 ] > >>

19, 31, 32, 111, 133, 142, 155, 177, 181, 183, 185, 186, 197, 207, 208, 213, 231, 279, 284, 308, 326, 335, 336, 338, 345, 348, 356, 387, 406, 462, 468, 471, 479, 495, 549, 551, 567, 627, 632, 640, 685, 801, 805, 808, 836, 847, 857, 860, 885, 909, 912, 941, 947, 948, 952, 959, 959, 968, 973, 987, 1029, 1056, 1071, 1192, 1208, 1234, 1268, 1363, 1385, 1401, 1409, 1443, 1471, 1490, 1493, 1499, 1508, 1535, 1536, 1559, 1571, 1576, 1636

Copyright (c) NBL @ Yonsei University

⑤ Configuration options.

Item	Description	Example
<b>Genome Build</b>	Human genome build	default : hg18
<b>SNP-gene dist. range</b>	Chromosomal distance range from SNP in searching for genes to assign p-value	default : 10000 bp
<b>Test p-value threshold range: log(P)</b>	Threshold range of SNP p-value for GWAB	default : -6 ~ -2
<b>Interval of log(P) threshold</b>	Step of log(P) threshold in the above test	default : 0.3

⑥ Click 'SUBMIT' button, then you can see the job status page

⑦ If you want to run GWAB using a test example, click 'FILL TEST EXAMPLE' button.

(This job may take 10 ~ 30 minutes depending on the status of the server.)

⑧ You can see the expected results for the test example by clicking 'SHOW EXPECTED RESULT' button.

## 2. STATUS OF GWAB JOB

To access GWAB result of your query, user should keep a record of the URL for the STATUS OF GWAB JOB page. Also, this status page will be refreshed in every 10 second. Once the analysis is completed, the status page will move to the GWAB result page automatically. In case you submitted your email as running parameter, you will receive the URL of Job status page via email.

### JOB STATUS

Bookmark this URL to access your query results in the future.  
In case you submitted your email address, you will receive the URL via email  
This status page will be refresh in every 10 sec.

GWAB job running status

SNP Mapping	P-value Extracting	Boosting & Validation
Done	Done	Done

②                      ③                      ④

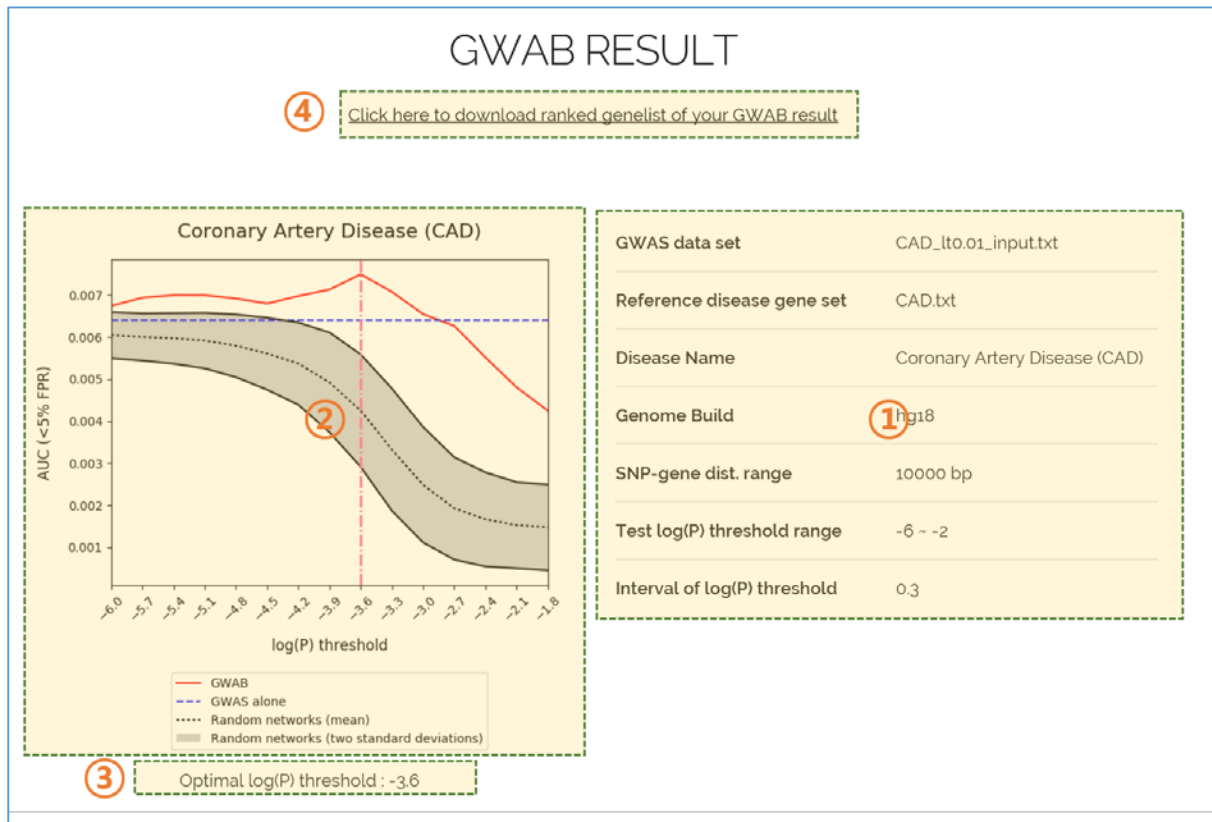
GWAB running parameters

GWAS data set	CAD_lto.01_input.txt
Reference disease gene set	CAD.txt
Disease Name	① Coronary Artery Disease (CAD)
Genome Build	hg18
SNP-gene dist. range	10000 bp
Test log(P) threshold range	-6 - -2
Interval of log(P) threshold	0.3

- ① This panel shows information for running parameters.
- ② To assign SNP to genes, GWAB performs distance-based gene mapping. Here, GWAB supports two versions of genome build, hg18 and hg19.
- ③ GWAB extract p-values from GWAS data set.
- ④ GWAB conducts boosting of GWAS p-values using co-functional links between genes. In addition, GWAB calculate the accuracy of retrieved reference disease gene set by 100 randomized networks by edge shuffling.

Each job status will be marked as either 'Not Yet', 'Running' or 'Done'.

### 3. INTERPRETATION OF RESULTS



① This panel shows information for running parameters.

② This plot shows the AUC score for FPR 5% (y-axis), indicating how much GWAB improved retrieval rate of reference disease gene set with the given log(P) threshold (x-axis). The plot also shows the baseline performance by GWAS alone and the range of 100 random performances to evaluate effectiveness and significance of the boosting, respectively.

a. GWAS alone

The baseline represents AUC score of the disease gene candidates prioritized by original GWAS p-value

b. Random networks

The mean and range of two standard deviations of AUC scores for candidate genes with boosting using 100 randomized HumanNet (version 2; unpublished) by edge

shuffling. If the AUC score by GWAB is higher than that by shuffled networks, as well as GWAS baseline, novel candidate genes are also likely to be related with corresponding disease.

- ③ To present the candidate genes prioritized by optimal boosting conditions, GWAB use  $\log(P)$  threshold for the highest AUC score.
- ④ For the optimal  $\log(P)$  threshold, users can download prioritized genes with their Entrez geneID, gene symbol, and GWAB score.

#### 4. PRE-CALCULATED GWAB

Disease Name	①	②	③
	Reference Genes and GWAS Data	Result	Publication
Alzheimer's disease (AD)	<a href="#">Download</a>	<a href="#">View Result</a>	PMID : <a href="#">24162737</a>
Coronary Artery Disease (CAD)	<a href="#">Download</a>	<a href="#">View Result</a>	PMID : <a href="#">21378990</a>
Crohn's disease (CD)	<a href="#">Download</a>	<a href="#">View Result</a>	PMID : <a href="#">21102463</a>
Rheumatoid arthritis (RA)	<a href="#">Download</a>	<a href="#">View Result</a>	PMID : <a href="#">20453842</a>
Schizophrenia (SZ)	<a href="#">Download</a>	<a href="#">View Result</a>	PMID : <a href="#">25056061</a>
Type 2 diabetes (T2D)	<a href="#">Download</a>	<a href="#">View Result</a>	PMID : <a href="#">22885922</a>
Ulcerative colitis (UC)	<a href="#">Download</a>	<a href="#">View Result</a>	PMID : <a href="#">21297633</a>

- ① Users can download GWAS data sets and reference disease gene set for the seven pre-calculated diseases.
- ② Users can see pre-calculated GWAB results and prioritized genes for all diseases.
- ③ The publication information for each GWAS