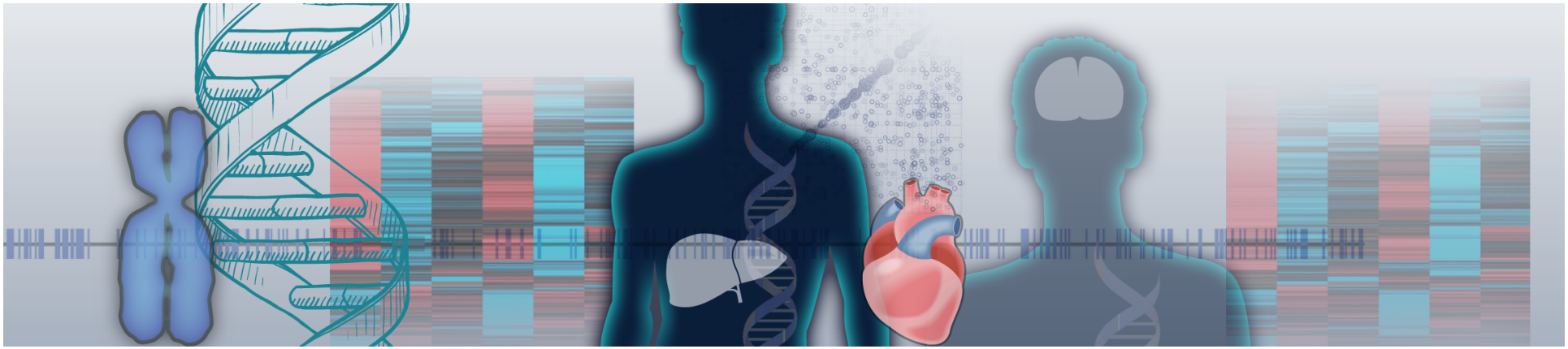


The Genotype-Tissue Expression (GTEx) pilot analysis: multi-tissue gene regulation in humans.

Manolis Dermitzakis
(University of Geneva, Switzerland - GTEx analysis PI)
on behalf of the GTEx consortium

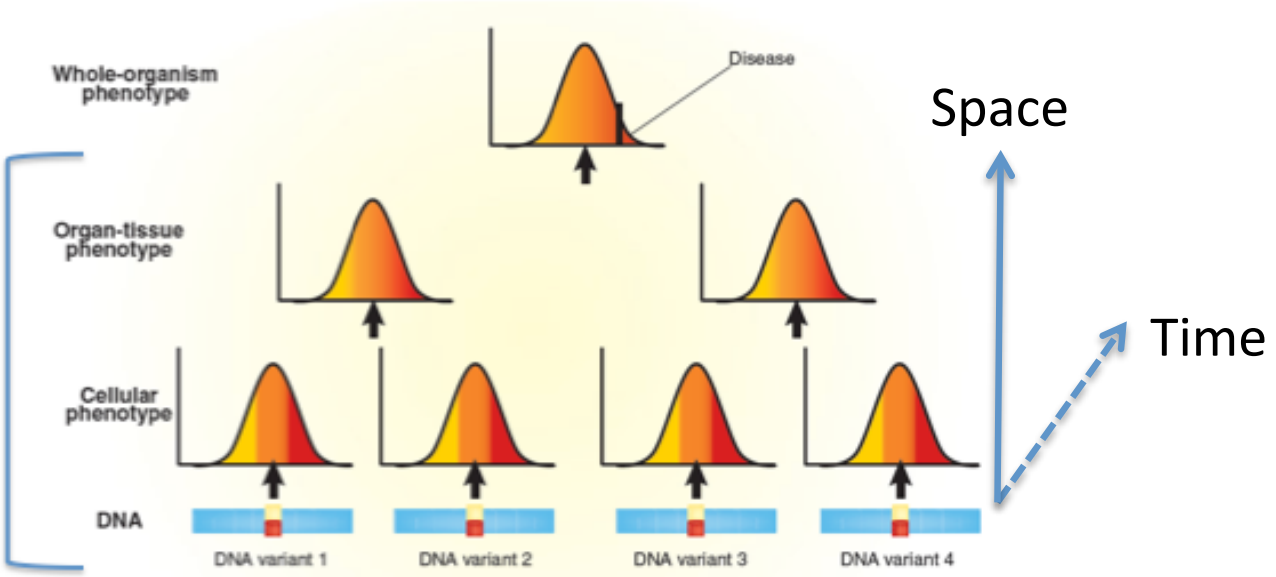
UNIVERSITÉ DE GENÈVE

GTEx

NIH
The Common Fund

Challenge: How do we go from trait-associated variants to biological mechanism?

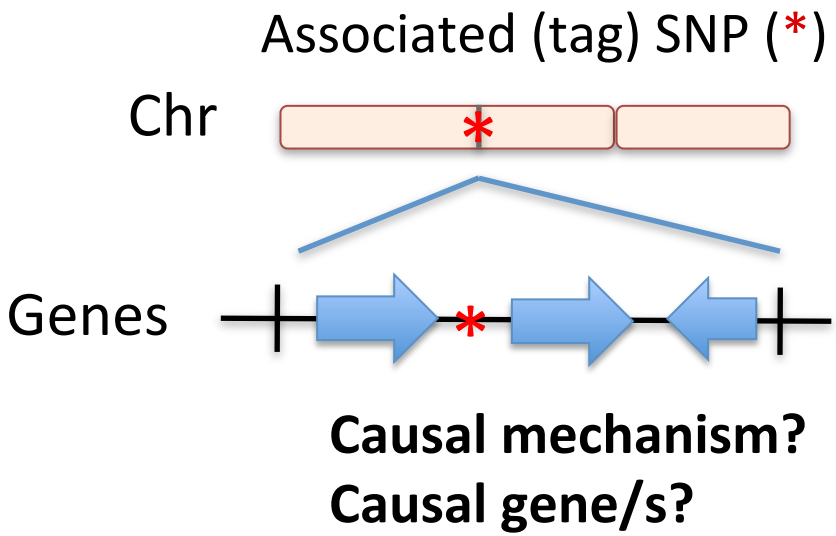
ENVIRONMENT
- Behavior
- Infections
- External factors
- etc



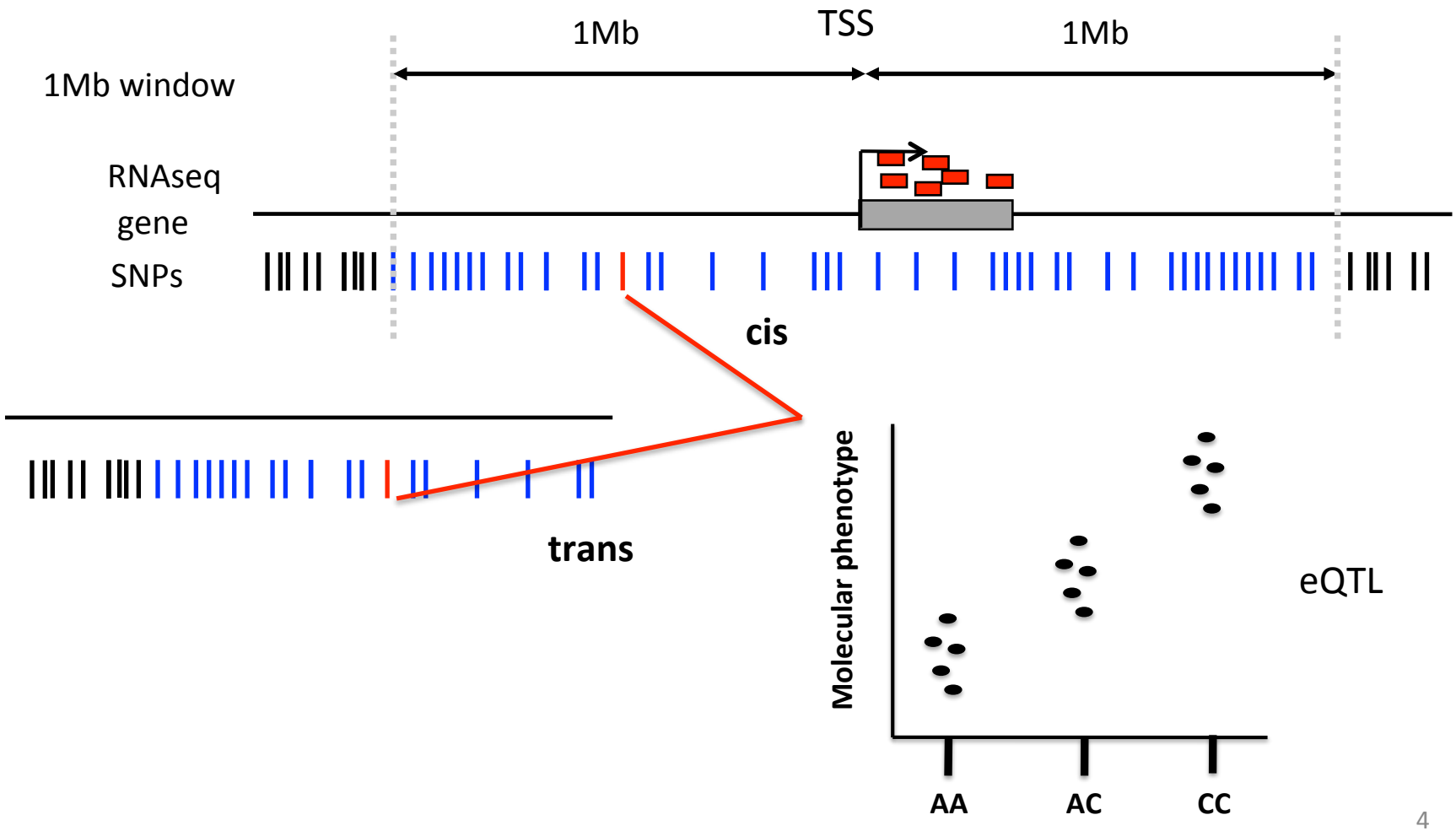
Background Rationale

Genome-wide association studies (**GWAS**) have identified **hundreds of common DNA variants** associated with multiple **complex diseases and traits**.

>2/3 GWAS SNPs lie in noncoding regions (e.g. intergenic, introns).



Expression Quantitative Trait Loci (eQTL)



Many studies show trait-associated SNPs enriched for eQTLs

OPEN ACCESS Freely available online

PLoS GENETICS

Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS

LCL eQTLs

Dan L. Nicolae[1,2,3], Eric Gamazon[1], Wei Zhang[1], Shiwei Duan[1¤], M. Eileen Dolan[1,2], Nancy J. Cox[1,2*]

1 Department of Medicine, University of Chicago, Chicago, Illinois, United States of America, 2 Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America, 3 Depar

OPEN ACCESS Freely available online

PLoS GENETICS

LCL eQTLs

Candidate Causal Regulatory Effects by Integration of Expression QTLs with Complex Trait Genetic Associations

Alexandra C. Nica[1,2], Stephen B. Montgomery[1,2], Antigone S. Dimas[1,2], Barbara E. Stranger[1,3], Claude Beazley[1], Inês Barroso[1], Emmanouil T. Dermitzakis[1,2*]

1 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom, 2 Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland, 3 Harvard Medical School/Brigham and Women's Hospital, Boston, Massachusetts, United States of America

Coanalysis of GWAS with eQTLs reveals disease-tissue associations

Hyunseok Peter Kang, M.D.[1], Alex A. Morgan, Ph.D.[1], Rong Chen, Ph.D.[1], Eric E. Schadt, Ph.D.[2], Atul J. Butte, M.D, Ph.D.[1]

[1] Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA

[2] Department of Genetics and Genome Sciences, Mount Sinai School of Medicine, New York, NY

Peripheral blood monocyte, liver and adipose eQTLs

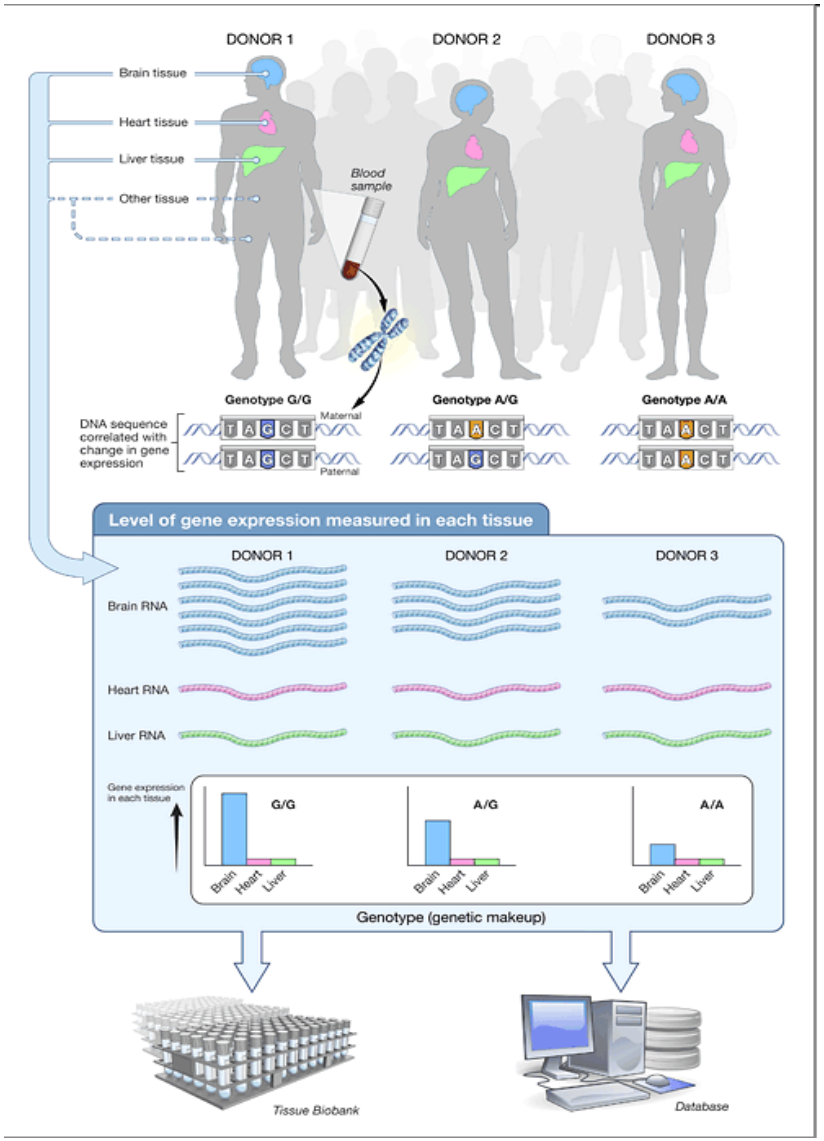
Challenges in using eQTLs to interpret disease associations

- Measuring eQTLs in disease-relevant tissues or cell types

- Most human tissue types are hard to obtain

- Large sample sizes are required for statistical power

GTEx = Genotype-Tissue Expression



GTEx GOALS:

- Atlas (database) of gene expression, regulation, and eQTLs from a wide range of non-diseased human tissues
- Biobank of tissues, DNA, RNA

ULTIMATE STUDY SIZE (by 1/2016):

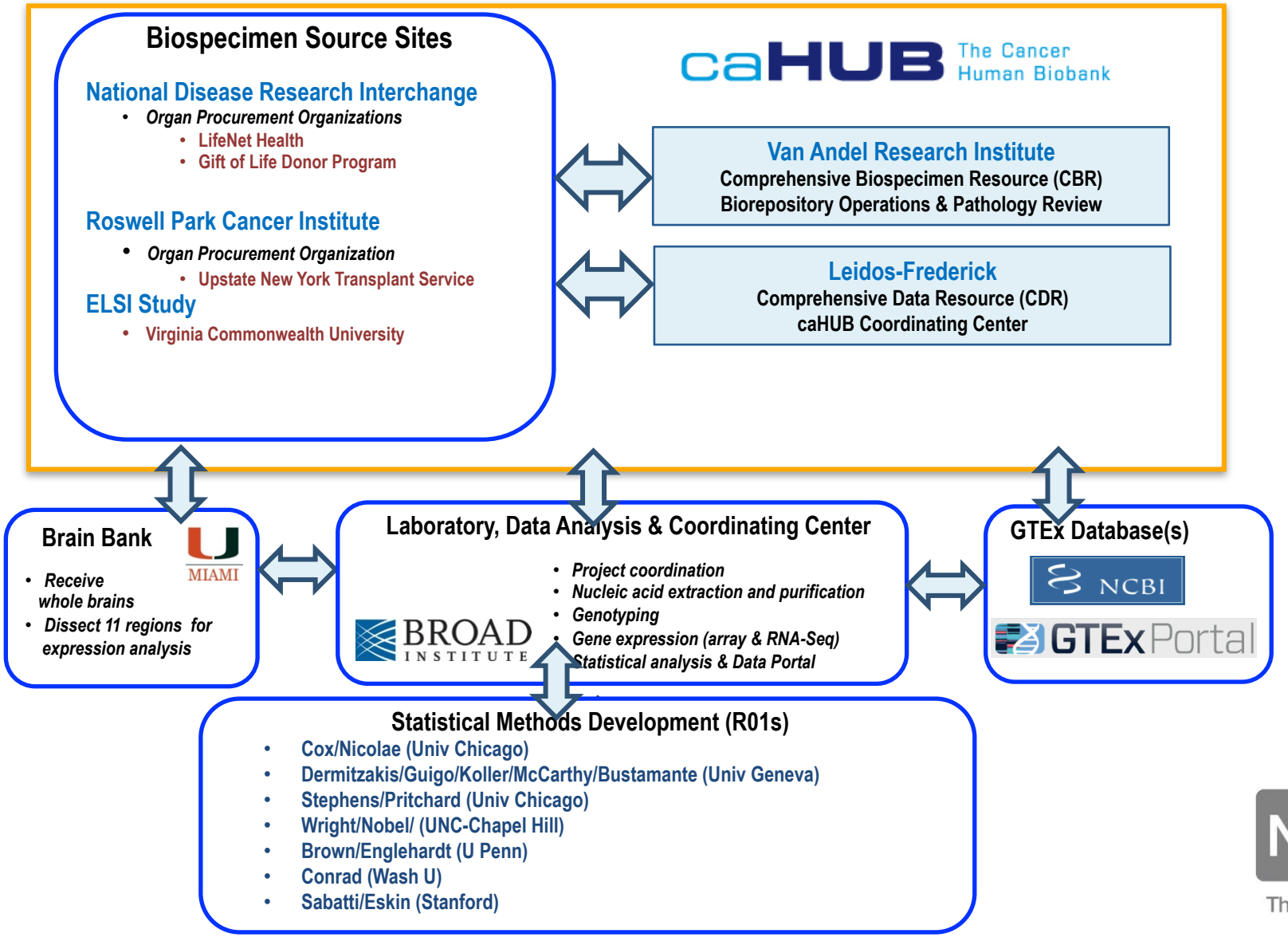
- 900 Postmortem Donors
- Whole exome sequencing
- Whole genome sequencing
- RNA-Seq of ~30 tissues/donor (>20,000 tissues)

NIH
The Common Fund

PILOT PHASE (in 2010):

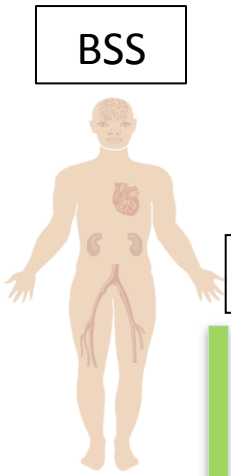
- 175 Postmortem Donors
- 1641 RNA-Seq of ~28 tissues/donor

Pilot GTEx Consortium



NIH The Common Fund

Laboratory Data Analysis and Coordinating Center GTEx Workflow

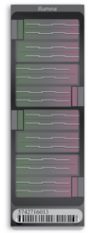


BSS

Donor

Bloods & Fresh Skin

1. Whole Blood → DNA
2. Whole Blood → EBV Culture
3. PAXgene Tube → RNA
4. Fresh Skin → Fibroblast Culture



Illumina 5M + exome

LDACC/Broad Institute

Tissues

CBR/Path

PAXgene Tissues → RNA

Brain

Brain Bank

Snap Frozen Tissues → RNA

Illumina RNA Seq

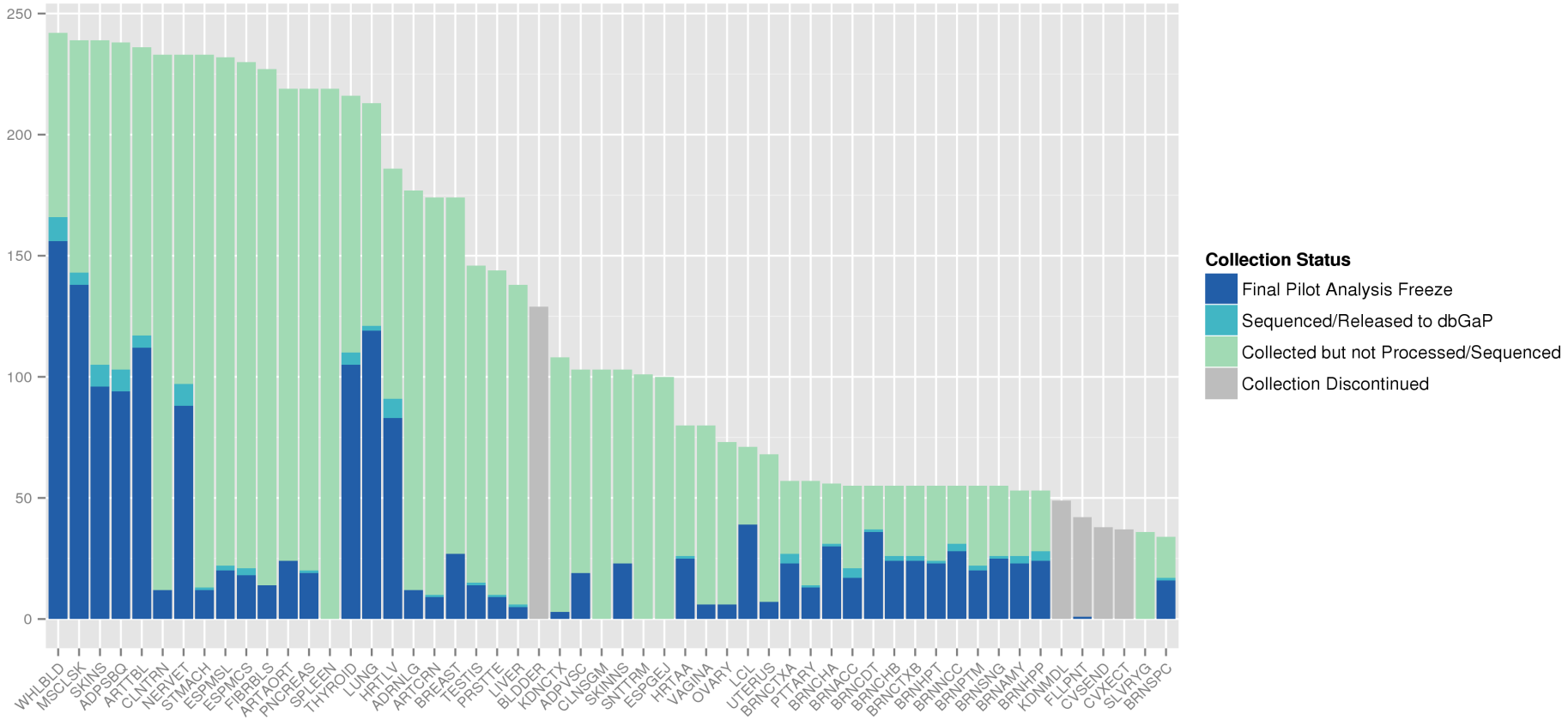
Analysis Pipeline

GTEx Analysis Working Group
dbGAP

PILOT Data Production:

1. Illumina 5 Million SNP chip + exome chip
2. Illumina TruSeq, 76bp paired-end to 50 million reads

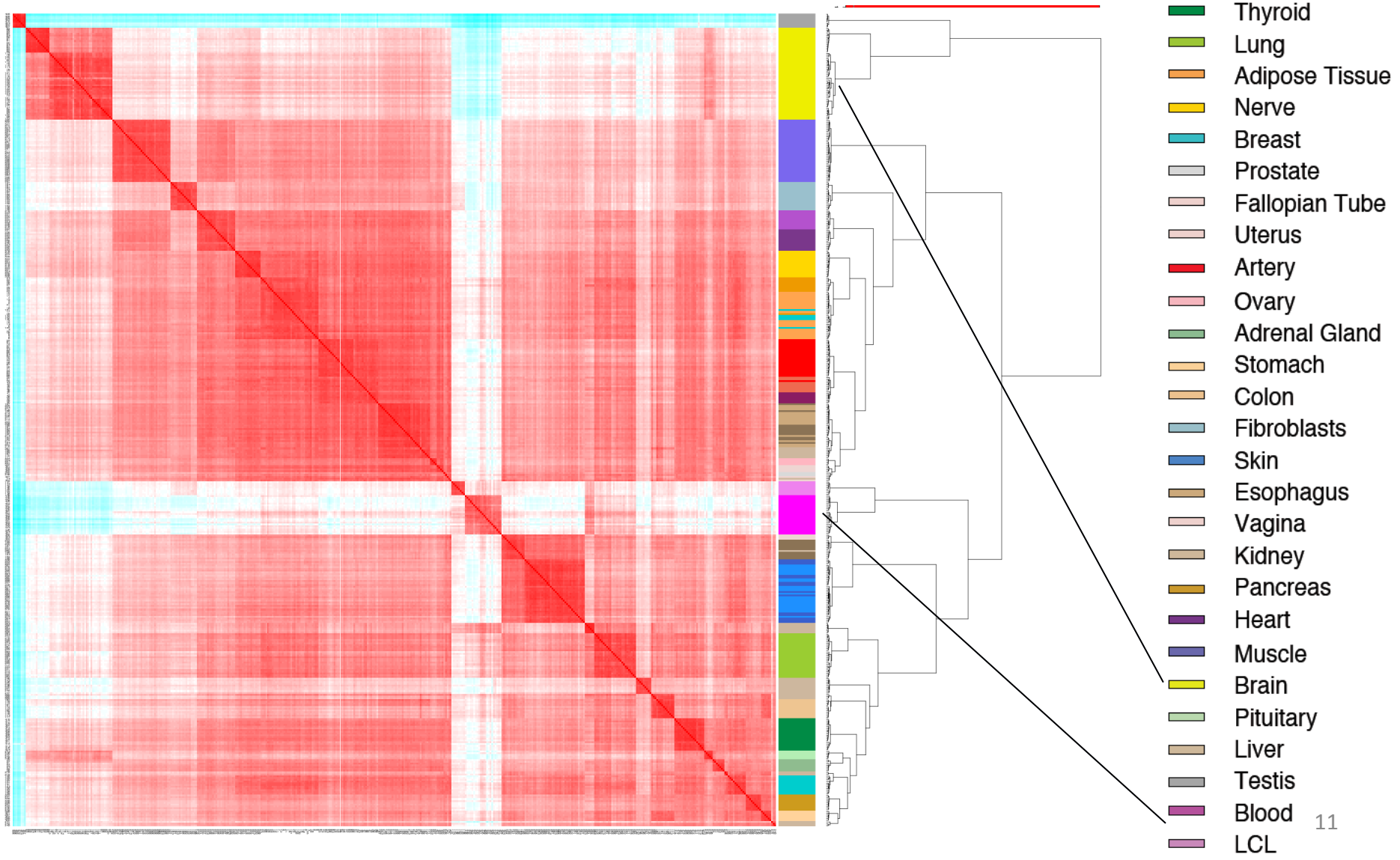
Pilot Phase Data

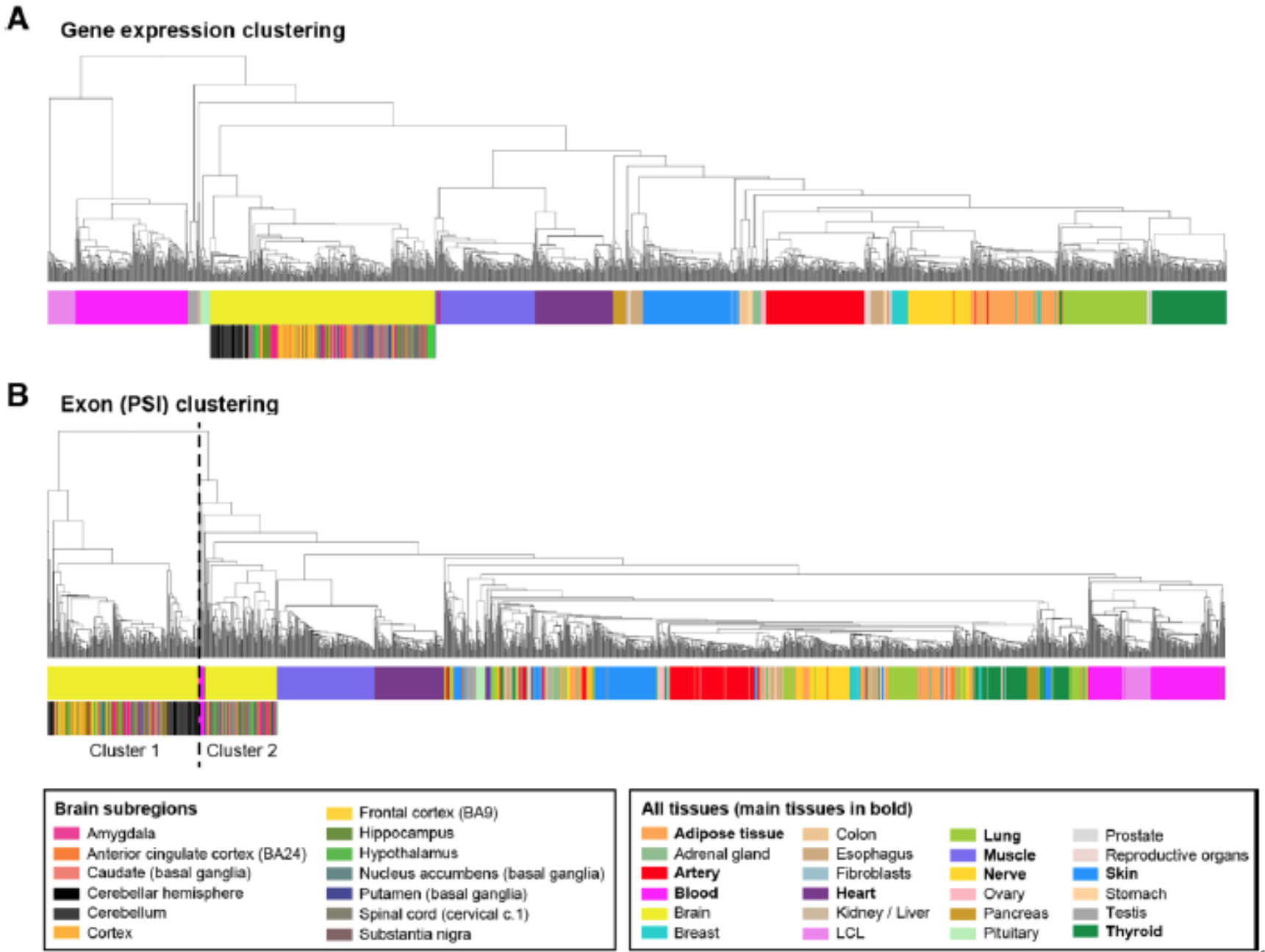


Pilot phase collected an average of ~ 28 tissue samples per donor from 54 distinct sites

Analysis Freeze = 175 Donors and 1,641 tissue RNA-seq samples

Clustering of Expression Data



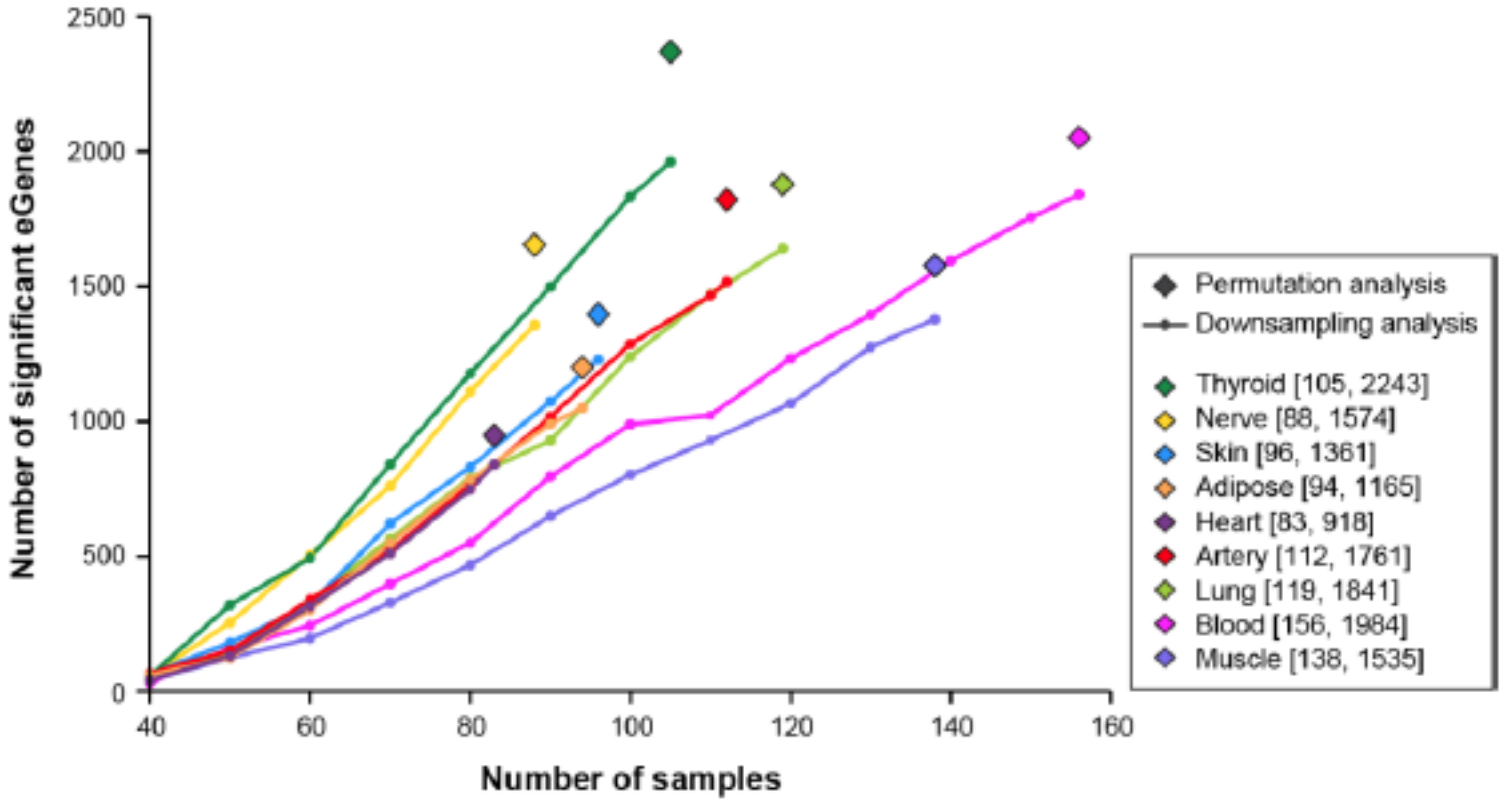

A
Gene expression clustering
B
Exon (PSI) clustering
Cluster 1
Cluster 2
Brain subregions
Amygdala
Anterior cingulate cortex (BA24)
Caudate (basal ganglia)
Cerebellar hemisphere
Cerebellum
Cortex
Frontal cortex (BA9)
Hippocampus
Hypothalamus
Nucleus accumbens (basal ganglia)
Putamen (basal ganglia)
Spinal cord (cervical c.1)
Substantia nigra
All tissues (main tissues in bold)
Adipose tissue
Adrenal gland
Artery
Blood
Brain
Breast
Colon
Esophagus
Fibroblasts
Heart
Kidney / Liver
LCL
Lung
Muscle
Nerve
Ovary
Pancreas
Pituitary
Prostate
Reproductive organs
Skin
Stomach
Testis
Thyroid

eQTL methodology

- 1 Mb (or 100Kb from TSS)
- PEER correction to remove undesirable effects (batch, other experimental etc)
- Matrix eQTL with permutations to discover eQTLs

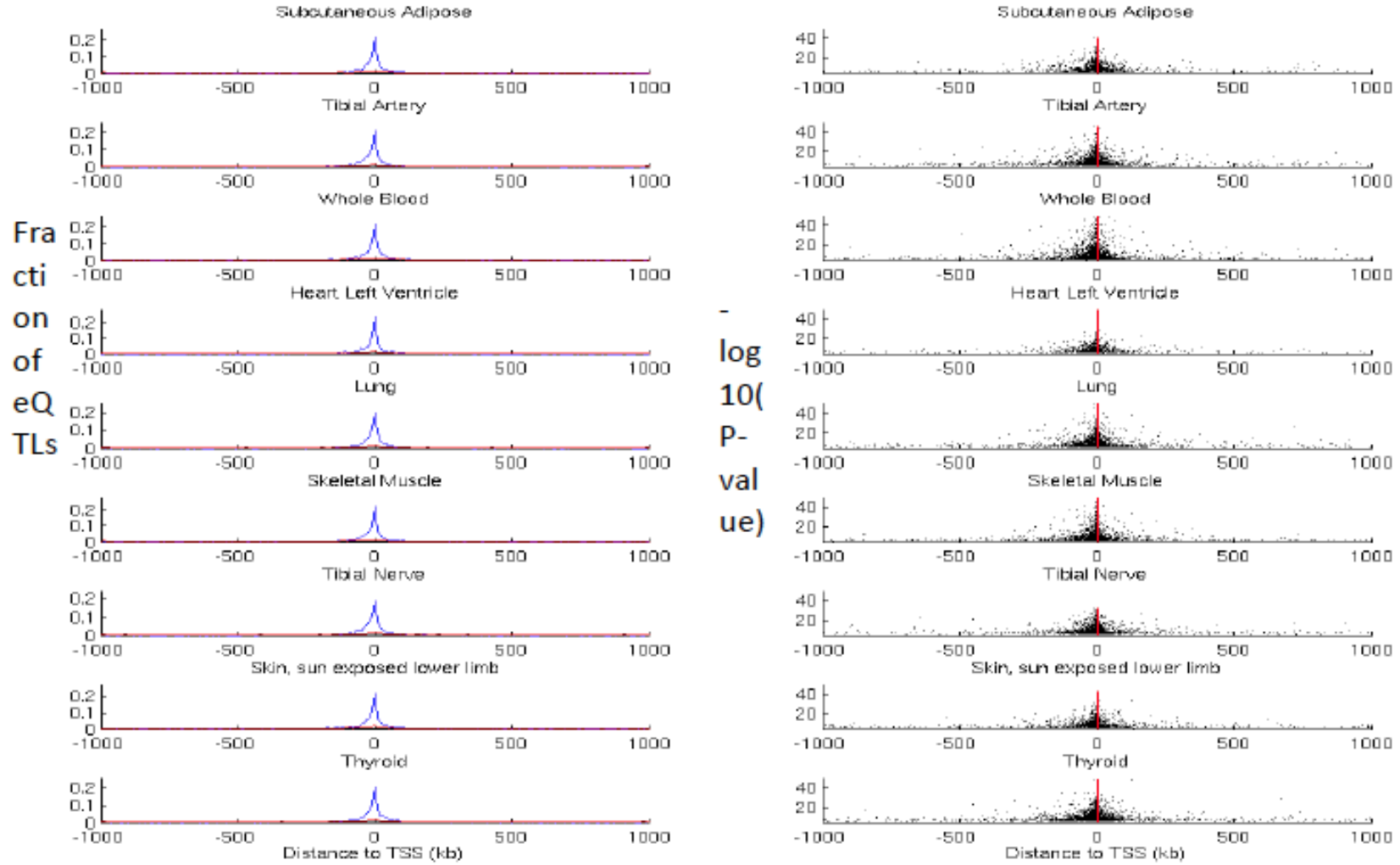
(now we use FastQTL developed by Olivier Delaneu)

eQTL discovery



A.

eQTL properties



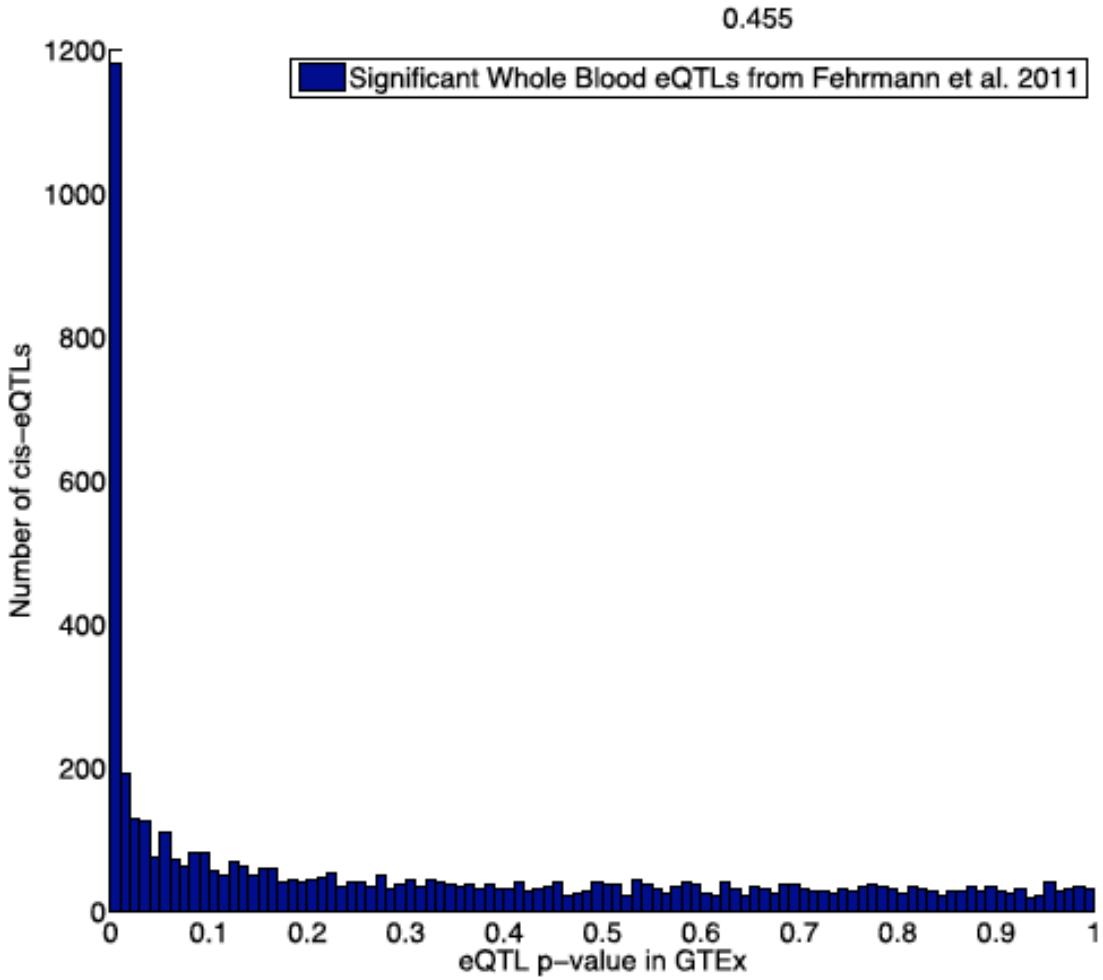
Significant best eQTLs per gene, FDR<5%

Non-significant best eQTLs per gene, FDR>5%

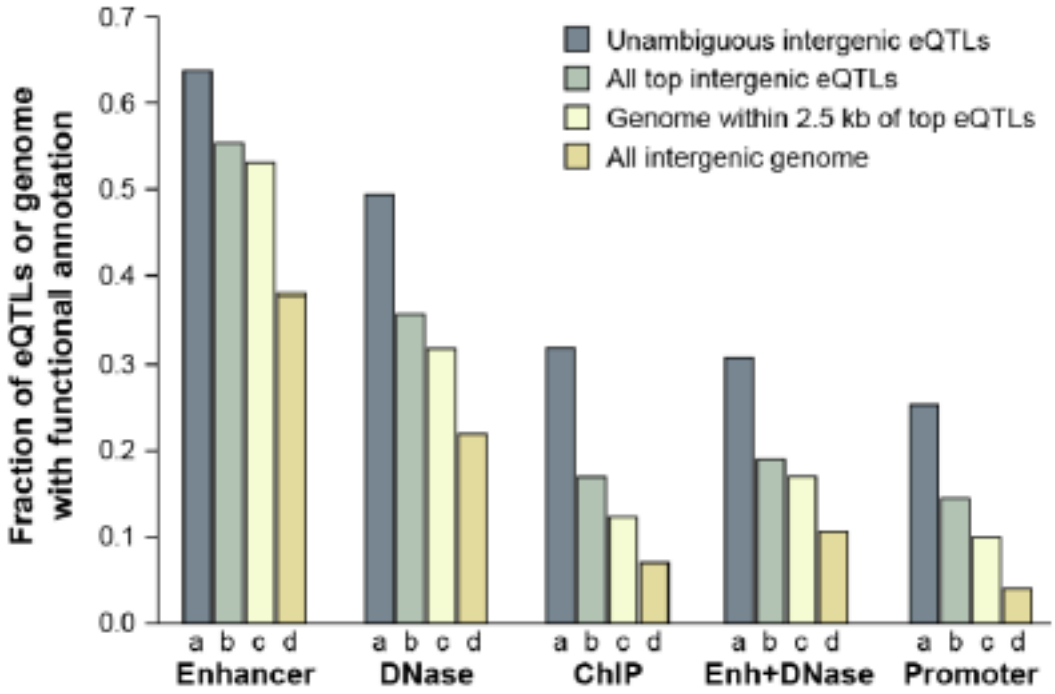
All SNP-gene pairs

eQTL replication

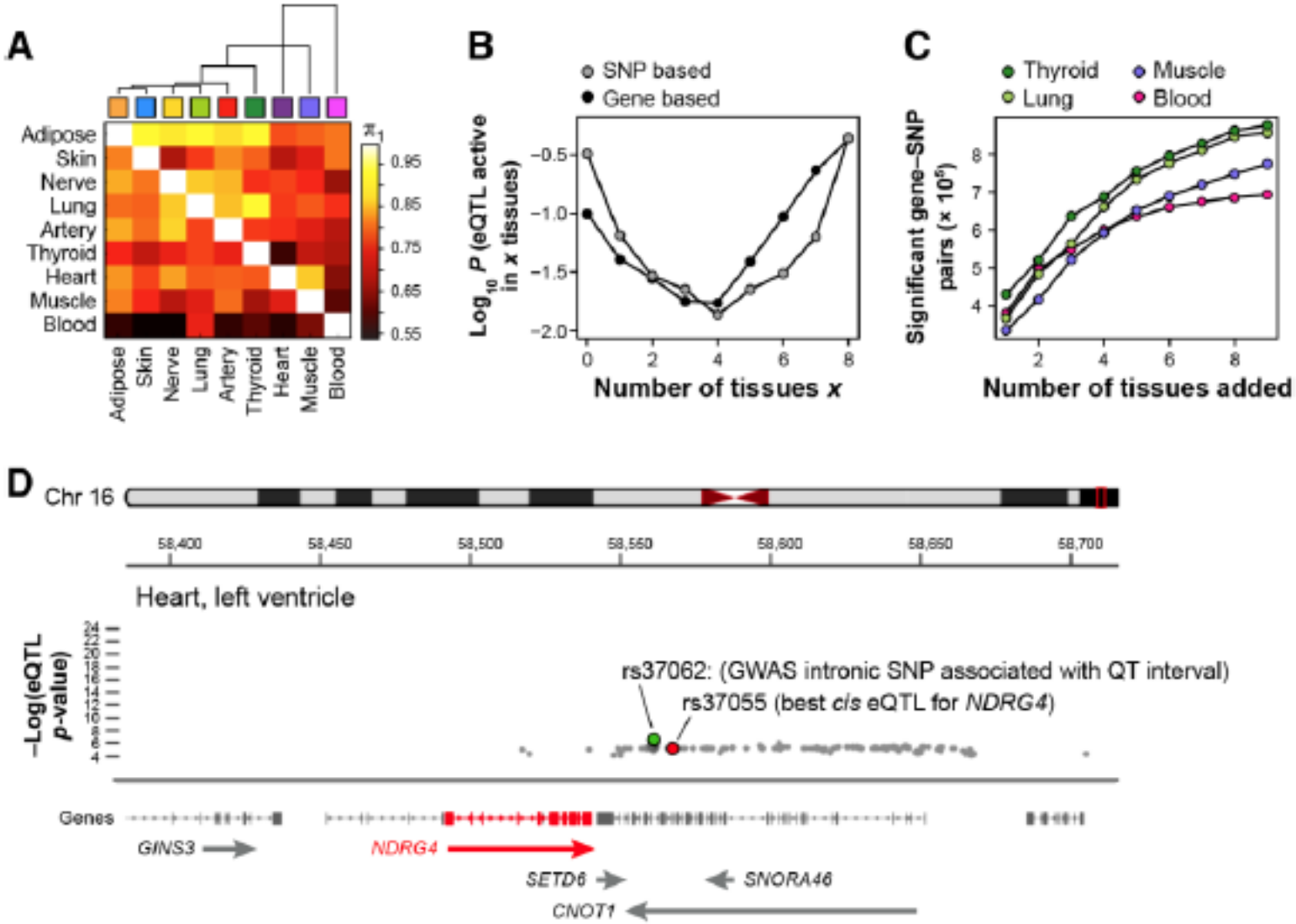
A.



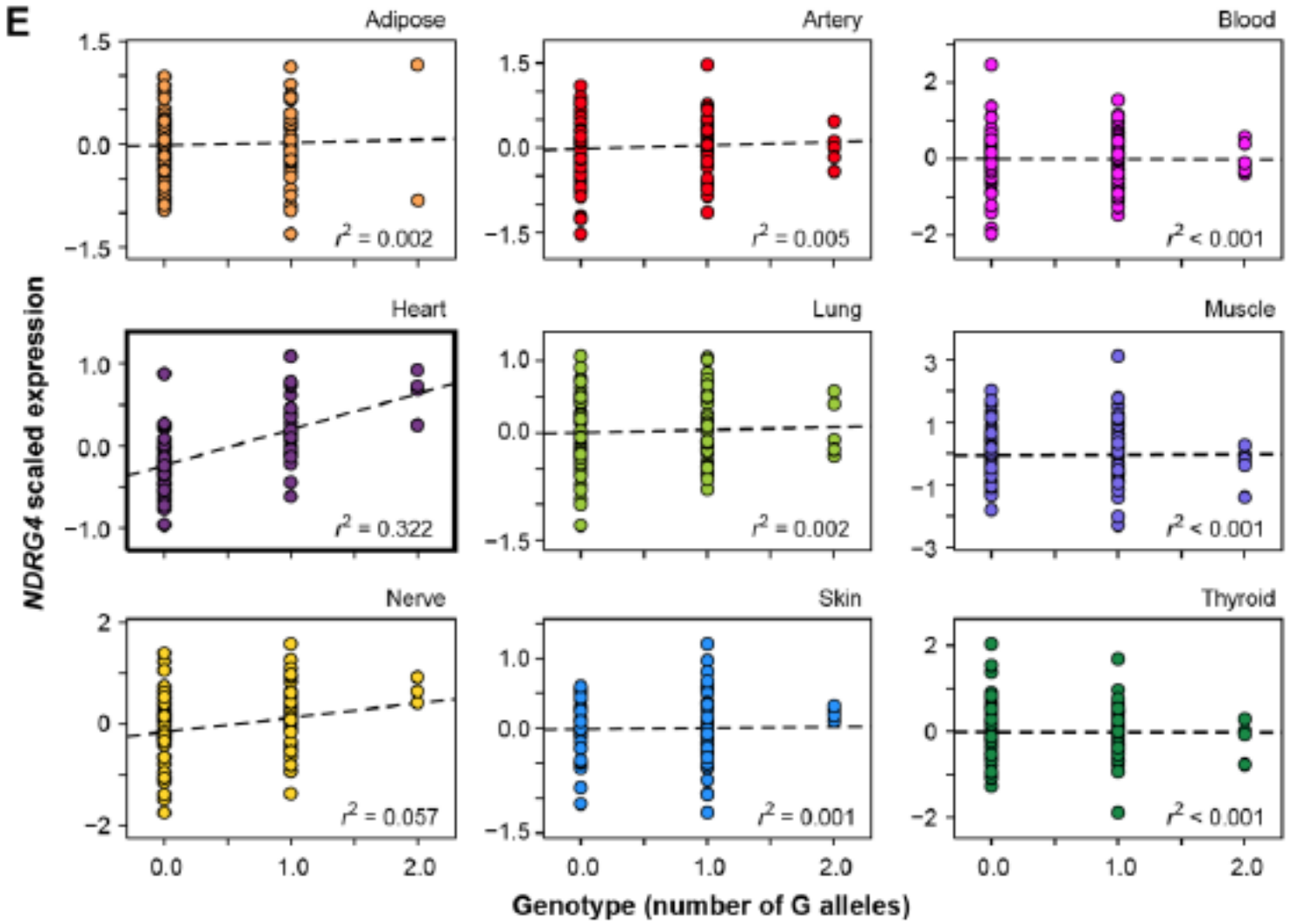
eQTL functional enrichment



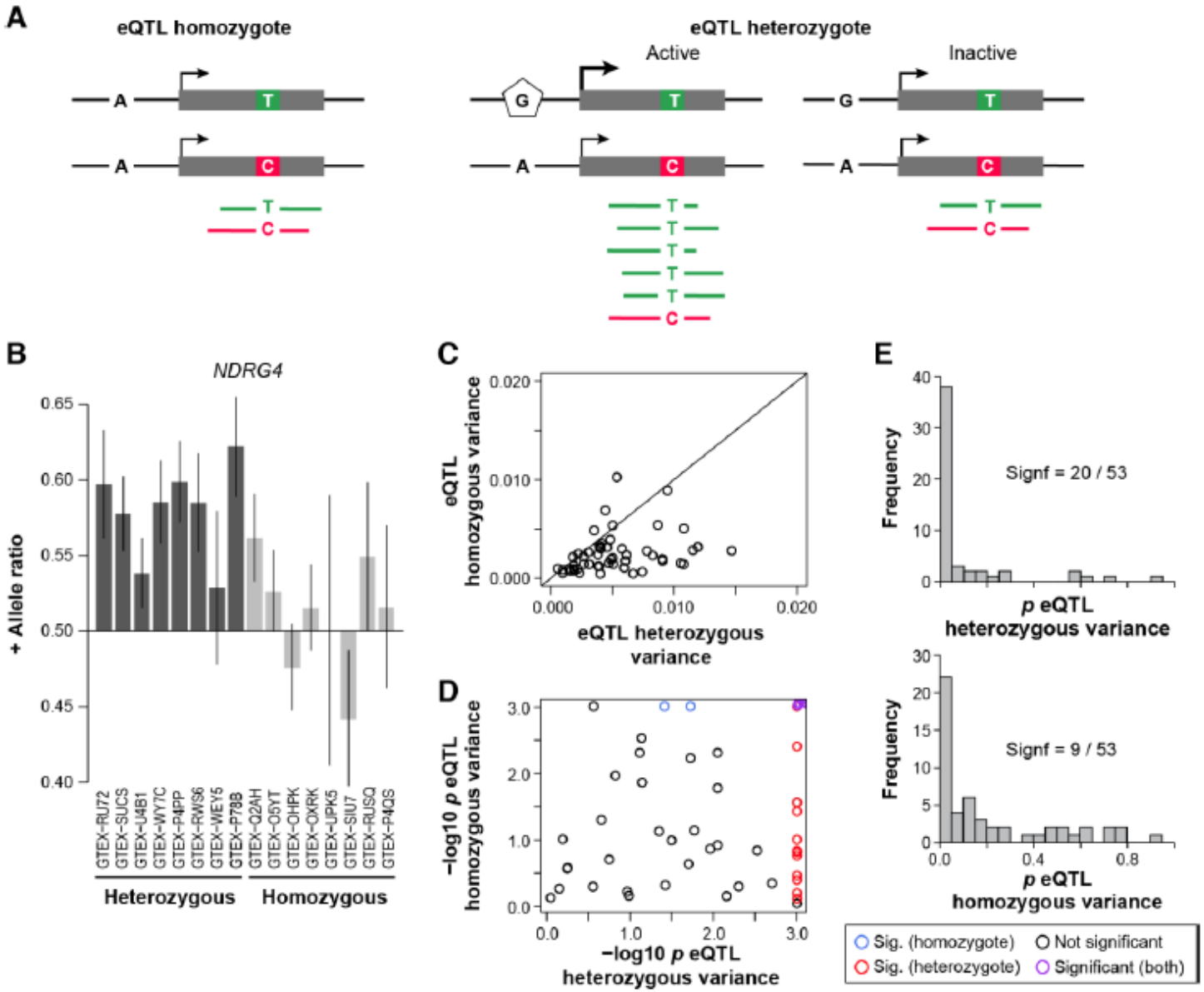
Multi-tissue eQTL discovery



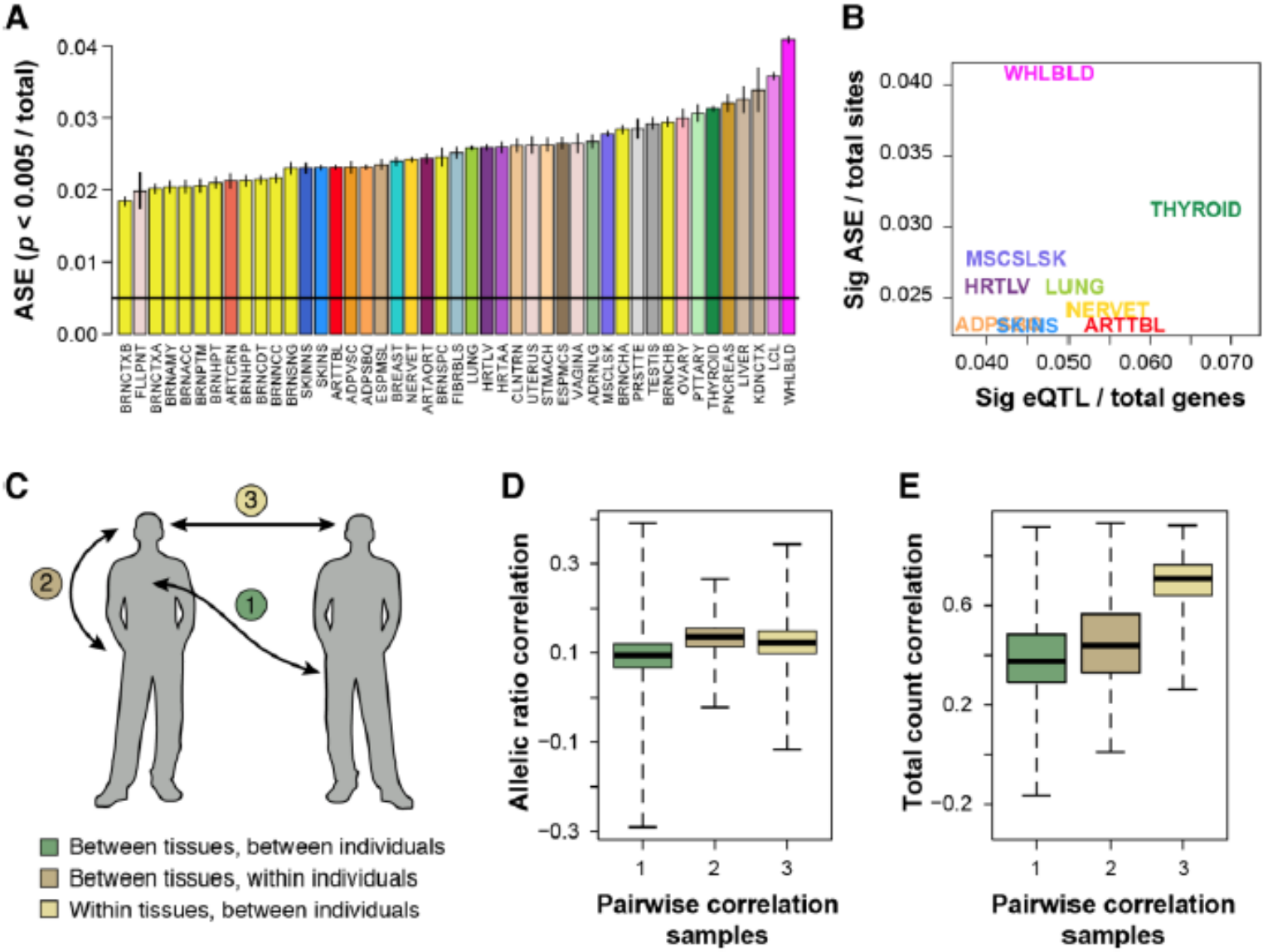
Multi-tissue eQTL discovery



Allele Specific Expression and eQTLs

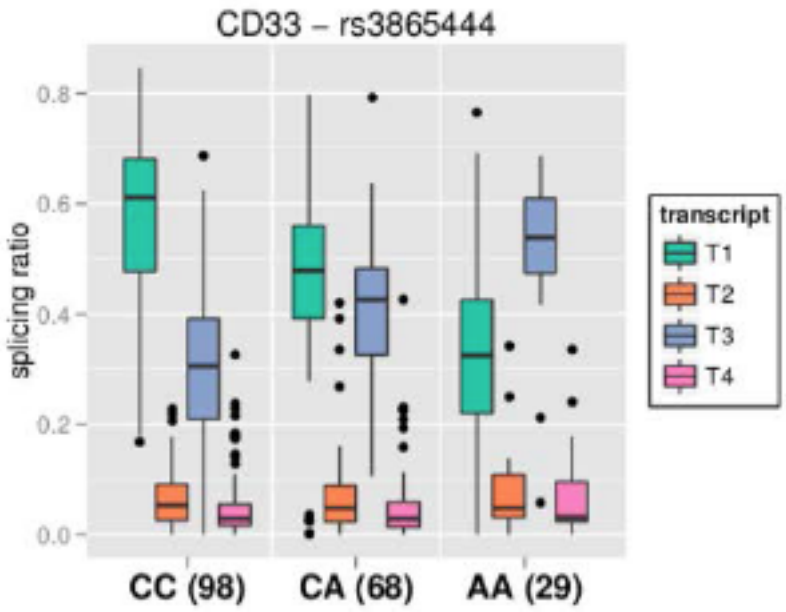


Allele Specific Expression

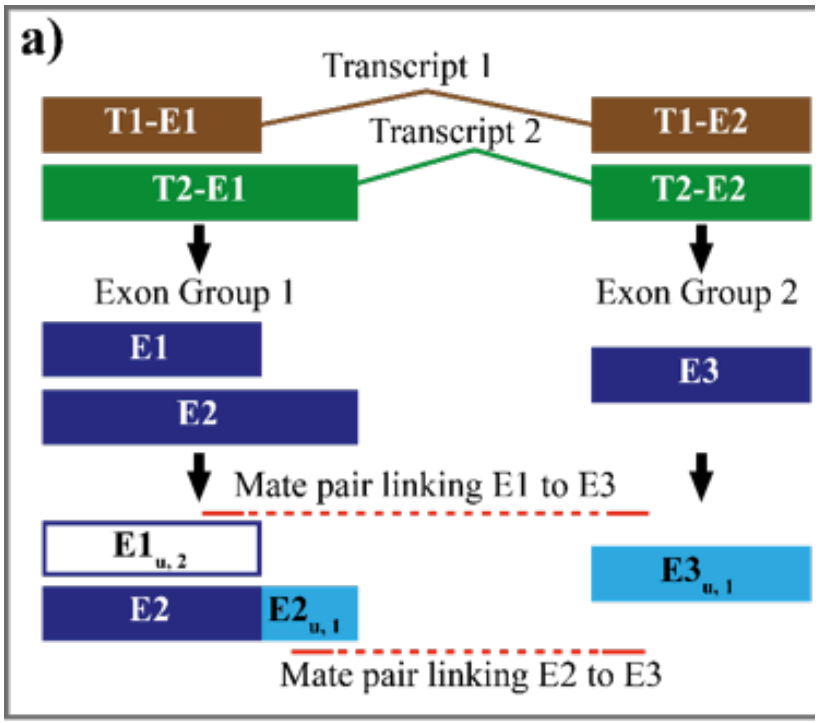


Alternative splicing QTLs

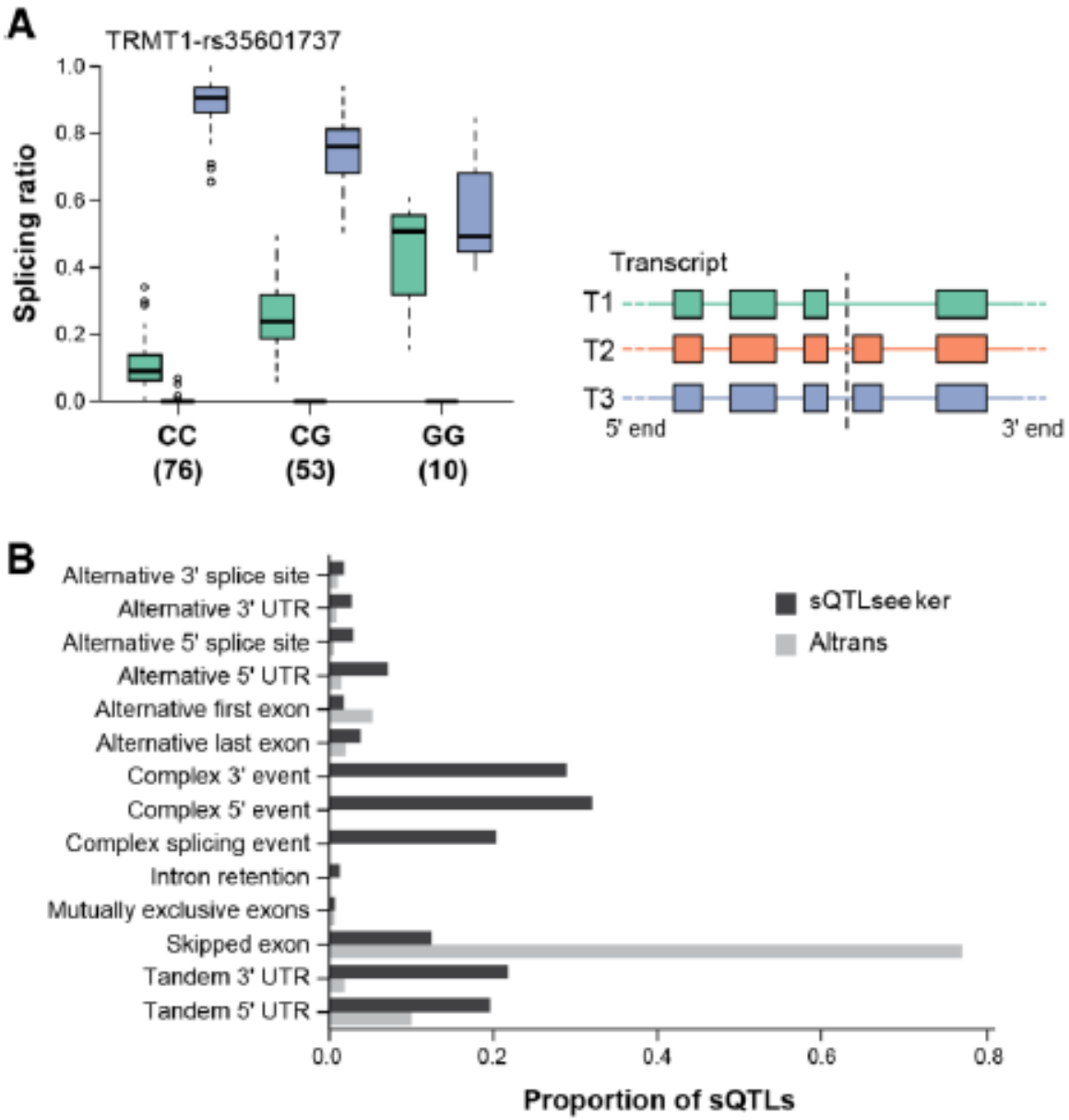
sQTL seeker



Altrans

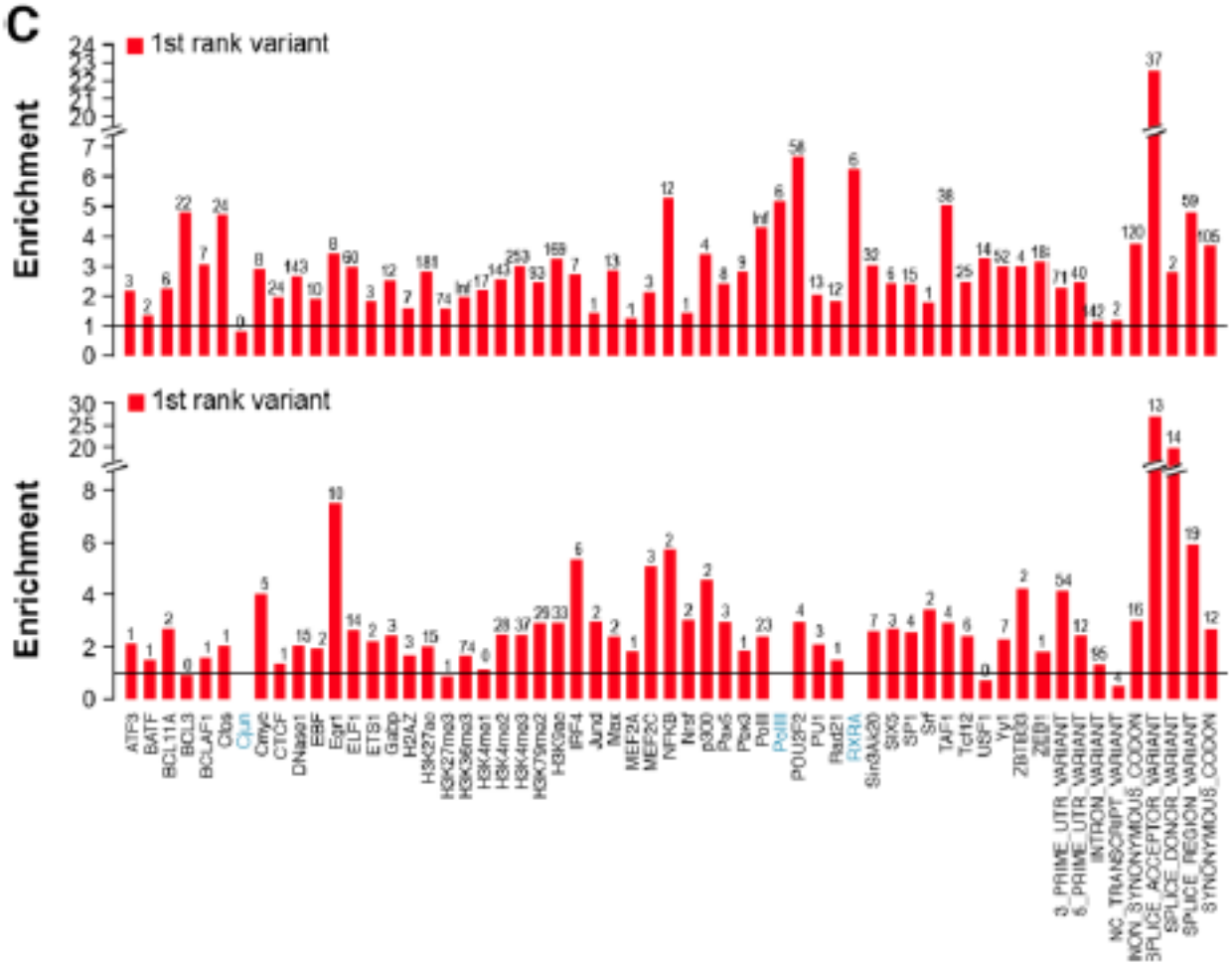


Alternative splicing QTLs

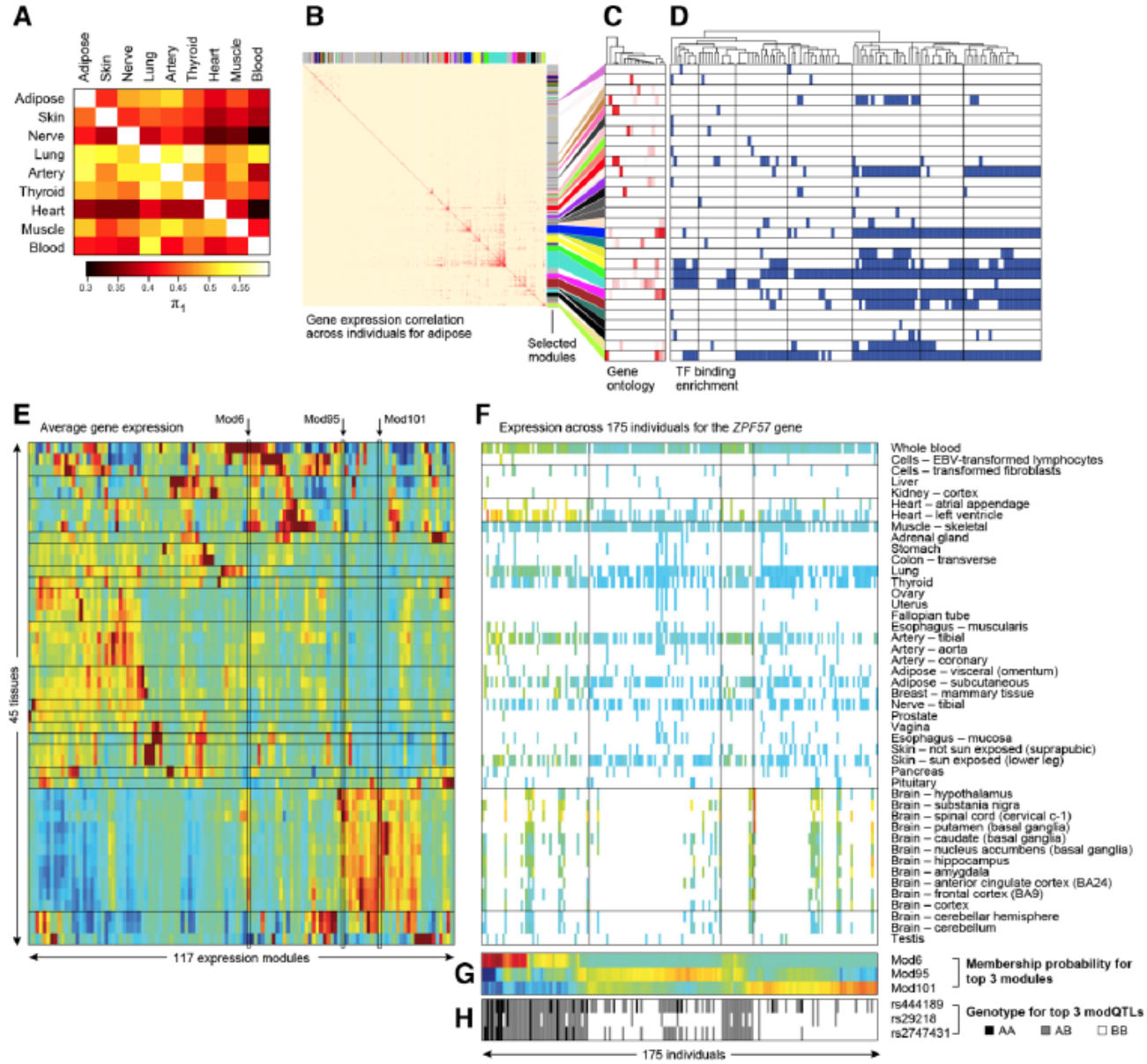


Alternative splicing QTLs

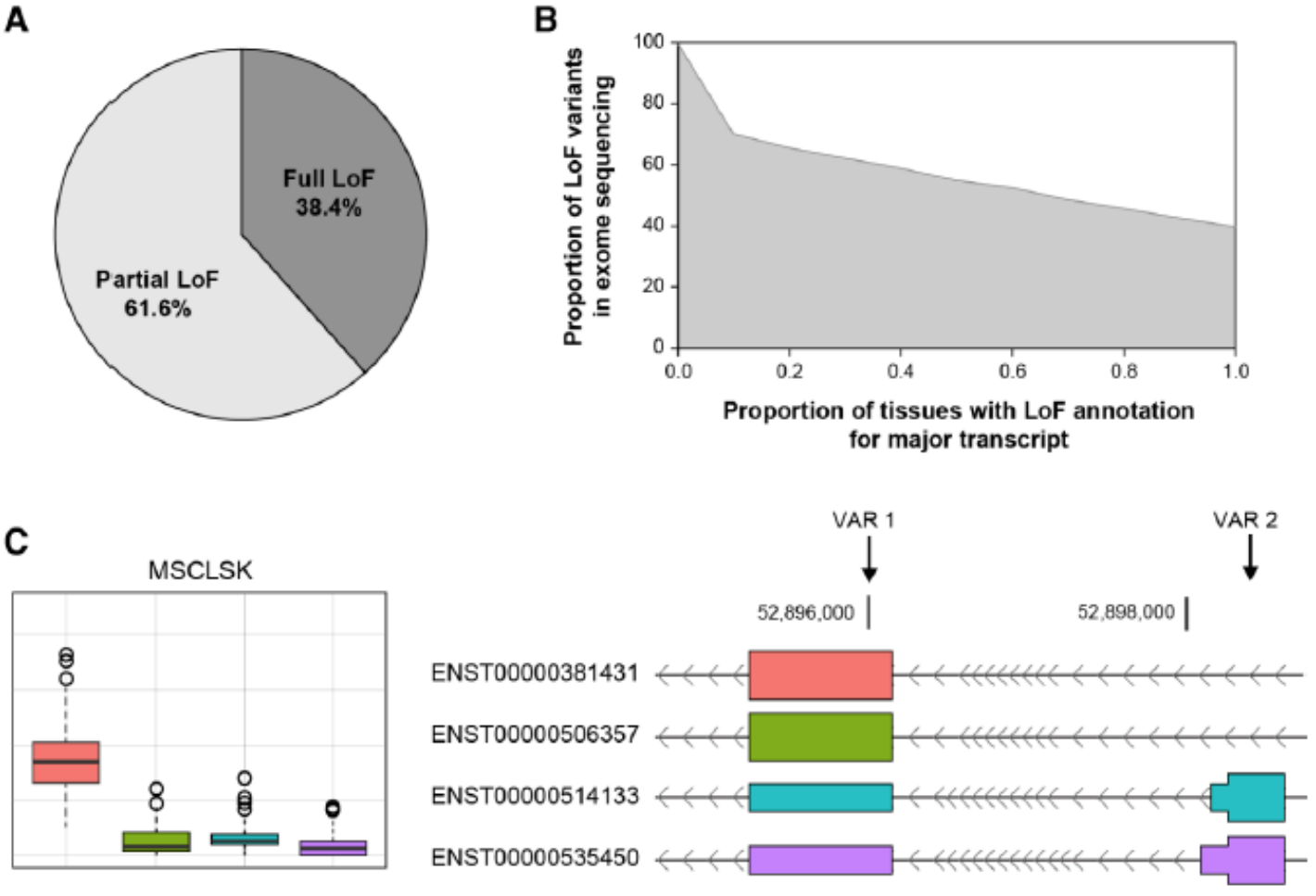
Functional element enrichment



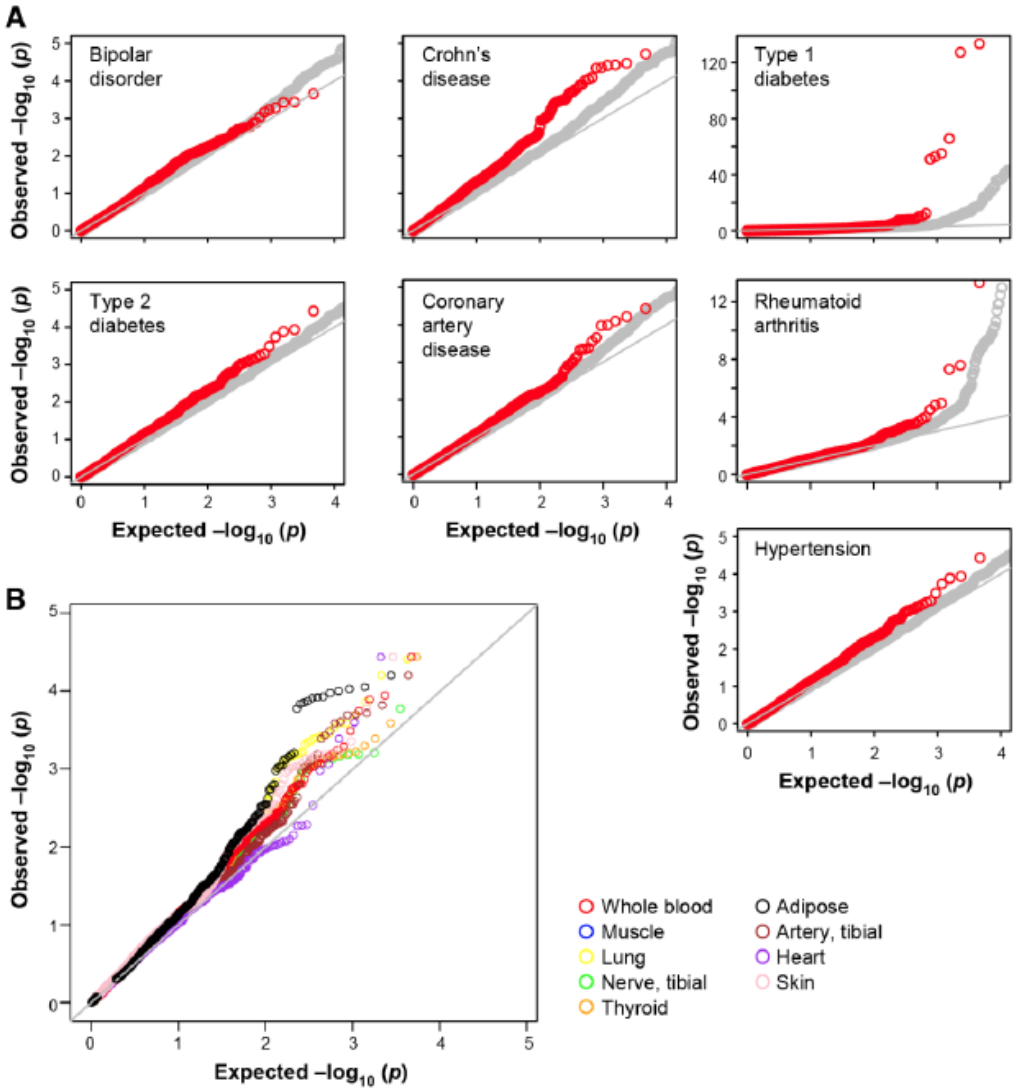
Network module QTLs



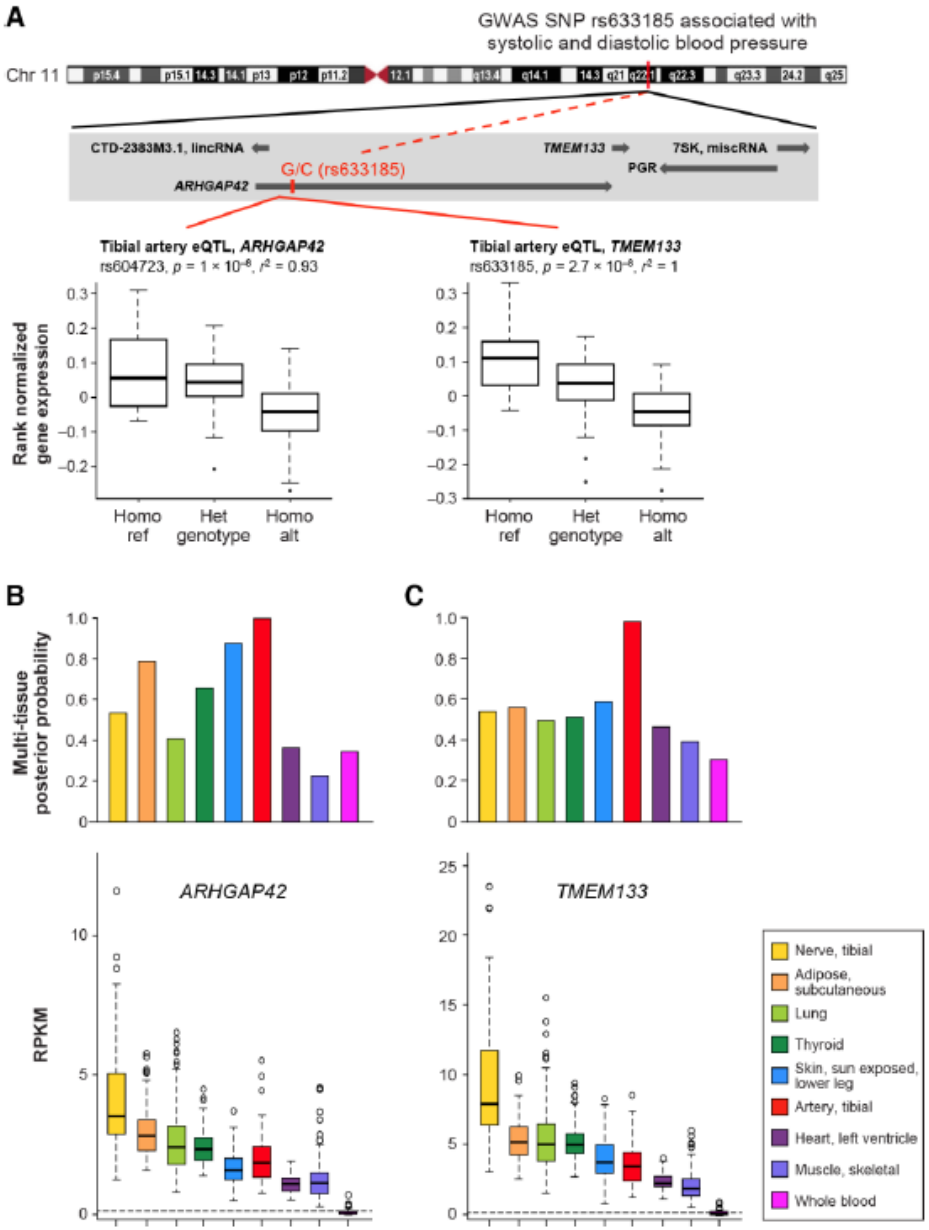
Truncating Variants in the transcriptome



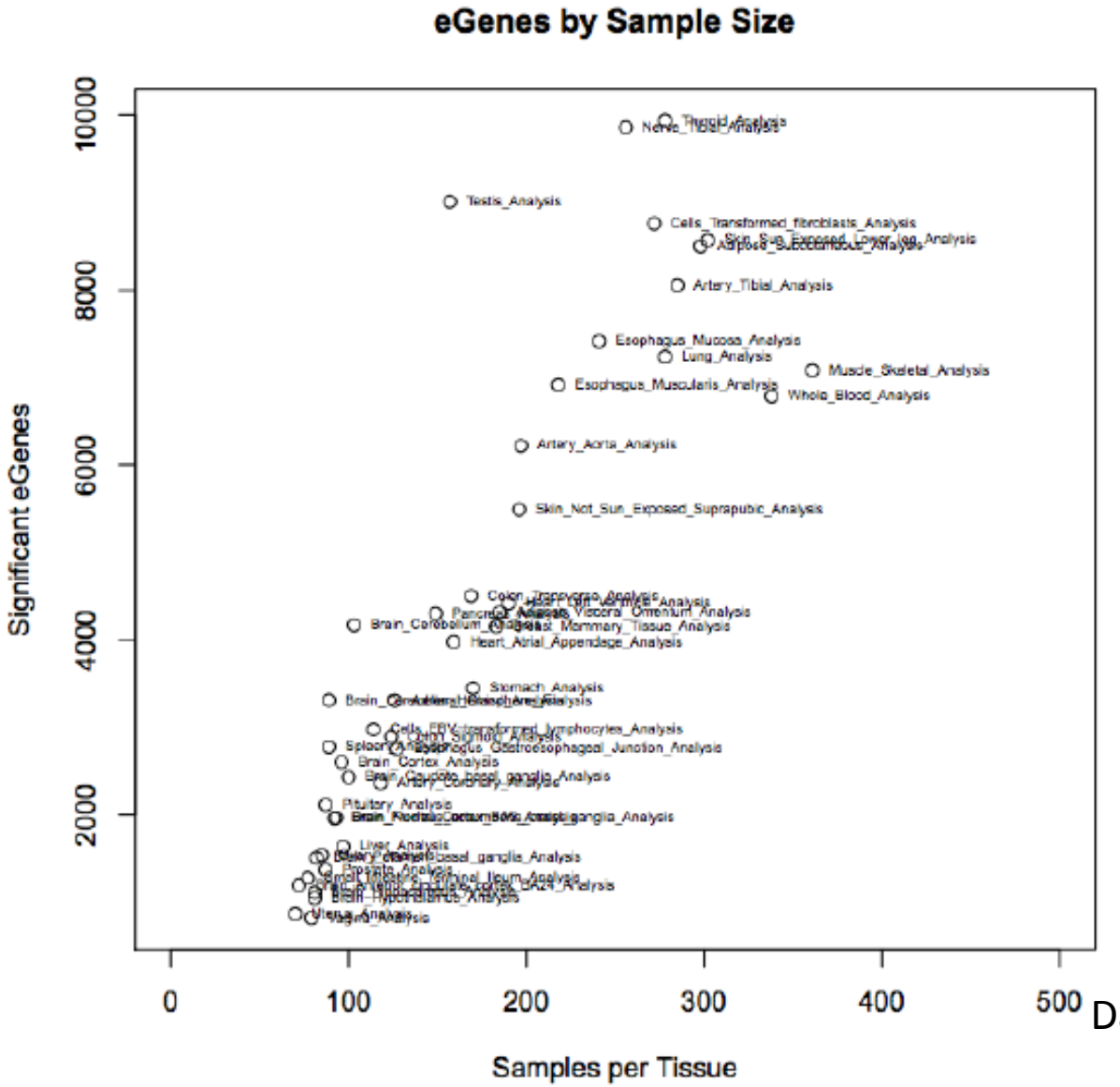
Link eQTLs to GWAS



Link eQTLs to GWAS



GTEx eQTLs in 44 tissues

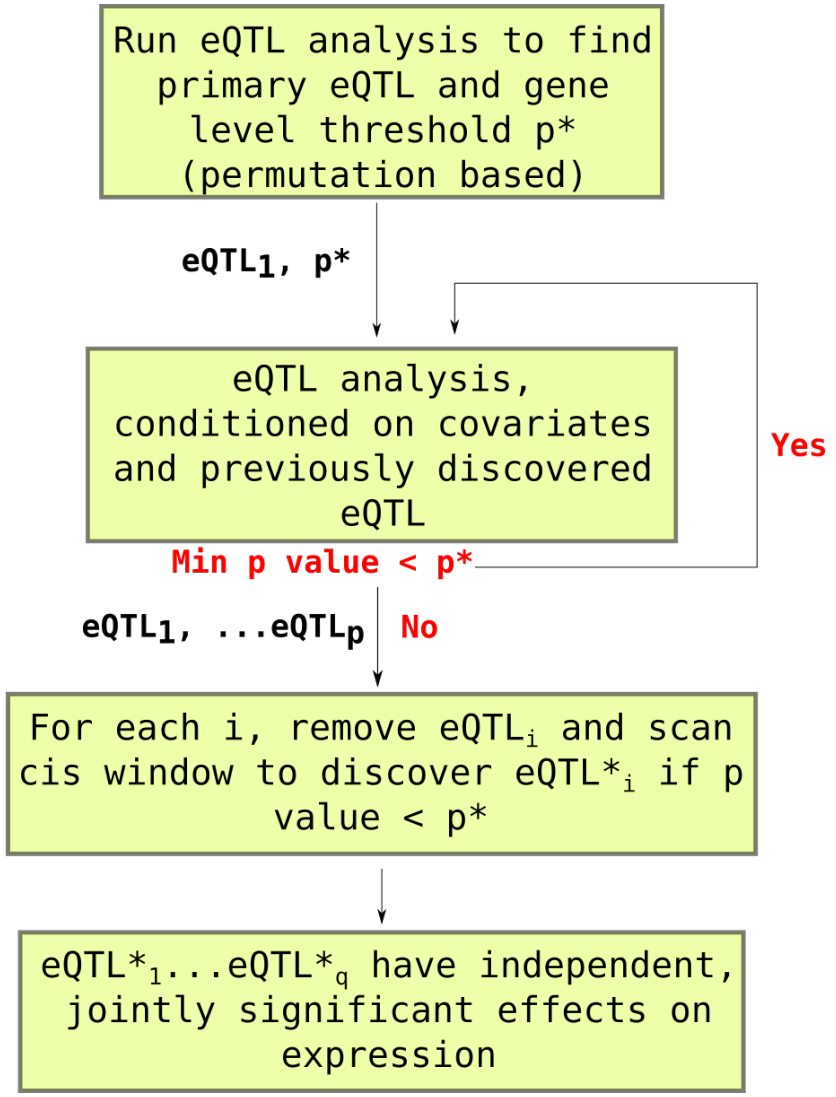


David Deluca

Multiple eQTLs per gene

Forward Stage

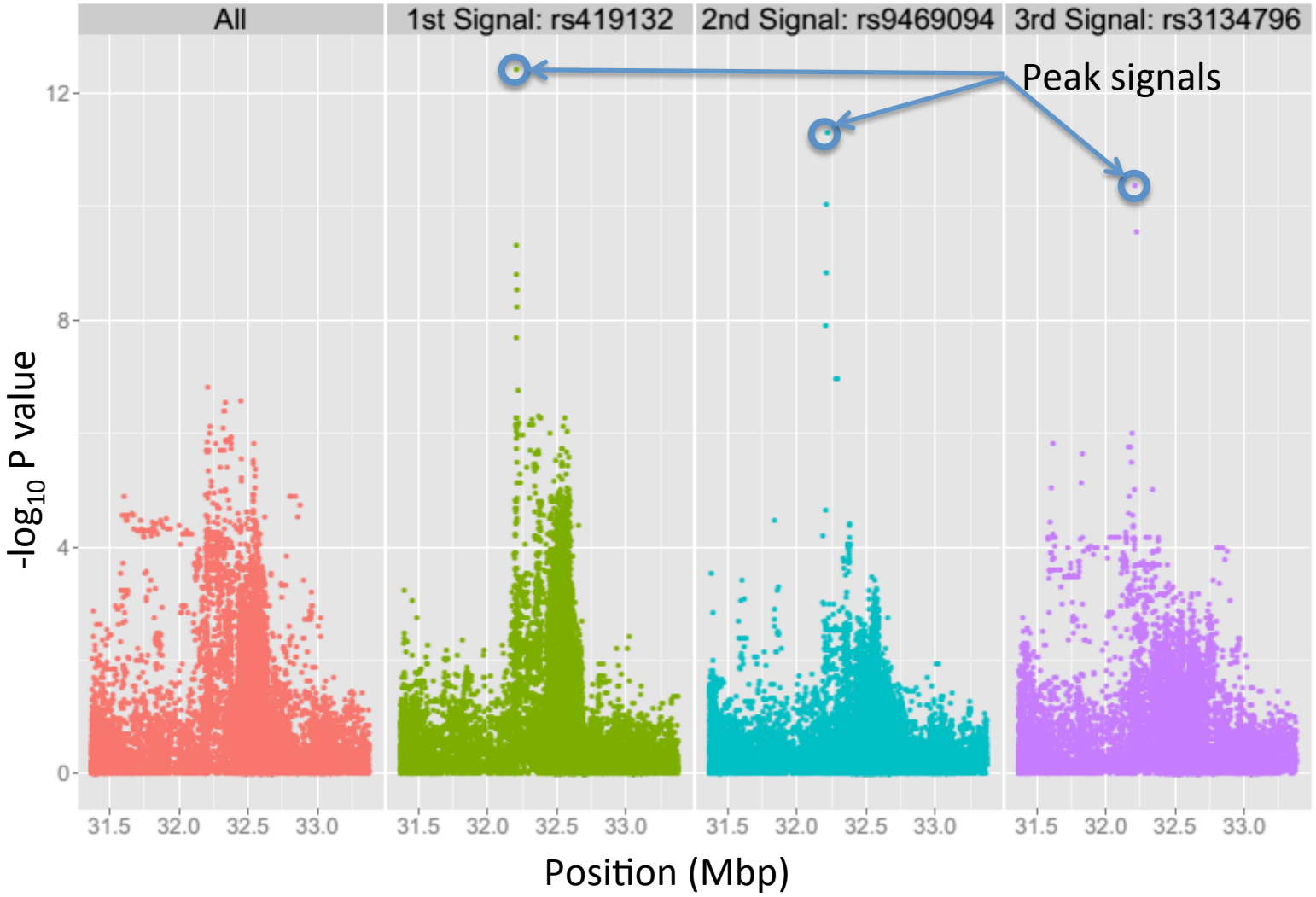
Backward stage



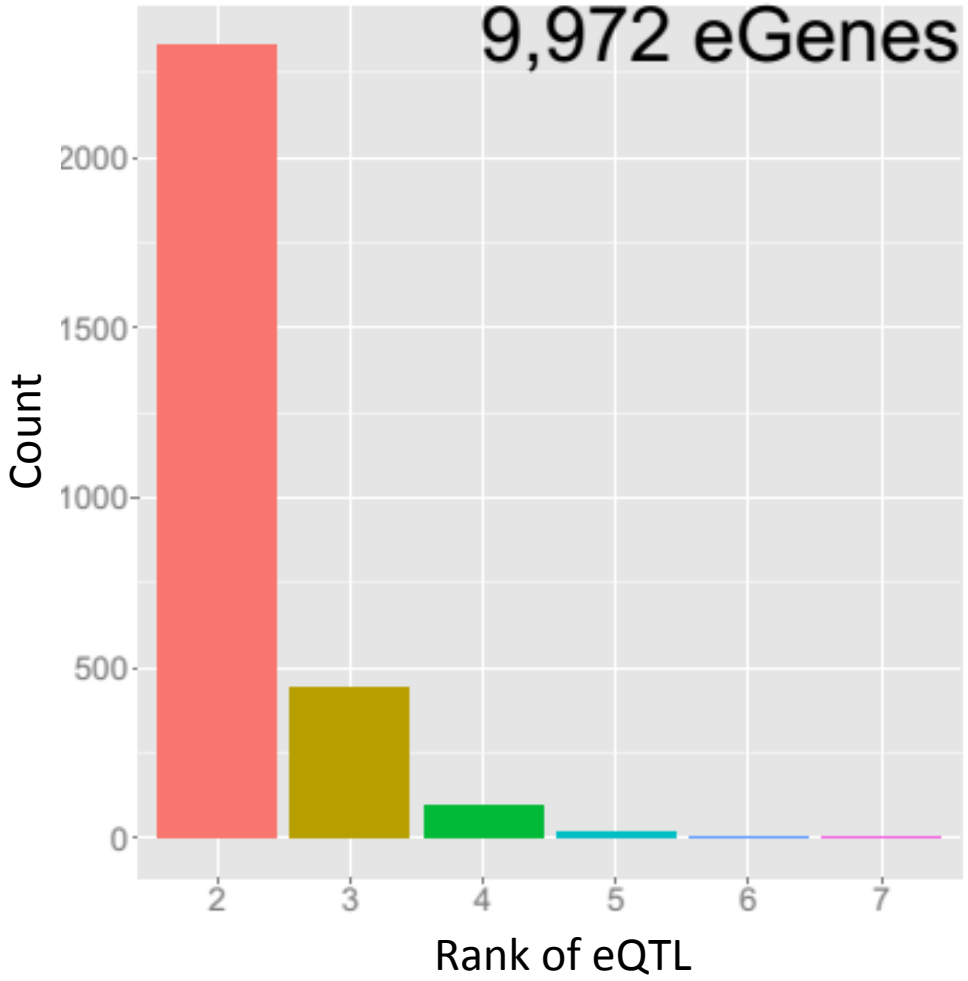
Andrew Brown, Olivier Delaneau

Example

We discovered 3 independent eQTLs for *BTNL2*, in the MHC region

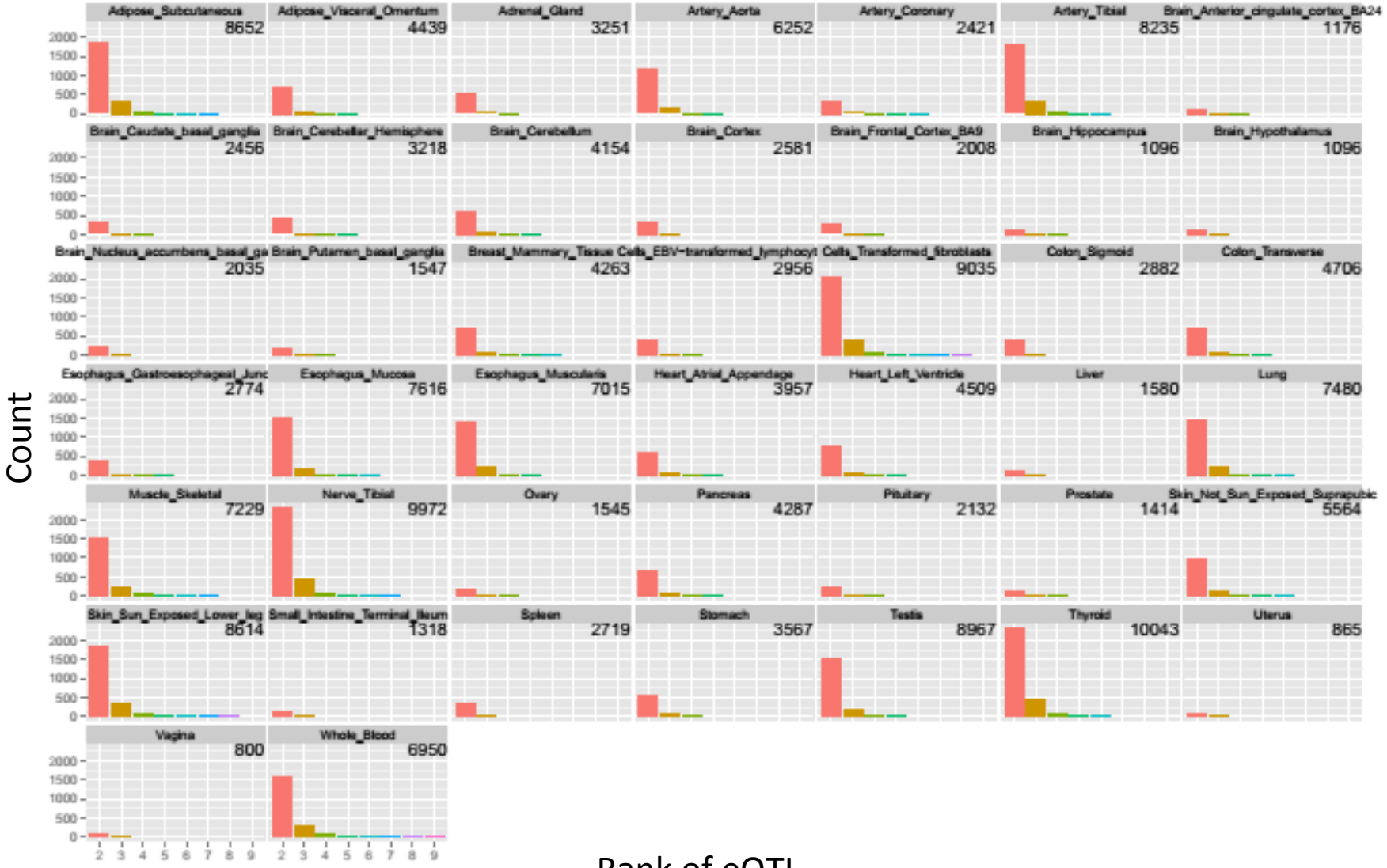


Number of additional eQTLs discovered in Nerve Tibial tissue

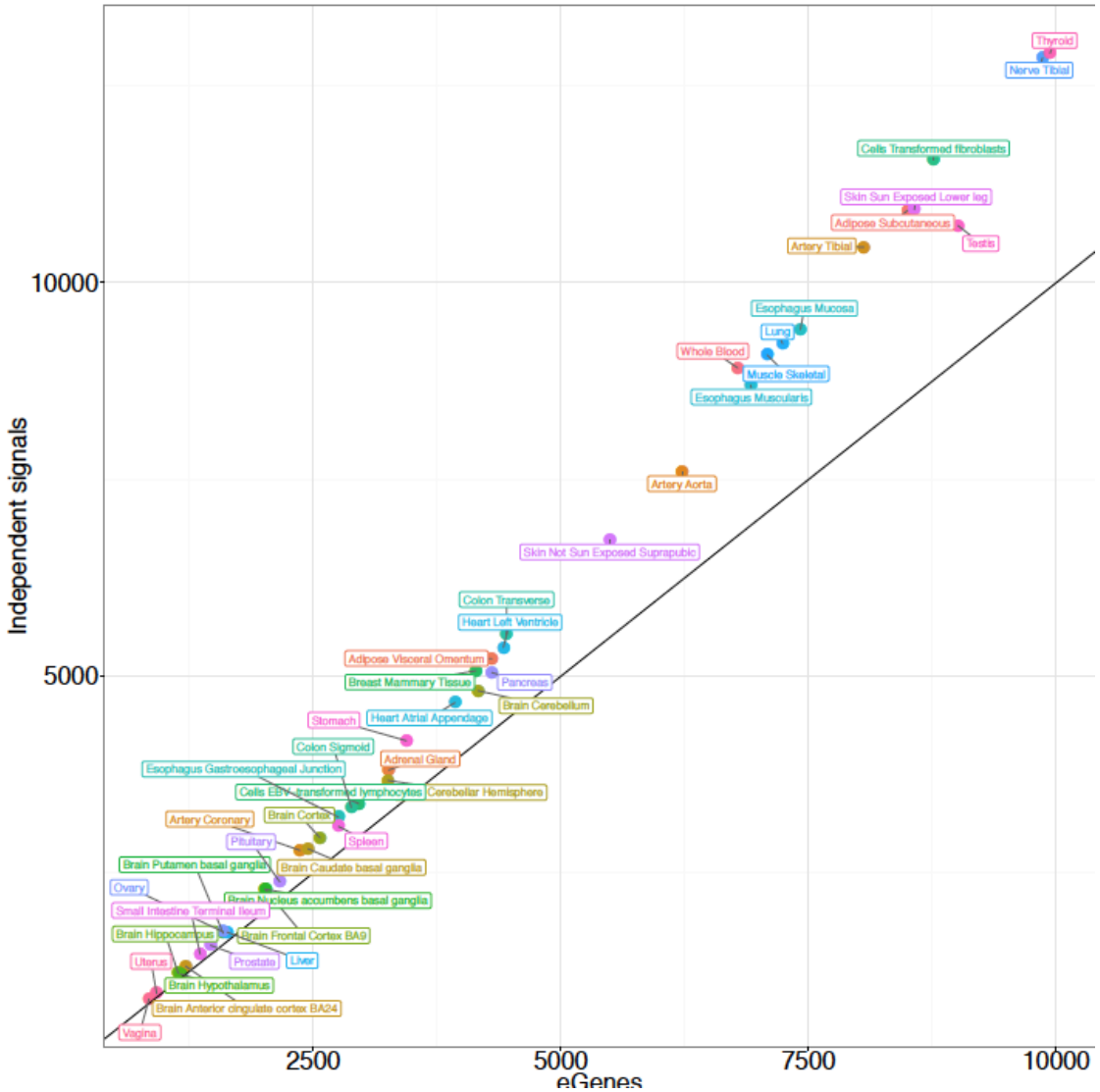


2284 Secondary signals in total

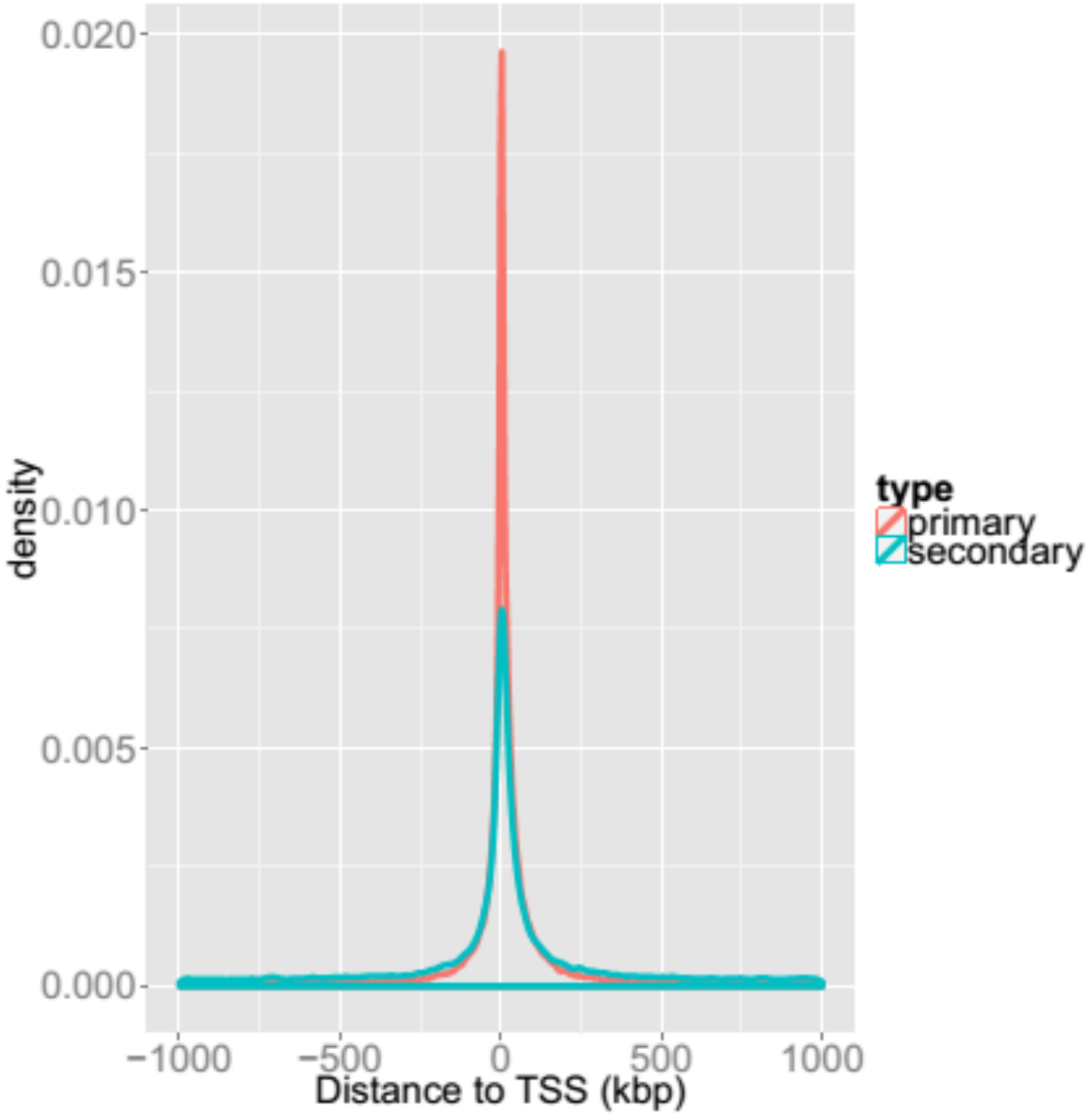
Number of additional eQTLs discovered in all tissues



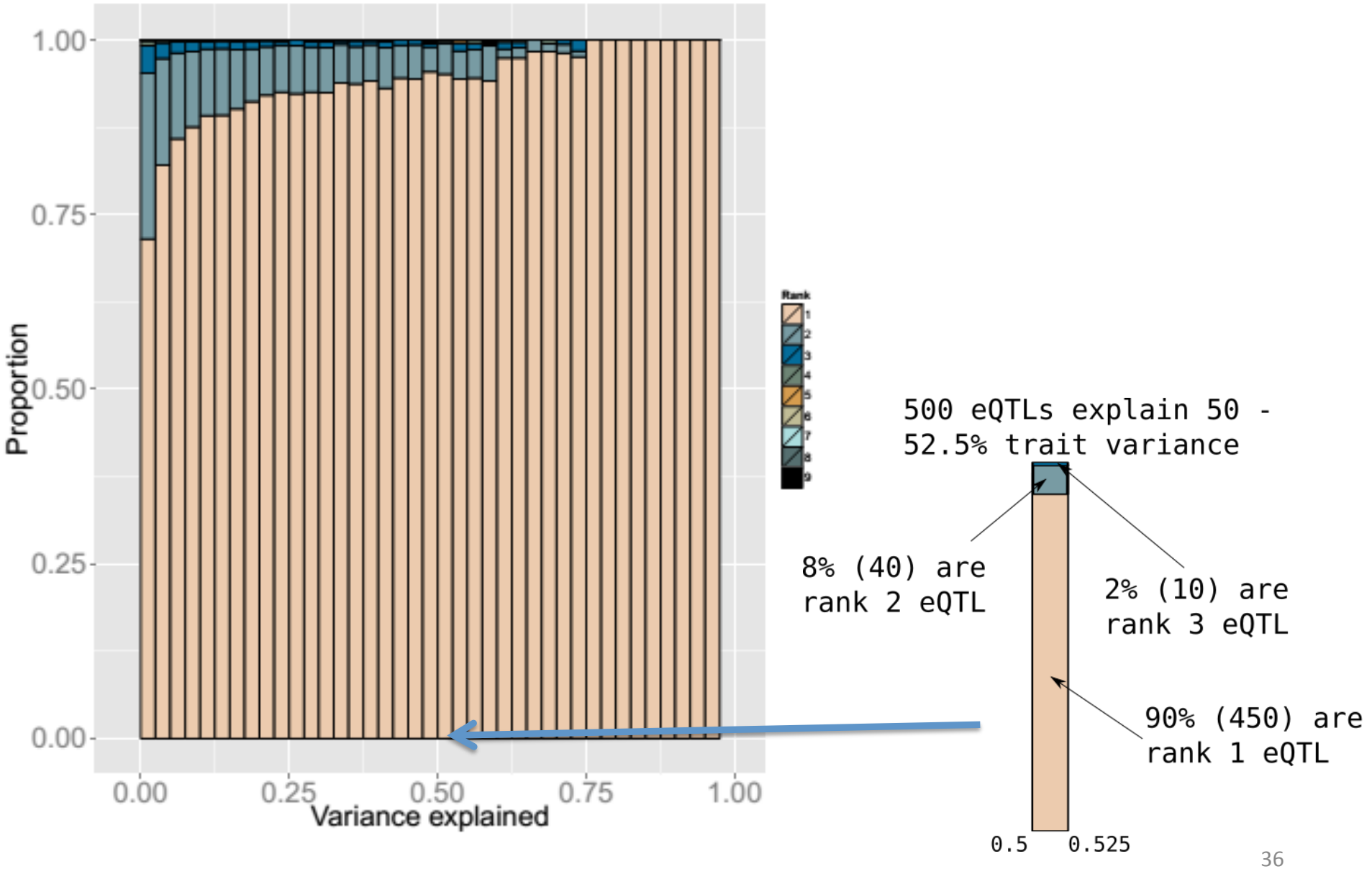
Additional signals discovered is a simple function of power



Secondary eQTL are less enriched around TSS



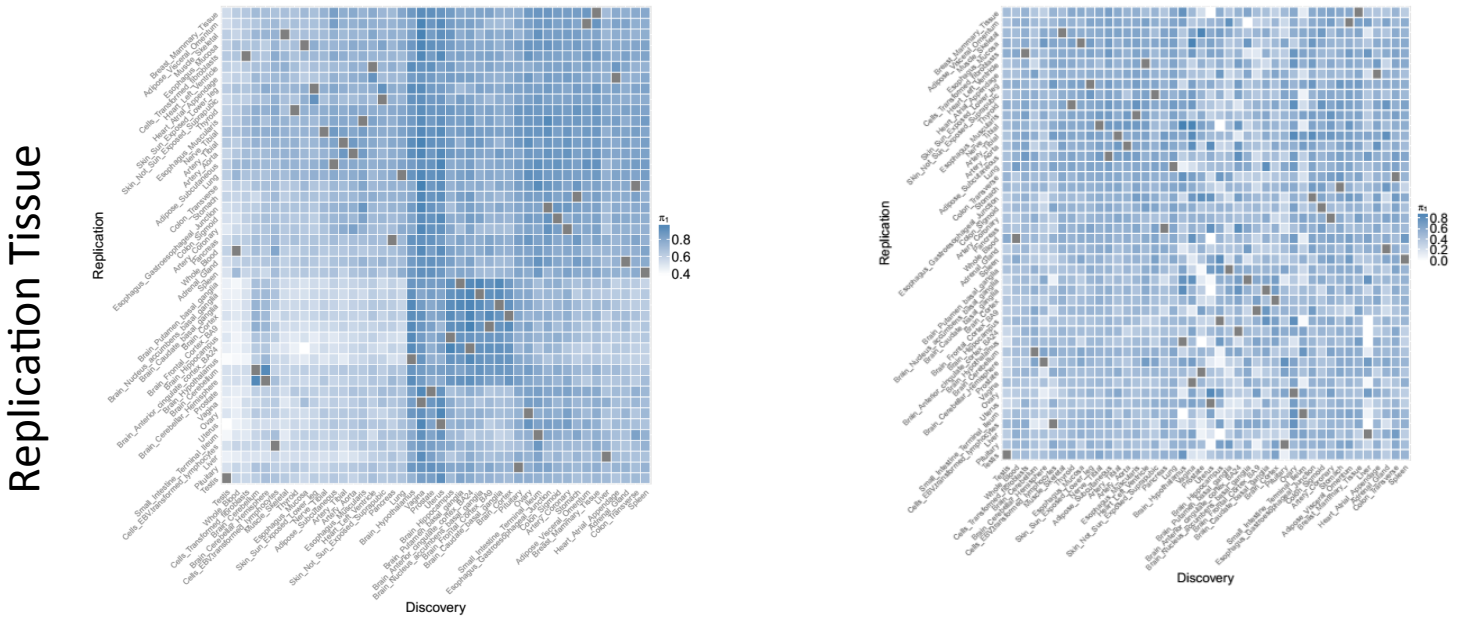
Variance explained by rank of eQTL



Replication in other tissues

Primary

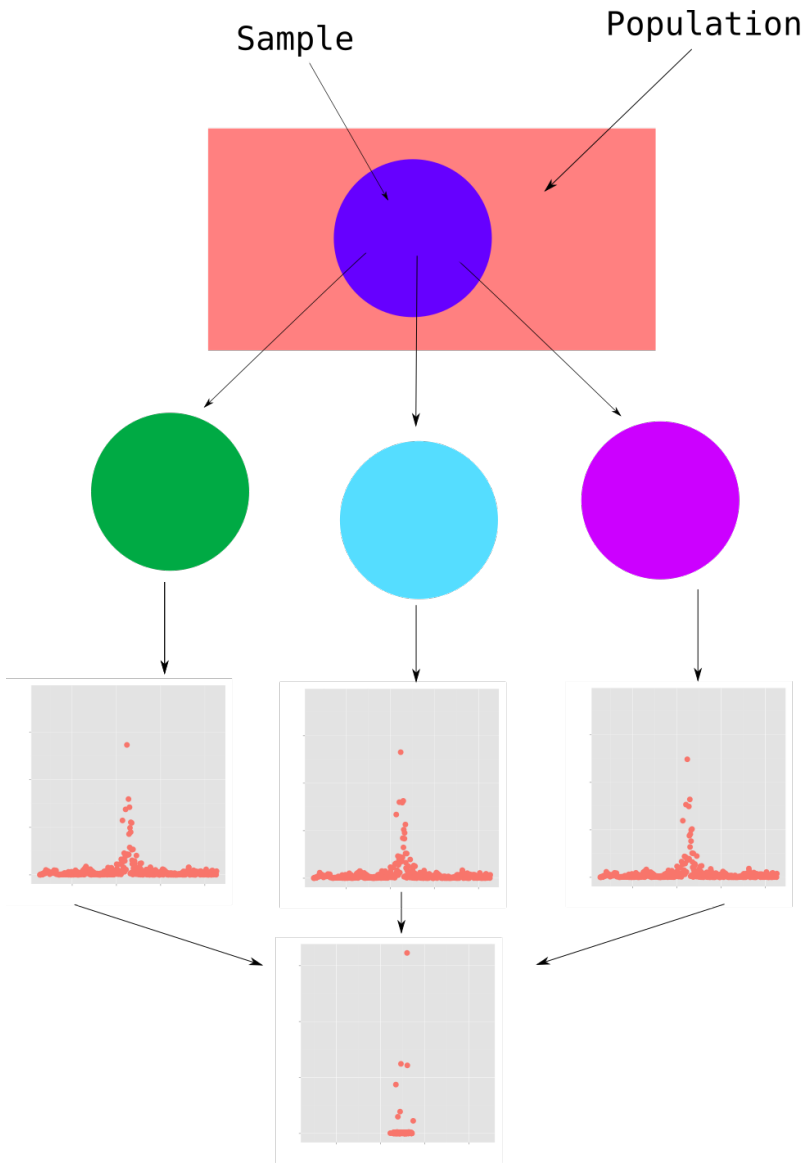
Secondary



Replication Tissue

Discovery Tissue

Mapping causal variants using resampling



Original experiment:
Sample from the population and estimate quantity of interest (Most likely causal variant)

Bootstrap:
Sample with replacement to mimic running the same experiment many times

Rerun analysis

Count ranks of SNPs to estimate likely causal ones

Andrew Brown

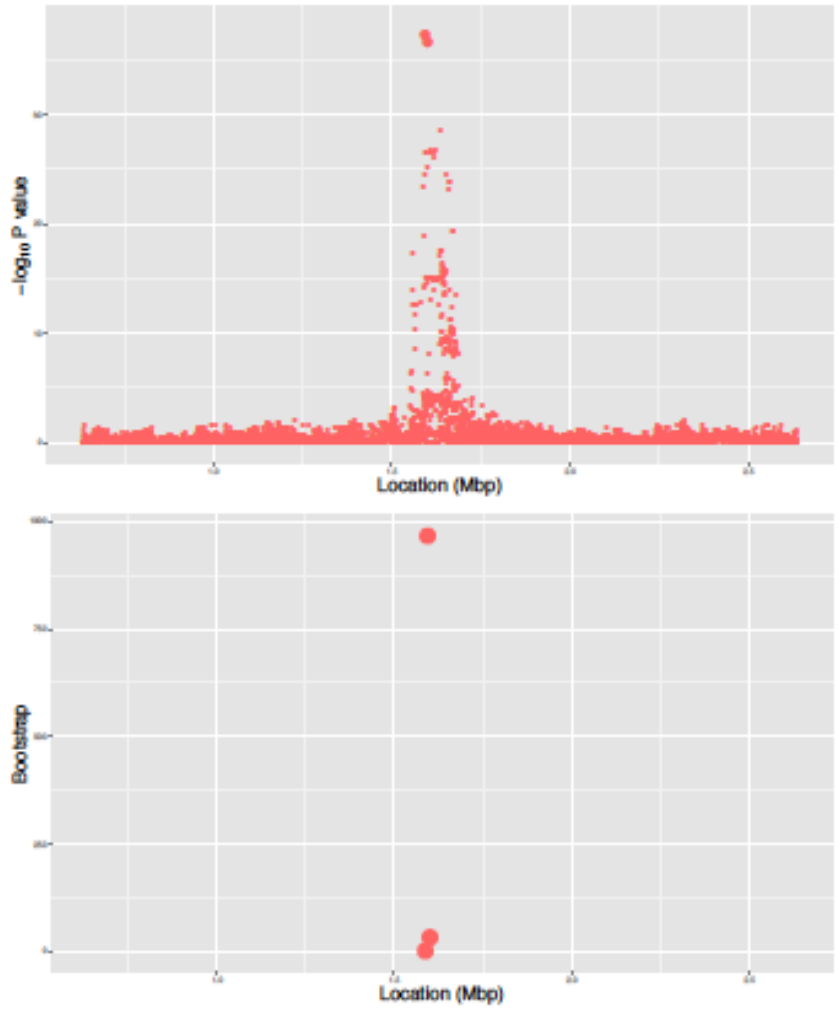
Applied to GTEx Data

- Bootstrap likelihoods are estimated as:

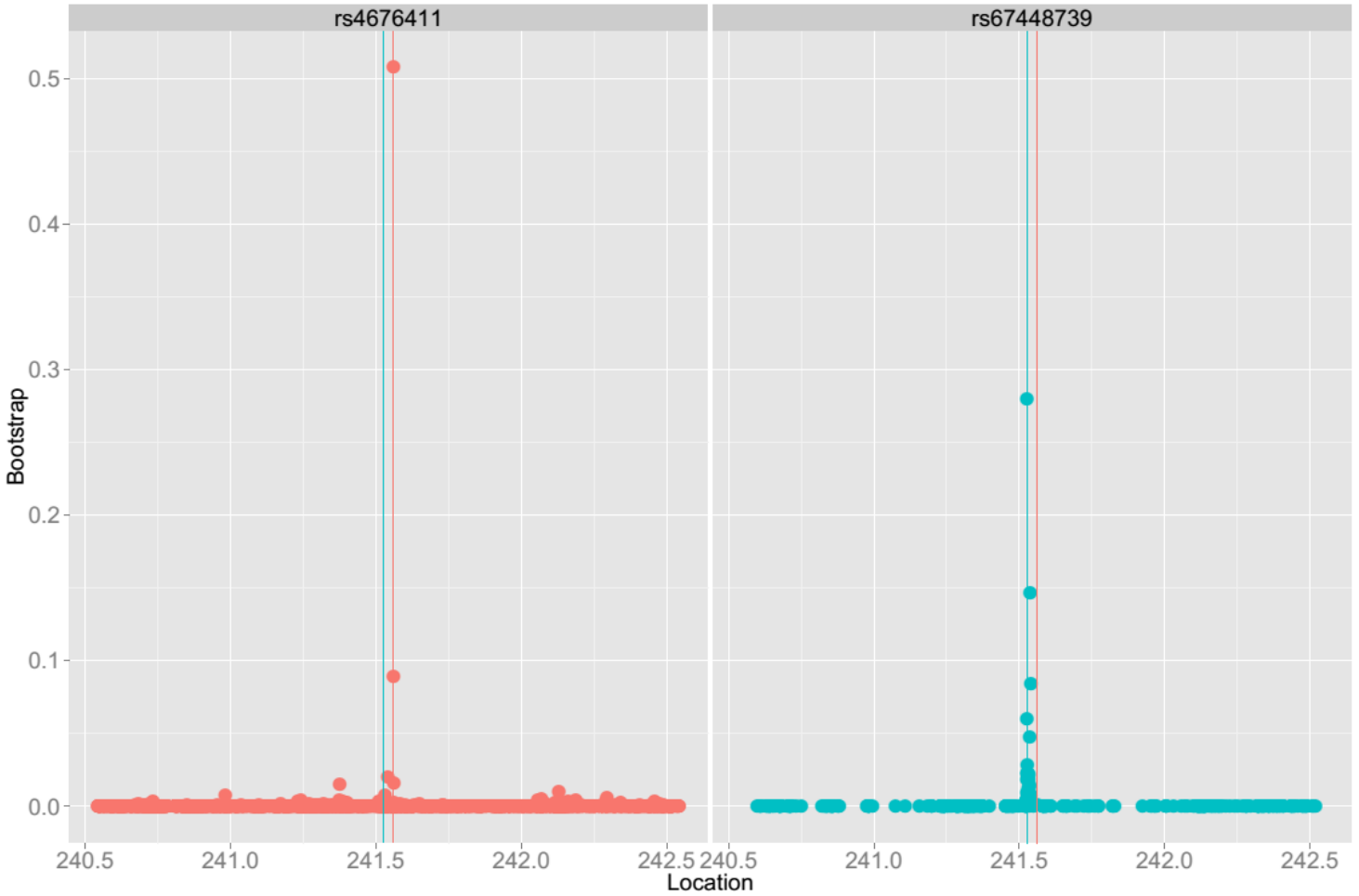
$$P(\text{SNP x is causal}) = \sum_{\text{ranks and bootstraps}} P(\text{SNP x in rank i}) P(\text{rank i is causal})$$

- Applied to 6 GTEx tissues: Brain caudate, Brain cortex, Lung, Muscle skeletal, Nerve Tibial and Testis.
- We use “single signal” phenotypes produced by conditional analysis.
- Also applied “pairwise bootstrap” to Nerve tibial and Muscle skeletal and to Brain caudate and Brain cortex together.

Resolving causal eQTL variants



Example applied to *GPR35* expression



Applying bootstrap to conditional phenotypes we can map causal variants separately

Regulatory Trait Concordance (RTC) Score

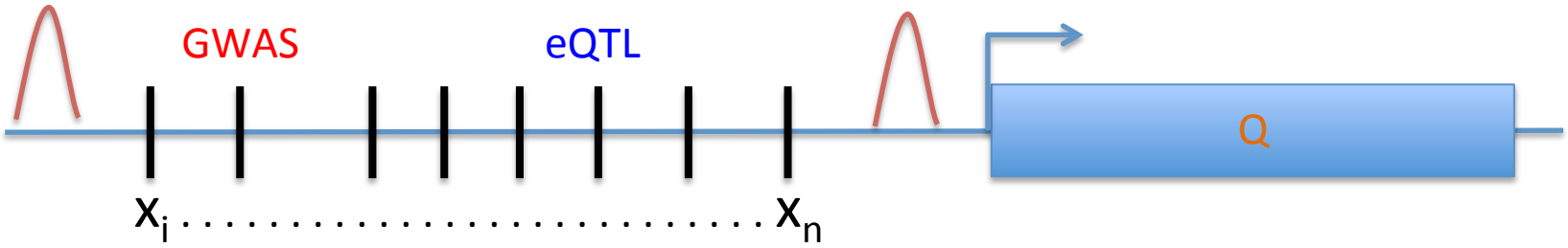
- For a GWAS variant falling into the same recombination coldspot interval with an eQTL, with *N* number of variants in a given coldspot:

  1. Correct the phenotype for each of the variants in the region separately by linear regression, resulting in *N* number of pseudo-phenotypes (residuals).
  2. Redo the eQTL variant association with all of these pseudo-phenotypes.
  3. Sort (descending) the resulting p-values and find the rank of the eQTL to $GWAS_{SNP}$-pseudo-phenotype among all eQTL to pseudo-phenotype associations.
  4. $RTC = Rank_{GWAS_{SNP}} / N$

- RTC scores ≥0.9 signify the eQTL and $GWAS_{SNP}$ are the same casual variant (determined empirically).

Nica A. et.al. PLOS Genetics 2010

Regulatory Trait Concordance (RTC) Score



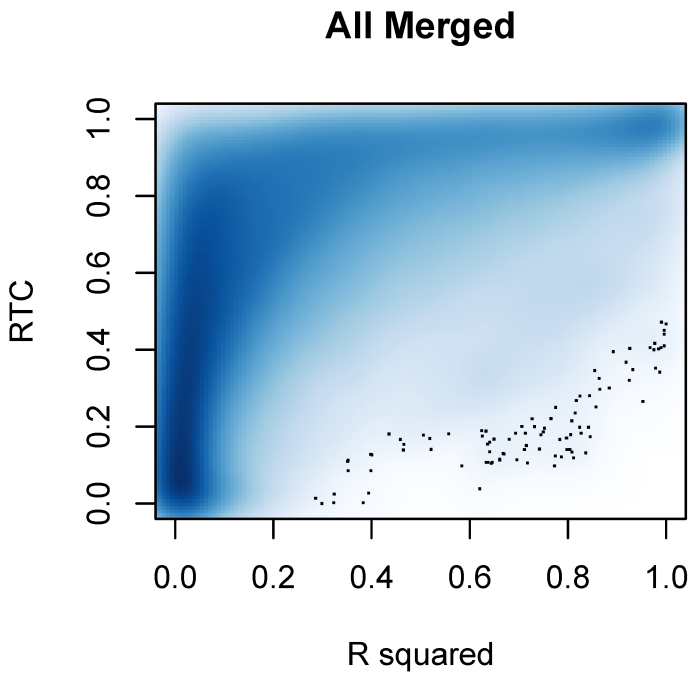
1. For x_i .. x_n:
   - Correct Q for X_i (linear regression) -> pseudo-phenotype X_iQ (residuals)

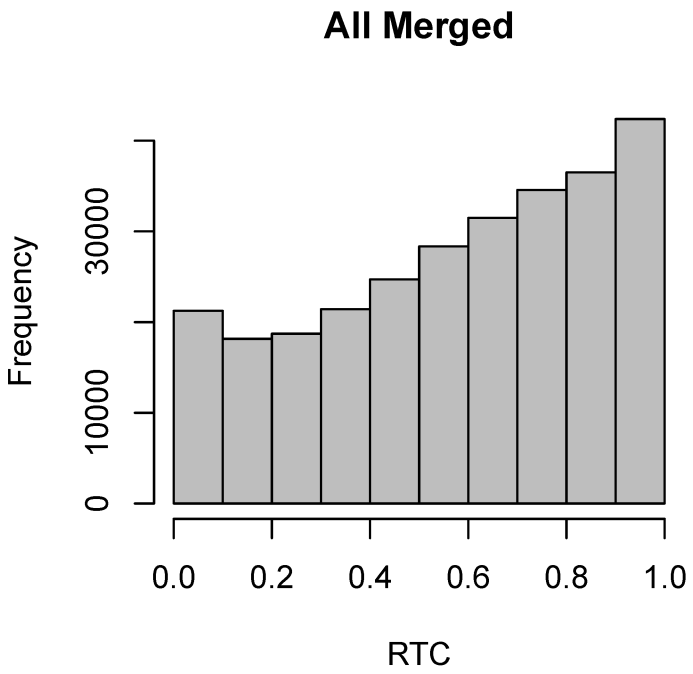
2. For x_i .. x_n:
   - Calculate p-value for eQTL-X_iQ association

3. Sort (ascending) p-values eQTL-$X_{i\ ..\ n}$Q and find the rank of eQTL-x_{GWAS}Q

4. RTC = $Rank_{GWAS}$ / *n*

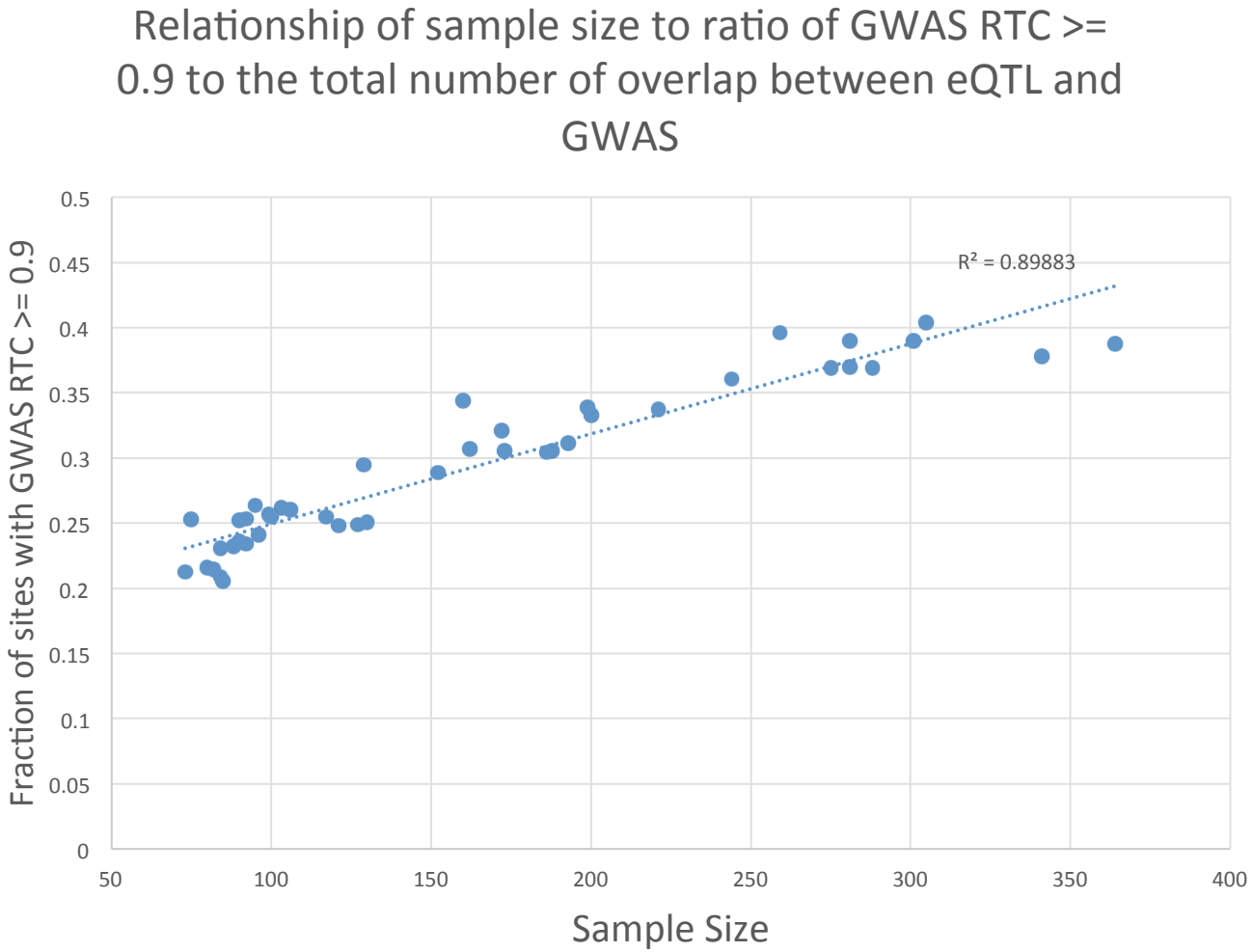
RTC score in the new 44 GTEx tissues



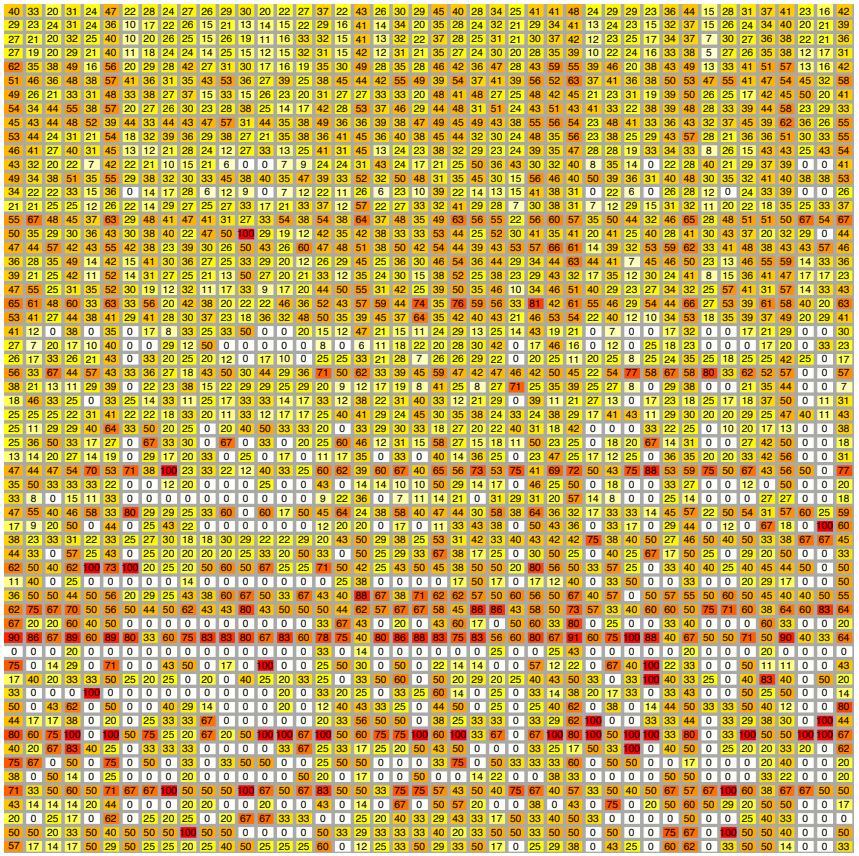


Halit Ongen

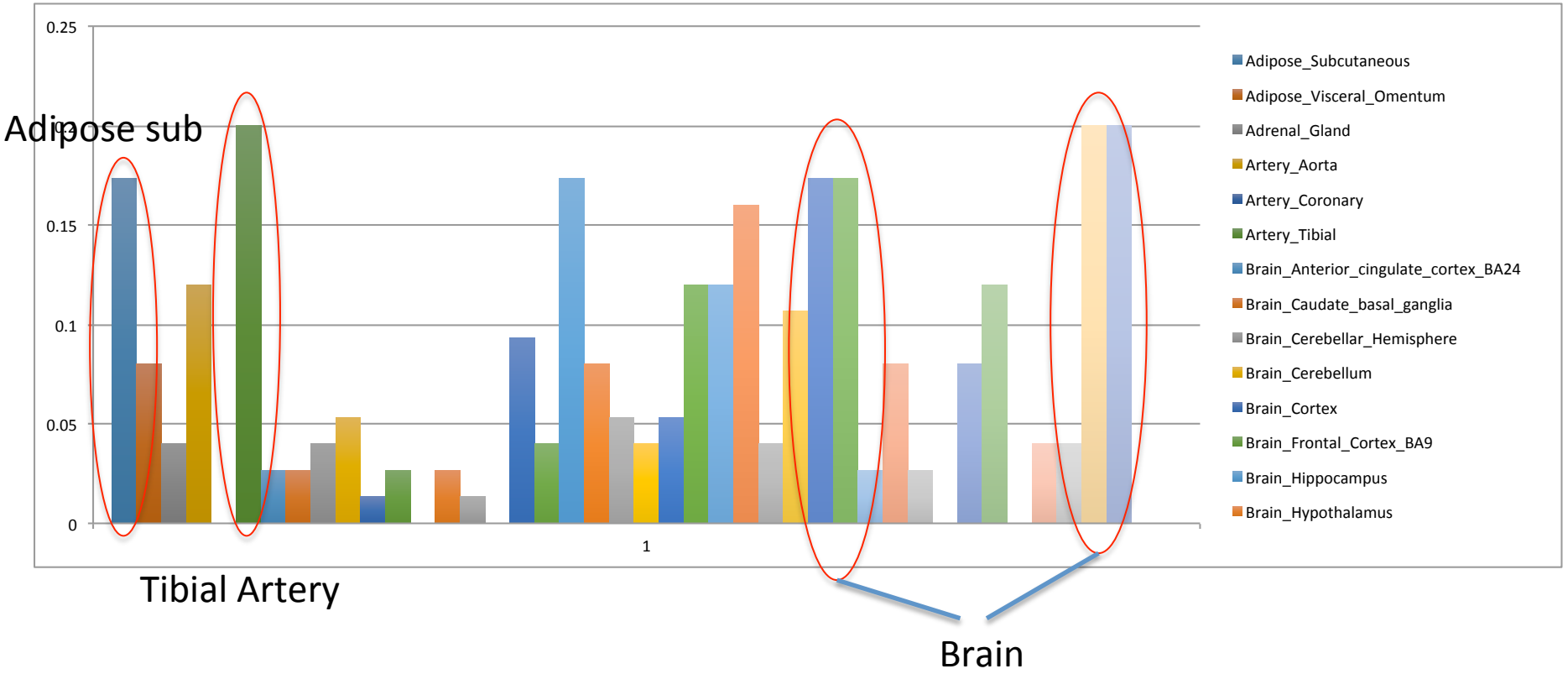
GWAS variants with RTC >= 0.9 by tissue



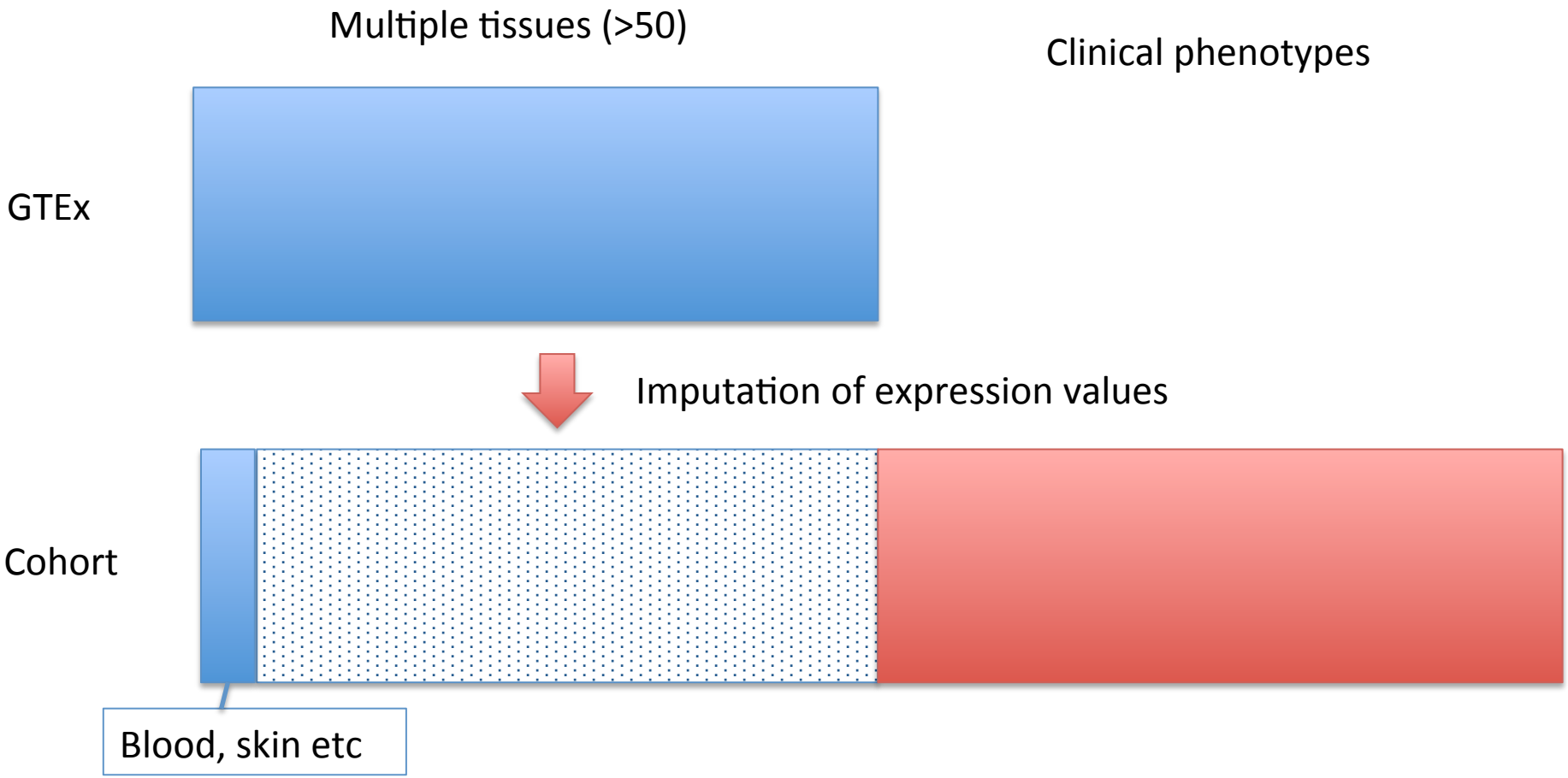
Tissue activity of GWAS in multiple tissues



Type II Diabetes tissue activity



Use of GTEx as reference



Summary

- Unprecedented catalogue of eQTLs in multiple tissues
- eQTL tissue specificity estimates
- First glimpse into effect of genetic variation in diverse tissues/organs of the human body
- GWAS and other disease variant interpretation and causal effects

Acknowledgments

The GTEx Consortium

Analysis Working Group:

LDACC: Kristin G. Ardlie[1], Gad Getz[1,2], David S. Deluca[1], Taylor R. Young[1], Ellen T. Gelfand[1], Ayellet V. Segrè[1], Timothy J. Sullivan[1], Casandra A. Trowbridge[1], Daniel G. MacArthur[1,3], Julian B. Maller[1,3], Taru Tukiainen[1,3], Monkol Lek[1,3], Manolis Kellis[1,4], Lucas D. Ward[1,4], Pouya Kheradpour[1,4], Joel Hirschhorn[1,5], Yan Meng[1], Cameron D. Palmer[1,5].

UNC/NCSU – Andrew B. Nobel[6], Ivan Rusyn[9], Fred A. Wright[8], Gen Li[6], Andrey A. Shablin[7], Yi-Hui Zhou[8].

U Geneva – Emmanouil T. Dermitzakis[10,11,12], Roderic Guigo[21,22,23], Daphne Koller[19], Mark I. McCarthy[16,17,18], Tuuli Lappalainen[10,11,12,13,14,15], Manuel A. Rivas[16], Alexis Battle[19,20], Sara Mostafavi[19], Jean Monlong[21,22,24], Pedro G. Ferreira[10,11,12], Michael Sammeth[21,22,25], Halit Ongen[10,11,12].

U Chicago – Nancy J. Cox[31], Dan L. Nicolae[31], Eric R. Gamazon[31], Anuar Konkashbaev[31].

U Chicago – Jonathan K. Pritchard[26,27,28], Matthew Stephens[26,35], Timothée Flutre[26], Xiaoquan Wen[30].

Harvard – Jun Liu[32], Jun Zhu[33,34], Zhidong Tu[33,34], Bin Zhang[33,34], Tao Huang[33,34], Quan Long[33,34], Luan Lin[33,34], Jialiang Yang[33,34].

Biospecimen and data collection, processing, quality control, storage, and pathological review:

caHUB Biospecimen Source Sites:

NDRI - Amanda Brown[36], Bernadette Mestichelli[36], Denee Tidwell[36], Edmund Lo[36], John T. Lonsdale[36], Jeffrey A. Thomas[36], Mike Salvatore[36], Saboor Shad[36].

Roswell Park - Christopher Choi[37], Barbara A. Foster[37], Michael T. Moser[37], Ellen Karasik[37], Kimberly Ramsey[37], Bryan M. Gillard[37].

Albert Einstein Medical Center – Nancy Young[38].

Gift of Life Donor Program - Rick Hasz[39].

LifeNet Health – Gary D. Walters[40].

UNYTS – Jason P. Bridge[41], Mark Miklos[41], Susan Sullivan[41].

caHUB ELSI Study:

VCU - Laura A. Siminoff[42], Laura K. Barker[42], Heather Traino[42], Magboeba Mosavel[42].

caHUB Comprehensive Biospecimen Resource:

Van Andel – Scott Jewel[43], Daniel C. Rohrer[43], Dana R. Valley[43].

caHUB Pathology Resource Center:

NCI - Philip Branton[44,46].

Leidos Biomedical Research Inc. - Leslie H. Sobin[45].

caHUB Comprehensive Data Resource:

Leidos Biomedical Research Inc. - Charles Shive[45], David Tabor[45], Liqun Qi[45], Pushpa Hariharan[45], Shenpei Wu[45].

caHUB Operations Management:

Leidos Biomedical Research Inc. - Anna M. Smith[45], Stephen A. Buia[45], Anita H. Undale[45], Karna L. Robinson[45], Nancy Roche[45], Kimberly M. Valentino[45], Angela Britton[45], Robin Burges[45], Debra Bradbury[45], Kenneth W. Hambright[45], John Seleski[45], Greg E. Korzeniewski[45].

Sapient Government Services - Kenyon Erickson[47].

Brain Bank Operations:

University of Miami – Deborah C. Mash[48], Margaret Basile[48], Yvonne Marcus[48], Jorge Tejada[48], Mehran Taherian[48], Chunrong Lu[48], Barnaby E. Robles[48].

Program Management:

NCI – Latarsha J. Carithers[46,49], Helen M. Moore[46,49], Ping Guan[44,46], Carolyn Compton[46,49], Sherilyn J. Sawyer[44,46], Joanne Peter Demchok[44,46], Jimmie B. Vaught[44,46], Chana A. Rabiner[44,46], Nicole C. Lockhart[44,46].

NHGRI – Simona Volpi[46,50], Jeff Struewing[46,50], Gary F. Temple[46,50], Joy Boyer[46,51], Deborah Colantuoni[46,50].

NIMH – Roger Little[46,52], Susan Koester[46,53].

NIH OD – Leslie K. Derr[46,54].

Writing Committee: Kristin G. Ardlie[1], Gad Getz[1,2], Fred A. Wright[8], Manolis Kellis[1,4], Simona Volpi[46,50], Emmanouil T. Dermitzakis[10,11,12].