

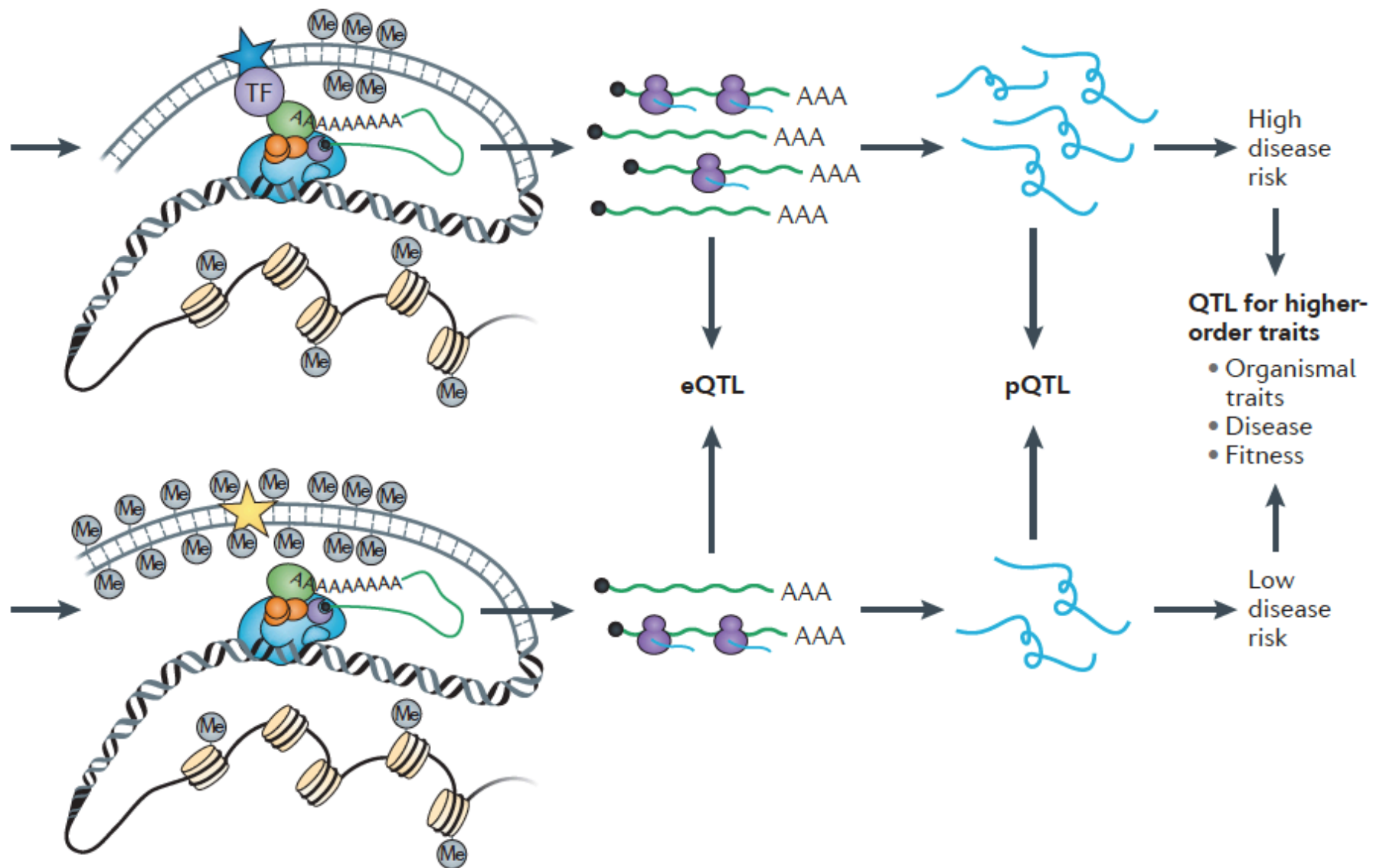
CBB 752
Final Project 1.2
Yale University

Compare the variants in Carl's genome with those found in the GTEx database. Use the results to predict gene expression in various tissues and better estimate the impact of noncoding variants in Carl's genome.

Writing

Describe eQTLs and the GTEx database.

How was the data present in GTEx generated and what scientific questions does it help answer?



C Local regulatory variation

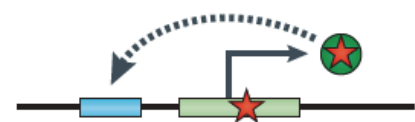
Cis-regulatory
(allele-specific)



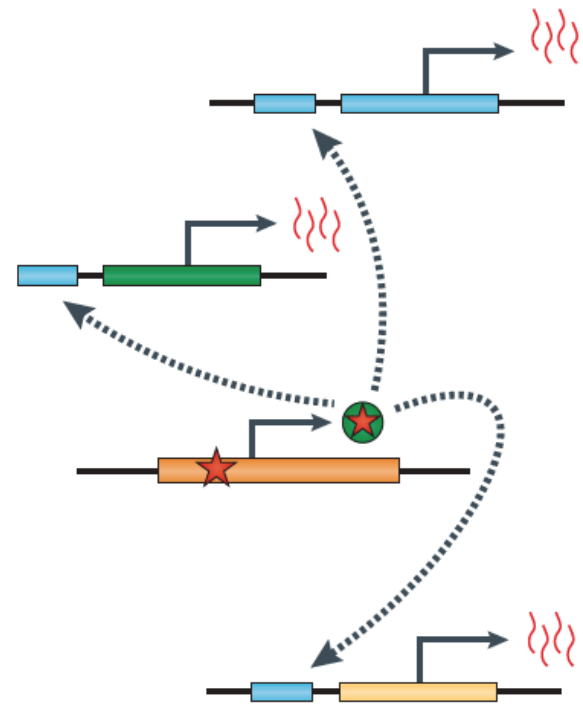
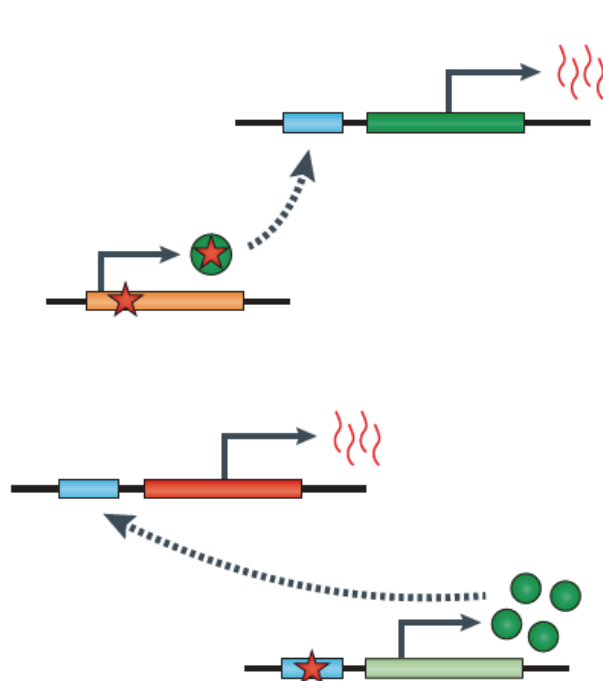
Neighbouring gene (*trans*)

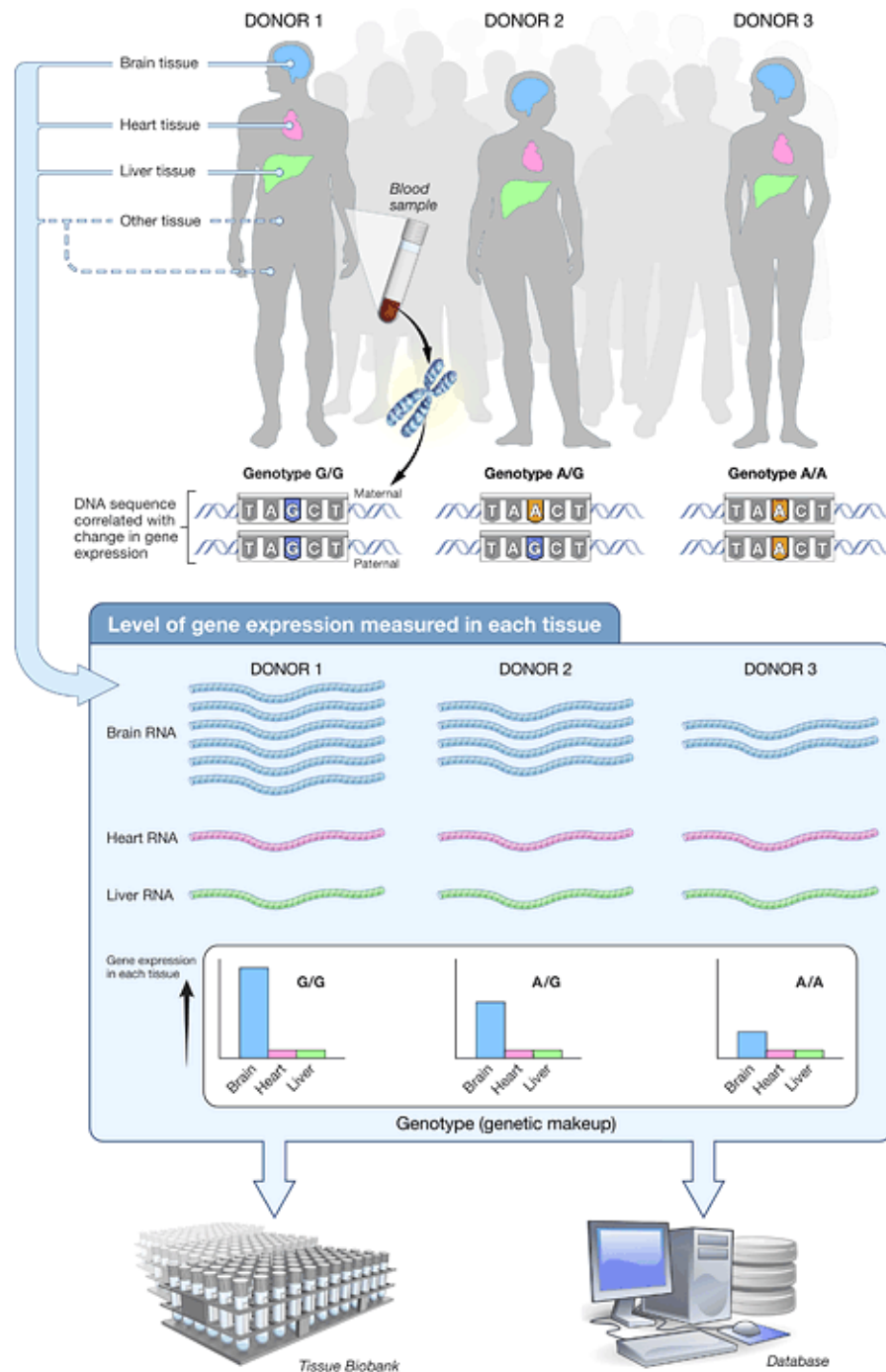


Autoregulation (*trans*)



Distant regulatory variation





Identifying eQTLs in the Zimmerome

- Combine genetic information (SNPs) and known eQTL sites (GTEx)
- Determine whether Carl possesses any of the variants of interest, which could influence protein amounts
- Generate predictions – Not confirmed without individual RNA-seq data

Brief Introduction: Coding

- Read in Zimmerome SNPs
- Find the overlapping SNP-gene-tissue pairs between Zimmerome SNPs and GTEx eQTL database
- Print out the overlapping SNP-gene-tissue pairs, intersection and statistics
- *Final1-2.2.py, final1-2.2.3.py, final1-2.x.py*

Result example

chr	pos	ref	alt	gene_id	tss_distance	pval_nominal	slope	slope_se	slope_fpk	slope_fpk_se	pval_nominal_threshold	min_pval_nominal	pval_beta	tissue	
1	739210	A	G	ENSG00000237491.4	25060	9.96E-08	-0.612927	0.112207	-0.0328812	0.00614083	7.76E-05	4.83E-13	4.89E-09	Whole Blood	
1	739528	G	A	ENSG00000237491.4	25378	1.04E-11	-0.831361	0.117406	-0.0464057	0.00638987	7.76E-05	4.83E-13	4.89E-09	Whole Blood	
1	752894	T	C	ENSG00000237491.4	38744	1.83E-05	-0.354281	0.0813353	-0.0186668	0.00445322	7.76E-05	4.83E-13	4.89E-09	Whole Blood	
1	739528	G	A	ENSG00000225880.4	-23374	7.99E-07	-0.504767	0.100086	-0.2397	0.0440575	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	752566	G	A	ENSG00000225880.4	-10336	9.46E-10	-0.416138	0.0658176	-0.1583	0.0296727	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	752721	A	G	ENSG00000225880.4	-10181	8.74E-10	-0.410651	0.064808	-0.157076	0.0292041	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	752894	T	C	ENSG00000225880.4	-10008	8.50E-08	-0.360696	0.0656582	-0.136818	0.0294824	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	753474	C	G	ENSG00000225880.4	-9428	4.62E-09	-0.389212	0.064442	-0.148809	0.028996	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	754503	G	A	ENSG00000225880.4	-8399	3.28E-09	-0.398022	0.0652317	-0.150661	0.0293869	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	754964	C	T	ENSG00000225880.4	-7938	4.44E-08	-0.383144	0.0681893	-0.140034	0.0307209	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	755775	A	G	ENSG00000225880.4	-7127	2.88E-09	-0.392338	0.0640578	-0.1468	0.0288912	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	756604	A	G	ENSG00000225880.4	-6298	2.29E-05	-0.308974	0.0718018	-0.10253	0.0322449	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	758144	A	G	ENSG00000225880.4	-4758	1.09E-07	-0.37022	0.0679944	-0.134632	0.0306125	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	759700	T	C	ENSG00000225880.4	-3202	1.11E-08	-0.385592	0.0655761	-0.1492	0.0294526	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	760912	C	T	ENSG00000225880.4	-1990	1.13E-08	-0.396459	0.0674781	-0.149974	0.0303642	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	761147	T	C	ENSG00000225880.4	-1755	4.52E-09	-0.40802	0.0675114	-0.153101	0.0304275	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	787685	G	T	ENSG00000225880.4	24783	1.26E-06	-0.328816	0.0664608	-0.119808	0.0298362	7.68E-05	4.35E-11	2.42E-07	Whole Blood	
1	886788	G	A	ENSG00000187961.9	-9179	1.19E-05	0.187878	0.0421644	0.197147	0.0491114	7.19E-05	3.47E-06	0.00829706	Whole Blood	
1	886817	C	T	ENSG00000187961.9	-9150	8.99E-06	0.19027	0.0421037	0.197754	0.0490751	7.19E-05	3.47E-06	0.00829706	Whole Blood	
1	891021	G	A	ENSG00000187961.9	-4946	6.35E-05	0.200311	0.0493664	0.211351	0.0574275	7.19E-05	3.47E-06	0.00829706	Whole Blood	
1	894573	G	A	ENSG00000187961.9	-1394	3.47E-06	0.20854	0.0440844	0.208635	0.0515295	7.19E-05	3.47E-06	0.00829706	Whole Blood	
1	895706	G	A	ENSG00000187961.9	-261	1.57E-05	0.194408	0.0442652	0.201694	0.0515803	7.19E-05	3.47E-06	0.00829706	Whole Blood	
1	901559	G	A	ENSG00000187961.9	5592	7.27E-06	0.203817	0.0446296	0.198985	0.0521976	7.19E-05	3.47E-06	0.00829706	Whole Blood	
1	902997	G	A	ENSG00000187961.9	7030	1.02E-05	0.189483	0.0422019	0.176451	0.0494474	7.19E-05	3.47E-06	0.00829706	Whole Blood	

Summary statistics

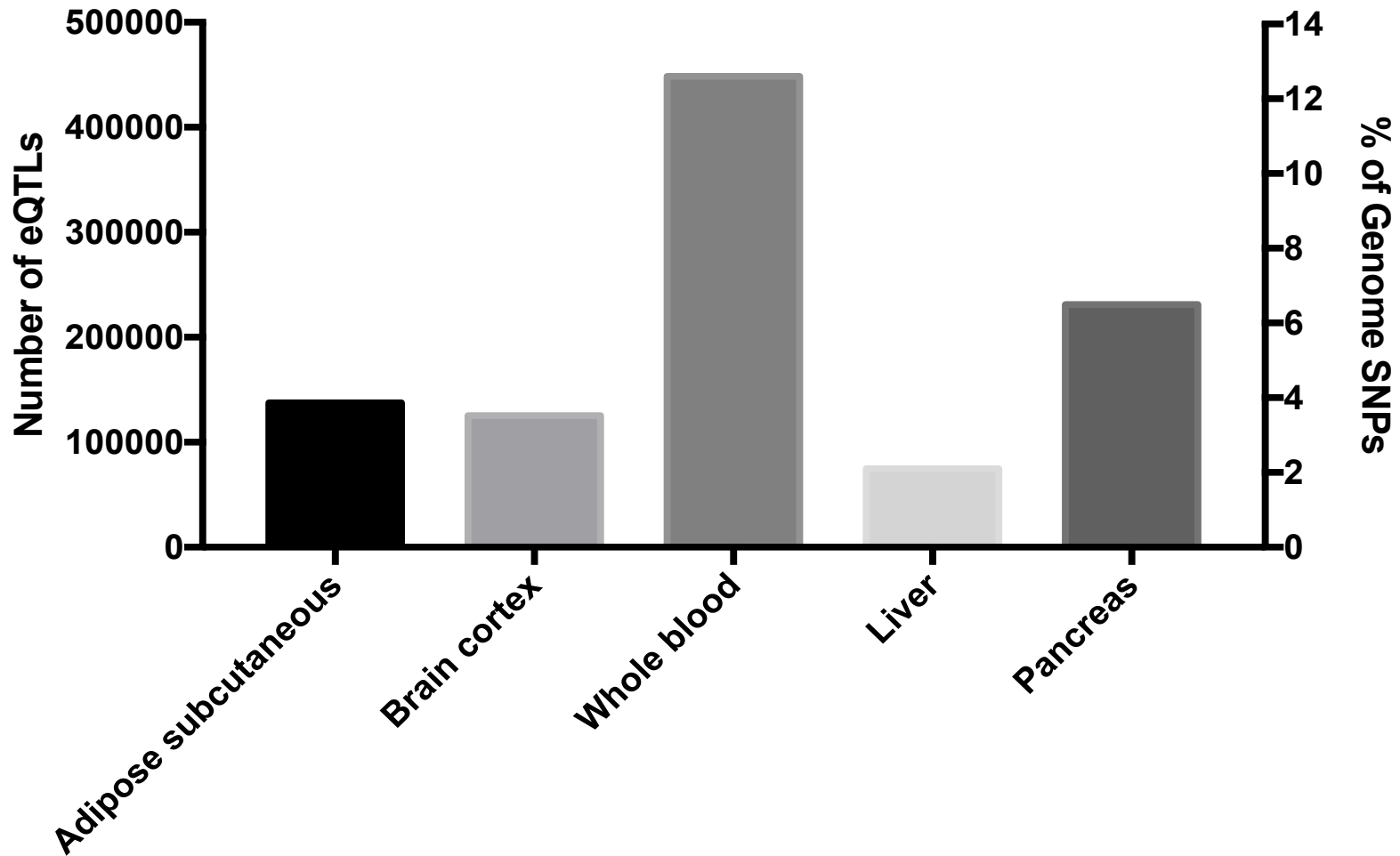
Tissue	#Z eQTLs	#total eQTLs	%GTEx	%Coverage	%Common
Adipose Subcutaneous	137526	1311216	10.5	3.93	15.2
Brain Cortex	125373	154580	81.1	3.59	7.4
Liver	74823	189024	39.6	2.14	11.2
Pancreas	231229	526676	43.9	6.61	5.4
Whole Blood	448413	1060536	42.2	12.8	4.3

Pipeline

Process the variants in Carl's genome to identify any eQTLs. Look at multiple tissues (make sure to include tissues that might be able to be tested in a noninvasive manner e.g. blood).

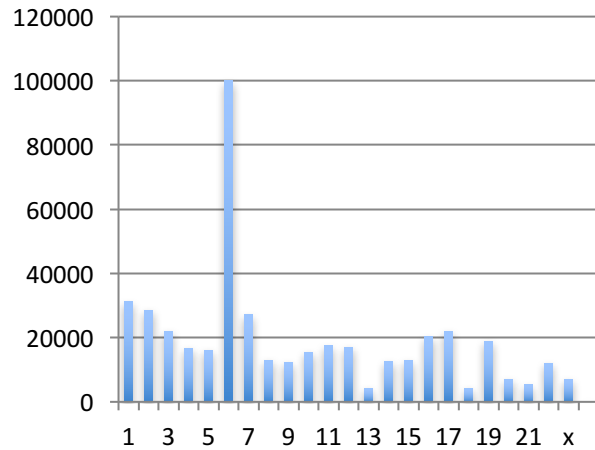
Yuyang Liu

eQTLs identified in different tissues

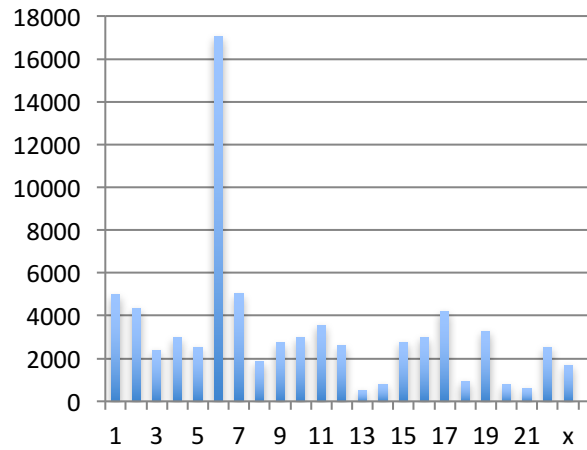


eQTL Chromosome Distribution

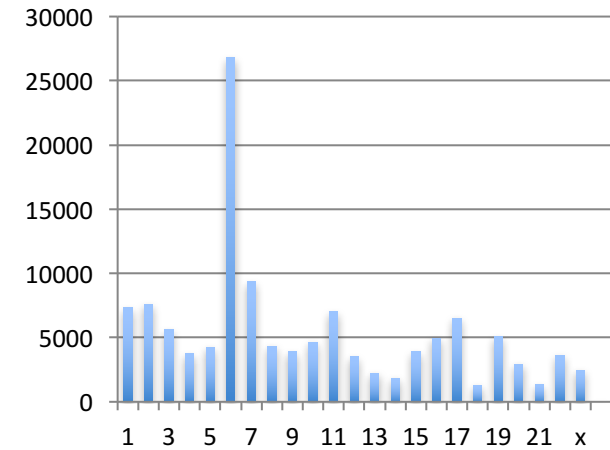
whole blood



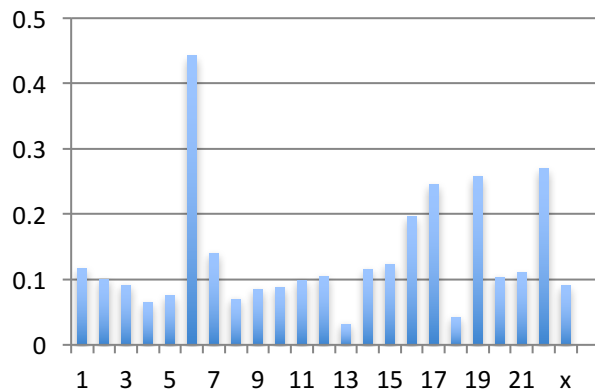
liver



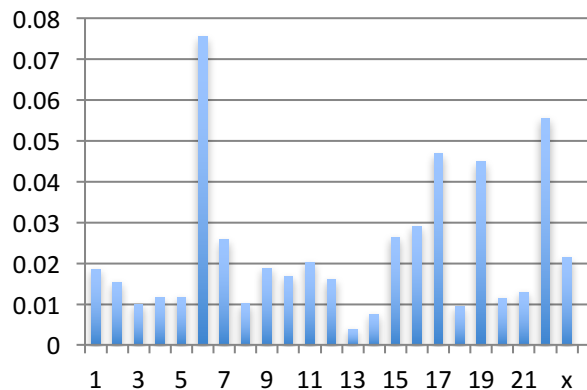
Brain Cortex



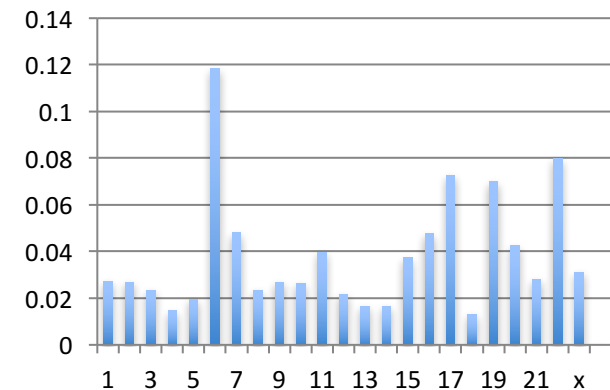
**whole blood normalized
to SNP number**



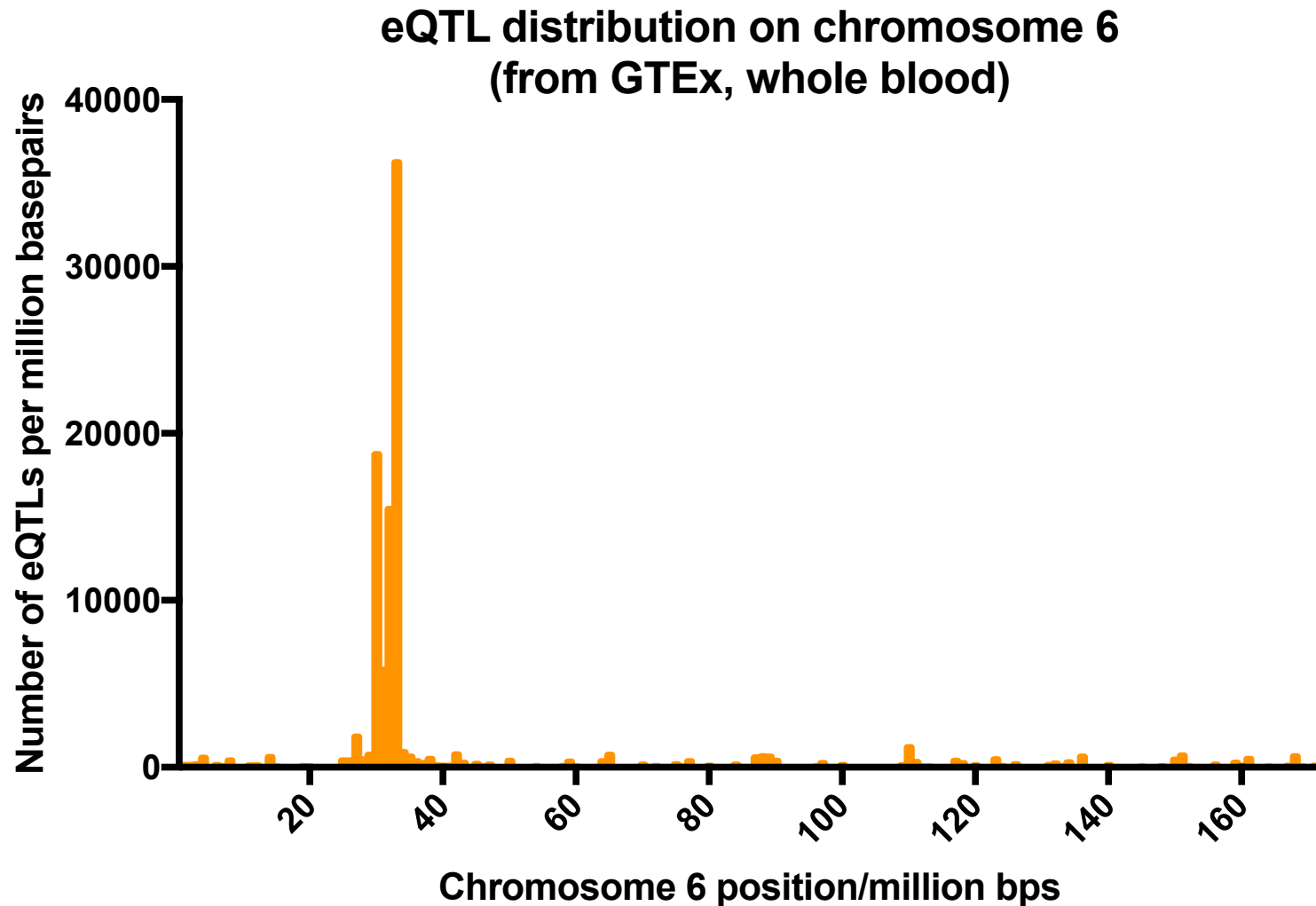
**Liver normalized to SNP
number**



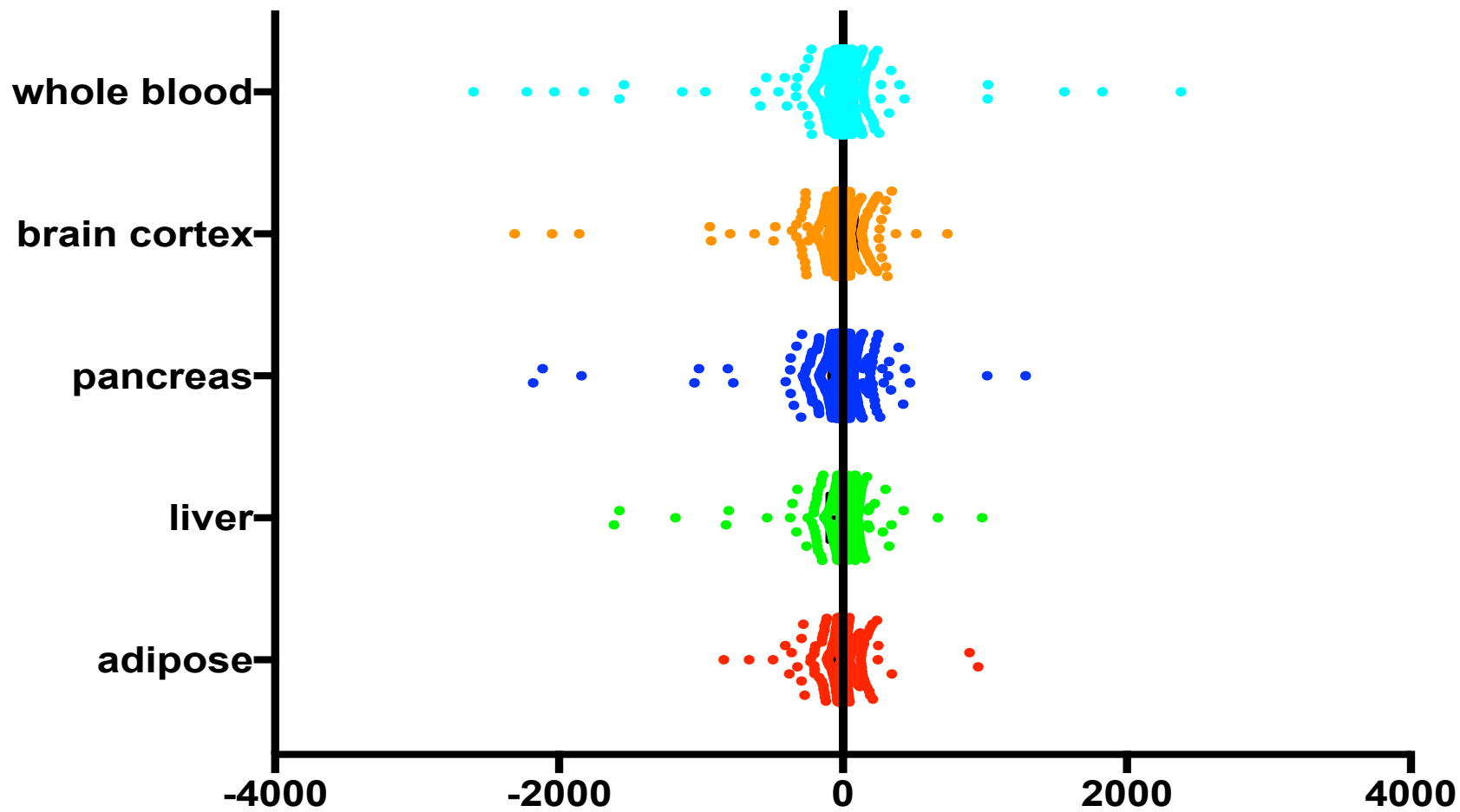
**Brain normalized to SNP
number**



A hotspot of eQTL on Chromosome 6



Distribution of cumulative effector size



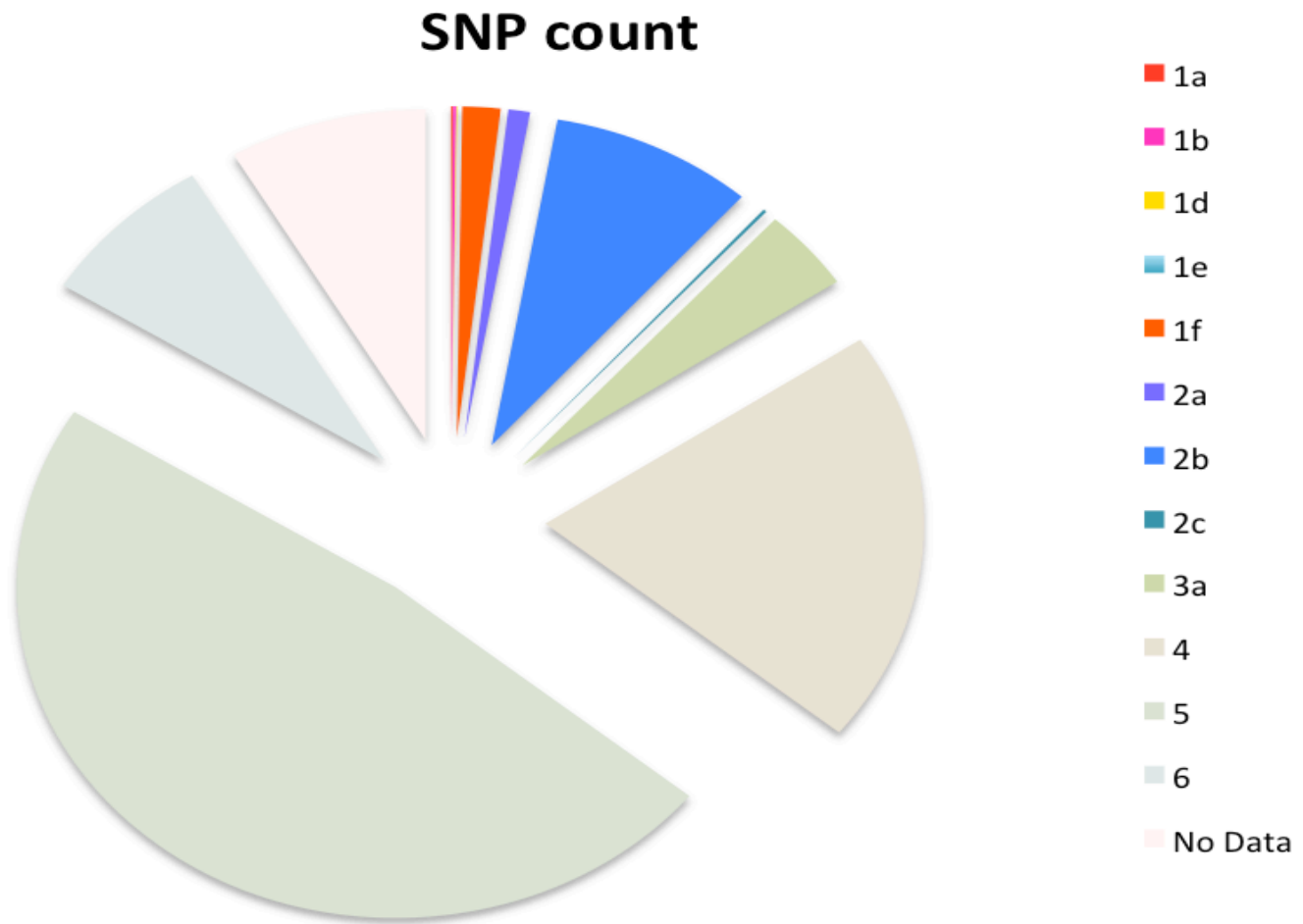
Sample GO analysis--pancreas

GO biological process complete	Homo sapiens - REFLIST (20972)	upload_1 (1661)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (P-value)
small molecule metabolic process (GO:0044281)	1792	198	141.93	+	1.4	1.47E-02
single-organism metabolic process (GO:0044710)	3582	357	283.7	+	1.26	1.71E-02
cellular metabolic process (GO:0044237)	8846	800	700.61	+	1.14	4.29E-03
metabolic process (GO:0008152)	9812	887	777.12	+	1.14	3.31E-04
biological_process (GO:0008150)	17266	1443	1367.48	+	1.06	2.40E-03
Unclassified (UNCLASSIFIED)	3706	218	293.52	-	0.74	0.00E+00
detection of stimulus (GO:0051606)	668	20	52.91	-	0.38	1.22E-03
sensory perception of chemical stimulus (GO:0007606)	529	14	41.9	-	0.33	3.69E-03
detection of stimulus involved in sensory perception (GO:0050906)	519	11	41.11	-	0.27	1.72E-04
detection of chemical stimulus involved in sensory perception (GO:0050907)	467	8	36.99	-	0.22	6.21E-05
detection of chemical stimulus (GO:0009593)	503	8	39.84	-	0.2	6.05E-06
sensory perception of smell (GO:0007608)	453	2	35.88	-	< 0.2	1.06E-09
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	423	1	33.5	-	< 0.2	5.97E-10

Regulome DB analysis

- First 3×10^6 nucleotide on chromosome 1 as input dataset
- Score higher than $1f$ are classified as eQTL

RegulomeDB scores

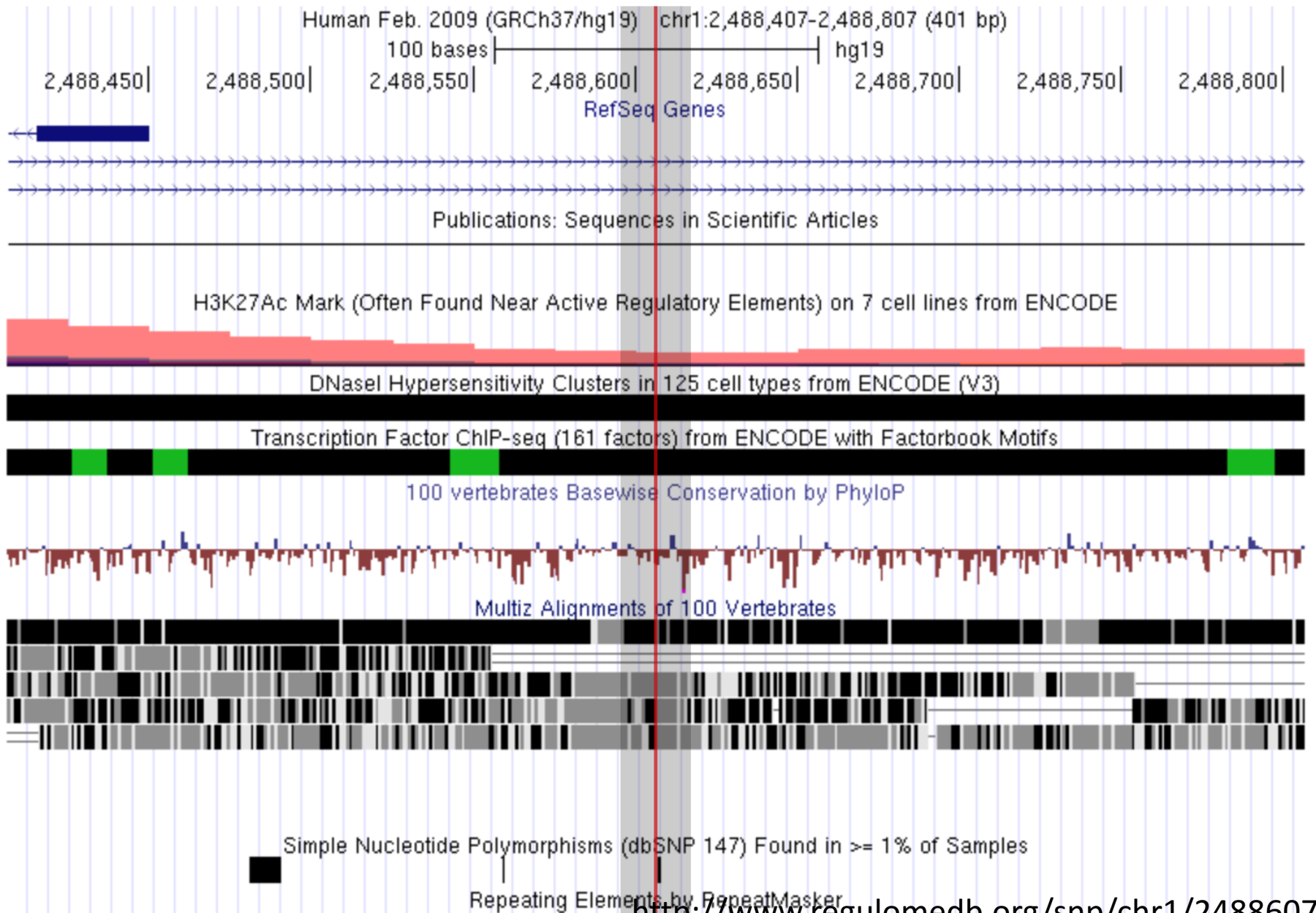


Example eQTL: rs1886730

- Score 1b
- Affects expression of TNFRSF14 in lymphoblastoid

Score: 1b

Likely to affect binding and linked to expression of a gene target

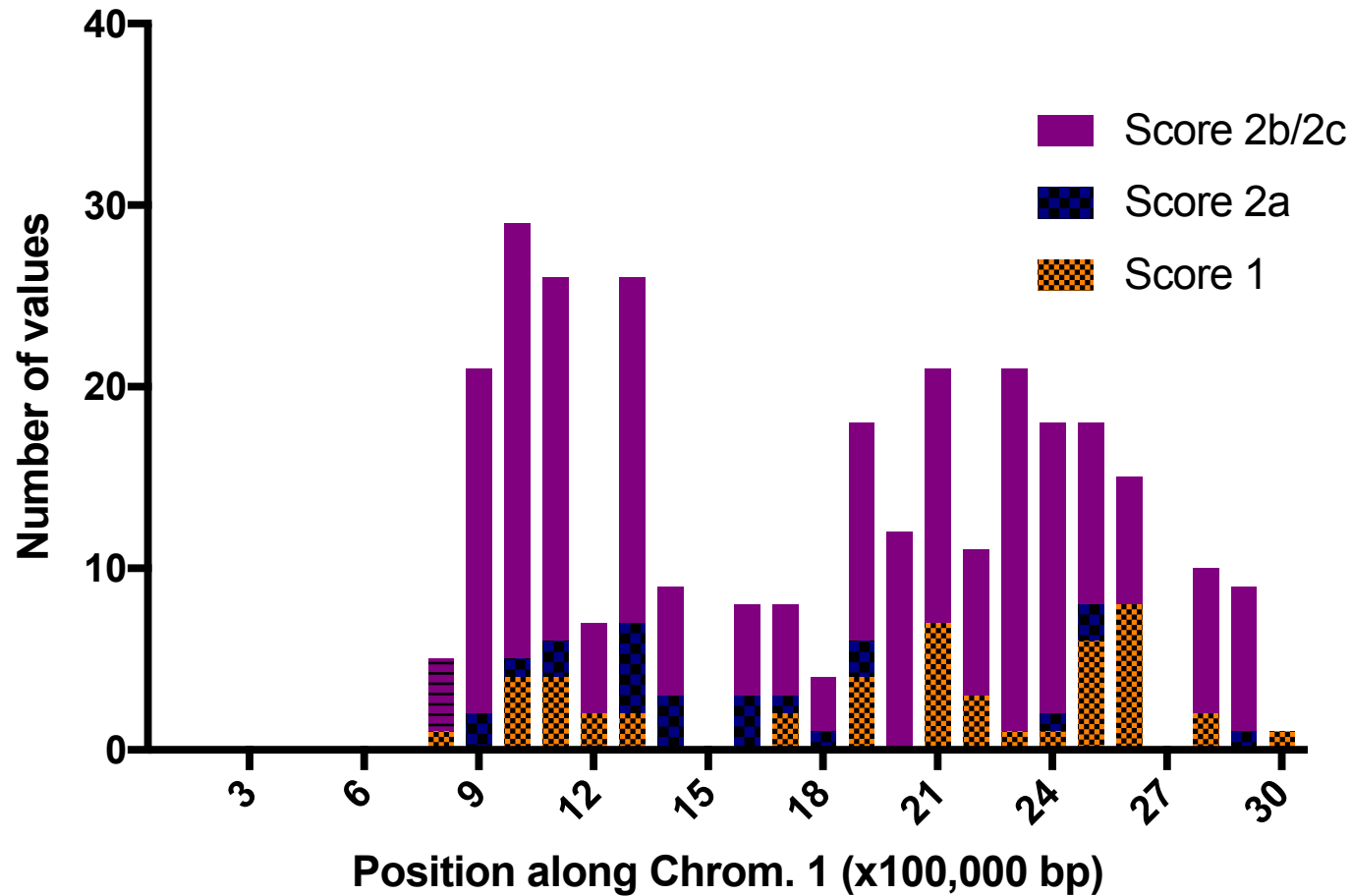


- Located within the binding site of FOXO6
- Evidence from multiple ChIP-Seq experiments
- Near active RSS

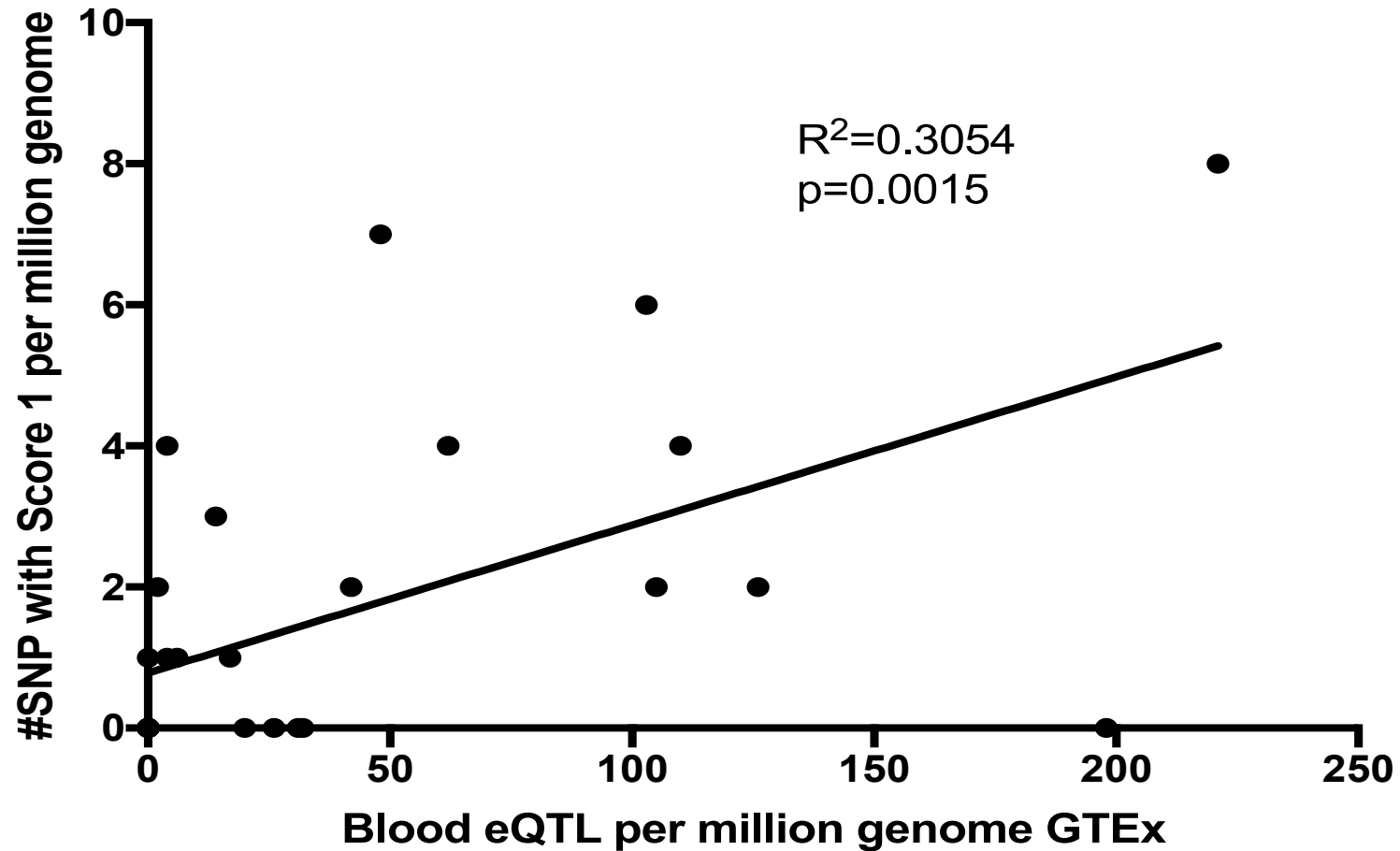


ChromHMM	chr1:2488600..2488800	Active TSS	Blood & T-cell	Primary T cells from cord blood	REMC
----------	-----------------------	------------	----------------	---------------------------------	------

Distribution of SNPs in genome region tested (from Regulome DB)



Correlation between eQTLs predicted by two methods



Summary

- Zimmerome SNPs are interestingly enriched in immune-related chromosome regions
- A small proportion of SNPs serve as eQTLs and regulate the expression of genes (mainly in the immune system)

Thank you!

Supplement

b

Genes sorted by genomic position

