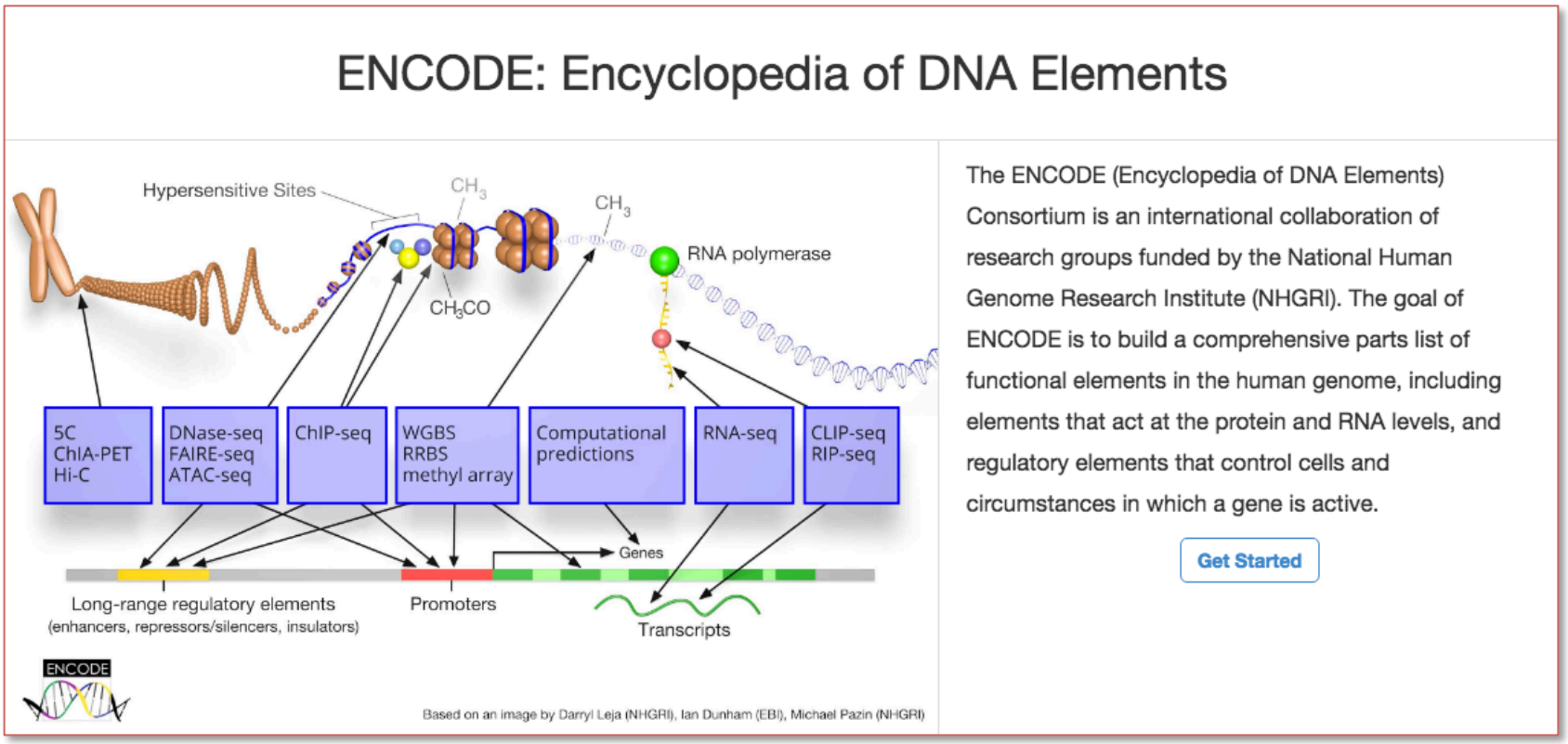
3D Genome Workshop - Overview of ENCODE



Dan Gilchrist
National Human Genome Research Institute
October 18, 2016

National Human Genome Research Institute

3D Genome Workshop - Overview of ENCODE

- The ENCODE Project
- 3D Data Access Through the ENCODE Portal
- The ENCODE Encyclopedia
- Tools for Investigating Genetic Variants

ENCODE Tutorials

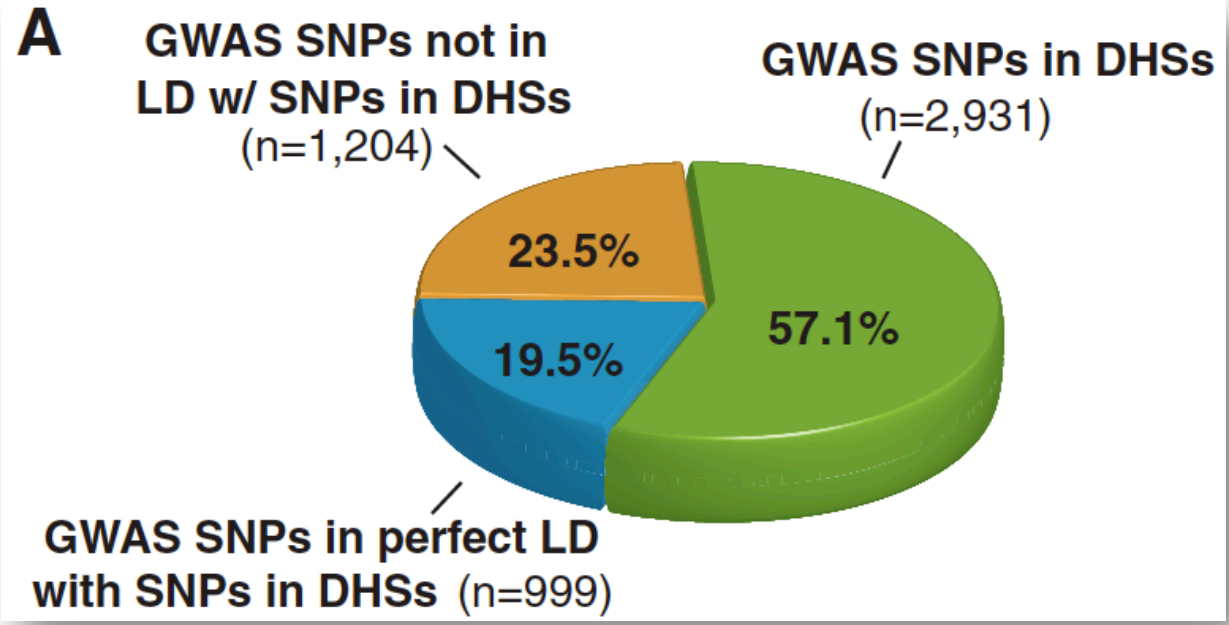
- ENCODE Tutorials

  https://www.encodeproject.org/tutorials/

  http://www.genome.gov/27553900

ENCODE: Annotating the non-coding genome

- Vast majority of common disease associations and heritability lie outside of coding regions
- Non-coding variants known to cause human diseases and alter human traits (FXS, ALS)



Stamatoyannopoulos, Science 337:1190, 2012

Goals of ENCODE

- Identify all candidate functional elements in the human and mouse genomes

  Promoters, enhancers, transcribed regions

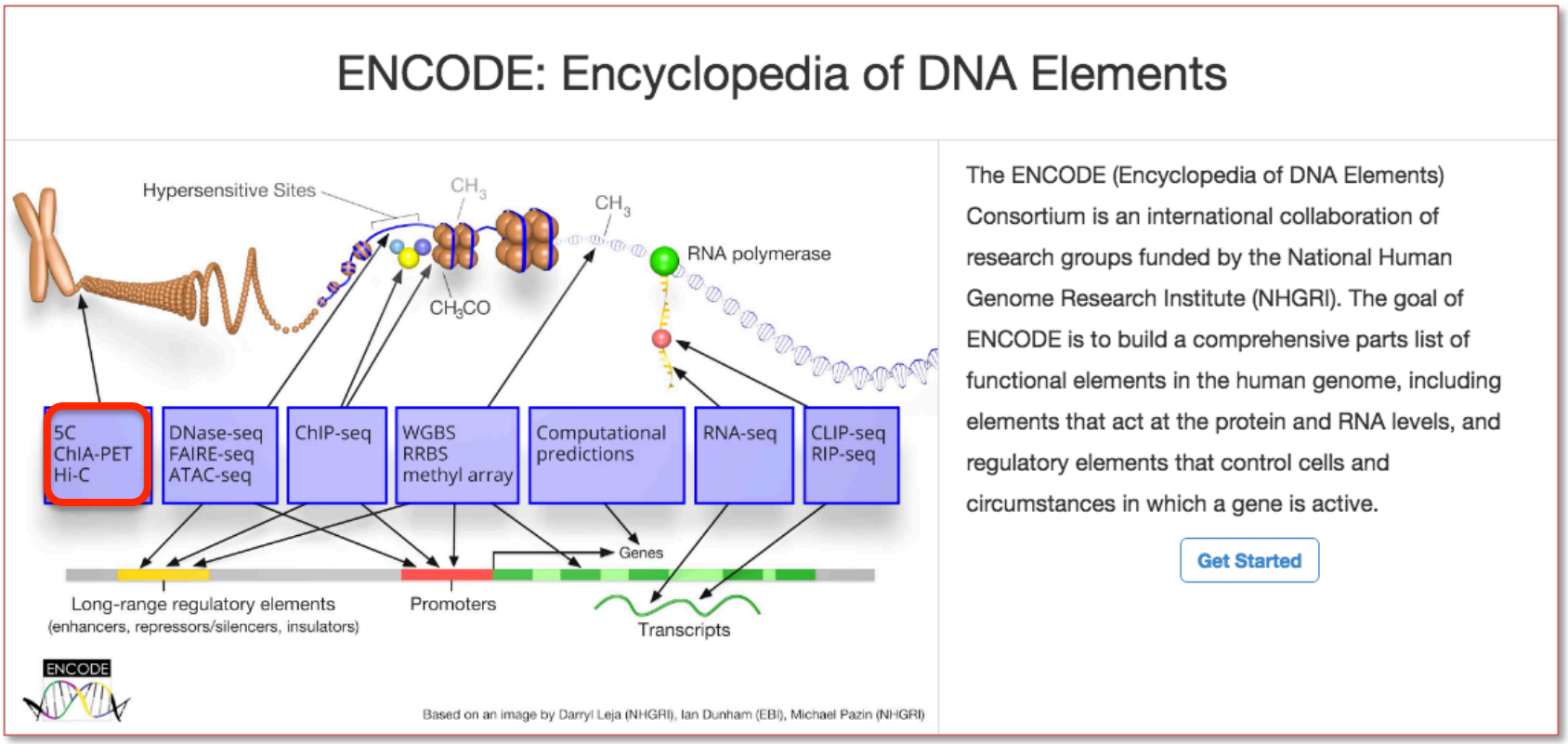
- Make resource freely available to the community for use in studies of:

  Genetic basis of disease

  Gene regulation

  Computational tool development

ENCODE Data Types

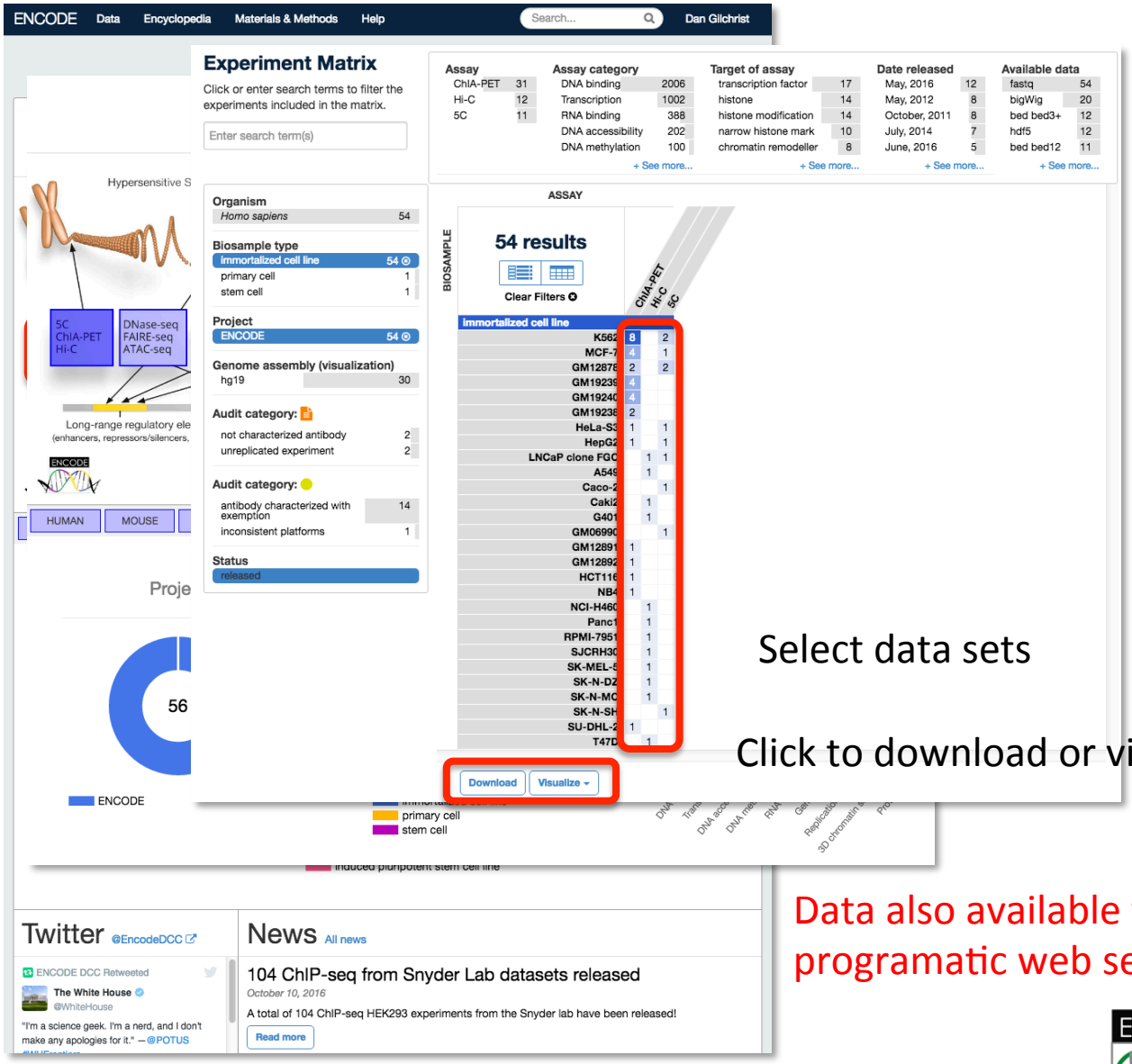


- 1000s of data sets
- 100s of biosamples

3D Genome Workshop - Overview of ENCODE

- The ENCODE Project
- 3D Data Access Through the ENCODE Portal
- The ENCODE Encyclopedia
- Tools for Investigating Genetic Variants

ENCODE Portal – Unrestricted data access



Select data sets

Click to download or visualize

Data also available through programatic web services

https://encodeproject.org

3D Genome Workshop - Overview of ENCODE

- The ENCODE Project
- 3D Data Access Through the ENCODE Portal
- The ENCODE Encyclopedia
- Tools for Investigating Genetic Variants

ENCODE Annotations – the Encyclopedia

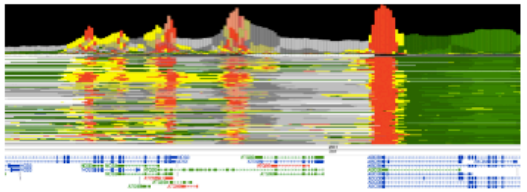
ENCODE Data Ency

Top Level Annotations

Chromatin states

Semi-automated genomic annotation methods such as ChromHMM and Segway take as input a panel of epigenomic data (including histone mark ChIP-seq and DNase-seq) in a particular cell type and use machine learning methods to simultaneously partition the genome into segments and assign chromatin states to these segments; the states are assigned such that two segments with the same state exhibit similar epigenomic patterns. The procedure is "semi-automated" because states are then manually compared with known biological information in order to designate each state as an enhancer-like, promoter-like, gene body, etc.

[Search]

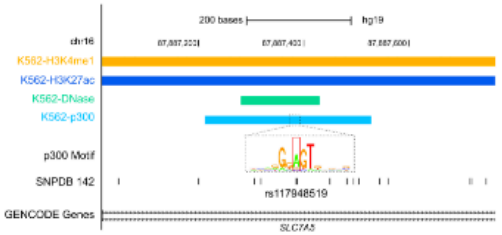


epilogos

Variant Annotation

Over the past decade, Genome Wide Association Studies (GWAS) have provided insights into how genetic variations contribute to human diseases. However, over 80% of the variants reported by GWAS are in noncoding regions of the genome and the mechanism of how they contribute to disease onset is unknown. By integrating data from the ENCODE project and other public sources, RegulomeDB and HaploReg are two resources developed by ENCODE labs to aid the research community in annotating GWAS variants. FunSeq2 is another ENCODE resource for annotating both germline and somatic variants, particularly in the noncoding regions of cancer genomes.

[RegulomeDB | HaploReg | FunSeq2]



Middle

Promoter-l

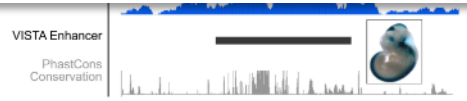
DNase hypersen
promoter functic
and H3K4me3 si
used to predict r
accuracy than D
human cell types
generated by the
types with only H
ranked them by
genomic locatio
WashU genome
[Visualize | | M

Enhancer-l

DNase hypersen
enhancer functic
and H3K27ac sig
tested on mouse
and H3K27ac in
mouse cell types
Roadmap Epigenomic consortia. For cell and tissues types with only H3K27ac or DNase data, we rank the peaks using the available data and make predictions of enhancer-like regions. You can query these enhancers by genomic locations, nearby genes, or SNPs, and visualize them in the UCSC and WashU genome browsers.

[Visualize | Method]

Enhancer-like genomic regions were tested on VISTA experimentally-validated enhancer elements: [VISTA]



Ground

Gene expres
The expression lev
types.
[Long RNA-seq D

Transcriptio
Peaks (enriched g
[Raw Data | Peaks
Visualize sequenc

Histone mar
Peaks of a variety
[Raw Data | Peaks

Open chrom
DNase I hypersens
experiments.
[Raw Data | Peaks

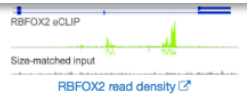
Topological
TADs ar
[Raw D
Visua

Promoter-e
Links between pro
experiments.
[Raw Data | Links

RNA binding protein occupancy (eCLIP-seq)

Peaks computed from eCLIP-seq data in human cell lines K562 and HepG2 for a large number of RNA Binding Proteins (RBPs).

[Raw Data | Peaks]



RBFOX2 read density

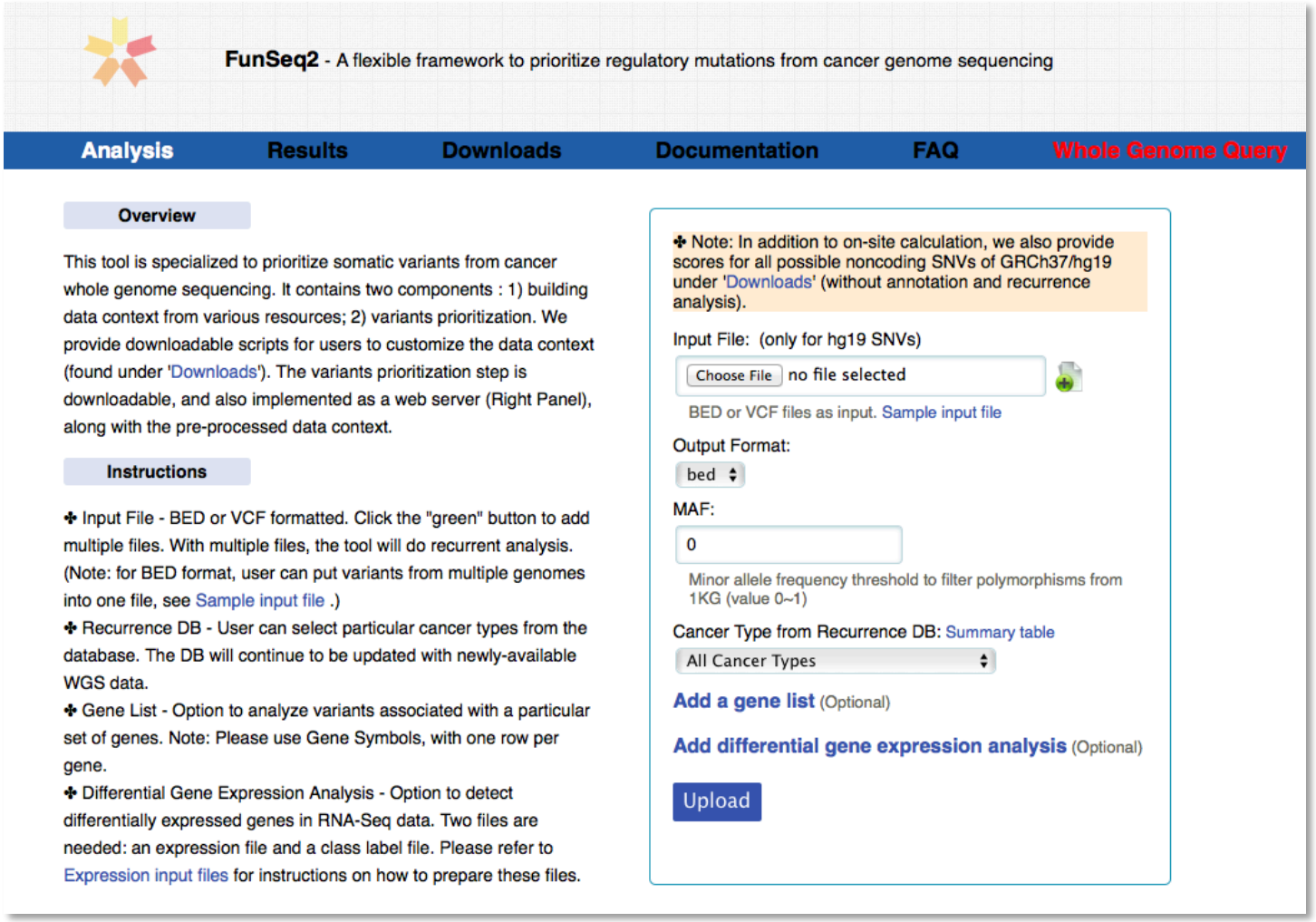
future plan

ENCODE

3D Genome Workshop - Overview of ENCODE

- The ENCODE Project
- 3D Data Access Through the ENCODE Portal
- The ENCODE Encyclopedia
- Tools for Investigating Genetic Variants

Tools for Variant Annotation – FunSeq2

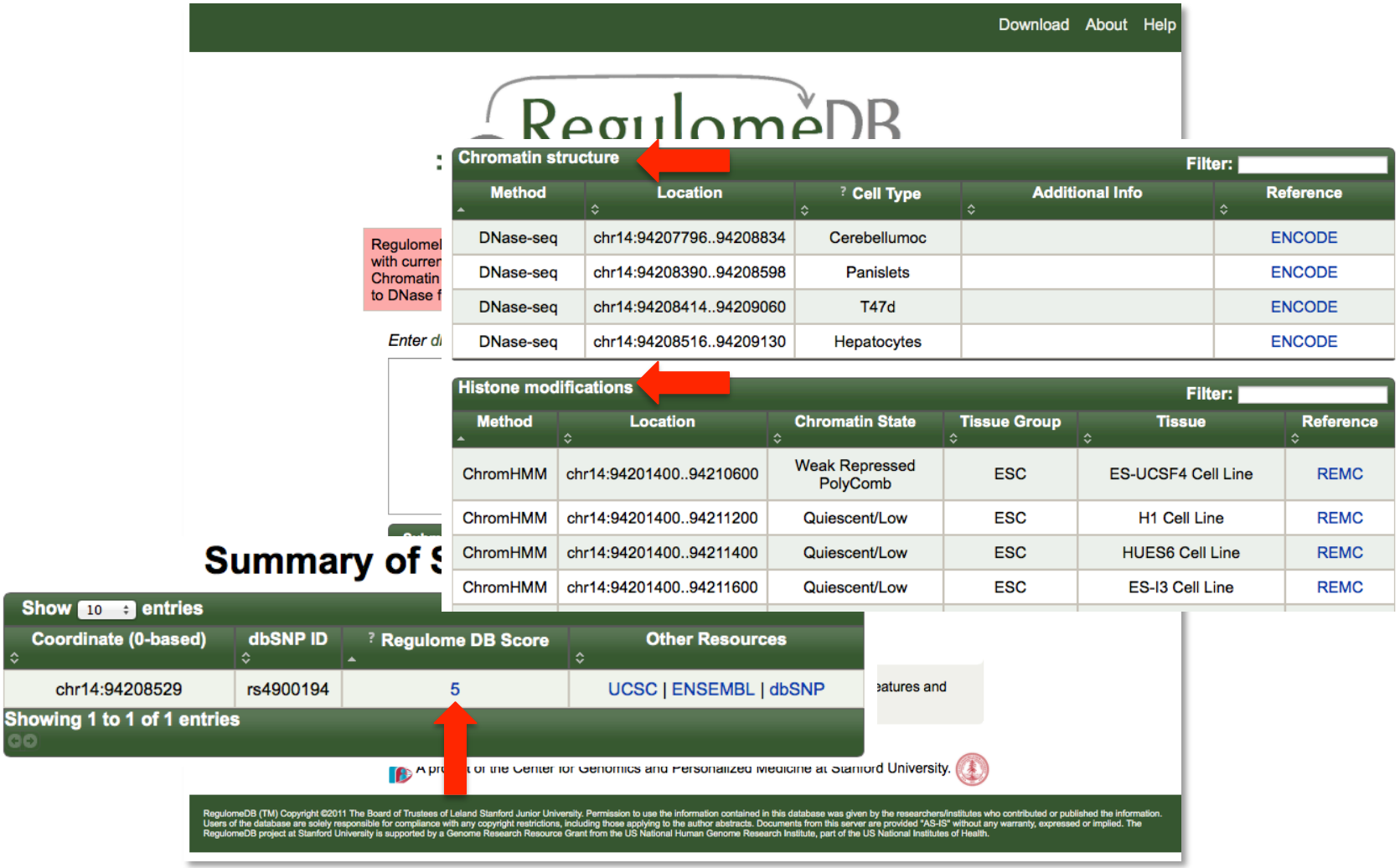


http://funseq2.gersteinlab.org/

Fu, Khurana, Gerstein, Genome Biology (15)10 2014

ENCODE

Tools for Variant Annotation - RegulomeDB

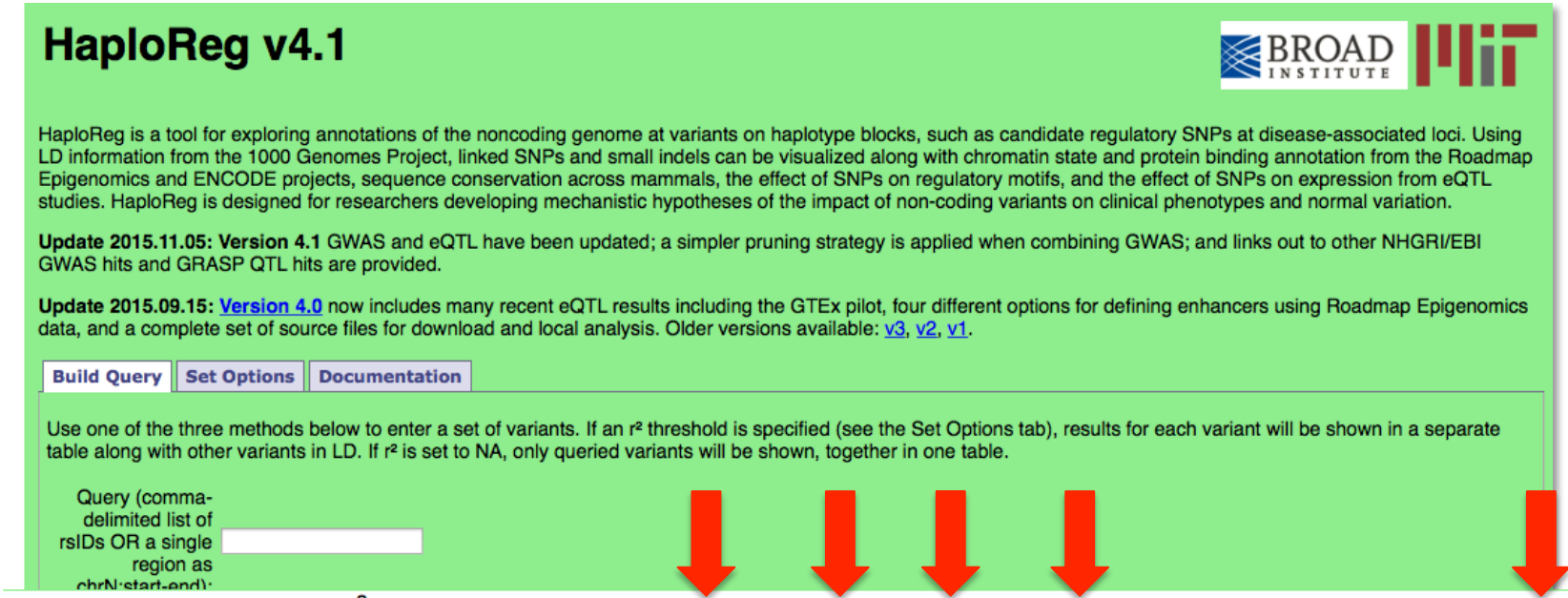


http://regulomedb.org/

Boyle, Cherry, Snyder, Genome Research 22-1790,2012

ENCODE

Tools for Variant Annotation - HaploReg



HaploReg v4.1

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 2015.11.05: Version 4.1 GWAS and eQTL have been updated; a simpler pruning strategy is applied when combining GWAS; and links out to other NHGRI/EBI GWAS hits and GRASP QTL hits are provided.

Update 2015.09.15: Version 4.0 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):

Query SNP: rs6575353 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	Motifs changed	NHGRI/EBI GWAS hits	GRASP QTL hits	Selected eQTL hits	GENCODE genes	dbSNP func annot
14	93736606	0.83	-0.96	rs77582682	G	A	0.03	0.06	0.00	0.08				9 tissues	CTCF,RAD21	4 altered motifs				PRIMA1	intronic
14	93738999	1	1	rs6575353	G	A	0.71	0.85	0.51	0.92						Sox,TCF4	1 hit	1 hit		PRIMA1	intronic
14	93739616	0.95	1	rs12896080	G	A	0.61	0.84	0.51	0.91						4 altered motifs				PRIMA1	intronic
14	93739927	0.83	1	rs4905084	A	G	0.60	0.83	0.51	0.90						4 altered motifs				PRIMA1	intronic
14	93742184	0.95	1	rs4900194	A	G	0.61	0.84	0.51	0.91			4 tissues	BRN		Gfi1,KAP1,Smad3		1 hit		PRIMA1	intronic
14	93743570	0.97	1	rs11160137	A	G	0.69	0.85	0.52	0.91			5 tissues	ESC,HRT		4 altered motifs				PRIMA1	intronic
14	93743836	0.85	0.93	rs12587586	T	G	0.59	0.84	0.52	0.91			6 tissues			Roaz				PRIMA1	intronic

www.broadinstitute.org/mammals/haploreg/

Ward and Kellis, Nucleic Acids Research 40-D930, 2011

Coming soon – ENCODE Element Visualizer!!!

Coming soon – Next Phase of ENCODE

- Starting in 2017
- Element mapping in normal and disease samples
- Characterizing function of candidate elements
- Likely to include expanded 3D data sets

ENCODE at ASHG2016

- Chat with ENCODE Data Coordination Center and NHGRI ENCODE Staff
  - NHGRI Booth #517
  - 2 - 3 PM Thursday 10/20

- ENCODE DCC Workshop - Open Science: Quality Assurance and Analysis of ChIP-seq Data Using the ENCODE Uniform Processing Pipeline
  - Ballroom B
  - 7:15 AM Friday 10/21

ENCODE Project Participants

Elise Feingold

Mike Pazin