Bioinformatic Tools for the Investigation of Tissue-specificity of Variant Effects

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

Tissue-specific effects:
eQTLs
3D confirmation
Transcriptional regulation (enhancers)

Tissue-specific tools

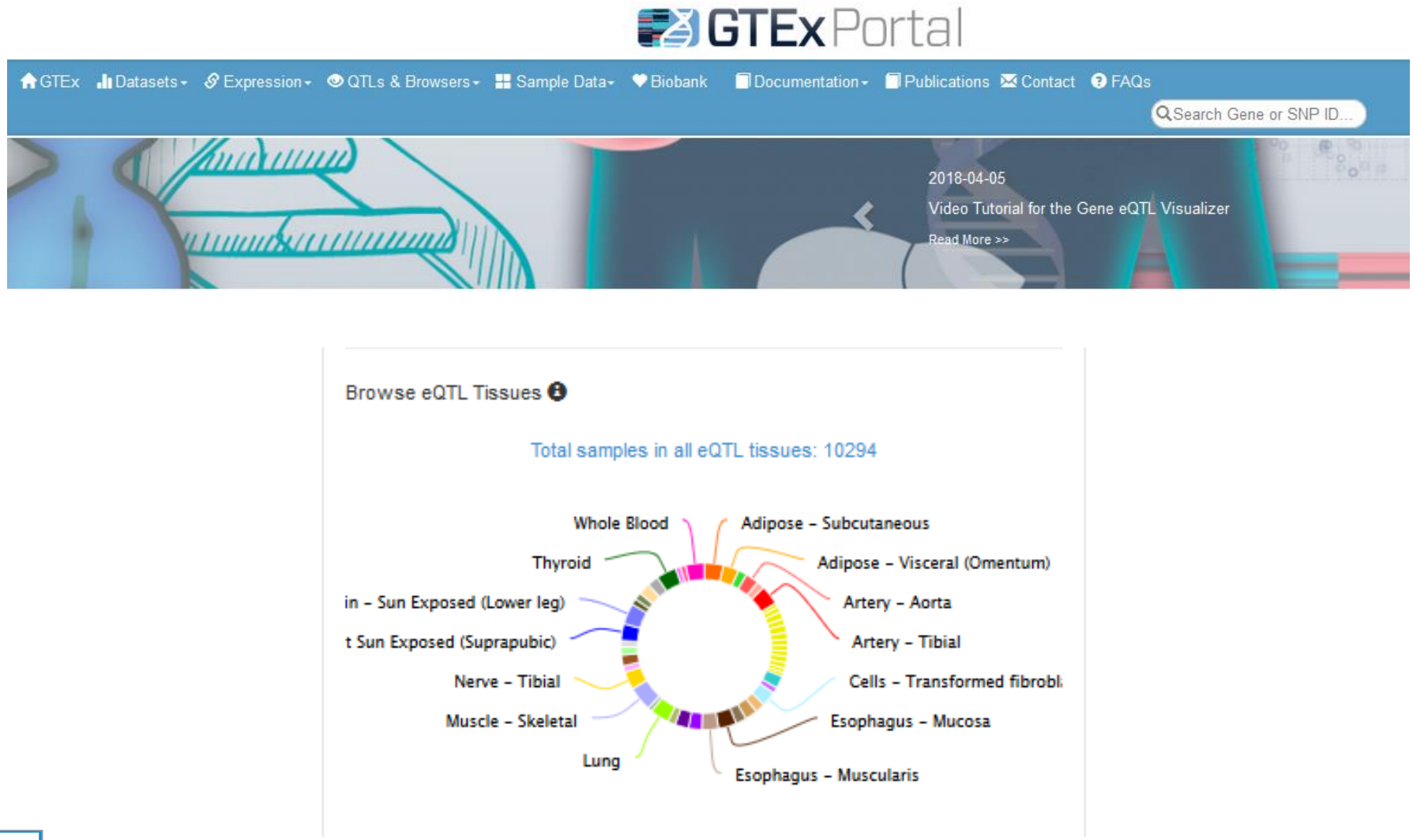
Tissue-specific Effects in Genome Biology

Gene expression is tissue-specific

DNA methylation and other epigenetic marks
Enhancer activity and enhancer-promoter interactions
miRNA expression
lncRNA expression
eQTL effects
Promoter interactions

Kingston University London

eQTL effects



Kingston University London

eQTL effects

rs12203592,IRF4,Whole_Blood

GTEx | Datasets | Expression | QTLs & Browsers | Sample Data | Biobank | Documentation | Publications | Contact | FAQs | Search Gene or SNP ID... | Hello, Mehmet Tevfik!

Test Your Own eQTLs

- Limited to 1000 entries.
- There must be only one entry per line.
- Each entry must be of the form: **VARIANT ID,GENE ID,TISSUE NAME.**

VARIANT ID: GTEx variant ID or dbSNP rs number.
GENE ID: Gencode ID, Ensembl ID, or HGNC Gene Symbol.
TISSUE NAME must use allowable tissue names.

An Example

Enter a comma-separated list containing SNP ID, Gene ID, and tissue name, e.g.: rs509556,ENSG00000248746.1,Muscle_Skeletal

rs12203592,IRF4,Whole_Blood

Calculate Your Own

Test Your Own SNP Gene Associations Results

Data Source: GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2)

Note: NES are the effect sizes in a normalized space where **MAGNITUDE HAS NO DIRECT BIOLOGICAL INTERPRETATION**. Allelic fold-change (aFC) is available for the top variant in the egene table.

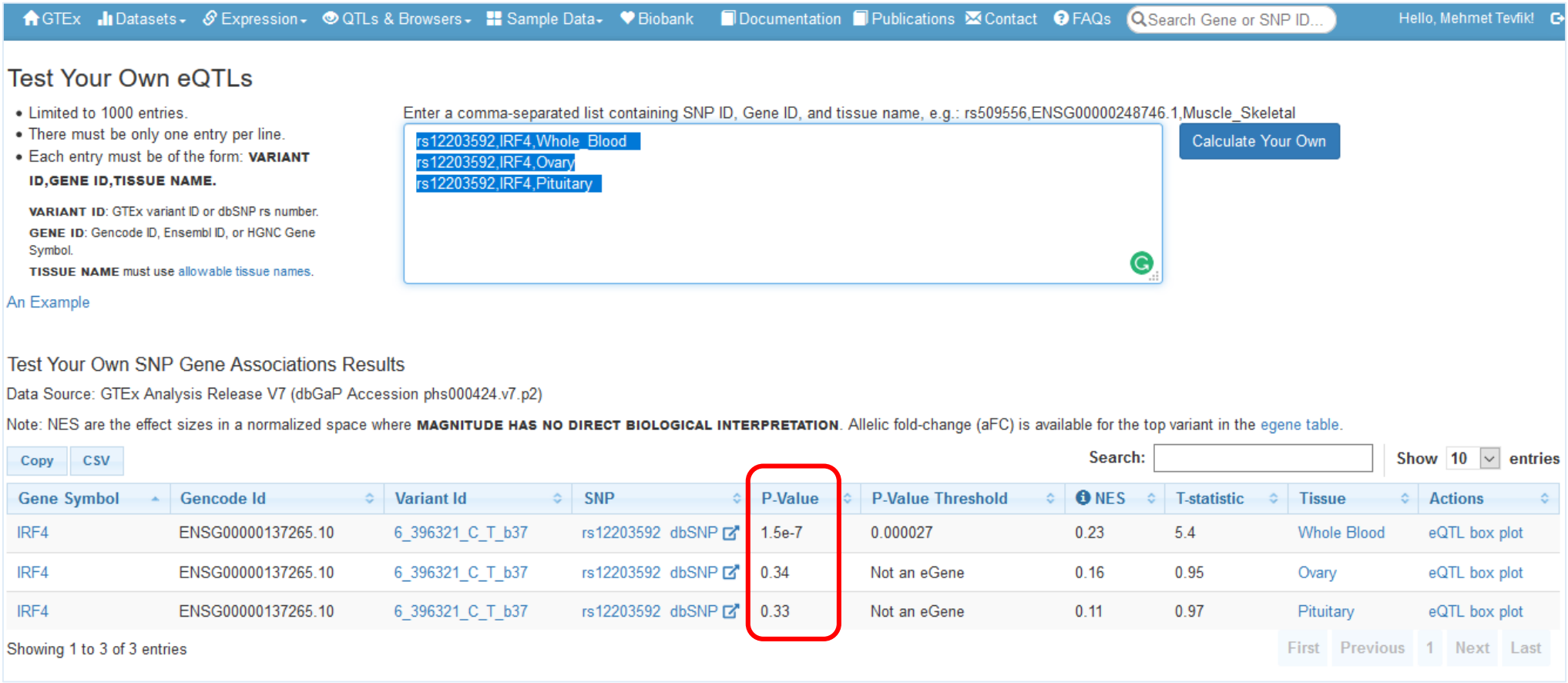
Copy | CSV | Search: | Show 10 entries

Gene Symbol	Gencode Id	Variant Id	SNP	P-Value	P-Value Threshold	NES	T-statistic	Tissue	Actions
IRF4	ENSG00000137265.10	6_396321_C_T_b37	rs12203592 dbSNP	1.5e-7	0.000027	0.23	5.4	Whole Blood	eQTL box plot

Kingston University London

eQTL effects

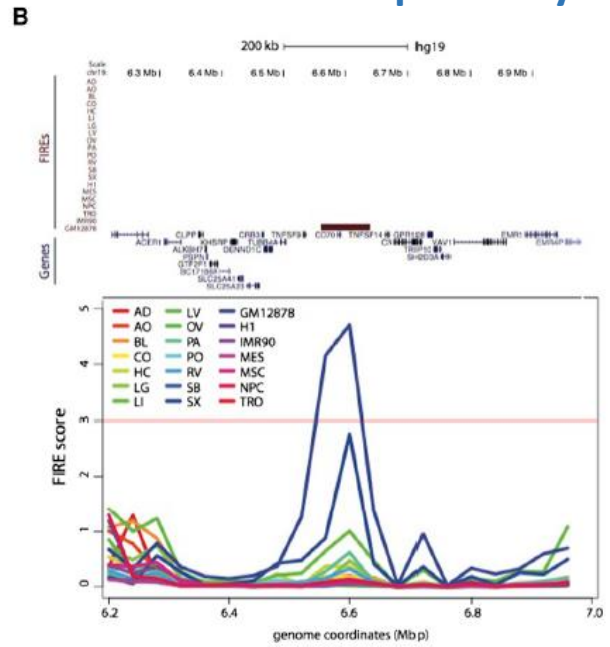
rs12203592,IRF4,Whole_Blood

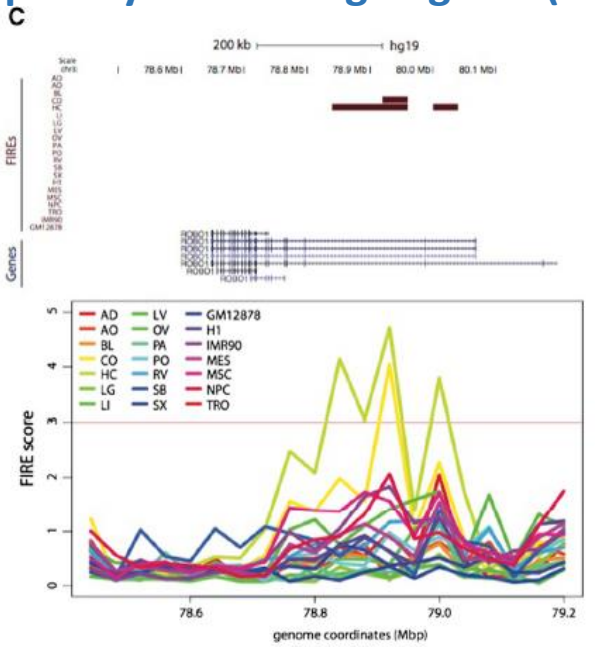


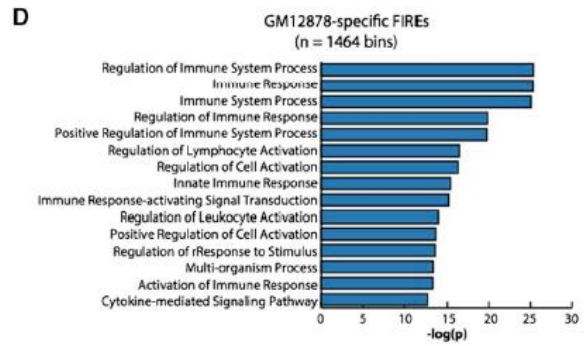
Kingston University London

Chromatin Interactions

Tissue Specificity of Frequently Interacting Regions (FIREs)







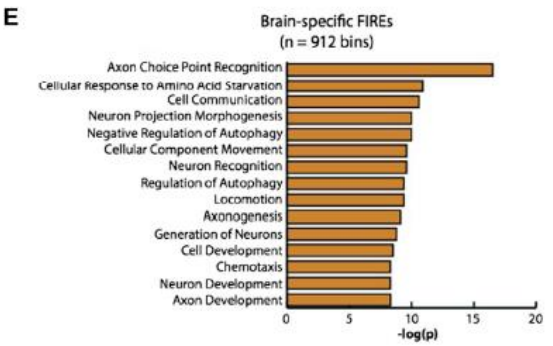


Figure 3. FIREs Are Tissue-type Specific and Enriched Near Genes Involved in Tissue Function

(A) At the top is a dendrogram resulting from a hierarchical clustering analysis using genome-wide FIRE scores for each sample. The y axis is the Euclidean distance between FIRE scores from any two samples. The heat map below shows a subset of FIRE bins (n = 8,371), corresponding to FIRE bins that are called as FIRE in only one or two samples. For ventricle tissues, brain tissues, IMR90/MSC, and H1/MES, FIREs specific to two samples are allowed in the definition of sample specific.

(B) Genome browser snapshot showing a GM12878-specific FIRE region (chr19:6,560,000–6,640,000) (top, maroon) in an 800-kb region around *CD70* (chr19:6,583,193–6,604,114). Below is a line plot of FIRE scores for each sample, showing the GM12878-specific FIRE peak (blue).

(C) Genome browser snapshot showing a brain-specific FIRE region (chr3:78,920,000–78,960,000), shared by CO and HC, in a 760-kb region within *ROBO1* (chr3:78,646,338–79,068,609). Below is a line plot of FIRE scores for each tissue showing CO (yellow) and HC (pea green) FIRE peaks.

(D) GREAT biological process analysis of genes surrounding GM12878-specific FIRE bins (n = 1,464 bins), showing biological processes highly related to immune functions. Plotted values are the −log10 of the Bonferroni-corrected binomial p values.

(E) Same as (D), except using genes surrounding brain (CO and HC) specific FIRE bins (n = 912 FIRE bins) showing several significant processes highly related to brain functionality. Plotted values are the −log10 of the Bonferroni-corrected binomial p values.

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

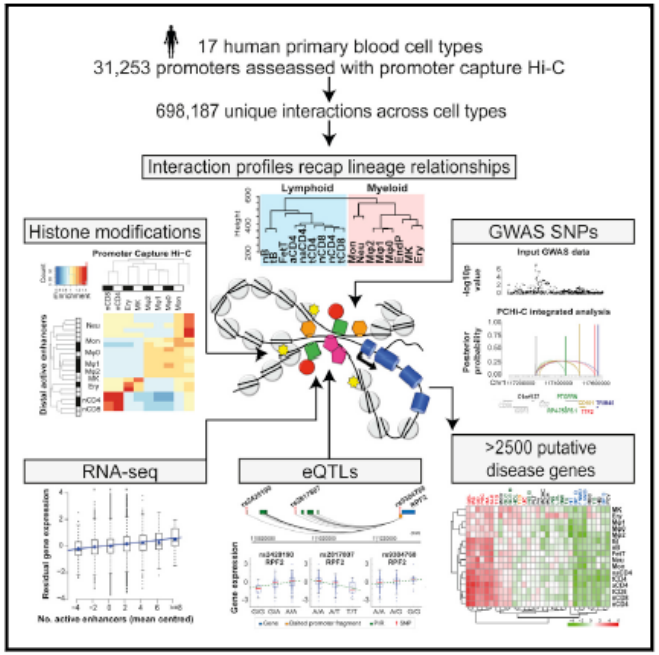
Anthony D. Schmitt,[1,2,12,13] Ming Hu,[3,12,14,*] Inkyung Jung,[1,15] Zheng Xu,[4,10,11] Yunjiang Qiu,[1,5] Catherine L. Tan,[1,13] Yun Li,[4] Shin Lin,[6] Yiing Lin,[7] Cathy L. Barr,[8] and Bing Ren[1,9,16,*]

Kingston University London

Chromatin Interactions

Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters

Graphical Abstract



Authors

Biola M. Javierre, Oliver S. Burren, Steven P. Wilder, ..., Chris Wallace, Mikhail Spivakov, Peter Fraser

Correspondence

mf471@cam.ac.uk (M.F.), cew54@medschl.cam.ac.uk (C.W.), mikhail.spivakov@babraham.ac.uk (M.S.), peter.fraser@babraham.ac.uk (P.F.)

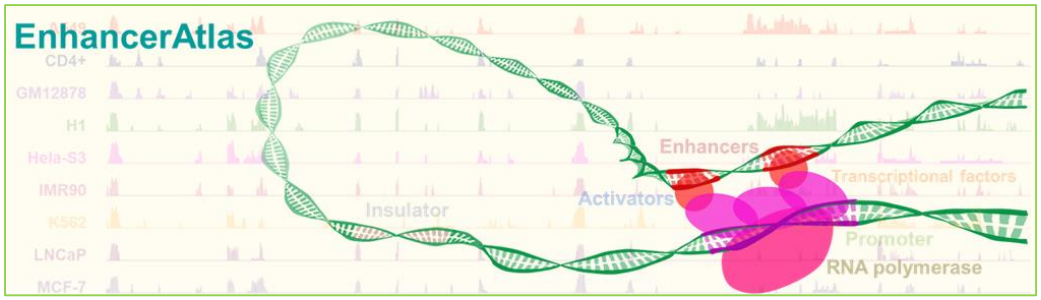
In Brief

This study deploys a promoter capture Hi-C approach in 17 primary blood cell types to match collaborating regulatory regions and identify genes regulated by noncoding disease-associated variants. Explore this and other papers at the Cell Press IHEC webportal at http://www.cell.com/consortium/IHEC.

Highlights

- High-resolution maps of promoter interactions in 17 human primary blood cell types
- Interaction patterns are cell type specific and segregate with the hematopoietic tree
- Promoter-interacting regions enriched for regulatory chromatin features and eQTLs
- Promoter interactions link non-coding GWAS variants with putative target genes

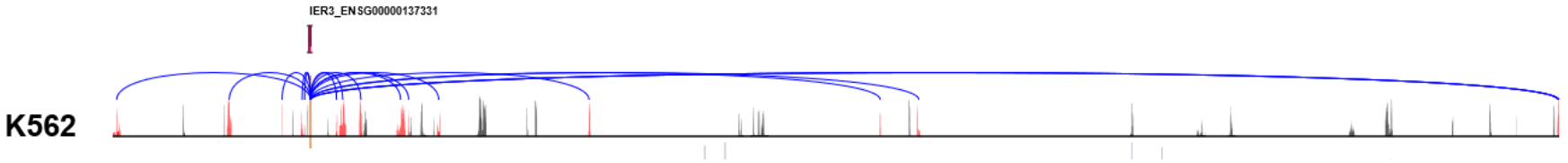
Kingston University London



EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types

Tianshun Gao [1,#], Bing He [5,6,#], Sheng Liu [1], Heng Zhu [2,3,4], Kai Tan [5,6,7,*] and Jiang Qian [1,4,*]

chr6: **30562310 bp ~ 31658310 bp**



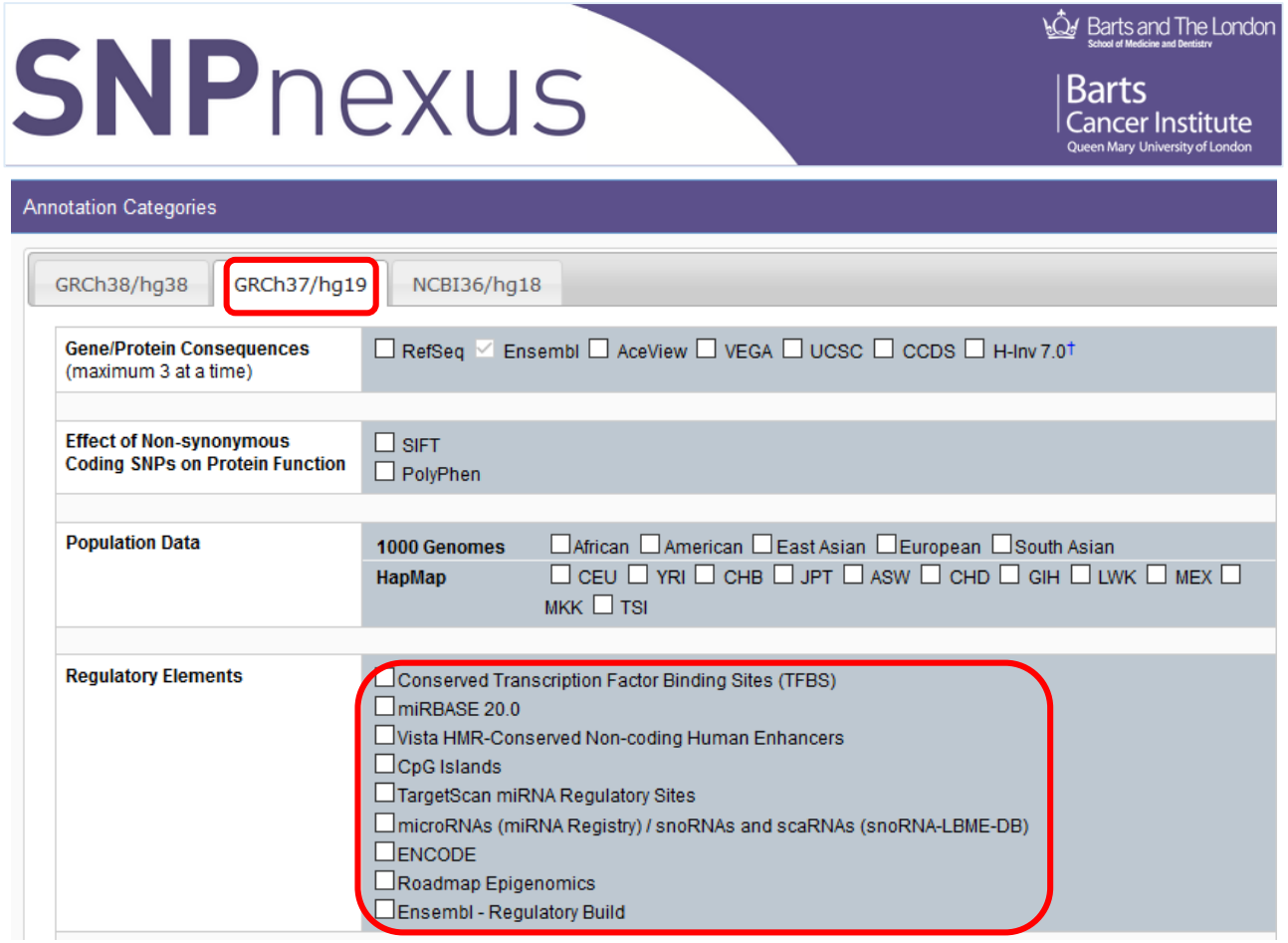
22 enhancers found to be associated with ENSG00000137331. Note that some enhancers are associated with different transcripts.

Enhancer ID	Enhancer Position	Transcript ID	Confidence Score	Conserved/specific
E34-10734	chr6:30562310-30568170	ENST00000259874	0.795	☆
E34-10734	chr6:30562310-30568170	ENST00000376377	0.795	☆
E34-10737	chr6:30648900-30652180	ENST00000259874	0.805	☆
E34-10738	chr6:30690460-30690690	ENST00000259874	0.91	☆
E34-10741	chr6:30704810-30706760	ENST00000259874	0.746	☆
E34-10741	chr6:30704810-30706760	ENST00000376377	0.746	☆
E34-10742	chr6:30707470-30707930	ENST00000259874	0.942	☆
E34-10742	chr6:30707470-30707930	ENST00000376377	0.942	☆
E34-10745	chr6:30731280-30732390	ENST00000259874	0.899	☆
E34-10747	chr6:30749160-30751580	ENST00000259874	0.961	☆
E34-10747	chr6:30749160-30751580	ENST00000376377	0.961	☆
E34-10749	chr6:30776790-30784040	ENST00000259874	0.757	☆
E34-10754	chr6:30808220-30810450	ENST00000259874	0.929	☆
E34-10754	chr6:30808220-30810450	ENST00000376377	0.929	☆
E34-10762	chr6:30922410-30924200	ENST00000259874	0.805	☆
E34-10765	chr6:31143560-31144270	ENST00000259874	0.714	☆
E34-10774	chr6:31656900-31657350	ENST00000259874	0.822	☆
E34-10775	chr6:31657480-31657850	ENST00000259874	0.823	☆
E34-10776	chr6:31657900-31658310	ENST00000259874	0.82	☆
E34-10746	chr6:30734130-30739040	ENST00000376377	0.838	☆

Kingston University London

Chromatin Marks

Tissue-specificity of regulatory disruption



Kingston University London

Results: http://snp-nexus.org/test/snpnexus_34909/results.html (2/3 April 2018 only)

Chromatin Marks

Tissue-specificity of regulatory disruption

SNPnexus

Barts and The London School of Medicine and Dentistry
Barts Cancer Institute
Queen Mary University of London

	ID	SNP	Chr	Region Start	Region End	Feature Type	Cell	Evidence
2	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	A549	true
3	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	DND-41	true
4	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	GM12878	true
5	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	H1ESC	false
6	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	HMEC	false
7	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	HSMM	true
8	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	HSMMtube	true
9	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	HUVEC	true
10	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	HeLa-S3	true
11	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	HepG2	true
12	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	IMR90	false
13	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	K562	true
14	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	Monocytes-CD14+	true
15	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	MultiCell	true
16	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	NH-A	false
17	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	NHDF-AD	true
18	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	NHEK	true
19	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	NHLF	false
20	chr6:32432077:G/C:1	rs9268905	chr6	32431919	32432518	CTCF Binding Site	Osteobl	true
21	chr6:32432181:A/G:1	rs9268906	chr6	32431919	32432518	CTCF Binding Site	A549	true
22	chr6:32432181:A/G:1	rs9268906	chr6	32431919	32432518	CTCF Binding Site	DND-41	true
23	chr6:32432181:A/G:1	rs9268906	chr6	32431919	32432518	CTCF Binding Site	GM12878	true
24	chr6:32432181:A/G:1	rs9268906	chr6	32431919	32432518	CTCF Binding Site	H1ESC	false
25	chr6:32432181:A/G:1	rs9268906	chr6	32431919	32432518	CTCF Binding Site	HMEC	false
26	chr6:32432181:A/G:1	rs9268906	chr6	32431919	32432518	CTCF Binding Site	HSMM	true
27	chr6:32432181:A/G:1	rs9268906	chr6	32431919	32432518	CTCF Binding Site	HSMMtube	true
28	chr6:32432181:A/G:1	rs9268906	chr6	32431919	32432518	CTCF Binding Site	HUVEC	true
29	chr6:32432181:A/G:1	rs9268906	chr6	32431919	32432518	CTCF Binding Site	HeLa-S3	true
30	chr6:32432181:A/G:1	rs9268906	chr6	32431919	32432518	CTCF Binding Site	HepG2	true
31	chr6:32432181:A/G:1	rs9268906	chr6	32431919	32432518	CTCF Binding Site	IMR90	false

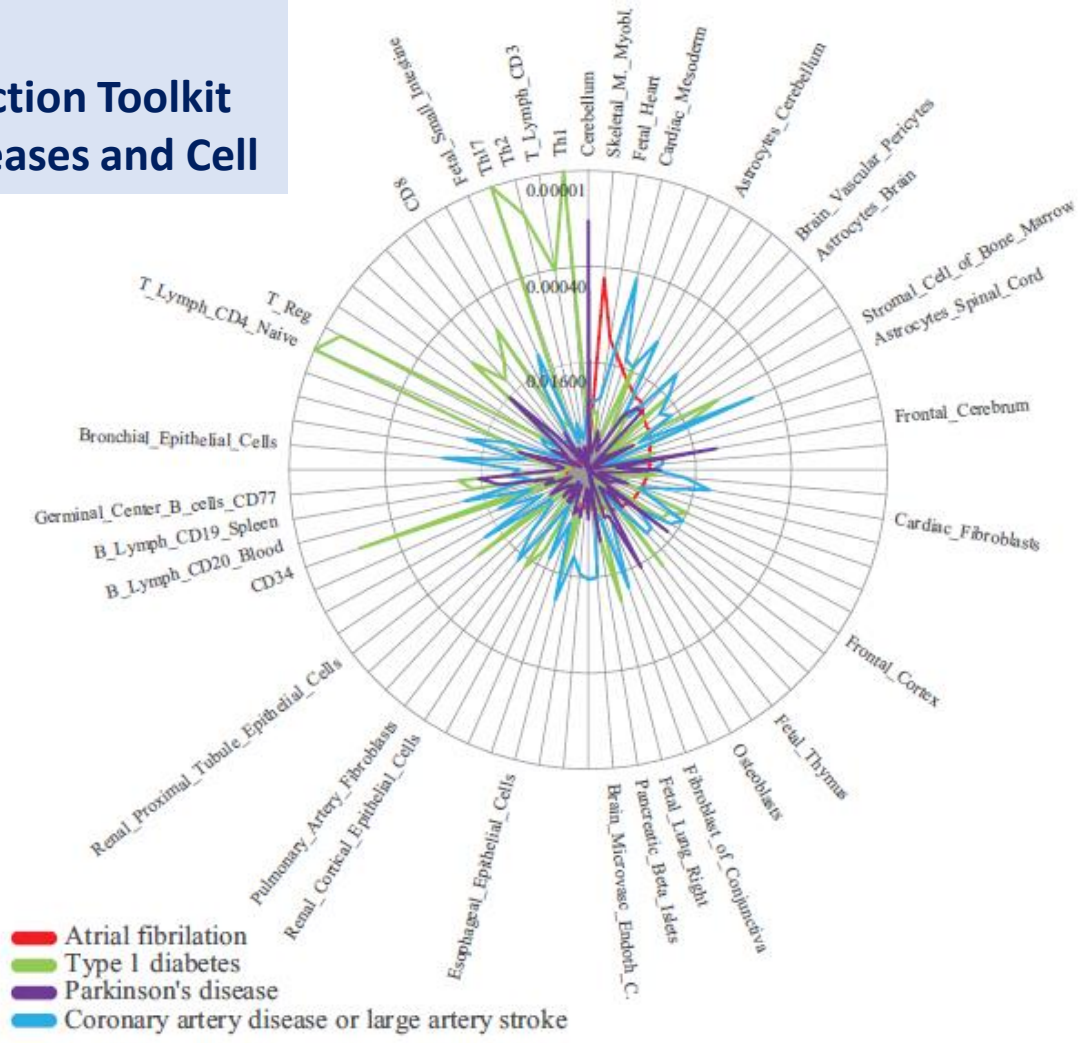
... | miRNA_snoRNA_scaRNA | ENCODE | Roadmap Epigenomics | **Regulatory Build (Ensembl)**

Kingston University London

Results: http://snp-nexus.org/test/snpnexus_34909/results.html (2/3 April 2018 only)

GARLIC

GARLIC:
GWAS-based Prediction Toolkit for Connecting Diseases and Cell



Kingston University London

Figure 2. GARLIC can be used to aetiologically connect human complex diseases and cell type-specific CRE maps. **(A)** GARLIC results obtained for a selected subset of diseases (rows) and CRE maps (columns) are shown as a heat map. The statistical connection between each disease and cell type is color coded according to GARLIC P-values, with the most and least significant connections represented in red and blue, respectively. **(B)** Radial plot summarizing the statistical connection between four selected diseases (indicated in the bottom left corner) and all cell types included in the GARLIC DB. The name of only a subset of all the investigated cell types is shown. Peaks closer to the outer border of the radial plot represent more significant connections, while those closer to the center are the least significant ones.


OXFORD
Human Molecular Genetics, 2017, Vol. 26, No. 4 742–752
doi: 10.1093/hmg/ddw423
Advance Access Publication Date: 22 December 2016
Original Article

ORIGINAL ARTICLE

GARLIC: a bioinformatic toolkit for aetiologically connecting diseases and cell type-specific regulatory maps

Miloš Nikolić[1,2], Argyris Papantonis[1] and Alvaro Rada-Iglesias[1,2,*]

GARLIC

Welcome to GARLIC Viewer!

This GARLIC version contains updated database from GWAS SNPs from 13.03.2017. and it has been converted to Hg19. Previos database version(s) can be found here.

Show values for all diseases and traits ☐

Show values for all cell types ☐

DNAse-seq regulatory map: -- Please select cell type --

Disease/Trait name: -- Please select disease or trait --

Show Clear

To begin, choose a cell type and/or a disease/trait using the option controls above

Kingston University London

GARLIC

Kingston University London

GARLIC: a bioinformatic toolkit for aetiologically connecting diseases and cell type-specific regulatory maps

Miloš Nikolić; Argyris Papantonis; Alvaro Rada-Iglesias

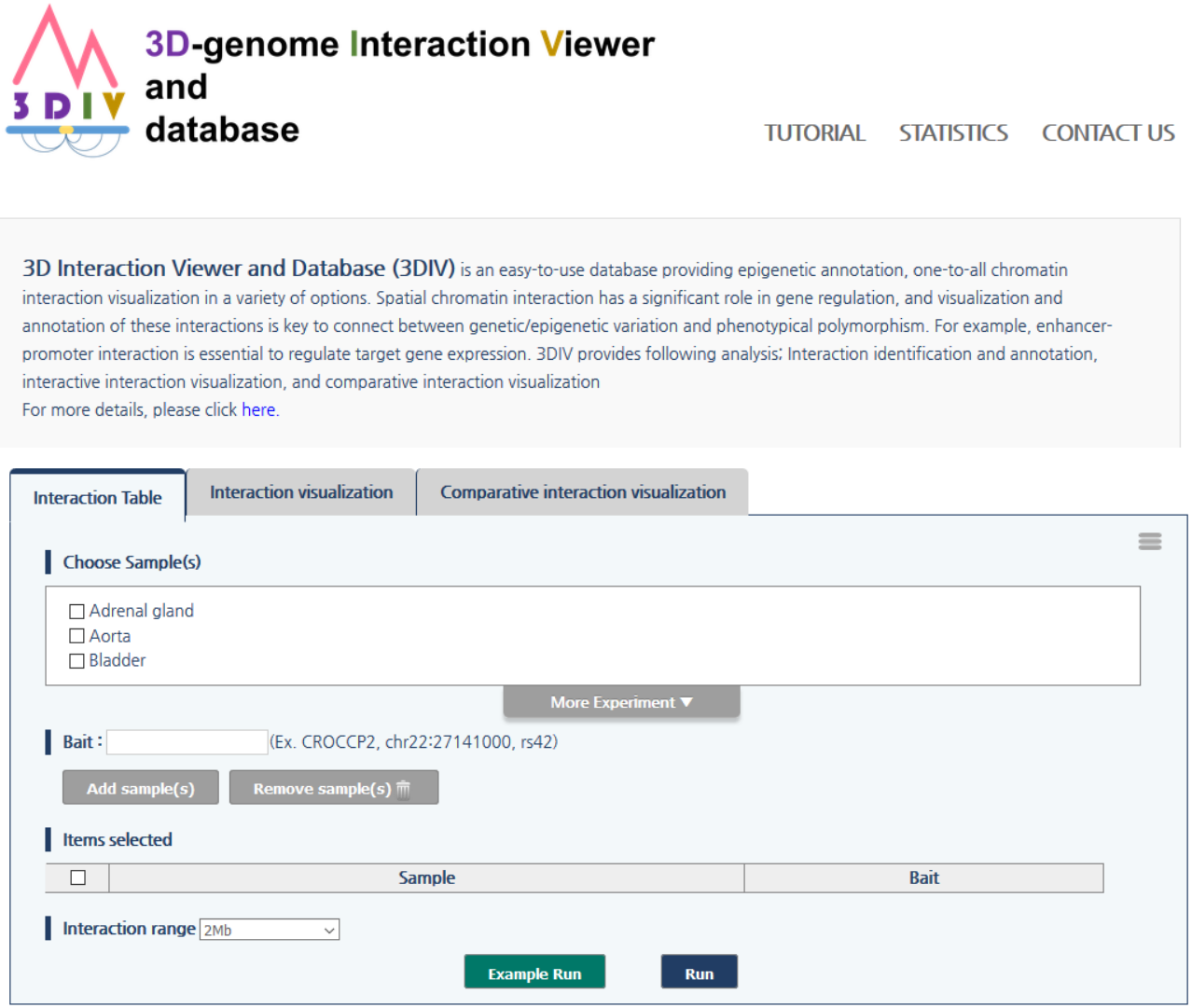
GARLIC

Kingston University London

GARLIC: a bioinformatic toolkit for aetiologically connecting diseases and cell type-specific regulatory maps

Miloš Nikolić; Argyris Papantonis; Alvaro Rada-Iglesias

Chromatin Interaction



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,*]

Enhancers for Each Gene

DATABASE
The Journal of Biological Databases and Curation

Database, 2017, 1–17
doi: 10.1093/database/bax028
Original article

Original article

GeneHancer: genome-wide integration of enhancers and target genes in GeneCards

Simon Fishilevich[1,†], Ron Nudel[1,†], Noa Rappaport[1], Rotem Hadar[1], Inbar Plaschkes[1], Tsippi Iny Stein[1], Naomi Rosen[1], Asher Kohn[2], Michal Twik[1], Marilyn Safran[1], Doron Lancet[1,*] and Dana Cohen[1,*]

Kingston University London

cepip

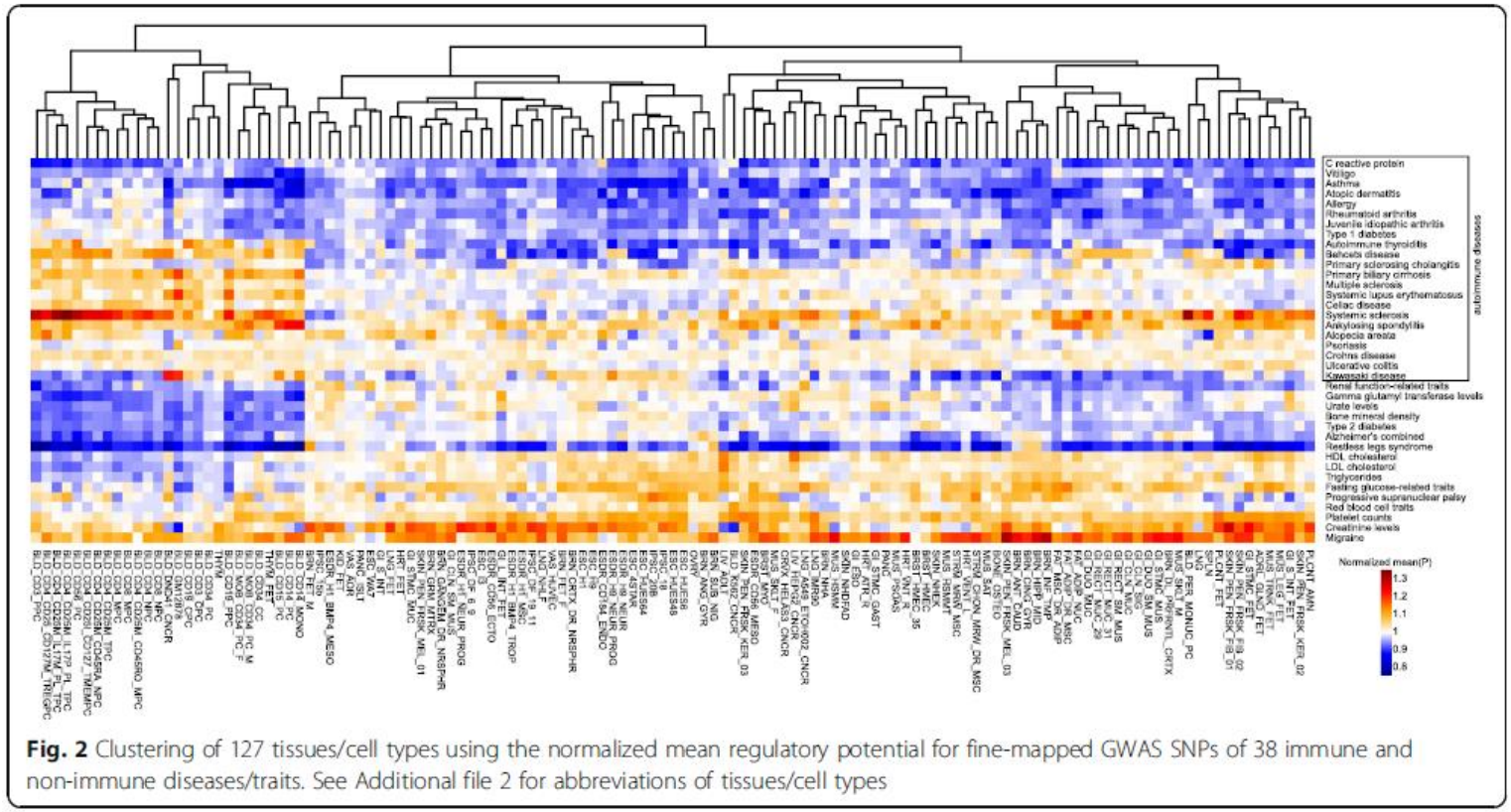


Fig. 2 Clustering of 127 tissues/cell types using the normalized mean regulatory potential for fine-mapped GWAS SNPs of 38 immune and non-immune diseases/traits. See Additional file 2 for abbreviations of tissues/cell types

cepip: Context-dependent epigenomic weighting for regulatory variant prioritization

References
Download
Installation
Example
Prioritization
Perl Version
FAQ?

Introduction

Majority of trait/disease associated variants identified by genome wide association studies (GWASs) locate in the regulatory regions. Since gene regul... highly context-specific, it remains challenging to fine-map and prioritize functional regulatory variants in a particular cell/tissue type and apply them to disease-associated genes detection. By connecting large-scale epigenome profiles to expression quantitative trait loci (eQTLs) in a wide range of human tissues/cell types, we identify combination of several critical chromatin features that predict variant regulatory potential. We develop a joint likelihood framework to measure the regulatory probability of genetic variants in a context-dependent manner. We show our model is superior to existing cell type-specific methods and exhibit significant GWAS signal enrichment. Using phenotypically relevant epigenomes to weight GWAS SNPs, we discover more disease-associated genes owing to regulatory changes and improve the statistical power in gene-based association test.

Li et al. Genome Biology (2017) 18:52
DOI 10.1186/s13059-017-1177-3

Genome Biology

METHOD Open Access

cepip: context-dependent epigenomic weighting for prioritization of regulatory variants and disease-associated genes

Mulin Jun Li[1,3,8*], Miaoxin Li[2,3,6,7*], Zipeng Liu[3,4*], Bin Yan[3,5], Zhicheng Pan[3,8], Dandan Huang[1], Qian Liang[1], Dingge Ying[3], Feng Xu[3,5], Hongcheng Yao[3,5], Panwen Wang[10], Jean-Pierre A. Kocher[10], Zhengyuan Xia[4], Pak Chung Sham[3,8], Jun S. Liu[9] and Junwen Wang[10,11*]

Kingston University London

Cell-specificity of Mutations

Kingston University London

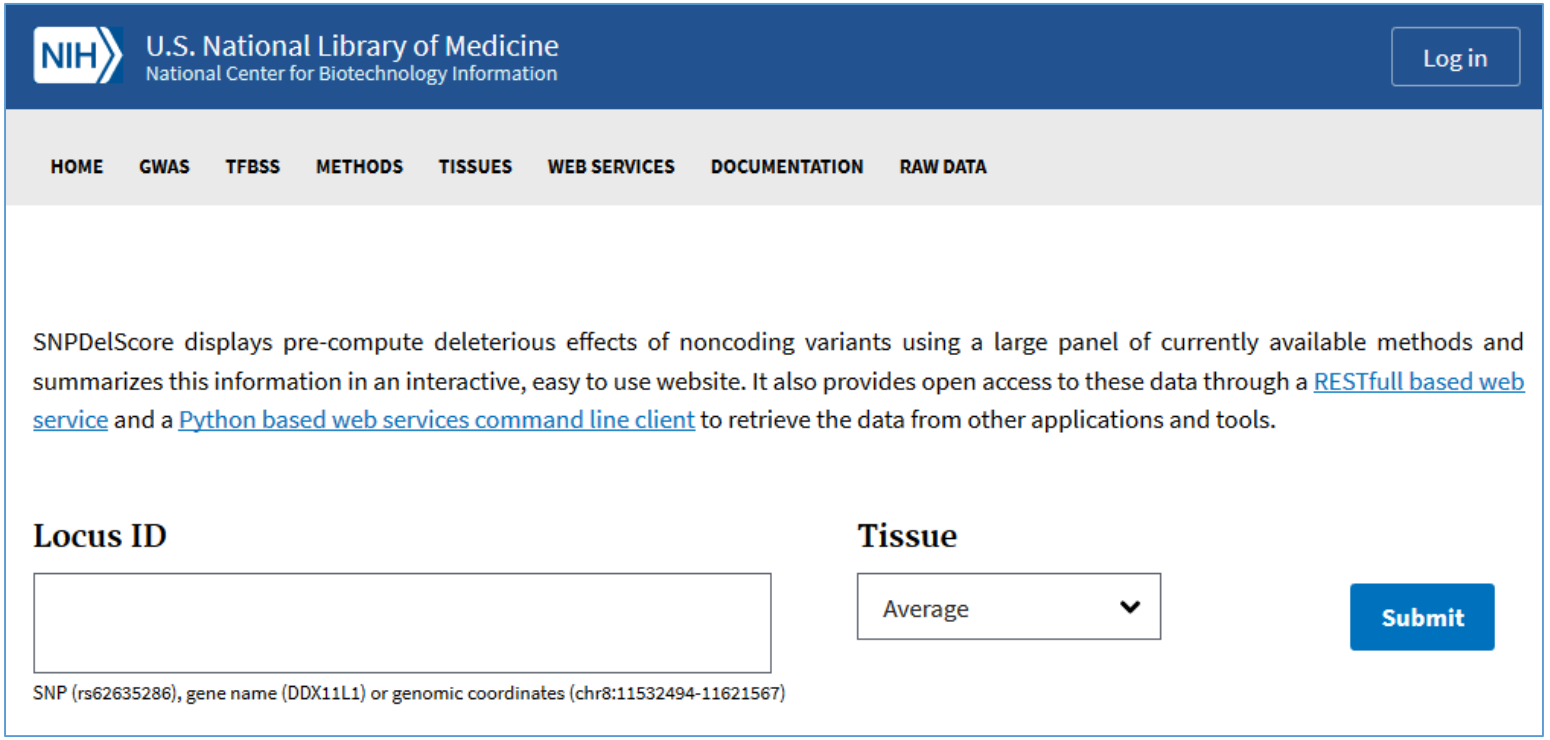
Published online 15 December 2016 *Nucleic Acids Research, 2017, Vol. 45, No. 5 2307–2317*
doi: 10.1093/nar/gkw1263

Quantifying deleterious effects of regulatory variants

Shan Li[1], Roberto Vera Alvarez[1], Roded Sharan[2], David Landsman[1] and Ivan Ovcharenko[1,*]

SNP Functionality Scores

Aggregate Functionality Scores
SNPDelScore



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*

Kingston University London

SNP Functionality Scores

Aggregate Functionality Scores
SNPDelScore

rs2395185

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

GRCh37.p13, chr6:32433167
Reference: G
Alternative: T
Location: intergenic

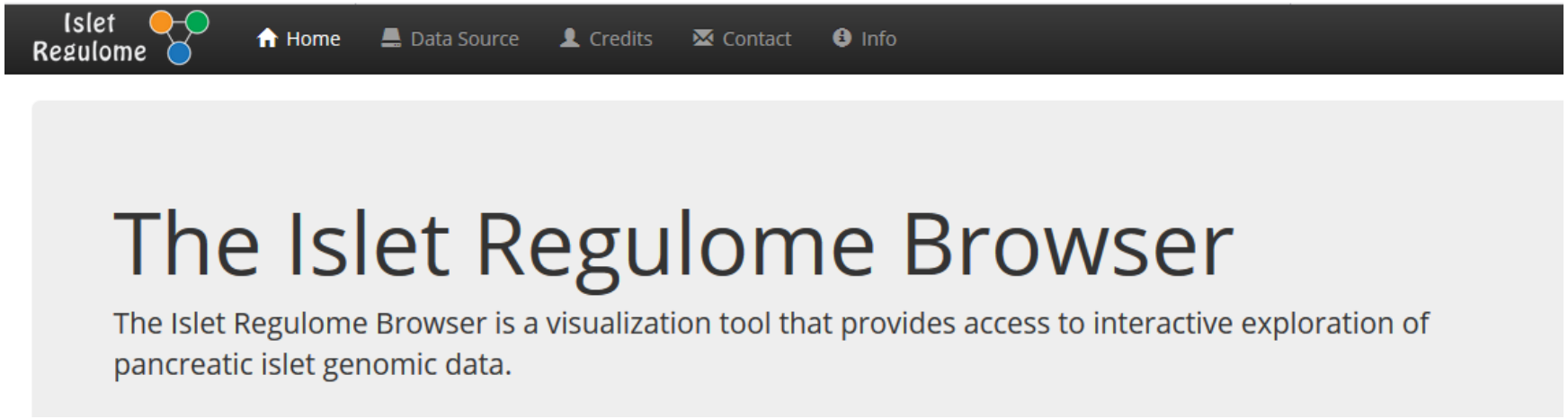
Deleterious SNP scores

All methods have been normalized between [0,1]

Tissue	Average	CAPE eQTL	CAPE dsQTL	deltaSVM	CATO	DeepSEA	CADD	LINSIGHT	P-Causal
Average	0.07	0.228	0.078	0.069	N/A	0	0.004	0.04	0.304
H1 Cells	0.146	0.126	0.476	-0.017	N/A	-0.001	N/A	N/A	N/A
H1 BMP4 Derived Mesendoderm Cultured Cells	0.059	0.098	0.168	-0.09	N/A	N/A	N/A	N/A	N/A
H1 BMP4 Derived Trophoblast Cultured Cells	0.137	0.217	0.057	N/A	N/A	N/A	N/A	N/A	N/A
H1 Derived Mesenchymal Stem Cells	0.087	0.094	0.089	0.078	N/A	N/A	N/A	N/A	N/A
H1 Derived Neuronal Progenitor Cultured Cells	0.159	0.249	0.089	0.138	N/A	N/A	N/A	N/A	N/A
H9 Cells	0.119	0.273	0.083	N/A	N/A	-0.001	N/A	N/A	N/A
IMR90 fetal lung fibroblasts Cell Line	0.077	0.076	0.078	N/A	N/A	N/A	N/A	N/A	N/A
iPS DF 6.9 Cells	0.155	0.392	0.072	N/A	N/A	-0	N/A	N/A	N/A
iPS DF 19.11 Cells	0.11	0.298	0.144	-0.111	N/A	N/A	N/A	N/A	N/A
Primary monocytes from peripheral blood	0.103	0.22	0.045	0.044	N/A	N/A	N/A	N/A	N/A
Primary B cells from peripheral blood	0.113	0.252	0.063	0.024	N/A	N/A	N/A	N/A	N/A
Primary T cells from peripheral blood	0.141	0.31	0.051	0.061	N/A	N/A	N/A	N/A	N/A
Primary Natural Killer cells from peripheral blood	0.151	0.253	0.049	N/A	N/A	N/A	N/A	N/A	N/A
Primary hematopoietic stem cells G-CSF-mobilized Female	0.101	0.308	0.04	-0.044	N/A	N/A	N/A	N/A	N/A

Kingston University London

Tissue-specific Tools

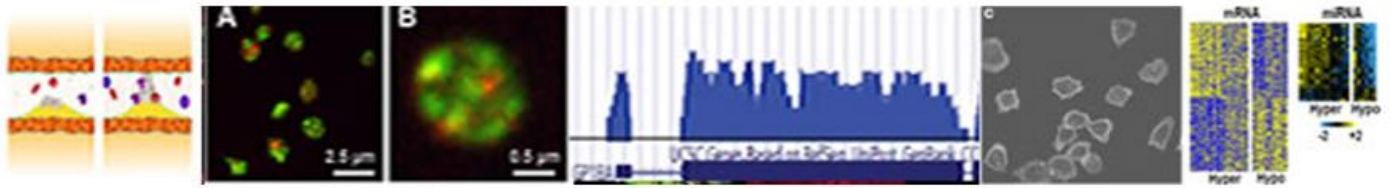


Kingston University London

Tissue-specific Tools

PLATELETOMICS

Interactive Results Tool version: 1.0



RNA Query

To query eQTL results please click HERE.

Note: Messenger RNAs were profiled using the Affymetrix GeneChip® Human Gene 1.0 ST Array. MicroRNAs were profiled using nCounter human miRNA assay kit v1 (Nanostring Technologies, Seattle, WA). Click [Here] to find your gene symbol or alternate official symbol. Click [Here] for a more complete description regarding limitations of this query .

(1) The Affymetrix microarray typically used more than one set of probes (Transcriptcluster ID) to quantify transcripts of a given gene; (2) many genes have multiple symbols and aliases, which may be a source of confusion; for comprehensive searching capacity, we list them all.

If you use information from this website, please refer to the following:

Human platelet microRNA-mRNA networks associated with age and gender revealed by integrated plateletomics. Simon, L. et al. Blood. doi: 10.1182/blood-2013-12-544692 (2014)

Search by Gene

Submit Reset

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