III. 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
Sekreterya: tibbibiyolojiad@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

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 Keleştemur
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 Profilleme: 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. Ahmet Okay Çağlayan
Bilim Üniversitesi Tıp Fak., Florence Nightingale Hastanesi, PHI Tech Bioinformatics

Yrd. Doç. Dr. Kaya Bilguvar
Yale Üniversitesi, Genom Analiz Merkezi Direktörü

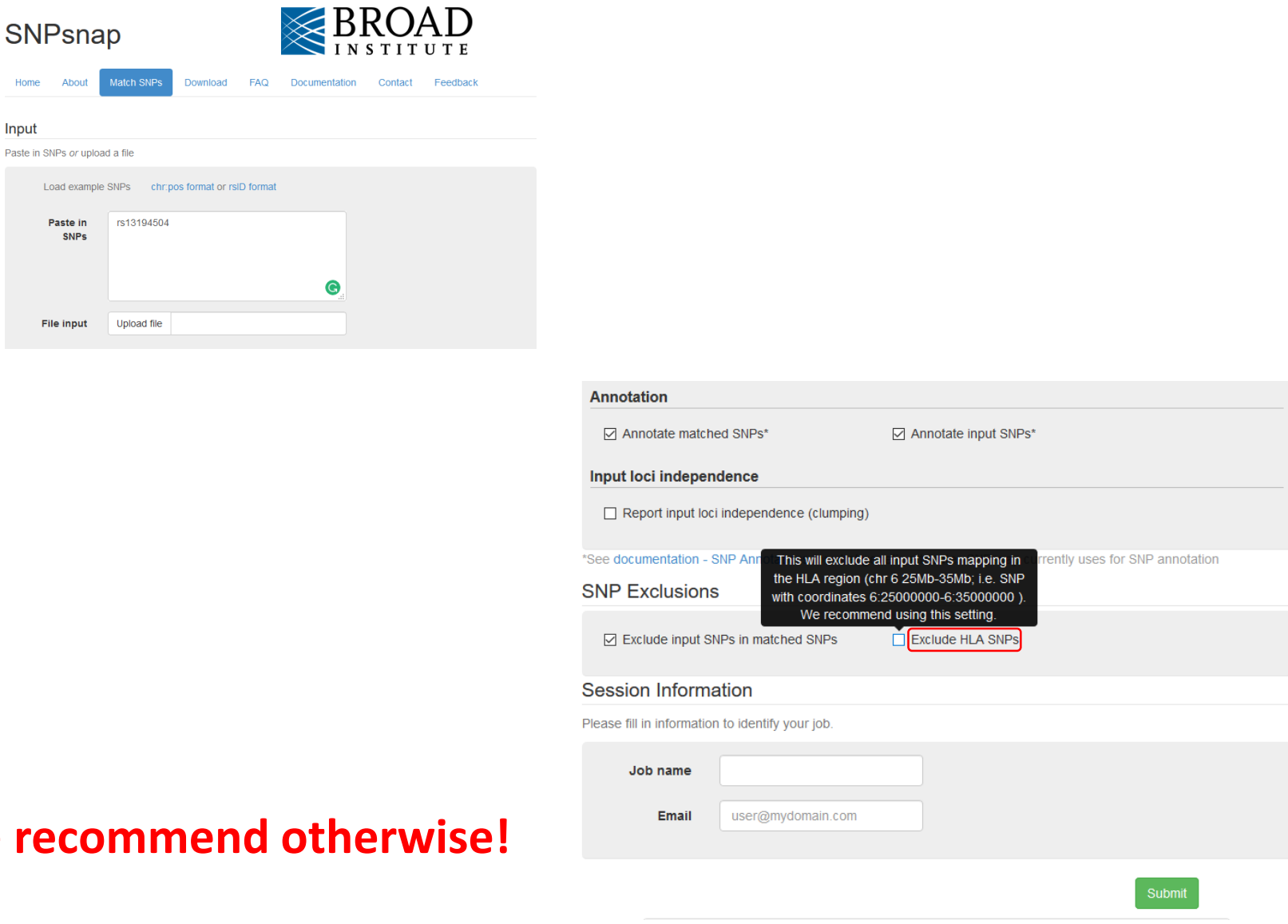
Dr. Zafer Yüksel
Centogene AG, Tıbbi Raporlama ve Tıbbi Biyoinformatik Direktörü - Tıbbi Genetik Uzmanı

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



We recommend otherwise!

HLA SNPs

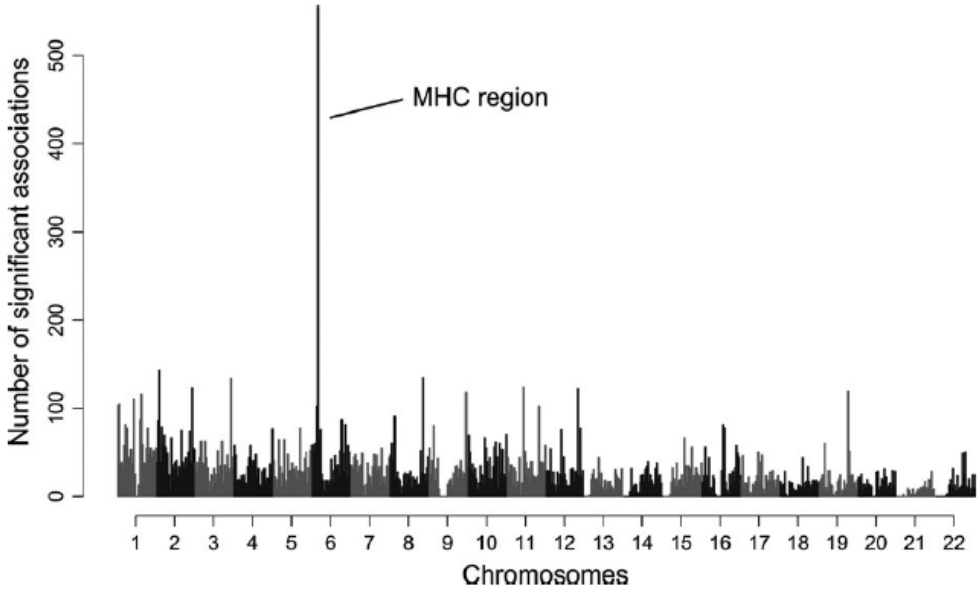


FIGURE 1 Number of significant GWAS associations along the genome. The chromosomal location of significant trait associations from GWAS (*N* = 18,682) is shown for all autosomes. Data from NHGRI GWAS catalog. Reproduced from "Lenz TL, Spirin V, Jordan DM, Sunyaev SR. Excess of Deleterious Mutations around 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/iji.12332

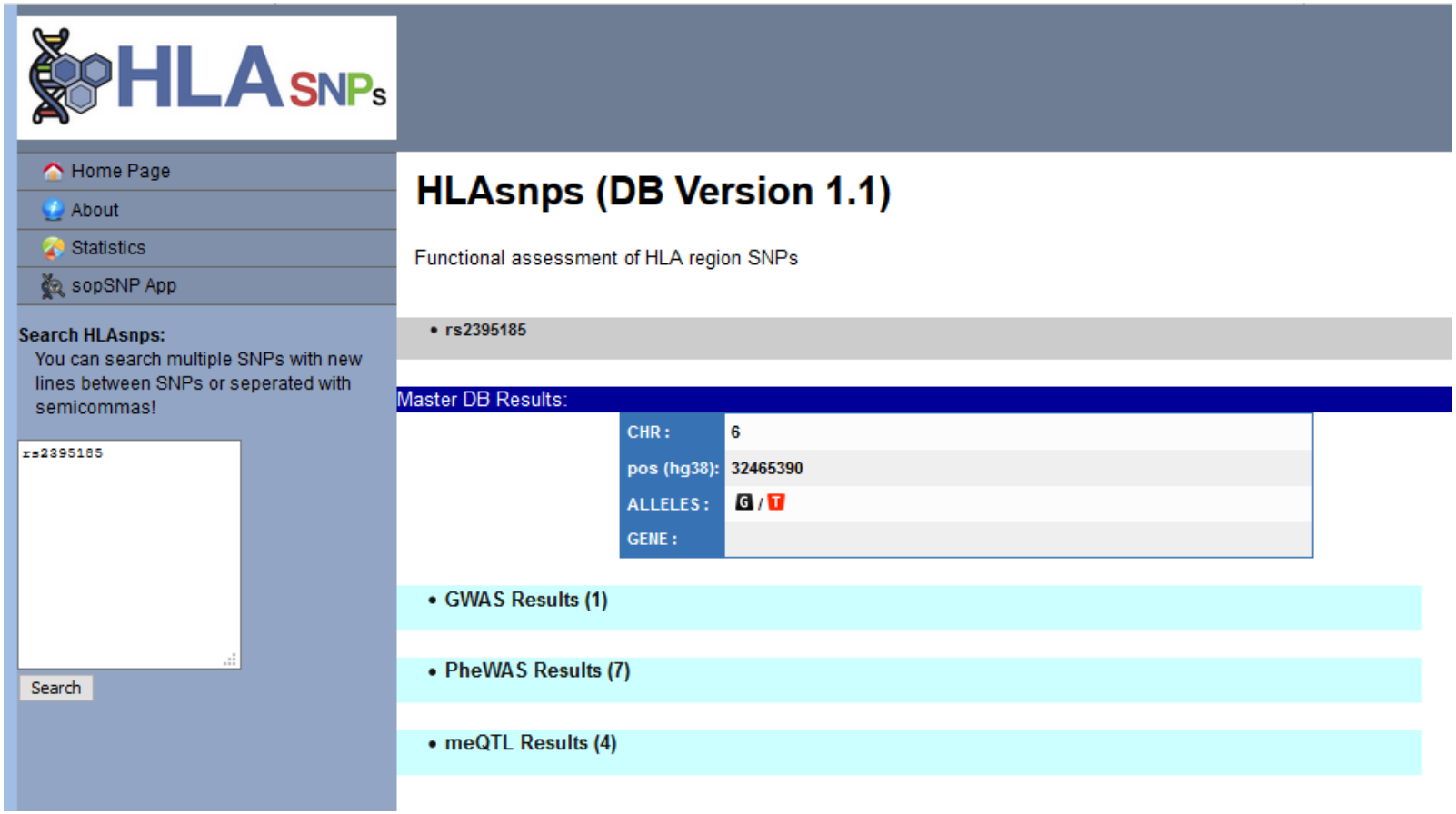
REVIEW

WILEY INTERNATIONAL JOURNAL OF IMMUNOGENETICS

What has GWAS done for HLA and disease associations?

A. E. Kennedy[1] | U. Ozbek[2,3] | M. T. Dorak[4]

HLASNPs (beta)



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

Kingston University London

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

Kingston University London

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?

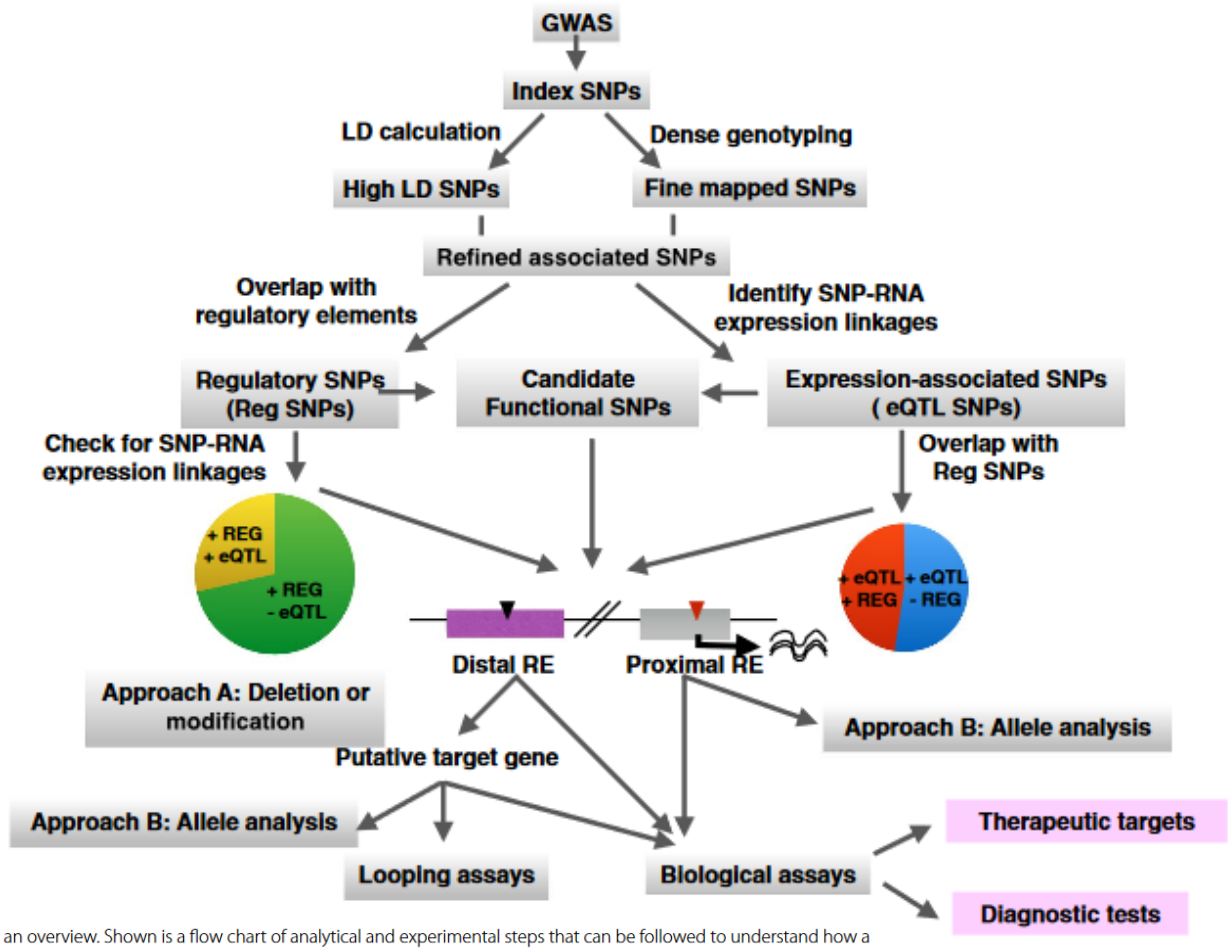


Fig. 1 Making sense of GWAS: an overview. Shown is a flow chart of analytical and experimental steps that can be followed to understand how a non-coding SNP can be associated with an increased risk for a specific disease. Index SNPs are identified using GWAS arrays and then expanded to a larger set of 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

Kingston University London

Tak and Farnham *Epigenetics & Chromatin* (2015) 8:57
DOI 10.1186/s13072-015-0050-4

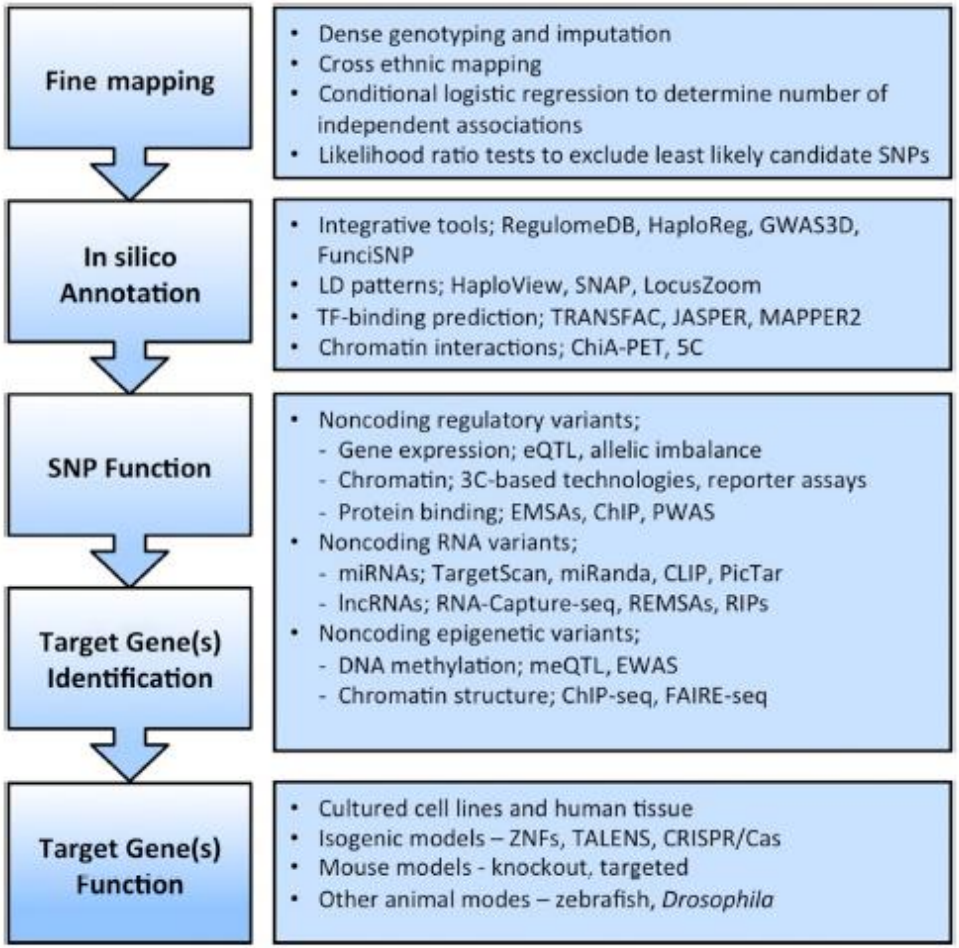
Epigenetics & Chromatin

REVIEW Open Access

Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome

Yu Gyoung Tak and Peggy J. Farnham*

How to Make Sense of GWAS?



I. Workflow for Functionally Analyzing and Interpreting GWAS Loci

Kingston University London

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, DNaseI hypersensitive site sequencing; MeDIP-seq, methylated DNA immunoprecipitation sequencing; MRE-seq, methylation-sensitive restriction enzyme sequencing; and RNA-PET, RNA paired-end-tag sequencing.
[a]Nonexhaustive list of examples.

Kingston University London

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?

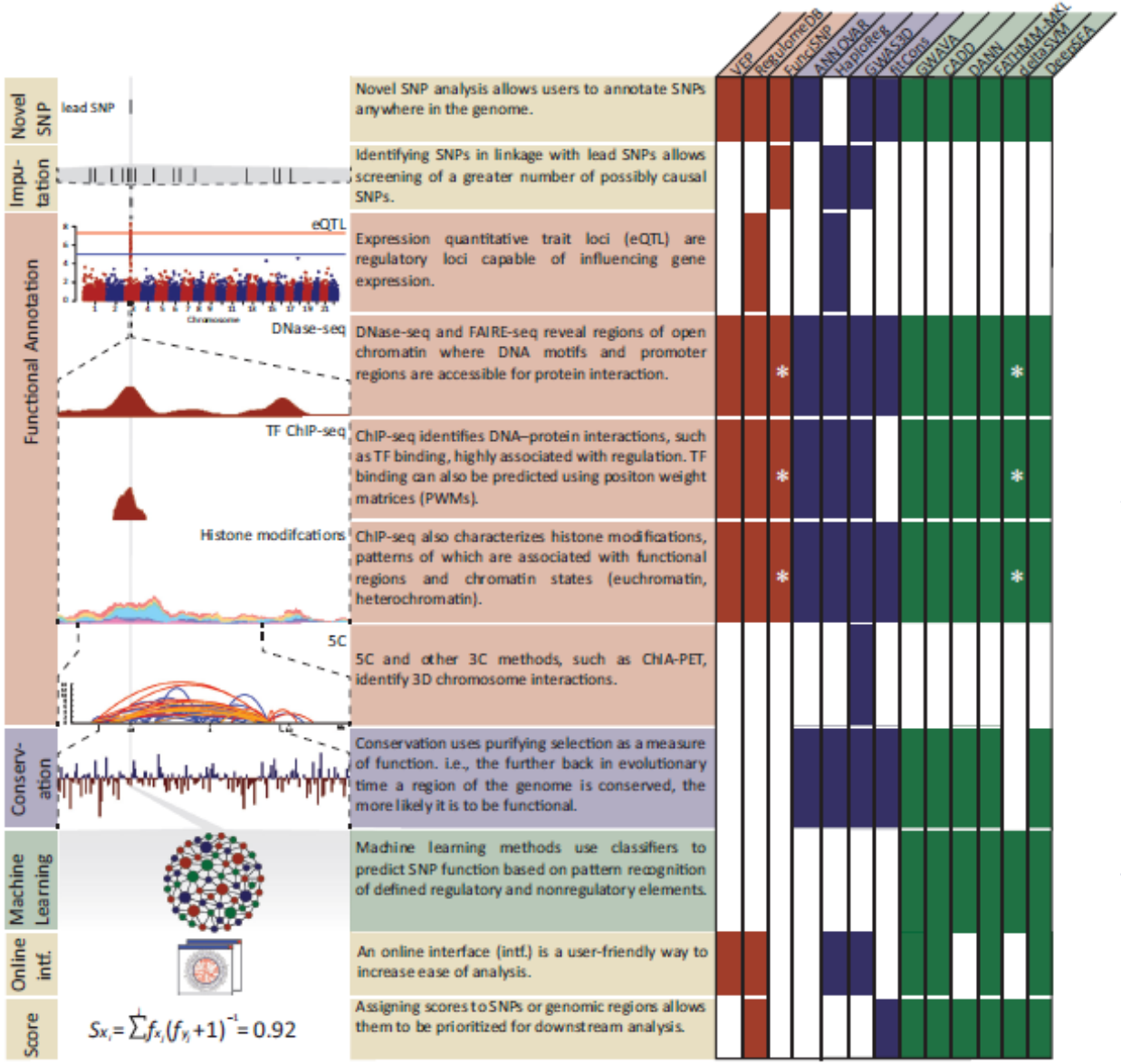
Genetic Study of Complex Diseases in the Post-GWAS Era

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	*SOX9* *SOX9*	Zhang et al., 2012
Prostate cancer	rs12653946		*IRX4*	Nguyen et al., 2012
Colorectal cancer	rs6983267	TCF7L2	*MYC*	Wasserman et al, 2010
Colorectal cancer	rs6983267	TCF4	*MYC*	Wright et al, 2010
Colorectal cancer	rs16888589	STAT	*EIF3H*	Pittman et al., 2010
Breast cancer	rs10069690	Splicing	*TRET*	Bojesen et al., 2013
Breast cancer	rs4784227	FOXA1	*TOX3*	Cowper-Sal lari et al., 2012
Breast cancer	rs4442975	FOXA1	*IGFBP5*	Ghoussaini et al. 2014
Breast cancer	rs554219	ELK4	*CCND1*	French et al., 2013
Breast cancer	rs7895676	Oct-1/Runx2	*FGFR2*	Meyer et al., 2008
Breast cancer	rs2981578	C/EBPβ	*FGFR2*	Meyer et al., 2008
Breast cancer	rs2981578	FOXA1	*FGFR2*	Meyer et al., 2013
Breast cancer	rs35054928	E2F1	*FGFR2*	Meyer et al., 2013
Renal cancer	11q13.3	HIF-2	*CCND1*	Schödel et al, 2012
Prostate cancer	rs339331	HOXB13	*RFX6*	Huang et al., 2014
T2D	rs10946398	Splicing	*CDKAL1*	Zhou et al., 2014
T2D/BMI	rs4684847	PRRX1	*PPARG2*	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.

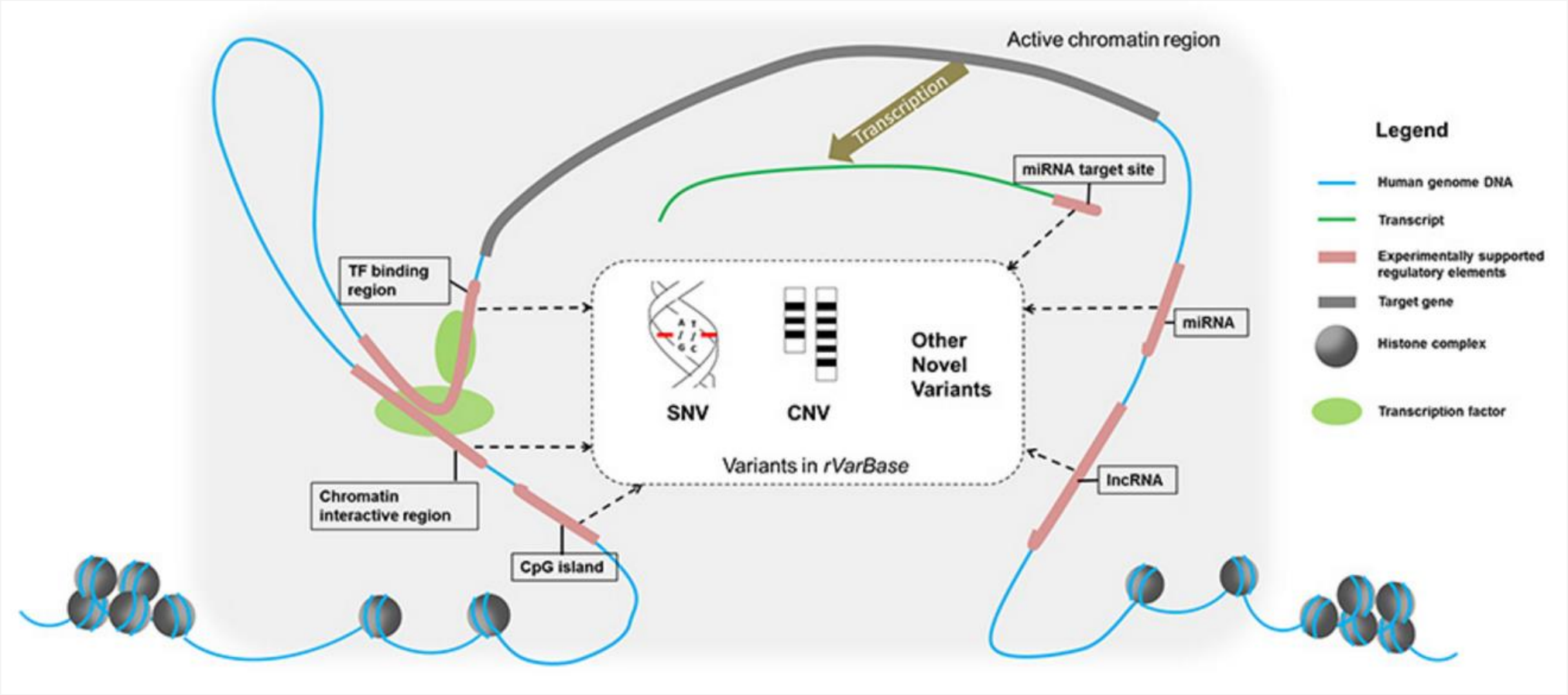
Abbreviations: 3C, chromosome conformation capture; 5C, chromosome conformation capture carbon copy; CADD, combined annotation-dependent depletion; ChIA-PET, chromatin interaction analysis by paired-end tag sequencing; DANN, deleterious annotation of genetic variants using neural networks; DNase-seq, 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.

Kingston University London

Mining the Unknown: Assigning Function to Noncoding Single Nucleotide Polymorphisms

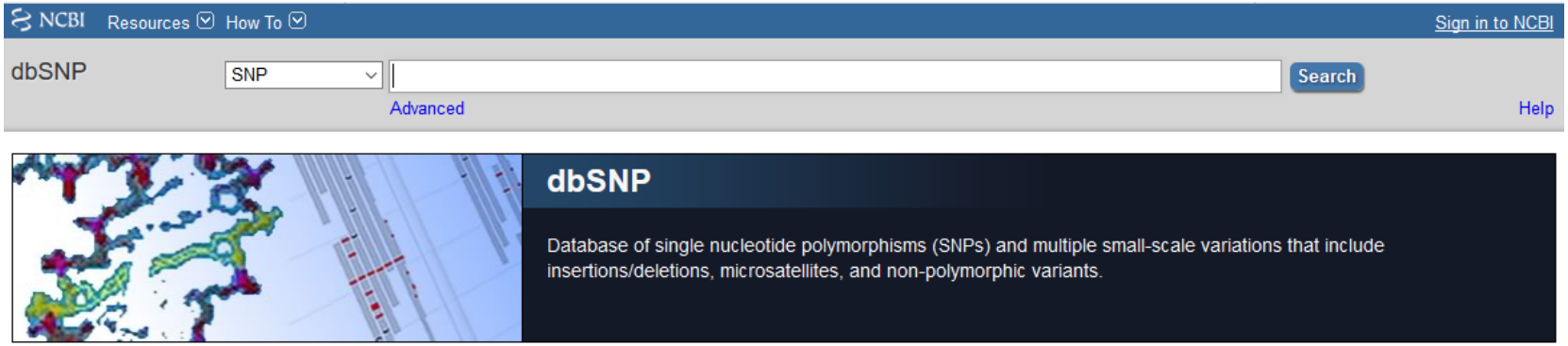
Sierra S. Nishizaki[1] and Alan P. Boyle[1,2,*]

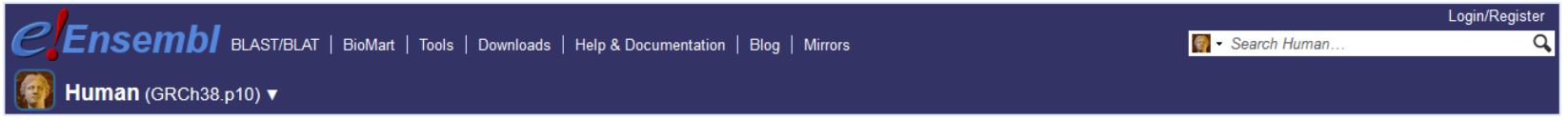
Assigning Function to SNPs



SNP Databases

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





Kingston University London

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 Wellderly; 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.

Kingston University London

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 Wang[1], Andrew V. Uzilov[1] 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 Wellderly [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. "Unique to study" counts variants that were observed only in that particular study. "Variants passed" 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

[a]dbNSFP contains hypothetical variants, see text

[b]ExAC includes samples from 1000 Genomes, ESP6500, and TCGA

[c]Note that data from HGMD, PharmGKB, UK10K diseases and TCGA germline are not visible to external users on the RVS website

[d]Counts for SwissVar refer to distinct amino acid changes. Further details on individual resources are provided in Additional file 4: Table S3

Kingston University London

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

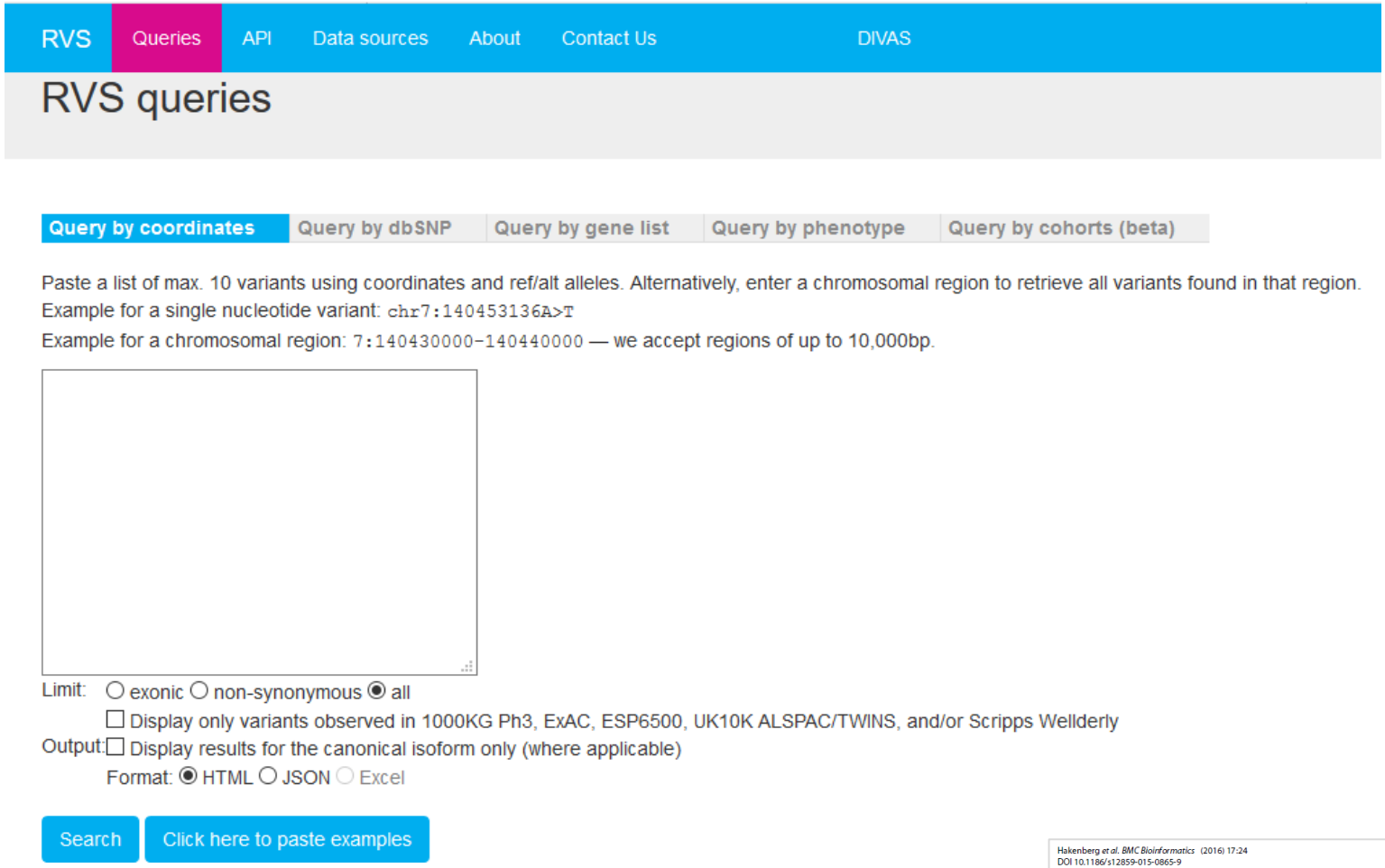
BMC Bioinformatics

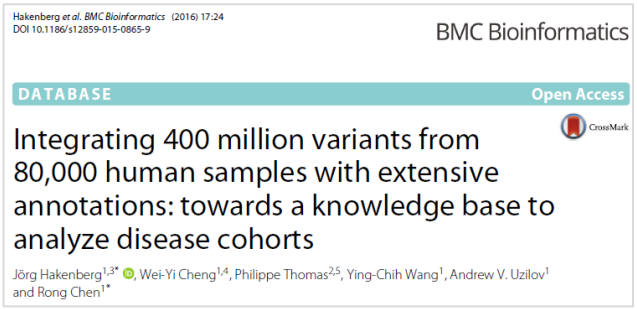
DATABASE Open Access

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 Wang[1], Andrew V. Uzilov[1] and Rong Chen[1*]

SNP Databases





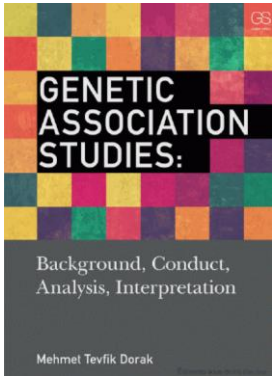
Kingston University London

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.

Kingston University London



eQTLs

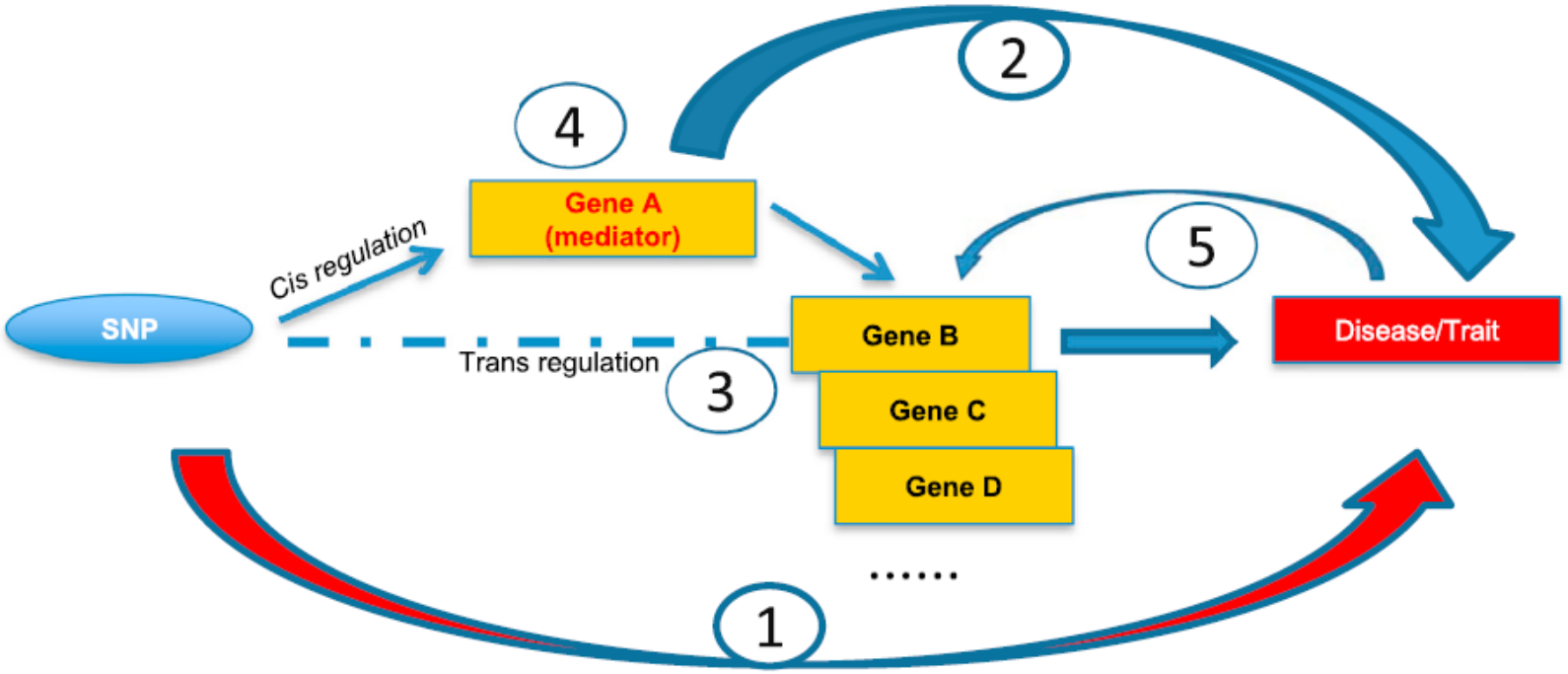


Figure 2. Mediation Mechanisms of eQTLs

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).

Kingston University London

Dynamic Role of *trans* Regulation of Gene Expression in Relation to Complex Traits

Chen Yao,[1,2] Roby Joehanes,[1,2,3] Andrew D. Johnson,[1,2] Tianxiao Huan,[1,2] Chunyu Liu,[1,2] Jane E. Freedman,[4] Peter J. Munson,[5] David E. Hill,[6,7] Marc Vidal,[6,7] and Daniel Levy[1,2,*]

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)

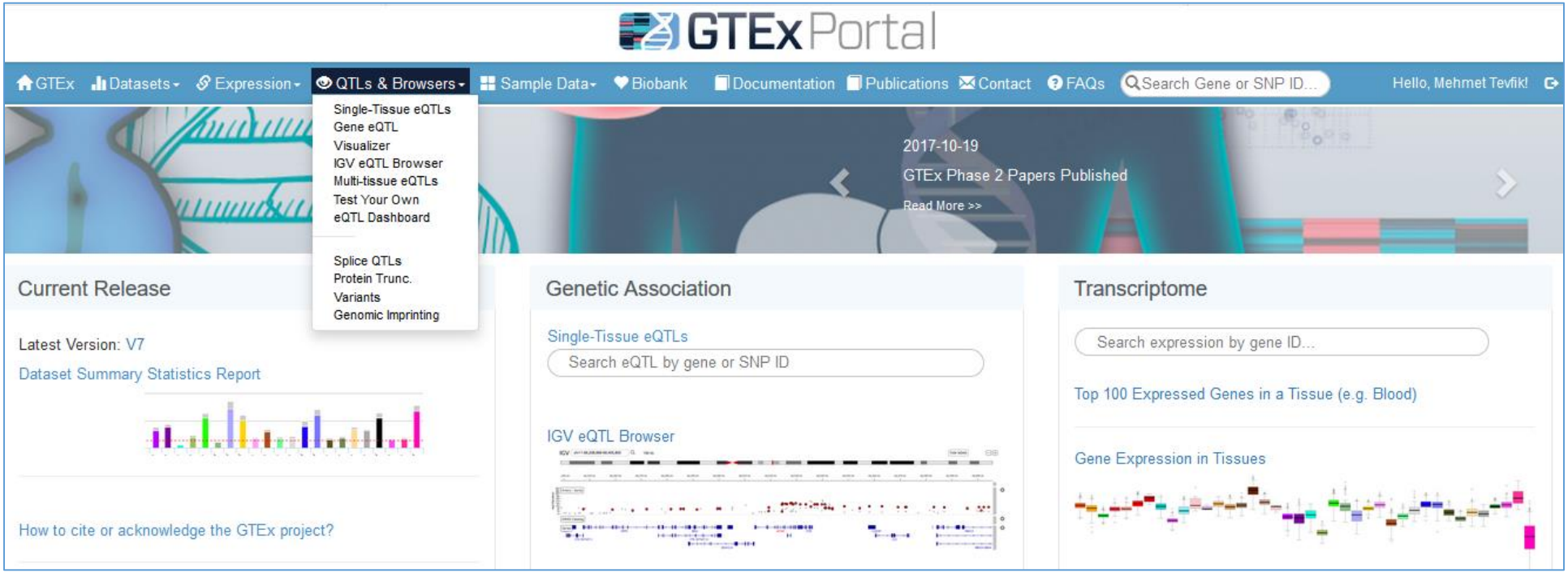
seeQTL (help / ref) Plateletomics (Platelet transcriptomics by gender) ImmGen eQTL Browser (mouse)

eCAVIAR: colocalization of GWAS signals and eQTLs (ref / manual) PancanQTL (ref)

DIVAN (ref)

Kingston University London

GTEx



Kingston University London

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

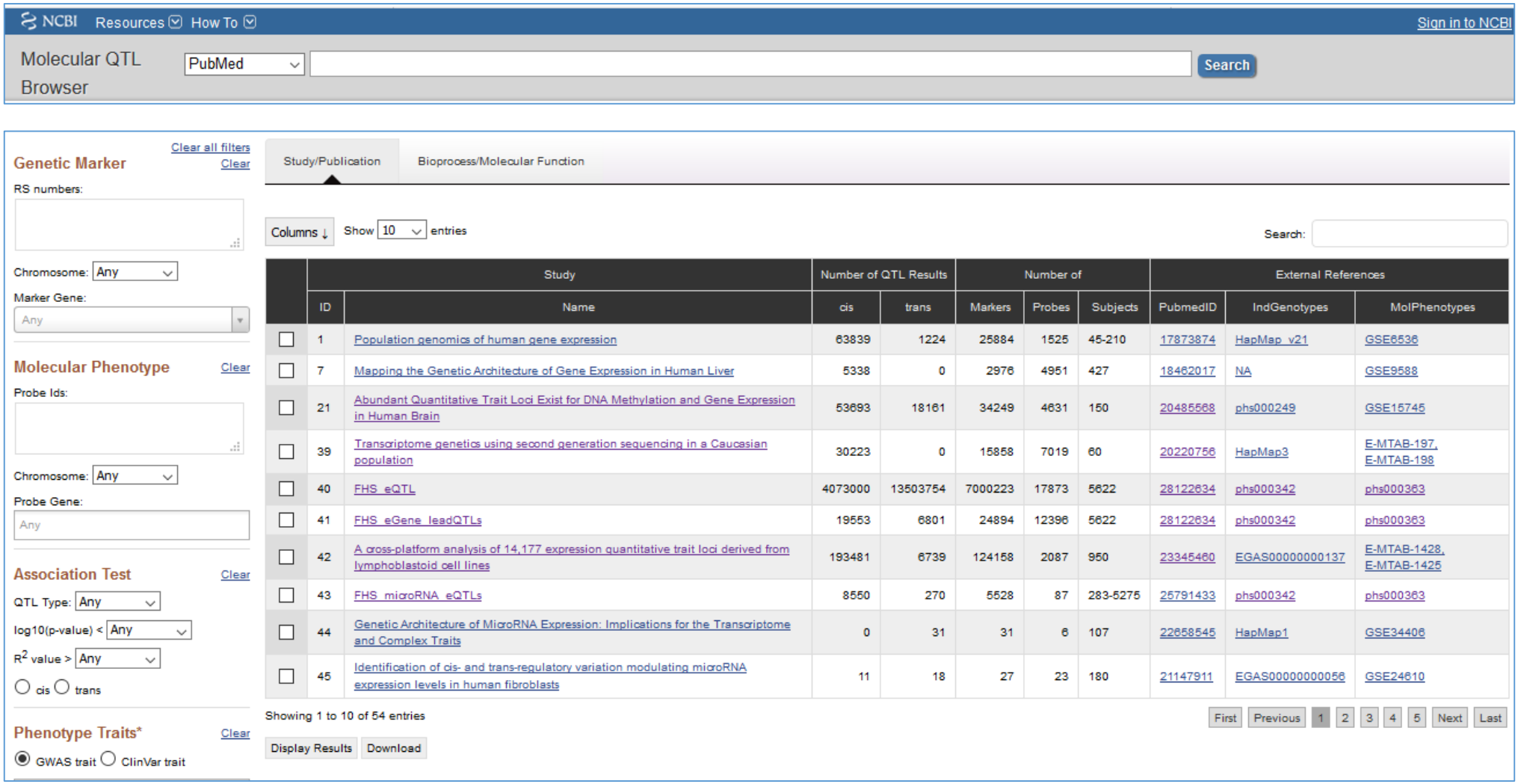
Query eQTL Results

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

Gene or SNP name: Search

Kingston University London

NCBI eQTL Browser (for FHS)



	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

Kingston University London

Trans-eQTLs

Table 1. *trans*-eQTL Hotspots

Hotspot Location (hg19)	Number of *trans*-eQTLs	Number of *trans*-eGenes Associated with Index eQTL	Directional Bias of *trans*-eGenes Associated with Index eQTL	Traits Associated in GWAS with Index eQTLs	*trans*-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	*STAT1/STAT2*
2: 60,708,597–60,725,451	14	14	−79%	fetal hemoglobin level	*NFAT/SP1*
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	*TCF3/ETS2*
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	*SP1/TCF3*
7: 50,423,963–50,562,361	19	76	−59%	childhood acute lymphoblastic leukemia	*PAX4*
12: 54,712,308–54,736,470	2	14	+79%	mean platelet volume	*ETS2*
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	*IRF8/IRF2*
17: 27,072,463–27,322,441	45	32	+55%	mean corpuscular volume	*E4F1*
17: 33,796,260–33,944,055	4	51	+75%	mean platelet volume	*ETS2/MAZ*

Plus sign (+) denotes the positive association; minus sign (−) denotes the negative association.
[a]Transcription 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.

Kingston University London

Dynamic Role of *trans* Regulation of Gene Expression in Relation to Complex Traits

Chen Yao,[1,2] Roby Joehanes,[1,2,3] Andrew D. Johnson,[1,2] Tianxiao Huan,[1,2] Chunyu Liu,[1,2] Jane E. Freedman,[4] Peter J. Munson,[5] David E. Hill,[6,7] Marc Vidal,[6,7] and Daniel Levy[1,2,*]

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 *Zhernakova 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 independant 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: Search

Kingston University London

mQTL Database | Home | Search | Download | Help

University of BRISTOL | MRC Integrative Epidemiology Unit

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.

Go »

Download

Download the entire ARIES mQTL database.

Go »

e meQTL | 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 e/meQTL Browse and the people behind it

Guide
Detailed documentation, guidelines and tutorials

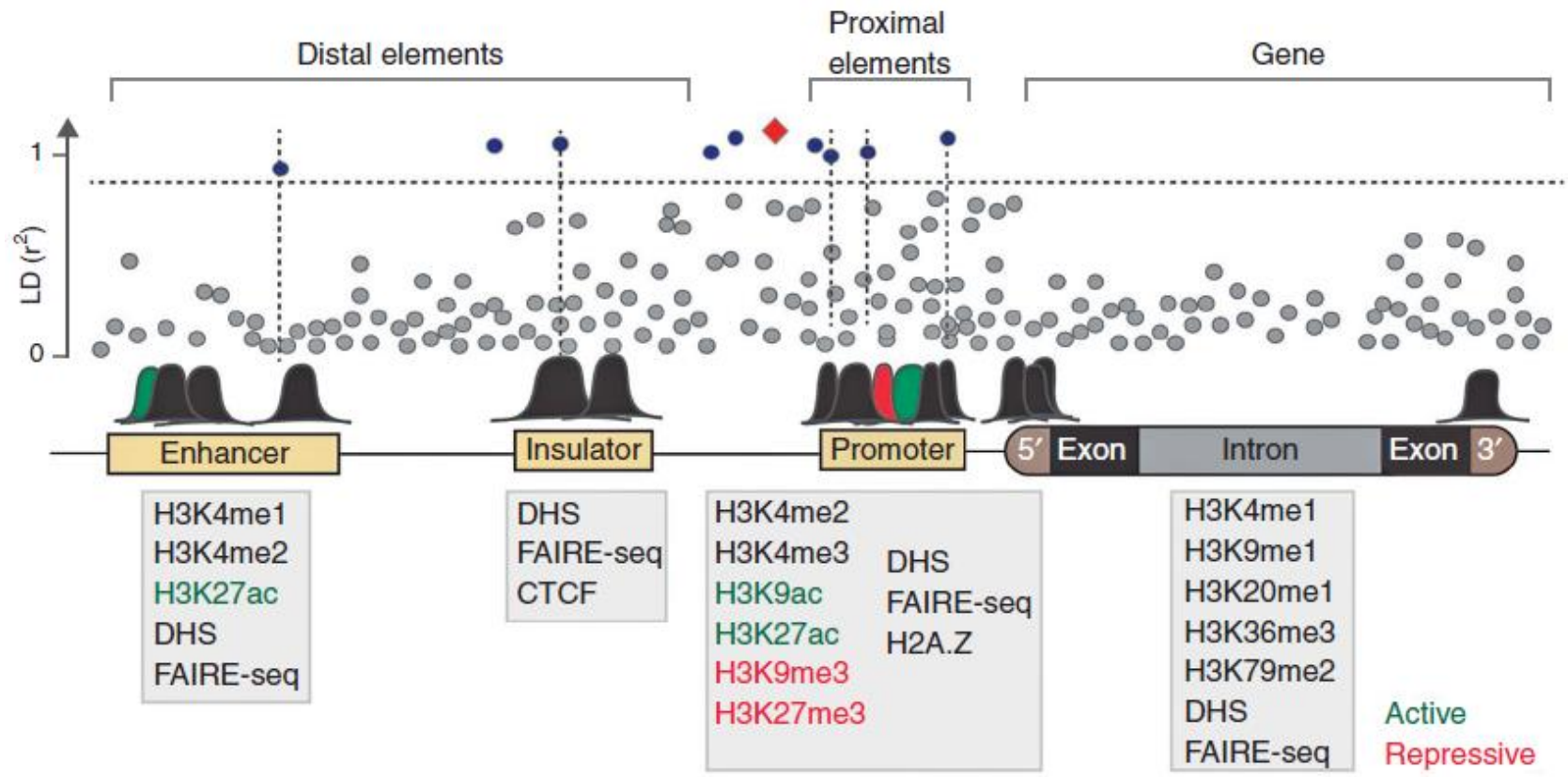
Community
Forum for questions and announcements

Videos
Tutorial videos, Q&A and event recordings

e meQTL Browser

Created and maitained by Dr. Jinyan HUANG.
jhuang at hsph dot harvard dot edu
Best Screen Resolution for browsing this websites is: 1280 x 800.

Chromatin Marks



Kingston University London

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Genetics & Development
ELSEVIER

Using chromatin marks to interpret and localize genetic associations to complex human traits and diseases

Gosia Trynka[1,2,3,4] and Soumya Raychaudhuri[1,2,3,4,5]

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): rs6964969

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

or, select a GWAS:

Submit Query

Query SNP: rs6964969 and variants with r^2 >= 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::CEBPalpha	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	rs28696237	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	

Kingston University London

Chromatin Marks

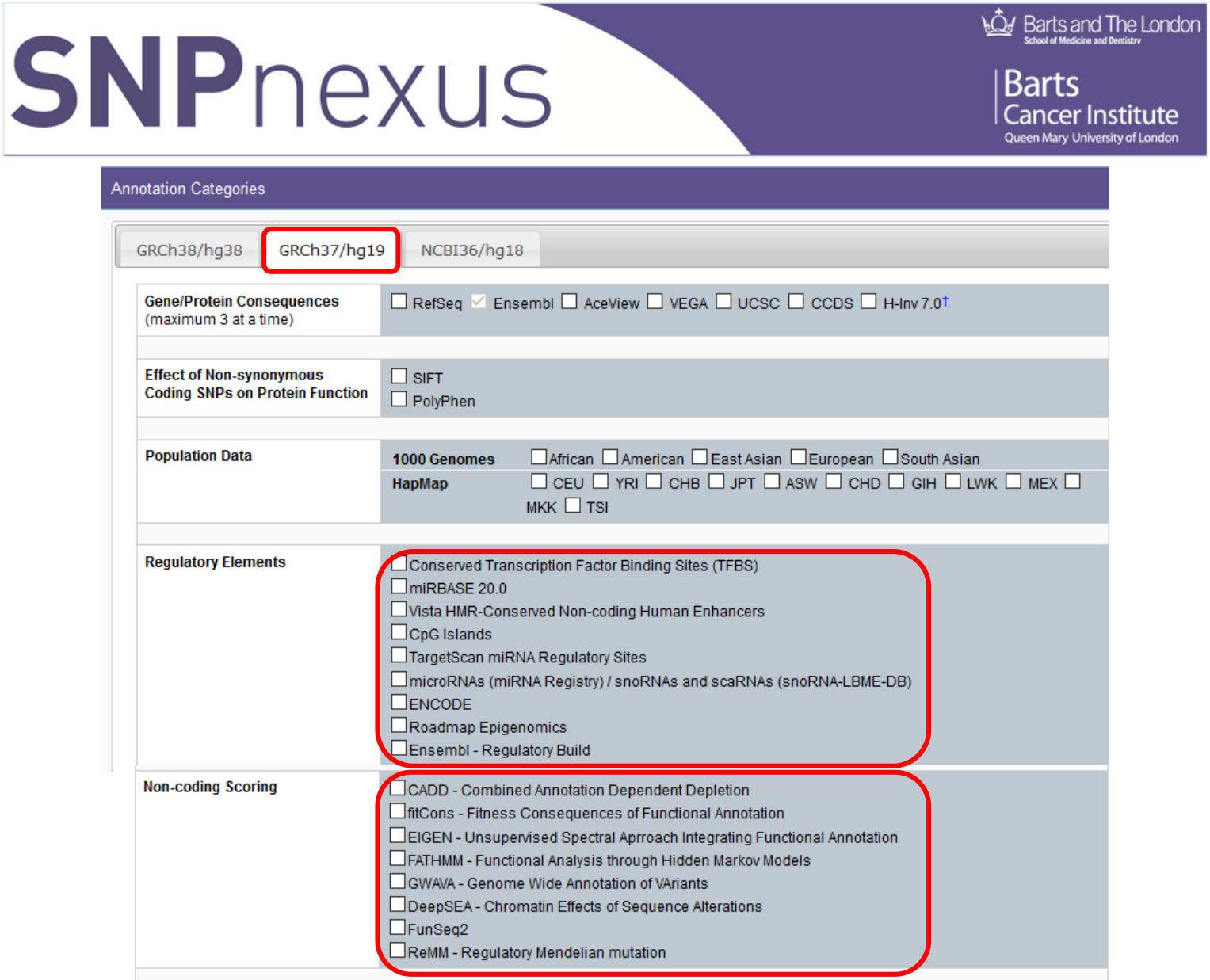
Updated in 2018

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

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

Kingston University London

Chromatin Marks



Kingston University London

Chromatin Interactions

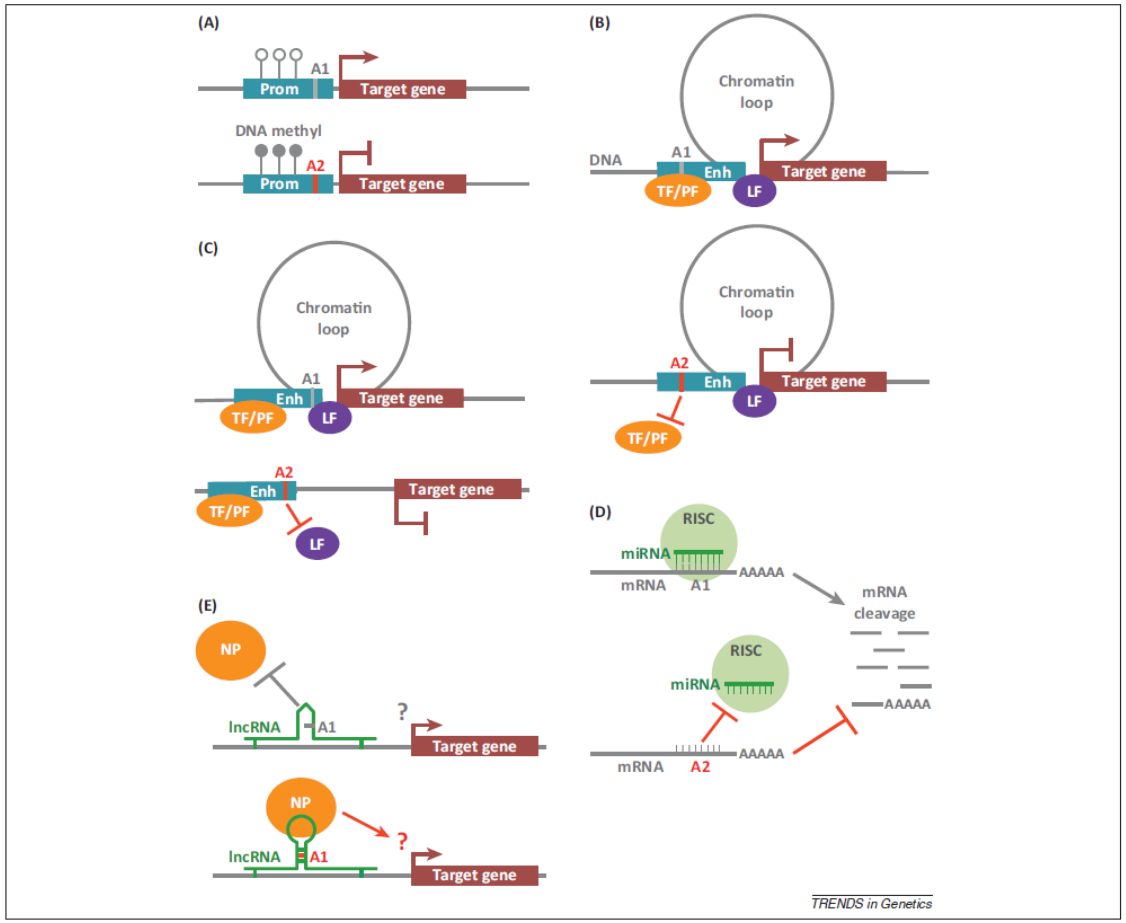


Figure 2. Non-coding genetic risk variants. **(A)** Genetic risk variants influence DNA methylation level at promoter regions. **(B)** Genetic risk variants modulate transcription factor binding to chromatin. **(C)** Genetic risk variants alter chromatin loop formation bridging enhancers and promoters. **(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

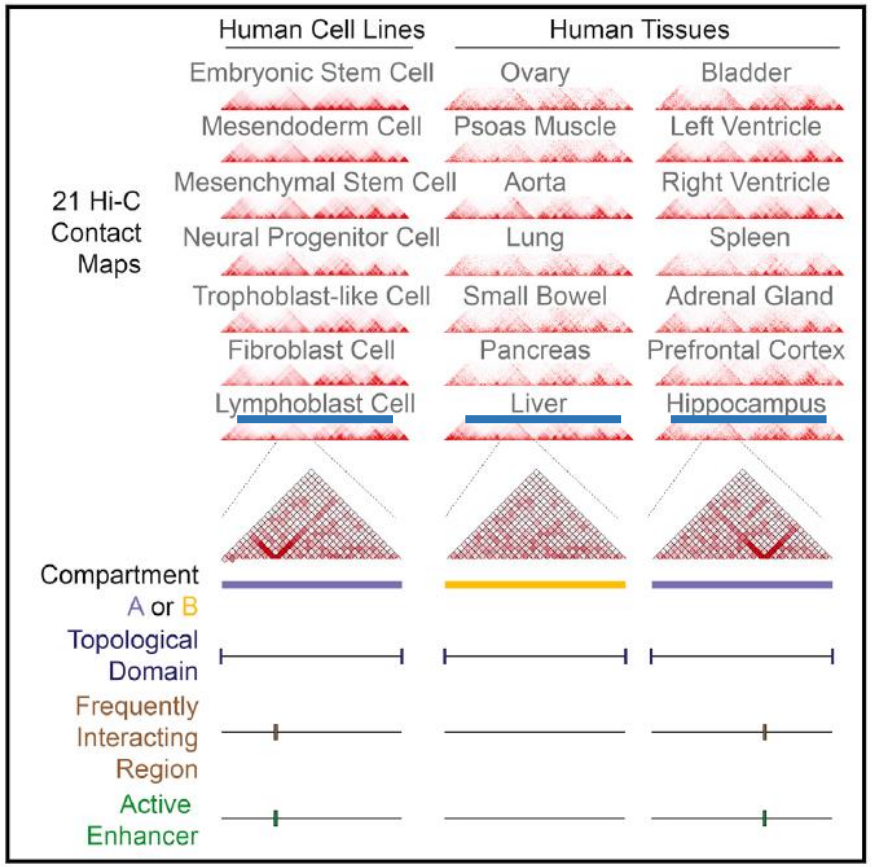
The human genome is organized in a 3D architecture which is thought to regulate a diverse set of DNA-templated processes [87–91]. This allows regulatory elements, such as enhancers and promoters, to interact physically through long-range chromatin interactions, or chromatin loops, to regulate gene expression [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).

Kingston University London

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

Kingston University London

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

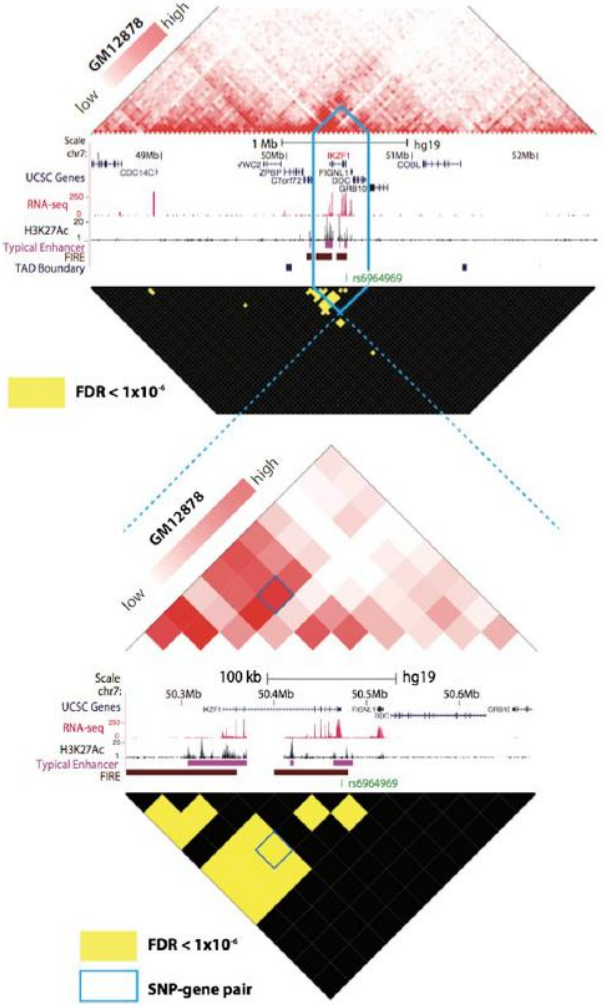
Anthony D. Schmitt,[1,2,12,13] Ming Hu,[3,12,14,*] Inkyung Jung,[1,15] Zheng Xu,[4,10,11] Yunjiang Qiu,[1,5] Catherine L. Tan,[1,13] Yun Li,[4] Shin Lin,[6] Yiing Lin,[7] Cathy L. Barr,[8] and Bing Ren[1,9,16,*]

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).

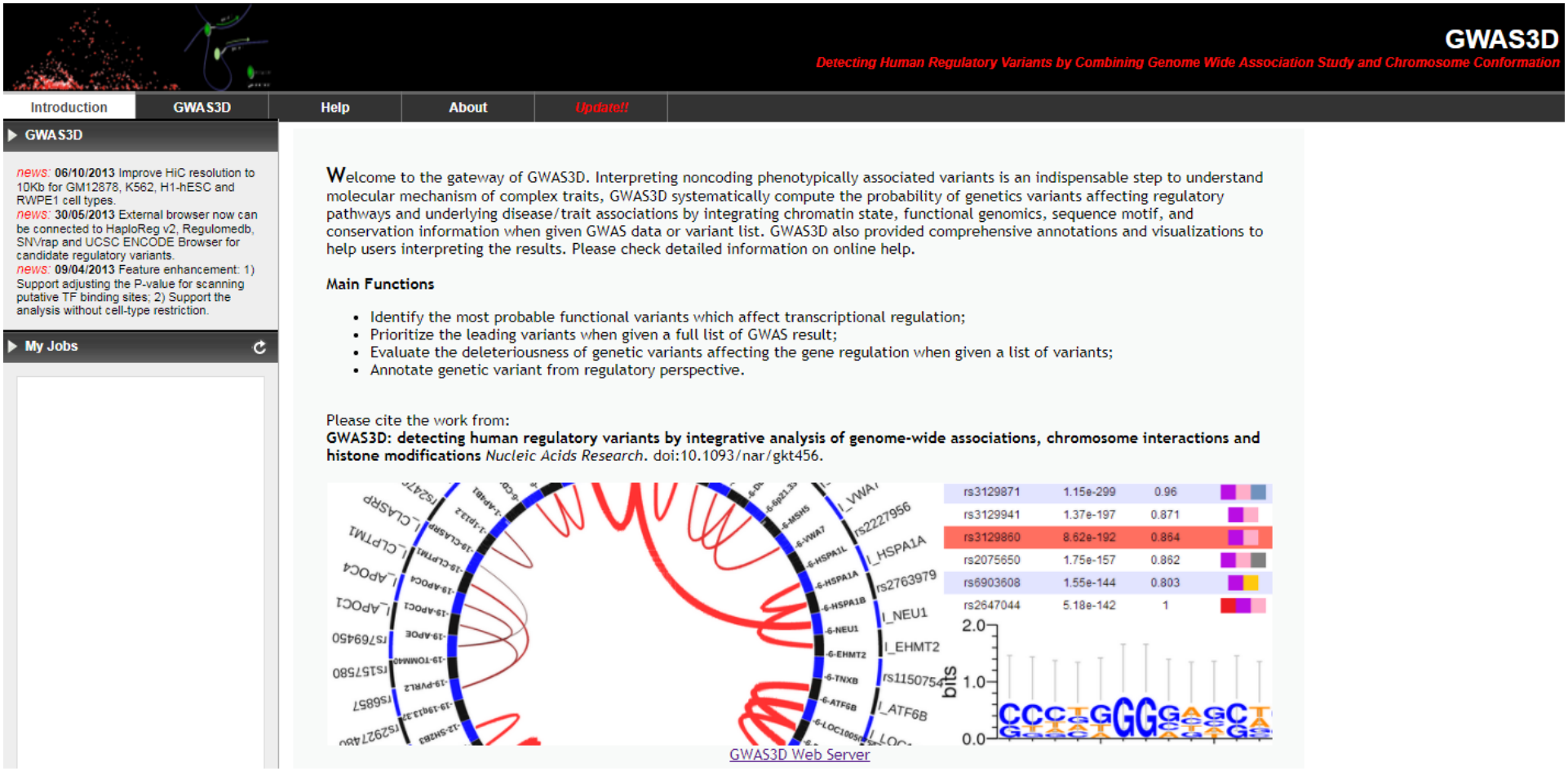


Kingston University London

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] Yiing Lin,[7] Cathy L. Barr,[8] and Bing Ren[1,9,16,*]

Chromatin Interactions

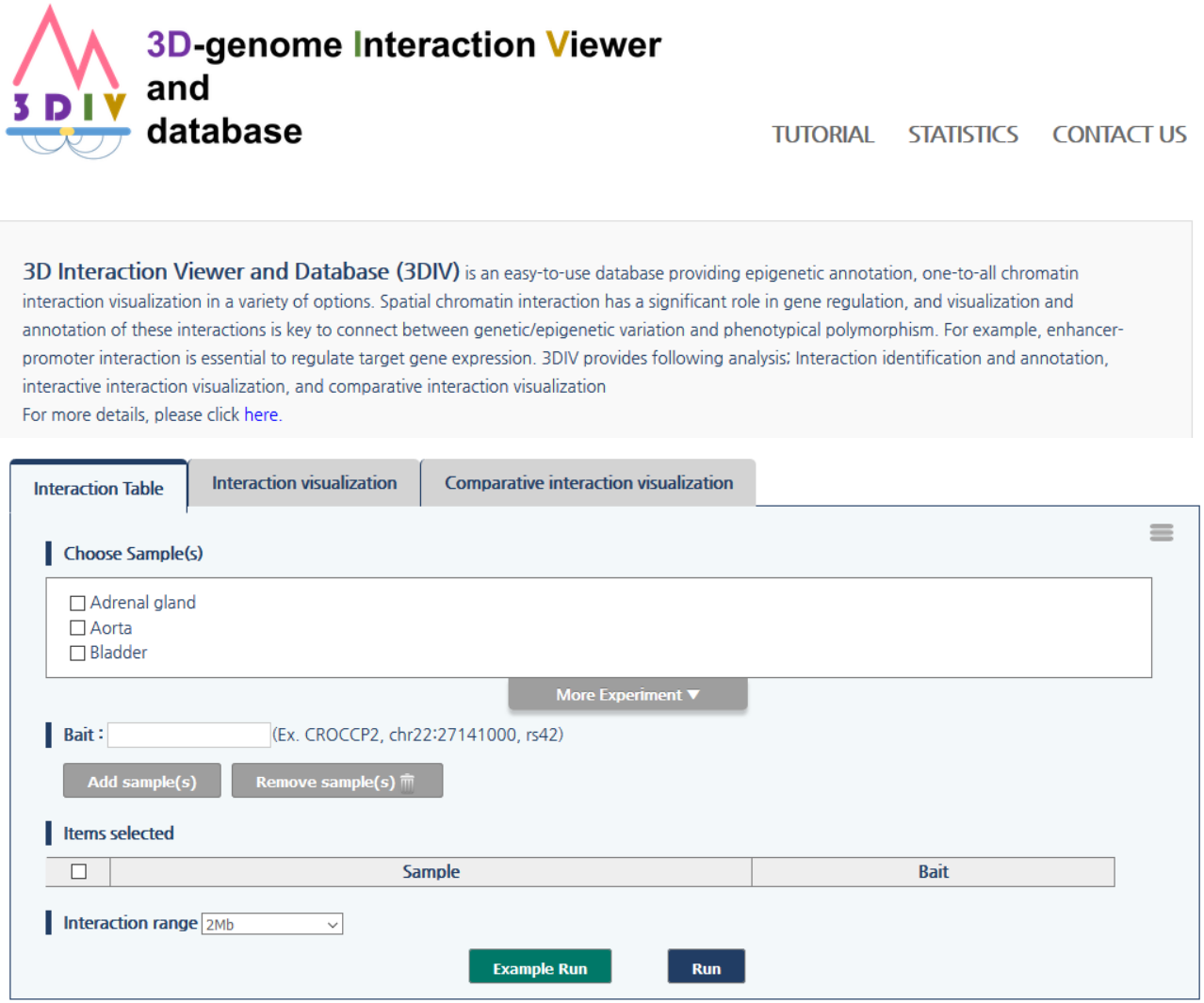


Kingston University London

GWAS3D: detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications

Mulin Jun Li[1,2], Lily Yan Wang[1,2], Zhengyuan Xia[2,3], Pak Chung Sham[2,4,5,6] and Junwen Wang[1,2,4,*]

Chromatin Interactions



Kingston University London

D52–D57 Nucleic Acids Research, 2018, Vol. 46, Database issue
doi: 10.1093/nar/gkx1017
Published online 2 November 2017

3DIV: A 3D-genome Interaction Viewer and database

Dongchan Yang[1,†], Insu Jang[2,†], Jinhyuk Choi[2], Min-Seo Kim[2], Andrew J. Lee[3], Hyunwoong Kim[3], Junghyun Eom[3], Dongsup Kim[1,*], Inkyung Jung[3,*] and Byungwook Lee[2,*]

SNP Functionality Scores

Table 1. Online Resources for Accessing Noncoding SNP Annotation Tools[a]

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]
FunciSNP	FunciSNP 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/FunciSNP.html	Coetzee *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 Kellis [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 circle plot mapping local 3D interactions. http://jiwanglab.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://compgen.bscb.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. ftp://ftp.sanger.ac.uk/pub/resources/software/gwava/	Ritchie *et al.* [25]
CADD	CADD trains on a linear kernel support vector matrix 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://cbcl.ics.uci.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.beerlab.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]

[a] Abbreviation: HGMD, Human Gene Mutation Database.

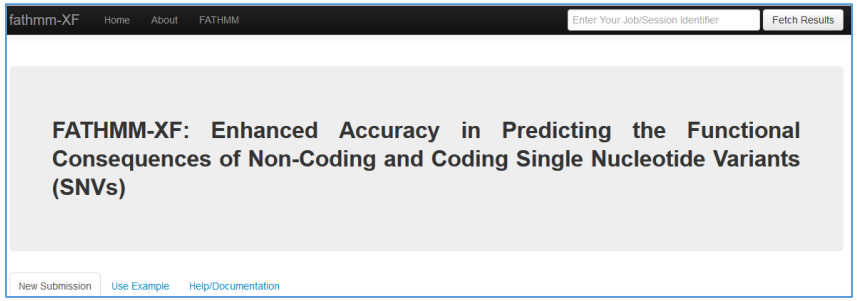
Kingston University London

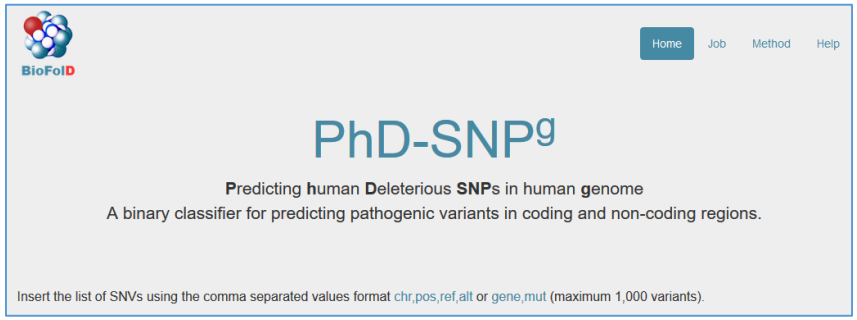
Mining the Unknown: Assigning Function to Noncoding Single Nucleotide Polymorphisms

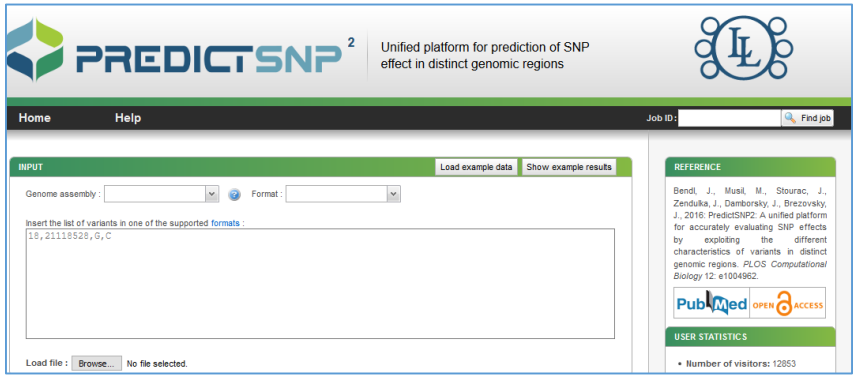
Sierra S. Nishizaki[1] and Alan P. Boyle[1,2,*]

SNP Functionality Scores

Aggregate Functionality Scores



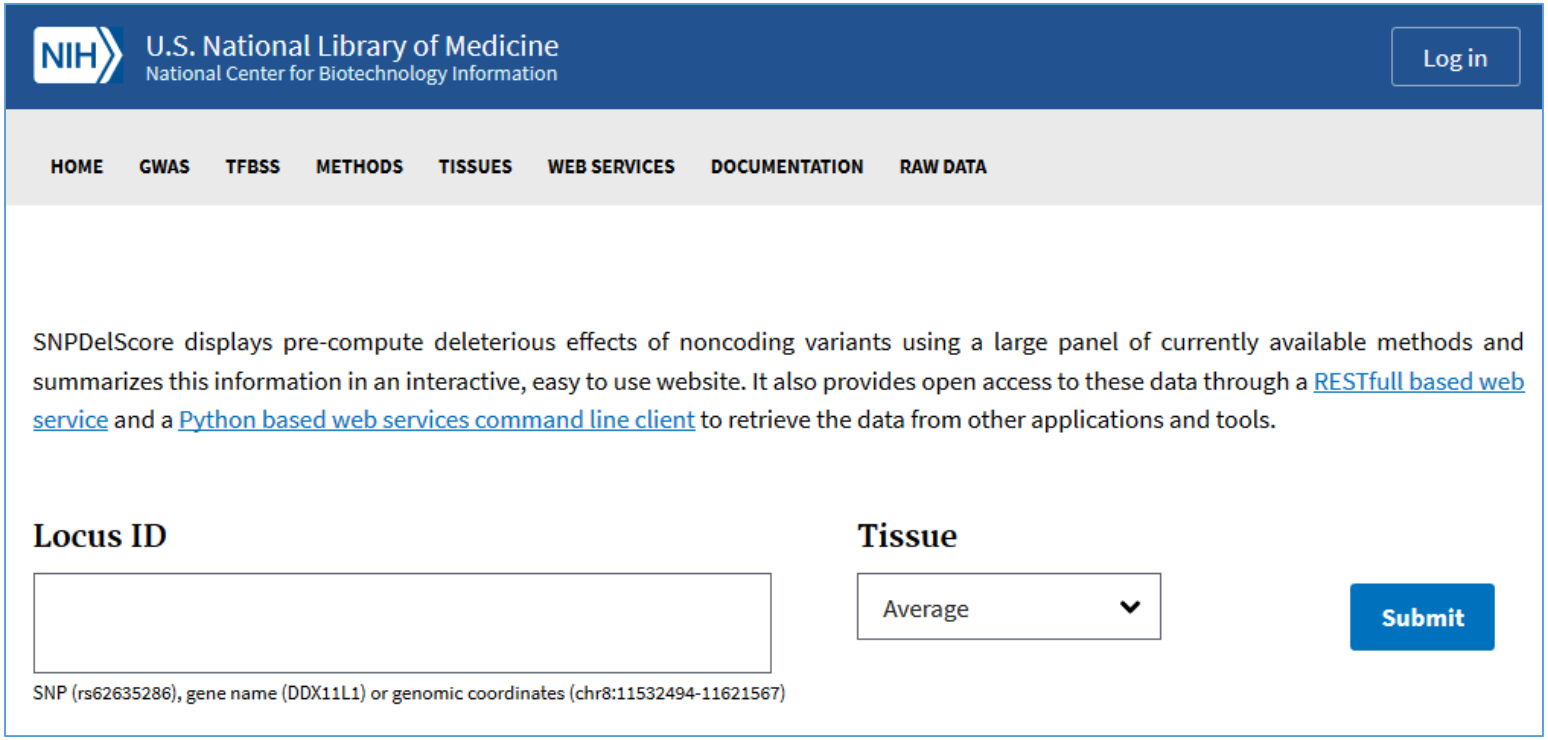




Kingston University London

SNP Functionality Scores

Aggregate Functionality Scores
SNPDelScore



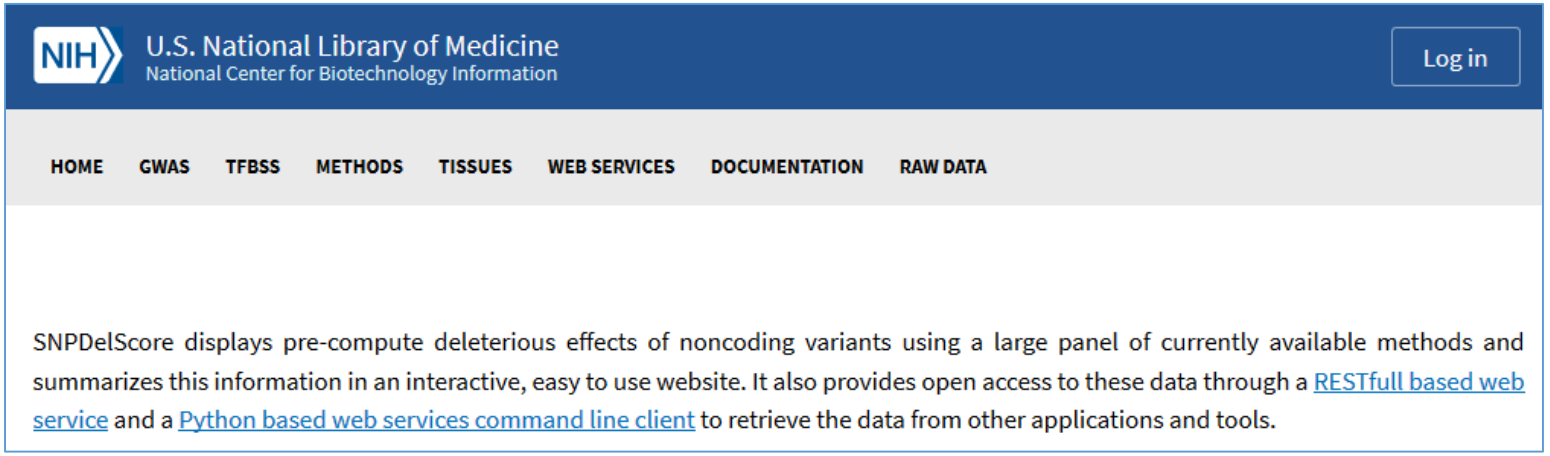
Kingston University London

SNPDelScore: combining multiple methods to score deleterious effects of noncoding mutations in the human genome

Roberto Vera Alvarez, Shan Li, David Landsman and Ivan Ovcharenko*

SNP Functionality Scores

Aggregate Functionality Scores
SNPDelScore



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

Kingston University London

SNPDelScore: combining multiple methods to score deleterious effects of noncoding mutations in the human genome

Roberto Vera Alvarez, Shan Li, David Landsman and Ivan Ovcharenko*

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

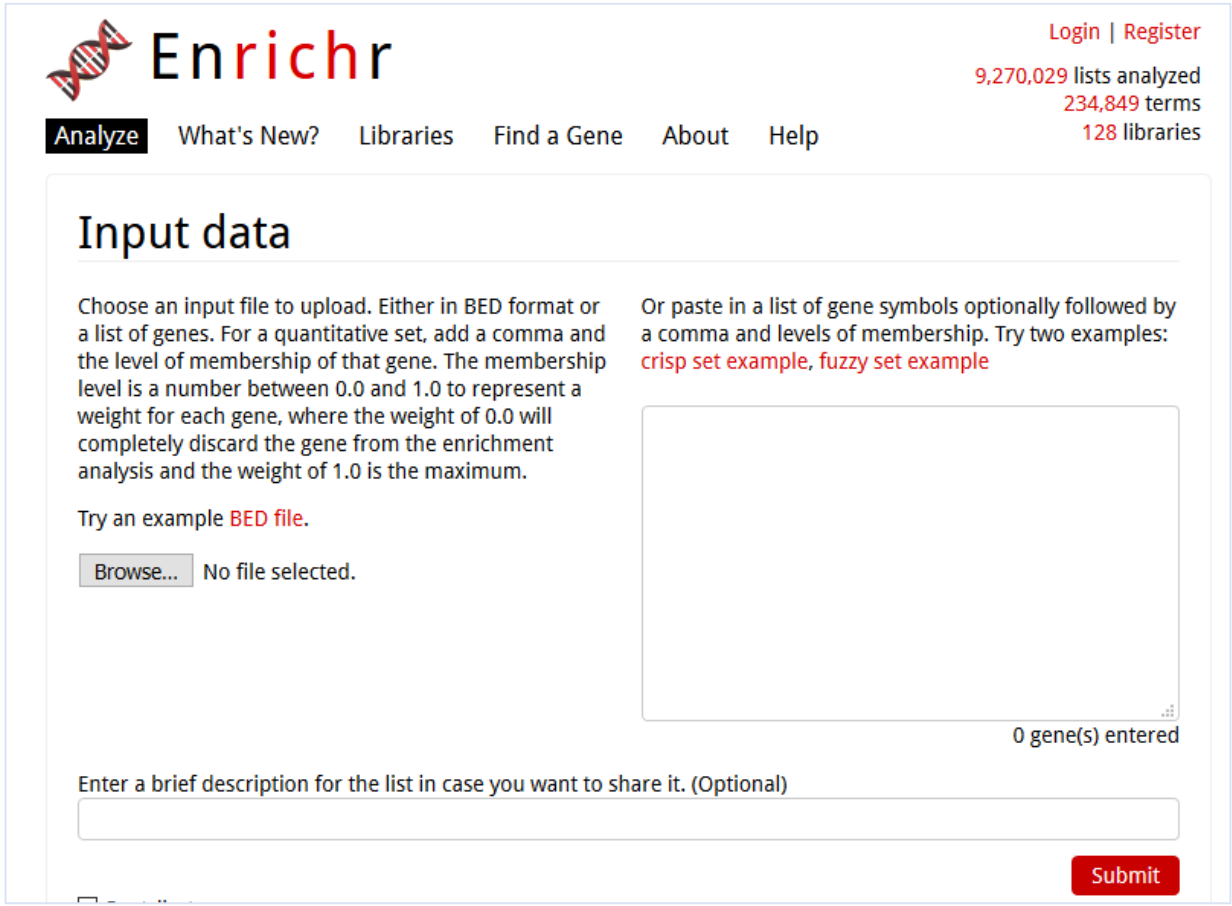
INVITED REVIEW

Gene set analysis for interpreting genetic studies

Tune H. Pers[1,2,*]

Kingston University London

Gene Set Enrichment Analysis



Kingston University London

W90–W97 Nucleic Acids Research, 2016, Vol. 44, Web Server issue Published online 3 May 2016
doi: 10.1093/nar/gkw377

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

Maxim V. Kuleshov[1], Matthew R. Jones[1], Andrew D. Rouillard[1], Nicolas F. Fernandez[1], Qiaonan Duan[1], Zichen Wang[1], Simon Koplev[1], Sherry L. Jenkins[1], Kathleen M. Jagodnik[2], Alexander Lachmann[1], Michael G. McDermott[1], Caroline D. Monteiro[1], Gregory W. Gundersen[1] 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-iterm pairs at one time in Cytoscape due to the limitation of cytoscape.js. However, Concept Map can visualize big data !).

Data Visulization:

eGIFT Gene Enrichment Analysis

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

Samples:

Entrez Sample | Uniprot Sample | Ensembl Sample | Gene symbol Sample | Gene-Iterm Sample

Input type:

○ Entrez ID ○ UniProt ID ○ Ensembl ID ○ Gene symbols ○ Symbol Iterm Pair

Custom Data

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

Sample: Sample

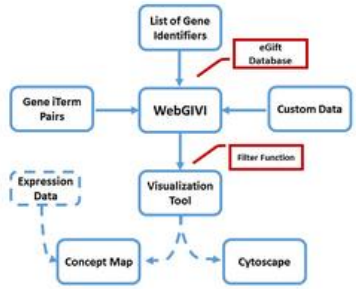
○ Custom Data

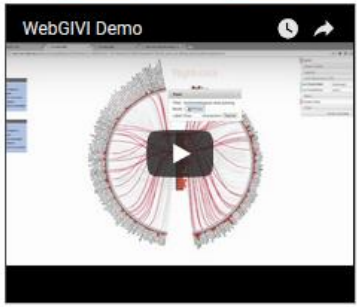
Name Your input List (optional)

Database you want to choose: ◉ eGIFT ○ Amigo2

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

Submit





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

BMC Bioinformatics

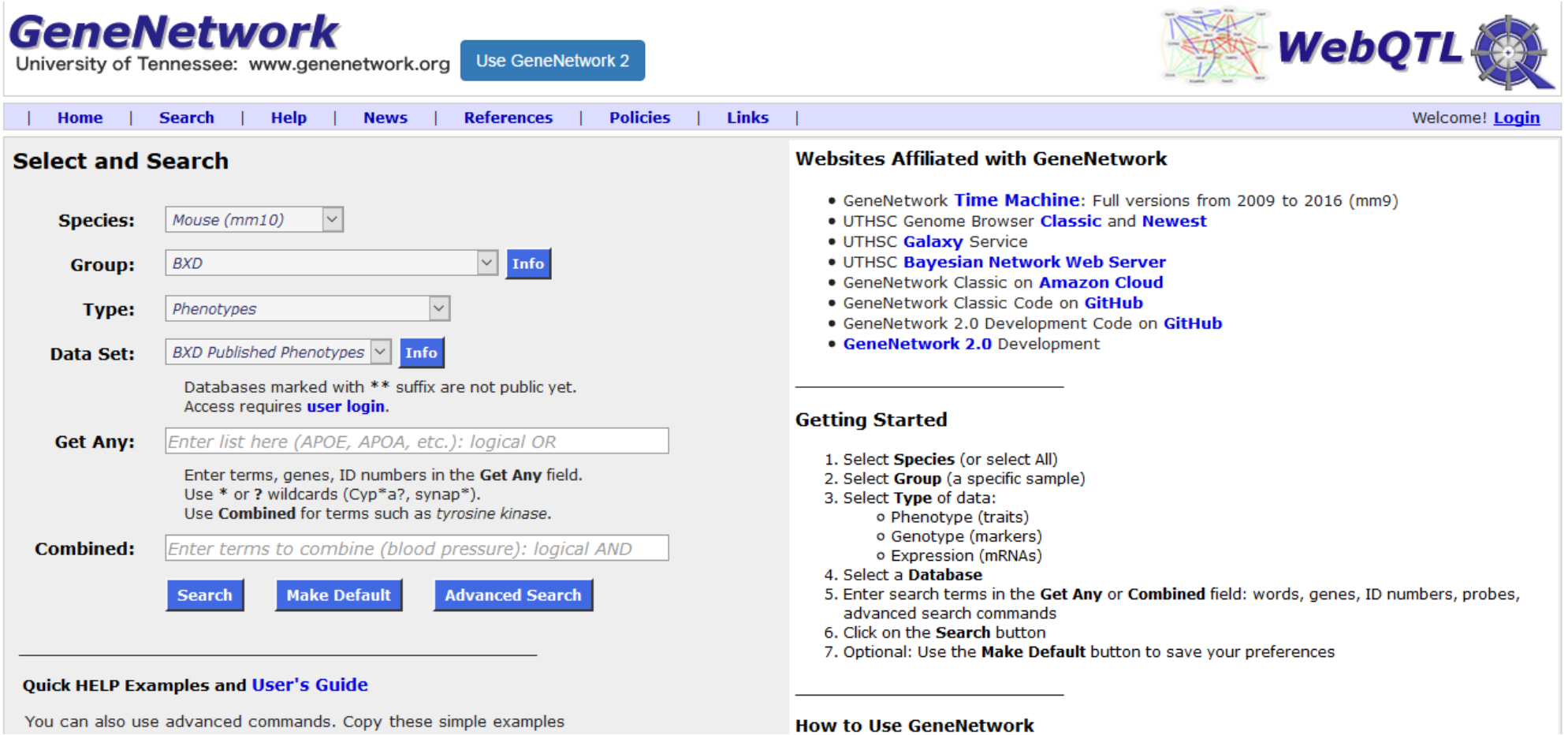
SOFTWARE Open Access

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

Liang Sun[1,6], Yongnan Zhu[2,3], A. S. M. Ashique Mahmood[4], Catalina O. Tudor[4], Jia Ren[5], K. Vijay-Shanker[4], Jian Chen[2] and Carl J. Schmidt[1*]

Kingston University London

Gene Networks



Kingston University London

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	

Kingston University London

Gene Networks

GeneNetwork
University of Tennessee: www.genenetwork.org

Use GeneNetwork 2

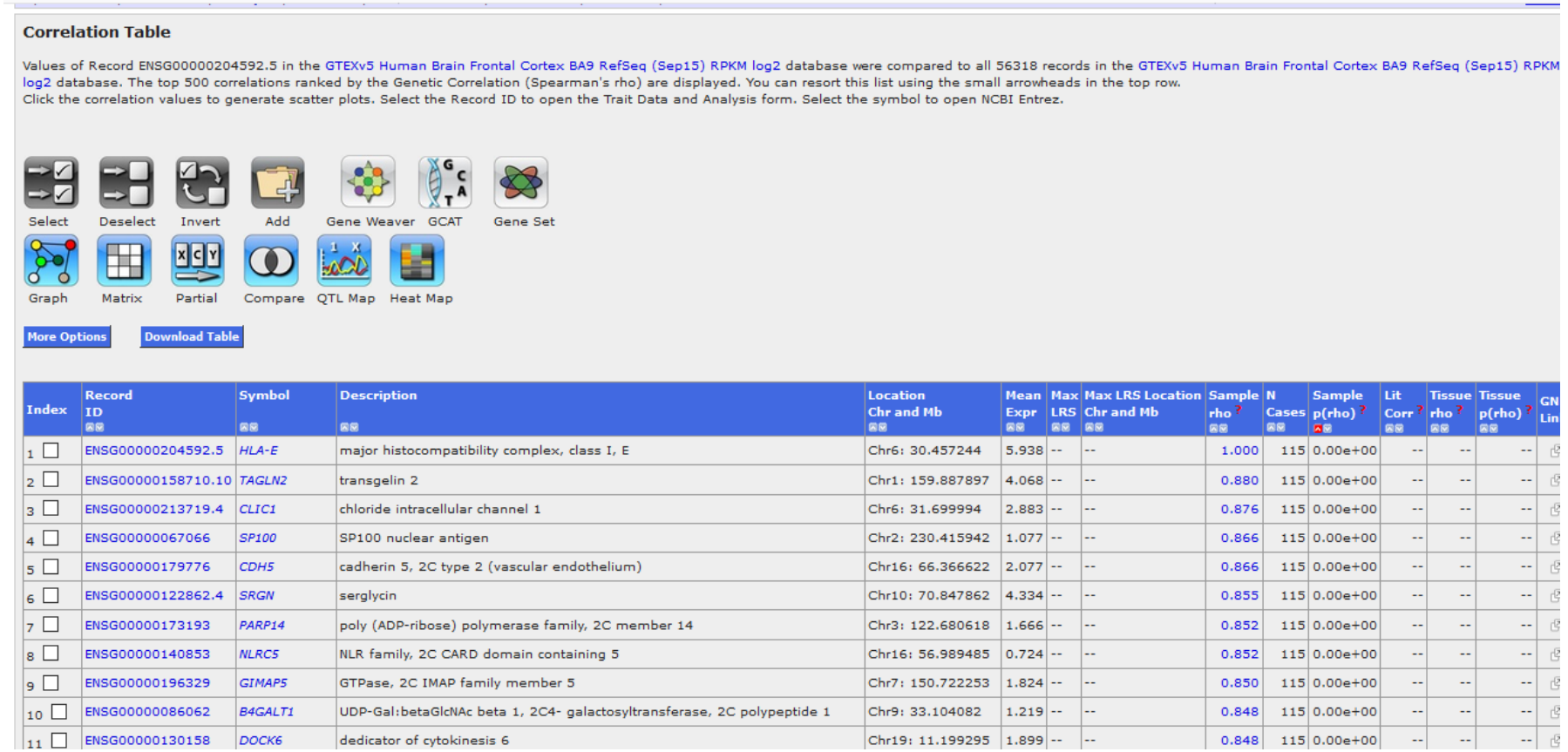
WebQTL

Kingston University London

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.

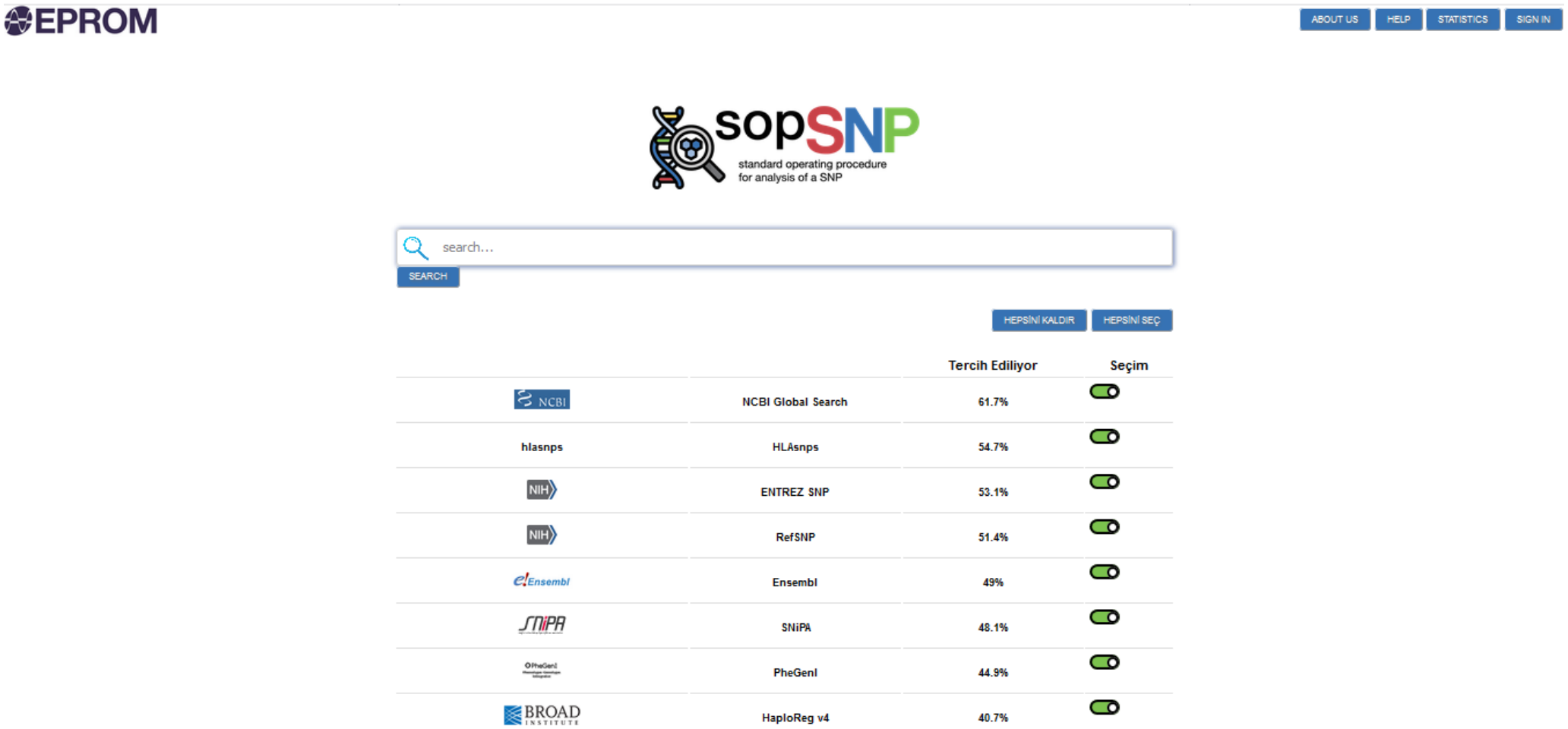
Select Deselect Invert Add Gene Weaver GCAT Gene Set
Graph Matrix Partial Compare QTL Map Heat Map

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	--	--	--	

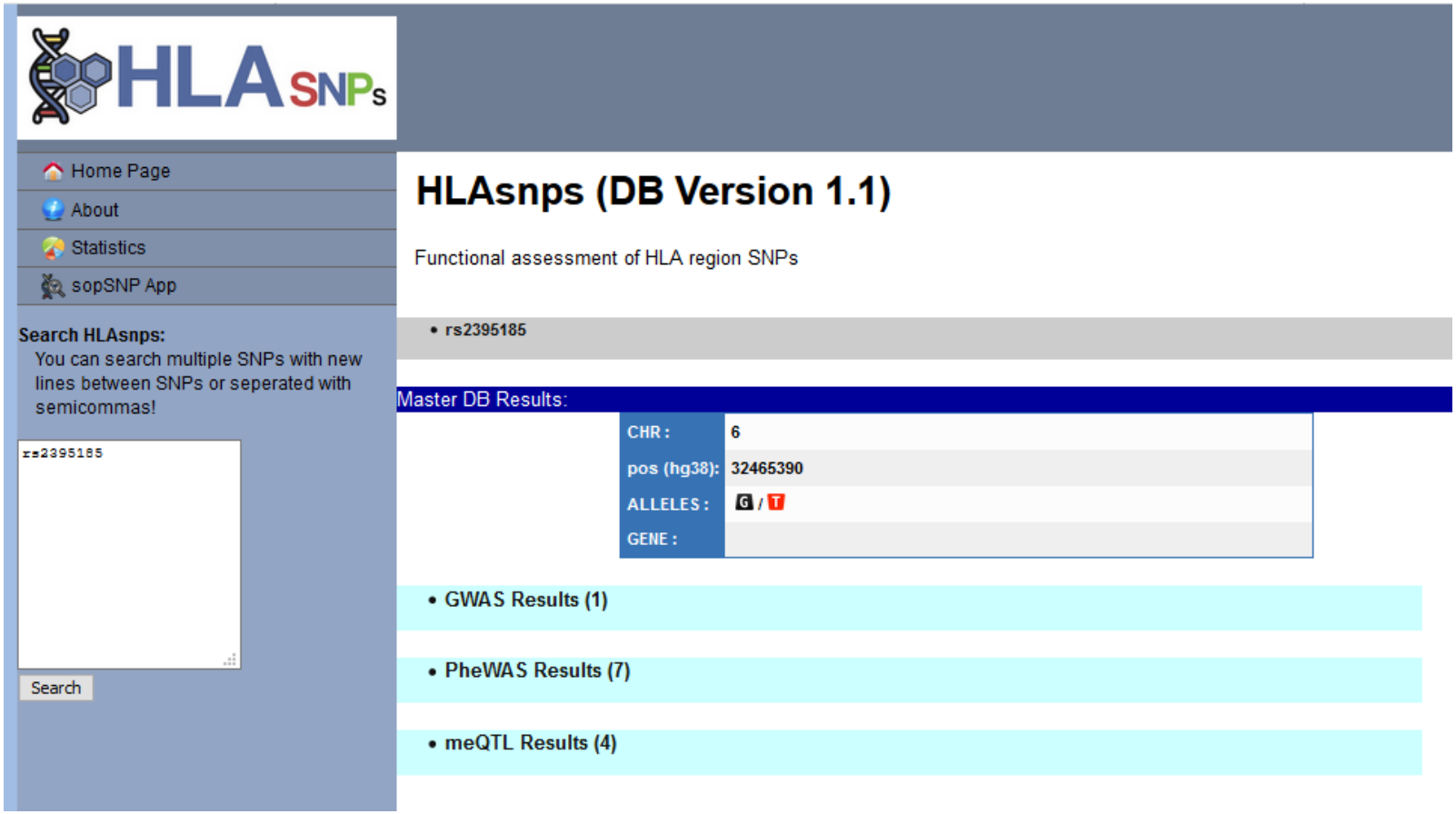
Kingston University London

sopSNP (beta)



Kingston University London

HLASNPs (beta)



Kingston University London

… 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

Kingston University London