Bioinformatics Tools for HLA

Mehmet Tevfik DORAK, MD PhD

School of Life Sciences, Pharmacy & Chemistry
Kingston University London

8th National Congress of Turkish Transplantation Immunology and Genetics Society *[TIGED2022]*
Antalya, 2022

Kingston University London

Outline

HLA and its unique features

Genome-wide bioinformatics tools

HLA-specific bioinformatics tools

HLA: Unique Features

Genome (3.2Gb)	xHLA Region (25.7 to 33.4Mb)	Comparison
Total No of Genes 60155	**Total No of Genes 674**	...
Protein-coding genes 19881	**Protein-coding genes 453**	32.7 vs 67.2% *P*<0.0001
Non-coding RNA Genes 25411	**Non-coding RNA Genes 54**	42.6 vs 8.0% *P*<0.0001
Long non-coding RNA genes 15877	**Long non-coding RNA genes 13**	26.4 vs 1.9% *P*<0.0001
Small non-coding RNA genes 9534	**Small non-coding RNA genes 7**	15.9 vs 1.0% *P*<0.0001
Pseudogenes 14467	**Pseudogenes 172**	24.0 vs 25.5% *P* = 0.37

xHLA makes up 0.24% of the genome, but contains 0.40% of all SNPs in the human genome

A systematic analysis of the gene and variation content of the extended HLA region

Ertan KANBUR, Mustafa DOGAN, Mehmet Tevfik DORAK

Presented at EFI 2017

Kingston University London

HLA: Unique Features

Heritability of DNA methylation

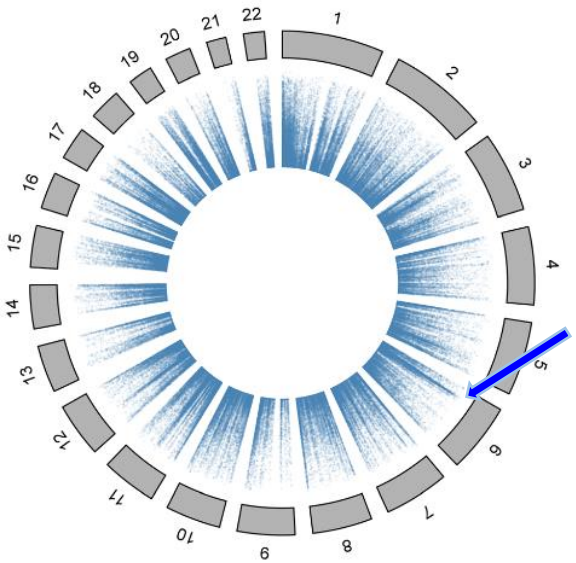


Figure 2 Distribution of genetic heritability estimates across the genome. The MHC region, which had the highest estimates of genetic heritability is clearly visible on chromosome 6. Telomeric regions show an increased density of probes with high genetic heritability, although this is primarily due to higher numbers of probes in these regions.

McRae *et al. Genome Biology* 2014, **15**:R73
http://genomebiology.com/2014/15/5/R73

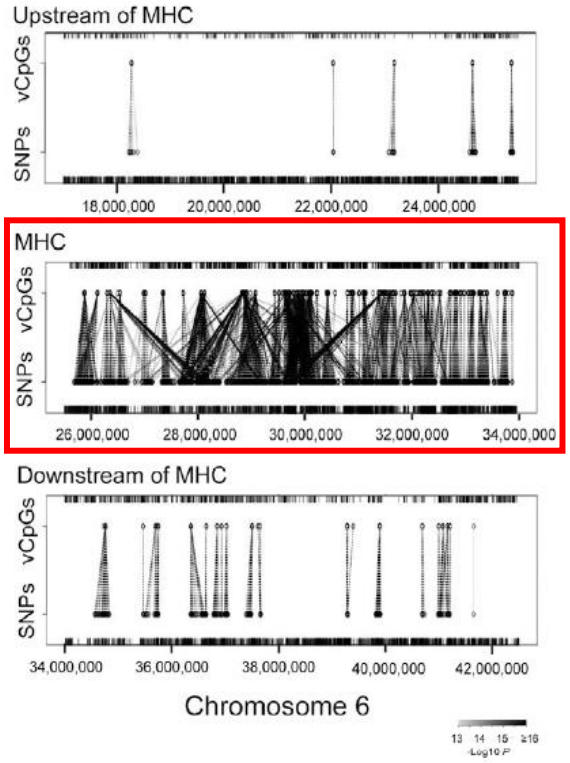
Genome Biology

RESEARCH Open Access

Contribution of genetic variation to transgenerational inheritance of DNA methylation

Allan F McRae[1,2*], Joseph E Powell[1,2], Anjali K Henders[3], Lisa Bowdler[3], Gibran Hemani[1,2], Sonia Shah[1,2], Jodie N Painter[3], Nicholas G Martin[3], Peter M Visscher[1,2†] and Grant W Montgomery[3†]

meQTL density



(E) Associations between CpG sites and SNPs upstream of (top panel), within (middle panel), or downstream of (bottom panel) the major histocompatibility complex (MHC) region. Each dashed line represents a significant association, and the shades of black indicate significance of the associations.

GeMes, Clusters of DNA Methylation under Genetic Control, Can Inform Genetic and Epigenetic Analysis of Disease

Yun Liu,[1,2,9] Xin Li,[1,2,9] Martin J. Aryee,[1,3] Tomas J. Ekström,[4,5] Leonid Padyukov,[4,6] Lars Klareskog,[4,6] Amy Vandiver,[1,2] Ann Zenobia Moore,[7] Toshiko Tanaka,[7] Luigi Ferrucci,[7] M. Daniele Fallin,[1,8,*] and Andrew P. Feinberg[1,2,*]

trans-eQTL/meQTL density

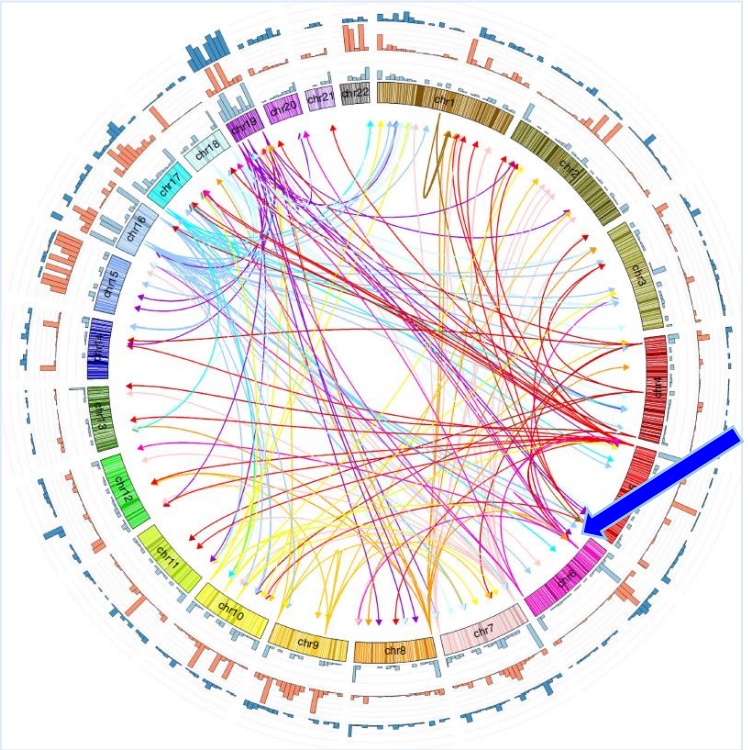


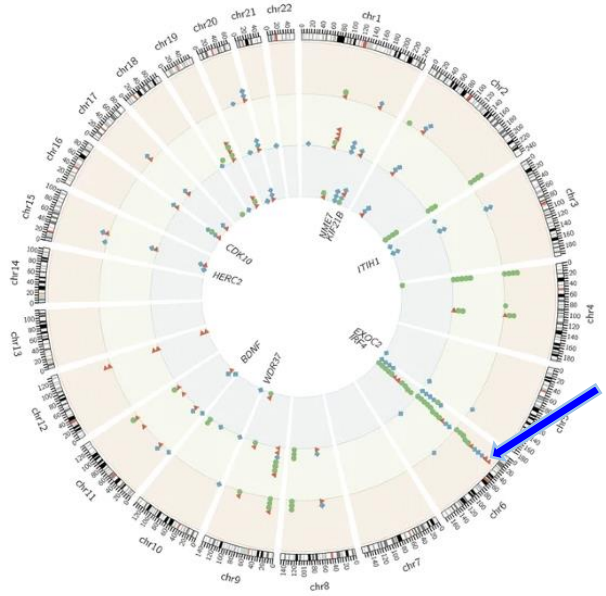
Figure 1 | Enrichment of features in regions harbouring SNPs involved in distal SNP–CpG associations. Outer histograms: number of SNPs involved in distal SNP–CpG associations (light blue), calculated in 7.5 Mb bins; number of piRNA sequences (orange); number of transcription factors (dark blue). Inner links: SNP regions associated with four or more CpG sites. Arrows are pointing from SNPs to the CpG sites they are associated with, and are coloured according to the chromosomes where the SNPs reside.

Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci

Mathieu Lemire[1], Syed H.E. Zaidi[1], Maria Ban[2], Bing Ge[3], Dylan Aïssi[4,5,6], Marine Germain[4,5,6], Irfahan Kassam[7], Mike Wang[1], Brent W. Zanke[8], France Gagnon[7], Pierre-Emmanuel Morange[9,10,11], David-Alexandre Trégouët[4,5,6], Philip S. Wells[8], Stephen Sawcer[2], Steven Gallinger[12,13], Tomi Pastinen[3] & Thomas J. Hudson[1,14,15]

Kingston University London

HLA: Unique Features



Overview of PheWAS associations in the genome after functional annotation. **a** This matrix shows the number of functional SNPs for their respective phenotype. **b** The *Circos plot* showing the PheWAS associations in different types of functional data. *Red triangles* represent the associations in the GWAS Catalog only, *green circles* represent GWAS Catalog associations replicated by PheWAS ($P < 0.05$), and *blue diamonds* represent new phenotype associations identified by PheWAS ($P < 4.6 \times 10^{-6}$ or FDR < 0.1)

including:

- **Schizophrenia**
- **Alzheimer disease**
- **Parkinson disease**
- **Lung cancer**
- **Hodgkin lymphoma**

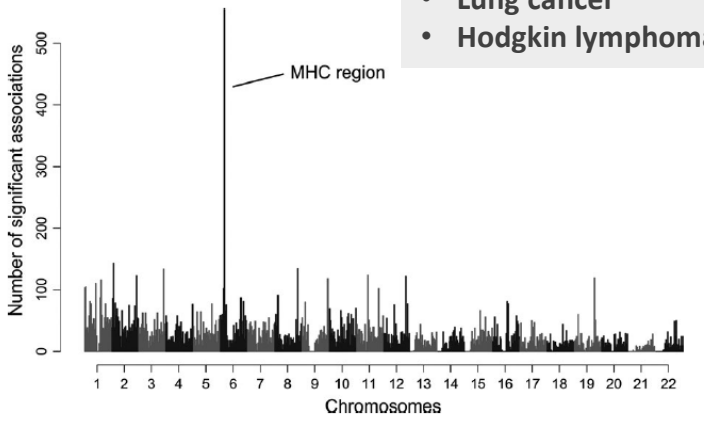


FIGURE 1 Number of significant GWAS associations along the genome. The chromosomal location of significant trait associations from GWAS (N = 18,682) is shown for all autosomes. Data from NHGRI GWAS catalog. Reproduced from "Lenz TL, Spirin V, Jordan DM, Sunyaev SR. Excess of Deleterious Mutations around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

Zhao et al. Genome Medicine (2018) 10:7
DOI 10.1186/s13073-018-0513-x

Genome Medicine

RESEARCH Open Access

An integrative functional genomics framework for effective identification of novel regulatory variants in genome–phenome studies

Junfei Zhao[1†], Feixiong Cheng[2,3†], Peilin Jia[1], Nancy Cox[4,5], Joshua C. Denny[5,6] and Zhongming Zhao[1,7*]

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]

Bioinformatics Tools: Genome-wide

GENCODE

Human | Mouse | How to access data | FAQ

Human

More about GENCODE Human

- Current human data
- Release history
- Statistics
- Data format
- FTP site

General stats

Total No of Genes	61544
Protein-coding genes	19988
Long non-coding RNA genes	18805
Small non-coding RNA genes	7567
Pseudogenes	14774

HGNC

Search symbols, keywords or IDs

Gene data | Tools | Downloads | VGNC | Contact us | More | Request symbol

Statistics & download files

Filter statistics and download files by chromosome

All chromosomes

Statistics

Locus Group	Total by Locus Group	Locus Type	Total by Locus Type
protein-coding gene	19220	gene with protein product	19220
non-coding RNA	8883	RNA, Y	4
		RNA, cluster	119
		RNA, long non-coding	5545
		RNA, micro	1912

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

Genome | Genome | Search | Limits | Advanced

Eukaryotic Annotation Home | Documentation | Annotated Genomes | Annotation Policy | Request Annotation

Eukaryotic genomes annotated at NCBI

Hundreds of eukaryotic genomes have been annotated by the NCBI Eukaryotic Genome Annotation Pipeline (see graphs). The latest annotation release available for each genome is shown in the tables below. The tables are organized by taxonomic group and provide links to the annotation report, FTP site, genome BLAST page, and Genome Data Viewer page.

Only completed annotations are shown here. Please browse the annotation runs currently in progress to see what will become available in a few days.

Show/Hide All

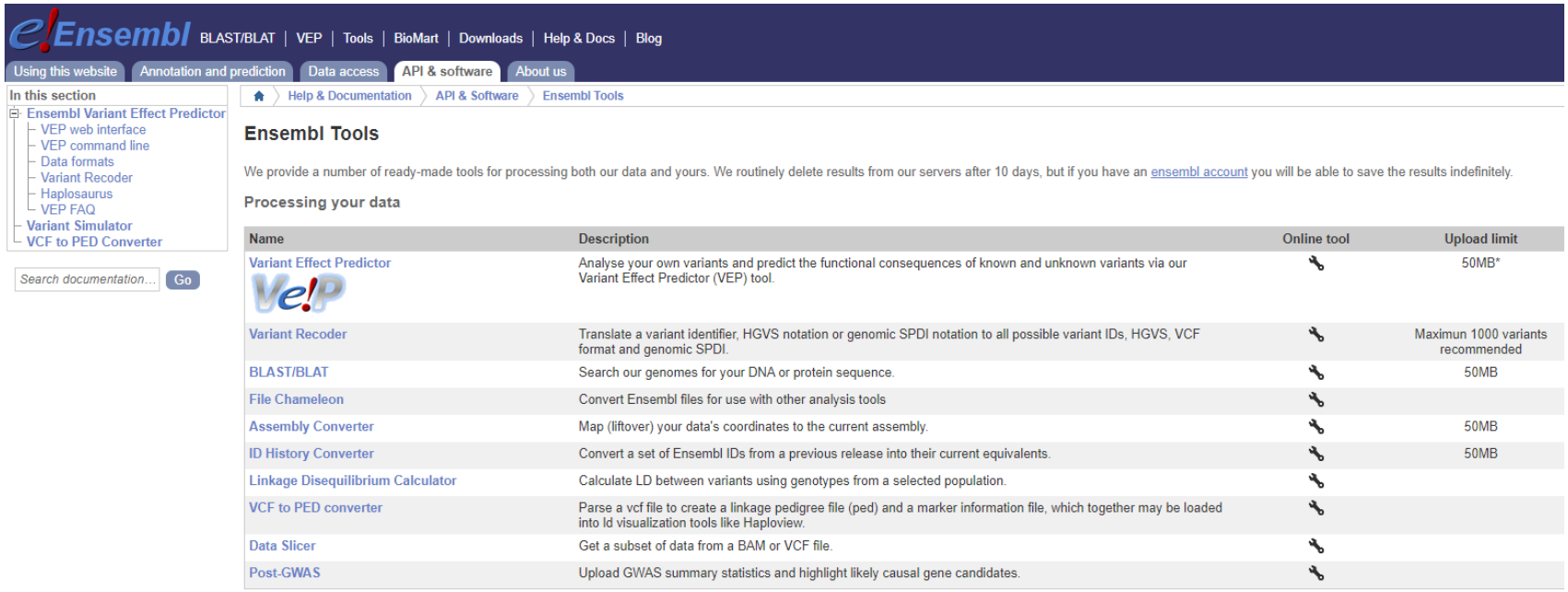
Featured (6)

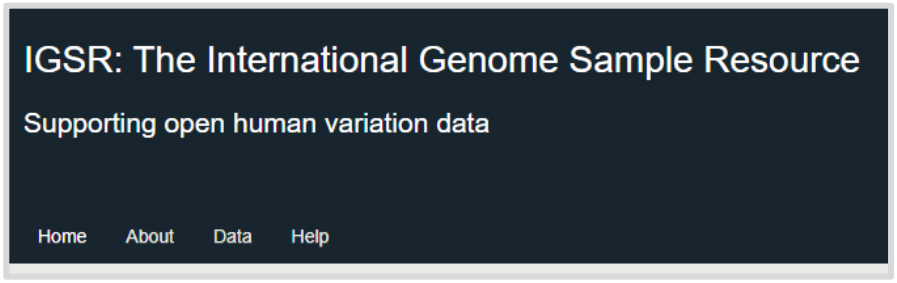
FTP - FTP Download B - Organism-specific BLAST AR - Annotation Report GDV - Genome Data Viewer

Species	RefSeq assembly(ies)	Annotation Release	Freeze Date	Release Date	Links
Homo sapiens (human)	GRCh38.p13 (GCF_000001405.39)	109.20211119	2021-11-19	2021-11-22	FTP B AR GDV
Mus musculus (house mouse)	GRCm39 (GCF_000001635.27)	109	2020-09-09	2020-09-22	FTP B AR GDV
Rattus norvegicus (Norway rat)	mRatBN7.2 (GCF_015227675.2)	108	2021-01-07	2021-01-21	FTP B AR GDV
Danio rerio (zebrafish)	GRCz11 (GCF_000002035.6)	106	2017-06-02	2017-06-26	FTP B AR GDV
Apis mellifera (honey bee)	Amel_HAv3.1 (GCF_003254395.2)	104	2018-09-13	2018-09-19	FTP B AR GDV
Zea mays (maize)	Zm-B73-REFERENCE-NAM-5.0 (GCF_902167145.1)	103	2020-08-09	2020-09-01	FTP B AR GDV

Kingston University London

Bioinformatics Tools: Genome-wide





1KG Project!

The International Genome Sample Resource

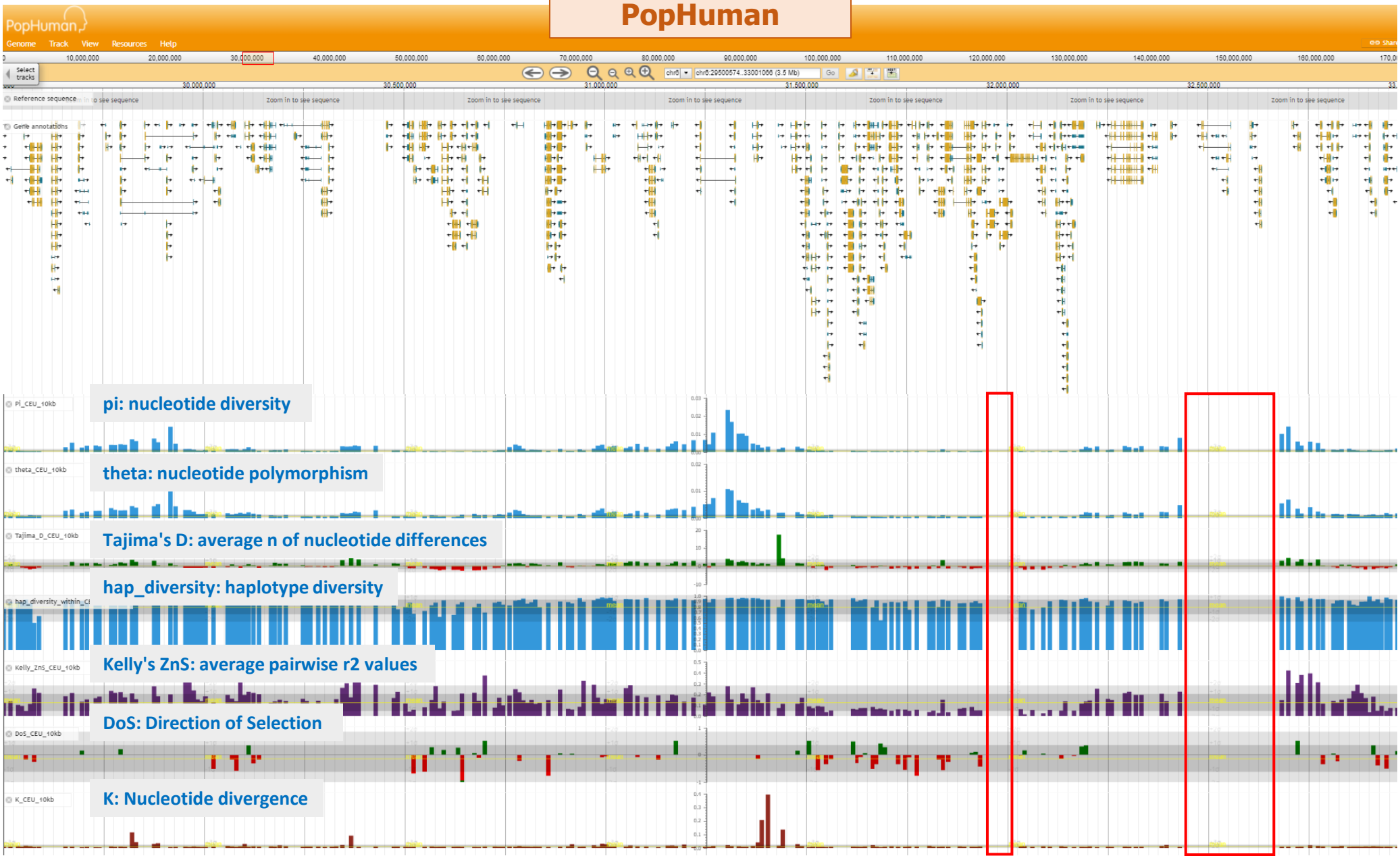
The 1000 Genomes Project created a catalogue of common human genetic variation, using openly consented samples from people who declared themselves to be healthy.

The International Genome Sample Resource (IGSR) maintains and shares the human genetic variation resources built by the 1000 Genomes Project.

Kingston University London

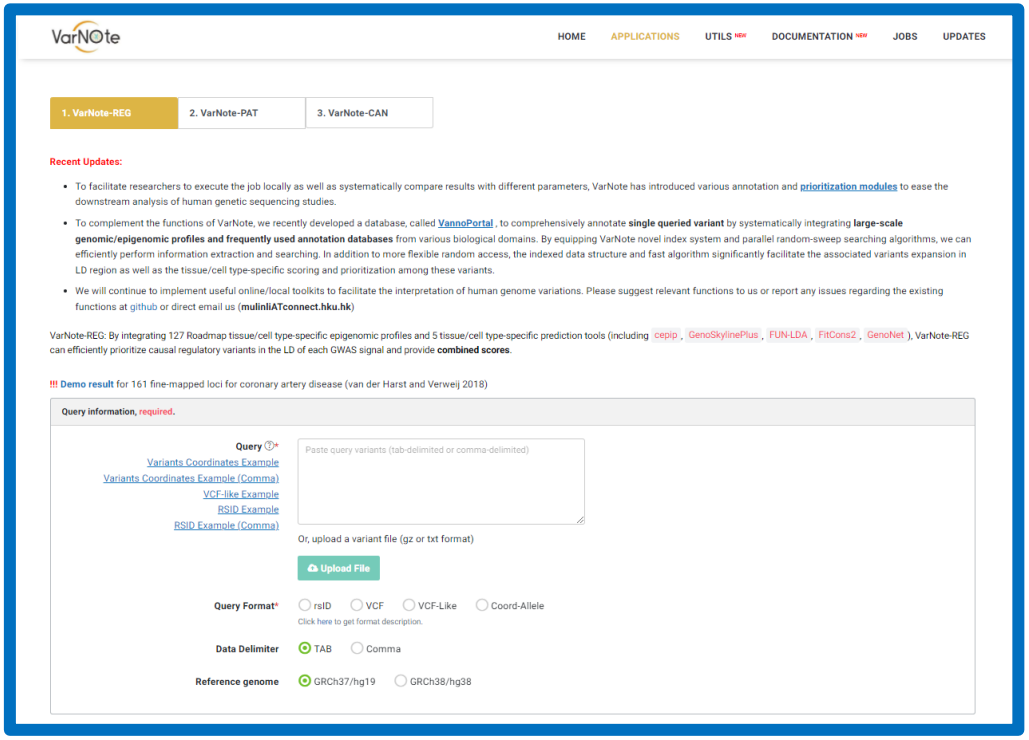
Bioinformatics Tools: Genome-wide

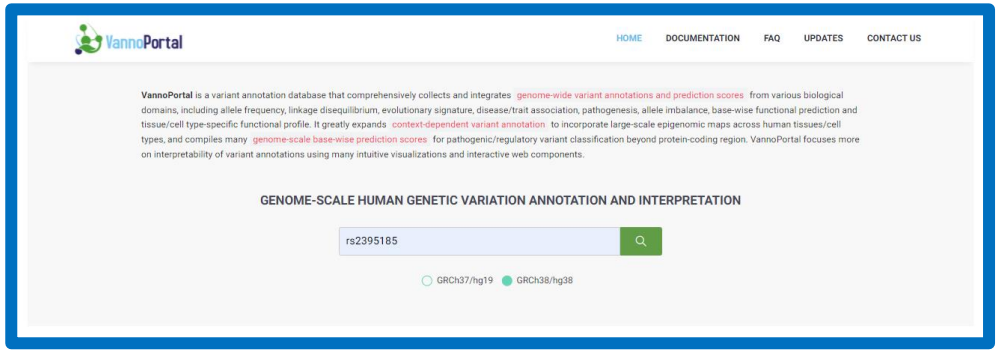
PopHuman

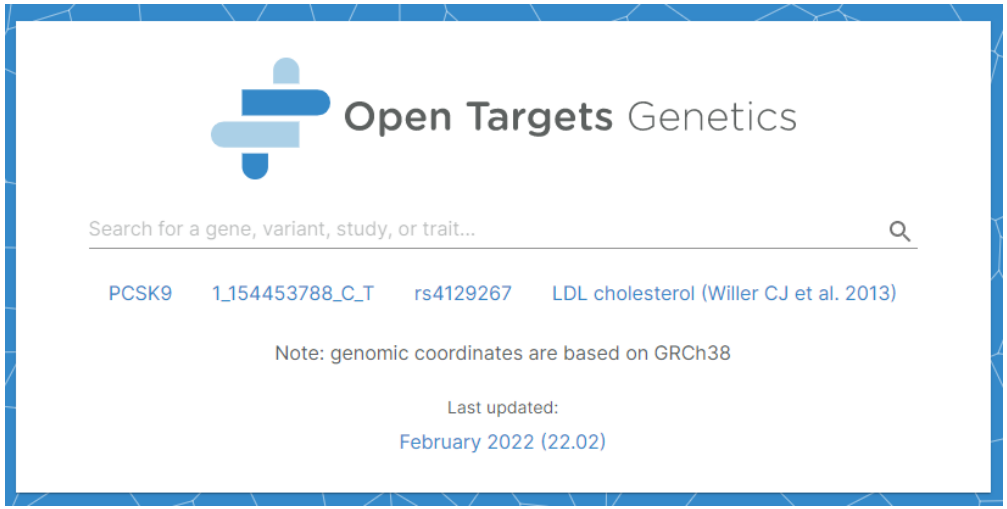


Bioinformatics Tools: Genome-wide

SNPs







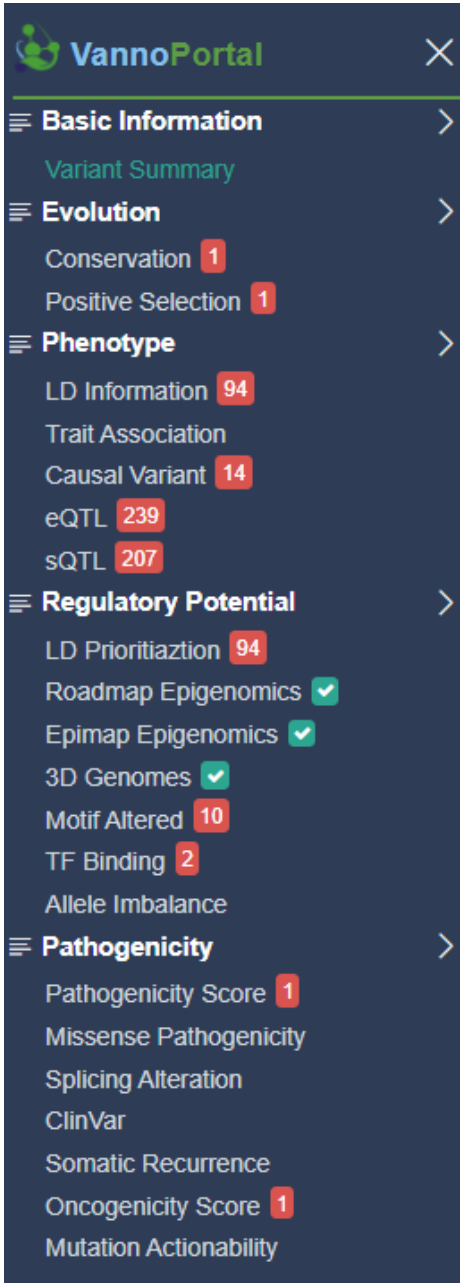
https://genetics.opentargets.org/variant/6_32465390_G_T

http://www.mulinlab.org/vportal/apir.html?q=rs2395185&g=hg38

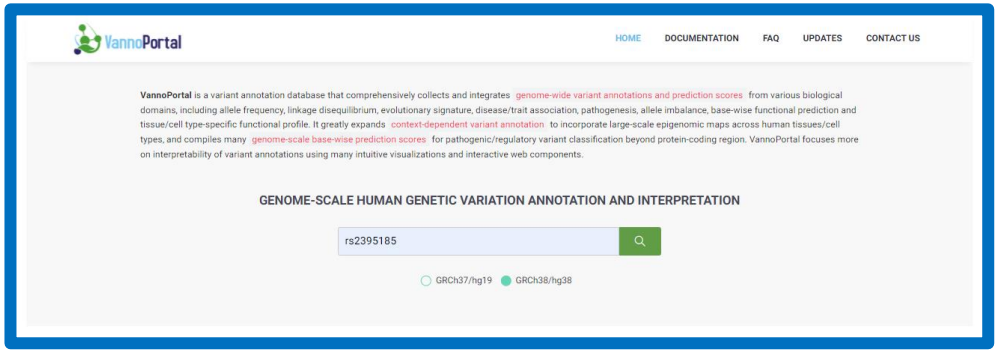
http://www.mulinlab.org/varnote/application.html#REG

Kingston University London

Bioinformatics Tools: Genome-wide



SNPs



Kingston University London

Bioinformatics Tools: Genome-wide

eQTLs

FIVEx: Visualization for Genotypes, Expressions, and Tissues

Currently hosting **eQTL** and **sQTL** data from **16 different studies** from the EBI eQTL Catalogue

Search for a variant, region, or gene: chr19:488506, rs10424907, or SHC2

eQTL sQTL

Search for: **Variant by position:** chr1:109274968 • **rsID:** rs12740374
Region: chr1:108774968-109774968 • **Gene:** *SORT1*
First time? **View the tutorial here** to see what FIVEx can do

eQTL Catalogue

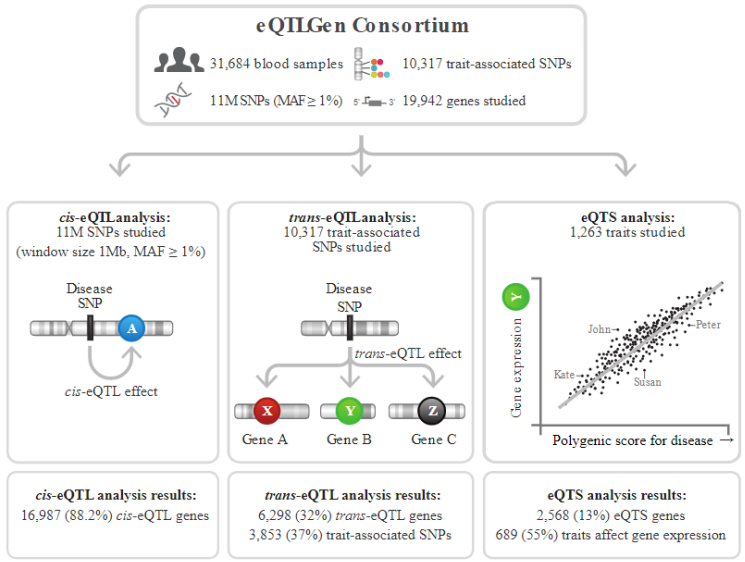
Expression and splicing QTLs recomputed from public datasets

eQTLGen

Cis-eQTLs *Trans*-eQTLs eQTS Replications Publications sc-eQTLGen

Welcome to the eQTLGen Consortium

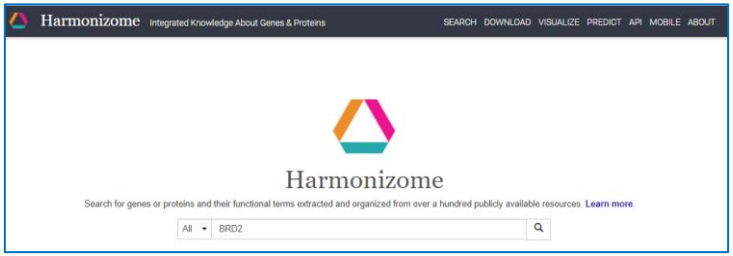
The eQTLGen Consortium has been set up to identify the downstream consequences of trait-related genetic variants. The consortium incorporates 37 datasets, with a total of 31,684 individuals. You can find the cis-eQTL, trans-eQTL, eQTS and replication results from our forthcoming paper on this website.



Kingston University London

Bioinformatics Tools: Genome-wide

Genes



Functional Associations

BRD2 has 4,992 functional associations with biological entities spanning 8 categories (molecular profile, organism, chemical, functional term, phrase or reference, disease, phenotype or trait, structural feature, cell line, cell type or tissue, gene, protein or microRNA) extracted from 80 datasets.

Click the + buttons to view associations for BRD2 from the datasets below.

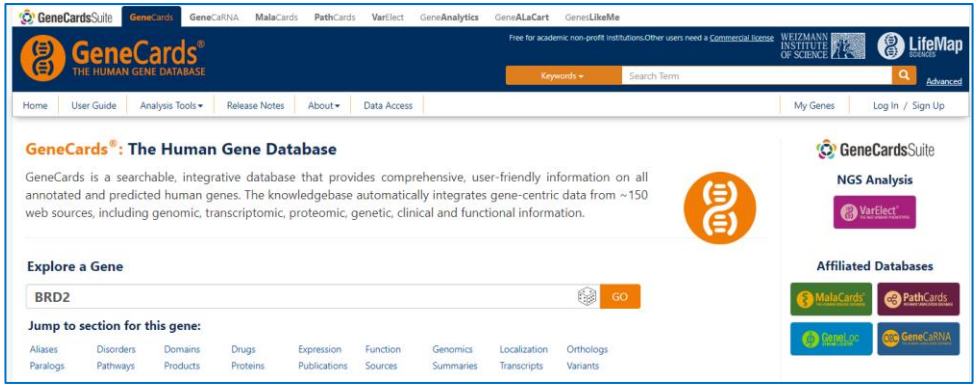
*If available, associations are ranked by **standardized value***

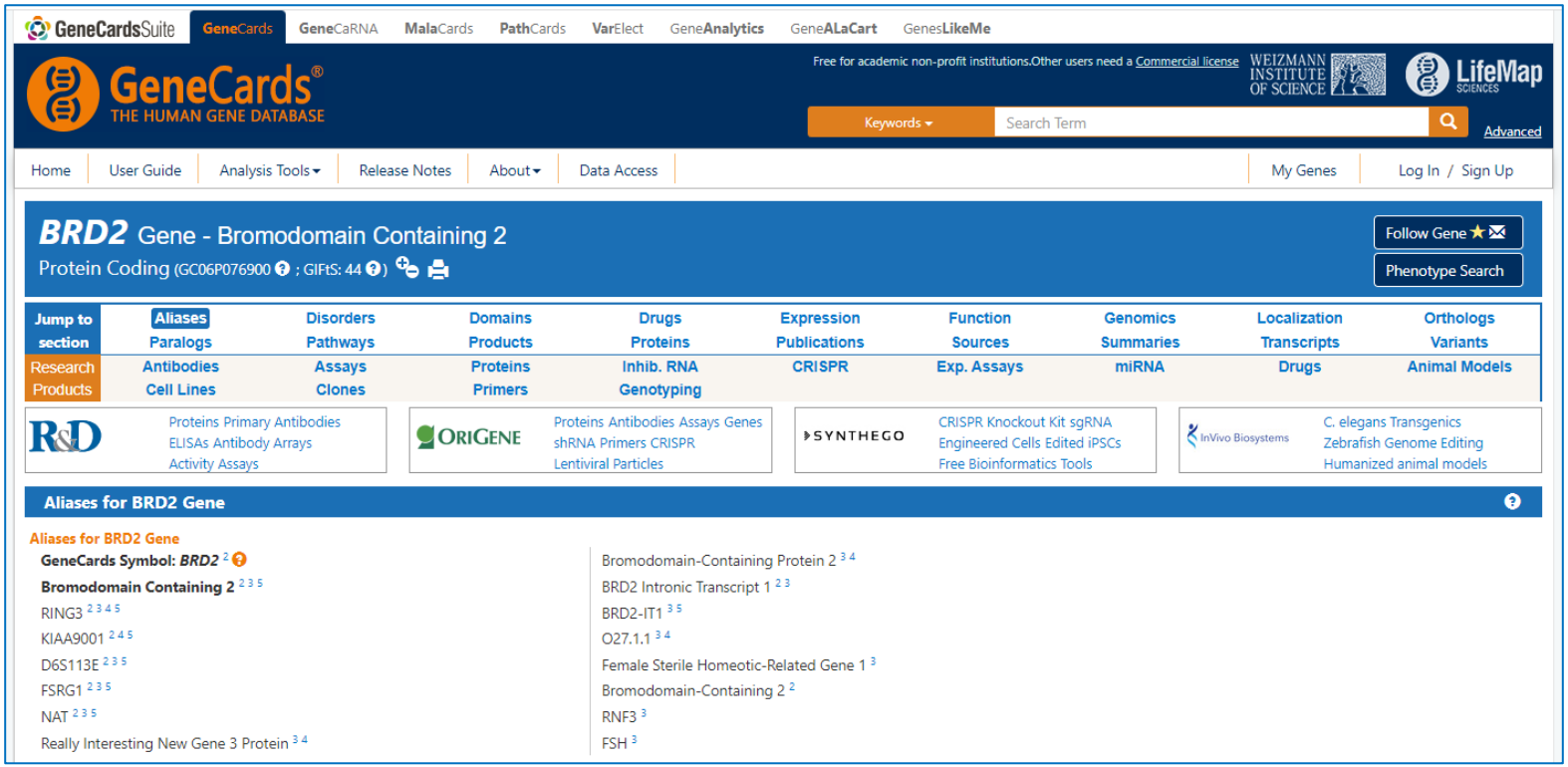
	Dataset	Summary
+	Achilles Cell Line Gene Essentiality Profiles	Cell lines with fitness changed by BRD2 gene knockdown relative to other cell lines from the Achilles Cell Line Gene Essentiality Profiles dataset.
+	Allen Brain Atlas Adult Human Brain Tissue Gene Expression Profiles	Tissues with high or low expression of BRD2 gene relative to other tissues from the Allen Brain Atlas Adult Human Brain Tissue Gene Expression Profiles dataset.
+	Allen Brain Atlas Adult Mouse Brain Tissue Gene Expression Profiles	Tissues with high or low expression of BRD2 gene relative to other tissues from the Allen Brain Atlas Adult Mouse Brain Tissue Gene Expression Profiles dataset.
+	Allen Brain Atlas Developing Human Brain Tissue Gene Expression Profiles by Microarray	Tissue samples with high or low expression of BRD2 gene relative to other tissue samples from the Allen Brain Atlas Developing Human Brain Tissue Gene Expression Profiles by Microarray dataset.
+	Allen Brain Atlas Developing Human Brain Tissue Gene Expression Profiles by RNA-seq	Tissue samples with high or low expression of BRD2 gene relative to other tissue samples from the Allen Brain Atlas Developing Human Brain Tissue Gene Expression Profiles by RNA-seq dataset.
+	Allen Brain Atlas Prenatal Human Brain Tissue Gene Expression Profiles	Tissues with high or low expression of BRD2 gene relative to other tissues from the Allen Brain Atlas Prenatal Human Brain Tissue Gene Expression Profiles dataset.
+	BioGPS Cell Line Gene Expression Profiles	Cell lines with high or low expression of BRD2 gene relative to other cell lines from the BioGPS Cell Line Gene Expression Profiles dataset.
+	BioGPS Human Cell Type and Tissue Gene Expression Profiles	Cell types and tissues with high or low expression of BRD2 gene relative to other cell types and tissues from the BioGPS Human Cell Type and Tissue Gene Expression Profiles dataset.
+	BioGPS Mouse Cell Type and Tissue Gene Expression Profiles	Cell types and tissues with high or low expression of BRD2 gene relative to other cell types and tissues from the BioGPS Mouse Cell Type and Tissue Gene Expression Profiles dataset.
+	CCLE Cell Line Gene CNV Profiles	Cell lines with high or low copy number of BRD2 gene relative to other cell lines from the CCLE Cell Line Gene CNV Profiles dataset.
+	CCLE Cell Line Gene Expression Profiles	Cell lines with high or low expression of BRD2 gene relative to other cell lines from the CCLE Cell Line Gene Expression Profiles dataset.
+	CCLE Cell Line Gene Mutation Profiles	Cell lines with BRD2 gene mutations from the CCLE Cell Line Gene Mutation Profiles dataset.
+	CHEA Transcription Factor Binding Site Profiles	Transcription factor binding site profiles with transcription factor binding evidence at the promoter of BRD2 gene from the CHEA Transcription Factor Binding Site Profiles dataset.
+	CHEA Transcription Factor Targets	Transcription factors binding the promoter of BRD2 gene in low- or high-throughput transcription factor functional studies from the CHEA Transcription Factor Targets dataset.
+	CMAP Signatures of Differentially Expressed Genes for Small Molecules	Small molecule perturbations changing expression of BRD2 gene from the CMAP Signatures of Differentially Expressed Genes for Small Molecules dataset.
+	COMPARTMENTS Curated Protein Localization Evidence Scores	Cellular components containing BRD2 protein from the COMPARTMENTS Curated Protein Localization Evidence Scores dataset.
+	COMPARTMENTS Text-mining Protein Localization Evidence Scores	Cellular components co-occuring with BRD2 protein in abstracts of biomedical publications from the COMPARTMENTS Text-mining Protein Localization Evidence Scores dataset.
+	COSMIC Cell Line Gene Mutation Profiles	Cell lines with BRD2 gene mutations from the COSMIC Cell Line Gene Mutation Profiles dataset.

Kingston University London

Bioinformatics Tools: Genome-wide

Genes





Kingston University London

Bioinformatics Tools: HLA-specific


TISSUE ANTIGENS
IMMUNE RESPONSE GENETICS

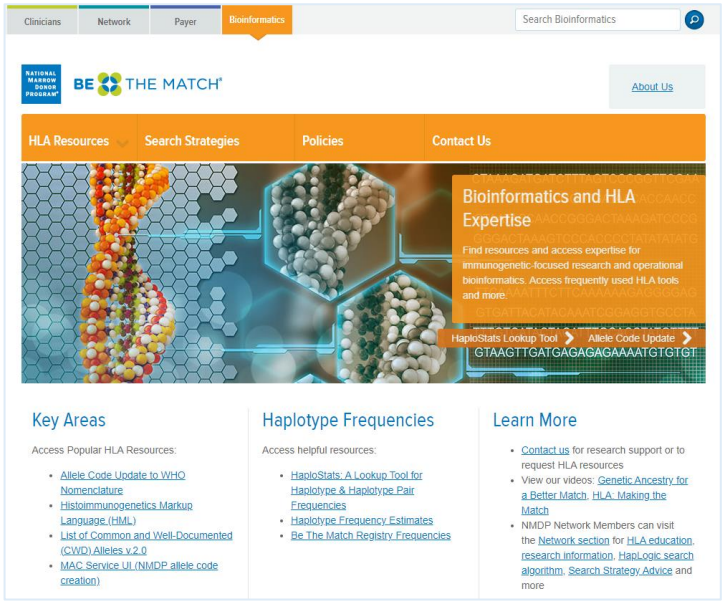
Tissue Antigens ISSN 0001-2815

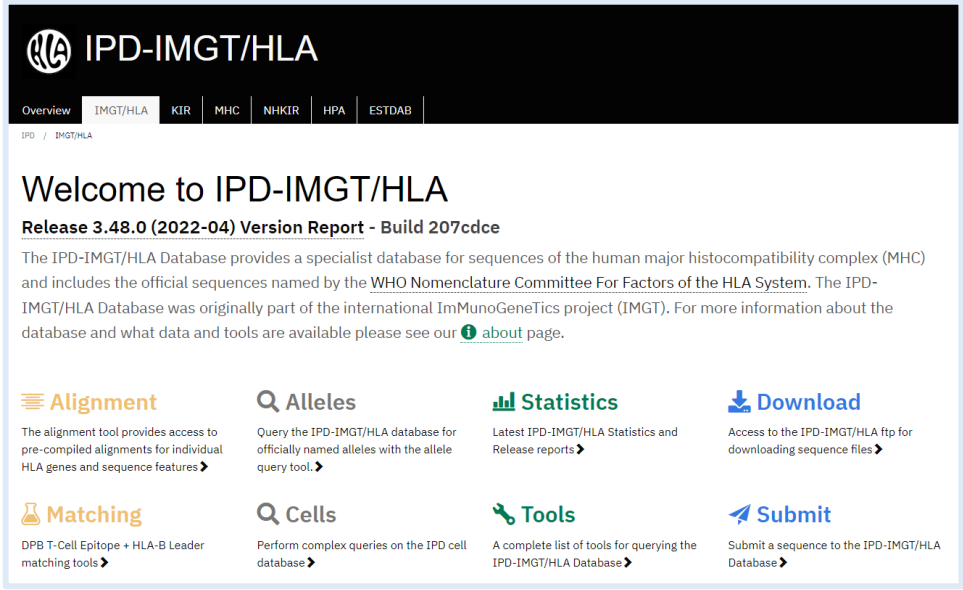
REVIEW ARTICLE

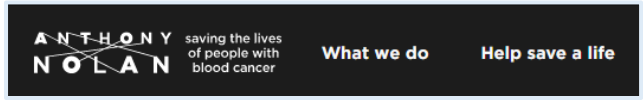
Bioinformatic Databases and Resources in the public domain to aid HLA research

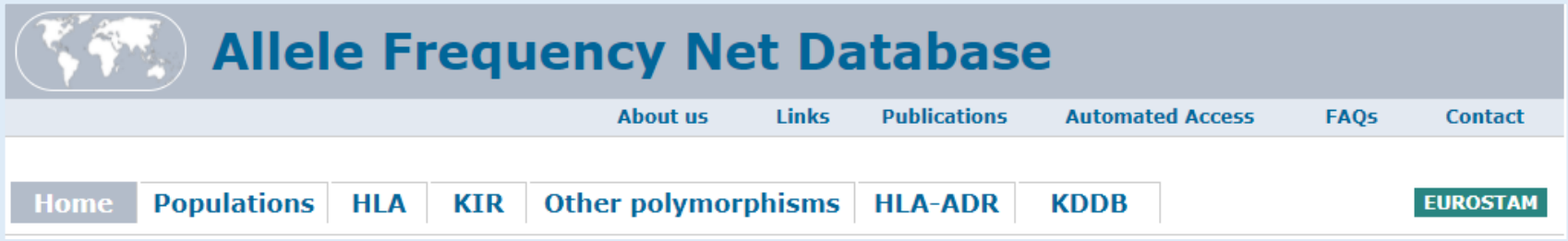
W. Helmberg

Department for Blood Group Serology and Transfusion Medicine, Medical University Graz, Graz, Austria









Kingston University London

Bioinformatics Tools: HLA-specific

HLA Data Handling

GENE[RATE] TOOLS

HLA-net > Gene[rate] Tools

May 2021 : Check-out our section with tools for binding and population diversity

TOOLS FOR HANDLING AND ANALYSING DATA WITH AMBIGUITIES

These programs expect **plain text files in UNIFORMAT v3** as input. A download link to the results file is provided once the execution is completed (on screen and/or by email). Read more about these tools (and see examples of input files) in the usage overview.

Please note that the results files are deleted after 48 hours, therefore if you have not saved your results file you have to run the programs again.

File conversions

Converts data files from different formats (tabulated, columnar, pdf files of Luminex reports, Arlequin projects) into UNIFORMAT, and vice versa.

LAUNCH

Uniformate

Checks UNIFORMAT files for formating errors and performs automatic translations of allele and haplotype names (new nomenclature, expansion of valid abbreviations,..).

LAUNCH

Basic statistics

Provides all the basic statistics for one and two-locus, including allele and haplotype frequencies, tests for Hardy-Weinberg, neutrality and linkage disequilibrium, and graphs.

LAUNCH

Haplotype frequency estimation

Calculates multi–locus haplotype frequencies (two or more loci) or allele frequencies under two alternative models (either Hardy-Weinberg equilibrium or inbreeding–like).

LAUNCH

Family phaser

Establishes haplotype phase for family data. Input is UNIFORMAT file with a modified identifier. There are no limits on the number of families included in each file.

LAUNCH

Phenotype

Calculates phenotypes from probe lists and reactivity data and returns a UNIFORMAT file.

LAUNCH

TISSUE ANTIGENS
IMMUNE RESPONSE GENETICS
Tissue Antigens ISSN 0001-2815

REVIEW ARTICLE

The *HLA-net GENE[RATE]* pipeline for effective HLA data analysis and its application to 145 population samples from Europe and neighbouring areas

J. M. Nunes, S. Buhler, D. Roessli, A. Sanchez-Mazas & the HLA-net 2013 collaboration†

Kingston University London

Bioinformatics Tools: HLA-specific

HLA SPREAD

HLA Web Resource for SNPs, Populations, Resources, ADRs, Diseases

Home | Visualization | Search | Team

Workflow

2910
Diseases

689
HLA alleles

163
Adverse Drug Reactions

286
Biomarker reports

Kingston University London

Dholakia et al. BMC Genomics (2022) 23:10
https://doi.org/10.1186/s12864-021-08239-0

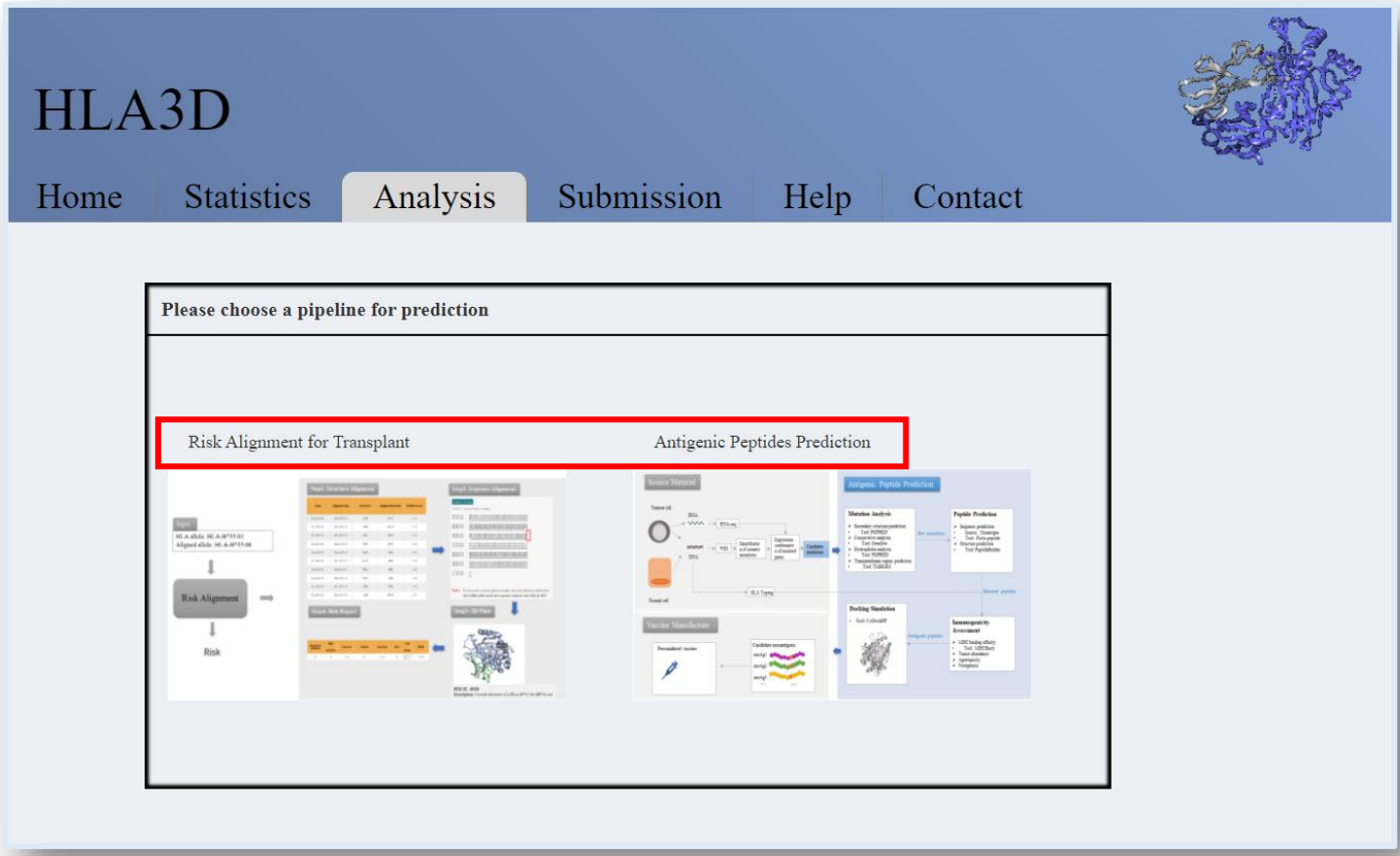
BMC Genomics

DATABASE Open Access

HLA-SPREAD: a natural language processing based resource for curating HLA association from PubMed abstracts

Dhwani Dholakia[1,2*†], Ankit Kalra[3†], Bishnu Raman Misir[4], Uma Kanga[5] and Mitali Mukerji[1,4,6*]

Bioinformatics Tools: HLA-specific



Key Points

- To provide comprehensive analysis of HLA for different populations, we developed HLA3D, a comprehensive toolkit that collected 1296 sequences, 256 PDB structures, 212 modelled structures, 120 000 frequency data, 73 000 associated literature, 39 000 disease-associated SNP of HLA and 1604 oncogenic mutations.
- Based on common HLA molecules in HLA3D, we qualified the HLA structure differences and obtained 100 000 RMSD records. In addition, we predicted 370 000 antigenic peptides with high affinity in common tumours, which helps to narrow the range of candidate neoantigens and promote subsequent experimental verification.
- By integrating the knowledge of differential immunogenicity in HLA sequences and structures, HLA3D established a risk alignment pipeline, providing users with the functions of structure alignment, sequence alignment, 3D-View and risk report, to help users predict the severity of aGVHD of mismatch HLA donors before transplantation.
- In view of the key characteristics that affect the immunogenicity of mutated peptides, HLA3D established an antigenic peptide prediction pipeline to provide users with a series of applications, such as mutation prediction, peptide prediction, immunogenicity assessment and docking simulation, to help users complete the prediction and analysis of immunogenic peptides.

OXFORD

"in press"

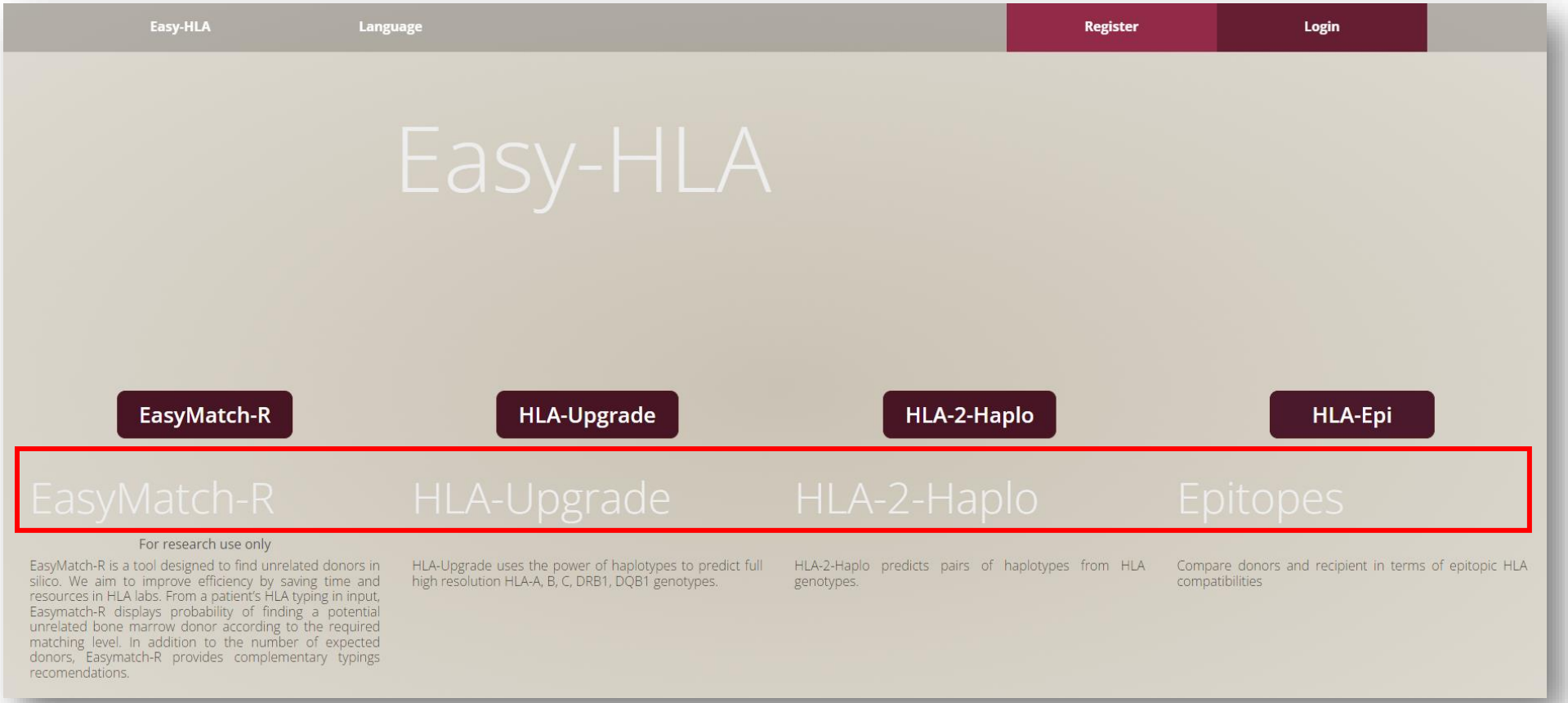
Briefings in Bioinformatics, 2022, 1–13
https://doi.org/10.1093/bib/bbac076
Problem Solving Protocol

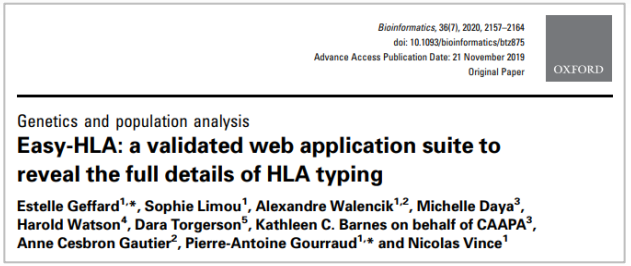
HLA3D: an integrated structure-based computational toolkit for immunotherapy

Xingyu Li†, Xue Lin†, Xueyin Mei, Pin Chen, Anna Liu, Weicheng Liang, Shan Chang and Jian Li

Kingston University London

Bioinformatics Tools: HLA-specific

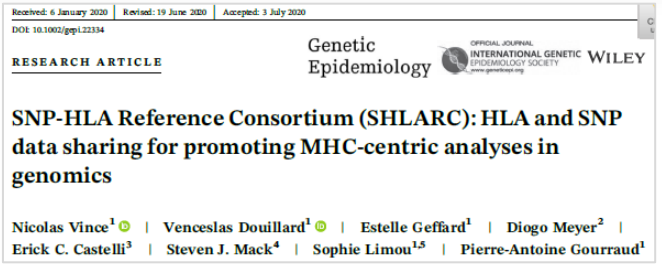



Bioinformatics, 36(7), 2020, 2157–2164
doi: 10.1093/bioinformatics/btz875
Advance Access Publication Date: 21 November 2019
Original Paper
OXFORD

Genetics and population analysis

Easy-HLA: a validated web application suite to reveal the full details of HLA typing

Estelle Geffard[1,*], Sophie Limou[1], Alexandre Walencik[1,2], Michelle Daya[3], Harold Watson[4], Dara Torgerson[5], Kathleen C. Barnes on behalf of CAAPA[3], Anne Cesbron Gautier[2], Pierre-Antoine Gourraud[1,*] and Nicolas Vince[1]


Received: 6 January 2020 | Revised: 19 June 2020 | Accepted: 3 July 2020
DOI: 10.1002/gepi.22334
RESEARCH ARTICLE
Genetic Epidemiology
OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY
WILEY

SNP-HLA Reference Consortium (SHLARC): HLA and SNP data sharing for promoting MHC-centric analyses in genomics

Nicolas Vince[1] | Venceslas Douillard[1] | Estelle Geffard[1] | Diogo Meyer[2] | Erick C. Castelli[3] | Steven J. Mack[4] | Sophie Limou[1,5] | Pierre-Antoine Gourraud[1]

Kingston University London

Bioinformatics Tools: HLA-specific

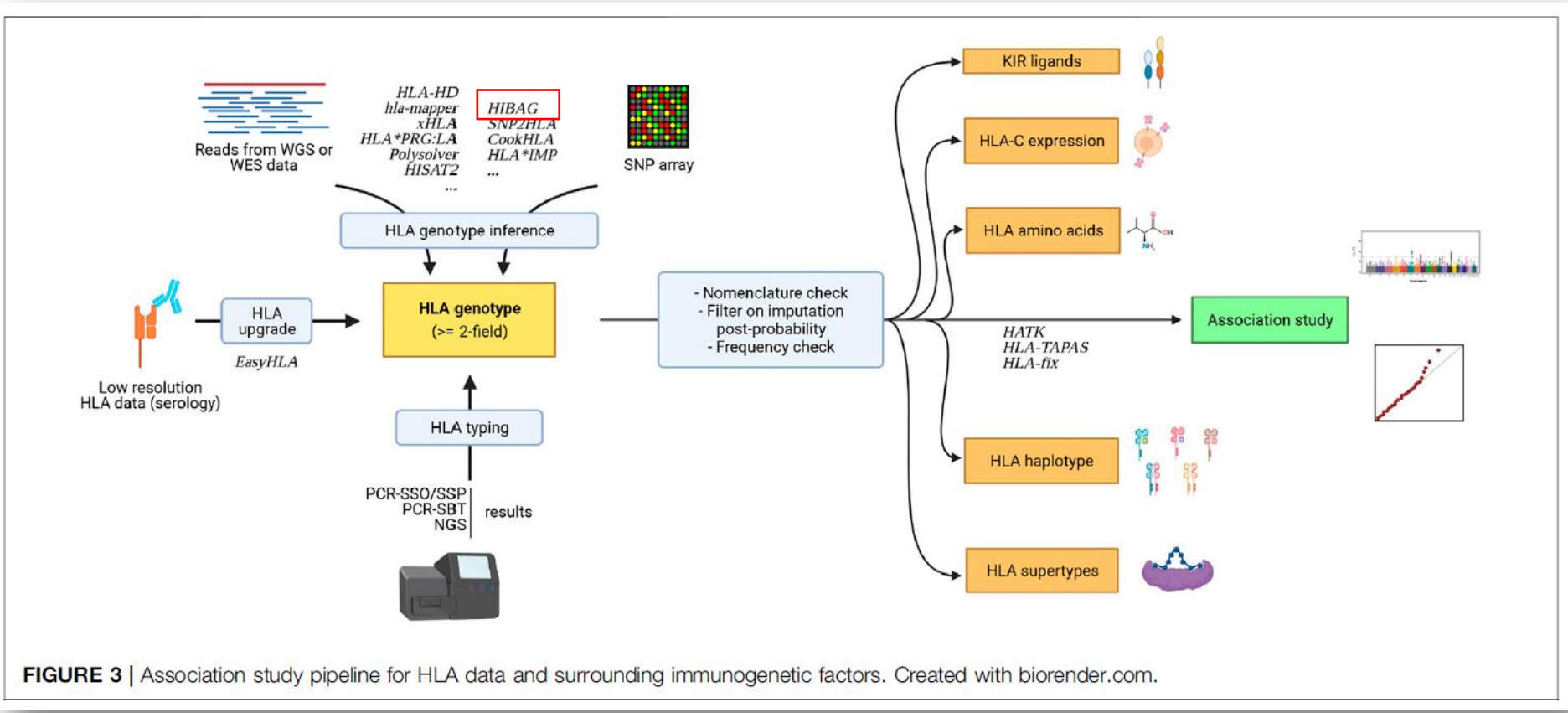


FIGURE 3 | Association study pipeline for HLA data and surrounding immunogenetic factors. Created with biorender.com.

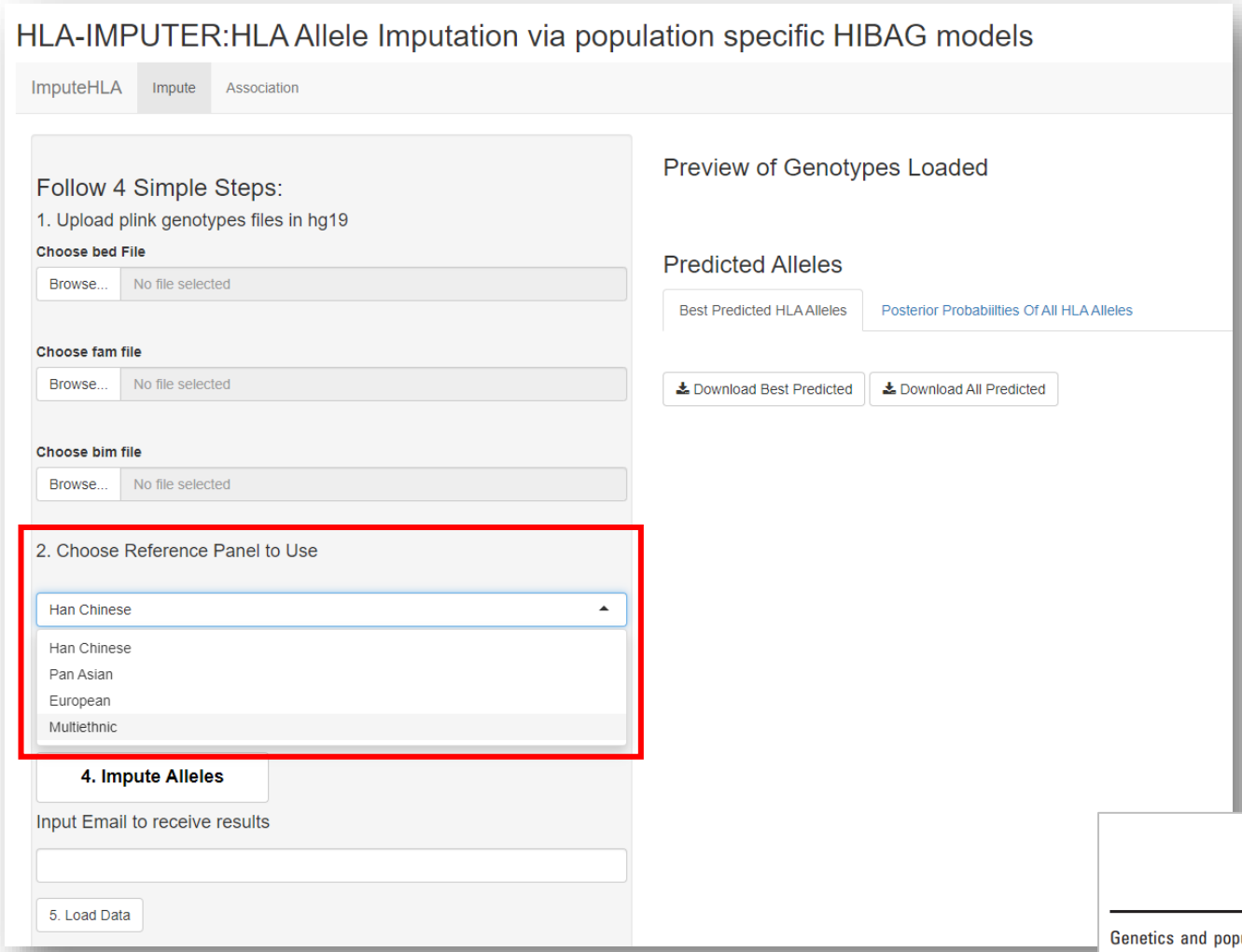
frontiers in Genetics
REVIEW
published: 02 December 2021
doi: 10.3389/fgene.2021.774916

Approaching Genetics Through the MHC Lens: Tools and Methods for HLA Research

Venceslas Douillard[1], Erick C. Castelli[2], Steven J. Mack[3], Jill A. Hollenbach[4,5], Pierre-Antoine Gourraud[1], Nicolas Vince[1†*] and Sophie Limou[1,6,*†] on behalf of the Covid-19|HLA & Immunogenetics Consortium and the SNP-HLA Reference Consortium

Kingston University London

Bioinformatics Tools: HLA-specific



Bioinformatics, 35(7), 2019, 1244–1246
doi: 10.1093/bioinformatics/bty730
Advance Access Publication Date: 31 August 2018
Applications Note

OXFORD

Genetics and population analysis

HLA-IMPUTER: an easy to use web application for HLA imputation and association analysis using population-specific reference panels

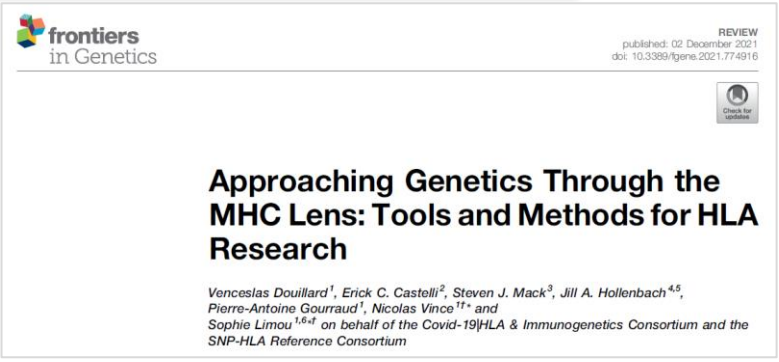
Jiangshan J. Shen[1,†], Chao Yang[2,†], Yong-Fei Wang[1], Ting-You Wang[1], Mengbiao Guo[1], Yu Lung Lau[1], Xuejun Zhang[2], Yujun Sheng[2,*] and Wanling Yang[1,*]

Kingston University London

Bioinformatics Tools: HLA-specific

TABLE 1 | Tools for HLA analyses.

HLA application name	Description	URL
Alphlard-nt (Hayashi et al., 2019)	Identification of somatic mutations in HLA molecules from whole-genome and exome data using Bayesian algorithms	—
BIGDAWG (Pappas et al., 2016)	Open-source R package for the case-control analysis of highly polymorphic data at the allele, haplotype and amino-acid level	https://CRAN.R-project.org/package=BIGDAWG
Easy-HLA (Geffard et al., 2020)	Website with HLA alleles haplotyping, upgrading and inference from HLA genotypes, prediction of HLA-C expression	http://hla.univ-nantes.fr/
HATK (Choi et al., 2021)	Open-source *Python* pipeline for HLA association studies, including tools for HLA data formatting	https://github.com/WansonChoi/HATK
HLA-check (Jeanmougin et al., 2017)	Perl tool evaluating the probability of accurate HLA genotype imputation by comparing it to SNP imputation in the exonic region of HLA.	https://github.com/mclegrand/HLA-check/
HLA-EMMA (Kramer et al., 2020)	Donor/recipient compatibility assessment based on solvent-accesible amino acids, based on intralocus comparisons	http://www.HLA-EMMA.com
HLA*fix*	Open-source R pipeline for HLA association studies. Performing SNP quality control steps, stratification, HLA imputation and representation of the results	https://univ-nantes.io/Nico_V/hlafix
HLAHapV (Osoegawa et al., 2016)	A Java-based HLA Haplotype Validator for quality assessments of HLA typing	https://github.com/nmdp-bioinformatics/ImmunogeneticDataTools
HLA-NET (Nunes et al., 2014)	Set of tools to manipulate HLA data, infer haplotypes, convert files format, and information about typing	https://hla-net.eu/
HLApers (Aguiar et al., 2020)	Genotyping and quantification of HLA expression from RNA-seq data	https://github.com/genevol-usp/HLApers
HLA-TAPAS (Luo et al., 2020)	Open-source *Python* pipeline for creation of reference panels and HLA association studies	https://github.com/immunogenomics/HLA-TAPAS
MergeReference (Cook and Han, 2017)	SNP2HLA compatible tool to concatenate multiple reference panels in order to gain accuracy during HLA imputation	http://software.buhmhan.com/MergeReference
pyHLA (Fan and Song, 2017)	Association analysis for HLA alleles in *Python* language	https://github.com/felixfan/PyHLA



Kingston University London

Bioinformatics Tools: HLA-specific

R Packages

MiDAS—Meaningful Immunogenetic Data at Scale

Maciej Migdal, Dan Fu Ruan, William F. Forrest, Amir Horowitz, Christian Hammer

HATK: HLA analysis toolkit

Wanson Choi, Yang Luo, Soumya Raychaudhuri, Buhm Han

Bioinformatics, Volume 37, Issue 3, 1 February 2021, Pages 416–418,
https://doi.org/10.1093/bioinformatics/btaa684

Home » Bioconductor 3.14 » Software Packages » immunotation

immunotation

platforms all | rank 1967 / 2083 | support 0 / 0 | in Bioc 1 year
build ok | updated before release | dependencies 68

DOI: 10.18129/B9.bioc.immunotation

Tools for working with diverse immune genes

Bioconductor version: Release (3.14)

MHC (major histocompatibility complex) molecules are cell surface complexes that present antigens to T cells. The repertoire of antigens presented in a given genetic background largely depends on the sequence of the encoded MHC molecules, and thus, in humans, on the highly variable HLA (human leukocyte antigen) genes of the hyperpolymorphic HLA locus. More than 28,000 different HLA alleles have been reported, with significant differences in allele frequencies between human populations worldwide. Reproducible and consistent annotation of HLA alleles in large-scale bioinformatics workflows remains challenging, because the available reference databases and software tools often use different HLA naming schemes. The package immunotation provides tools for consistent annotation of HLA genes in typical immunoinformatics workflows such as for example the prediction of MHC-presented peptides in different human donors. Converter functions that provide mappings between different HLA naming schemes are based on the MHC restriction ontology (MRO). The package also provides automated access to HLA alleles frequencies in worldwide human reference populations stored in the Allele Frequency Net Database.

Author: Katharina Imkeller [cre, aut]

Kingston University London

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Bioinformatics Tools

BIOINFORMATICS TOOLS

(for Genetic Epidemiologists)

Mehmet Tevfik DORAK, MD, PhD

Open Access Reviews / Journals

ENCODE Resources (Pazin MJ. CSHP 2015); Database Tools (Bianco, Genomics 2013; Zou, 2015); UCSC Browser (Mangan, 2014)

Journals: Nucleic Acids Research (NAR); NAR Genomics & Bioinformatics; Database; Bioinformatics

Nucleic Acids Research Databases Catalog

Scientific Data

Article Collections

ENCODE (Web Focus) FANTOM5 (RIKEN) IHEC Epigenome RoadMap TCGA GTEx

F1000Research: Bioconductor

Societies

ISCB ISMB AMIA ECCB

Online Suites

Galaxy (101 - Support) ExPASy (ref) Taverna 2.5.0/Archive (ref) Apache-Taverna BioMart (ref) Artemis (manual) Babelomics (ref / tutorial)

Bioconductor (manual) BioPython (tutorial & cookbook / PDF)

EDGE: Empowering the Development of Genomics Expertise (ref)

Biomarts

BioTools: Biomart

Ensembl BioMart (YouTube tutorial) UCSC Table Browser GWAS Central GWAS Mart (tutorial) SPS Mart (allele freqs) Fantom5

Genome Browsers

BioTools: Genome Visualization

UCSC Human Genome BioInformatics GENCODE FANTOM5 (ref / tutorial / wiki / user guide (pp.12-19)) - ZENBU - SLIDEBASE (tutorial)

Ensembl Genome Browser (Human) Vega Genome Browser (Human)

NIH RoadMap Epigenomic Browser (tutorial) WashU EpiGenome Human Epigenome Atlas NCBI Epigenomics (help) NGSmethDB (ref)

1KG (Browser - Quick Start Guide)-(NCBI) PopHuman (ref)

UK10K (ref) GENCODE Browser SNiPA (gnomAD) genome Aggregation Browser (ref)

CNV Browser DECIPHER eQTL Browser seeQTL GTEx eQTL Browser

lncRBase NonCode FANTOM CAT (lncRNA Atlas) Browser (ref) Integrative Genomics (ref)

VISTA Enhancer Browser Swiss Regulon Reactome Pathway Browser ImmunoBase Browser ImmunPop Browser (ref)

NCBI Genome Data Viewer (GDV) (ref / tutorial)

General Resources

BioTools OmicX

NCBI Global Search GoPubMed HuGE LabMeeting NCBI Tutorials / Handbook / Help Manual NIH Helix NCBI Tools for Data Mining NCI Data Analysis Tools

Sanger: Software / Databases & Tools EMBL-EBI Swiss Institute of Bioinformatics: ExPASy JJ Wang Lab

NCBI Gene .. GeneCards .. GeneMap .. GenAtlas .. iHOP .. BioGPS Gene Annotation .. Harmonizome (ref)

NCBI (tutorials on YouTube) .. BLAST .. MimicMe (ref) .. e-PCR .. In Silico PCR .. GeneNote .. GeneLoc .. CGAP-GAI .. Aceview

ENCODE (ref / tutorial1 / tutorial2 / cell types) GENCODE RegulomeDB FANTOM/Gateway (papers / software) Broad Institute Portals

Database of Genomic Variants (CNV) CNV inspector Copy Number Variation Resources DECIPHE

OMIM .. OMIM Map .. NCBI Genetic Association Database (GAD) [Now Retired] .. HuGE Navigator .. Variant N

Gene Exp Omnibus -GEO (NCBI) .. NCBI dbGaP .. ArrayExpress (EBI) .. SOURCE (genomics tool)

http://www.dorak.info/mtd/bioinf.html

Kingston University London

Bioinformatics Tools for HLA

The amount of biological data related to the HLA genes and HLA region is ever increasing

Cell/tissue-specific, validated, high quality data should be exploited at every opportunity to find answers to research questions

To do that, user friendly tools and apps are available

There is, however, no substitute for expert advice from a trained bioinformatician

Experimental verification is also still needed

Kingston University London

Thank You

Kingston University London