

GWAS3D User Manual

v1.3

Table of contents

Introduction	3
System Requirements	4
Welcome	4
Features	4
Create New Run	5
GWAS3D Result Overview	8
GWAS3D Result Details	11
Download Result.....	18
Retrieve Jobs.....	19

Introduction

Interpreting noncoding common phenotypically associated variants is an indispensable step to understand molecular mechanism of complex traits (1). The Encyclopedia of DNA Elements (ENCODE) project identified a comprehensive map of functional elements in active chromatin states by advanced techniques such as ChIP-seq, DNase-seq, bisulfate sequencing, chromosome conformation capture, etc (2). Recent work showed that the associated SNPs detected by genome wide association study (GWAS) are significantly enriched in those regulatory regions including many transcriptional factor binding sites (TFBSs), histone modification marked regions, Dnase I hypersensitive sites (DHSs) and expression quantitative trait loci (eQTLs) (3). Also, those regulatory elements can engage in long range looping interactions to exert elaborate gene regulation. Genetic variation in those noncoding regions may affect the function of cis acting elements or distal interactions and finally contribute to complex phenotype (4). Therefore, combinatory analysis of GWAS data and profile of functional elements to capture regulatory variants in a particular disease/trait is needed.

GWAS3D (<http://jiwanglab.org/gwas3d>) systematically compute the probability of genetics variants affecting regulatory pathways and underlying disease/trait associations by integrating chromatin state, functional genomics, sequence motif, and conservation information when given GWAS data or variant list. We first collected and curated genome-wide chromosome conformation (5C, Hi-C, ChIA-PET) data, enhancer/insulator/promoter marks (H3K4me1, H3K27ac, P300, CTCF, DHS) and chromHMM predicted functional elements in 16 different cell types. Using those regulatory regions, we mapped genetic variants to them and evaluated the binding affinity changes of regulatory factors by scanning 73 ENCODE-motifs. Finally, we combined original GWAS signal, risk haplotype, binding affinity significance and conservation information to prioritize the genetic variants. GWAS3D also provided comprehensive annotations and visualizations to help users interpreting the results.

Main Functions:

- Identify the most probable functional variant associated with interested trait in given risk loci;
- Prioritize the leading variants when given a full list of GWAS result;
- Evaluate the deleteriousness of genetic variants affecting the gene regulation without any prior-effect;
- Annotate genetic variant from regulatory perspective.

1. Ward, L.D. and Kellis, M. (2012) Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol*, **30**, 1095-1106.
2. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Fritze, S., Harrow, J., Kaul, R. *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.
3. Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S. and Snyder, M. (2012) Linking disease associations with regulatory information in the human genome. *Genome Res*, **22**, 1748-1759.
4. Sanyal, A., Lajoie, B.R., Jain, G. and Dekker, J. (2012) The long-range interaction landscape of gene promoters. *Nature*, **489**, 109-113.

System Requirements

GWAS3D is best accessed using the Google Chrome web browser. It has been tested to work with Mozilla Firefox and Safari and Internet Explorer 9. Not all functions are available with Internet Explorer 8, due to a lack of HTML5 support by IE. It don't support the old version of IE under 8.

Since GWAS3D uses many JavaScript features and libraries and will display batch of dataset in one web page, it has some requirements about the hardware configuration.

Recommend configuration: two cores CPU and 2G memory.

Welcome

This document aims to introduce the usage and function of GWAS3D. In order to access the public site, please visit <http://jjwanglab.org/gwas3d>.

Please check the site for the most up-to-date versions of the user manual.

Features

1. Detecting regulatory (Non-coding) human genetic variants specially;
2. Wide-spread cell types supporting;
3. Cell type specific chromatin state data including signals of distal interaction, enhancer, promoter, insulator and predicted functional elements etc;
4. 73 ENCODE, TRANSFAC, Jasper motifs of TF families including hundreds of known and novel motifs;
5. Enhanced regulatory variants prioritization;
6. Comprehensive variants annotation;
7. Multiply association/SNV list formats support;
8. Useful user defined parameters, nice visualization and usability.

Create New Run

To perform a new run for your GWAS association result or SNV list, please follow:

1. Enter the name of the investigated study.
2. Specify your E-mail Address to retrieve your job, a notification will be sent to your assigned mailbox.

Please specify the name of your **study**:

Please specify your **E-mail Address** to retrieve your job:

3. Select an input format for GWAS result, GWAS3D currently support four different formats including Plink-like format, VCF-like format, single SNP Id and genomic coordinates.

Before association file is inputted, please notice that our system is based on the latest homo species genome assembly version (hg19/GRCh37) and dbSNP 137. The input variants coordinates should be consistent with hg19 (if have). While, the SNP identification is no special restriction about version, we will convert SNPs to dbSNP 137 automatically.

It is encouraged to use association data with P-value as input, variant list without association P-value is acceptable.

Input File Format Help

GWAS3D supports following input format, please use VCF format if you want to input allele information:

Plink-like format: Chromosome,dbSNP Id,Position[,P-value]
VCF-like format: Chromosome,Position,SNP Id,RefAllele,Alt Allele[,P-value]
Variants Coordinates: Chromosome,Position[,P-value]
Single SNP Id: dbSNP Id[,P-value]

We accept user provided regions for variants inclusion:
Specific Region: Chromosome,Start,End

Variants should be separated by newline, fields should be separated by tab or comma.
A example is available by clicking the example link.

4. Choose input text or upload a input file.

Input Format: ?

PLINK-like Map

Upload Association/SNPs File (<50M)

選擇檔案 未選擇檔案

Paste Association/SNPs List

[example \(Plink-like\)](#)

[example \(VCF-like\)](#)

[example \(dbSNP\)](#)

[example \(Only genomic coordinates\)](#)

5. Input data without association P-value or define P-value cutoff.

Select if your data is pure SNV list.

The P-value cutoff refers to the maximal P-value cutoff, variants with P-value larger than the cutoff will be discarded.

Investigated population (HapMap I+II+III) for computing the synthetic association.

Without Association P-value: ?

Pvalue Cutoff: ?

Without GWAS Association Statistics

1.0 x 10⁻⁵

6. Select SNP data set and population.

HapMap I+II+III

SNP Data Set: ?

HapMap I+II+III

CEU: European American(CEPH)
ASW: African ancestry in Southwest USA
CHB: Han Chinese in Beijing, China
CHD: Chinese in Metropolitan Denver, Colorado

Population: ?

1000 Genomes Pilot 1

SNP Data Set: ?

1000 Genomes Pilot 1

EUR: European
AFR: African
ASN: East Asian
AMR: Ad Mixed American

Population: ?

7. Define a cutoff for haplotype checking

The *RSquare* value of linkage disequilibrium will be restricted between 0.6 to 1 with leading SNP.

Variant Linkage Option

Linkage Disequilibrium Cutoff (R^2): ?

0.8

8. Select a investigated cell type.

Specify the target cell type related to the GWAS traits or your objective. Please select "without cell type restriction" when you don't make sure the phenotype information or no matched cell type.

Cell Type Option

Specific Cell Type: ?

K562
GM12878
H1-hESC
HeLa-S3
IMR90
MCF-7
GM06990
CD4+

9. Select TF motifs used as binding affinity scanning.

Support mutiple selection.

Regulatory Elements Option

ENCODE TF Familay Motifs: ?

use "Ctrl" for mutiple selection in IE
clear all select all

AP-1
AP-2
ATF3
BAF155
BATF
BCL
BDP1
BHLHE40

10. Only scan known motifs.

Only consider the known motifs of selected factors.

Known Motifs: ?

Using Known Motifs for Scanning

11. Define TF binding site P-value.

The P-value cutoff of putative TF binding site scanning. Strict setting may reduce the false positive.

Binding Site P-value ?

x 10-

12. Define promoter range.

The range of promoter region should be restricted between [100, 10000] upstream and [50, 5000] downstream of TSS.

Promoter Range (Up/Down): ?

/

13. Input a user customized region if wanted.

If defined, system only consider the variants in those regions.

Specific Region: ? [example](#)

14. Define maximal variants and interactions could be displayed in the plotting.

Plotting Option

Plotting Variant Size: ?

Plotting Interaction Size: ?

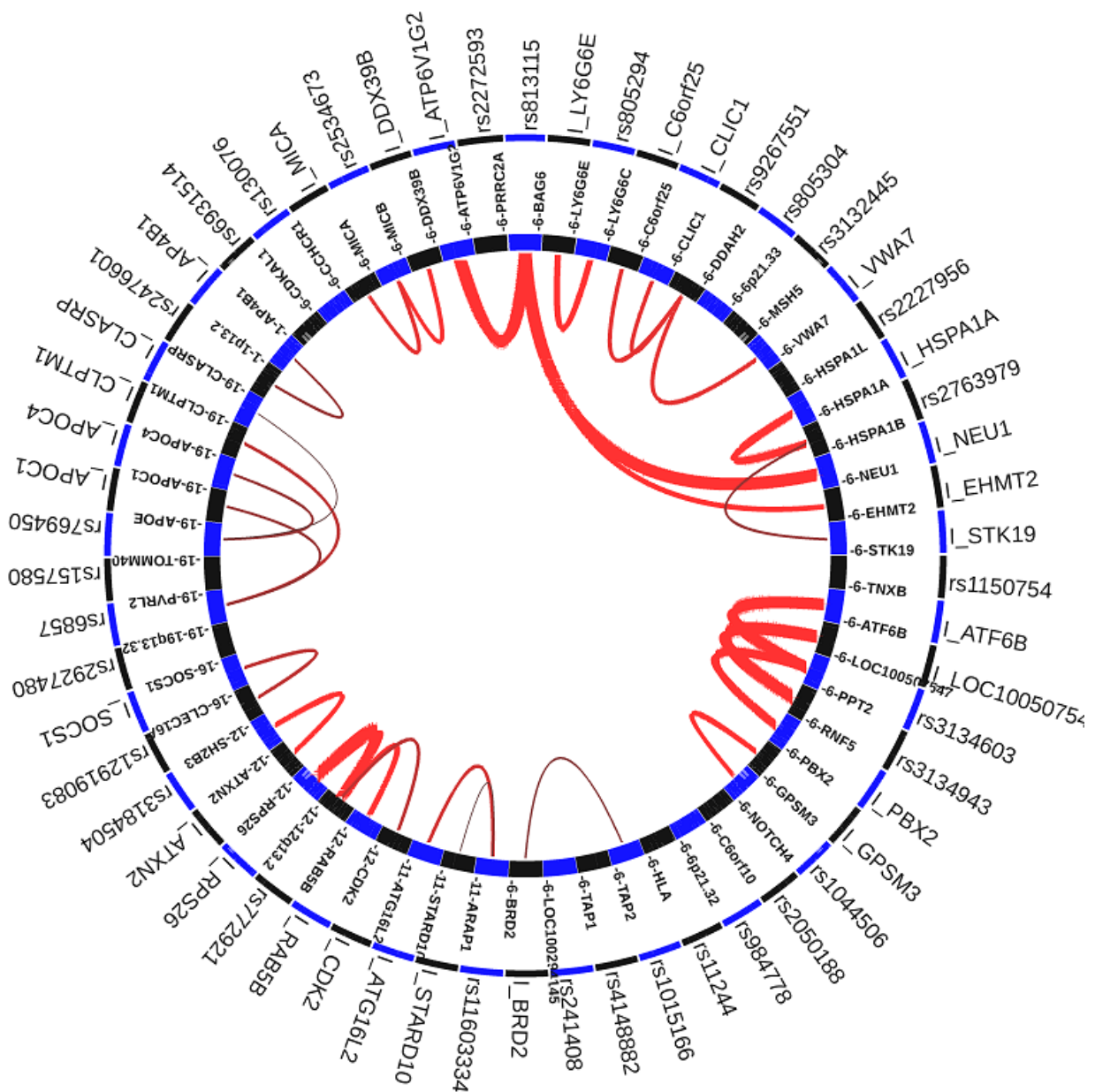
After preparing the parameters, please make sure all required information is filled. Then click the "submit" button, the job will be submitting to web server.

GWAS3D Result Overview

1. Circos-style regulatory variants visualization

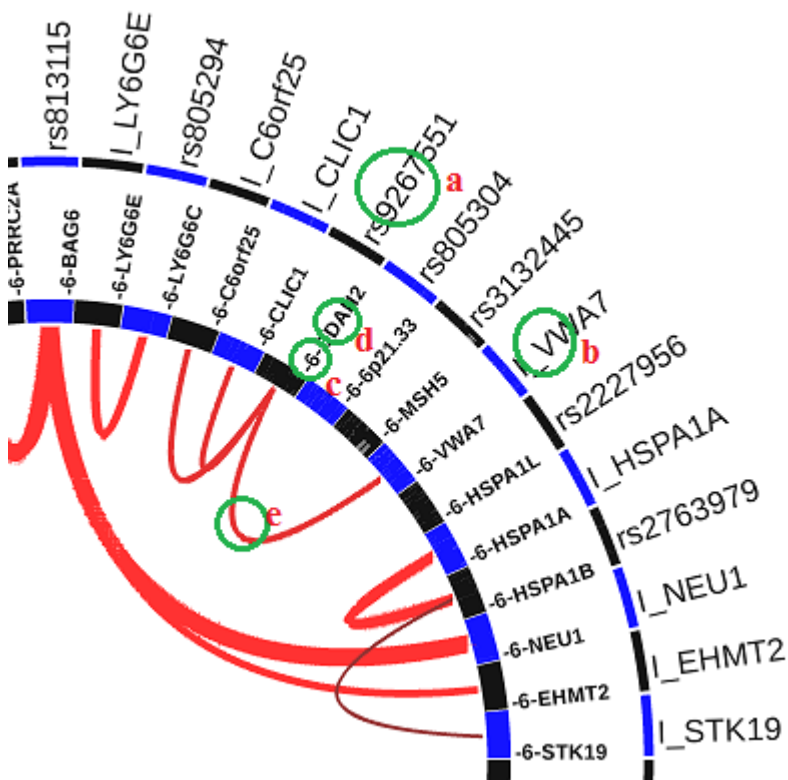
Entering your workspace by clicking the finished job, system will first display a Circos-style graph with some interactive attributes.

1.1 Circos-style plotting for variants visualization displays broad horizontal area genome-widely. Top variants with highest regulatory signals and distal interaction regions are displayed in the outer circle. The genes or genomic locations connected to respective SNVs will be showed in the inner circle.



1.2 Several information can be easily read from the plotting, from the outer to inner, there are significant regulatory variants and distal interaction regions, genes and genomic location, chromosome number and distal interaction indicator line. For example (following figure shown),

GWAS SNP rs9267551 is detected as a significant regulatory variant (a), this variant located on the 5' UTR region of *DDAH2* (d) in chromosome 6 (c). One of important regulatory features for this variant, which can be viewed from the plotting, is the region of rs9267551 located has a long-range interaction signal to another locus near *VWA7*, (b, interactive elements with significant regulatory variant will start with I_) the red line indicated this signal (e) and the intensity of interaction is represented by wideness.



2. GWAS3D signal statistics

Right overview panel will summary the GWAS3D result which indicate the number of significant variants detected, the number of variants have been detected having significant TFBS affinities, the number of variants have been detected affecting the long range interaction, the number of variants have direct effect by GWAS leading SNPs and the number of variants have indirect effect by high LD of GWAS leading SNPs.

Overview	Download
GWAS3D Statistics	
Based on the <i>K562</i> Cell Line and <i>CEU</i> Population, GWAS3D detected:	
360 variants have regulatory signals (affinity change affects promoter activity or distal interaction) in this study.	
360 variants have been detected having significant TFBS affinities changes.	
322 variants have been detected affecting the long range interaction.	
98 variants have direct effect by GWAS leading SNPs.	
262 variants have indirect effect by high LD of GWAS leading SNPs.	

3. Panel switch

GWAS3D Result Details

1. Significant regulatory variants table viewer

GWAS3D provides a table showing significant regulatory variants information detecting by GWAS3D algorithm. In this table, only one top variant in a LD block of leading SNV will be reported, but user can check other significant variants on same LD from "LD signal tab" of right information panel.

1.1 Variant information in table

Significant variants are sorted by their final P-value (**a**) and some variants will be highlighted with orange background if [final P-value/original P-value > 10E5] (**b**). However, some variants can rank at the top area although the final P-value is below than original P-value since those variants are significant enough in GWAS association test. The selected variant will be marked by red background (**c**). Variant with different GWAS3D signals can be represented by color markers which indicate different function elements or chromatin status (**d**) including Leading variant, Significant TFBS affinity, distal interaction, promoter region of a gene, putative enhancer region, CTCF binding region, regulatory region annotated by ChromHMM and GERP++ conservation element (**e**).

SNPID	POS	LOCUS	FINAL_P	LEADSNP	ORIG_P	R ²	STATUS
rs813115	6:31620020	BAG6	9.32e-127	rs1046089	1.28e-124	0.958	
rs805304	6:31698088	6p21.33	1.24e-115	rs707928	8.89e-111	0.957	
rs3134603	6:32126002	PPT2	1.29e-113	rs3134603	3.36e-116	1	
rs3134943	6:32147761	RNF5	1.88e-113	rs3134954	1.4e-108	0.894	
rs4148882	6:32816958	TAP1	3.82e-110	rs3101942	3.09e-110	0.813	
rs1015166	6:32798731	TAP2	6.9e-107	rs1015166	8.68e-107	1	

Leading variant;
 Significant TFBS affinity;
 Mapping on distal interaction;
 Mapping on promoter region of a gene;
 Mapping on putative enhancer region;
 Mapping on CTCF binding region;
 Mapping on regulatory region annotated by ChromHMM;
 Mapping on GERP++ conservation element.

1.2 Table operation

Table can be searched (**a**), adjusted entries (**b**), scrolled (**c**), paged (**d**) under user intervention.

SNPID	POS	LOCUS	FINAL_P	LEADSNP	ORIG_P	R ²	STATUS
rs984778	6:32400088	6p21.32	6.9e-299	rs3129871	1.15e-299	0.96	
rs3135377	6:32385399	6p21.32	4.21e-197	rs3129941	1.37e-197	0.871	
rs28490179	6:32519005	6p21.32	6.08e-191	rs3129860	8.62e-192	0.864	
rs6857	19:45392254	PVRL2	6.78e-160	rs2075650	1.75e-157	0.862	
rs2050188	6:32339897	C6orf10	1.85e-144	rs6903608	1.55e-144	0.803	
rs2647044	6:32667910	6p21.32	1.35e-139	rs2647044	5.18e-142	1	
rs2272593	6:31601344	PRRC2A	4.38e-138	rs3130617	6.29e-135	1	
rs1044506	6:32172065	NOTCH4	6.2e-129	rs3132946	4.92e-126	0.893	
rs2227956	6:31778272	HSPA1L	1.98e-128	rs9267649	3.11e-126	0.951	
rs1573646	6:32731624	HLA-DQB2	5.41e-101	rs2301271	2.38e-100	0.851	

Showing 1 to 50 of 360 entries

2. Significant regulatory variants information panel

6 related information about selected significant variant can be checked from different tabs of right panel including Variant Summary, Binding Affinity, GWAS3D Signals, Genomic Elements, LD Signals and External Browsers.

Variant Summary	Binding Affinity	GWAS3D Signals	Genomic Elements	LD Signals	External Browser	
------------------------	-------------------------	-----------------------	-------------------------	-------------------	-------------------------	--

2.1 Variant summary

2.1.1 General information

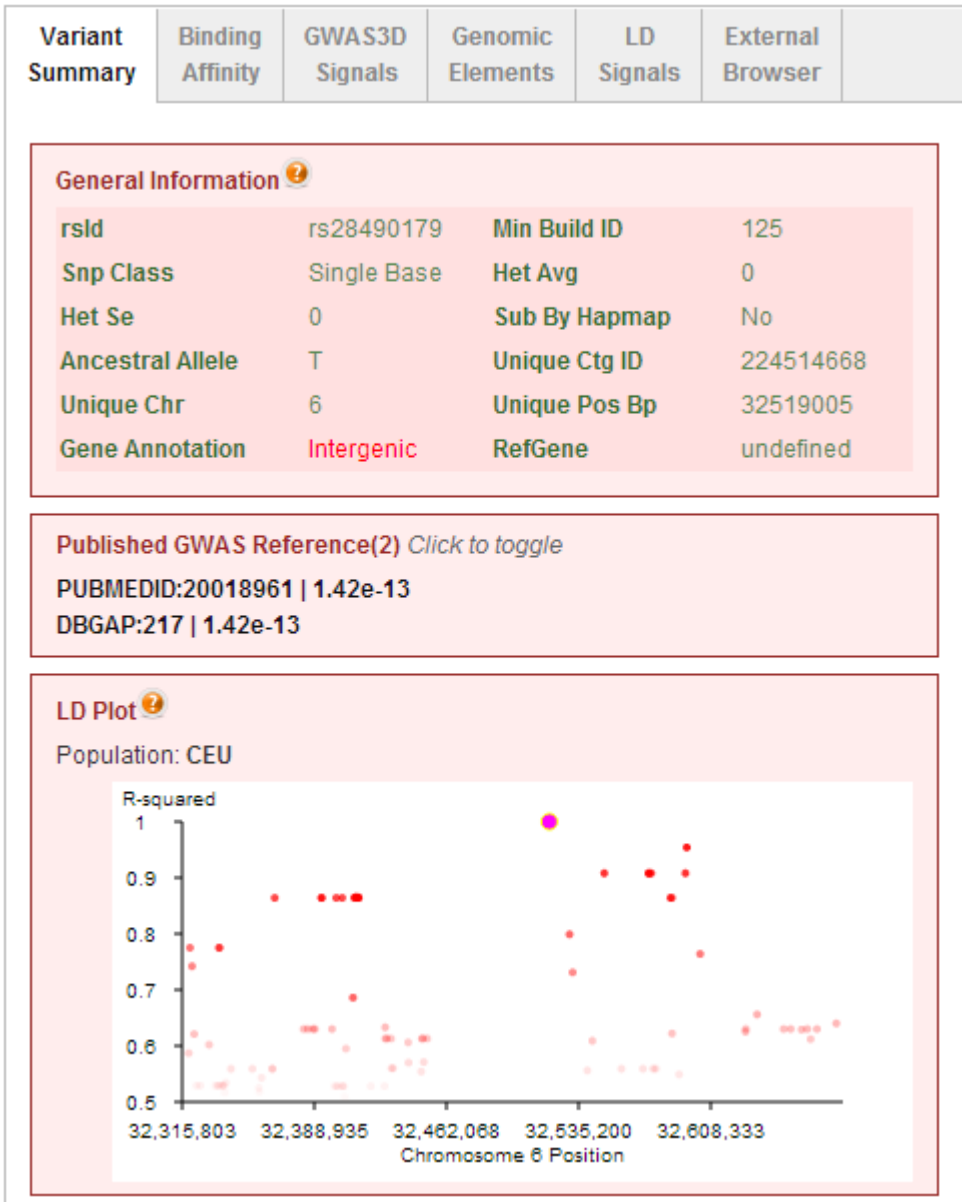
Report the variant basic information for target SNP such as allele frequency, SNP attributes.

2.1.2 Published GWAS Reference

Report the reference or publication if this SNP is reported as significant effect in current GWAS.

2.1.3 LD plot

Hapmap LD information of this variant for investigated population.



2.2 Binding Affinity

GWAS3D will report the significant binding affinity changes for different alleles of target variant by scanning defined factor motifs on variant locus, top 5 factors will be reported with detailed factor and binding information. For example (following figure shown), a known motif of TF *Pou2f2* has a significant binding affinity change between allele T and allele G in the 15th position of that motif. The score and P-value indicate the degree and probability of affinity change, which will be used for GWAS3D prioritization.

Variant Summary	Binding Affinity	GWAS3D Signals	Genomic Elements	LD Signals	External Browser
<p>Long Range Interactions (5C,ChIA-PET,Hi-C)(4) Click to toggle K562 chr6:32128373-32148684 chr6:32116233-32128353 K562 chr6:32128373-32148684 chr6:32093775-32107518 K562 chr6:32128373-32148684 chr6:32154777-32168235 K562 chr6:32128373-32148684 chr6:32934756-32956220</p>					
<p>Gene Promoters (RefGene/User Defined Region)(0)</p>					
<p>Enhancers (H3K4me1,H3K27ac,P300,DHS)(0)</p>					
<p>Insulators (CTCF)(0)</p>					
<p>ChromHMM Markers (Promoter,Enhancer,Insulator)(1) Click to toggle K562 chr6:32146622-32148222</p>					
<p>Conservation Regions (GERP++ Elements)(1) Click to toggle chr6:32147316-32148173</p>					

2.4 Genomic Elements

Comprehensive genomic mapping annotations are provided to indicate which functional element connect to target variant.

2.4.1 Reference Gene

Gene annotation from NCBI Refseq.

2.4.2 Ensemble Gene

Gene annotation from Ensemble.

2.4.3 Known Gene

Gene annotation from UCSC.

2.4.4 Small RNA

snoRNA and miRNA annotations from UCSC.

2.4.5 MicroRNA Target

TargetScan generated miRNA target site predictions.

2.4.6 Long Non-coding RNA



Human long non-coding RNA from re-annotated microarray studies.

2.4.7 HapMap eQTL

HapMap eQTL consensus

2.4.8 ENCODE transcripts and functional elements

ENCODE gene annotation

Variant Summary	Binding Affinity	GWAS3D Signals	Genomic Elements	LD Signals	External Browser
Reference Gene  (0) Click to toggle					
Ensemble Gene  (0) Click to toggle					
Known Gene  (0) Click to toggle					
Small RNA  (0) Click to toggle					
MicroRNA Target (From TargetScan)  (0) Click to toggle					
Long Non-coding RNA(0) Click to toggle					
HapMap eQTL Consensus(0) Click to toggle					
GENCODE V12 Annotation(0) Click to toggle					
GENCODE V12 2wayconspseudos(0) Click to toggle					
GENCODE V12 PolyAs(0) Click to toggle					
GENCODE V12 Long Noncoding RNAs(0) Click to toggle					
GENCODE V12 tRNAs(0) Click to toggle					

2.5 LD Signals

First top variant in a LD block of leading SNV will be reported in the table viewer, here user can check other significant variants on same LD.

Variant Summary	Binding Affinity	GWAS3D Signals	Genomic Elements	LD Signals	External Browser	
Search: <input type="text"/>						
SNPID	POS	LOCUS	FINAL_P	LEADSNP	ORIG	
rs3135377	6:32385399	6p21.32	4.21e-197	rs3129941	1.37e	
rs3132958	6:32297901	C6orf10	1.5e-196	rs3129941	1.37e	
rs4959027	6:32383050	6p21.32	1.99e-195	rs3129941	1.37e	
rs3132959	6:32298942	C6orf10	1.1e-194	rs3129941	1.37e	
Showing 1 to 4 of 4 entries						

2.6 External Browser

GWAS3D provides some useful external browser for annotating and ranking the genetic variant especially for Non-coding variant.

2.6.1 GWASrap QuickRap

Genetic variants representation, annotation and prioritization in the post-GWAS era. Cell Res. 2012 doi: 10.1038/cr.2012.106.

2.6.2 Regulomedb

Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res. 2012 doi: 10.1101/gr.137323.112.*

2.6.3 UCSC ENCODE Browser

An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 6;489(7414):57-74.

Variant Summary	Binding Affinity	GWAS3D Signals	Genomic Elements	LD Signals	External Browser	
<div style="border: 1px solid #ccc; padding: 10px; margin-bottom: 5px;"> Annotate/Prioritize Current SNV on GWASrap QuickRap query </div> <div style="border: 1px solid #ccc; padding: 10px; margin-bottom: 5px;"> Rank/Annotate Current SNV on ENCODE elements (Regulomedb) query </div> <div style="border: 1px solid #ccc; padding: 10px;"> Annotate Current SNV on UCSC ENCODE Genome Browser query </div>						

Download Result

All significant information detected by GWAS3D can be downloaded from the "download tab" of GWAS3D PLOT panel

User can download the significant variants information (one per LD), all significant variants information and GWAS3D plotting in this tab.

The GWAS3D download file contain following information:

(CHRPOS/SNP ID/GENOTYPE/LOCUS/FINAL PVALUE/LEAD SNP/LEADSNP PVALUE/RSQUARE/SIGNALS)

Overview	Download	
Significant variants information (One per LD):	download	
All significant variants information:	download	
GWAS3D plotting:	download	

Retrieve Jobs

There are three ways to retrieve your submit job in GWAS3D.

1. Received by E-mail.

Please fill right E-mail address for the notification in the input page.

Please specify your **E-mail Address** to retrieve your job:

2. Check from a fixed link.

GWAS3D provides a encrypted link for retrieving your job.

GWAS3D Jobs

GWAS3D 

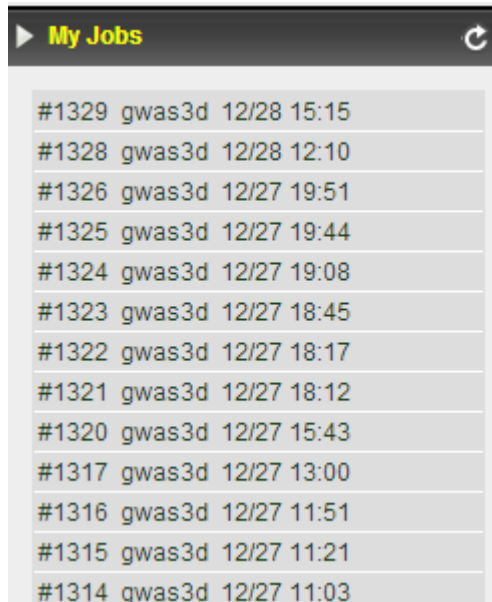
You can retrieve your job from <http://jjwanglab.org/gwas3d/Z3dhcZNLTEzNTY2ODEwMzk>


SGE Queue:

Job ID	Job Name	Run Time(s)	Status
1330	gwas3d-1356681039	Waiting	pending

3. Check from workspace cookies in client browser.

GWAS3D provides a cookies mechanism with your used web browser, it will help you manage all of your submit jobs.



My Jobs 		
#1329	gwas3d	12/28 15:15
#1328	gwas3d	12/28 12:10
#1326	gwas3d	12/27 19:51
#1325	gwas3d	12/27 19:44
#1324	gwas3d	12/27 19:08
#1323	gwas3d	12/27 18:45
#1322	gwas3d	12/27 18:17
#1321	gwas3d	12/27 18:12
#1320	gwas3d	12/27 15:43
#1317	gwas3d	12/27 13:00
#1316	gwas3d	12/27 11:51
#1315	gwas3d	12/27 11:21
#1314	gwas3d	12/27 11:03