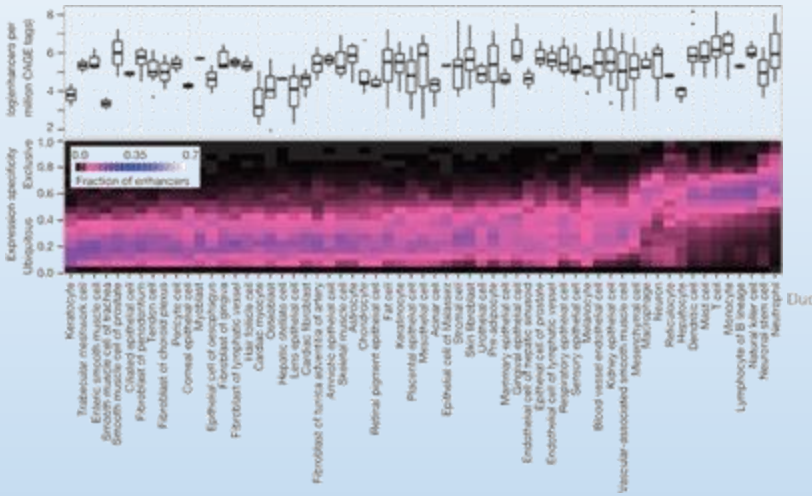
Using INFERNO to characterize the effects of noncoding AD variants

Alex Amlie-Wolf

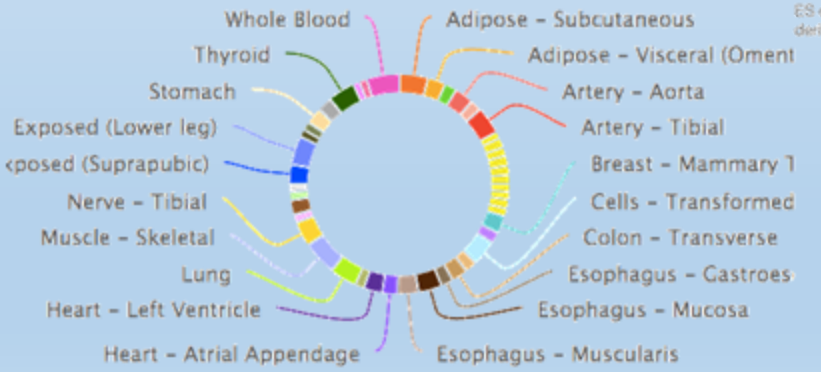
Genomics & Computational Biology PhD Student, Li-San Wang lab

ADGC 01/30/2018

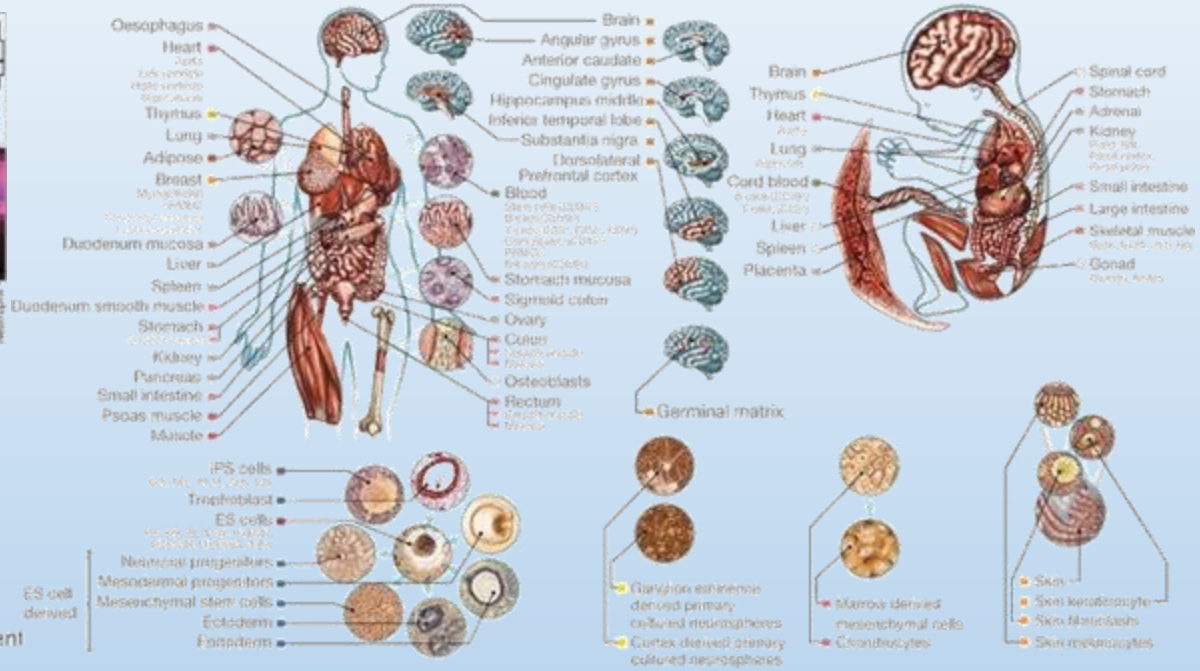
Large-scale functional genomics data sources



FANTOM5 eRNA Measurements



GTEx eQTLs



Roadmap Epigenomics Consortium

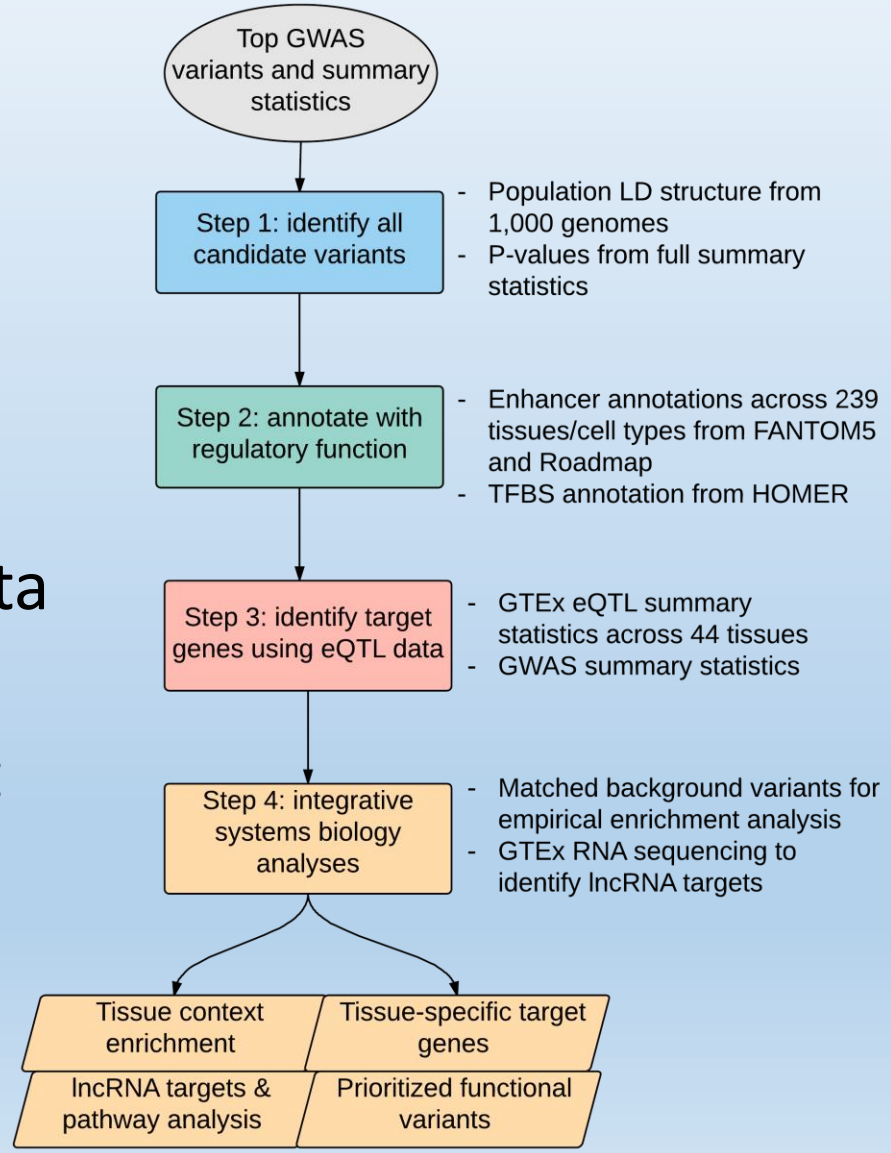
ENCODE

DASHR

Database of small human noncoding RNAs

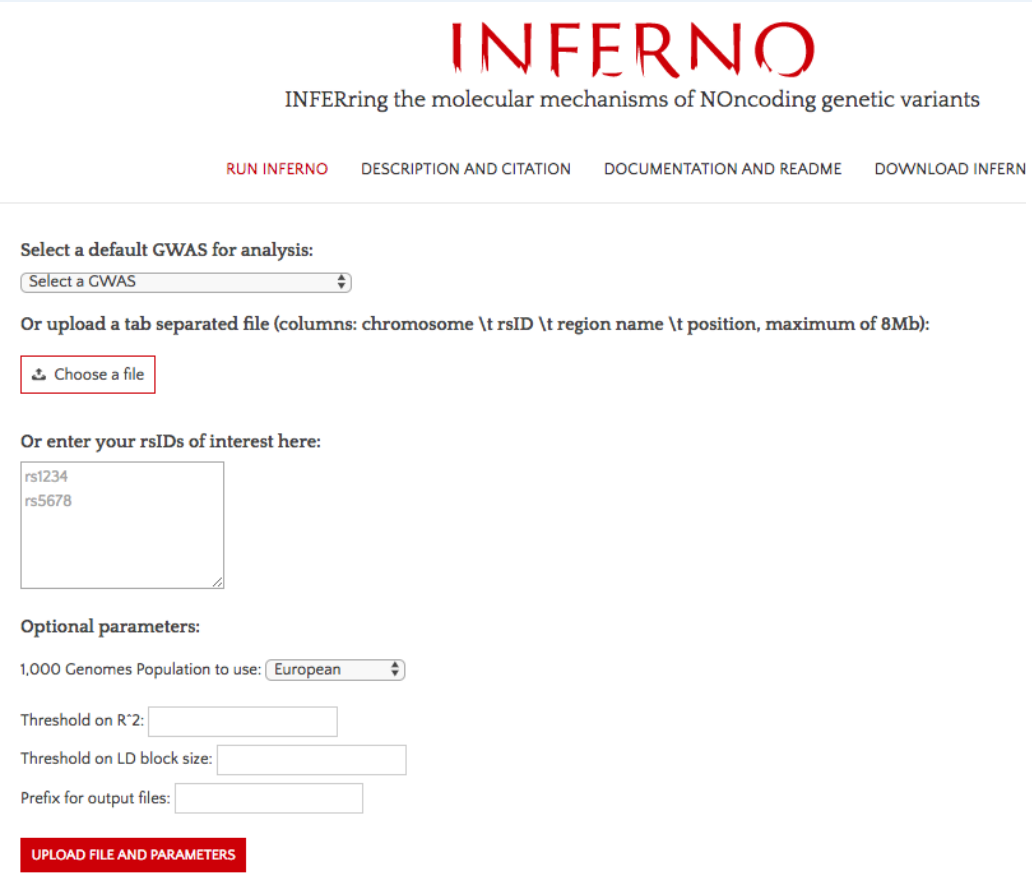
INFERNO: INFERring the molecular mechanisms of NOncoding genetic variants

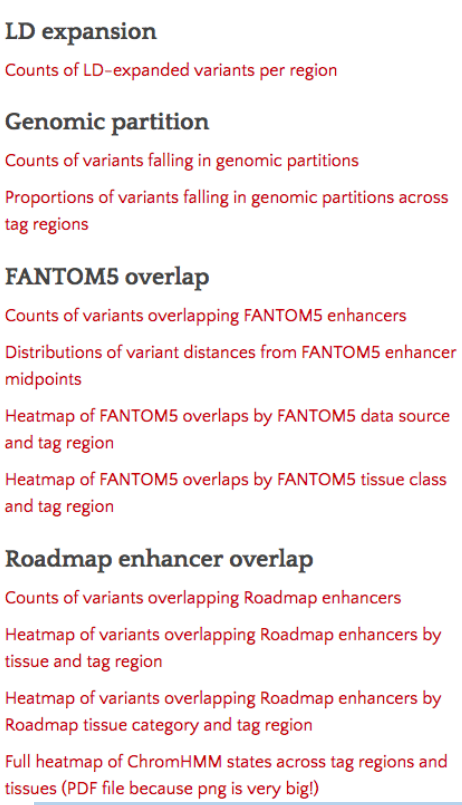
- Take any GWAS top hits and characterize:
  - Tissue context
  - Regulatory mechanism
  - Target genes
- Tissue categorization is key to data integration
- Identify variants with concordant enhancer, TFBS, and co-localized eQTL support in a tissue context

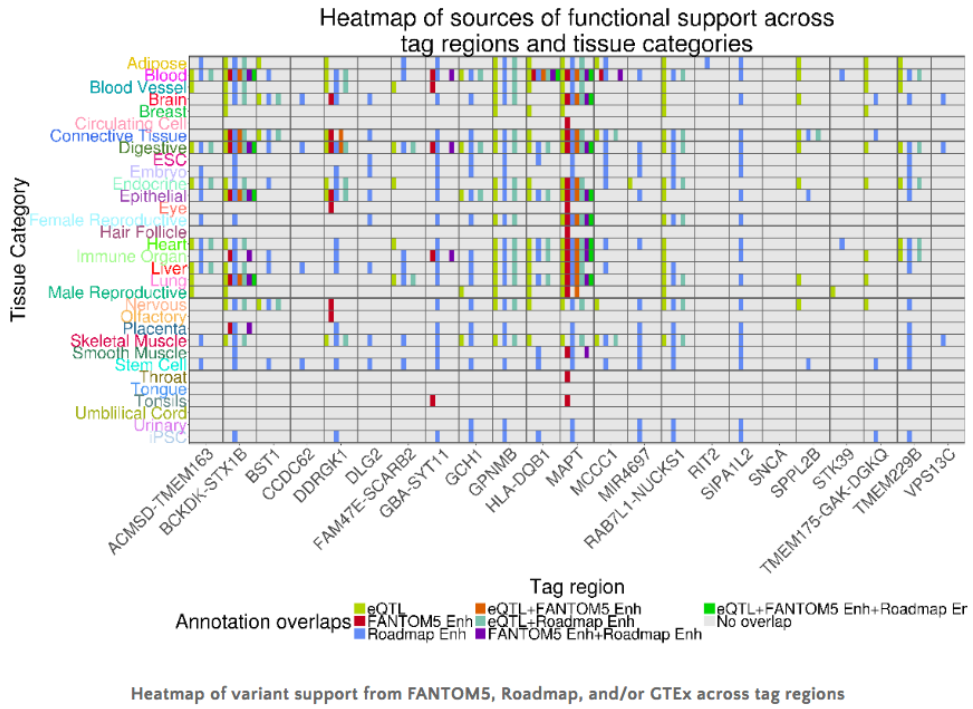


Open source pipeline implemented in R, Python, and bash
https://bitbucket.org/alexamlie/INFERNO
https://www.biorxiv.org/content/early/2017/10/30/211599

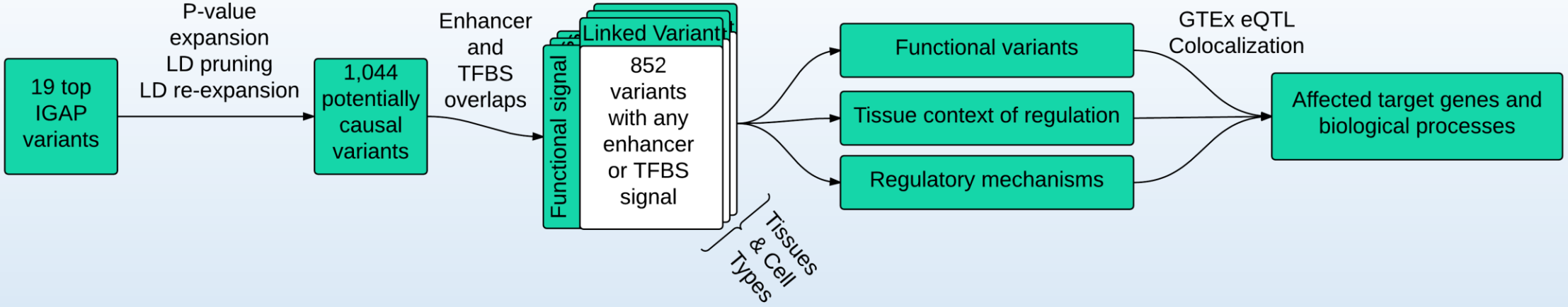
INFERNO web server



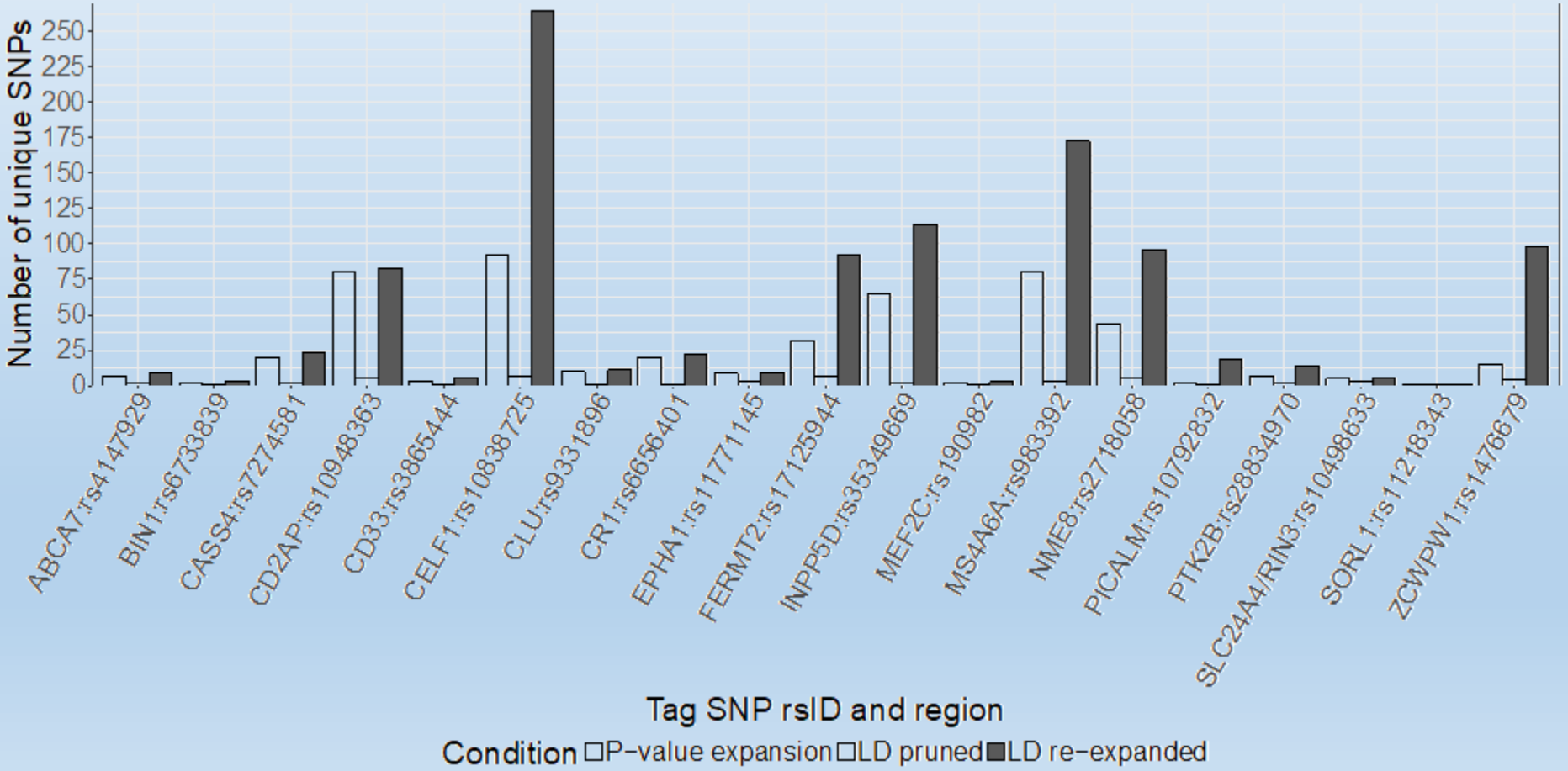




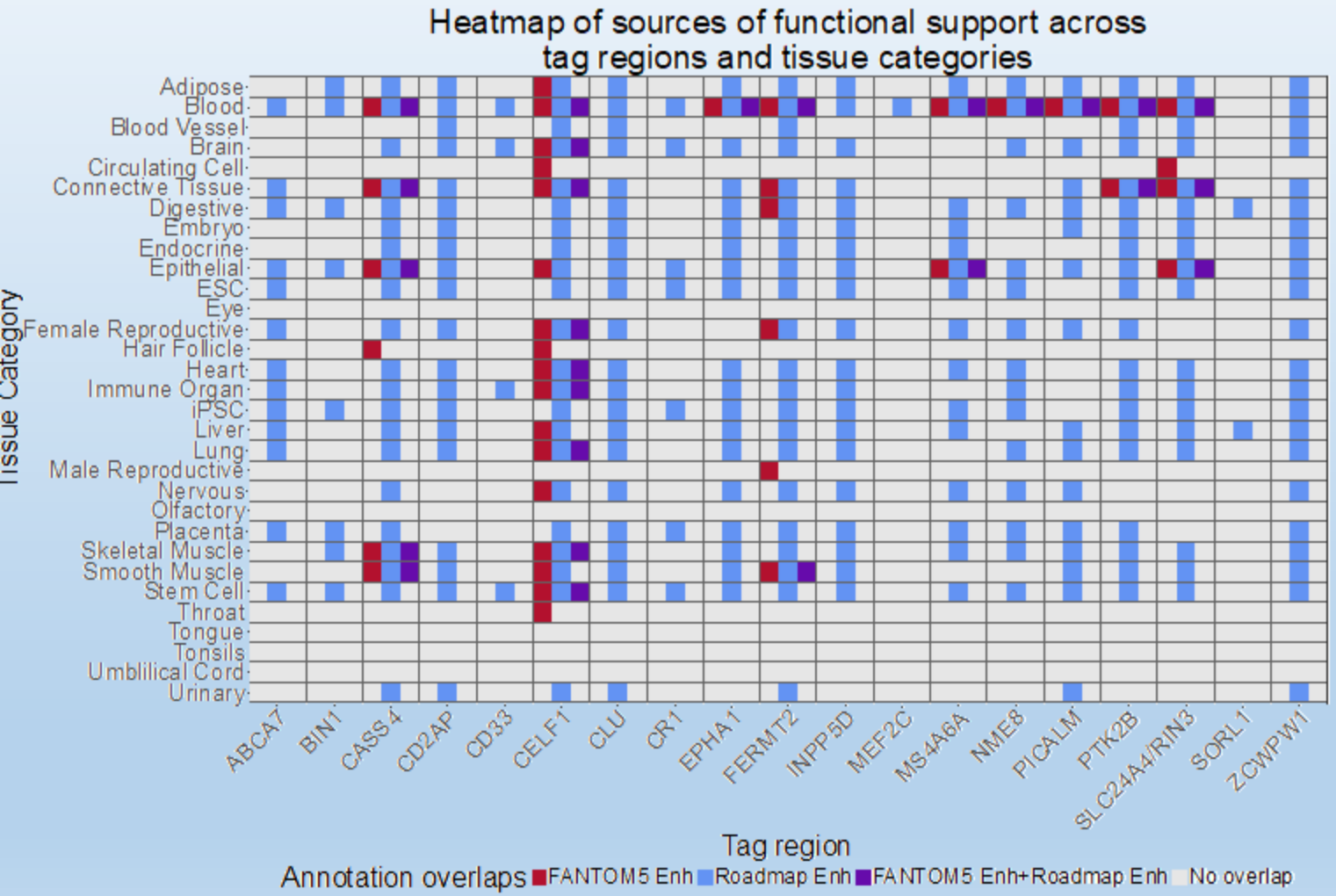
- http://inferno.lisanwanglab.org/
- Accepts top variants from any GWAS, expands into LD blocks and performs integrative enhancer+TFBS overlap and direct eQTL overlap analysis
- Full pipeline including Bayesian co-localization can be run on bsub-based servers

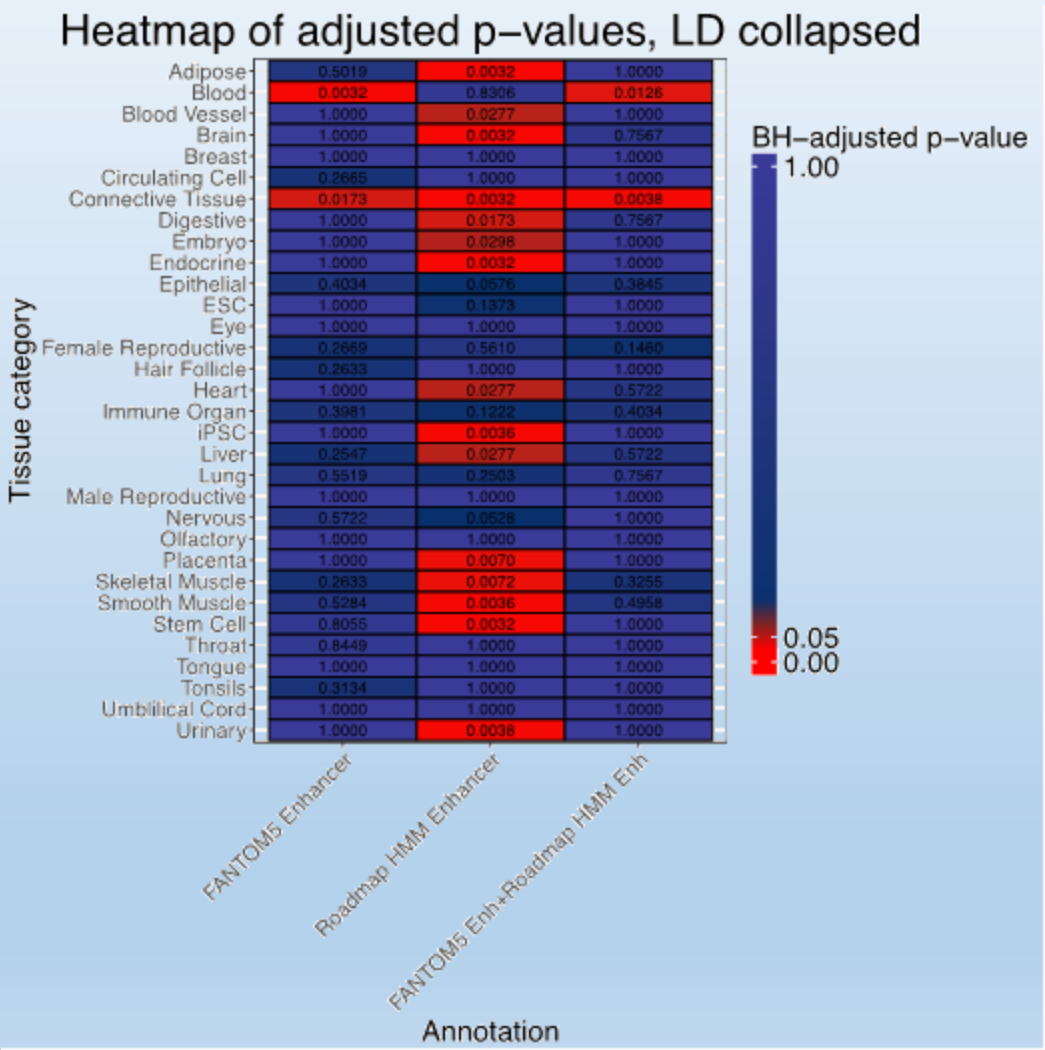


- Starting from 19 IGAP phase-I tag SNPs, grab all others within 500 kb with p-value within an order of magnitude and same effect direction
- Then LD prune these and re-expand into full LD blocks



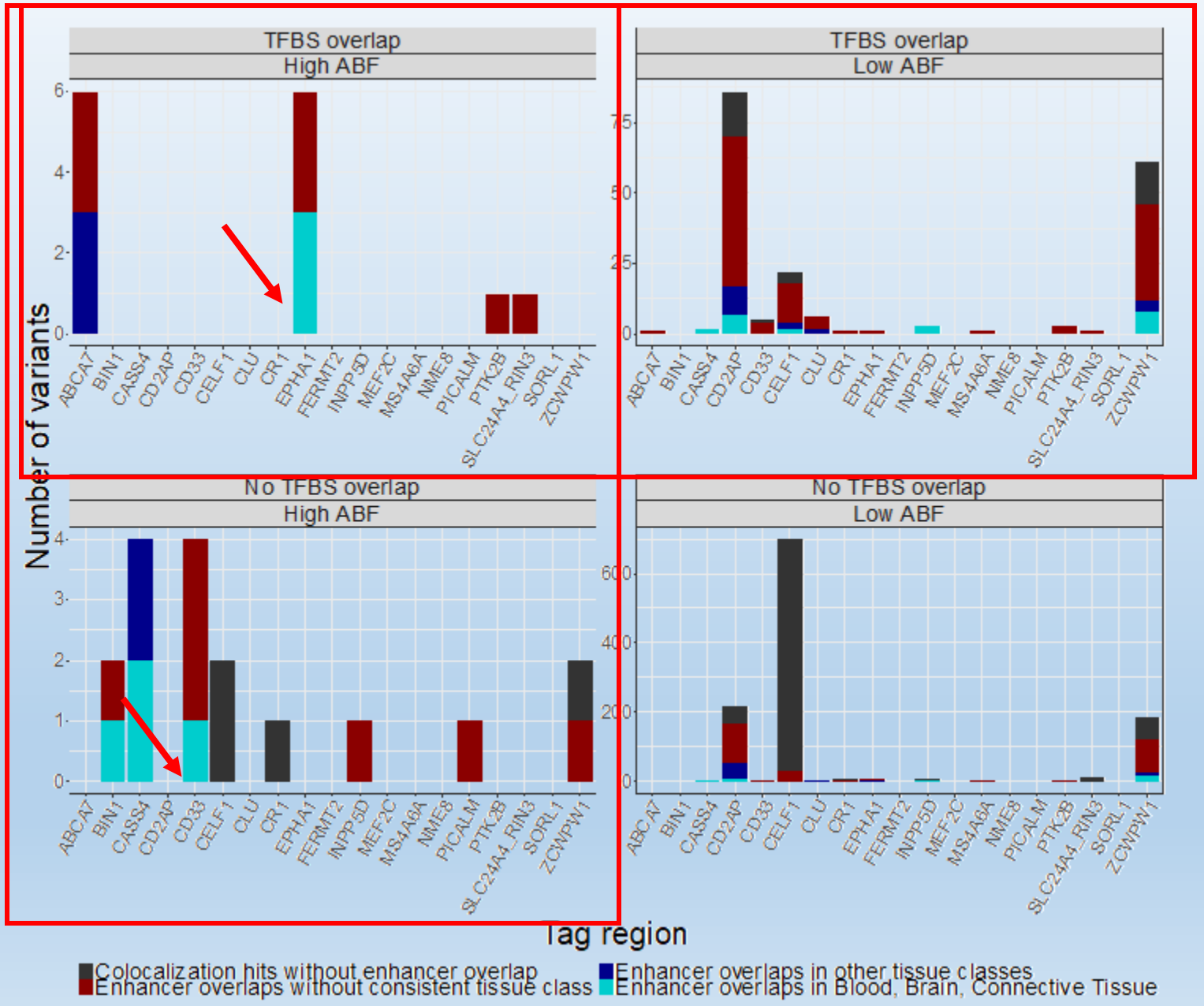
Enhancer enrichment in IGAP





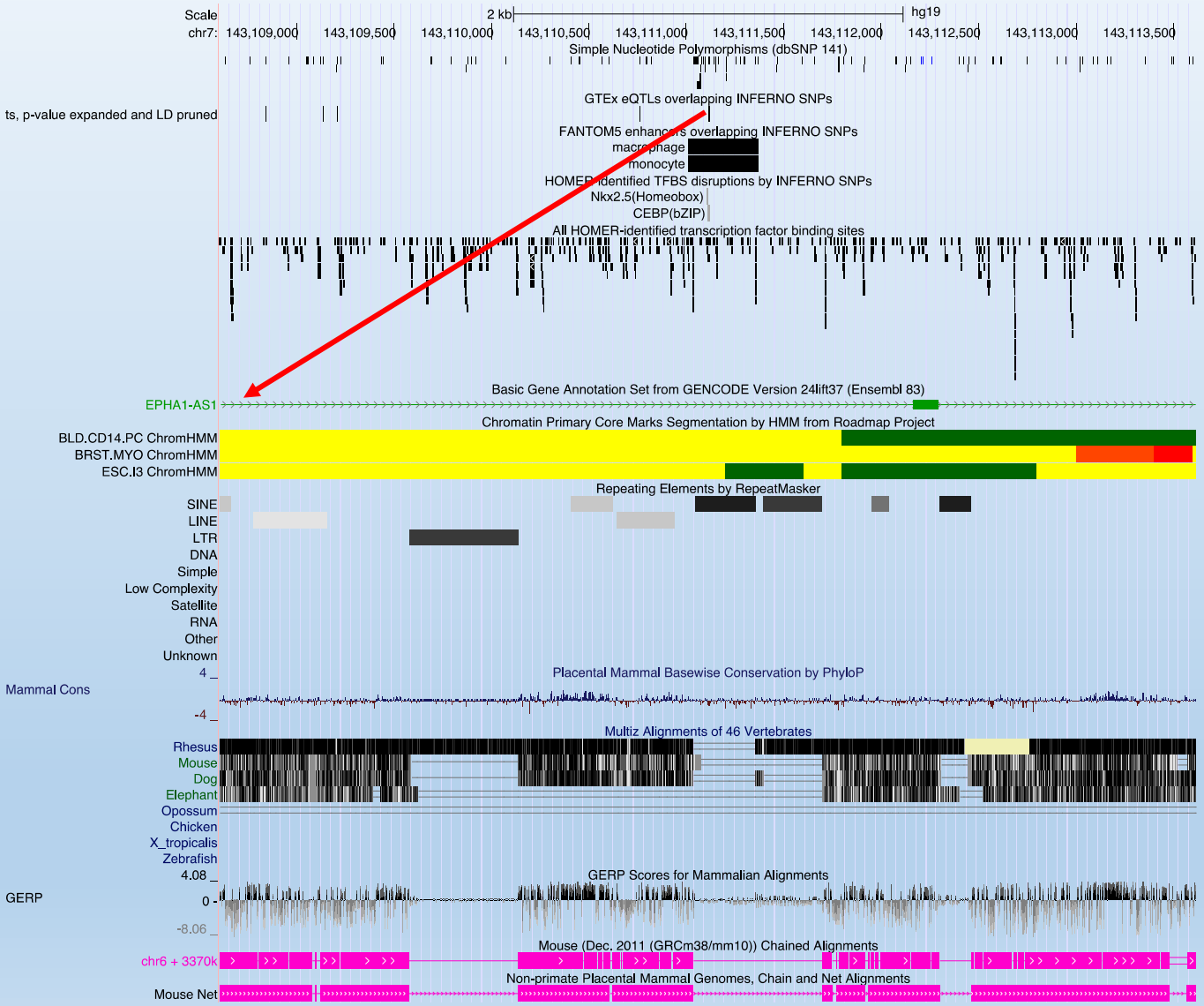
Two approaches for prioritizing signals

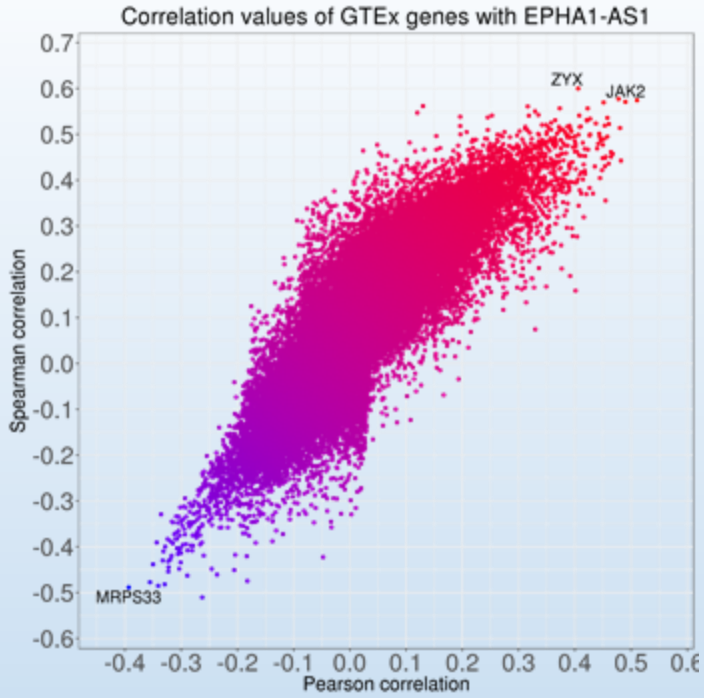
- Detected 154 strongly co-localized eQTL – GWAS signals across 15 tag regions
- Define variant sets accounting for >= 0.5 of cumulative probability underlying each co-localization signal
  - Each variant has an approximate Bayes factor (ABF) for its probability of being causal
- Prioritize variants that have enhancer overlaps from same tissue class as eQTL signal
- Then further prioritize by one or both:
  - Variants overlapping TFBSs
  - Variants with individually high ABFs

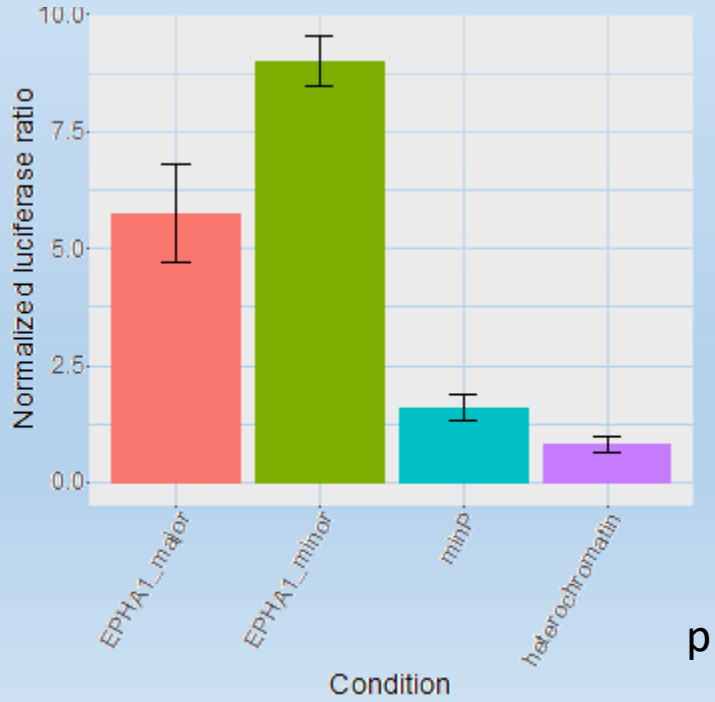


Tag Region	Affected mechanism and evidence	Direction of effect
ABCA7	Digestive system regulation of ABCA7 and CNN2, high ABF variant	Risk allele ↑ABCA7 and CNN2 expression
BIN1	Lymphocyte regulation of BIN1, high ABF variant	Risk allele ↑BIN1 expression
CASS4	HOXD13-mediated enhancer with blood eQTL for CASS4 High ABF variant for digestive CASS4 eQTL	Protective allele ↓blood ↑fibroblasts
CD2AP	Strong homeobox TF disruption in enhancer for **RP11-385F7.1** in blood and brain, affecting GTPase signaling	Risk allele ↓brain / other tissues
CD33	Whole blood regulation of CD33, tag variant colocalized with eQTL, high ABF variant	Protective tag variant ↓CD33 expression
CELF1	Brain signal for **RP11-750H9.5**, moderate TF disruption, affecting immune regulatory hub	Risk allele ↓lncRNAs
EPHA1	Very strong ABF for rs11765305 affecting **EPHA1-AS1** (→JAK2) and two taste receptor signals in blood (monocytes) with strengthened CEBPB motif	Protective allele ↑↑ *EPHA1-AS1* expression
INPP5D	Blood signal for INPP5D, strong disruption of Homeobox TFs and moderate on other TFs	Risk allele ↓ *INPP5D* expression
ZCWPW1	One SNP strongly disrupts several motifs and colocalizes with GTEx brain eQTLs for GAL3ST4, PVRIG, and STAG3	Protective allele Varying regulatory effects in tissues & genes (prioritized variants downregulate target genes)

EPHA1 region



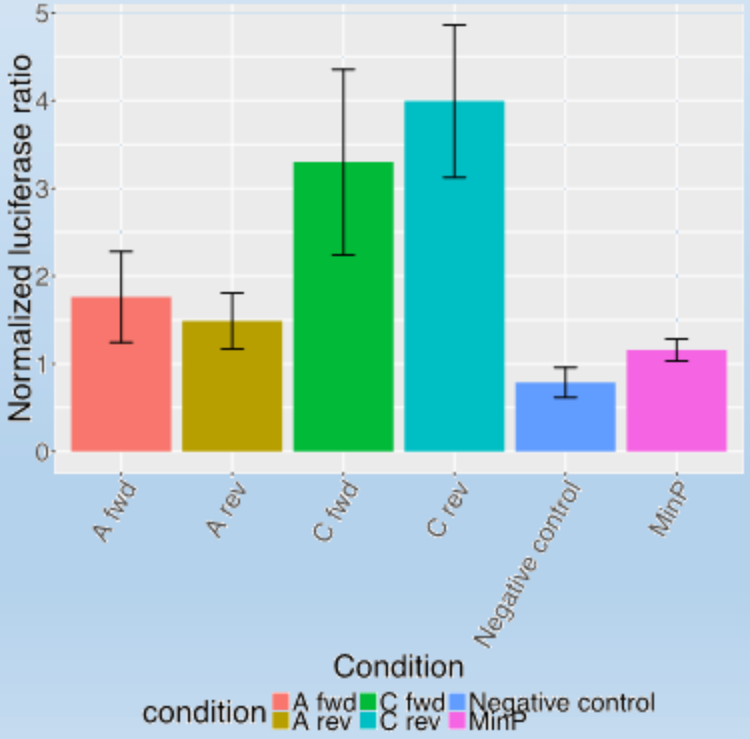




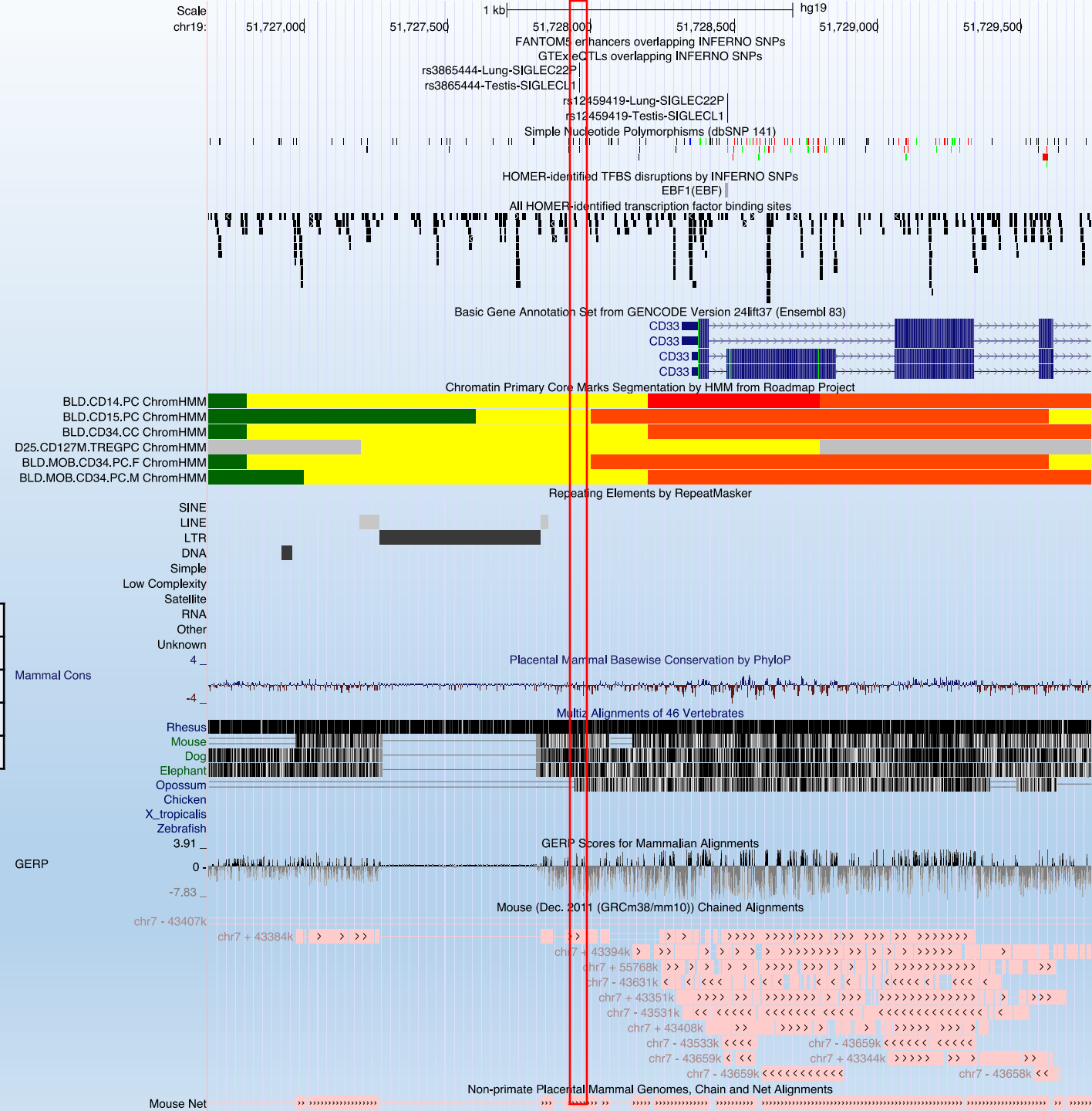
p = 0.018

CD33 region

- Variant rs386544 is tag variant and underlies whole blood eQTL for CD33 ($P(H_4) = 0.966$, ABF = 0.508)
- This variant has been quite well studied
- Risk allele C: increases in CD33 brain protein level, increases in microglia expression level, decrease in a specific isoform in brain, and increases in monocyte expression level

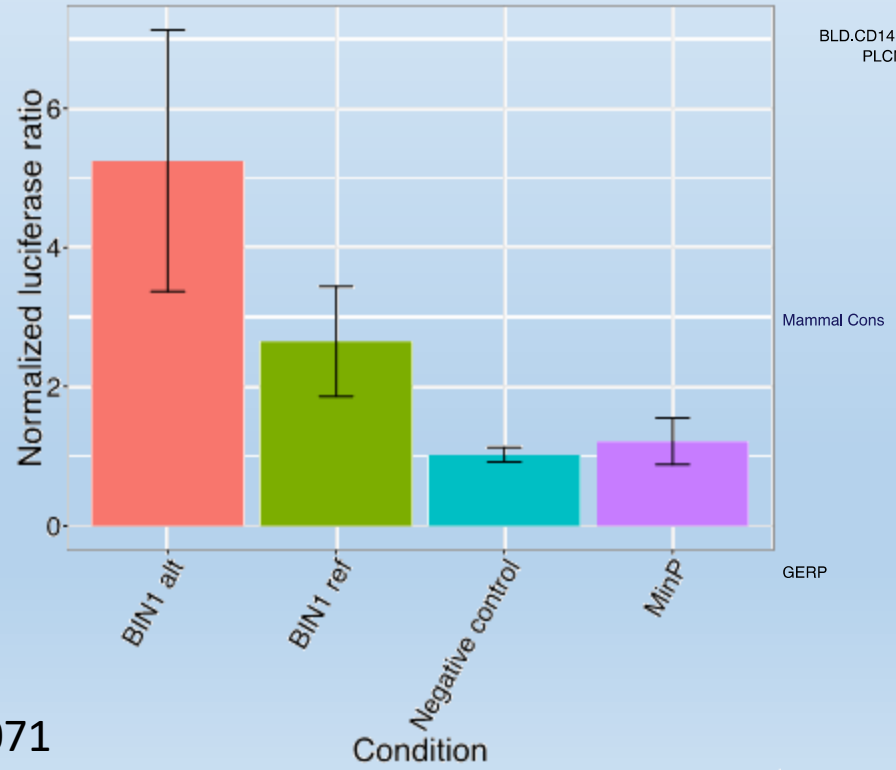


Vec 1	Vec 2	p
A fwd	C fwd	0.067
A fwd	C rev	0.015
A rev	C fwd	0.048
A rev	C rev	0.015

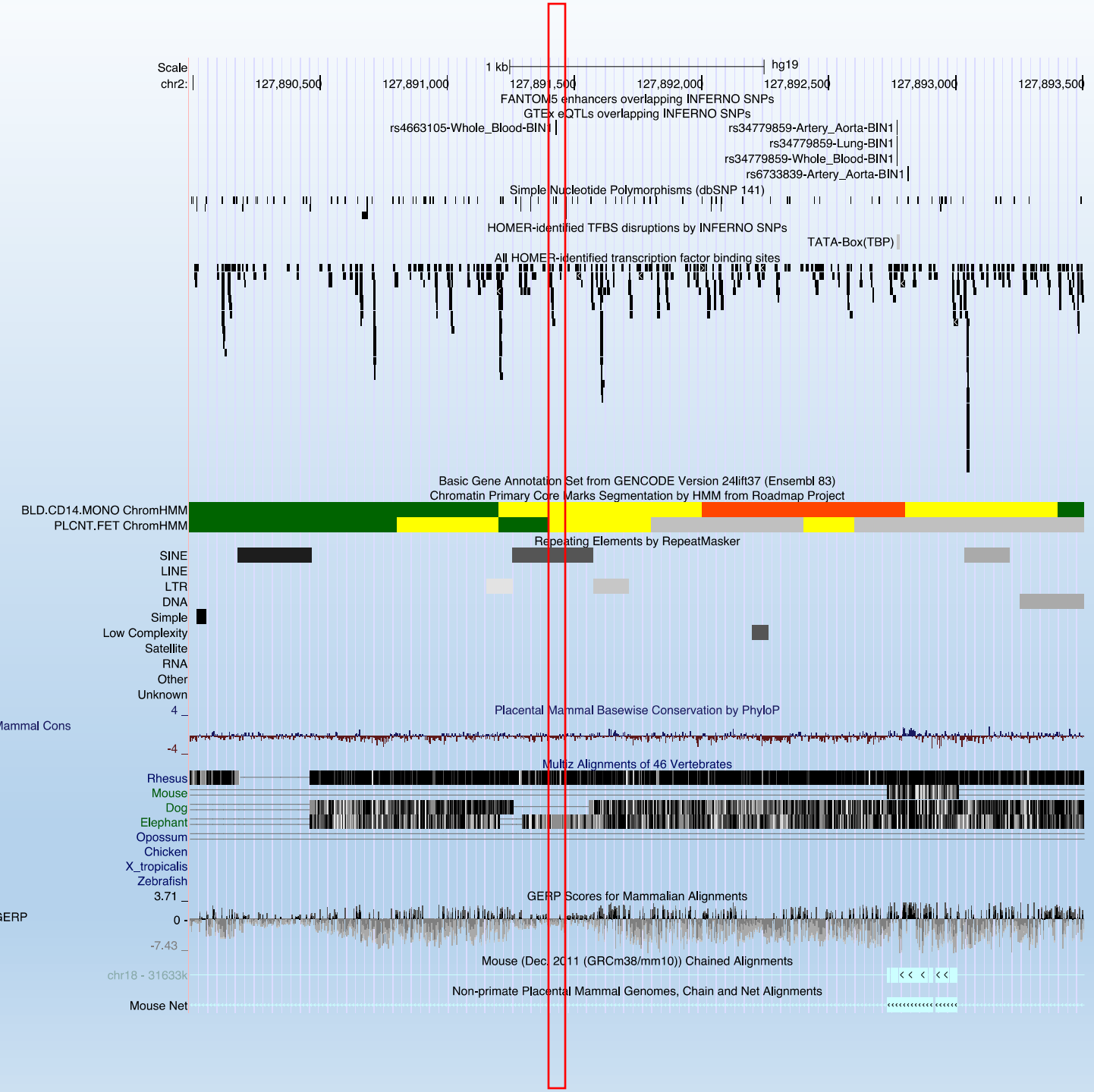


BIN1 region

- Variant rs4663105 is a few kb away from BIN1
- Whole blood BIN1



p = 0.071



Acknowledgements

- **Li-San Wang**
- **Jerry Schellenberg**
- **Mitchell Tang**
- Elizabeth Mlynarski
- Beth Dombrowski
- Jessica Way
- Fanny Leung
- Nikolaos Vrettos
- Yi-Fan Chou
- Pavel Kuksa
- Jacky Wang
- Adam Naj
- Mingyao Li
- Edward Lee
- Casey Brown
- Barbara Engelhardt
- T32 AG000255-18: Training in Age Related Neurodegenerative Diseases (**Virginia Lee**)

Brain-related traits with noncoding signals

Phenotype	Acronym	Number of input variants	Citation
Alzheimer's Disease	**IGAP**	**19 significant**	**Lambert et al., 2013**
Frontotemporal Dementia	FTD	7 significant	Ferrari et al., 2014
Corticobasal Degeneration	CBD	7 significant	Kouri et al., 2015
Amyotrophic Lateral Sclerosis	ALS	7 significant	van Rheenen et al., 2016
Parkinson's Disease	PD	23 significant	Nalls et al., 2014
Progressive Supranuclear Palsy	PSP	7 significant, 12 suggestive	Wang et al., unpublished
Schizophrenia	**SCZ**	**111 significant**	**Ripke et al., 2014**
Attention Deficit Hyperactivity Disorder	ADHD	7 suggestive	Neale et al., 2010
Autism Spectrum Disorder	ASD	51 suggestive	Preliminary results from PGC
Bipolar Disorder	BIP	43 significant	Sklar et al., 2011
Major Depressive Disorder	MDD	11 suggestive	Ripke et al., 2013

MAPT region complexity

- 55 eQTL genes in 44 different tissues
- ~3k variants in expanded H1/H2 haplotype
  - Overlap 11 unique FANTOM5 enhancers and 4,150 unique Roadmap ChromHMM enhancers