



İTF Tıbbi Biyoloji AD
Uygulamalı Biyoinformatik Kursu
02-03 Nisan 2018

Prof. Dr. M. Tevfik DORAK- Liverpool Hope University, UK
Ö. Orkun SUREL- EPROM Ind. Ltd. Co.

Doç. Dr. Mustafa DOĞAN- İTÜ Elektrik-Elektronik BD
Dr. Sedat KARADENİZ- Kadir Has Üniv, Biyoinformatik Bölümü

Yer: İstanbul Tıp Fakültesi, Hulusi Behçet Kitaplığı, Pembe Salon

Sekreteryası: tibbiyolojiad@gmail.com

Tel: 212/6351168



**Kingston
University**
London

Aim

**To Make Sense of Genetic Variants
Associated with Disease Susceptibility
&**

**To Increase Familiarity with Python
as an Encouragement to Learn a Coding Language**

Aim

This course does not cover:

Bioinformatics algorithms

or

Next-generation sequencing data analysis

Biyoinformatik Forumu-2018

Genom Dizileme ve Biyoinformatik

11 MAYIS 2018 CUMA

08:00 - 17:30

GEBZE TEKNİK ÜNİVERSİTESİ

ONLINE
KAYIT



biyoinformatikforumu.org

KONULUĞU VE KONUKLAR



SPONSÖRLER



BİYOİNFORMATİK FORUMU - PROGRAM

11 Mayıs 2018, Cuma

- 8:00 – 9:00 Kayıt ve Tanışma
9:00 – 9:30 **Açılış Konuşmaları**
Yrd. Doç. Dr. Saliha Durmuş
Gebze Teknik Üniversitesi, Biyomühendislik Bölümü & PHI Tech Bioinformatics
Prof. Dr. Haluk Görgün
Gebze Teknik Üniversitesi Rektörü
- 9:30 – 10:10 **Konuşma I: Türkiye Sağlık Enstitüleri Başkanlığı (TÜSEB)**
Prof. Dr. Fahrettin Keleştemür
TÜSEB Başkanı
- 10:10 – 10:50 **Konuşma II: Türk Genom Projesi**
Prof. Dr. Ali Osman Kılıç
TÜSEB Türkiye Biyoteknoloji Enstitüsü Başkanı
- 10:50 – 11:20 Kahve arası
- 11:20 – 11:50 **Konuşma III: Yeni Nesil Dizileme ve Biyoinformatik**
Doç. Dr. Tunahan Çakır
Gebze Teknik Üniversitesi, Biyomühendislik Bölümü & PHI Tech Bioinformatics
- 11:50 – 12:40 **Konuşma IV: Klinik için Genom Profillemeye:**
Seçenekler, Zamanlama, Beklentiler, Zorluklar
Yrd. Doç. Dr. Kaya Bilguvar
Yale Üniversitesi, Genom Analiz Merkezi Direktörü
- 12:40 – 14:10 Yemek arası
- 14:10 – 15:00 **Konuşma V: Ekzom/Genom Dizileme ve Klinik Uygulamalara Örnekler**
Dr. Zafer Yüksel
Centogene AG, Tıbbi Raporlama ve Tıbbi Biyoinformatik Direktörü - Tıbbi Genetik Uzmanı
- 15:00 – 15:30 Kahve arası
- 15:30 – 17:30 **Panel – Türkiye'de Genom Dizileme ve Biyoinformatik**
Moderatör: Prof. Dr. Zehra Oya Uyguner
İstanbul Üniversitesi Tıp Fakültesi, Tıbbi Genetik Anabilim Dalı
Panelistler:
Prof. Dr. Almet Okay Çağlayan
Bilim Üniversitesi Tıp Fak., Florence Nightingale Hastanesi, PHI Tech Bioinformatics
Yrd. Doç. Dr. Kaya Bilguvar
Yale Üniversitesi, Genom Analiz Merkezi Direktörü
Dr. Zafer Yüksel
Centogene AG, Tıbbi Raporlama ve Tıbbi Biyoinformatik Direktörü - Tıbbi Genetik Uzmanı

Schedule

Day 1

Bioinformatics tools for the assessment of genetic variants

**Bioinformatic analysis of experimentally verified functional
SNP associations with disease**

**Bioinformatic tools for the investigation of tissue specificity of
variant effects**

Online tools: Standard operating procedures for gene and variant analysis

Applications and practice on participant projects

Day 2

Introduction to Python as a coding language for bioinformatics

What can you do with programming languages?

What is Python and why Python?

Python basics

Simple Python codes

Demonstrations of more advanced methods

Questions & answers

HLA SNPs

SNPsnap



[Home](#) [About](#) [Match SNPs](#) [Download](#) [FAQ](#) [Documentation](#) [Contact](#) [Feedback](#)

Input

Paste in SNPs or upload a file

[Load example SNPs](#) [chr:pos format or rsID format](#)

Paste in
SNPs

rs13194504



File input

Upload file

Annotation

☒ Annotate matched SNPs*

☒ Annotate input SNPs*

Input loci independence

☐ Report input loci independence (clumping)

*See [documentation - SNP Annotation](#). This will exclude all input SNPs mapping in the HLA region (chr 6 25Mb-35Mb; i.e. SNP with coordinates 6:25000000-6:35000000). We recommend using this setting.

SNP Exclusions

☒ Exclude input SNPs in matched SNPs

☐ Exclude HLA SNPs

Session Information

Please fill in information to identify your job.

Job name

Email

user@mydomain.com

Submit

We recommend otherwise!

HLA SNPs

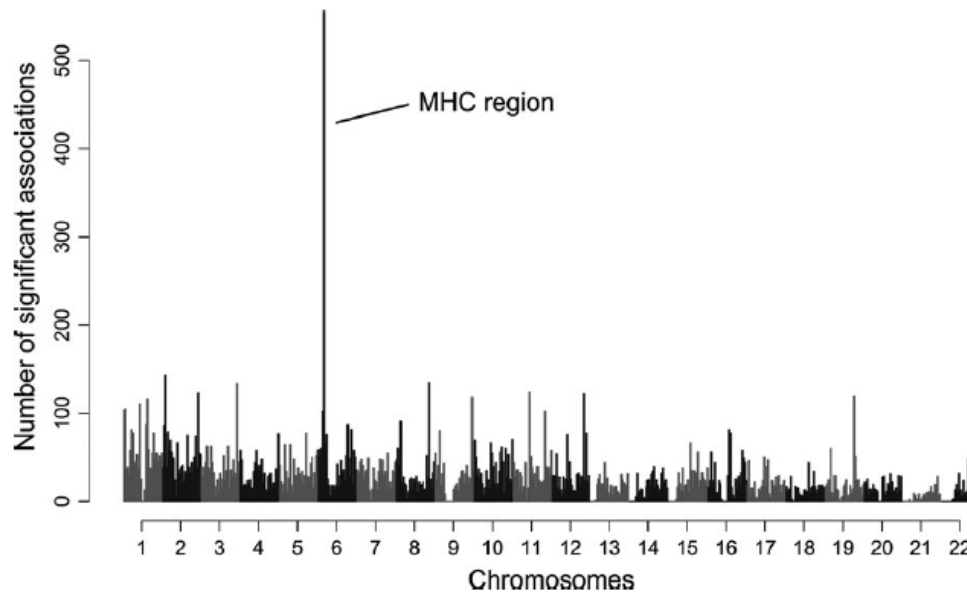


FIGURE 1 Number of significant GWAS associations along the genome. The chromosomal location of significant trait associations from GWAS ($N = 18,682$) is shown for all autosomes. Data from NHGRI GWAS catalog. Reproduced from "Lenz TL, Spirin V, Jordan DM, Sunyaev SR. Excess of Deleterious Mutations around HLA Genes Reveals Evolutionary Cost of Balancing Selection. *Mol Biol Evol* 2016;33(10):2555-64. <https://doi.org/10.1093/molbev/msw127>" by permission of Oxford University Press on behalf of the Society for Molecular Biology and Evolution

.... despite already showing the highest number of disease associations, the true extent of the involvement of the MHC region in disease genetics may not have been uncovered.

Received: 9 March 2017 | Revised: 16 June 2017 | Accepted: 20 July 2017

DOI: 10.1111/jji.12332

REVIEW

WILEY INTERNATIONAL JOURNAL OF IMMUNOGENETICS

What has GWAS done for HLA and disease associations?

A. E. Kennedy¹ | U. Ozbek^{2,3} | M. T. Dorak⁴ 

HLASNP_s (beta)



Home Page

About

Statistics

sopSNP App

Search HLAsnp_s:

You can search multiple SNPs with new lines between SNPs or separated with semicommas!

rs2395185

Search

HLAsnp_s (DB Version 1.1)

Functional assessment of HLA region SNPs

• rs2395185

Master DB Results:

CHR :	6
pos (hg38):	32465390
ALLELES :	G / T
GENE :	

• GWAS Results (1)

• PheWAS Results (7)

• meQTL Results (4)

Course Material

Genome Biology for Genetic Epidemiologists

Mehmet Tevfik DORAK

Genome Biology Abbreviations

Genome Biology Notes

I. Genome Biology and Applied Bioinformatics Course

(28/29 March 2016, Istanbul)

II. Genome Biology and Applied Bioinformatics Course

(17/18 April 2017, Istanbul)

III. Genome Biology and Applied Bioinformatics Course

(2/3 April 2018, Istanbul)

Genetic Epidemiology

Bioinformatics Tools for Genetic Epidemiologists

Mehmet Tevfik DORAK, MD, PhD

[Genetics](#) [Epidemiology](#) [Biostatistics](#) [HLA](#) [MHC](#) [Glossary](#) [Homepage](#)

<http://www.dorak.info/genbiol>

[HLA](#) [MHC](#) [Genetics](#) [Evolution](#) [Biostatistics](#) [Population Genetics](#) [Genetic Epidemiology](#) [Homepage](#)

BIOINFORMATICS TOOLS

(for Genetic Epidemiologists)

Mehmet Tevfik DORAK, M.D., Ph.D.

Day 1 - Lecture 1

Bioinformatics Tools for the Assessment of Genetic Variants

Mehmet Tevfik DORAK, MD PhD
Kingston University London
U.K.

Faculty of
Science,
Engineering
and Computing

Schedule

Day 1

Bioinformatics tools for the assessment of genetic variants

**Bioinformatic analysis of experimentally verified functional
SNP associations with disease**

**Bioinformatic tools for the investigation of tissue specificity of
variant effects**

Online tools: Standard operating procedures for gene and variant analysis

Applications and practice on participant projects

Outline

How to Make Sense of GWAS?

Assigning Function to SNPs

SNP Databases

eQTLs

Chromatin Marks

Chromatin Interactions

SNP Functionality Scores

Gene Set Enrichment Analysis

Gene Networks

How to Make Sense of GWAS?

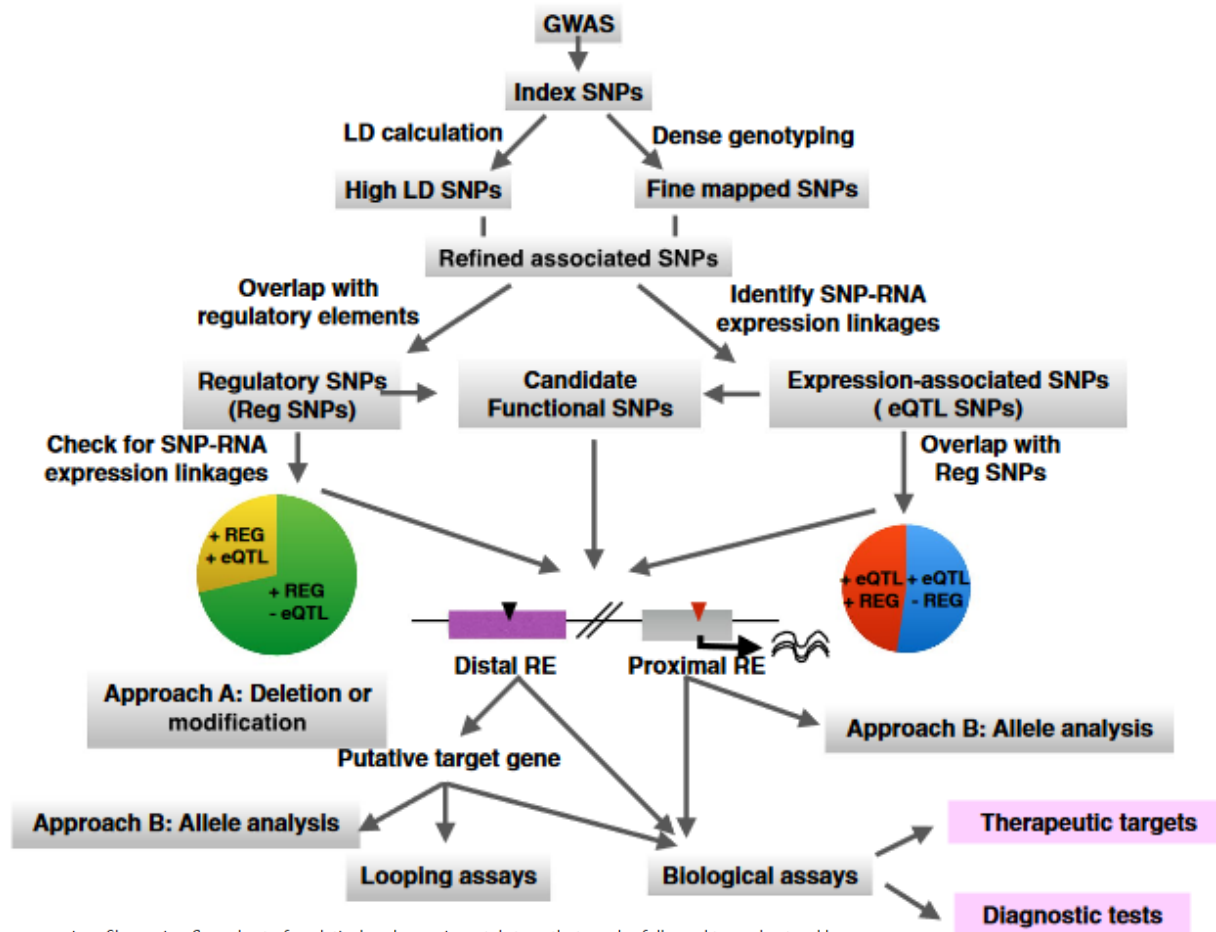
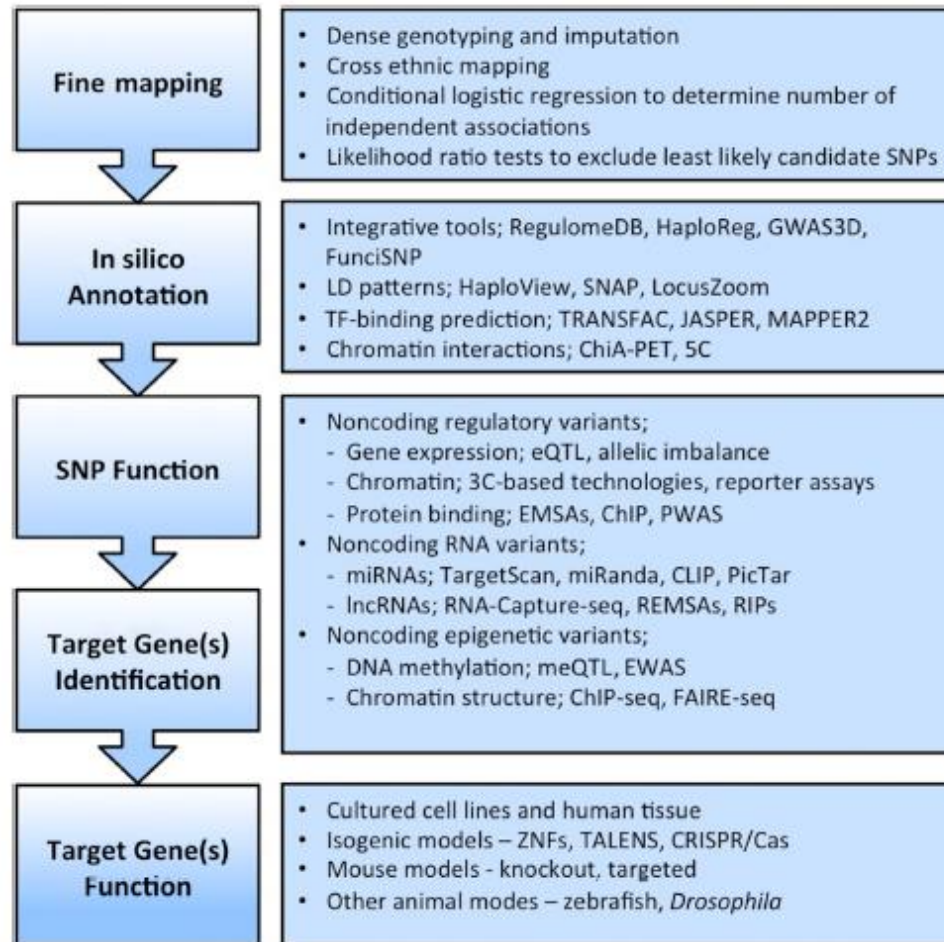


Fig. 1 Making sense of GWAS: an overview. Shown is a flow chart of analytical and experimental steps that can be followed to understand how a non-coding SNP can be associated with an increased risk for a specific disease. Index SNPs are identified using GWAS arrays and then expanded to a larger set of SNPs (termed Refined Associated SNPs) using LD scores and fine-mapping. These Refined Associated SNPs are then prioritized using functional annotation to identify Regulatory SNPs (Reg SNPs) or linkage to allele-specific gene expression to identify eQTL SNPs, producing a set of Candidate Functional SNPs. The Candidate Functional SNPs can either be studied directly or further refined by testing the Regulatory SNPs for possible SNP-RNA linkages or by testing the eQTL SNPs for functional annotation. If a Candidate Functional SNP (yellow arrowhead) lies within a distal regulatory element, it can be deleted or modified using genomic nucleases or epigenomic toggle switches (Approach A); putative target genes are then identified using RNA-seq. Distal regulatory elements that cause changes in gene expression when deleted or modified can then be studied using allele-specific analyses (Approach B); promoters harboring risk-associated SNPs (pink arrowhead) can be directly studied using Approach B. As described in the text, cells deleted for the distal regulatory elements can be used to identify an appropriate phenotypic assay for analysis of the candidate target genes. Then, the genes that show expression changes that are linked to distal SNPs and the genes regulated by the promoter SNPs can be studied using those biological assays to identify possible therapeutic targets and/or candidates for diagnostic tests. Finally, looping assays can be performed to distinguish direct from indirect targets of the distal regulatory elements. It is important to note that a gene whose expression is indirectly affected by a non-coding SNP could be a more important diagnostic or therapeutic target than the directly affected gene

How to Make Sense of GWAS?



1. Workflow for Functionally Analyzing and Interpreting GWAS Loci

REVIEW

Beyond GWASs: Illuminating the Dark Road from Association to Function

Stacey L. Edwards,^{1,2,4,*} Jonathan Beesley,^{1,4} Juliet D. French,^{1,2,4} and Alison M. Dunning^{3,4}

How to Make Sense of GWAS?

Table 2. Computational Tools and Resources for the Analysis of GWAS Loci

Feature	Description	Significance	Experimental Approach	Bioinformatic Tools and Online Resources ^a
Open chromatin	nucleosome-depleted chromatin	sequences harboring regulatory signals	DNase-seq, FAIRE sequencing	ENCODE, NIH Roadmap Epigenomics Project, RegulomeDB, HaploReg, FunciSNP
TF-binding prediction	short DNA consensus recognition sequence characteristic of a particular DNA-binding protein	computationally predicted TF recognition site	position weight matrices	TRANSFAC, JASPAR, MAPPER2
DNA-protein interaction	short DNA sequence associated with a DNA-binding protein after precipitation with a specific antibody	physical protein-nucleic-acid binding (note: no direct evidence of activity)	ChIP-seq, DNase footprinting	ENCODE, NRCistrome, RegulomeDB, HaploReg
DNA methylation	methylation of cytosine residues in CpG dinucleotides	repression of gene expression	methylation array, bisulphite sequencing, MeDIP-seq, MRE-seq	ENCODE, NIH Roadmap Epigenomics Project, MethDB, EpiGraph
RNA expression	detection and measurement of transcribed RNA	coding RNA, noncoding RNA, alternative splicing	RNA-seq, RNA-PET, CAGE	ENCODE, Gene Expression Omnibus, Galaxy
Histone modifications	specific posttranslational modifications of particular histone protein residues are associated with various regulatory activities	H3K4me1: promoters and enhancers H3K4me2: promoters and enhancers H3K4me3: promoters H3K79me2: transcription transition H3K27ac: active regulatory region H3K9ac: promoters H3K9me1: active chromatin H3K9me3: repressed chromatin	ChIP-seq	ENCODE, NIH Roadmap Epigenomics Project, NRCistrome, RegulomeDB, HaploReg, ChromHMM, GWAS3D, Segway, ChroMoS
Chromatin interactions	long-range physical interactions between distal genomic regions	contact between regulatory motifs, such as tissue-specific enhancers and promoters	3C, 4C, 5C, 6C, Hi-C, ChIA-PET	GWAS3D, Hi-C Project, ChIA-PET Browser

Abbreviations are as follows: 3C, chromosome conformation capture; 4C, circular 3C; 5C, carbon-copy 3C; 6C, combined 3C-ChIP-cloning; CAGE, cap analysis gene expression; DNase-seq, DNase hypersensitive site sequencing; MeDIP-seq, methylated DNA immunoprecipitation sequencing; MRE-seq, methylation-sensitive restriction enzyme sequencing; and RNA-PET, RNA paired-end-tag sequencing.

^aNonexhaustive list of examples.

REVIEW

Beyond GWASs: Illuminating the Dark Road from Association to Function

Stacey L. Edwards,^{1,2,4,*} Jonathan Beesley,^{1,4} Juliet D. French,^{1,2,4} and Alison M. Dunning^{3,4}

How to Make Sense of GWAS?

Genetic Study of Complex Diseases in the Post-GWAS Era

Qingyang Huang*

Table 1

Functional mechanisms of GWAS-associated SNPs

Functional category	Computational tool	Experimental approach
RNA splicing	ESE finder, RESCUE-ESE	RNA-seq
Transcription factor binding	TRANSFAC, JASPER, MAPPER2	EMSAs ChIP, eQTL, reporter assays
DNA methylation	MethDB, EpiGraph, ENCODE	MeDIP-seq, MRE-seq, methylation array, bisulphate sequencing
miRNAs-mRNA interaction	miRNASNP, mrSNP, SNPinfo, MirSNP, miRdSNP	Reporter assays, Western blot, RNAi
LncRNA expression or structure	LincSNP, LncRNASNP 3dRNA	RNA-seq, RIP, microarray

How to Make Sense of GWAS?

Genetic Study of Complex Diseases in the Post-GWAS Era

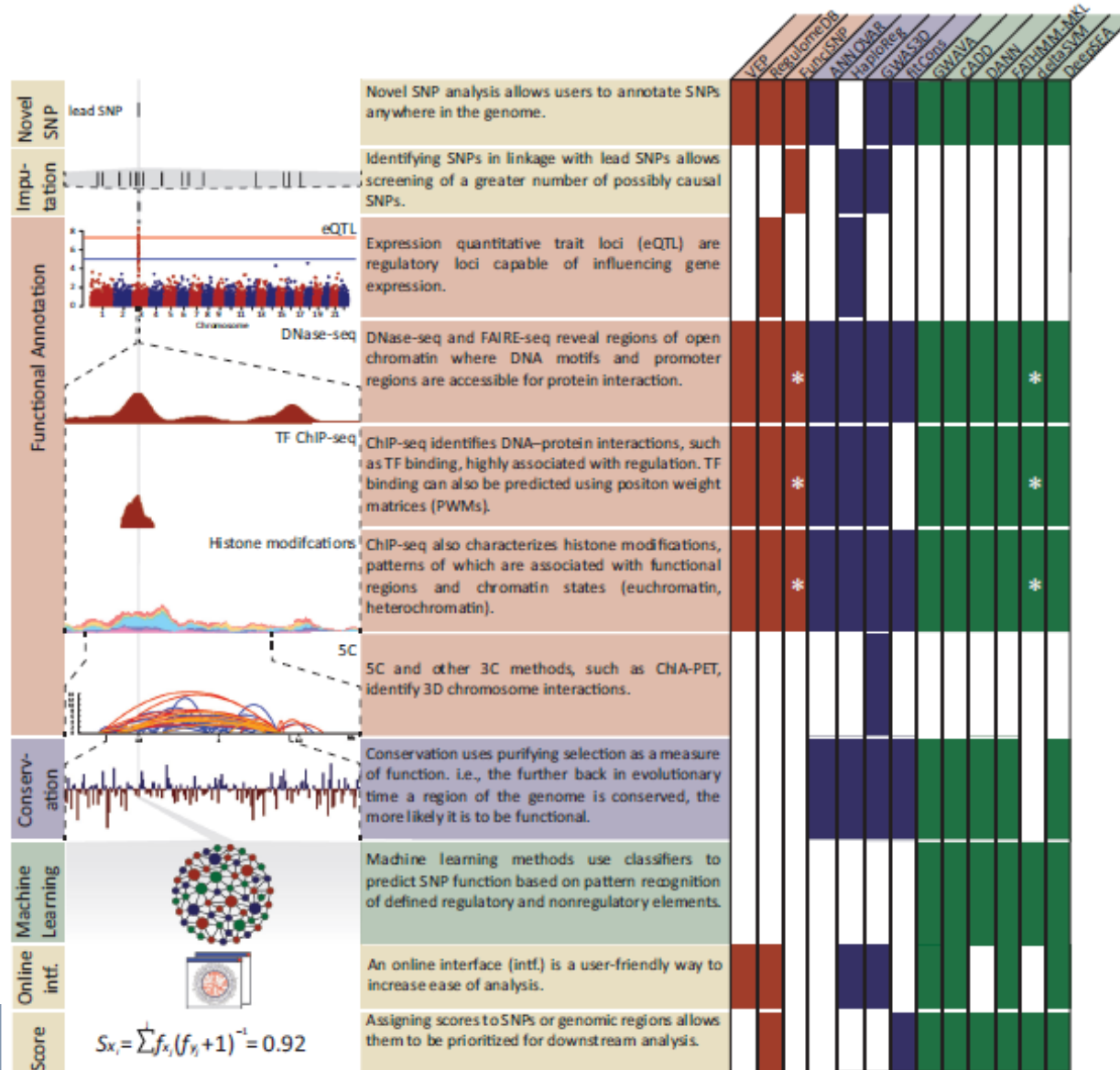
Qingyang Huang*

Table 2

Successfully characterized GWAS functional SNPs

Disease or phenotype	SNP	Function ^a	Target gene of SNP	Reference
Prostate cancer	rs11986220	FOXA1		Jia et al., 2009
Prostate cancer	rs8072254/ rs1859961	AR FOXA1/AP-1	<i>SOX9</i> <i>SOX9</i>	Zhang et al., 2012
Prostate cancer	rs12653946		<i>IRX4</i>	Nguyen et al., 2012
Colorectal cancer	rs6983267	TCF7L2	<i>MYC</i>	Wasserman et al, 2010
Colorectal cancer	rs6983267	TCF4	<i>MYC</i>	Wright et al, 2010
Colorectal cancer	rs16888589	STAT	<i>EIF3H</i>	Pittman et al., 2010
Breast cancer	rs10069690	Splicing	<i>TRET</i>	Bojesen et al., 2013
Breast cancer	rs4784227	FOXA1	<i>TOX3</i>	Cowper-Salari et al., 2012
Breast cancer	rs4442975	FOXA1	<i>IGFBP5</i>	Ghoussaini et al. 2014
Breast cancer	rs554219	ELK4	<i>CCND1</i>	French et al., 2013
Breast cancer	rs7895676	Oct-1/Runx2	<i>FGFR2</i>	Meyer et al., 2008
Breast cancer	rs2981578	C/EBPβ	<i>FGFR2</i>	Meyer et al., 2008
Breast cancer	rs2981578	FOXA1	<i>FGFR2</i>	Meyer et al., 2013
Breast cancer	rs35054928	E2F1	<i>FGFR2</i>	Meyer et al., 2013
Renal cancer	11q13.3	HIF-2	<i>CCND1</i>	Schödel et al, 2012
Prostate cancer	rs339331	HOXB13	<i>RFX6</i>	Huang et al., 2014
T2D	rs10946398	Splicing	<i>CDKAL1</i>	Zhou et al., 2014
T2D/BMI	rs4684847	PRRX1	<i>PPARG2</i>	Claussnitzer et al, 2014

Assigning Function to SNPs



Data and Tools Used to Analyze Noncoding Variants.

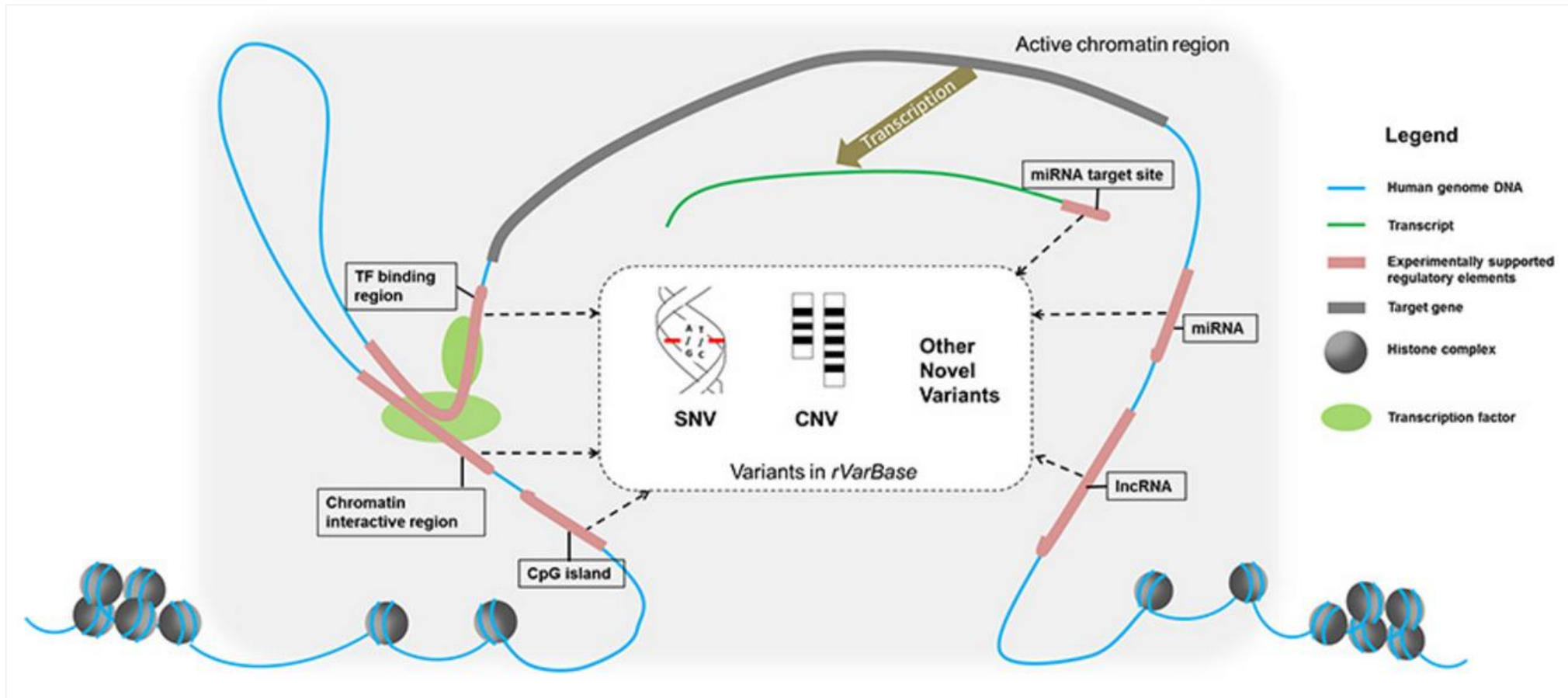
Single nucleotide polymorphism (SNP) aligned with functional (red) and conservation (blue) data, machine learning methods (green), and tool features (yellow). Each tool discussed in this perspective is labeled with annotation types used in its noncoding variant analysis platform. * represents optional input data sets supplied by the user.

Abbreviations: 3C, chromosome conformation capture; 5C, chromosome conformation capture carbon copy; CADD, combined annotation-dependent depletion; ChIA-PET, chromatin interaction analysis by paired-end tag sequencing; DANN, deleterious annotation of genetic variants using neural networks; DNase-seq, DNase I hypersensitive sites sequencing; eQTL, expression quantitative trait loci; FAIRE, formaldehyde-assisted isolation of regulatory elements; FunciSNP, Functional Identification of SNPs; GWAVA, genome-wide annotation of variants; TF, transcription factor; VEP, Variant Effect Predictor.

Mining the Unknown: Assigning Function to Noncoding Single Nucleotide Polymorphisms

Sierra S. Nishizaki¹ and Alan P. Boyle^{1,2,*}

Assigning Function to SNPs



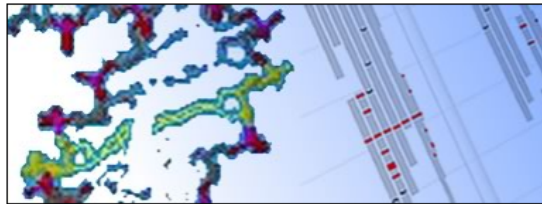
SNP Databases

dbSNP (ENTREZ SNP) 1KG Project / Ensembl Reference Variant Store

NCBI Resources How To Sign in to NCBI

dbSNP SNP Search

Advanced Help

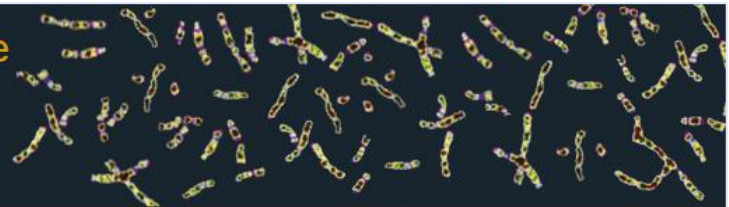


dbSNP

Database of single nucleotide polymorphisms (SNPs) and multiple small-scale variations that include insertions/deletions, microsatellites, and non-polymorphic variants.

IGSR: The International Genome Sample Resource

Providing ongoing support for the 1000 Genomes Project data



Home About Data Portal Analysis Contact Browser FAQ

Search 1000genomes



BLAST/BLAT BioMart Tools Downloads Help & Documentation Blog Mirrors

Login/Register



Human (GRCh38.p10) ▼

Search Human...



SNP Databases

RVS

Queries

API

Data sources

About

Contact Us

DIVAS

the Reference Variant Store

The Reference Variant Store holds information on more than 520 million genetic variants using various methods of annotation. Sources for variants in RVS are re-sequencing projects such as the 1000 Genomes, ESP6500, UK10K, TCGA, Scripps Welllderly; clinical annotation databases such as ClinVar and HGMD; and hypothetical, amino-acid changing single-point mutations from dbNSFP. We provide annotations on low level effects, phenotypes and diseases, population frequencies, and predictive functional impact scores.

Available queries

Our query interface is available via the "Queries" menu. Entry points are searches by specific coordinate, dbSNP, or gene; or a comparison of cohorts based on populations in RVS. The first four queries are also wrapped in our [REST API](#).

Query by dbSNP: enter dbSNP IDs (rs-numbers) to search RVS for matching variants.

Query by coordinates: enter specific chromosomal locations with alternate alleles to search RVS for genetic variants; or enter chromosomal regions to search RVS for all variants falling into those regions.

Query by gene: enter one or more genes and, optionally, genetic variants to search RVS.

Query by phenotype: enter a phenotype or disease name, and optionally a gene to search.

Query by cohorts: choose two cohorts from the available datasets and find the differences in allele frequencies with respect to selected genes.



Integrating 400 million variants from 80,000 human samples with extensive annotations: towards a knowledge base to analyze disease cohorts

Jörg Hakenberg^{1,3*}, Wei-Yi Cheng^{1,4}, Philippe Thomas^{2,5}, Ying-Chih Wang¹, Andrew V. Uzilov¹ and Rong Chen^{1*}

SNP Databases

Table 1 Number of variants imported from various external resources

Study	Variant sites	Variants	Unique to study	Variants passed	Samples
1000 Genomes [1]	81,195,126	81,693,252	57,400,612	all	2,504
ESP6500 [2]	1,982,177	1,998,204	184,225	all	6,503
UK10K [47] ALSPAC/TWINS	37,258,978	37,560,436	6,155,493	all	2,432
UK10K with disease ^c	9,391,582	11,177,227	8,847,466	9,969,036	4,888
TCGA [4] germline ^c	200,691,728	219,533,884	90,884,769	n/a	4,224
TCGA somatic	876,970	890,172	696,754	all	4,205
Scripps Welllderly [48]	76,144,271	91,947,469	63,331,143	53,303,437	534
ExAC ^b [3]	9,579,712	10,450,724	6,581,946	8,811,372	63,352
MSSM BioBank genotyping	849,806	849,806	0	all	11,210
In-house resequencing study	29,326,393	29,671,729	10,134,258	23,610,572	142
Total observed	358,152,122	399,404,510	244,216,666	>217,796,115	82,558 ^b
Other resources:					
dbNSFP ^a [18]	30,523,109	89,617,785	73,561,239	—	—
ClinVar [12]	101,317	104,455	31,694	—	—
OMIM [49]	10,863	10,913	—	—	—
COSMIC [50]	1,483,983	1,525,243	—	—	—
PharmGKB ^c [51]	672	684	—	—	—
SwissVar ^d	(77,047)	(84,649)	(34,198)	—	—
HGMD ^c [13]	125,744	133,464	32,178	—	—
Literature mining	—	890,665	—	—	—
Total observed + other	388,902,292	472,965,749	317,841,777	>217,796,115	82,558

The first block refers to sequencing/genotyping studies, the second to sample-independent annotation databases. ^aUnique to study^a counts variants that were observed only in that particular study. ^bVariants passed^b refers to variants that passed quality metrics as defined by the particular study, at least one sample has to pass; n/a: individual sample quality metrics not available. Totals exclude duplicates seen in different studies. Variants in annotation databases are included only if they can be mapped to precise coordinates and allele. Since a large proportion of the variants discovered by literature mining are given at the protein level only, they were not compared to other studies

^adbNSFP contains hypothetical variants, see text

^bExAC includes samples from 1000 Genomes, ESP6500, and TCGA

^cNote that data from HGMD, PharmGKB, UK10K diseases and TCGA germline are not visible to external users on the RVS website

^dCounts for SwissVar refer to distinct amino acid changes. Further details on individual resources are provided in Additional file 4: Table S3

Hakenberg et al. BMC Bioinformatics (2016) 17:24
DOI 10.1186/s12859-015-0865-9

BMC Bioinformatics

DATABASE

Open Access



Integrating 400 million variants from 80,000 human samples with extensive annotations: towards a knowledge base to analyze disease cohorts

Jörg Hakenberg^{1,3*}, Wei-Yi Cheng^{1,4}, Philippe Thomas^{2,5}, Ying-Chih Wang¹, Andrew V. Uzilov¹ and Rong Chen^{1*}

SNP Databases

[RVS](#)[Queries](#)[API](#)[Data sources](#)[About](#)[Contact Us](#)[DIVAS](#)

RVS queries

[Query by coordinates](#)[Query by dbSNP](#)[Query by gene list](#)[Query by phenotype](#)[Query by cohorts \(beta\)](#)

Paste a list of max. 10 variants using coordinates and ref/alt alleles. Alternatively, enter a chromosomal region to retrieve all variants found in that region.

Example for a single nucleotide variant: chr7:140453136A>T

Example for a chromosomal region: 7:140430000-140440000 — we accept regions of up to 10,000bp.

Limit: ☐ exonic ☐ non-synonymous ☒ all

☐ Display only variants observed in 1000KG Ph3, ExAC, ESP6500, UK10K ALSPAC/TWINS, and/or Scripps Welllderly

Output: ☐ Display results for the canonical isoform only (where applicable)

Format: ☒ HTML ☐ JSON ☐ Excel

[Search](#)[Click here to paste examples](#)

Hakenberg et al. BMC Bioinformatics (2016) 17:24
DOI 10.1186/s12859-015-0865-9

BMC Bioinformatics

DATABASE

[Open Access](#)



Integrating 400 million variants from 80,000 human samples with extensive annotations: towards a knowledge base to analyze disease cohorts

Jörg Hakenberg^{1,3*}, Wei-Yi Cheng^{1,4}, Philippe Thomas^{2,5}, Ying-Chih Wang¹, Andrew V. Uzilov¹ and Rong Chen^{1*}

eQTLs

eQTL effects are the most common mediators of GWAS signals

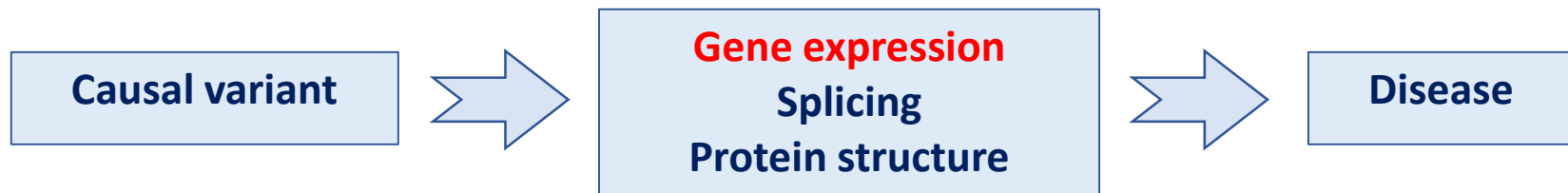


Figure 1.1. Intermediate mechanisms mediating causal variant's effect on disease susceptibility. Genetic variants modify disease risk by causing changes in gene expression (most common), splicing process or protein structure.

eQTLs

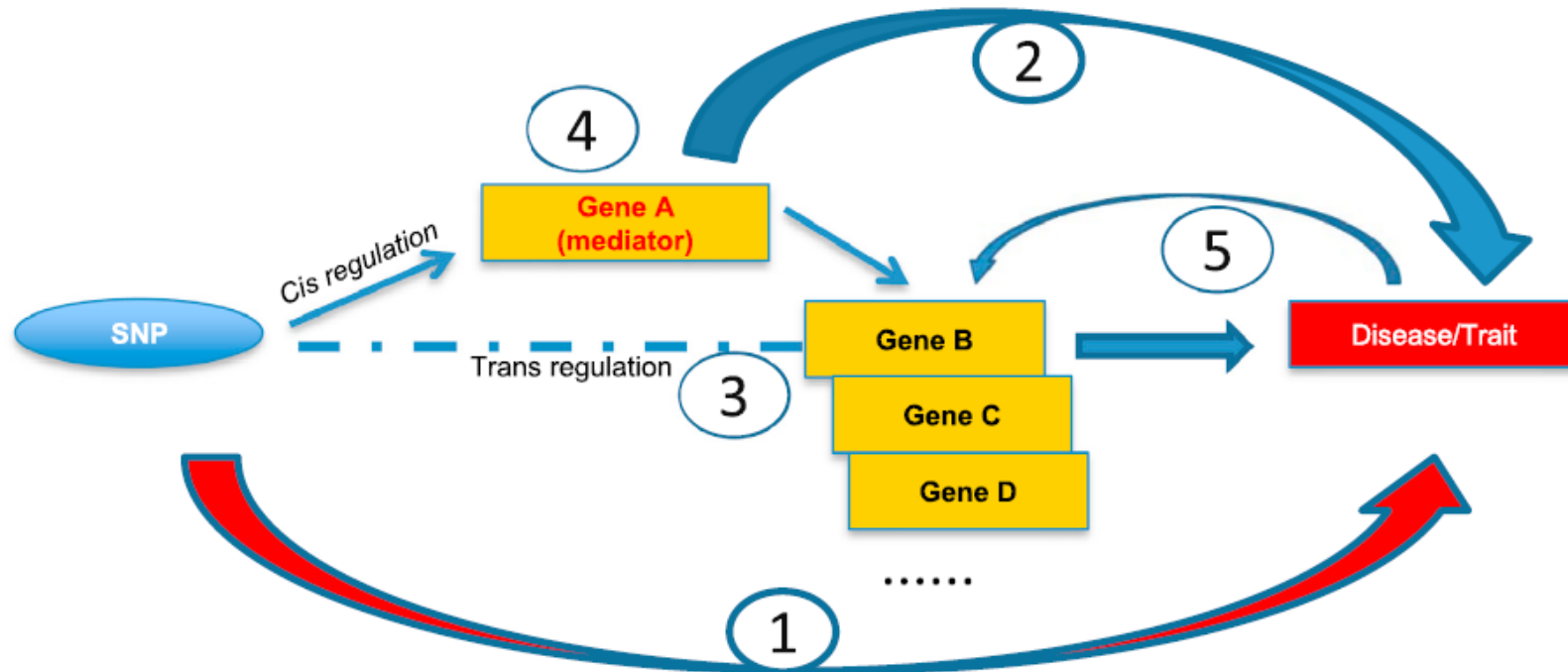


Figure 2. Mediation Mechanisms of eQTLs

Genetic variants can affect traits through the following mechanisms: (1) missense SNP affects protein structure/function; (2) non-coding SNP affects gene expression (*cis*); (3) non-coding SNP affects remote (*trans*) gene expression directly or by (4) *cis*-eGene mediation of the *trans*-eQTL-*trans*-eGene association; or (5) reverse causality (trait has feedback effect on gene expression).

eQTL Databases

Most commonly used eQTL databases:

GTEx (multiple tissues)

Blood eQTL (meta-analysis)

CBI Molecular QTL Browser (incl. FHS data) (mega-analysis)

eQTL / meQTL Resources

[GTEx](#) .. [NCBI Molecular QTL Browser](#) (ref) .. [SCAN](#) (tutorial) .. [PheGenI](#) .. [QTLminer](#) (ref) .. [eQTL Databases](#)
[eQTL Resources @ Pritchard Lab](#) (browser) .. [Blood eQTL Browser](#) .. [BIOS QTL Browser](#) (ref1 / ref2) .. [eQTL Catalog \(NESDA/NTR\)](#) (Ref)
[CAGE eQTL Browser](#) (ref) [ImmunPop eQTL Browser](#) (ref) [GWAS2Genes](#) (ref)
[mQTLdb](#) (ref) .. [e/meQTL](#) .. [reQTL](#) (ref) .. [HSM \(Haplotype-specific Methylation-associated SNPs\)](#) (ref)
[FIRE \(Functional Inference of genetic variants that Regulate gene Expression\)](#) (ref)
[seeQTL](#) (help / ref) [Plateletomics](#) (Platelet transcriptomics by gender) [ImmGen eQTL Browser](#) (mouse)
[eCAVIAR: colocalization of GWAS signals and eQTLs](#) (ref / manual) [PancanQTL](#) (ref)
[DIVAN](#) (ref)

[Home](#)
[GTE_x](#)
[Datasets](#)
[Expression](#)
[QTLs & Browsers](#)
[Sample Data](#)
[Biobank](#)
[Documentation](#)
[Publications](#)
[Contact](#)
[FAQs](#)

Hello, Mehmet Tervik!

- Single-Tissue eQTLs
- Gene eQTL Visualizer
- IGV eQTL Browser
- Multi-tissue eQTLs
- Test Your Own eQTL Dashboard
- Splice QTLs
- Protein Trunc.
- Variants
- Genomic Imprinting

2017-10-19

GTE_x Phase 2 Papers Published

[Read More >>](#)

Current Release

Latest Version: V7

[Dataset Summary Statistics Report](#)

[How to cite or acknowledge the GTE_x project?](#)

Genetic Association

[Single-Tissue eQTLs](#)

[IGV eQTL Browser](#)

Transcriptome

[Top 100 Expressed Genes in a Tissue \(e.g. Blood\)](#)

[Gene Expression in Tissues](#)

Blood eQTL Browser

Blood eQTL browser

This web page accompanies the manuscript titled '[Systematic identification of trans-eQTLs as putative drivers of known disease associations](#)', by Westra *et al*, which has been published in Nature Genetics. If you want to use any of the *cis*- or *trans*-eQTL results displayed on this page in your publication, please cite this paper as indicated below. For further questions, contact the corresponding author: lude@ludesign.nl

Download eQTL Results

You can download the full *cis*- and *trans*-eQTLs, detected at a false-discovery rate of 0.50:

[Cis-eQTLs \(FDR 0.5\)](#)

[Trans-eQTLs \(FDR 0.5\)](#)

How to cite

If you use the eQTLs present on this website in your paper or research, please cite our work: [Download citation directly from Nature Genetics](#)

Query eQTL Results

Or, you can query the *cis*- and *trans*-eQTLs below (examples: rs7807018 or VWCE):

Gene or SNP name:

NCBI eQTL Browser (for FHS)

NCBI Resources How To

Sign in to NCBI

Molecular QTL
Browser

PubMed

Search

Genetic Marker

[Clear all filters](#)
[Clear](#)

RS numbers:

Chromosome: Any

Marker Gene:

Molecular Phenotype

[Clear](#)

Probe Ids:

Chromosome: Any

Probe Gene:

Association Test

[Clear](#)

QTL Type: Any

log10(p-value) < Any

R² value > Any

☐ cis ☐ trans

Phenotype Traits*

[Clear](#)

☒ GWAS trait ☐ ClinVar trait

Study/Publication

Bioprocess/Molecular Function

Columns Show 10 entries

Search:

	Study		Number of QTL Results		Number of			External References		
	ID	Name	cis	trans	Markers	Probes	Subjects	PubmedID	IndGenotypes	MolPhenotypes
☐	1	Population genomics of human gene expression	63839	1224	25884	1525	45-210	17873874	HapMap_v21	GSE6536
☐	7	Mapping the Genetic Architecture of Gene Expression in Human Liver	5338	0	2976	4951	427	18462017	NA	GSE9588
☐	21	Abundant Quantitative Trait Loci Exist for DNA Methylation and Gene Expression in Human Brain	53693	18161	34249	4631	150	20485568	phs000249	GSE15745
☐	39	Transcriptome genetics using second generation sequencing in a Caucasian population	30223	0	15858	7019	60	20220756	HapMap3	E-MTAB-197 E-MTAB-198
☐	40	FHS eQTL	4073000	13503754	7000223	17873	5622	28122634	phs000342	phs000363
☐	41	FHS eGene leadQTLs	19553	6801	24894	12396	5622	28122634	phs000342	phs000363
☐	42	A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines	193481	6739	124158	2087	950	23345460	EGAS00000000137	E-MTAB-1428 E-MTAB-1425
☐	43	FHS microRNA eQTLs	8550	270	5528	87	283-5275	25791433	phs000342	phs000363
☐	44	Genetic Architecture of MicroRNA Expression: Implications for the Transcriptome and Complex Traits	0	31	31	6	107	22658545	HapMap1	GSE34406
☐	45	Identification of cis- and trans-regulatory variation modulating microRNA expression levels in human fibroblasts	11	18	27	23	180	21147911	EGAS00000000056	GSE24610

Showing 1 to 10 of 54 entries

First Previous 1 2 3 4 5 Next Last

Display Results Download

Trans-eQTLs

Table 1. *trans*-eQTL Hotspots

Hotspot Location (hg19)	Number of <i>trans</i> -eQTLs	Number of <i>trans</i> -eGenes Associated with Index eQTL	Directional Bias of <i>trans</i> -eGenes Associated with Index eQTL	Traits Associated in GWAS with Index eQTLs	<i>trans</i> -eGene Enrichment in TF Motifs ^a
1: 205,187,981–205,244,972	10	10	+64%	platelet count	NA
1: 248,039,451	1	12	–58%	red blood cell count	<i>STAT1/STAT2</i>
2: 60,708,597–60,725,451	14	14	–79%	fetal hemoglobin level	<i>NEAT/SP1</i>
3: 50,093,209	1	24	+100%	age at menarche	NA
3: 56,849,749–56,865,776	2	126; 84	–94%; +92%	platelet count; mean platelet volume	<i>TCF3/ETS2</i>
6: 135,411,228–13,543,5501	13	22	–55%	fetal hemoglobin	NA
6: 139,840,693–139,844,429	13	48	–70%	erythrocyte count	<i>SP1/TCF3</i>
7: 50,423,963–50,562,361	19	76	–59%	childhood acute lymphoblastic leukemia	<i>PAX4</i>
12: 54,712,308–54,736,470	2	14	+79%	mean platelet volume	<i>ETS2</i>
12: 111,884,608–112,610,714	9	13	–62%	LDL cholesterol; blood pressure; asthma	NA
16: 57,061,189–57,061,189	2	10	–100%	HDL cholesterol	<i>IRF8/IRF2</i>
17: 27,072,463–27,322,441	45	32	+55%	mean corpuscular volume	<i>E4F1</i>
17: 33,796,260–33,944,055	4	51	+75%	mean platelet volume	<i>ETS2/MAZ</i>

Plus sign (+) denotes the positive association; minus sign (–) denotes the negative association.

^aTranscription factors whose motifs were matched with promoter regions [–2 kb, 2 kb] around transcription start site of the *trans*-eGenes; NA, no TF target enrichment.

meQTL Databases

BIOS QTL browser

This web page accompanies the manuscripts titled "[Disease variants alter transcription factor levels and methylation of their binding sites](#)", by *Bonder et al* and "[Identification of context-dependent expression quantitative trait loci in whole blood](#)", by *Zhemakova et al*, both have been published to Nature Genetics. For further questions, contact the corresponding author: lude@ludesign.nl

Download meQTL results

You can download the independent top *cis*- and *trans*-meQTL and eQTMs, detected at a false-discovery rate of 0.05:

[Cis-meQTLs independent top effects](#)

[Cis-eQTMs independent top effects](#)

[Trans-meQTLs top effects](#)

You can also download the full results of the primary cis-meQTL mapping at a false-discovery rate of 0.05:

[Full list of primary cis-meQTLs](#)

Download eQTL results

You can download the *cis*-eQTLs detected at a false-discovery rate of 0.05:

[Cis-eQTLs Gene-level independent top effects with context specific effects](#)

[Cis-eQTLs Gene-level all primary effects](#)

[Cis-eQTLs Exon-level independent top effects](#)

[Cis-eQTLs Exon-ratio independent top effects](#)

[Cis-eQTLs PolyA-ratio independent top effects](#)

Query results

Or, you can query our independent eQTL, meQTL and eQTMs below (examples: rs3774937, cg10154826, ENSG00000226979 or LTA):

Gene or SNP name:

mQTL Database

Home

Search

Download

Help



mQTLdb

Large-scale genome-wide DNA methylation analysis of 1,000 mother-child pairs at serial time points across the life-course (ARIES).

[Learn more about ARIES »](#)

Search

Perform a quick search for mQTL across the ARIES mQTL database.

Download

Download the entire ARIES mQTL database.



Home

About

eQTL

meQTL

EPigenetics

Network

Real time eQTL

Links

With 400 children from families recruited through a proband with asthma, we have performed a genome-wide analysis of common genetic variation controlling differential expression of transcript isoform. Based on the genotyped 301,173 SNPs, 6,032,596 SNPs are imputed using 1000 genome CEU template. Expression quantitative trait loci from measurements were identified in Epstein-Barr virus-transformed lymphoblastoid cell lines...

[Learn more →](#)



About

Overview of the eMeQTL. Browse and the people behind it



Guide

Detailed documentation, guidelines and tutorials



Community

Forum for questions and announcements



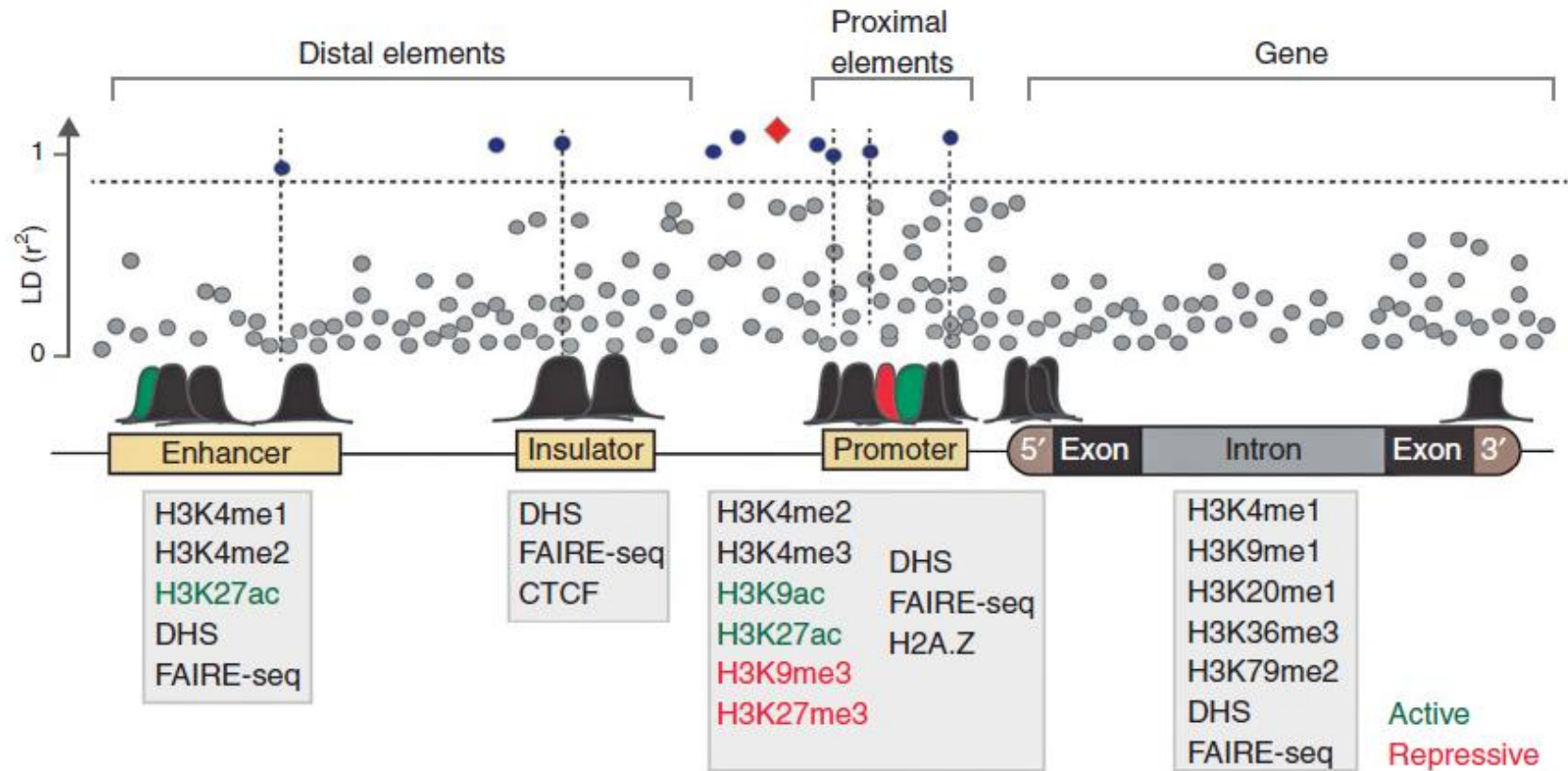
Videos

Tutorial videos, Q&A and event recordings



Created and maintained by Dr. Jinyan HUANG.
z.j.huang@hsph.harvard.edu
Best Screen Resolution for browsing this websites is: 1280 x 800.

Chromatin Marks



Chromatin Marks

HaploReg v4



HaploReg is a tool for exploring annotations of the noncoding genome at variants on haplotype blocks, such as candidate regulatory SNPs at disease-associated loci. Using LD information from the 1000 Genomes Project, linked SNPs and small indels can be visualized along with chromatin state and protein binding annotation from the Roadmap Epigenomics and ENCODE projects, sequence conservation across mammals, the effect of SNPs on regulatory motifs, and the effect of SNPs on expression from eQTL studies. HaploReg is designed for researchers developing mechanistic hypotheses of the impact of non-coding variants on clinical phenotypes and normal variation.

Update 2015.09.15: Version 4 now includes many recent eQTL results including the GTEx pilot, four different options for defining enhancers using Roadmap Epigenomics data, and a complete set of source files for download and local analysis. Older versions available: [v3](#), [v2](#), [v1](#).

[Build Query](#) [Set Options](#) [Documentation](#)

Use one of the three methods below to enter a set of variants. If an r^2 threshold is specified (see the Set Options tab), results for each variant will be shown in a separate table along with other variants in LD. If r^2 is set to NA, only queried variants will be shown, together in one table.

Query (comma-delimited list of rsIDs OR a single region as chrN:start-end):

or, upload a text file (one refSNP ID per line):

or, select a GWAS:

Query SNP: **rs6964969** and variants with $r^2 \geq 0.8$

chr	pos (hg38)	LD (r ²)	LD (D')	variant	Ref	Alt	AFR freq	AMR freq	ASN freq	EUR freq	siPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	eQTL results	Motifs changed	GENCODE genes	dbSNP func annot
7	50398132	0.96	0.99	rs62447205	A	G	0.23	0.22	0.13	0.31			BLD	4 tissues			4 altered motifs	IKZF1	intronic
7	50398606	0.96	0.99	rs11978267	A	G	0.22	0.22	0.13	0.31		LNG	BLD			7 eQTL results	Foxp1,Pax-6	IKZF1	intronic
7	50398997	0.96	0.99	rs6973210	G	A	0.23	0.22	0.13	0.31			BLD	BLD,PANC		7 eQTL results	4 altered motifs	IKZF1	intronic
7	50399099	0.97	0.99	rs6960400	A	G	0.23	0.22	0.13	0.31			BLD	BLD,BLD			ATF3,Maf,NF-E2	IKZF1	intronic
7	50401254	0.98	0.99	rs10278451	G	T	0.23	0.22	0.13	0.31						7 eQTL results	Sox	IKZF1	3'-UTR
7	50401853	0.98	0.99	rs11552047	C	T	0.22	0.22	0.12	0.31			BLD				Nr2e3,Zec	IKZF1	3'-UTR
7	50402283	0.98	0.99	rs11980379	T	C	0.23	0.22	0.12	0.31			BLD	THYM		8 eQTL results	NF-E2	IKZF1	3'-UTR
7	50402678	0.98	0.99	rs200338223	C	CT	0.23	0.22	0.12	0.31			BLD				12 altered motifs	IKZF1	3'-UTR
7	50402680	0.96	0.99	rs33999320	T	TC	0.23	0.22	0.12	0.31			BLD				10 altered motifs	IKZF1	3'-UTR
7	50402906	0.98	0.99	rs4132601	T	G	0.23	0.22	0.12	0.31			BLD			8 eQTL results		IKZF1	3'-UTR
7	50403915	0.99	1	rs11980407	G	A	0.22	0.22	0.12	0.30			BLD			7 eQTL results	Pou5f1	IKZF1	3'-UTR
7	50404626	1	1	rs62445866	G	A	0.22	0.22	0.12	0.30			BLD	12 tissues	CTCF,RAD21		NRSF	IKZF1	3'-UTR
7	50405144	1	1	rs58923657	C	T	0.22	0.22	0.12	0.30		BLD	BLD, THYM, SPLN	BLD,BLD			CHOP::CEBPa	42bp 3' of IKZF1	
7	50405553	1	1	rs6964969	A	G	0.23	0.22	0.12	0.30		BLD	4 tissues	BLD,BLD,BLD	NFKB	7 eQTL results	Pdx1,ZBTB33	451bp 3' of IKZF1	
7	50405592	0.97	1	rs6956014	T	C	0.22	0.22	0.12	0.30		BLD	4 tissues	BLD	NFKB		Myc,RFX5,SREBP	490bp 3' of IKZF1	
7	50406172	1	1	rs28462675	A	G	0.22	0.22	0.12	0.30			4 tissues	BLD,BLD				1.1kb 3' of IKZF1	
7	50407229	0.98	1	rs150935798	CA	C	0.11	0.21	0.12	0.30			4 tissues	PLCNT,THYM			BDP1,SREBP	2.1kb 3' of IKZF1	
7	50407232	0.98	1	rs200342481	GC	G	0.11	0.21	0.12	0.30			4 tissues	PLCNT,THYM			HDAC2,NF-I	2.1kb 3' of IKZF1	
7	50407623	1	1	rs10264390	T	C	0.22	0.22	0.12	0.30				BLD,THYM,BLD		7 eQTL results		2.5kb 3' of IKZF1	
7	50408133	0.99	1	rs28686237	C	G	0.18	0.22	0.12	0.30		SPLN	BLD, THYM					3kb 3' of IKZF1	
7	50409446	0.97	1	rs10230978	G	A	0.18	0.22	0.13	0.30							4 altered motifs	4.3kb 3' of IKZF1	
7	50409515	0.98	1	rs10272724	T	C	0.18	0.22	0.13	0.30				THYM,PANC		7 eQTL results	9 altered motifs	4.4kb 3' of IKZF1	
7	50409816	0.97	0.99	rs17133805	T	G	0.18	0.22	0.13	0.30		BLD	6 tissues	BLD,BLD,BLD			HNF4	4.7kb 3' of IKZF1	
7	50409913	0.98	0.99	rs62445869	G	A	0.12	0.21	0.13	0.30		BLD	6 tissues	5 tissues	TCF12		6 altered motifs	4.8kb 3' of IKZF1	
7	50409989	0.97	0.99	rs17133807	G	A	0.19	0.22	0.13	0.30		BLD	6 tissues	6 tissues	4 bound proteins	7 eQTL results	CTCF,HEN1	4.9kb 3' of IKZF1	
7	50410929	0.89	0.97	rs1110701	A	G	0.27	0.25	0.13	0.32			5 tissues	4 tissues		7 eQTL results	4 altered motifs	5.8kb 3' of IKZF1	

Chromatin Marks

The screenshot displays the SNPnexus web application interface. At the top, the 'SNPnexus' logo is on the left, and the 'Barts and The London School of Medicine and Dentistry' and 'Barts Cancer Institute Queen Mary University of London' logos are on the right. A navigation bar below the logos contains links for 'Home', 'About', 'User Guide', 'News', 'Citation', and 'Contact'. The main content area is divided into two sections: 'User details' and 'Query Options'. The 'User details' section has two optional input fields: 'Email address (optional):' and 'Dataset name (optional):'. The 'Query Options' section features a 'Query Type' dropdown menu set to 'Batch Query'. Below this, there is a 'Batch Query' section with a link to '[Input format]' and a link to '[Load Example]'. To the right of these links is a large text area labeled 'Paste in your query (upto 100K SNPs/InDels):'. Below the text area, there is a '-- OR --' separator and a prompt 'Please specify a file (upto 100K SNPs/InDels)' followed by a 'Browse...' button and the text 'No file selected.'.

SNPnexus

Barts and The London
School of Medicine and Dentistry

Barts
Cancer Institute
Queen Mary University of London

Home About User Guide News Citation Contact

User details

Email address (optional):

Dataset name (optional):

Query Options

Query Type Batch Query

Batch Query
[\[Input format\]](#) [\[Load Example\]](#)

Paste in your query (upto 100K SNPs/InDels):

-- OR --
Please specify a file (upto 100K SNPs/InDels) [Browse...](#) No file selected.

Updated in 2018

**A practical guide for the functional
annotation of genetic variations using
SNPnexus**

Abu Z. Dayem Ullah, Nicholas R. Lemoine and Claude Chelala

Chromatin Marks

SNPnexus

Barts and The London
School of Medicine and Dentistry

Barts
Cancer Institute
Queen Mary University of London

Annotation Categories

GRCh38/hg38

GRCh37/hg19

NCBI36/hg18

Gene/Protein Consequences (maximum 3 at a time)

☐ RefSeq ☒ Ensembl ☐ AceView ☐ VEGA ☐ UCSC ☐ CCDS ☐ H-Inv 7.0[†]

Effect of Non-synonymous Coding SNPs on Protein Function

☐ SIFT
☐ PolyPhen

Population Data

1000 Genomes ☐ African ☐ American ☐ East Asian ☐ European ☐ South Asian
HapMap ☐ CEU ☐ YRI ☐ CHB ☐ JPT ☐ ASW ☐ CHD ☐ GIH ☐ LWK ☐ MEX ☐
MKK ☐ TSI

Regulatory Elements

☐ Conserved Transcription Factor Binding Sites (TFBS)
☐ miRBASE 20.0
☐ Vista HMR-Conserved Non-coding Human Enhancers
☐ CpG Islands
☐ TargetScan miRNA Regulatory Sites
☐ microRNAs (miRNA Registry) / snoRNAs and scaRNAs (snoRNA-LBME-DB)
☐ ENCODE
☐ Roadmap Epigenomics
☐ Ensembl - Regulatory Build

Non-coding Scoring

☐ CADD - Combined Annotation Dependent Depletion
☐ fitCons - Fitness Consequences of Functional Annotation
☐ EIGEN - Unsupervised Spectral Approach Integrating Functional Annotation
☐ FATHMM - Functional Analysis through Hidden Markov Models
☐ GWAVA - Genome Wide Annotation of Variants
☐ DeepSEA - Chromatin Effects of Sequence Alterations
☐ FunSeq2
☐ ReMM - Regulatory Mendelian mutation

Chromatin Interactions

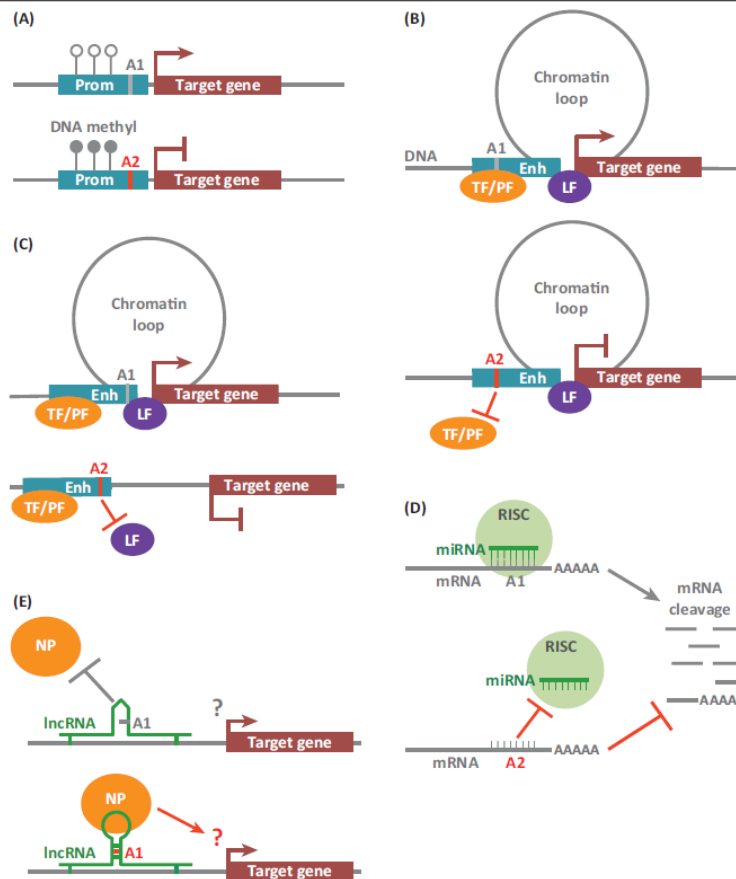


Figure 2. Non-coding genetic risk variants. (A) Genetic risk variants influence DNA methylation level at promoter regions. (B) Genetic risk variants modulate transcription factor binding to chromatin. (C) Genetic risk variants alter chromatin loop formation bridging enhancers and promoters. (D) Genetic risk variants influence the repression effect of miRNAs. (E) Genetic risk variants influence the interaction of lncRNAs with target proteins. Abbreviations: Enh, enhancer; LF, chromatin looping factor; lncRNA, long non-coding RNA; miRNA, microRNA; NP, nuclear protein; PF, pioneer factor; Prom, promoter; RISC, RNA-induced silencing complex; TF, transcription factor.

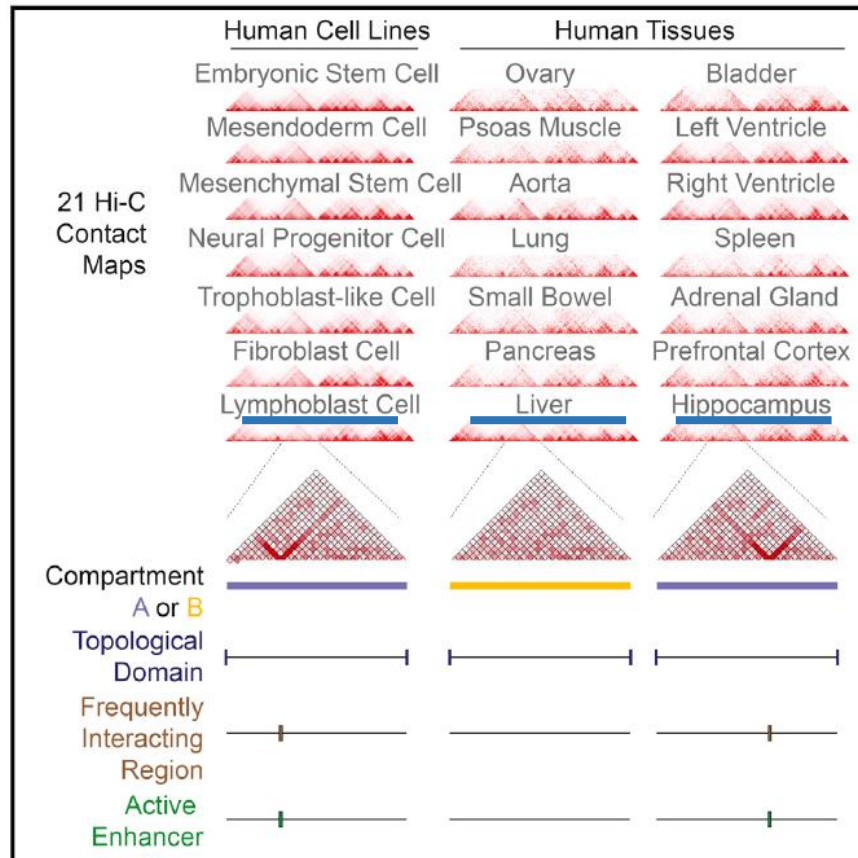
Genetic risk variants can alter chromatin loop formation bridging enhancers and promoters

The human genome is organized in a 3D architecture which is thought to regulate a diverse set of DNA-templated processes [87–91]. This allows regulatory elements, such as enhancers and promoters, to interact physically through long-range chromatin interactions, or chromatin loops, to regulate gene expression [39,42]. The human pigmentation-associated SNP, rs12913832, imposes allele-specific chromatin loop formation [92]. The rs12913832 SNP resides in an enhancer 21 kb upstream of the *OCA2* (oculocutaneous albinism II) pigment gene [92]. The T allele of this SNP favors chromatin loops to the *OCA2* gene compared to the C allele and is associated with a darker pigmentation in melanocytes [92]. Specific DNA binding proteins, including the cohesin and mediator complex as well as the insulator protein CTCF (CCCTC-binding factor), promote chromatin loop formation [93–95]. Although the rs12913832 SNP is the only genetic risk variant known to modulate chromatin loop formation, variants altering the DNA affinity for looping factors will likely also result in allele-specific chromatin loop formation (Figure 2C).

Laying a solid foundation for Manhattan – ‘setting the functional basis for the post-GWAS era’

Xiaoyang Zhang^{1*,†}, Swneke D. Bailey^{2,3,†}, and Mathieu Lupien^{2,3,4}

Chromatin Interactions



In Brief

Schmitt et al. analyze Hi-C maps in 21 human cell lines and primary tissues and uncover a class of genome organizational features termed FIREs. FIREs are local interaction hotspots, highly tissue-specific, and correspond to active enhancers. We discuss the implications of our findings for the study of gene regulation and disease. Explore the Cell Press IHEC web portal at <http://www.cell.com/consortium/IHEC>.

Highlights

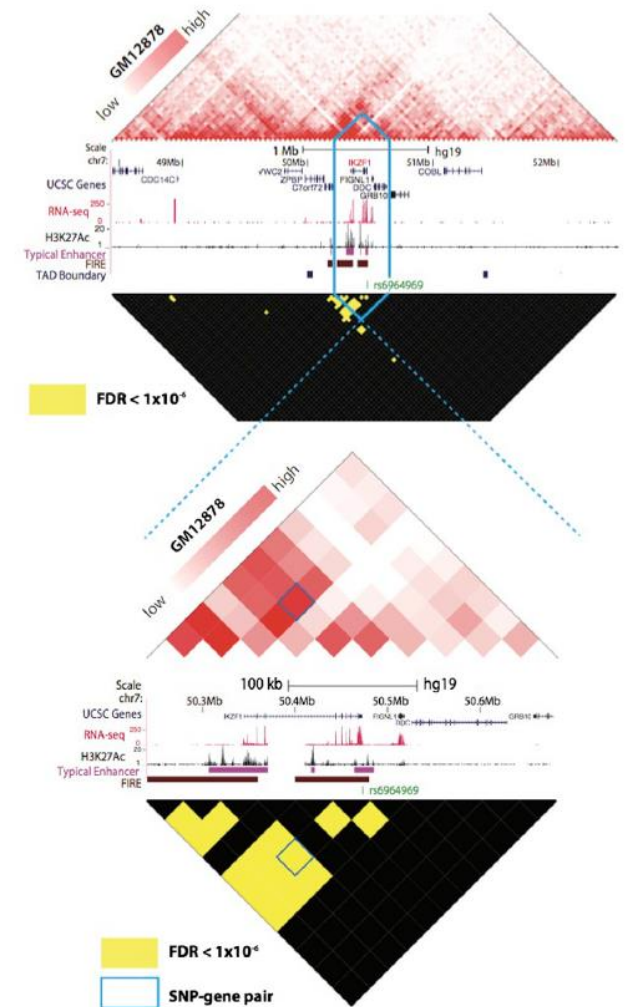
- Integrative analysis of chromatin architecture in a broad set of human tissues
- FIREs are an architectural feature of chromatin organization
- FIREs are enriched for super-enhancers and show tissue-specific chromatin interactions
- FIRE formation is partially dependent on CTCF and the Cohesin complex

Chromatin Interactions

Disease Associations

FIREs Are Enriched for Disease-Associated SNPs

Our analyses have indicated that FIREs are enriched for active enhancers and super-enhancers (Figures 4A–4D; Figures S3B, S3C, S3F, S3G, S3N, and S3O). Because typical and super-enhancers contain a significant proportion of disease-associated SNPs (Hnisz et al., 2013), we further investigated the overlap between FIREs and disease-associated SNPs. First, we mapped 4,327 previously annotated disease-associated non-coding SNPs to FIREs defined in each cell line and tissue (see Supplemental Experimental Procedures) (Hnisz et al., 2013). Consistent with previous results (Hnisz et al., 2013), we observed 7.06 and 3.76 SNPs per megabase, and among 354 GM12878 FIREs overlapped with super-enhancers and 2,800 GM12878 FIREs overlapped with typical enhancers, respectively (Figure S5A). Surprisingly, among 1,615 GM12878 FIREs that do *not* overlap an annotated enhancer, we also observed 3.33 SNPs per megabase, which is ~2.3-fold higher than the genome-wide SNP density (1.42 SNPs per megabase) (Figure S5A). Importantly, these SNPs would not be captured by directly overlapping super-enhancers or typical enhancers with disease-associated SNPs (Hnisz et al., 2013).



A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome

Anthony D. Schmitt,^{1,2,12,13} Ming Hu,^{3,12,14,*} Inkyung Jung,^{1,15} Zheng Xu,^{4,10,11} Yunjiang Qiu,^{1,5} Catherine L. Tan,^{1,13} Yun Li,⁴ Shin Lin,⁶ Ying Lin,⁷ Cathy L. Barr,⁸ and Bing Ren^{1,9,16,*}

Chromatin Interactions

GWAS3D
Detecting Human Regulatory Variants by Combining Genome Wide Association Study and Chromosome Conformation

Introduction | GWAS3D | Help | About | Update!!

GWAS3D

news: 06/10/2013 Improve HiC resolution to 10Kb for GM12878, K562, H1-hESC and RWPE1 cell types.

news: 30/05/2013 External browser now can be connected to HaploReg v2, Regulomedb, SNVrap and UCSC ENCODE Browser for candidate regulatory variants.

news: 09/04/2013 Feature enhancement: 1) Support adjusting the P-value for scanning putative TF binding sites; 2) Support the analysis without cell-type restriction.

My Jobs

Welcome to the gateway of GWAS3D. Interpreting noncoding phenotypically associated variants is an indispensable step to understand molecular mechanism of complex traits, 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. GWAS3D also provided comprehensive annotations and visualizations to help users interpreting the results. Please check detailed information on online help.

Main Functions

- Identify the most probable functional variants which affect transcriptional regulation;
- Prioritize the leading variants when given a full list of GWAS result;
- Evaluate the deleteriousness of genetic variants affecting the gene regulation when given a list of variants;
- Annotate genetic variant from regulatory perspective.

Please cite the work from:
GWAS3D: detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications *Nucleic Acids Research*. doi:10.1093/nar/gkt456.

rs3129871	1.15e-299	0.96	
rs3129941	1.37e-197	0.871	
rs3129860	8.62e-192	0.864	
rs2075650	1.75e-157	0.862	
rs6903608	1.55e-144	0.803	
rs2647044	5.18e-142	1	

[GWAS3D Web Server](#)

Chromatin Interactions



**3D-genome Interaction Viewer
and
database**

[TUTORIAL](#) [STATISTICS](#) [CONTACT US](#)

3D Interaction Viewer and Database (3DIV) is an easy-to-use database providing epigenetic annotation, one-to-all chromatin interaction visualization in a variety of options. Spatial chromatin interaction has a significant role in gene regulation, and visualization and annotation of these interactions is key to connect between genetic/epigenetic variation and phenotypical polymorphism. For example, enhancer-promoter interaction is essential to regulate target gene expression. 3DIV provides following analysis: Interaction identification and annotation, interactive interaction visualization, and comparative interaction visualization. For more details, please click [here](#).

Interaction Table

Interaction visualization

Comparative interaction visualization

Choose Sample(s)

☐ Adrenal gland

☐ Aorta

☐ Bladder

More Experiment ▼

Bait :

(Ex. CROCCP2, chr22:27141000, rs42)

Add sample(s)

Remove sample(s)

Items selected

☐

Sample

Bait

Interaction range

2Mb

▼

Example Run

Run

SNP Functionality Scores

Table 1. Online Resources for Accessing Noncoding SNP Annotation Tools*

Tool	URL	Refs
VEP	VEP incorporates annotations from the Ensembl database, allowing it to make predictions genome-wide as well as predict tissue-specific activity for 13 human cell lines. http://www.ensembl.org/info/docs/tools/vep/script/index.html	McLaren et al. [17]
RegulomeDB	RegulomeDB uses a heuristic scoring system to catalog the likelihood that a given SNP or indel resides in a functional region, using functional data from over 100 cell types. http://regulomedb.org	Boyle et al. [18]
FuncSNP	FuncSNP is an R/Bioconductor package that employs user input annotations to prioritize SNPs, allowing users to customize their annotations to query a cell type of interest. http://www.bioconductor.org/packages/release/bioc/html/FuncSNP.html	Cootner et al. [19]
ANNOVAR	ANNOVAR is a command line tool that uses region-based annotations to annotate noncoding variants and insertions and deletions (indels), in addition to comparing them to known variation databases. http://annovar.openbioinformatics.org	Wang et al. [21]
HaploReg	HaploReg is a searchable repository for SNPs and indels from the 1000 Genomes Project, providing a summary of known annotations for variants within an LD block. http://www.broadinstitute.org/mammals/haploreg/haploreg.php	Ward and Kelle [22]
GWAS3D	GWAS3D evaluates SNPs and indels by analyzing their 3D chromosomal interactions and disruptions to TF binding affinity. It outputs scores as well as a drde plot mapping local 3D interactions. http://jwanglab.org/gwas3d	Li et al. [23]
fitCons	fitCons uses the INSIGHT method to predict the probability that SNPs will influence fitness by screening for signatures positive and negative selection using data from three cell types. http://comp.gen.bs.cornell.edu/fitCons/	Gulko et al. [24]
GWAVA	GWAVA trains on a random forest algorithm using disease mutations from HGMD and control variants from the 1000 genomes project to predict if queried variants are functional. http://ftp.sanger.ac.uk/pub/resources/software/gwava/	Ritchie et al. [25]
CADD	CADD trains on a linear kernel support vector machine using simulated variants as deleterious variants and alleles fixed between human and chimpanzee as control variants. http://cadd.gs.washington.edu	Kircher et al. [26]
DANN	DANN trains on a nonlinear learning neural network algorithm using the same training set data (fixed alleles vs. simulated variants) as CADD. https://dbclips.ud.edu/public_data/DANN/	Quang et al. [29]
FATHMM-MKL	FATHMM-MKL implements a kernel-based classifier to estimate complex nonlinear patterns using HGMD pathogenic and 1000 Genomes Project control variant training set data. http://fathmm.biocompute.org.uk	Shihab et al. [30]
deltaSVM	deltaSVM uses a gapped k-mer support vector machine to estimate the effect of a variant in a cell-type-specific manner. http://www.boerslab.org/deltaSVM/	Lee et al. [31]
DeepSEA	DeepSEA uses a multilayered hierarchical structured deep learning-based sequence model to predict functional SNPs with single nucleotide sensitivity using ENCODE and Roadmap Epigenomics data. http://deepsea.princeton.edu/job/analysis/create/	Zhou and Troyanskaya [32]

* Abbreviation: HGMD, Human Gene Mutation Database.

Mining the Unknown:
Assigning Function to
Noncoding Single Nucleotide
Polymorphisms

Sierra S. Nishizaki¹ and Alan P. Boyle^{1,2,*}

SNP Functionality Scores

Aggregate Functionality Scores

fathmm-XF Home About FATHMM

Enter Your Job/Session Identifier Fetch Results

FATHMM-XF: Enhanced Accuracy in Predicting the Functional Consequences of Non-Coding and Coding Single Nucleotide Variants (SNVs)

New Submission Use Example Help/Documentation

BioFolD Home Job Method Help

PhD-SNP^g

Predicting human Deleterious SNPs in human genome
A binary classifier for predicting pathogenic variants in coding and non-coding regions.

Insert the list of SNVs using the comma separated values format chr,pos,ref,alt or gene,mut (maximum 1,000 variants).

 **PREDICTSNP²** Unified platform for prediction of SNP effect in distinct genomic regions 

Home Help Job ID: Find job

INPUT Load example data Show example results

Genome assembly: Format:

Insert the list of variants in one of the supported formats:

18,21118528,G,C

Load file: Browse... No file selected.

REFERENCE

Bendi, J., Musil, M., Stourac, J., Zedduka, J., Demborsky, J., Brezovsky, J., 2016. PredictSNP2: A unified platform for accurately evaluating SNP effects by exploiting the different characteristics of variants in distinct genomic regions. PLOS Computational Biology 12: e1004962.

USER STATISTICS

• Number of visitors: 12853

SNP Functionality Scores

Aggregate Functionality Scores SNPDeIScore

 U.S. National Library of Medicine
National Center for Biotechnology Information

Log in

HOME GWAS TFBSS METHODS TISSUES WEB SERVICES DOCUMENTATION RAW DATA

SNPDeIScore displays pre-compute deleterious effects of noncoding variants using a large panel of currently available methods and summarizes this information in an interactive, easy to use website. It also provides open access to these data through a [RESTfull based web service](#) and a [Python based web services command line client](#) to retrieve the data from other applications and tools.

Locus ID

SNP (rs62635286), gene name (DDX11L1) or genomic coordinates (chr8:11532494-11621567)

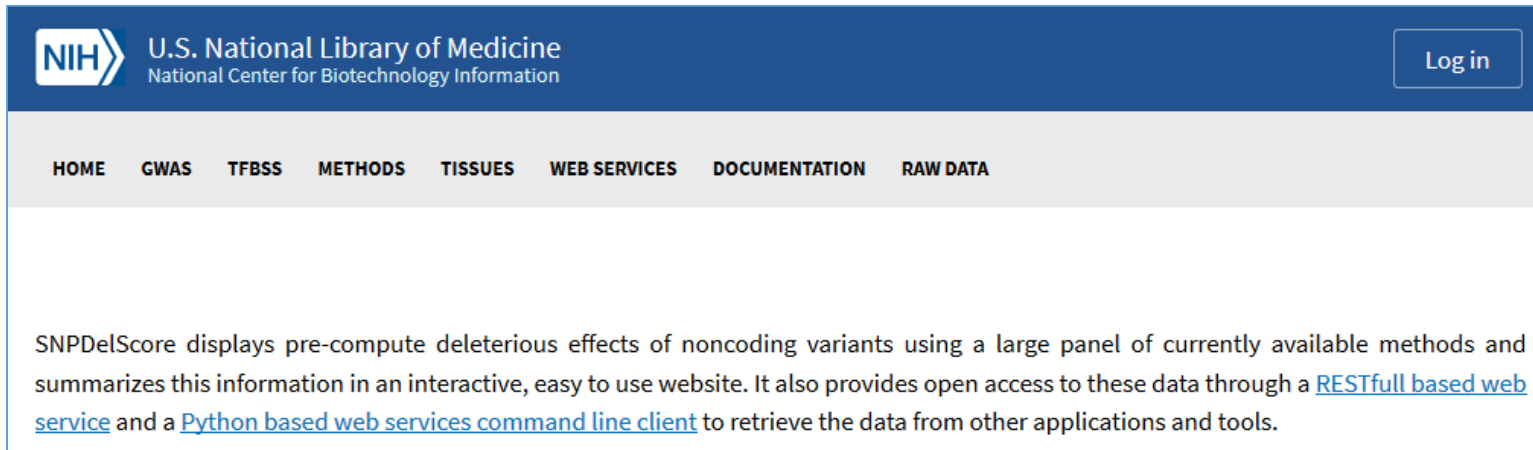
Tissue

Average ▼

Submit

SNP Functionality Scores

Aggregate Functionality Scores SNPDeIScore



The screenshot shows the top section of the SNPDeIScore website. At the top left is the NIH logo (U.S. National Library of Medicine, National Center for Biotechnology Information). To the right is a 'Log in' button. Below this is a navigation bar with links: HOME, GWAS, TFBSS, METHODS, TISSUES, WEB SERVICES, DOCUMENTATION, and RAW DATA. The main content area below the navigation bar contains a paragraph describing the tool: 'SNPDeIScore displays pre-compute deleterious effects of noncoding variants using a large panel of currently available methods and summarizes this information in an interactive, easy to use website. It also provides open access to these data through a [RESTfull based web service](#) and a [Python based web services command line client](#) to retrieve the data from other applications and tools.'

<https://www.ncbi.nlm.nih.gov/research/snpdelscore/snp/rs2395185>

From SNPs to Target Genes

eQTL targets

meQTLs targets

Missense SNPs/splice variants: the same gene

Chromatin looping contacts (promoters)

Gene Set Enrichment Analysis

Once a target gene list is generated, it should be examined for enrichment for certain characteristics.



Human Molecular Genetics, 2016, Vol. 25, No. R2 R133–R140

doi: 10.1093/hmg/ddw249

Advance Access Publication Date: 10 August 2016

Invited Review

INVITED REVIEW

Gene set analysis for interpreting genetic studies

Tune H. Pers^{1,2,*}

Gene Set Enrichment Analysis

 **Enrichr**

[Login](#) | [Register](#)

9,270,029 lists analyzed
234,849 terms
128 libraries

Analyze | [What's New?](#) | [Libraries](#) | [Find a Gene](#) | [About](#) | [Help](#)

Input data

Choose an input file to upload. Either in BED format or a list of genes. For a quantitative set, add a comma and the level of membership of that gene. The membership level is a number between 0.0 and 1.0 to represent a weight for each gene, where the weight of 0.0 will completely discard the gene from the enrichment analysis and the weight of 1.0 is the maximum.

Try an example [BED file](#).

No file selected.

Or paste in a list of gene symbols optionally followed by a comma and levels of membership. Try two examples: [crisp set example](#), [fuzzy set example](#)

0 gene(s) entered

Enter a brief description for the list in case you want to share it. (Optional)

Enrichr: a comprehensive gene set enrichment analysis web server 2016 update

Maxim V. Kuleshov¹, Matthew R. Jones¹, Andrew D. Rouillard¹, Nicolas F. Fernandez¹, Qiaonan Duan¹, Zichen Wang¹, Simon Koplev¹, Sherry L. Jenkins¹, Kathleen M. Jagodnik², Alexander Lachmann¹, Michael G. McDermott¹, Caroline D. Monteiro¹, Gregory W. Gundersen¹ and Avi Ma'ayan^{1,*}

Gene Set Enrichment Analysis

WebGIVI^{tool}

Visualization of Gene and iTerm

Home

eGIFT

Tutorial

Contact

WebGIVI can accept a gene list that will be used to retrieve a gene symbol and iTerm list. This list can be resubmitted to visualize the gene-iTerm pairs using Cytoscape or Concept Map. The resulting list can be edited to remove iTerms that are not of interest to the user. Visualized Graphs can also be saved as PNG format. (NOTE: You may not be able to visualize more than 1000 node-item pairs at one time in Cytoscape due to the limitation of cytoscape.js. However, Concept Map can visualize big data!).

Data Visualization:

eGIFT Gene Enrichment Analysis

Please input your gene IDs(at least 2 IDs) OR gene symbols OR gene item pairs below

Samples:

Entrez Uniprot Ensembl Gene symbol Gene-Item

Input type:

☐ Entrez ID ☐ UniProt ID ☐ Ensembl ID ☐ Gene symbols ☐ Symbol Item Pair

Custom Data

Users can visualize their two-column tab delimited data (e.g. gene pathway pair data)

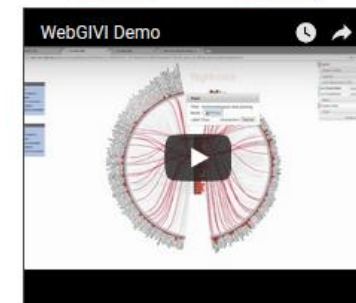
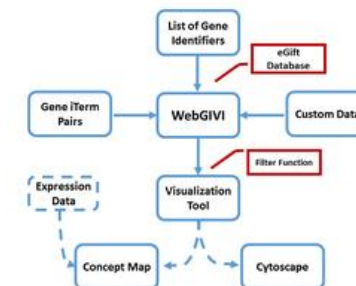
Sample:

☐ Custom Data

Name Your input List (optional)

Database you want to choose: ☒ eGIFT ☐ Amigo2

Organism (only for Amigo2): (Please note all provided sample IDs are from organism Gallus gallus)



Sun et al. BMC Bioinformatics (2017) 18:237
DOI 10.1186/s12859-017-1664-2

BMC Bioinformatics

SOFTWARE

Open Access

WebGIVI: a web-based gene enrichment analysis and visualization tool

Liang Sun^{1,6}, Yongnan Zhu^{2,3}, A. S. M. Ashique Mahmood⁴, Catalina O. Tudor⁴, Jia Ren⁵, K. Vijay-Shanker⁴, Jian Chen² and Carl J. Schmidt^{1*}



Gene Networks

GeneNetwork

University of Tennessee: www.genenetwork.org

Use GeneNetwork 2



WebQTL

[Home](#) | [Search](#) | [Help](#) | [News](#) | [References](#) | [Policies](#) | [Links](#)

Welcome! [Login](#)

Select and Search

Species:

Group: [Info](#)

Type:

Data Set: [Info](#)

Databases marked with ** suffix are not public yet.
Access requires [user login](#).

Get Any:

Enter terms, genes, ID numbers in the **Get Any** field.
Use * or ? wildcards (Cyp*a?, synap*).
Use **Combined** for terms such as *tyrosine kinase*.

Combined:

[Search](#)

[Make Default](#)

[Advanced Search](#)

Quick HELP Examples and [User's Guide](#)

You can also use advanced commands. Copy these simple examples

Websites Affiliated with GeneNetwork

- GeneNetwork [Time Machine](#): Full versions from 2009 to 2016 (mm9)
- UTHSC Genome Browser [Classic](#) and [Newest](#)
- UTHSC [Galaxy](#) Service
- UTHSC [Bayesian Network Web Server](#)
- GeneNetwork Classic on [Amazon Cloud](#)
- GeneNetwork Classic Code on [GitHub](#)
- GeneNetwork 2.0 Development Code on [GitHub](#)
- [GeneNetwork 2.0](#) Development

Getting Started

1. Select **Species** (or select All)
2. Select **Group** (a specific sample)
3. Select **Type** of data:
 - Phenotype (traits)
 - Genotype (markers)
 - Expression (mRNAs)
4. Select a **Database**
5. Enter search terms in the **Get Any** or **Combined** field: words, genes, ID numbers, probes, advanced search commands
6. Click on the **Search** button
7. Optional: Use the **Make Default** button to save your preferences

How to Use GeneNetwork

Gene Networks

GeneNetwork

University of Tennessee: www.genenetwork.org

Use GeneNetwork 2



WebQTL

Search Results

Details and Links

GeneNetwork searched the **GTExv5 Human Brain Frontal Cortex BA9 RefSeq (Sep15) RPKM log2 Database** for all records that match the term *HLA-E*. GeneNetwork found a total of **72** records.

Records

To add a group of **Record IDs** to your Trait Collection, use the **Index** checkboxes and click the **Add** button. To analyze any single record click on its **Record ID**.



Select Deselect Invert Add

[Download Table](#)

Index	Record ID	Symbol	Description	Location Chr and Mb	Mean Expr	Max LRS ?	Max LRS Location Chr and Mb	Add ?
39	☐ ENSG00000223534.1	HLA-DQB1-AS1	HLA-DQB1-AS1	Chr6: 32.628132	0.193	N/A	N/A	
40	☐ ENSG00000232629.4	HLA-DQB2	major histocompatibility complex, class II, DQ beta 2	Chr6: 32.723875	0.108	N/A	N/A	
41	☐ ENSG00000226030.1	HLA-DQB3	major histocompatibility complex, class II, DQ beta 3	Chr6: 32.698535	0.000	N/A	N/A	
42	☐ ENSG00000204287	HLA-DRA	major histocompatibility complex, 2C class II, 2C DR alpha	Chr6: 32.439842	3.757	N/A	N/A	
43	☐ ENSG00000196126.6	HLA-DRB1	major histocompatibility complex, class II, DR beta 1	Chr6: 32.546546	2.537	N/A	N/A	
44	☐ ENSG00000198502.5	HLA-DRB5	major histocompatibility complex, class II, DR beta 5	Chr6: 32.485120	1.054	N/A	N/A	
45	☐ ENSG00000229391.3	HLA-DRB6	major histocompatibility complex, class II, DR beta 6 (pseudogene)	Chr6: 32.520490	0.235	N/A	N/A	
46	☐ ENSG00000196301.3	HLA-DRB9	major histocompatibility complex, class II, DR beta 9 (pseudogene)	Chr6: 32.427598	0.016	N/A	N/A	
47	☐ ENSG00000204592.5	HLA-E	major histocompatibility complex, class I, E	Chr6: 30.457244	5.938	N/A	N/A	
48	☐ ENSG00000204642	HLA-F	major histocompatibility complex, 2C class I, 2C F	Chr6: 29.722775	1.876	N/A	N/A	
49	☐ ENSG00000214922	HLA-F-AS1	HLA-F antisense RNA 1	Chr6: 29.726601	0.344	N/A	N/A	
50	☐ ENSG00000204632.7	HLA-G	major histocompatibility complex, class I, G	Chr6: 29.794744	0.093	N/A	N/A	
51	☐ ENSG00000206341.6	HLA-H	major histocompatibility complex, class I, H (pseudogene)	Chr6: 29.856244	1.075	N/A	N/A	
52	☐ ENSG00000204622.6	HLA-J	major histocompatibility complex, class I, J (pseudogene)	Chr6: 29.975966	0.720	N/A	N/A	

Gene Networks

GeneNetwork

University of Tennessee: www.genenetwork.org

Use GeneNetwork 2



WebQTL

Trait Data and Analysis for Record ID ENSG00000204592.5

Details and Links

Gene Symbol: *HLA-E*
Aliases: *EA1.2; EA2.1; HLA-6.2; MHC; QA1*
Description: major histocompatibility complex, class I, E
Location: Chr 6 @ 30.457244 Mb on the plus strand
Target Score:
Species and Group: Human, GTEx_V5
Database 3: [GTExv5 Human Brain Frontal Cortex BA9 RefSeq \(Sep15\) RPKM log2](#)
Resource Links: [Gene](#) [OMIM](#) [UniGene](#) [HomoloGene](#)
[BioGPS](#) [STRING](#) [PANTHER](#) [Gemma](#) [ABA](#) [EBI GWAS](#) [Wiki-Pi](#)



Add



Find



Verify



GeneWiki

+ Basic Statistics

- Calculate Correlations

Sample r **Literature r** **Tissue r**

Database: [GTExv5 Human Brain Frontal Cortex BA9 RefSeq \(Sep15\) RPKM log2](#)

Return: [top 500](#)

Pearson ☐ Spearman Rank ☒

Compute

Gene Networks

GeneNetwork

University of Tennessee: www.genenetwork.org

Use GeneNetwork 2



WebQTL



Correlation Table

Values of Record ENSG00000204592.5 in the [GTExv5 Human Brain Frontal Cortex BA9 RefSeq \(Sep15\) RPKM log2](#) database were compared to all 56318 records in the [GTExv5 Human Brain Frontal Cortex BA9 RefSeq \(Sep15\) RPKM log2](#) database. The top 500 correlations ranked by the Genetic Correlation (Spearman's rho) are displayed. You can resort this list using the small arrowheads in the top row. Click the correlation values to generate scatter plots. Select the Record ID to open the Trait Data and Analysis form. Select the symbol to open NCBI Entrez.



Select



Deselect



Invert



Add



Gene Weaver



GCAT



Gene Set



Graph



Matrix



Partial



Compare



QTL Map



Heat Map

[More Options](#)

[Download Table](#)

Index	Record ID	Symbol	Description	Location Chr and Mb	Mean Expr	Max LRS	Max LRS Location Chr and Mb	Sample rho	N Cases	Sample p(rho)	Lit Corr	Tissue rho	Tissue p(rho)	GN Lin
1	☐ ENSG00000204592.5	HLA-E	major histocompatibility complex, class I, E	Chr6: 30.457244	5.938	--	--	1.000	115	0.00e+00	--	--	--	🔗
2	☐ ENSG00000158710.10	TAGLN2	transgelin 2	Chr1: 159.887897	4.068	--	--	0.880	115	0.00e+00	--	--	--	🔗
3	☐ ENSG00000213719.4	CLIC1	chloride intracellular channel 1	Chr6: 31.699994	2.883	--	--	0.876	115	0.00e+00	--	--	--	🔗
4	☐ ENSG00000067066	SP100	SP100 nuclear antigen	Chr2: 230.415942	1.077	--	--	0.866	115	0.00e+00	--	--	--	🔗
5	☐ ENSG00000179776	CDH5	cadherin 5, 2C type 2 (vascular endothelium)	Chr16: 66.366622	2.077	--	--	0.866	115	0.00e+00	--	--	--	🔗
6	☐ ENSG00000122862.4	SRGN	serglycin	Chr10: 70.847862	4.334	--	--	0.855	115	0.00e+00	--	--	--	🔗
7	☐ ENSG00000173193	PARP14	poly (ADP-ribose) polymerase family, 2C member 14	Chr3: 122.680618	1.666	--	--	0.852	115	0.00e+00	--	--	--	🔗
8	☐ ENSG00000140853	NLRC5	NLR family, 2C CARD domain containing 5	Chr16: 56.989485	0.724	--	--	0.852	115	0.00e+00	--	--	--	🔗
9	☐ ENSG00000196329	GIMAP5	GTPase, 2C IMAP family member 5	Chr7: 150.722253	1.824	--	--	0.850	115	0.00e+00	--	--	--	🔗
10	☐ ENSG00000086062	B4GALT1	UDP-Gal:betaGlcNAc beta 1, 2C4- galactosyltransferase, 2C polypeptide 1	Chr9: 33.104082	1.219	--	--	0.848	115	0.00e+00	--	--	--	🔗
11	☐ ENSG00000130158	DOCK6	dedicator of cytokinesis 6	Chr19: 11.199295	1.899	--	--	0.848	115	0.00e+00	--	--	--	🔗

sopSNP (beta)

[SEARCH](#)[HEPSİNİ KALDIR](#)[HEPSİNİ SEÇ](#)

		Tercih Ediliyor	Seçim
	NCBI Global Search	61.7%	☒
hlsnps	HLAsnps	54.7%	☒
	ENTREZ SNP	53.1%	☒
	RefSNP	51.4%	☒
	Ensembl	49%	☒
	SNiPA	48.1%	☒
	PheGenI	44.9%	☒
	HaploReg v4	40.7%	☒

HLASNP_s (beta)



Home Page

About

Statistics

sopSNP App

Search HLA_{snp}s:

You can search multiple SNPs with new lines between SNPs or separated with semicommas!

rs2395185

Search

HLAsnp_s (DB Version 1.1)

Functional assessment of HLA region SNPs

• rs2395185

Master DB Results:

CHR :	6
pos (hg38):	32465390
ALLELES :	G / T
GENE :	

• GWAS Results (1)

• PheWAS Results (7)

• meQTL Results (4)

... Looking forward

Day 1

Bioinformatics tools for the assessment of genetic variants

**Bioinformatic analysis of experimentally verified functional
SNP associations with disease**

**Bioinformatic tools for the investigation of tissue specificity of
variant effects**

Online tools: Standard operating procedures for gene and variant analysis

Applications and practice on participant projects