Integrative functional analysis identifies enhancers and lncRNAs perturbed by LOAD-associated genetic variants

Alexandre Amlie-Wolf
Li-San Wang lab, University of Pennsylvania

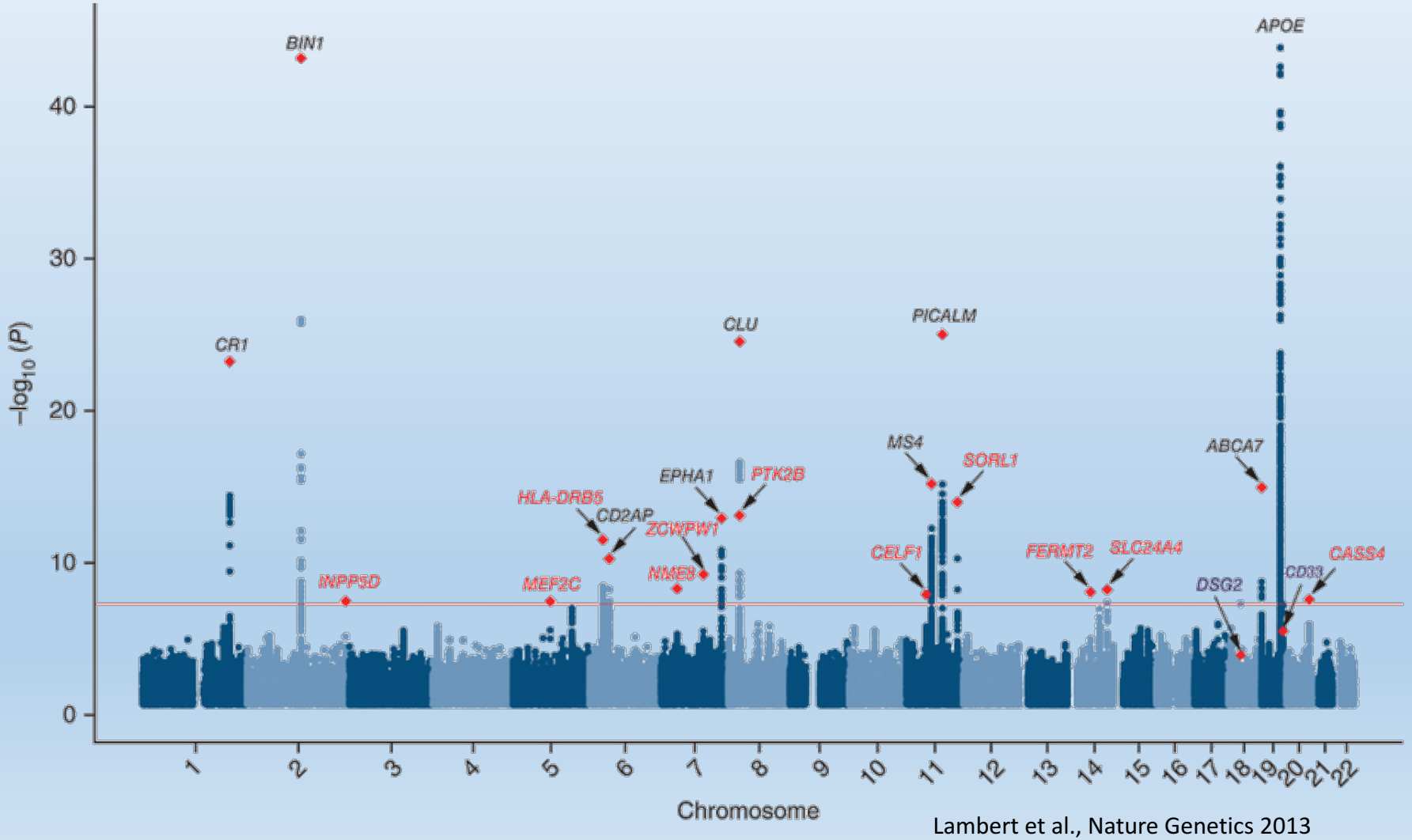
02/16/17 ADGC meeting
Cleveland, OH

Outline

- Background:
  - Noncoding regulatory elements
  - Alzheimer's disease genetics
- Description of INFERNO
- IGAP top hits enhancer overlap analysis
- GTEx eQTL colocalization
  - *EPHA1* region results
  - *CD2AP* region results
  - *CELF1* region results
- Conclusions

LOAD Genetics

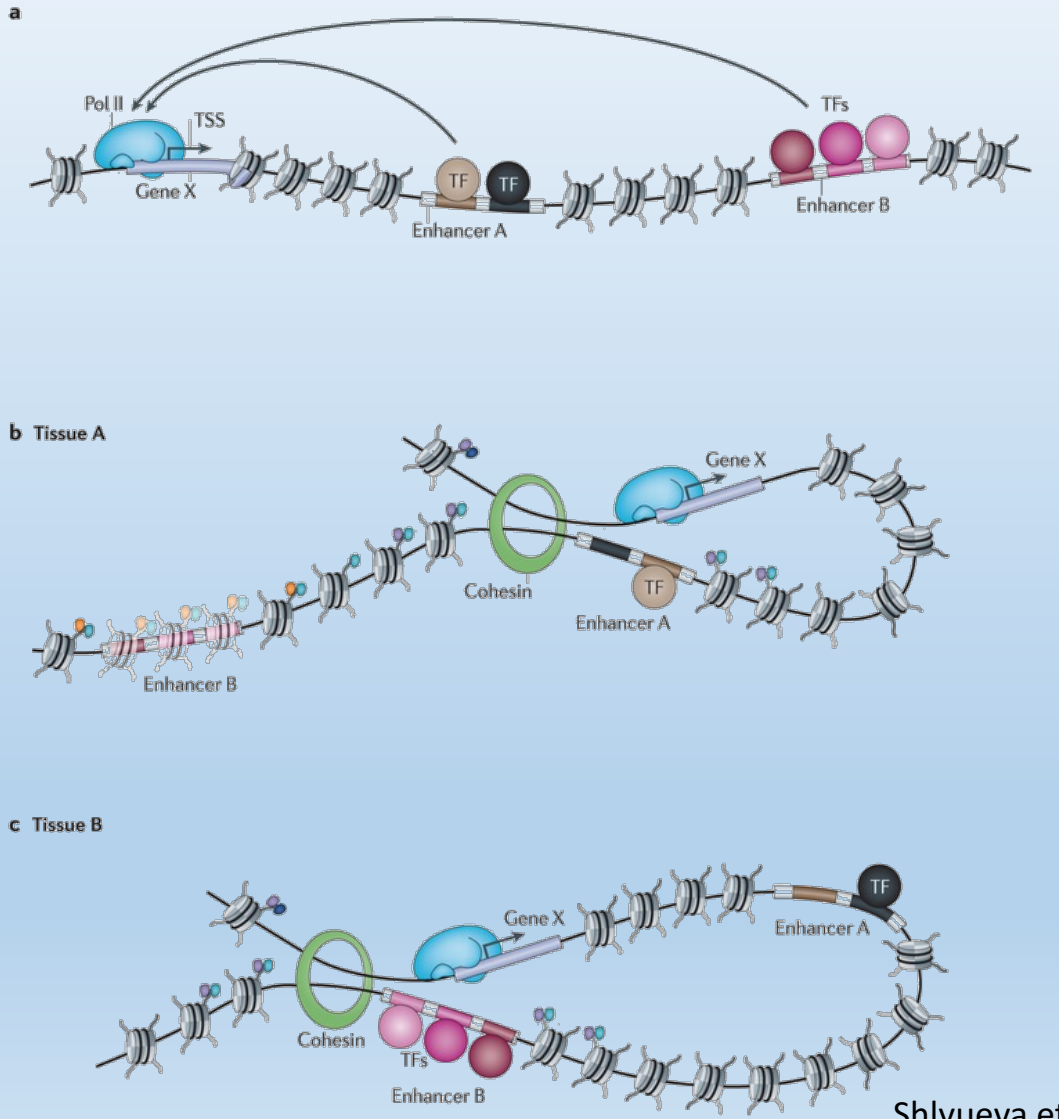
- Twin-based studies suggest 60-70% heritability
- GWAS identified dozens of associated common variants
- Almost all reside in the noncoding genome



Lambert et al., Nature Genetics 2013

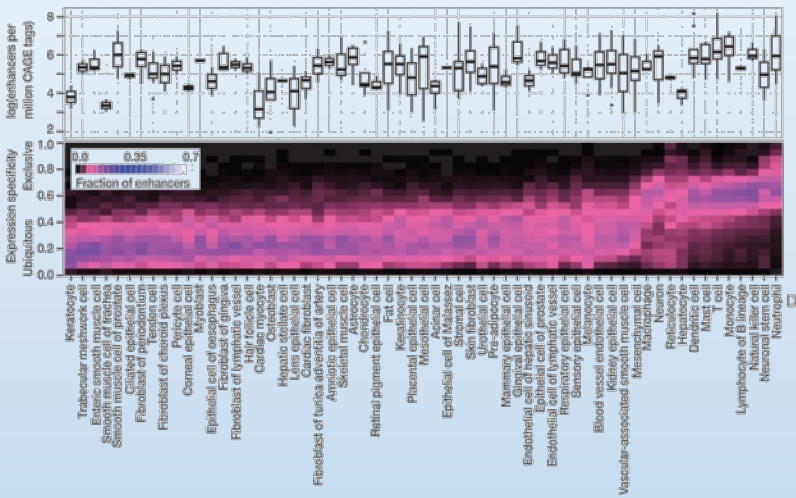
Transcriptional enhancers

- Enhancers modulate target gene transcription from any distance or relative position
- Tissue context is important: some are tissue specific and some are ubiquitous

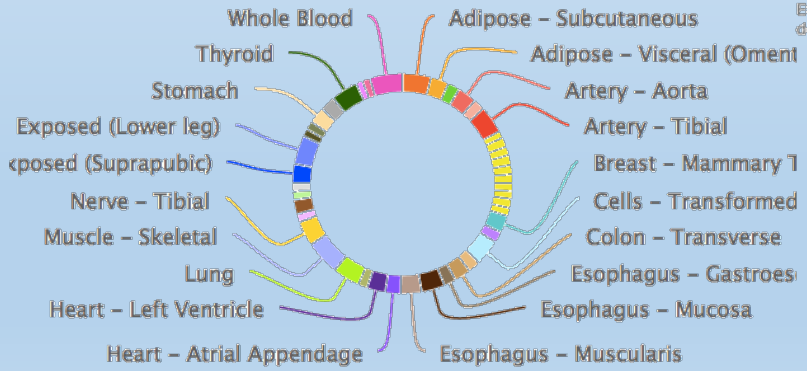


Shlyueva et al., NRG 2014

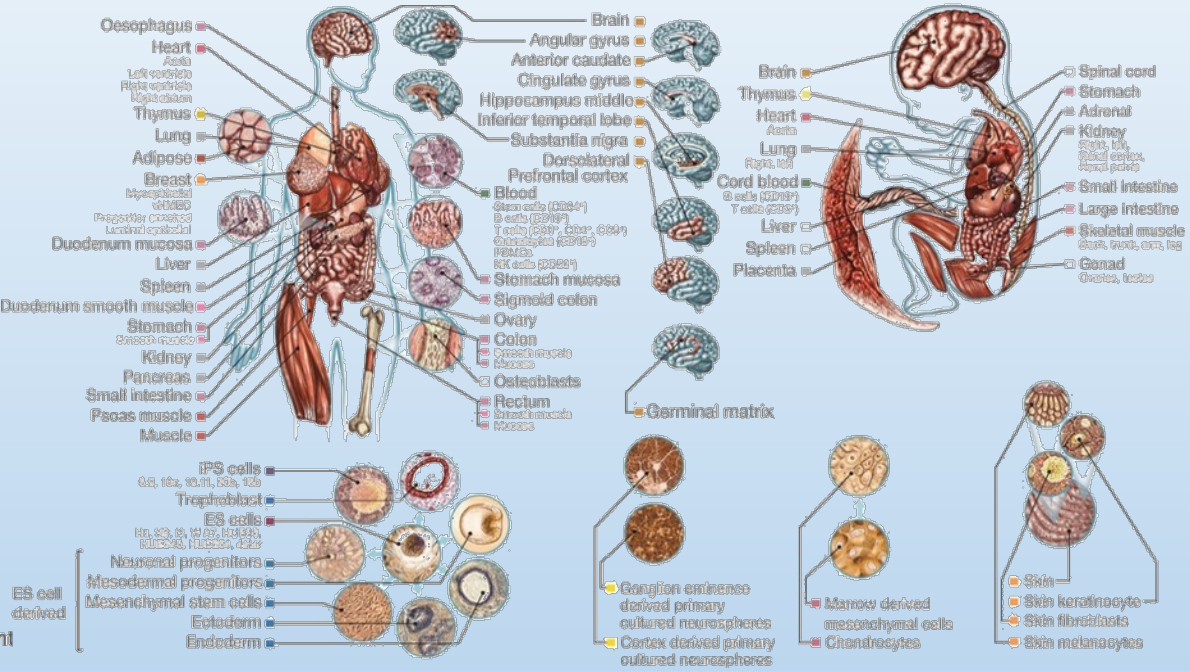
Large-scale functional genomics data sources



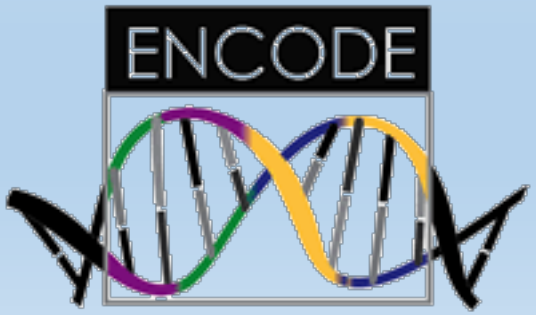
FANTOM5 eRNA Measurements



GTEx eQTLs

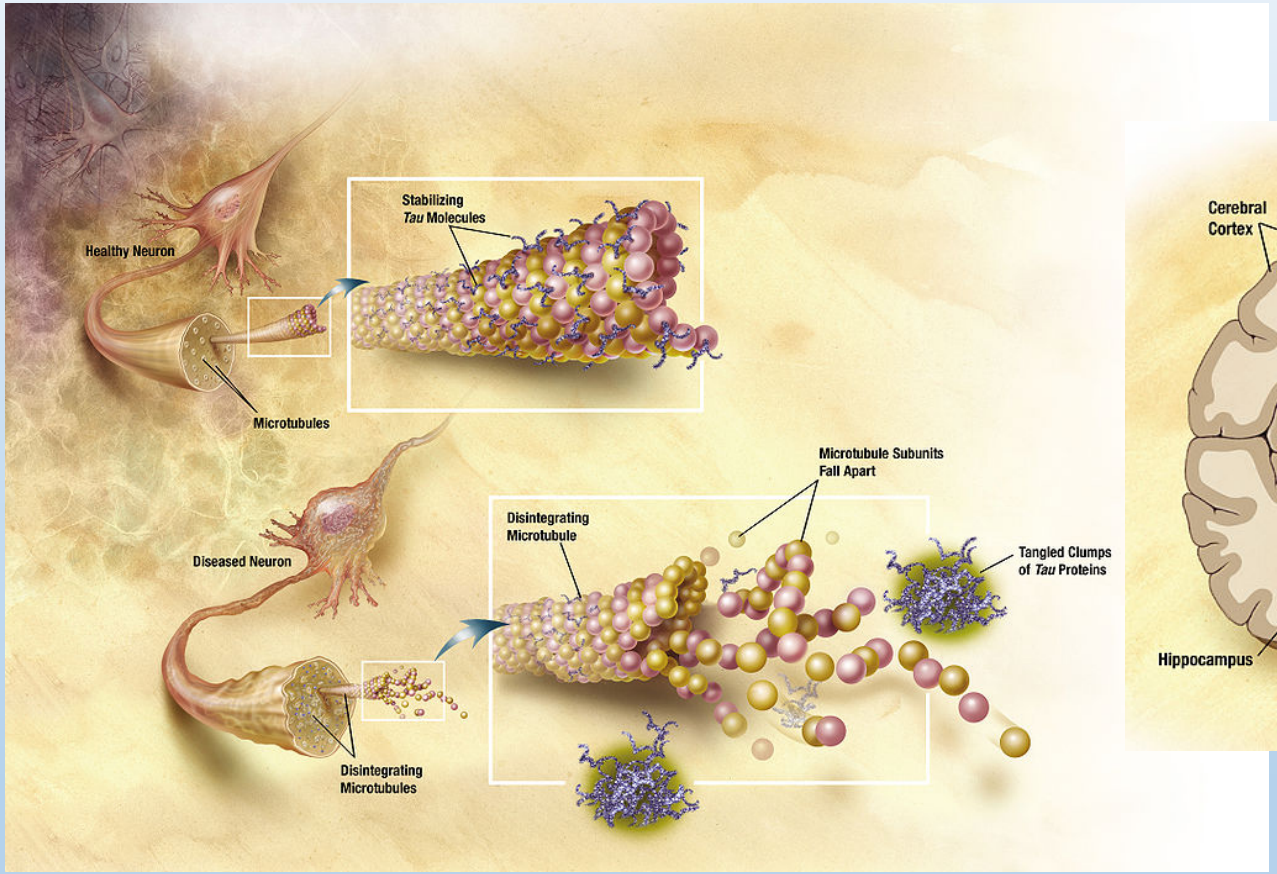


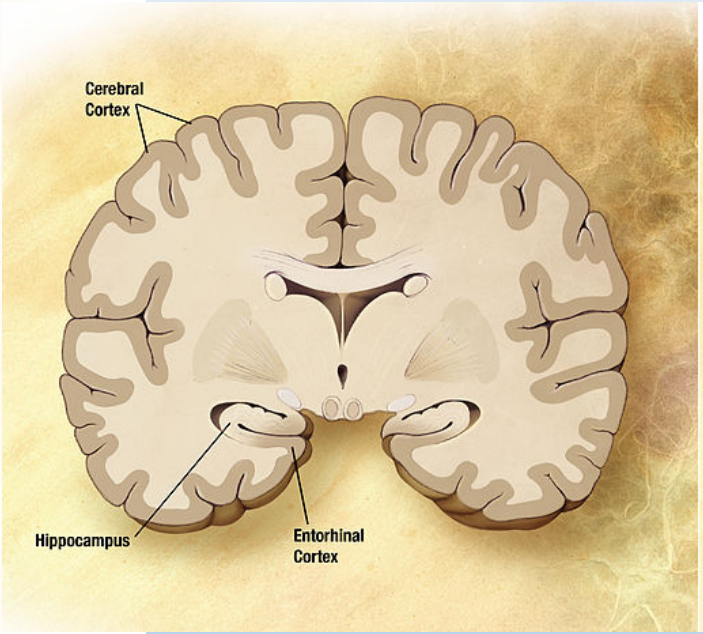
Roadmap Epigenomics Consortium

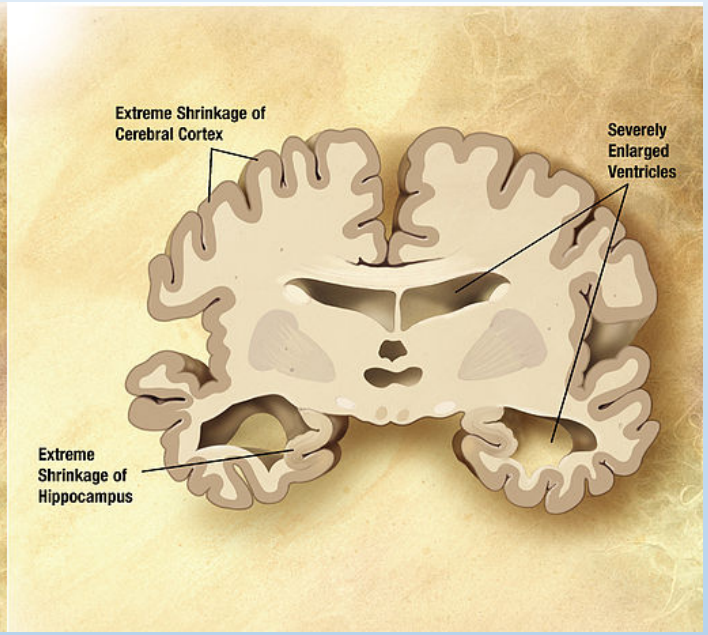


Database of small human noncoding RNAs

Alzheimer's disease (AD)



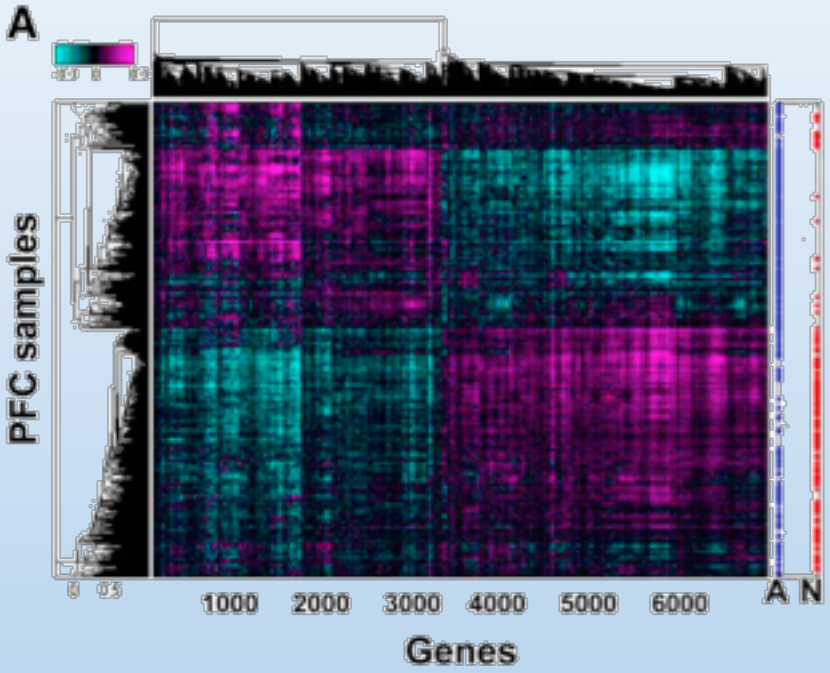




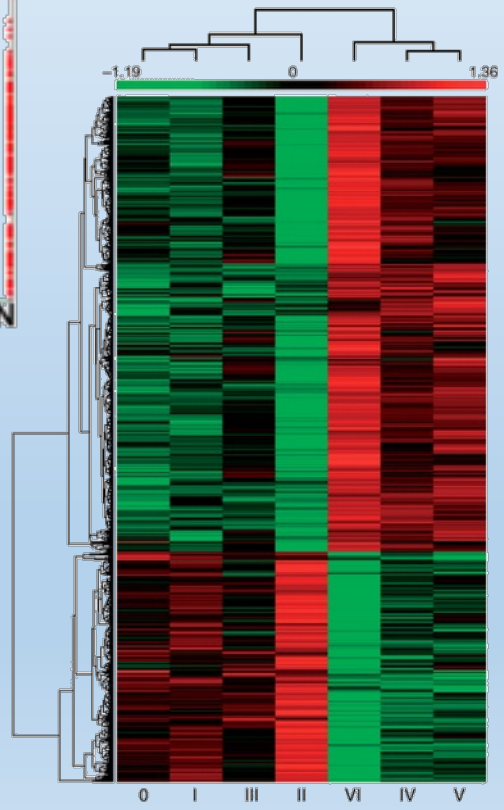
Wikipedia / NIA

- Late-onset AD (LOAD) is heritable and incurable neurodegenerative disease
- Characterized by Aß plaques and tau neurofibrillary tangles

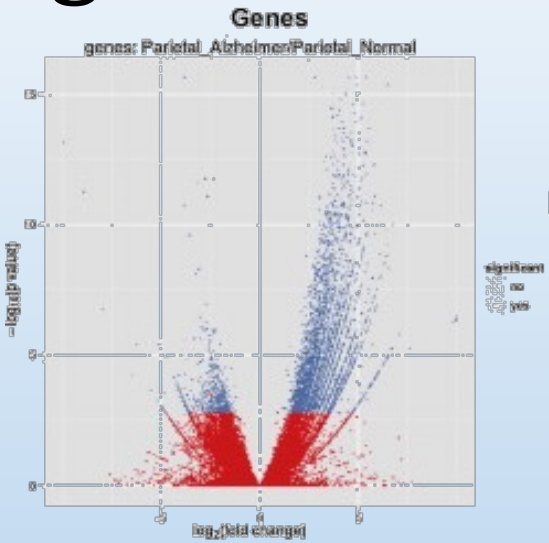
Gene expression dysregulation in AD



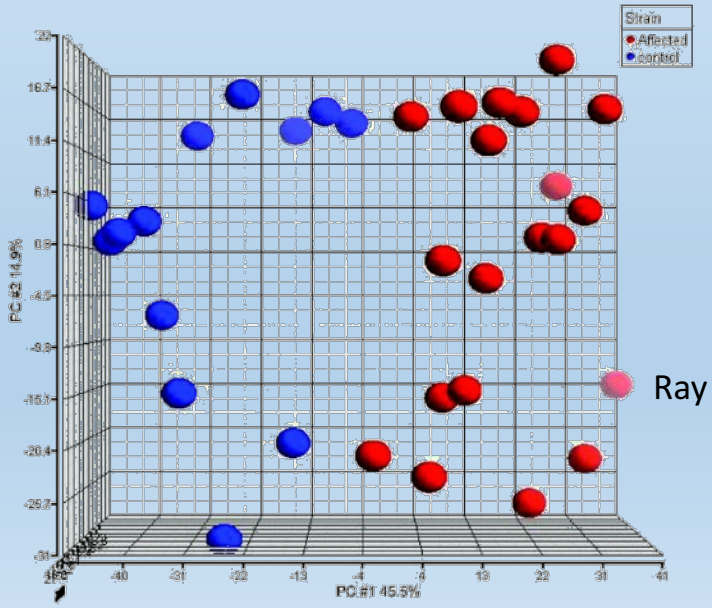
Zhang et al., Cell 2013



Bossers et al., Brain 2010



Mills et al., Neuroscience Letters 2013

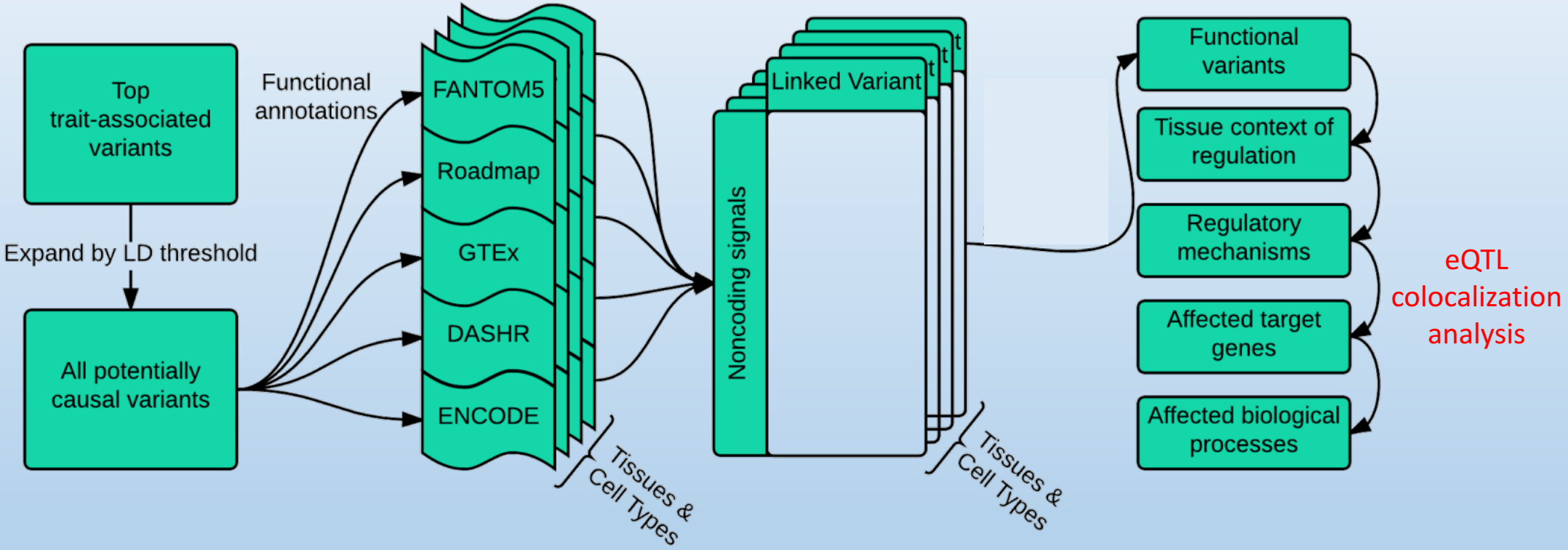


Ray et al., Genome Biology 2008

From genetic signal to regulatory process

- Hypothesis: **noncoding genetic variants associated with LOAD affect transcriptional regulatory elements**
- Goal: Identify functional variants, characterize:
  - Affected regulatory mechanisms
  - Tissue context
  - Target genes and direction of effect
- Approach:
  1. Build pipeline to overlap with diverse functional data across tissues
  2. Develop model for functional variant prioritization and characterization by integrating signal from diverse enhancer-related data sources
  3. Colocalization with eQTL signals to identify target genes

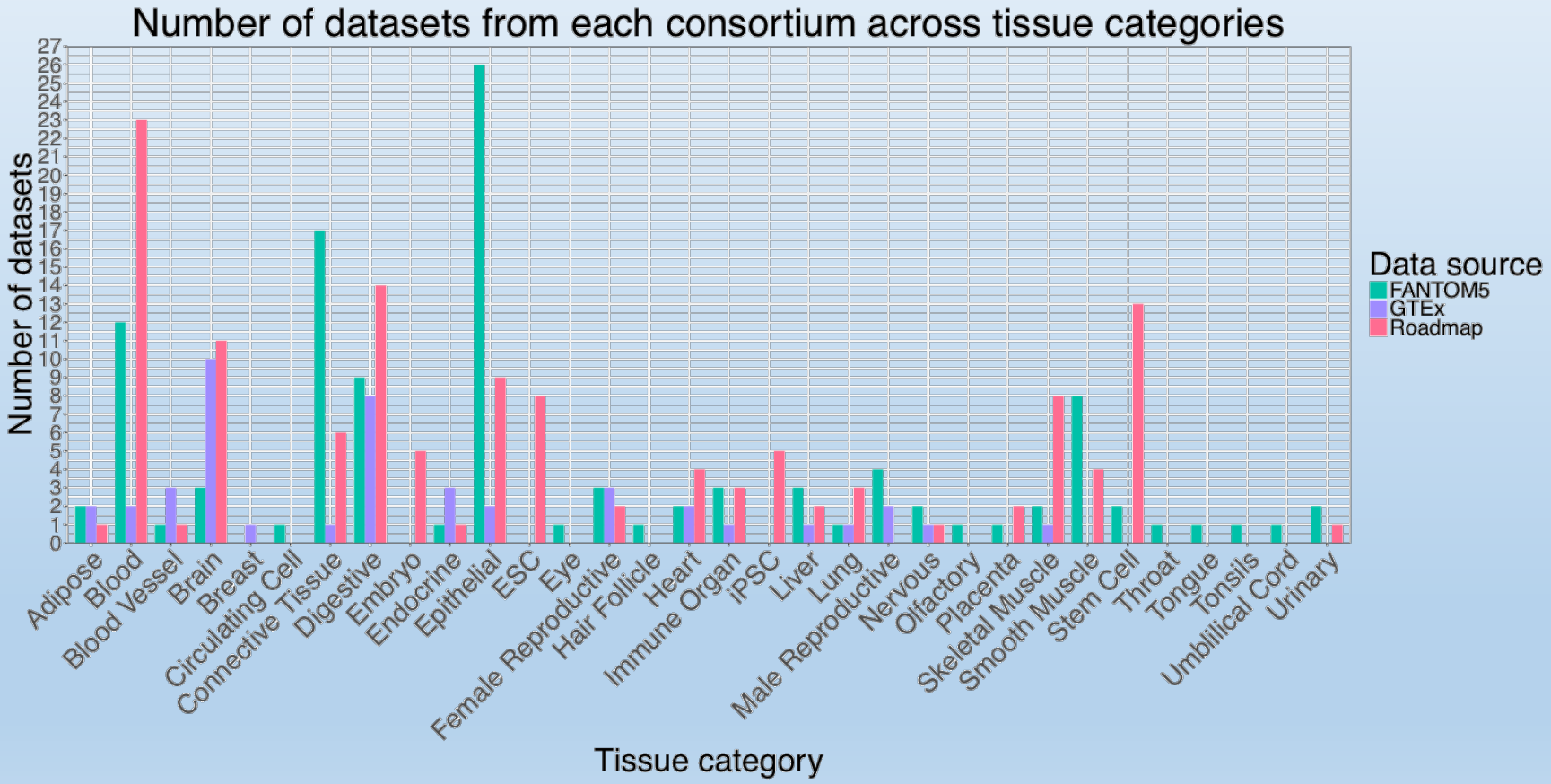
INFERNO – INFERring the molecular mechanisms of NOncoding genetic variants



- Need to account for LD structure, expand set of potential variants to include all potentially causal ones and focus on **interpretability of tissue context**
- Reproducible pipeline implemented in Python, R, and bash; will be a web server

Tissue categorization is key for integration

- Roadmap: 130 tissues and cell types
- FANTOM5: 112 tissue facets
- GTEx: 44 tissues

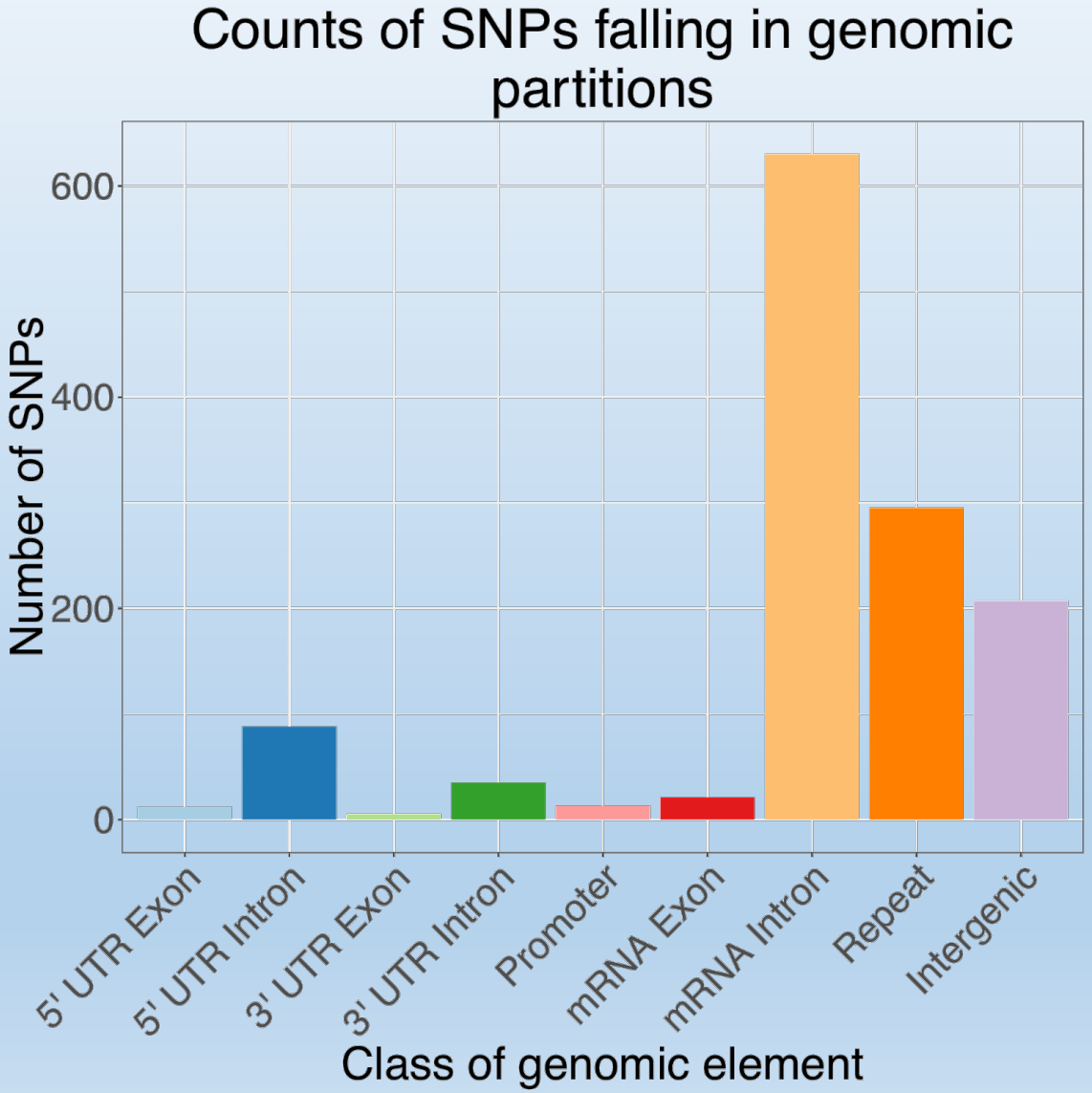


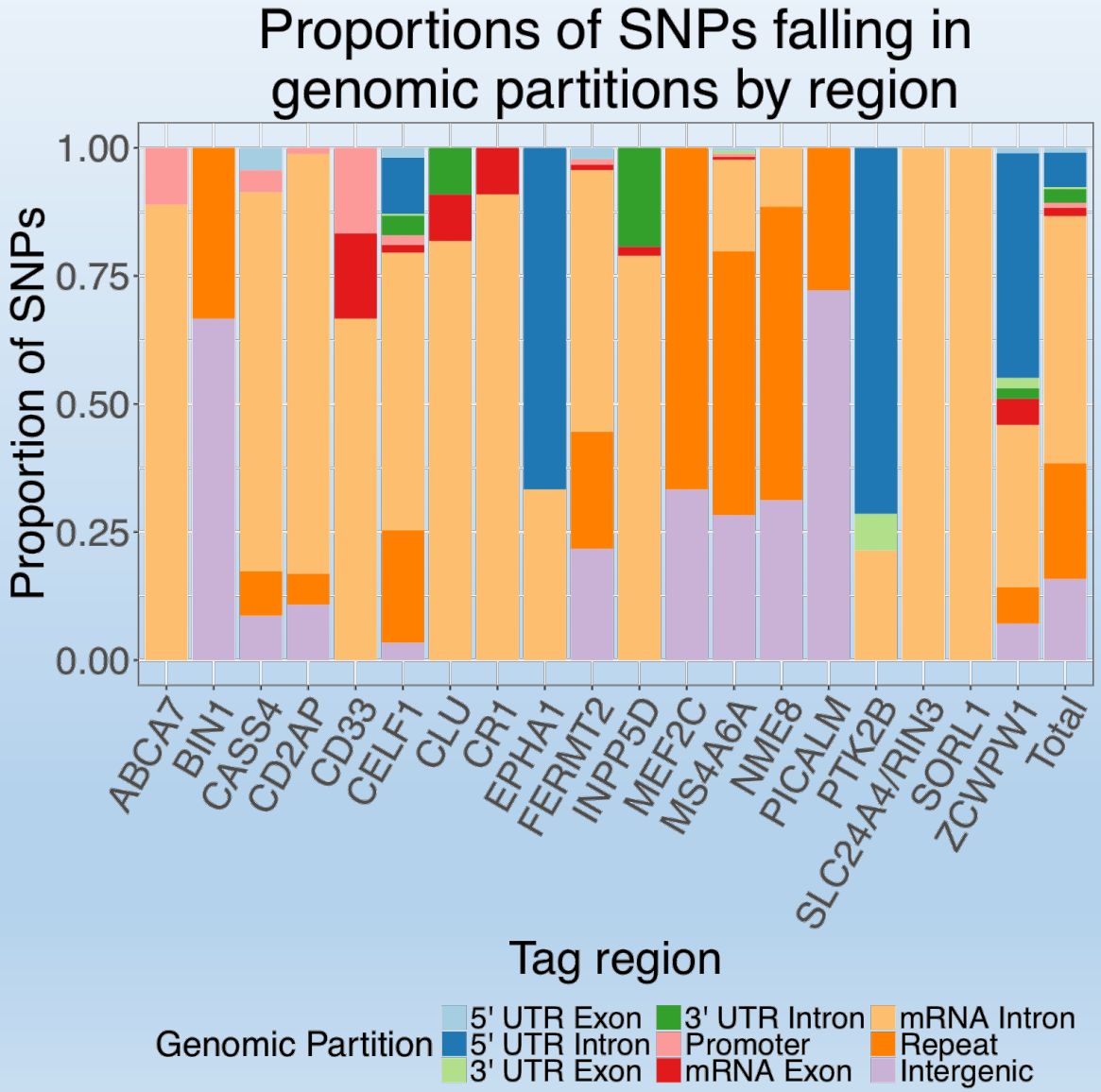
IGAP genetics data sources

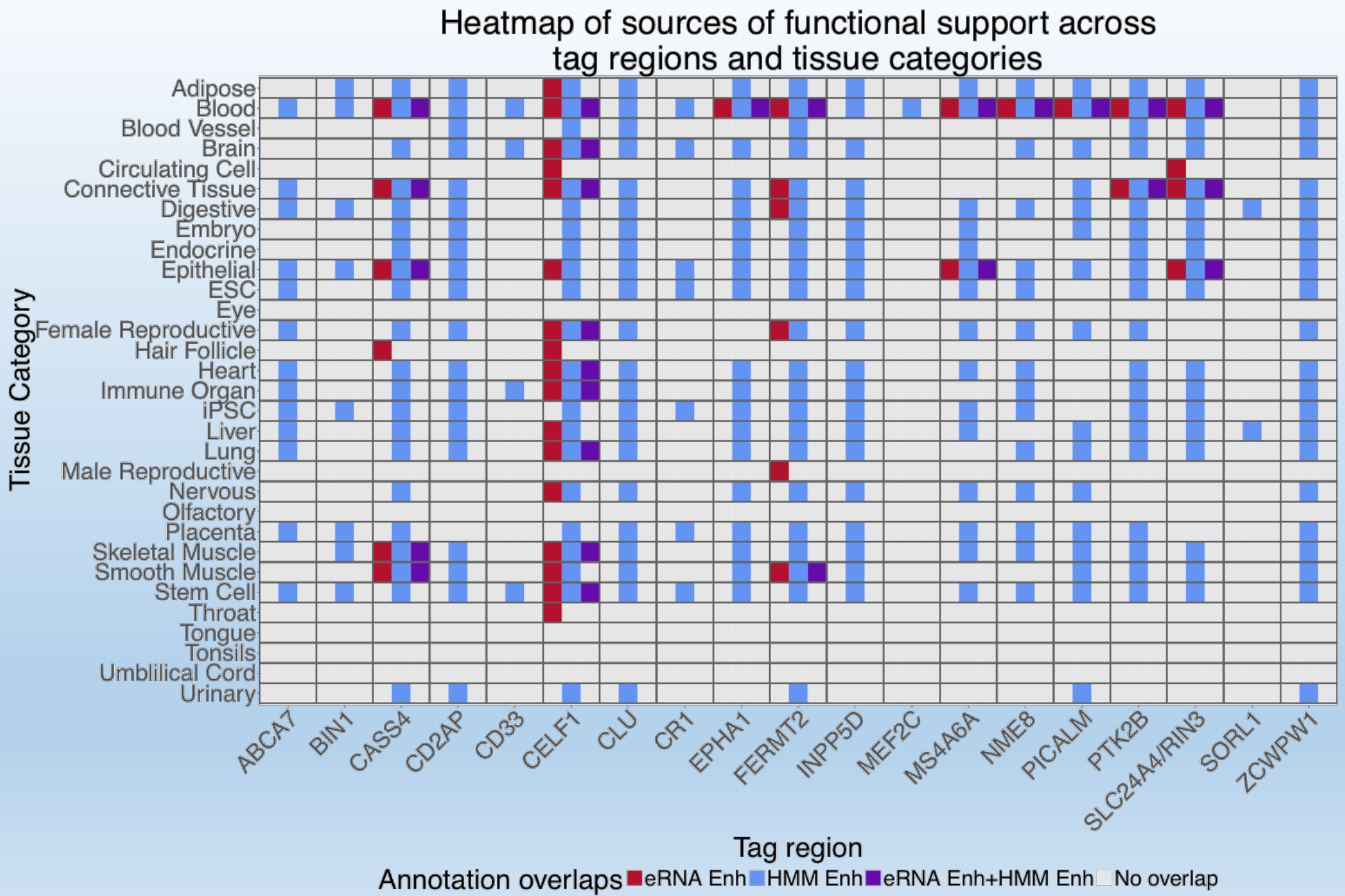
- Starting from tag SNPs, grab all others within 500 kb with p-value within an order of magnitude and same effect direction
- Then LD prune these and expand into full LD blocks

Tag variant	Gene region	# p-value expanded	# LD pruned	# LD expanded	IGAP Phase 1 Odds Ratio
rs4147929	*ABCA7*	7	2	9	1.14
rs35349669	*BIN1*	2	1	3	1.21
rs7274581	*CASS4*	20	2	23	0.87
rs10948363	*CD2AP*	80	5	83	1.10
rs3865444	*CD33*	3	1	6	0.91
rs10838725	*CELF1*	92	7	264	1.08
rs28834970	*CLU*	10	1	11	0.86
rs6656401	*CR1*	20	1	22	1.17
rs11771145	*EPHA1*	9	3	9	0.90
rs17125944	*FERMT2*	32	7	92	1.13
rs6733839	*INPP5D*	65	2	114	1.07
rs190982	*MEF2C*	2	1	3	0.92
rs983392	*MS4A6A*	80	3	173	0.90
rs1476679	*TXNDC3 / NME8*	44	5	96	0.93
rs10792832	*PICALM*	2	1	18	0.88
rs9331896	*PTK2B*	7	2	14	1.10
rs10498633	*SLC24A4*	5	3	5	0.90
rs11218343	*SORL1*	1	1	1	0.76
rs2718058	*ZCWPW1*	15	4	98	0.92
Total		**496**	**52**	**1044**	

Majority of expanded SNPs are noncoding

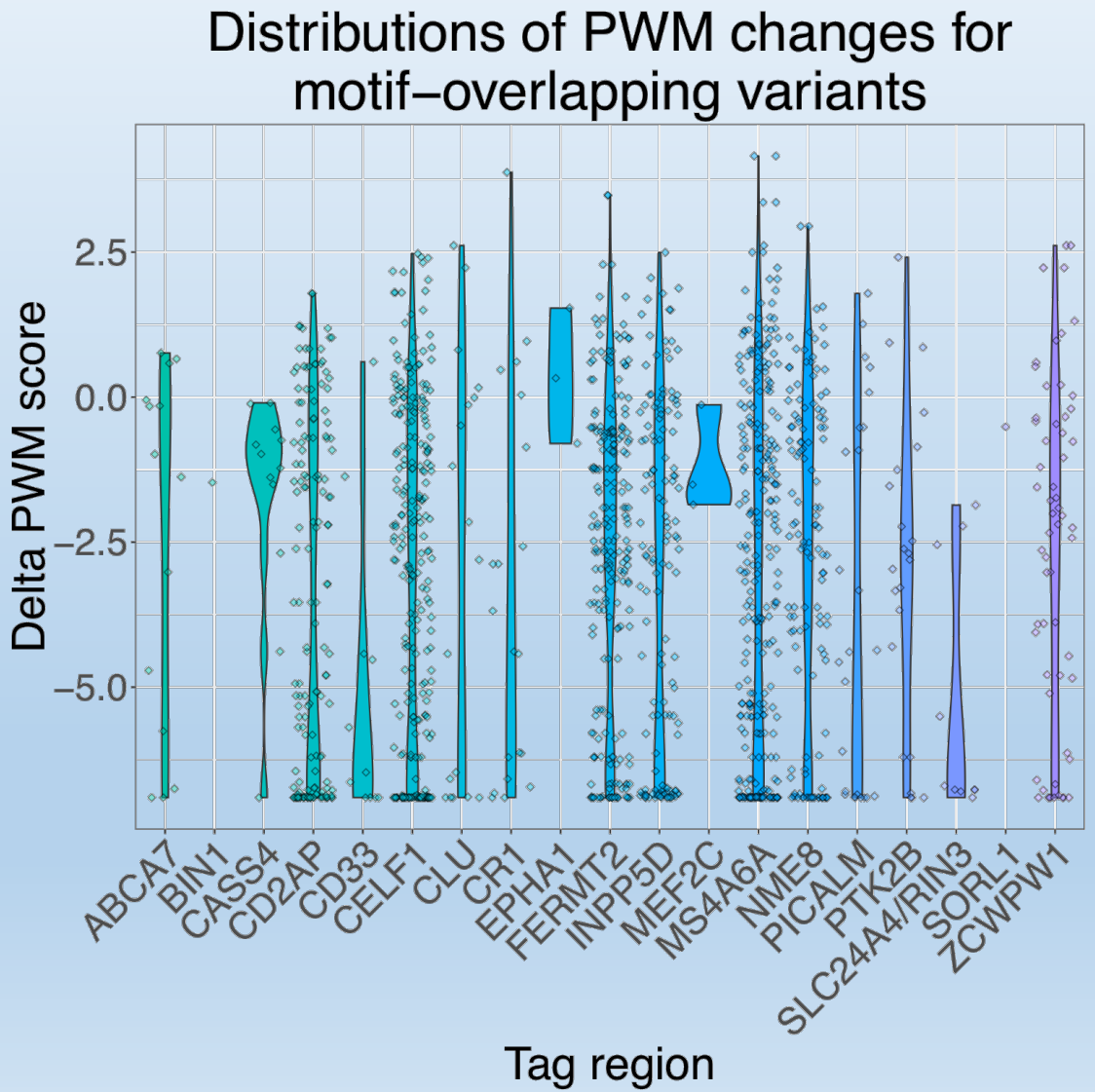





Heatmap of sources of functional support across tag regions and tissue categories
Tissue Category
Adipose
Blood
Blood Vessel
Brain
Circulating Cell
Connective Tissue
Digestive
Embryo
Endocrine
Epithelial
ESC
Eye
Female Reproductive
Hair Follicle
Heart
Immune Organ
iPSC
Liver
Lung
Male Reproductive
Nervous
Olfactory
Placenta
Skeletal Muscle
Smooth Muscle
Stem Cell
Throat
Tongue
Tonsils
Umblilical Cord
Urinary
ABCA7
BIN1
CASS4
CD2AP
CD33
CELF1
CLU
CR1
EPHA1
FERMT2
INPP5D
MEF2C
MS4A6A
NME8
PICALM
PTK2B
SLC24A4/RIN3
SORL1
ZCWPW1
Tag region
Annotation overlaps
eRNA Enh
HMM Enh
eRNA Enh+HMM Enh
No overlap

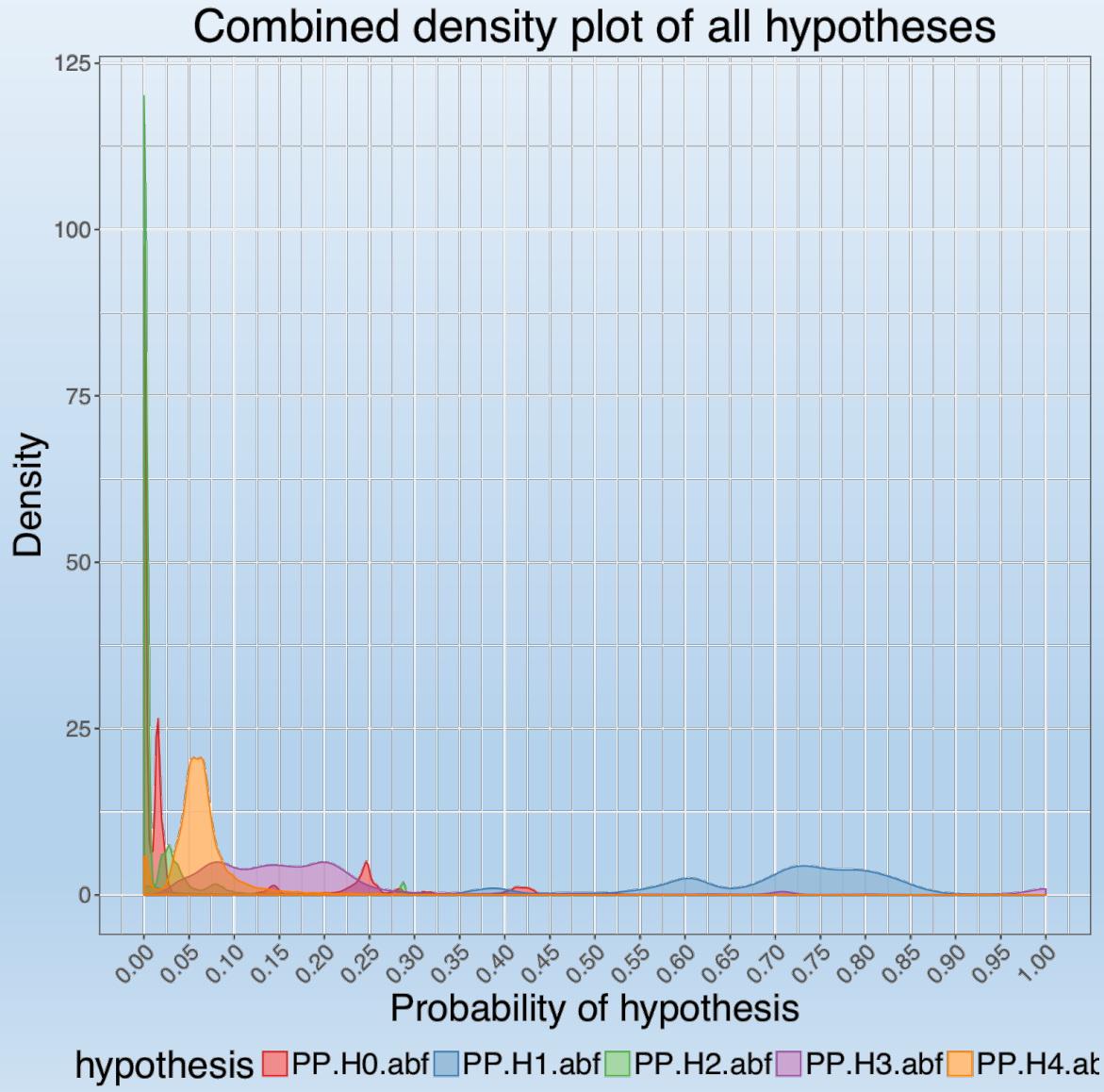
Overlap with HOMER sequence-based TFBSs

- Calculate ΔPWM score (reference allele score – alternate allele score)
- 451 / 1,044 variants overlap TFBSs for 191 TFs
- 1,398 total overlaps – 1,152 disrupt, 245 create stronger site, 1 is the same

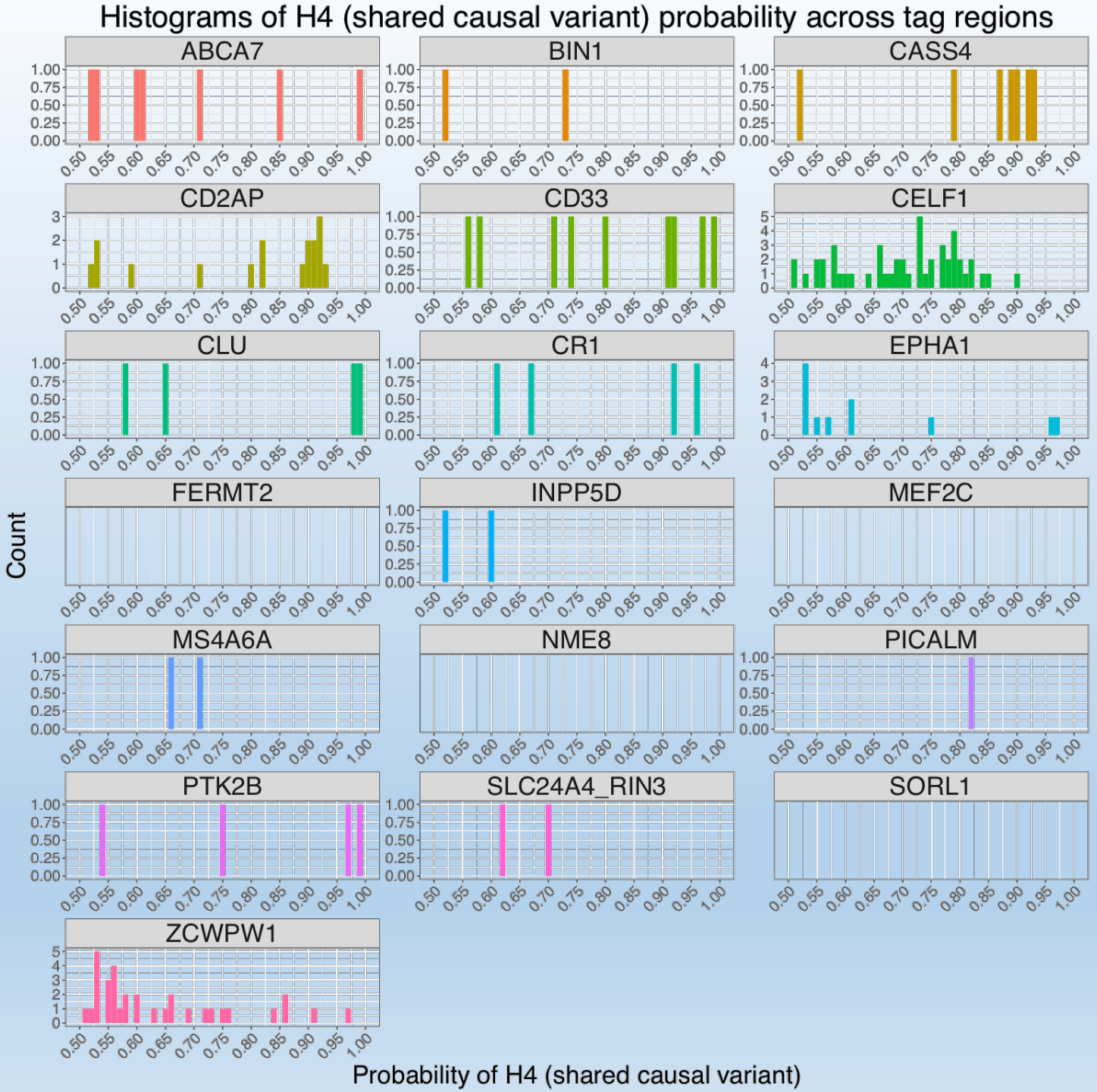


GTEx – IGAP colocalization analysis

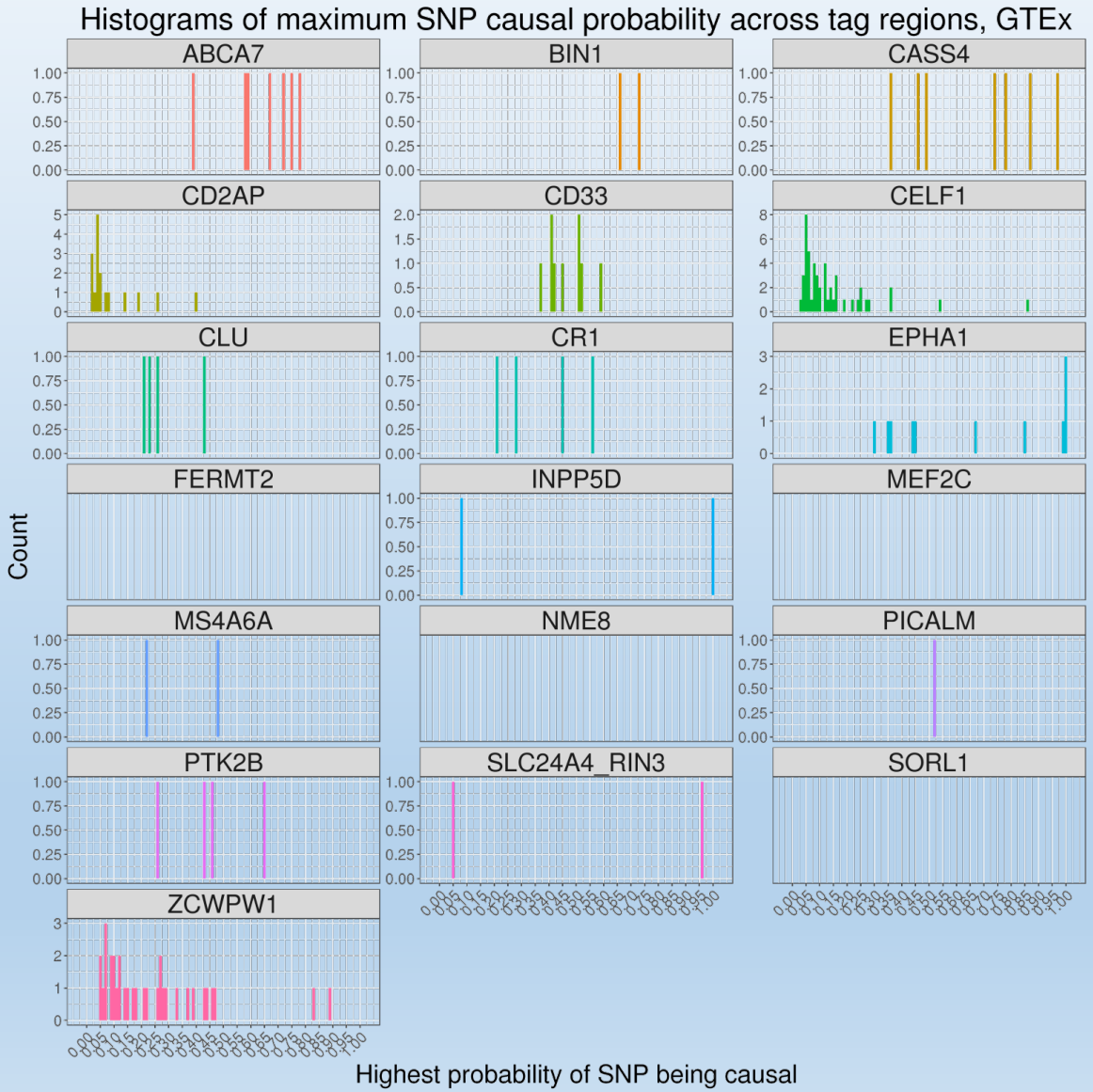
- Using COLOC R package
  - H_0 : no causal variant in either
  - H_1 : causal variant in GWAS, not for eQTL
  - H_2 : causal variant for eQTL, not for GWAS
  - H_3 : separate causal variants for GWAS and eQTL
  - **H_4 : shared causal variant for GWAS and eQTL signal**
- In each tag region and GTEx tissue, take all the genes tested for the tag SNP then perform colocalization analysis
  - all the SNPs tested for eQTL with each gene and all SNPs within 500 kb of the tag variant

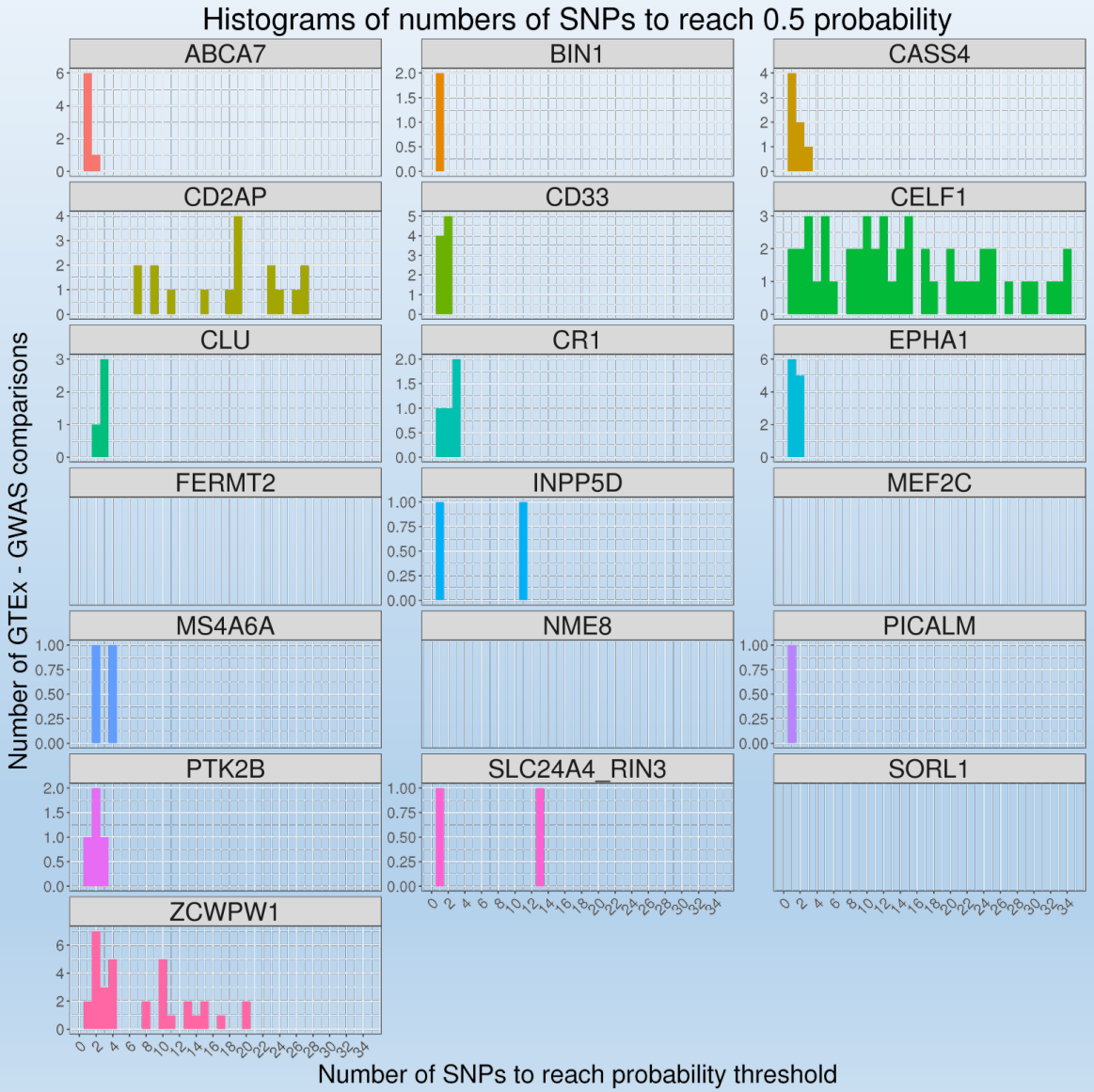


- 154 eQTL gene – GWAS comparisons have $P(H_4) > 0.5$ across 15 tag regions
- 77 unique SNPs, 26 not in p-value and LD-expanded set

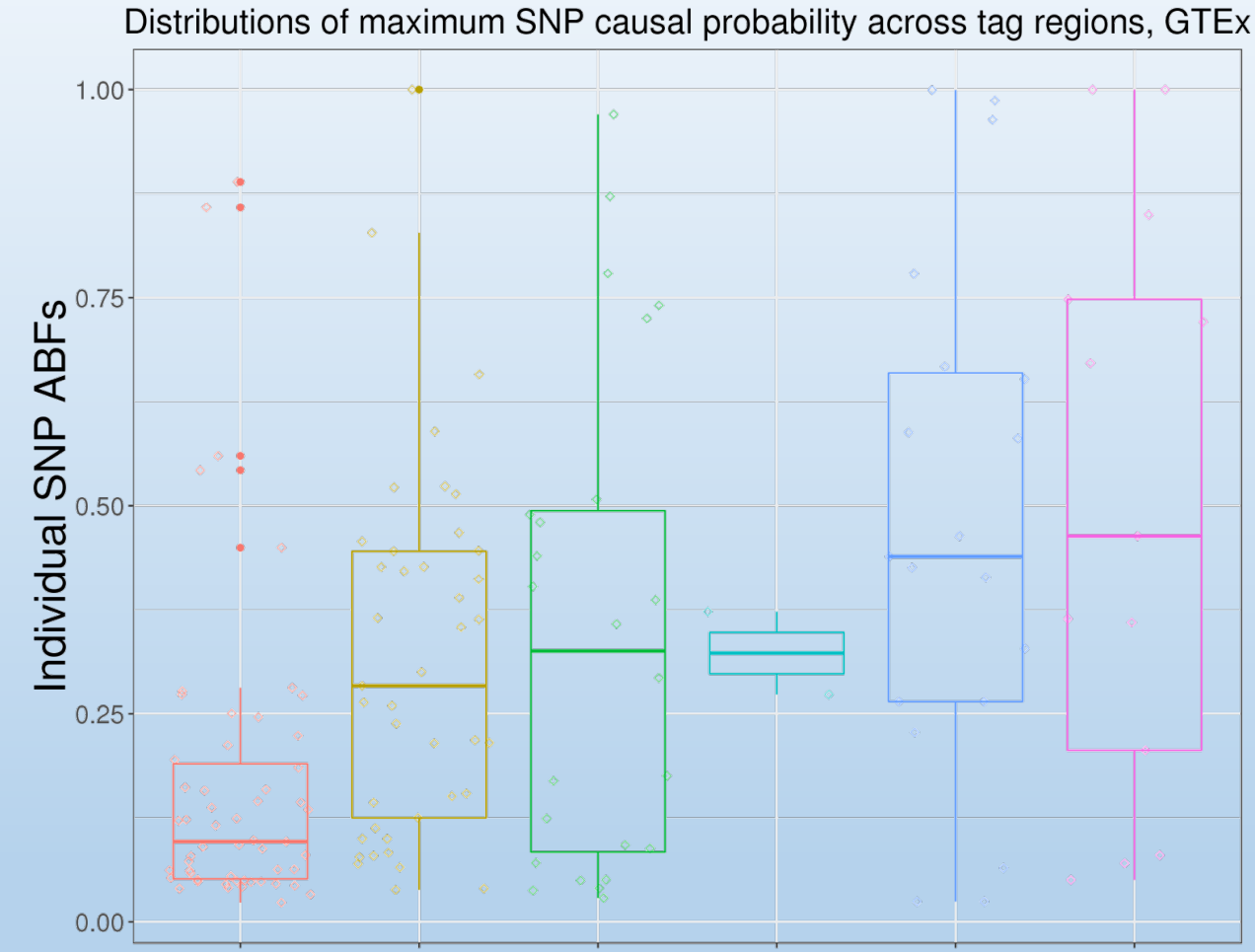


Expanded GTEx sets by ABF



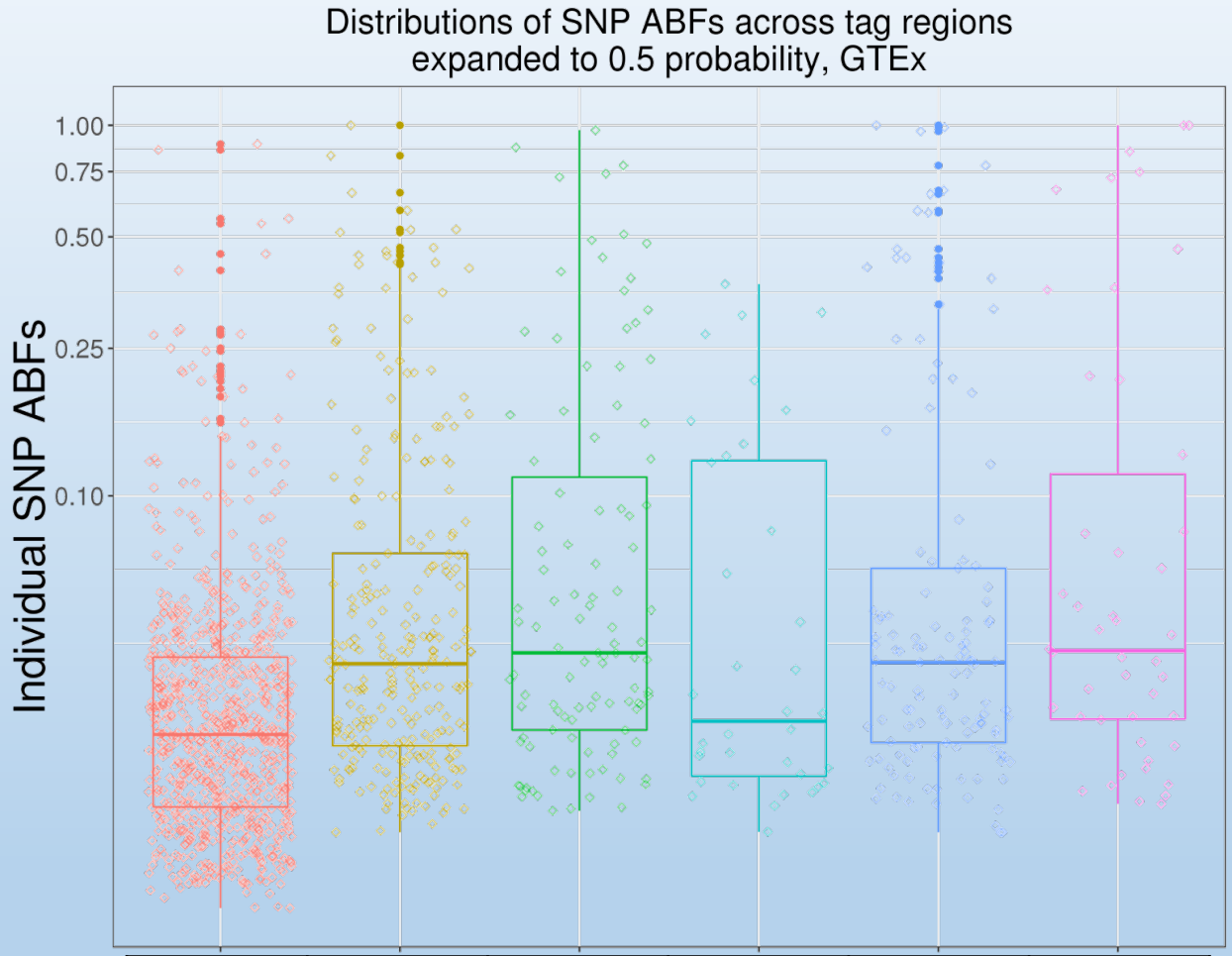


SNP Probability by enhancer+motif overlap



Motif	X	X	X	✓	✓	✓
Enhancer	X	✓	✓	X	✓	✓
Tissue match	X	X	✓	X	X	✓

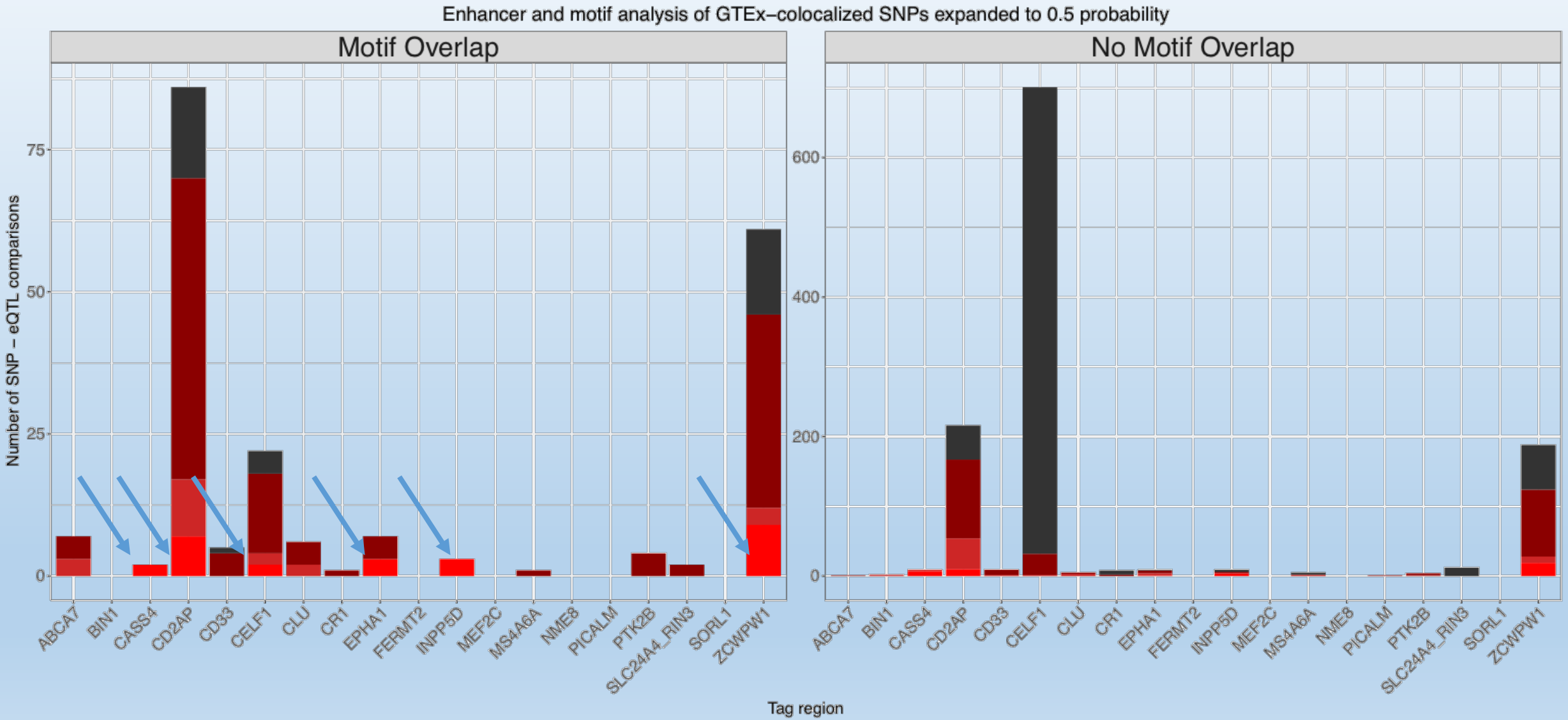
Type of motif and enhancer overlaps



X	X	X	✓	✓	✓
X	✓	✓	X	✓	✓
X	X	✓	X	X	✓

Type of motif and enhancer overlaps

26 eQTLs overlap motif and enhancer in blood, brain, connective tissue, or nervous categories

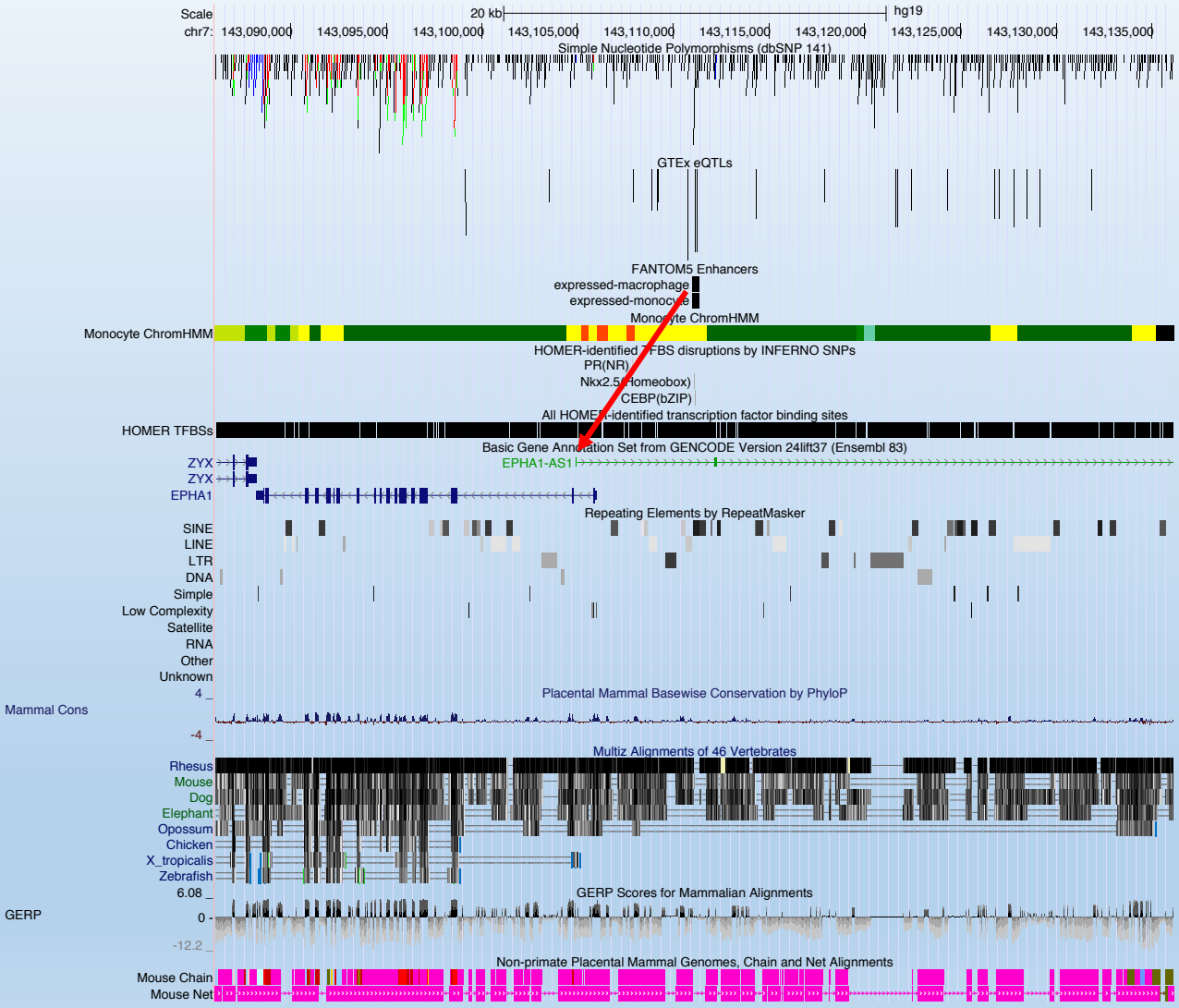


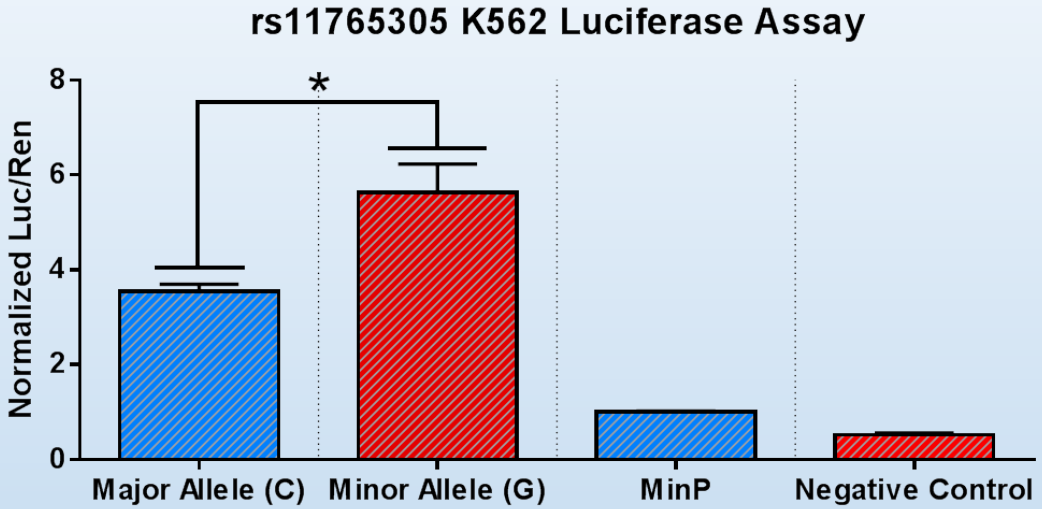
Enhancer and motif signals

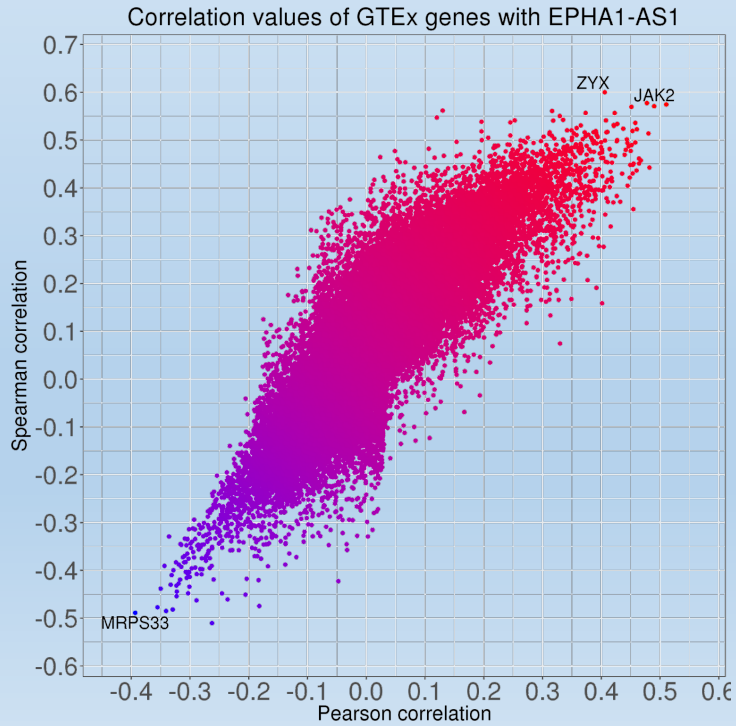
tag_region	GTEx tissue	GTEx tissue class	eQTL target	PP.H4.abf	# expanded SNPs	rsID	SNP ABF	FANTOM5	# Roadmap Classes
CASS4	Whole_Blood	Blood	CASS4	0.918	2	rs6014722	0.464	None	7
CASS4	Whole_Blood	Blood	RPL39P	0.874	3	rs6014722	0.129	None	7
CD2AP	Brain_Cerebellar_Hemisphere	Brain	RP11-385F7.1	0.916	19	rs9367279	0.025	None	13
CD2AP	Cells_EBV-transformed_lymphocytes	Blood	RP11-385F7.1	0.594	24	rs9367279	0.029	None	13
CD2AP	Cells_EBV-transformed_lymphocytes	Blood	RP11-385F7.1	0.594	24	rs4715019	0.025	None	1
CD2AP	Cells_EBV-transformed_lymphocytes	Blood	RP11-385F7.1	0.594	24	rs2151974	0.017	None	1
CD2AP	Cells_EBV-transformed_lymphocytes	Blood	RP11-385F7.1	0.594	24	rs9473126	0.016	None	1
CD2AP	Cells_EBV-transformed_lymphocytes	Blood	RP11-385F7.1	0.594	24	rs9473123	0.016	None	2
CD2AP	Cells_Transformed_fibroblasts	Connective Tissue	CD2AP	0.822	7	rs9367279	0.039	None	13
CELF1	Brain_Cerebellar_Hemisphere	Brain	RP11-750H9.5	0.851	6	rs7947450	0.064	None	2
CELF1	Cells_Transformed_fibroblasts	Connective Tissue	SLC39A13	0.589	10	rs2293579	0.054	None	3
EPHA1	Whole_Blood	Blood	EPHA1-AS1	0.533	1	rs11765305	1.000	Blood	3
EPHA1	Whole_Blood	Blood	TAS2R60	0.533	1	rs11765305	1.000	Blood	3
EPHA1	Whole_Blood	Blood	TAS2R62P	0.549	1	rs11765305	0.850	Blood	3
INPP5D	Whole_Blood	Blood	INPP5D	0.516	11	rs7558417	0.080	None	9
INPP5D	Whole_Blood	Blood	INPP5D	0.516	11	rs7570320	0.079	None	3
INPP5D	Whole_Blood	Blood	INPP5D	0.516	11	rs7568027	0.025	None	1
ZCWPW1	Brain_Caudate_basal_ganglia	Brain	PVRIG	0.558	14	rs1727138	0.044	None	13
ZCWPW1	Brain_Caudate_basal_ganglia	Brain	STAG3	0.523	20	rs1727138	0.018	None	13
ZCWPW1	Brain_Cerebellar_Hemisphere	Brain	PVRIG	0.548	17	rs1727138	0.030	None	13
ZCWPW1	Brain_Cerebellum	Brain	GAL3ST4	0.529	20	rs1727138	0.018	None	13
ZCWPW1	Brain_Hypothalamus	Brain	PVRIG	0.584	15	rs1727138	0.015	None	13
ZCWPW1	Brain_Nucleus_accumbens_basal_ganglia	Brain	PVRIG	0.560	13	rs1727138	0.046	None	13
ZCWPW1	Brain_Nucleus_accumbens_basal_ganglia	Brain	STAG3	0.573	10	rs1727138	0.047	None	13
ZCWPW1	Brain_Putamen_basal_ganglia	Brain	PVRIG	0.551	13	rs1727138	0.050	None	13
ZCWPW1	Nerve_Tibial	Nervous	STAG3	0.562	10	rs6948729	0.042	None	4

Region	Tissue	Tissue class	Target gene	P(H_4)	SNP	SNP ABF	FANTOM5	# Roadmap Classes	Transcription Factors	ΔPWM Scores
CASS4	Whole Blood	Blood	CASS4	0.918	rs6014722	0.464	None	7	HOXD13	-0.1
CD2AP	Cells EBV-transformed lymphocytes	Blood	**RP11-385F7.1**	0.594	rs9367279	0.029	None	13	CArG; CArG	-1.38 -1.38
CD2AP	Brain Cerebellar Hemisphere	Brain	**RP11-385F7.1**	0.916	rs9367279	0.025	None	13	CArG; CArG	-1.38 -1.38
CELF1	Brain Cerebellar Hemisphere	Brain	**RP11-750H9.5**	0.851	rs7947450	0.064	None	2	Arnt:Ahr(bHLH)	-2.22
EPHA1	Whole Blood	Blood	**EPHA1-AS1**	0.533	rs11765305	1.000	Blood	3	Nkx2.5; CEBP(bZIP)	-0.8; 1.54
INPP5D	Whole Blood	Blood	INPP5D	0.516	rs7568027	0.025	None	1	Hoxc9; Lhx3; Lhx3	-4.93 -2.39 -2.4
ZCWPW1	Brain Cerebellum	Brain	GAL3ST4	0.529	rs1727138	0.018	None	13	CArG; TATA-Box(TBP); Mef2c; Mef2a; CArG; Cdx2	-4.06 -6.87 -4.85 -6.14 -3.88; -6.9
ZCWPW1	Brain Caudate basal ganglia	Brain	PVRIG	0.558	rs1727138	0.044	None	13	"	"
ZCWPW1	Brain Cerebellar Hemisphere	Brain	PVRIG	0.548	rs1727138	0.030	None	13	"	"
ZCWPW1	Brain Hypothalamus	Brain	PVRIG	0.584	rs1727138	0.015	None	13	"	"
ZCWPW1	Brain Nucleus accumbens basal ganglia	Brain	PVRIG	0.560	rs1727138	0.046	None	13	"	"
ZCWPW1	Brain Putamen basal ganglia	Brain	PVRIG	0.551	rs1727138	0.050	None	13	"	"
ZCWPW1	Brain Caudate basal ganglia	Brain	STAG3	0.523	rs1727138	0.018	None	13	"	"
ZCWPW1	Brain Nucleus accumbens basal ganglia	Brain	STAG3	0.573	rs1727138	0.047	None	13	"	"

EPHA1 region



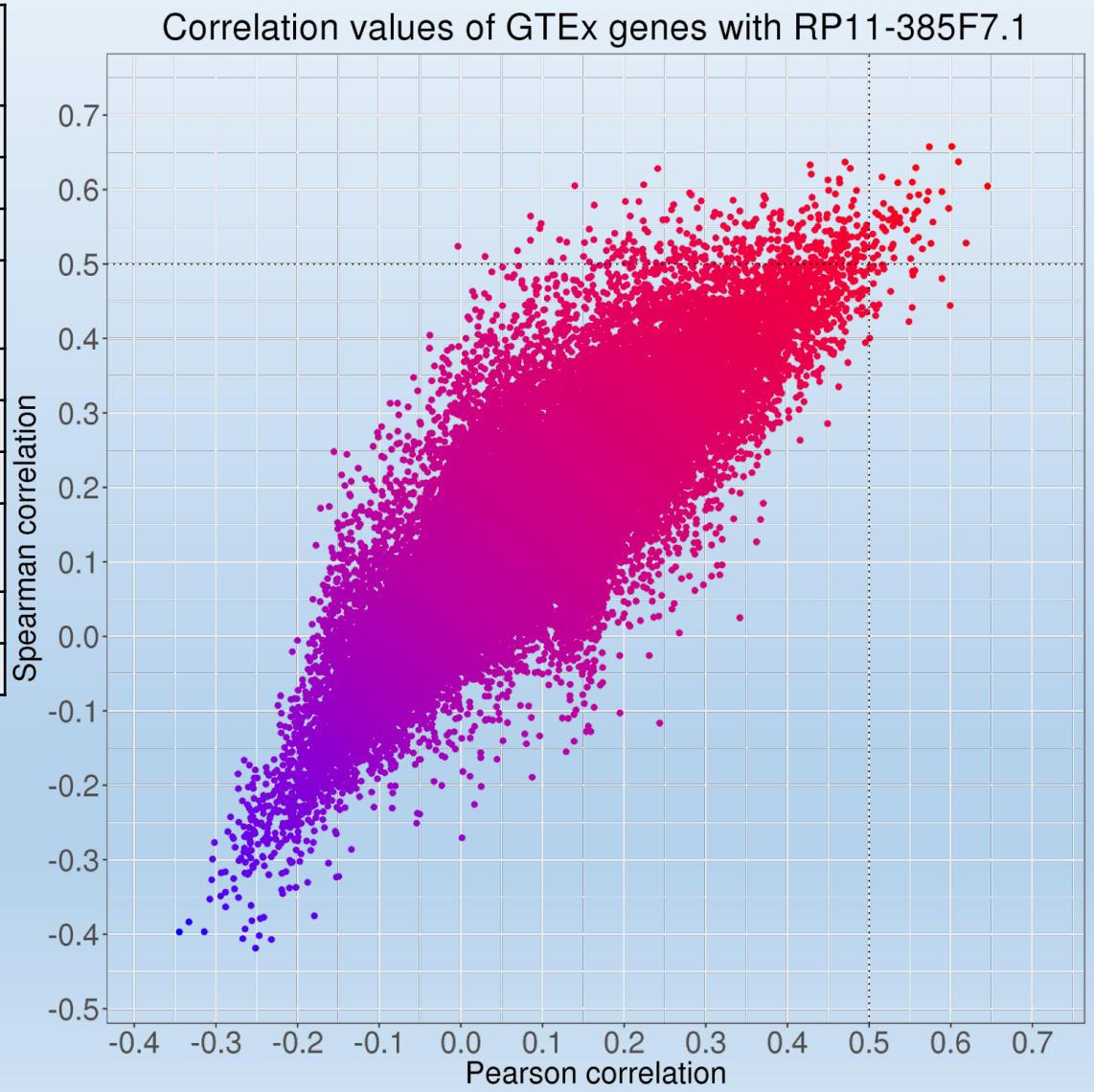




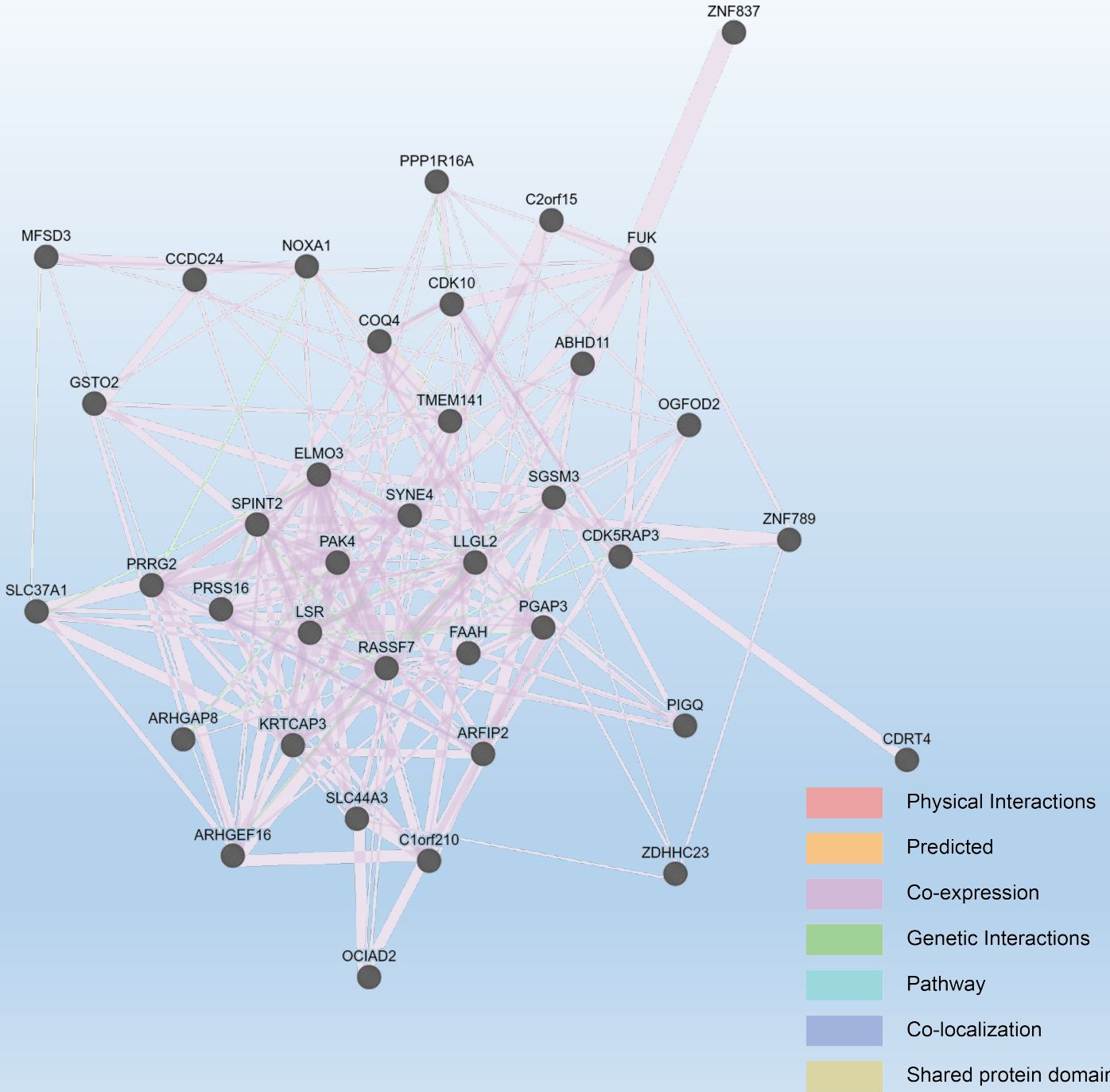
CD2AP region

chr	gene	Pearson correlation	Spearman correlation	Average correlation
chr9	COQ4	0.601	0.658	0.630
chr8	PPP1R16A	0.645	0.605	0.625
chr9	NOXA1	0.609	0.637	0.623
chr1	FAAH	0.574	0.657	0.616
chr10	RP11-18I14.10	0.557	0.629	0.593
chr19	SPPL2B	0.589	0.597	0.593
chr7	ABHD11	0.597	0.575	0.586
chr1	CCDC24	0.573	0.597	0.585
chr2	C2ORF15	0.553	0.610	0.581
chr16	FUK	0.571	0.585	0.578

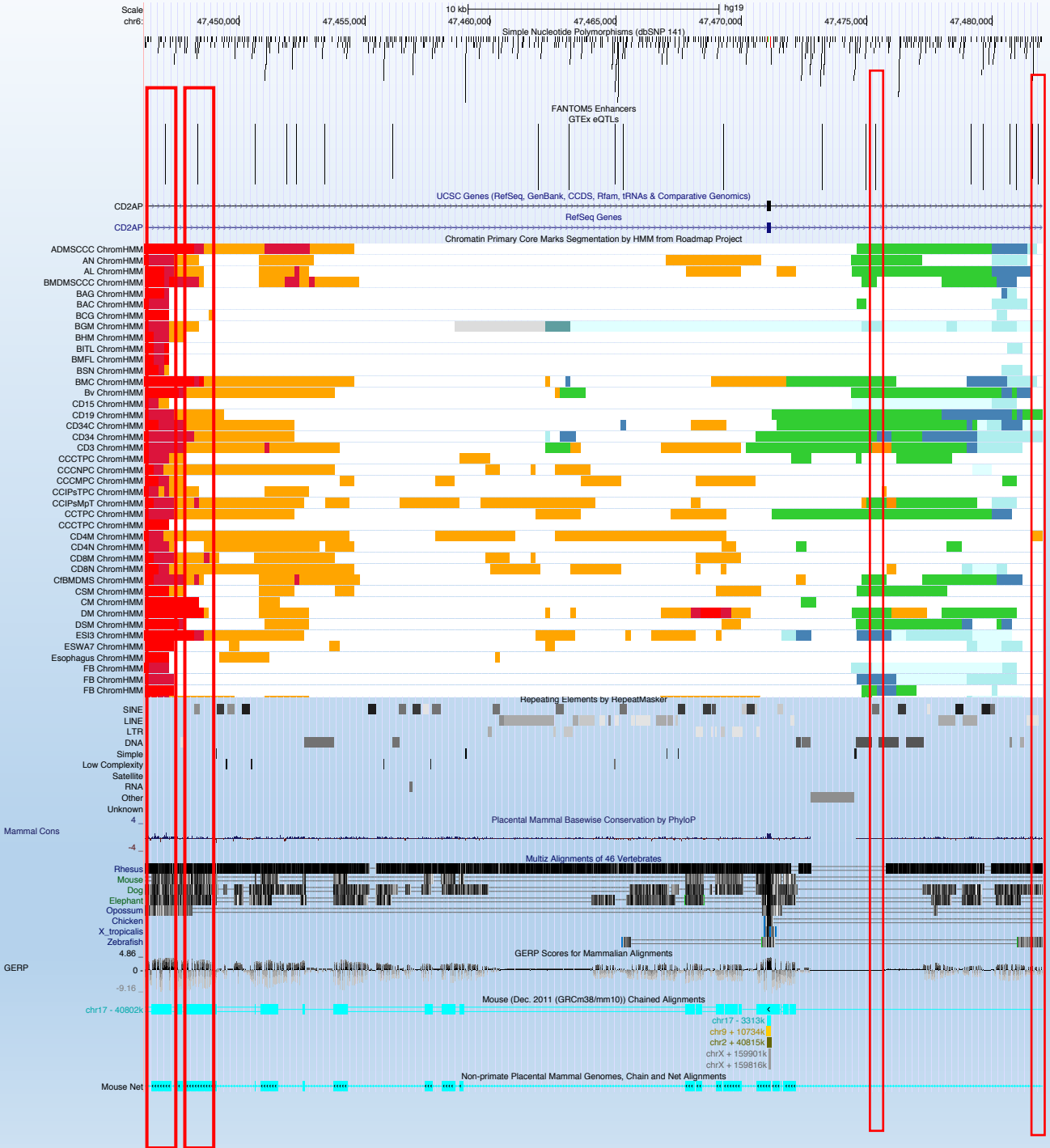
- 53 genes have >= 0.5 Pearson and Spearman correlation
- Enrichment of GTPase binding pathways

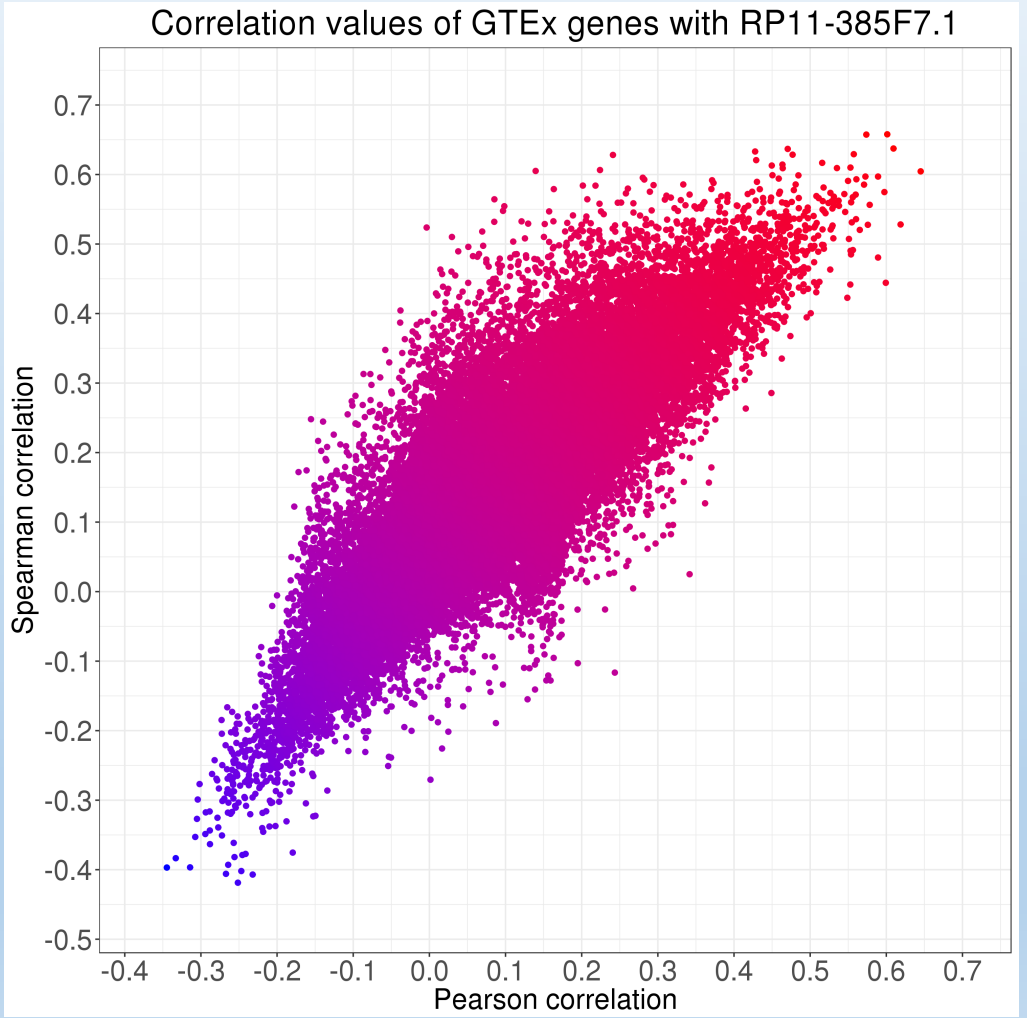


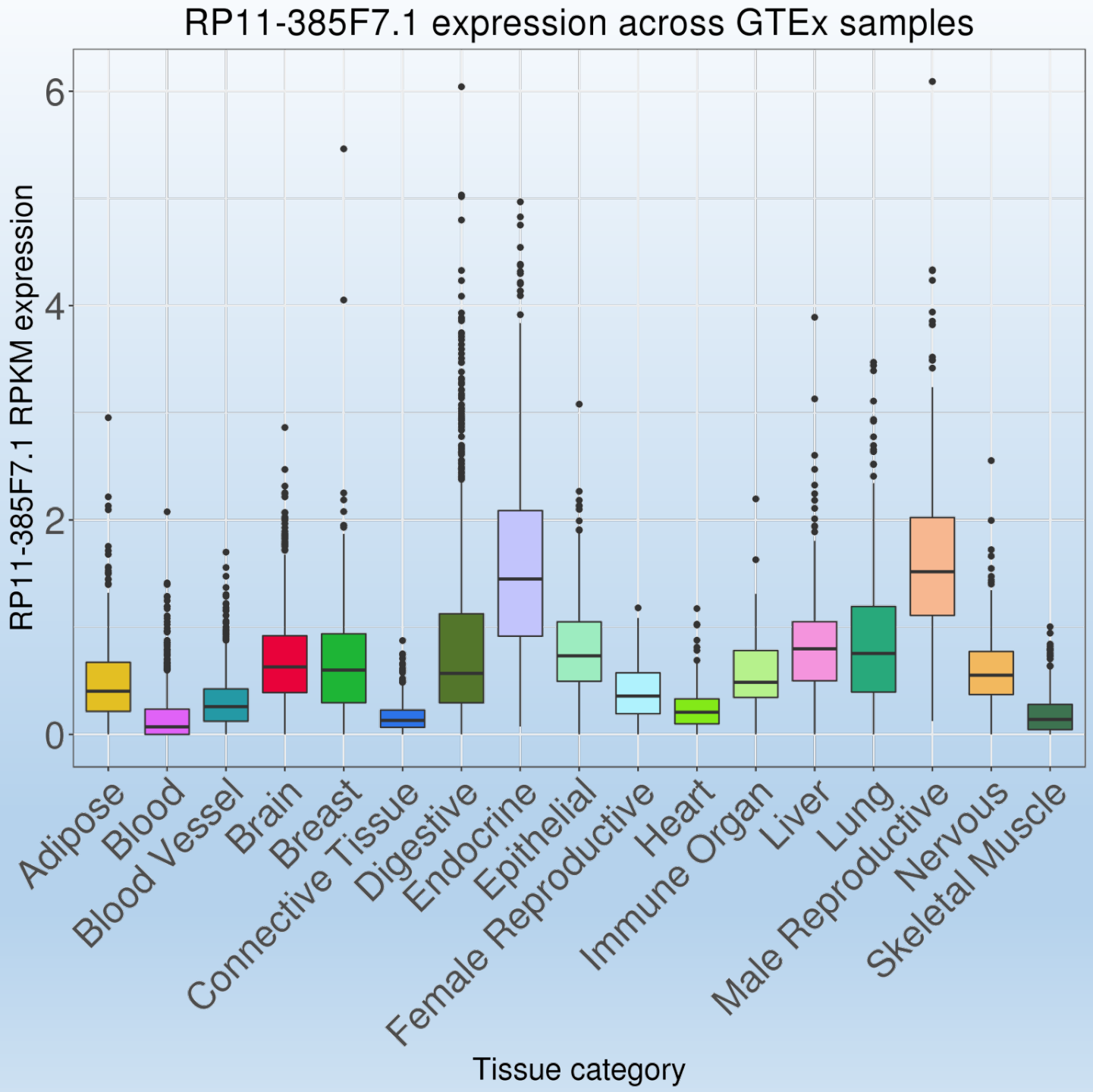
Function	FDR	Genes in network	Genes in genome
small GTPase binding	0.213	4	90
GTPase binding	0.213	4	97
Rho GTPase binding	0.213	3	35
Ras GTPase binding	0.213	4	79



CD2AP region



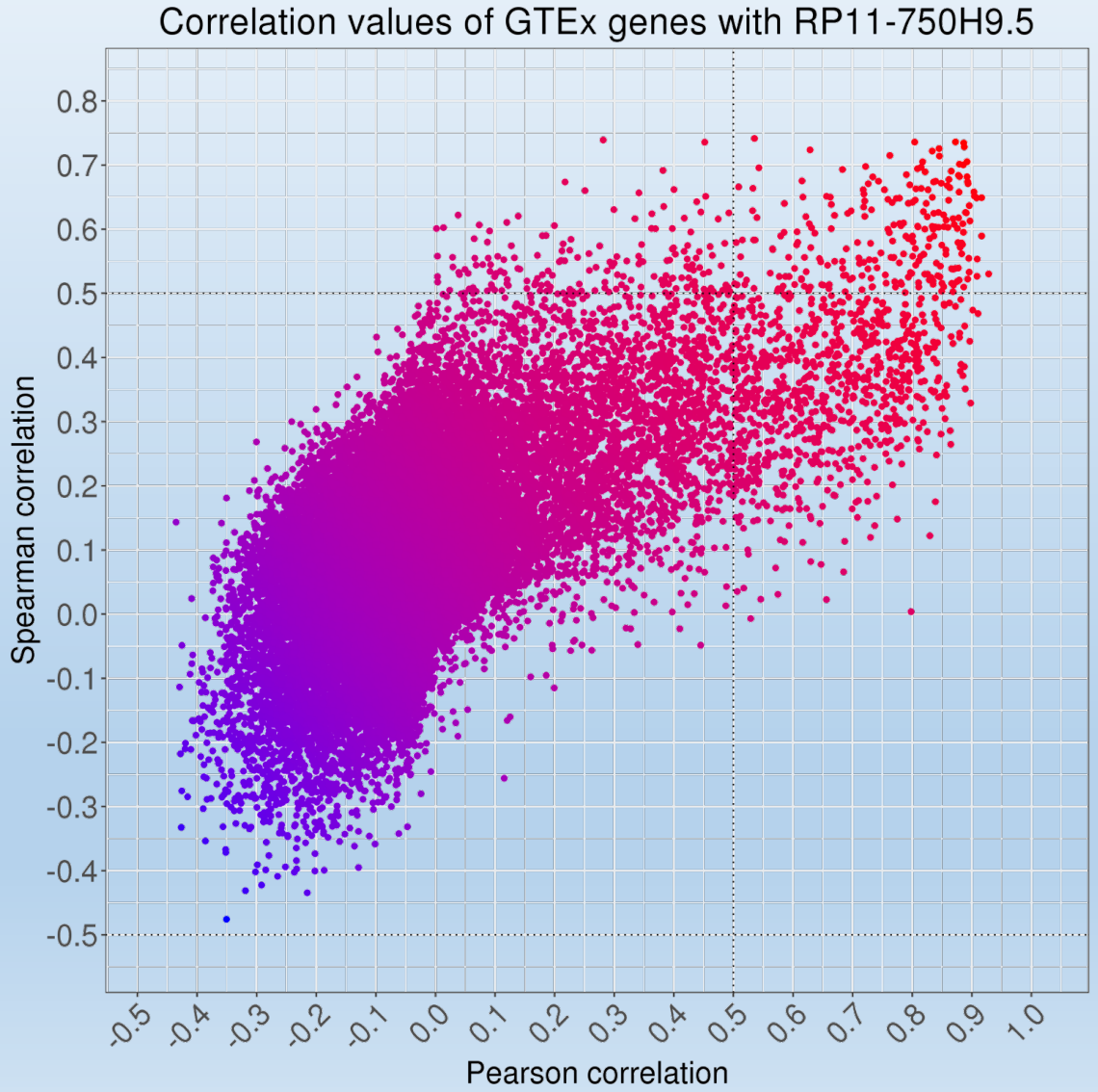



RP11-385F7.1 expression across GTEx samples
RP11-385F7.1 RPKM expression
0
2
4
6
Adipose
Blood
Blood Vessel
Brain
Breast
Connective Tissue
Digestive
Endocrine
Epithelial
Female Reproductive
Heart
Immune Organ
Liver
Lung
Male Reproductive
Nervous
Skeletal Muscle
Tissue category

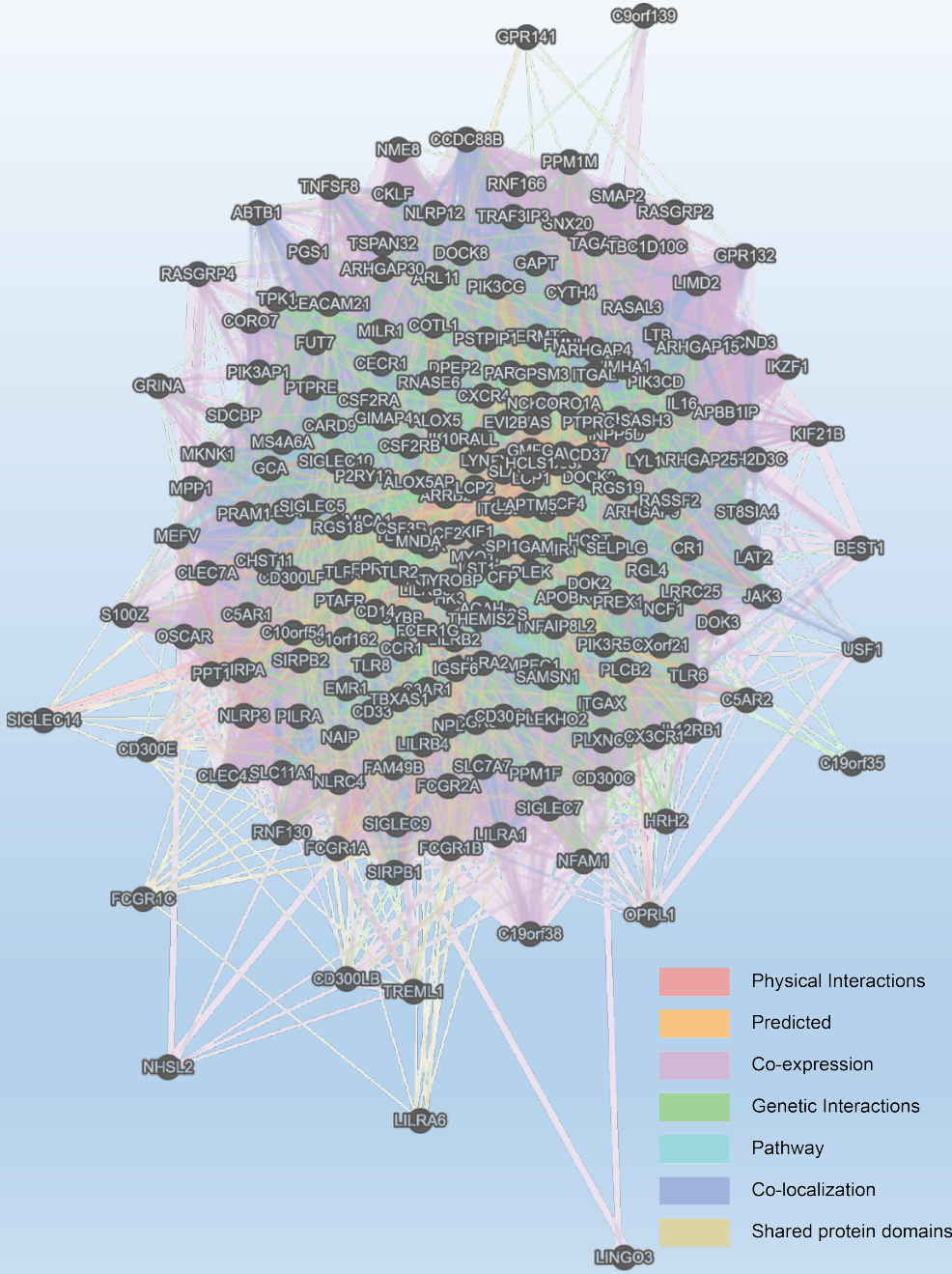
CELF1 region

chr	gene	Pearson correlation	Spearman correlation	Mean correlation
chr17	ARRB2	0.886	0.734	0.810
chrX	CSF2RA	0.887	0.728	0.808
chr16	ITGAM	0.873	0.736	0.804
chr16	IGSF6	0.892	0.705	0.799
chr19	LILRA1	0.886	0.700	0.793
chr17	CD300LF	0.879	0.701	0.790
chr20	HCK	0.892	0.682	0.787
chr3	P2RY13	0.845	0.726	0.785
chr20	SIRPB1	0.893	0.674	0.783
chr12	BIN2	0.916	0.649	0.783

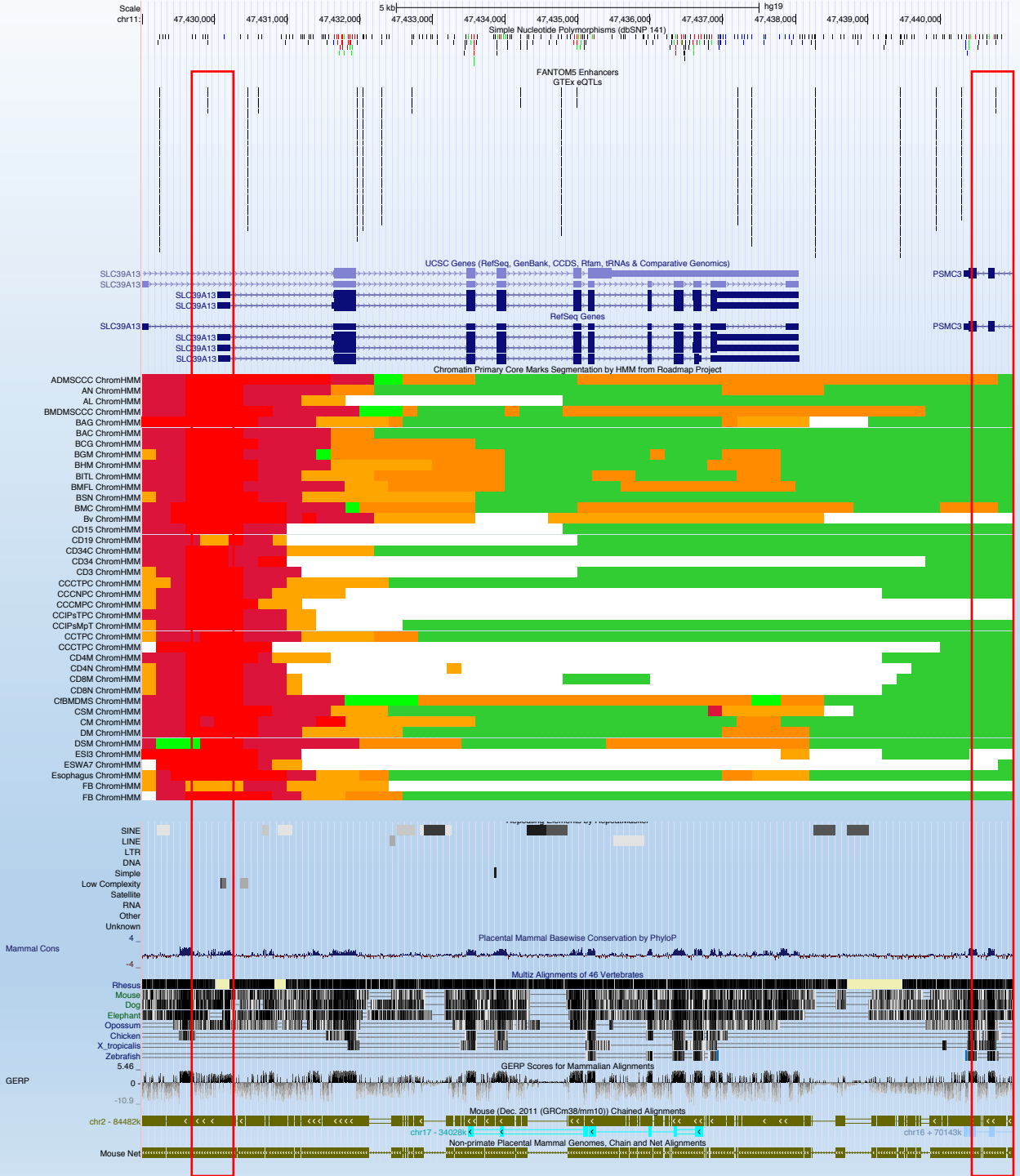
- 252 genes have >= 0.5 Spearman and Pearson correlation
- Dozens of enriched pathways including immune- and leukocyte-related and signaling pathways

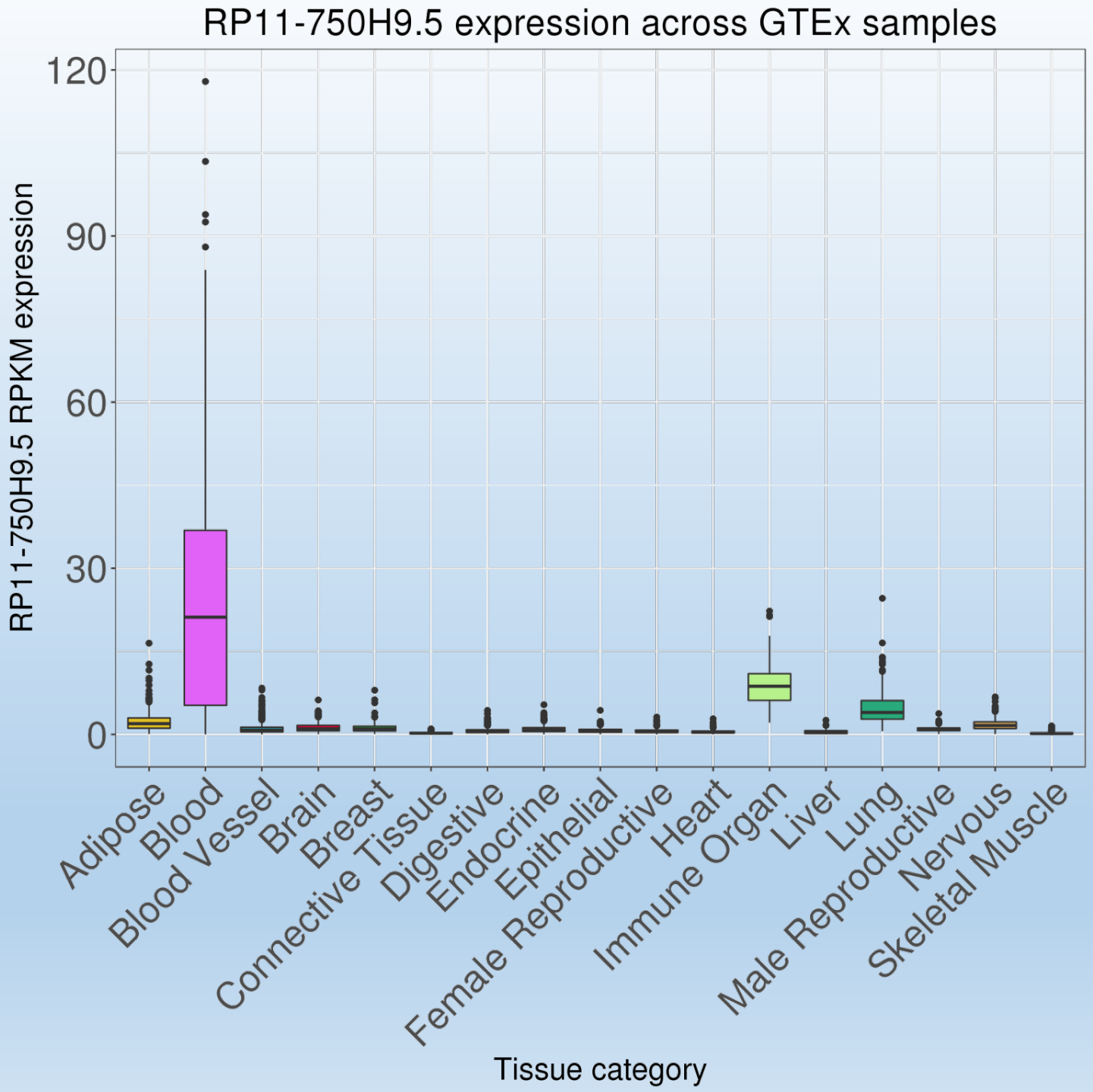


Function	FDR	Genes in network	Genes in genome
leukocyte migration	2.18E-14	26	230
immune response-activating cell surface receptor signaling pathway	8.84E-12	21	176
phagocytosis	4.42E-11	18	129
immune response-regulating cell surface receptor signaling pathway	2.55E-10	24	297
inflammatory response	3.98E-08	21	282
myeloid leukocyte activation	7.62E-08	12	71
myeloid leukocyte migration	3.79E-07	12	82
leukocyte chemotaxis	5.42E-07	13	106
cell chemotaxis	7.00E-07	15	157
B cell receptor signaling pathway	1.18E-06	8	28
leukocyte activation involved in immune response	1.45E-06	12	96
cell activation involved in immune response	1.45E-06	12	96
leukocyte mediated immunity	2.39E-06	13	124
mast cell degranulation	2.39E-06	6	12
leukocyte degranulation	2.86E-06	7	21



- CELF1 region
- First red box: rs7947450 is brain eQTL for RP11-750H9.5, antisense lncRNA
- Second red box: rs2293579 is fibroblast eQTL for SLC39A13, zinc transporter transmembrane protein, mutations associated with connective tissue syndrome (“elastic skin”)




RP11-750H9.5 expression across GTEx samples
RP11-750H9.5 RPKM expression
120
90
60
30
0
Adipose
Blood
Blood Vessel
Brain
Breast
Connective Tissue
Digestive
Endocrine
Epithelial
Female Reproductive
Heart
Immune Organ
Liver
Lung
Male Reproductive
Nervous
Skeletal Muscle
Tissue category

High-level summary of results

Tag Region	Affected mechanism and evidence
CASS4	HOXD13-mediated enhancer with eQTL for **CASS4**
CD2AP	Strong homeobox TF disruption in enhancer for **RP11-385F7.1** in blood and brain, affecting oxygen species
CELF1	Brain signal for **RP11-750H9.5**, moderate TF disruption, affecting immune regulatory hub
EPHA1	Very strong ABF for rs11765305 affecting **EPHA1-AS1** (→**JAK2**) and two taste receptor signals in blood (monocytes) with strengthened CEBP motif
INPP5D	Blood signal for **INPP5D**, strong disruption of Homeobox TFs and moderate on other TFs
ZCWPW1	One SNP strongly disrupts several motifs and colocalizes with GTEx brain eQTLs for **GAL3ST4**, **PVRIG**, and **STAG3**

Conclusions and next steps

- Large-scale annotation across diverse data types and tissues yields insights into regulatory mechanisms affected by noncoding AD variants
- Several lncRNA signals suggest that identifying enhancer targets is only the first step to understanding regulatory dysfunction
- Need to validate enhancer activity and effect on target gene expression in relevant cell types
- Eventually want to identify conserved enhancers for mouse CRISPR studies

Acknowledgements

- **Li-San Wang**
- **Jerry Schellenberg**
- **Mitchell Tang**
- Elizabeth Mlynarski
- Beth Dombrowski
- Jessica King
- Fanny Leung
- Pavel Kuksa
- Barry Slaff
- Jacky Wang
- Adam Naj

- Thesis Committee:
  - Mingyao Li (Chair)
  - Jerry Schellenberg
  - Edward Lee
  - Casey Brown
  - Barbara Engelhardt (Princeton)

- T32 AG000255-18: Training in Age Related Neurodegenerative Diseases (**Virginia Lee**)

Perelman School of Medicine
University of Pennsylvania

Motif signals

tag_region	GTEx tissue class	eQTL target	PP.H4.abf	# expanded SNPs	rsID	SNP ABF	num_motif_hits	tfs	delta_pwms
CASS4	Blood	CASS4	0.918	2	rs6014722	0.464	1	HOXD13	-0.1
CASS4	Blood	RPL39P	0.874	3	rs6014722	0.129	1	HOXD13	-0.1
CD2AP	Brain	RP11-385F7.1	0.916	19	rs9367279	0.025	2	CArG	-1.38, -1.38
CD2AP	Blood	RP11-385F7.1	0.594	24	rs9367279	0.029	2	CArG	-1.38, -1.38
CD2AP	Blood	RP11-385F7.1	0.594	24	rs4715019	0.025	18	Lhx3, Lhx2, Nkx6.1, Nanog, Isl1, Lhx3, Lhx2, Nkx6.1, Eomes, Lhx3, Lhx2, Nkx6.1, Nanog, Isl1, Lhx3, Lhx2, Nkx6.1, Eomes	-6.78, -6.81, -5.82, -6.45, -6.85, -6.73, -6.74, -6.9, -5.08, -6.78, -6.81, -5.82, -6.45, -6.85, -6.73, -6.74, -6.9, -5.08
CD2AP	Blood	RP11-385F7.1	0.594	24	rs2151974	0.017	2	NFAT	0.43, 0.43
CD2AP	Blood	RP11-385F7.1	0.594	24	rs9473126	0.016	2	Sox3	-1.55, -1.55
CD2AP	Blood	RP11-385F7.1	0.594	24	rs9473123	0.016	4	Klf4 (Zf), KLF5 (Zf), Klf4 (Zf), KLF5 (Zf)	1.79, 0.14, 1.79, 0.14
CD2AP	Connective Tissue	CD2AP	0.822	7	rs9367279	0.039	2	CArG , CArG	-1.38, -1.38
CELF1	Brain	RP11-750H9.5	0.851	6	rs7947450	0.064	1	Arnt:Ahr (bHLH)	-2.22
CELF1	Connective Tissue	SLC39A13	0.589	10	rs2293579	0.054	2	AR-halfsite, AR-halfsite	-2.84, -6.9
EPHA1	Blood	EPHA1-AS1	0.533	1	rs11765305	1.000	2	Nkx2.5 , CEBP (bZIP)	-0.8, 1.54
EPHA1	Blood	TAS2R60	0.533	1	rs11765305	1.000	2	Nkx2.5 , CEBP (bZIP)	-0.8, 1.54
EPHA1	Blood	TAS2R62P	0.549	1	rs11765305	0.850	2	Nkx2.5 , CEBP (bZIP)	-0.8, 1.54

Motif signals cont'd

tag_region	GTEx tissue class	eQTL target	PP.H4.abf	# expanded SNPs	rsID	SNP ABF	num_motif_hits	tfs	delta_pwms
INPP5D	Blood	INPP5D	0.516	11	rs7558417	0.080	1	RXR,DR1	-3.27
INPP5D	Blood	INPP5D	0.516	11	rs7570320	0.079	1	Fox:Ebox	-0.73
INPP5D	Blood	INPP5D	0.516	11	rs7568027	0.025	3	Hoxc9, Lhx3, Lhx3	-4.93, -2.39, -2.4
ZCWPW1	Brain	PVRIG	0.558	14	rs1727138	0.044	6	CArG, TATA-Box, Mef2c, Mef2a, CArG, Cdx2	-4.06, -6.87, -4.85, -6.14, -3.88, -6.9
ZCWPW1	Brain	STAG3	0.523	20	rs1727138	0.018	6	CArG, TATA-Box, Mef2c, Mef2a, CArG, Cdx2	-4.06, -6.87, -4.85, -6.14, -3.88, -6.9
ZCWPW1	Brain	PVRIG	0.548	17	rs1727138	0.030	6	CArG, TATA-Box, Mef2c, Mef2a, CArG, Cdx2	-4.06, -6.87, -4.85, -6.14, -3.88, -6.9
ZCWPW1	Brain	GAL3ST4	0.529	20	rs1727138	0.018	6	CArG , TATA-Box, Mef2c, Mef2a, CArG, Cdx2	-4.06, -6.87, -4.85, -6.14, -3.88, -6.9
ZCWPW1	Brain	PVRIG	0.584	15	rs1727138	0.015	6	CArG, TATA-Box, Mef2c, Mef2a, CArG, Cdx2	-4.06, -6.87, -4.85, -6.14, -3.88, -6.9
ZCWPW1	Brain	PVRIG	0.560	13	rs1727138	0.046	6	CArG ,TATA-Box , Mef2c , Mef2a , CArG , Cdx2	-4.06, -6.87, -4.85, -6.14, -3.88, -6.9
ZCWPW1	Brain	STAG3	0.573	10	rs1727138	0.047	6	CArG , TATA-Box , Mef2c , Mef2a , CArG , Cdx2	-4.06, -6.87, -4.85, -6.14, -3.88, -6.9
ZCWPW1	Brain	PVRIG	0.551	13	rs1727138	0.050	6	CArG , TATA-Box , Mef2c , Mef2a , CArG , Cdx2	-4.06, -6.87, -4.85, -6.14, -3.88, -6.9
ZCWPW1	Nervous	STAG3	0.562	10	rs6948729	0.042	5	Nanog , Lhx3 , Nkx6.1 , GRHL2, GRHL2	-3.02, -1.55, -6.9, -1.55, -1.74