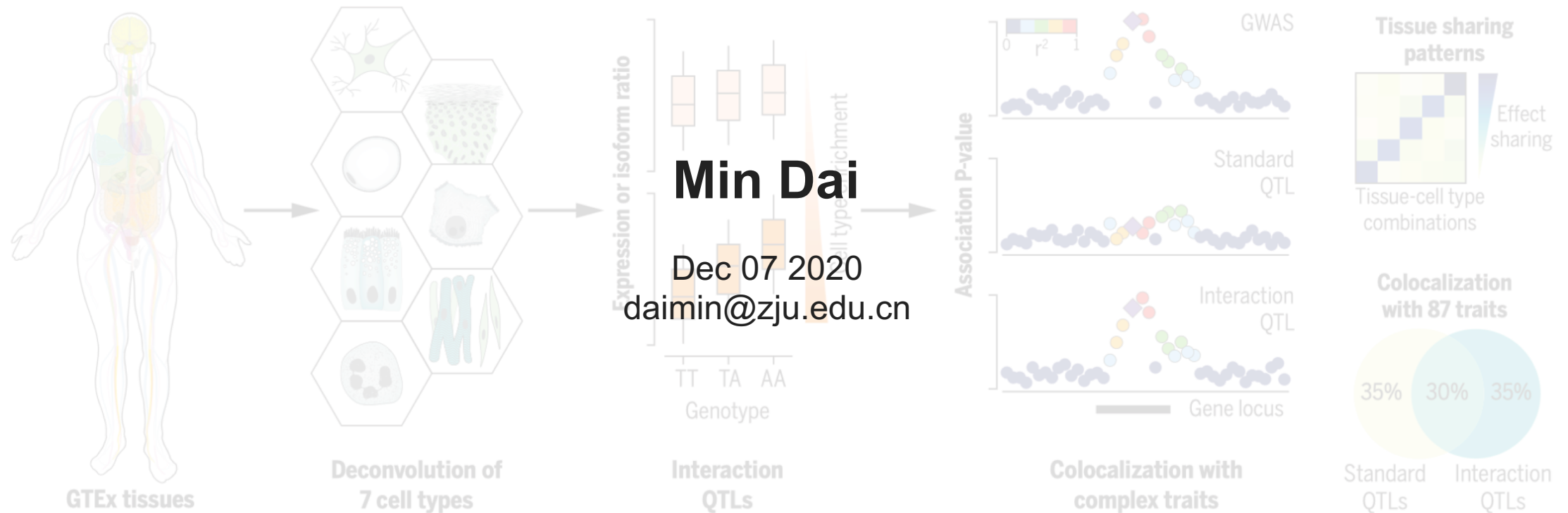


Cell type-specific genetic regulation of gene expression across human tissues



Outline

- Background
- Results
 - Identifying cell types *in silico* in bulk tissue
 - Mapping cell type–interaction eQTLs and sQTLs
 - Cell type ieQTLs contribute to tissue specificity
 - GWAS and tissue-specific eQTLs and sQTLs
- Summary

SHARE

RESEARCH ARTICLE



Cell type-specific genetic regulation of gene expression across human tissues

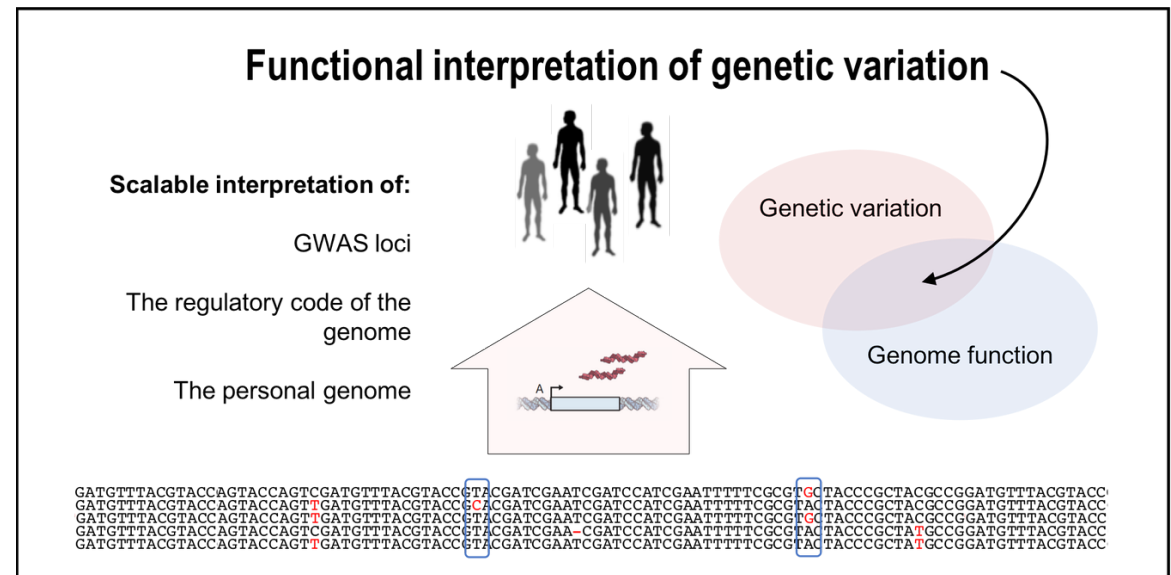
Sarah Kim-Hellmuth^{1,2,3,*†‡}, François Aguet^{4,†}, Meritxell Oliva^{5,6}, Manuel Muñoz-Aguirre^{7,8}, Silva Kasela^{2,3}, Valentin Wucher⁷, Stephane E. Castel^{2,3}, Andrew R. Hamel^{4,9}, Ana Viñuela^{10,11,12,13}, Amy L. Roberts¹⁰, Serghei Mangul^{14,15}, Xiaoquan Wen¹⁶, Gao Wang¹⁷, Alvaro N. Barbeira⁵, Diego Garrido-Martín⁷, Brian B. Nadel¹⁸, Yuxin Zou¹⁹, Rodrigo Bonazzola⁵, Jie Quan²⁰, Andrew Brown^{11,21}, Angel Martinez-Perez²², José Manuel Soria²², GTEx Consortium§, Gad Getz^{4,23,24}, Emmanouil T. Dermitzakis^{11,12,13}, Kerrin S. Small¹⁰, Matthew Stephens¹⁷, Hualin S. Xi²⁵, Hae Kyung Im⁵, Roderic Guigó^{7,26}, Ayellet V. Segrè^{4,9}, Barbara E. Stranger^{5,27}, Kristin G. Ardlie⁴, Tuuli Lappalainen^{2,3,*}

Kim-Hellmuth, S. *et al.* *Science* **369**, (2020).

Assoc. Prof. Tuuli Lappalainen (Corresponding author)



- Focuses on functional genetic variation in human populations and its contribution to traits and diseases
- A core faculty member at the New York Genome Center, and an Associate Professor with tenure in Columbia University



RESEARCH ARTICLE

HUMAN GENOMICS

Cell type-specific genetic regulation of gene expression across human tissues

Sarah Kim-Hellmuth^{1,2,3*}†‡, **François Aguet**⁴†, Meritxell Oliva^{5,6}, Manuel Muñoz-Aguirre^{7,8},
Silva Kasela^{2,3}, Valentin Wucher⁷, Stephane E. Castel^{2,3}, Andrew R. Hamel^{4,9}, Ana Viñuela^{10,11,12,13},
Amy L. Roberts¹⁰, Serghei Mangul^{14,15}, Xiaoquan Wen¹⁶, Gao Wang¹⁷, Alvaro N. Barbeira⁵,
Diego Garrido-Martín⁷, Brian B. Nadel¹⁸, Yuxin Zou¹⁹, Rodrigo Bonazzola⁵, Jie Quan²⁰,
Andrew Brown^{11,21}, Angel Martinez-Perez²², José Manuel Soria²², GTEx Consortium§, Gad Getz^{4,23,24},
Emmanouil T. Dermitzakis^{11,12,13}, Kerrin S. Small¹⁰, Matthew Stephens¹⁷, Hualin S. Xi²⁵,
Hae Kyung Im⁵, Roderic Guigó^{7,26}, Ayellet V. Segre^{4,9}, Barbara E. Stranger^{5,27},
Kristin G. Ardlie⁴, **Tuuli Lappalainen**^{2,3*}

Teamwork from GTEx Consortium

Authors

Lead Analysts†: François Aguet^{1‡}, Alvaro N. Barbeira², Rodrigo Bonazzola², Andrew Brown^{3,4}, Stephane E. Castel^{5,6}, Brian Jo^{7,8}, Silva Kasela^{5,6}, Sarah Kim-Hellmuth^{5,6,9}, Yanyu Liang², Meritxell Oliva^{2,10}, Princy Parsana¹¹

Analysts†: Elise D. Flynn^{5,6}, Laure Fresard¹², Eric R. Gamazon^{13,14,15,16}, Andrew R. Hamel^{17,1}, Yuan He¹⁸, Farhad Hormozdiari^{19,1}, Pejman Mohammadi^{5,6,20,21}, Manuel Muñoz-Aguirre^{22,23}, YoSon Park^{24,25}, Ashis Saha¹¹, Ayellet V. Segre^{1,17}, Benjamin J. Strober¹⁸, Xiaoquan Wen²⁶, Valentin Wucher²²

Manuscript Working Group†: François Aguet¹, Kristin G. Ardlie¹, Alvaro N. Barbeira², Alexis Battle^{18,11}, Rodrigo Bonazzola², Andrew Brown^{3,4}, Christopher D. Brown²⁴, Stephane E. Castel^{5,6}, Nancy Cox¹⁶, Sayantan Das²⁶, Emmanouil T. Dermitzakis^{3,27,28}, Barbara E. Engelhardt^{7,8}, Elise D. Flynn^{5,6}, Laure Fresard¹², Eric R. Gamazon^{13,14,15,16}, Diego Garrido-Martín²², Nicole R. Gay²⁹, Gad A. Getz^{1,30,31}, Roderic Guigó^{22,32}, Andrew R. Hamel^{17,1}, Robert E. Handsaker^{33,33,35}, Yuan He¹⁸, Paul J. Hoffman⁵, Farhad Hormozdiari^{19,1}, Hae Kyung Im², Brian Jo^{7,8}, Silva Kasela^{5,6}, Seva Kashin^{33,34,35}, Sarah Kim-Hellmuth^{5,6,9}, Alan Kwong²⁶, Tuuli Lappalainen^{5,6}, Xiao Li¹, Yanyu Liang², Daniel G. MacArthur^{34,36}, Pejman Mohammadi^{5,6,20,21}, Stephen B. Montgomery^{12,29}, Manuel Muñoz-Aguirre^{22,23}, Meritxell Oliva^{2,10}, YoSon Park^{24,25}, Princy Parsana¹¹, John M. Rouhana^{17,1}, Ashis Saha¹¹, Ayellet V. Segre^{1,17}, Matthew Stephens³⁷, Barbara E. Stranger^{2,38}, Benjamin J. Strober¹⁸, Ellen Todres¹, Ana Viñuela^{39,3,27,28}, Gao Wang³⁷, Xiaoquan Wen²⁶, Valentin Wucher²², Yuxin Zou⁴⁰

Analysis Team Leaders†: François Aguet¹, Alexis Battle^{18,11}, Andrew Brown^{3,4}, Stephane E. Castel^{5,6}, Barbara E. Engelhardt^{7,8}, Farhad Hormozdiari^{19,1}, Hae Kyung Im², Sarah Kim-Hellmuth^{5,6,9}, Meritxell Oliva^{2,10}, Barbara E. Stranger^{2,38}, Xiaoquan Wen²⁶

Senior Leadership†: Kristin G. Ardlie¹, Alexis Battle^{18,11}, Christopher D. Brown²⁴, Nancy Cox¹⁶, Emmanouil T. Dermitzakis^{3,27,28}, Barbara E. Engelhardt^{7,8}, Gad A. Getz^{1,30,31}, Roderic Guigó^{22,33}, Hae Kyung Im², Tuuli Lappalainen^{5,6}, Stephen B. Montgomery^{12,29}, Barbara E. Stranger^{2,38}

Manuscript Writing Group: François Aguet¹, Hae Kyung Im², Alexis Battle^{18,11}, Kristin G. Ardlie¹, Tuuli Lappalainen^{5,6}

†Alphabetical order

‡First author

Scientific question

- **Existing problem:** the functional characterization of QTLs has been limited by the heterogeneous cellular composition of GTEx tissue samples
- **Question:** how to uncover the cellular specificity of regulatory effects underlying complex traits?

Background

Background

- The GTEx projects study the genetic regulation of the transcriptome
- The GTEx Consortium has built an extensive catalog of expression and splicing quantitative trait loci in cis (cis-eQTLs and cis-sQTLs, respectively) across a large range of tissues
- There exist many *in silico* cell type deconvolution methods for estimating cell type abundance in bulk RNA sequencing data
- The single-cell RNA sequencing methods are not yet scalable to sample sizes and coverage sufficient to achieve power comparable with that of bulk eQTL studies

Identifying cell types *in silico* in bulk tissue

METHOD

Open Access

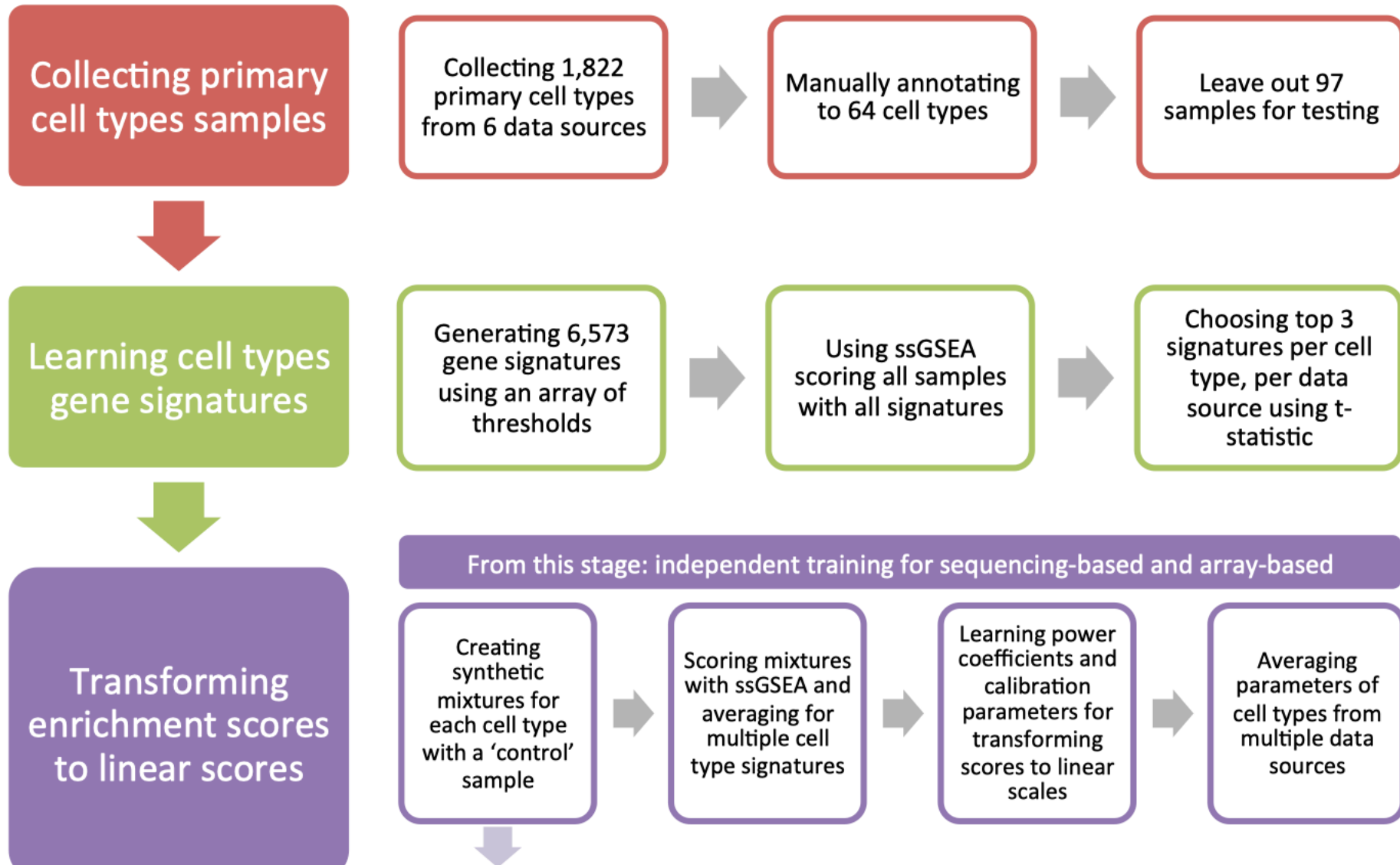
xCell: digitally portraying the tissue cellular heterogeneity landscape



Dvir Aran^{*}, Zicheng Hu and Atul J. Butte^{*} 

The xCell development pipeline

- Identify gene signatures for each cell type and then calculate enrichment scores for bulk datasets



Detection of cell type–specific effects on gene expression

- **One sentence summary:** The enrichment of seven cell types is calculated across GTEx tissues, enabling mapping of cell type–interaction QTLs for expression and splicing by testing for significant interactions between genotypes and cell type enrichments, and linking these QTLs to complex trait associations enables discovery of >50% more colocalizations compared with standard QTLs and reveals the cellular specificity of traits.

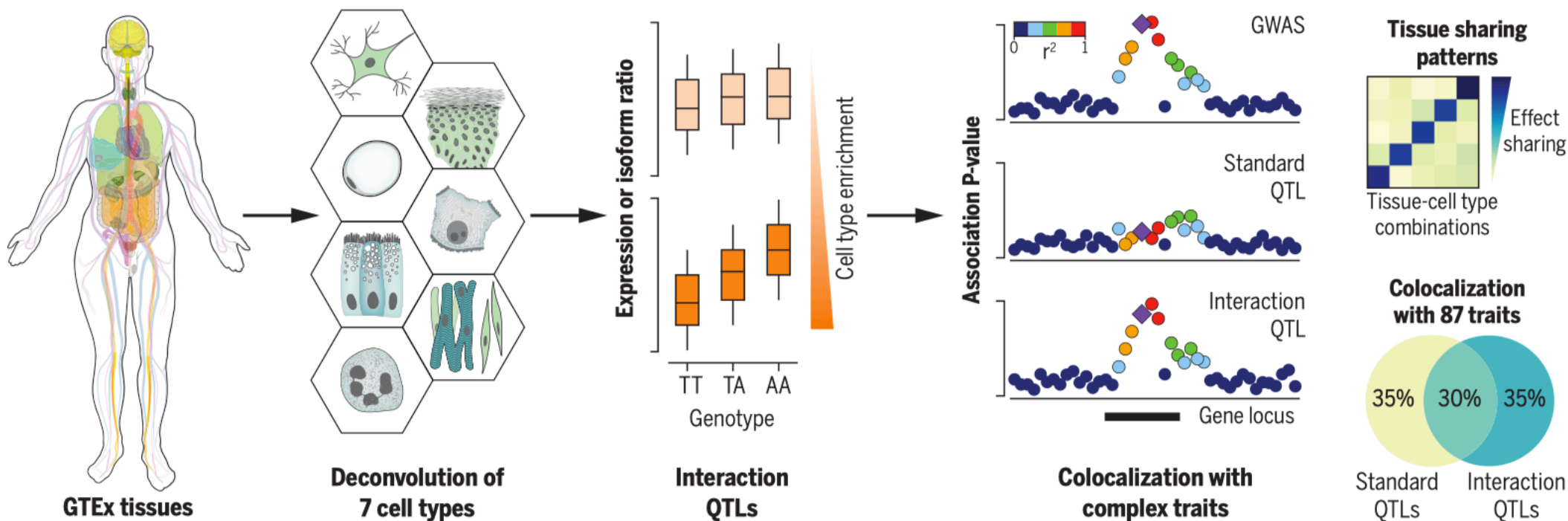
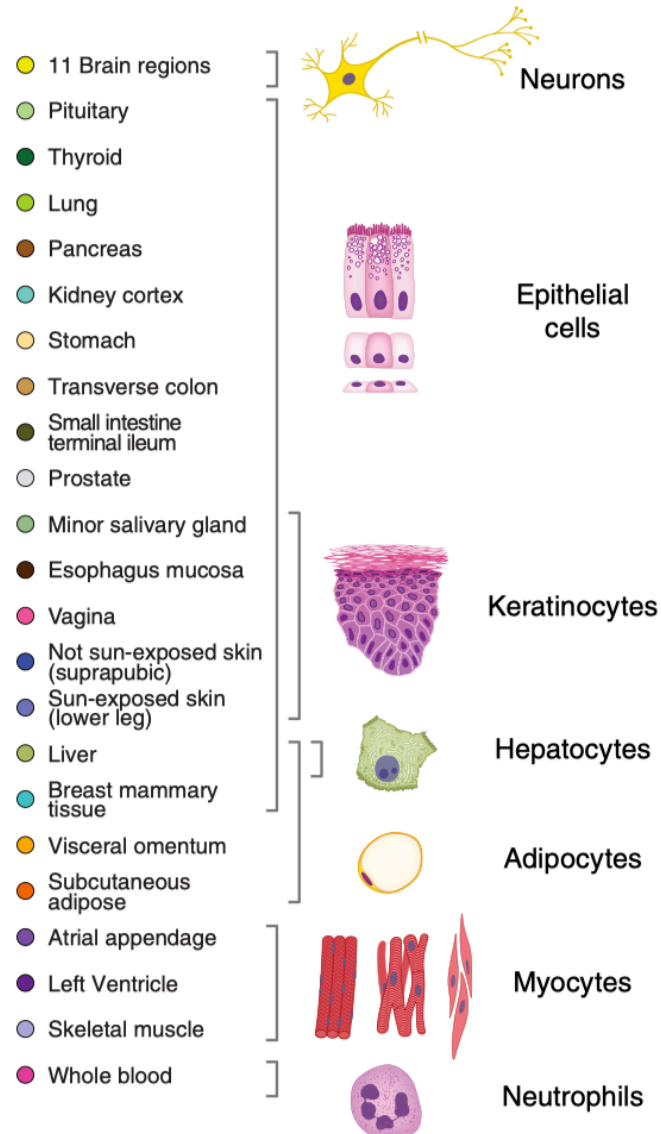


Illustration of 43 cell type–tissue pairs included in this work

- Cell types with median xCell enrichment score >0.1 within a tissue were used



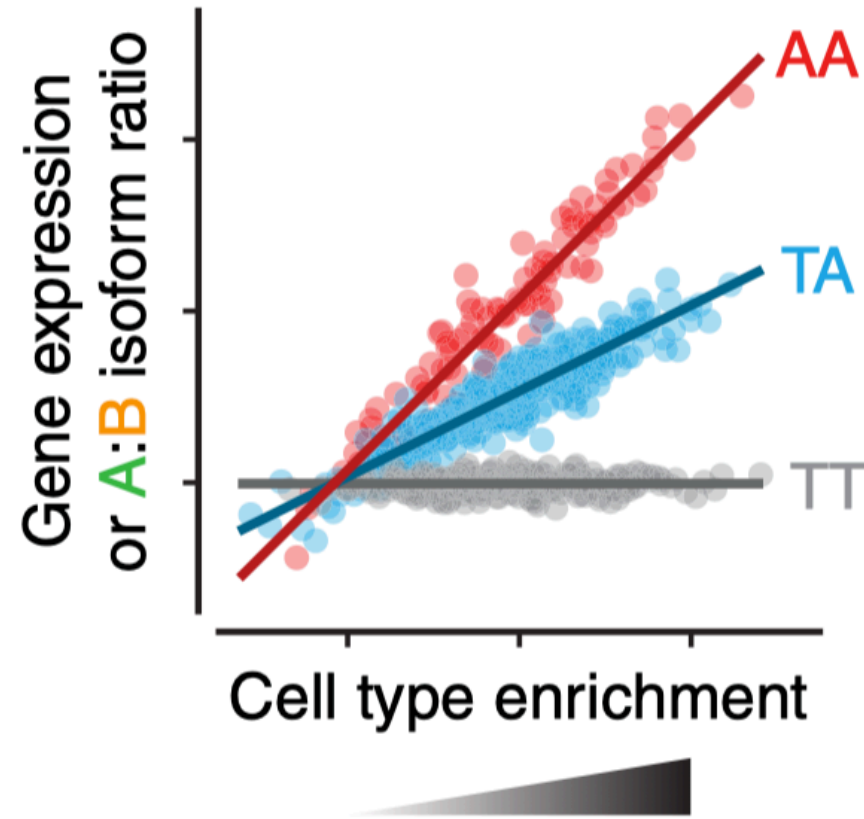
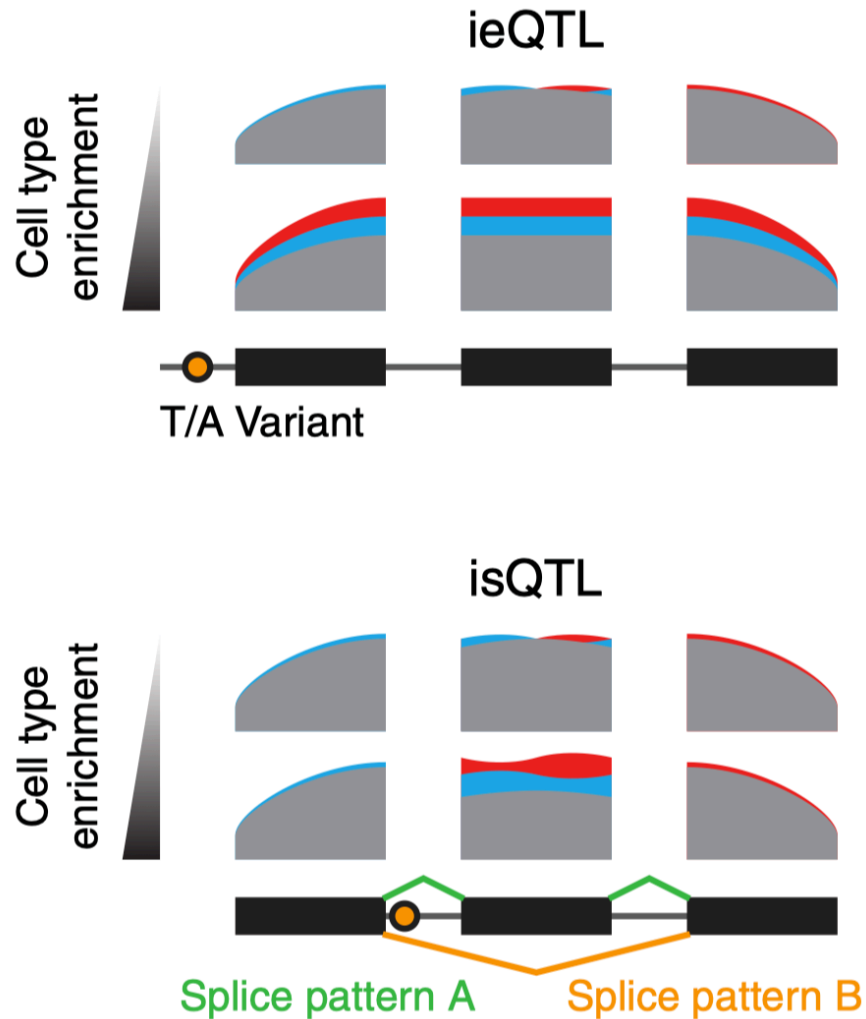
Summary of tissues used in this work

Tissue	Abbreviation	Tissue	Abbreviation
Adipose - Subcutaneous	ADPSBQ	Heart - Atrial Appendage	HRTAA
Adipose - Visceral (Omentum)	ADPVSC	Heart - Left Ventricle	HRTLTV
Brain - Amygdala	BRNAMY	Kidney - Cortex	KDNCTX
Brain - Anterior cingulate cortex (BA24)	BRNACC	Liver	LIVER
Brain - Caudate (basal ganglia)	BRNCDT	Lung	LUNG
Brain - Cerebellar Hemisphere	BRNCHB	Minor Salivary Gland	SLVRYG
Brain - Cerebellum	BRNCHA	Muscle - Skeletal	MSCLSK
Brain - Cortex	BRNCTXA	Pancreas	PNCREAS
Brain - Frontal Cortex (BA9)	BRNCTXB	Pituitary	PTTARY
Brain - Hippocampus	BRNHPP	Prostate	PRSTTE
Brain - Hypothalamus	BRNHPT	Skin - Not Sun Exposed (Suprapubic)	SKINNS
Brain - Nucleus accumbens (basal ganglia)	BRNNCC	Skin - Sun Exposed (Lower leg)	SKINS
Brain - Putamen (basal ganglia)	BRNPTM	Small Intestine - Terminal Ileum	SNTTRM
Brain - Spinal cord (cervical c-1)	BRNSPC	Stomach	STMACH
Brain - Substantia nigra	BRNSNG	Thyroid	THYROID
Breast - Mammary Tissue	BREAST	Vagina	VAGINA
Colon - Transverse	CLNTRN	Whole Blood	WHLBLD
Esophagus - Mucosa	ESPMCS		

Mapping cell type–interaction eQTLs and sQTLs

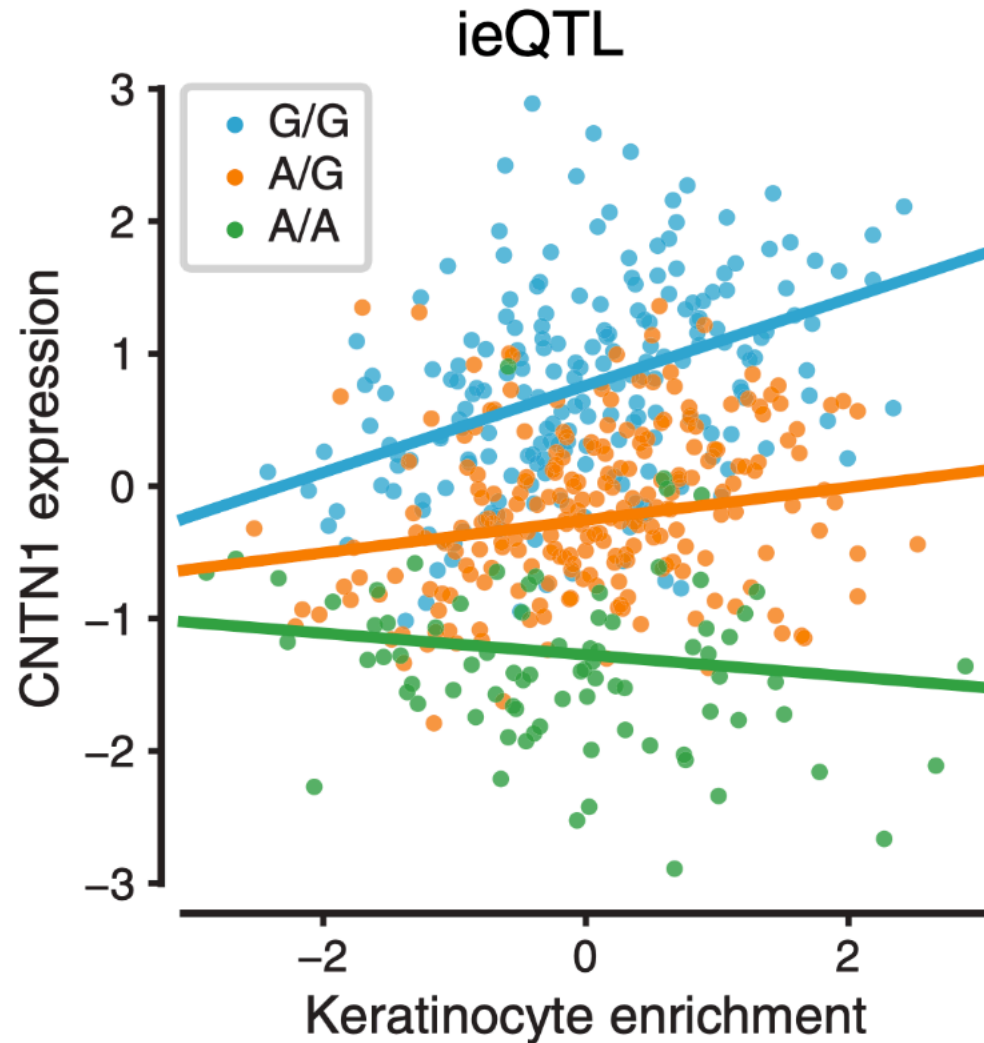
Mapping cell type–interaction eQTLs and sQTLs

- Cell type-interaction QTLs: cis-eQTLs and cis-sQTLs whose effect varies depending on the enrichment of the cell type

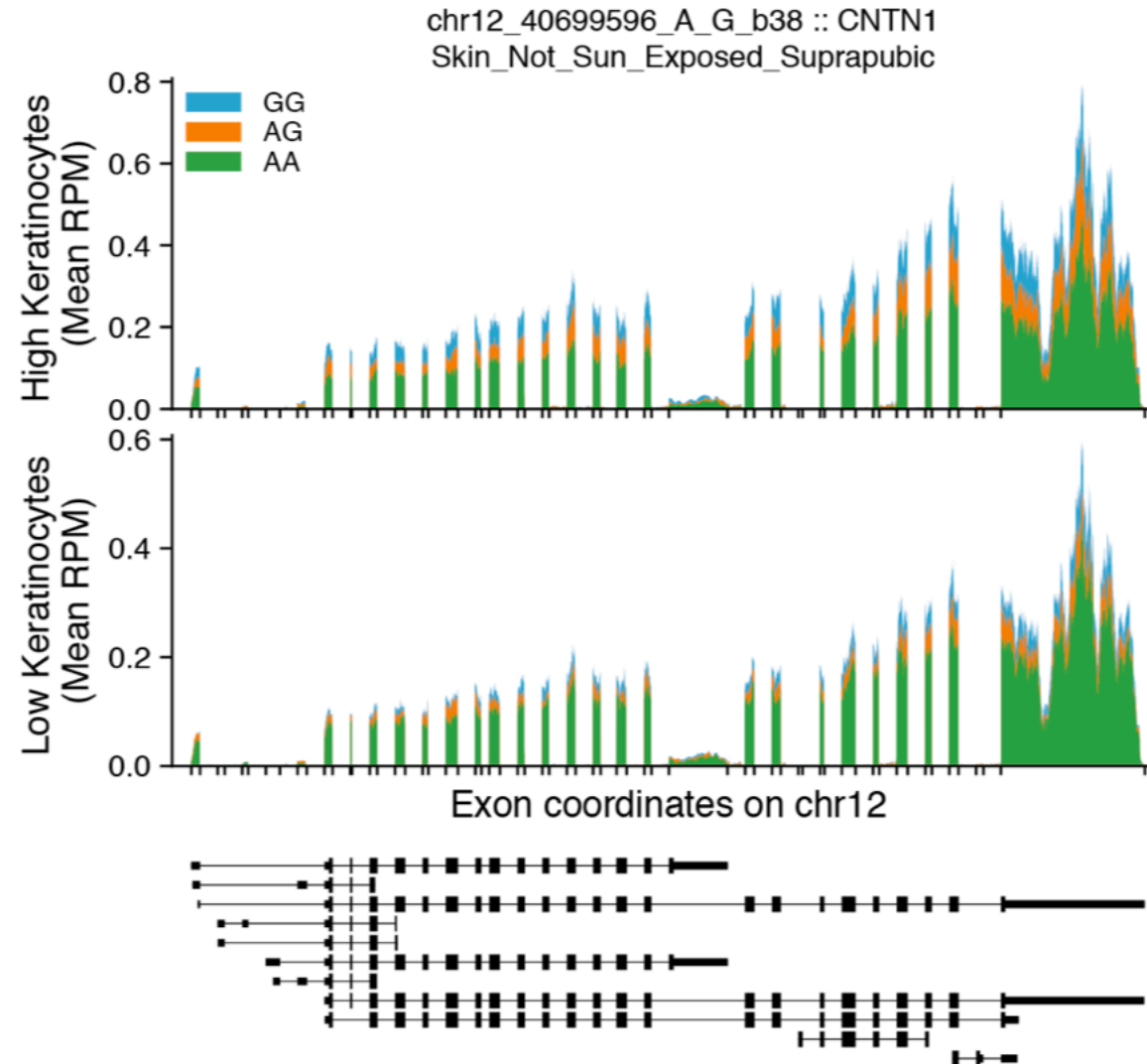


An example of cell type ieQTL

- The *CNTN1* eQTL effect in skin unexposed to the Sun is associated with keratinocyte abundance ($P = 4.1 \times 10^{-19}$)

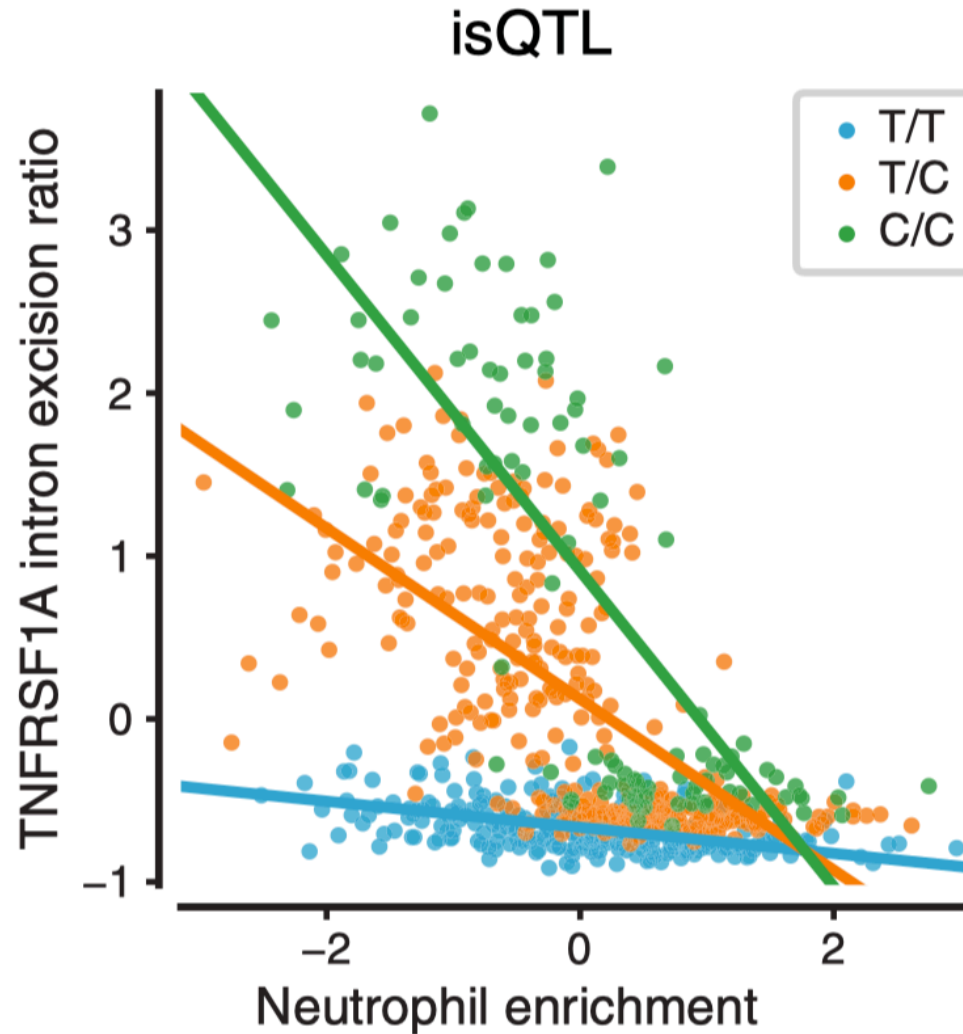


The eQTL for *CNTN1* is most strongly observed in samples with high keratinocyte enrichment

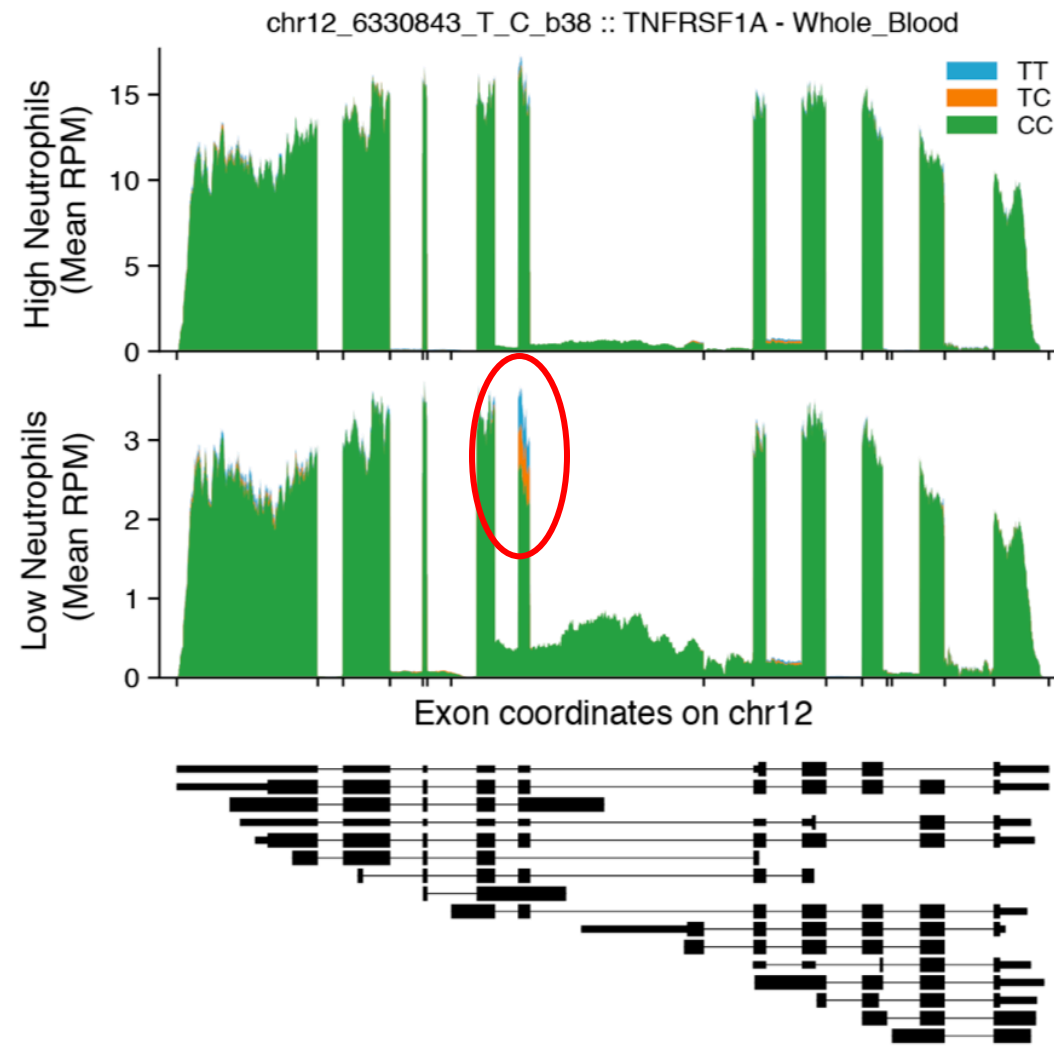


An example of cell type isQTL

- The *TNFRSF1A* sQTL effect in whole blood is associated with neutrophil abundance but is only detected in samples with lower neutrophil abundances ($P = 6.7 \times 10^{-78}$).

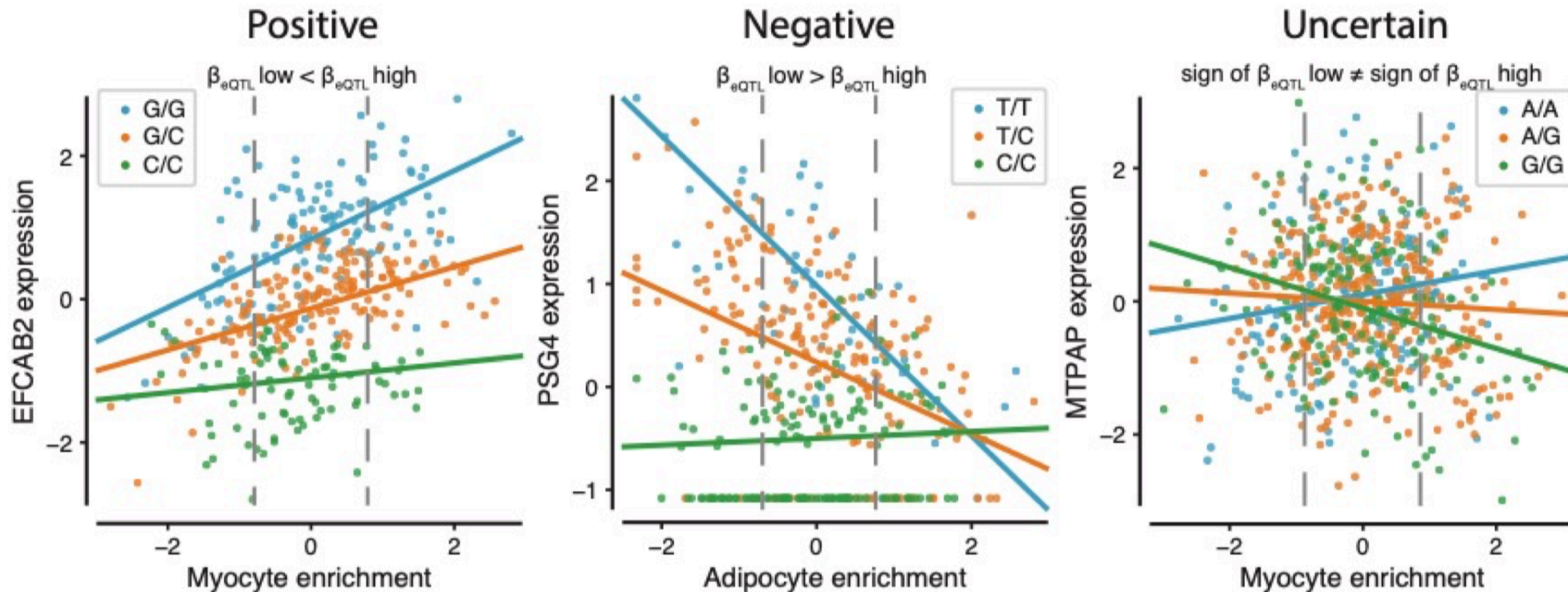


The eQTL for *TNFRSF1A* is only observed in samples with low neutrophil enrichment



Correlation between ieQTL effects and cell type enrichment scores

- The QTL effect of ieQTLs and isQTLs can increase or decrease as a function of cell type enrichment
- 56% are positive, 19% are negative
- ieQTL examples that are either positively (left) or negatively (middle) correlated with cell type estimates, or where cell type correlation is uncertain (right)

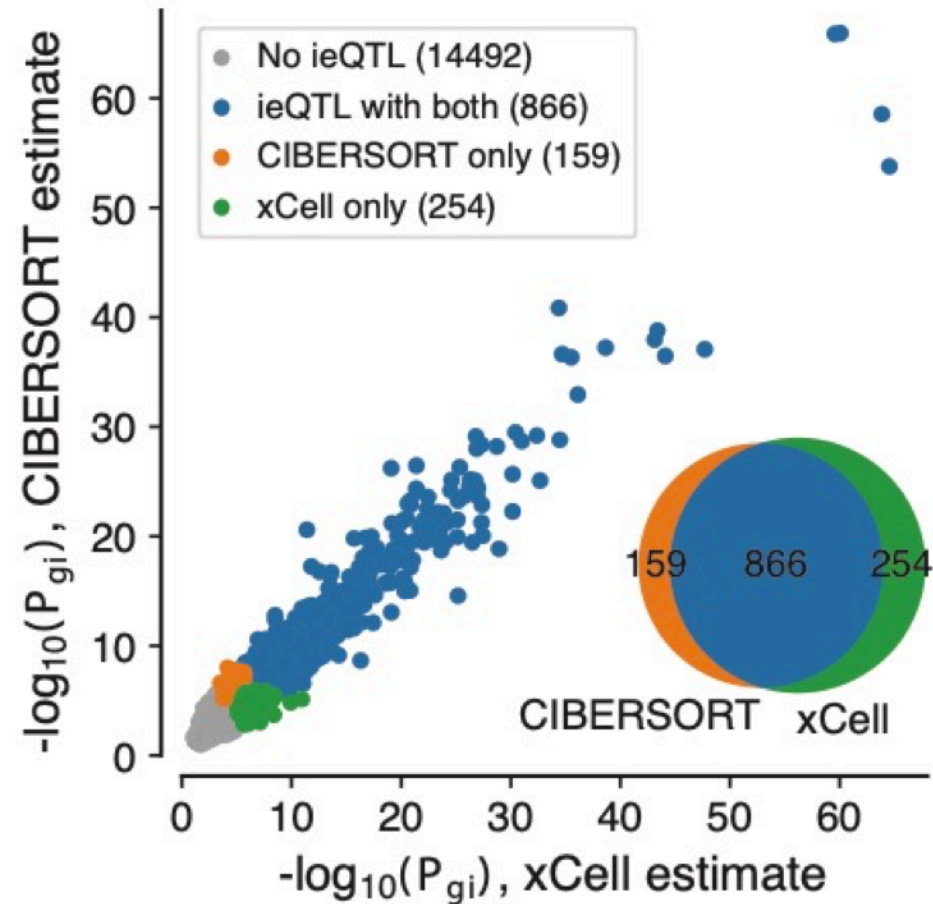


Caution

These cell type iQTLs pinpoint the cellular specificity of QTLs that might not necessarily be specific to the tested cell type but may also capture QTL effects of correlated (or anticorrelated) cell types.

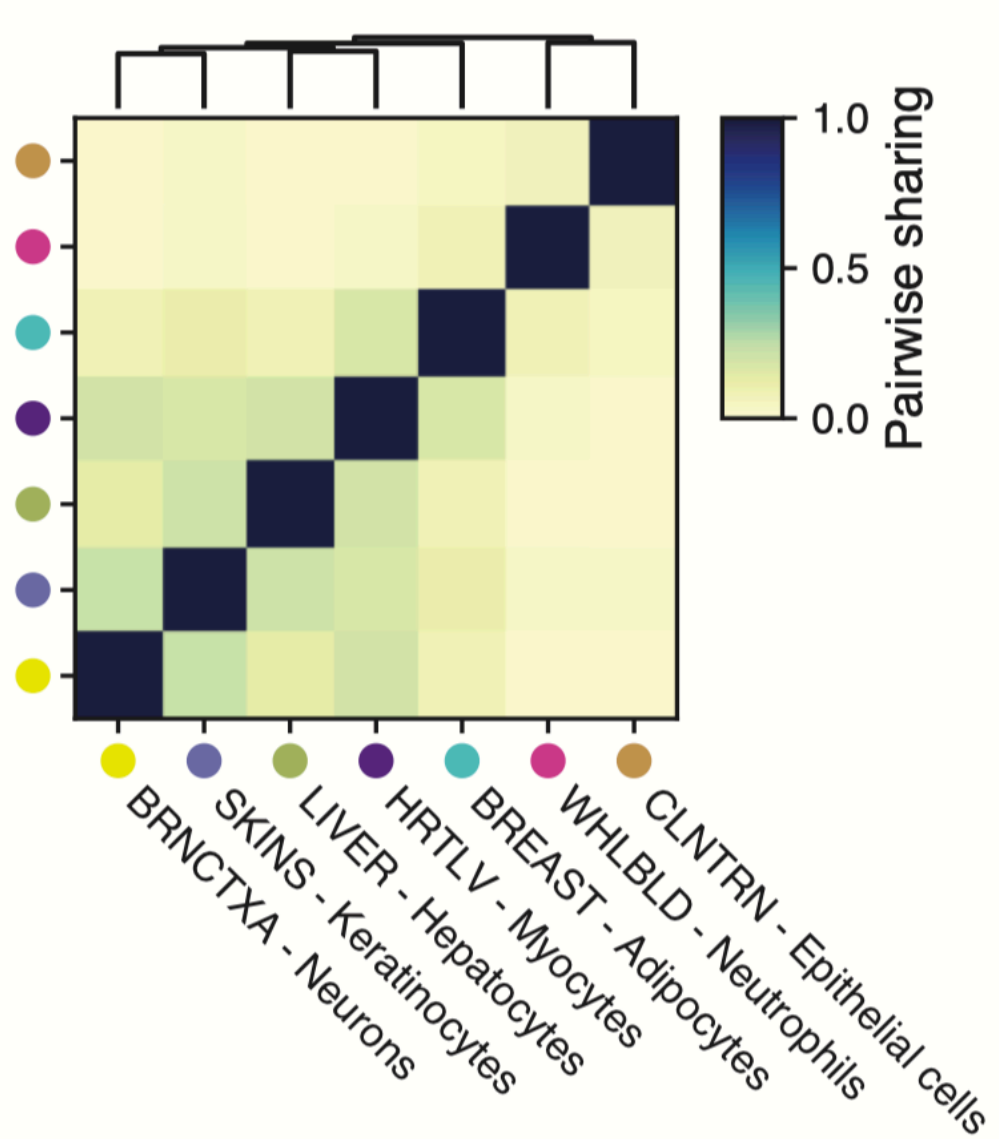
The identification of ieQTL and isQTL is robust to cell type deconvolution method

- Neutrophil ieQTL mapping using xCell estimates vs. CIBERSORT estimates
- Each dot corresponds to the p-value of the top-associated variant per gene



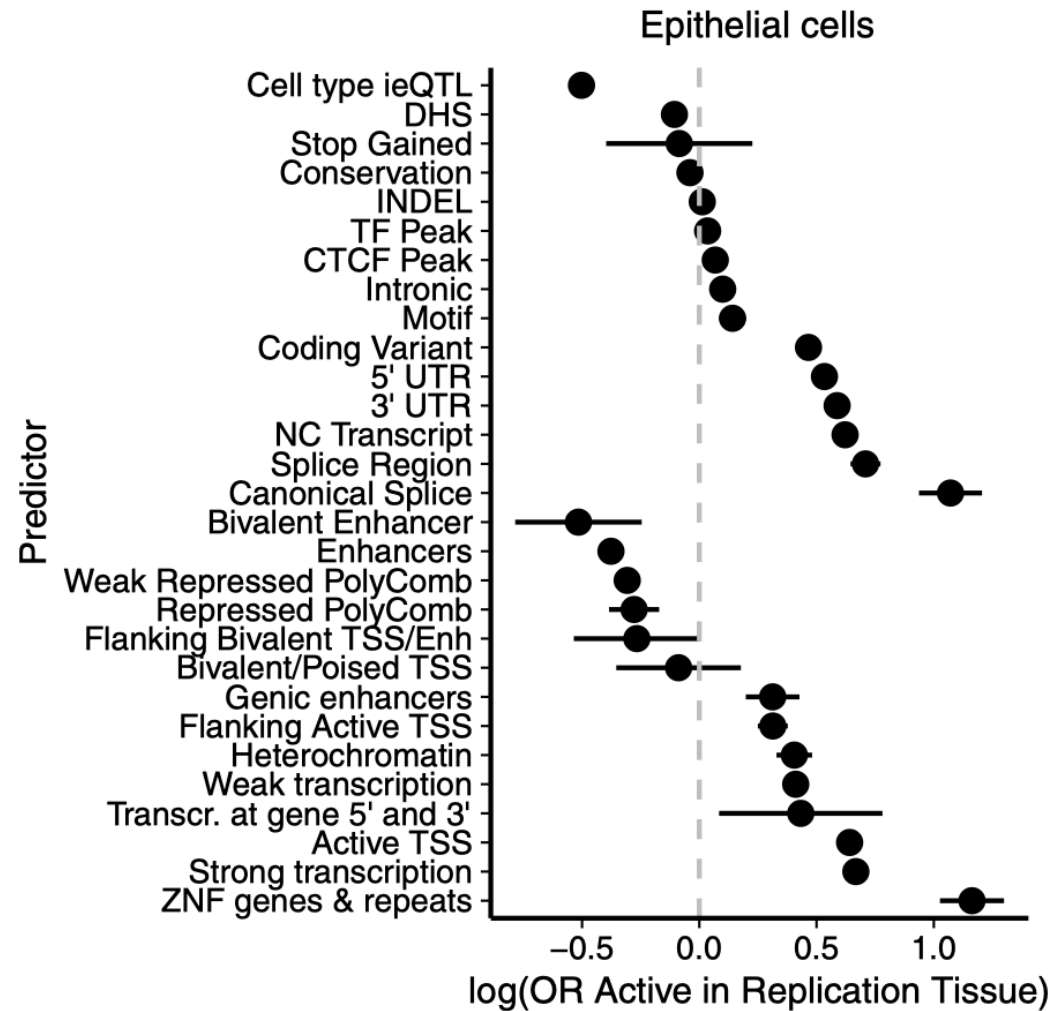
**Cell type ieQTLs contribute to tissue
specificity**

ieQTLs for one cell type were generally not ieQTLs for other cell types



Cell type ieQTLs contribute to cis-eQTL tissue specificity

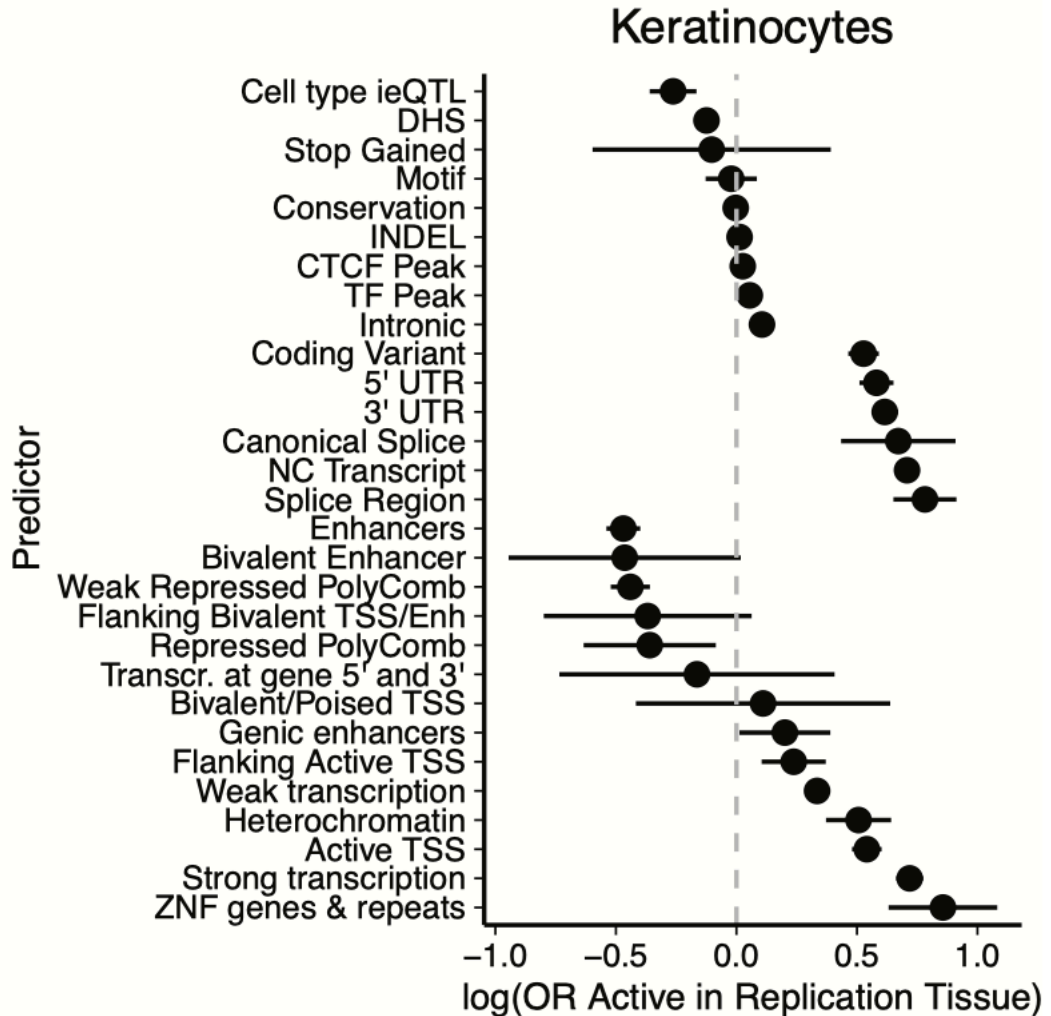
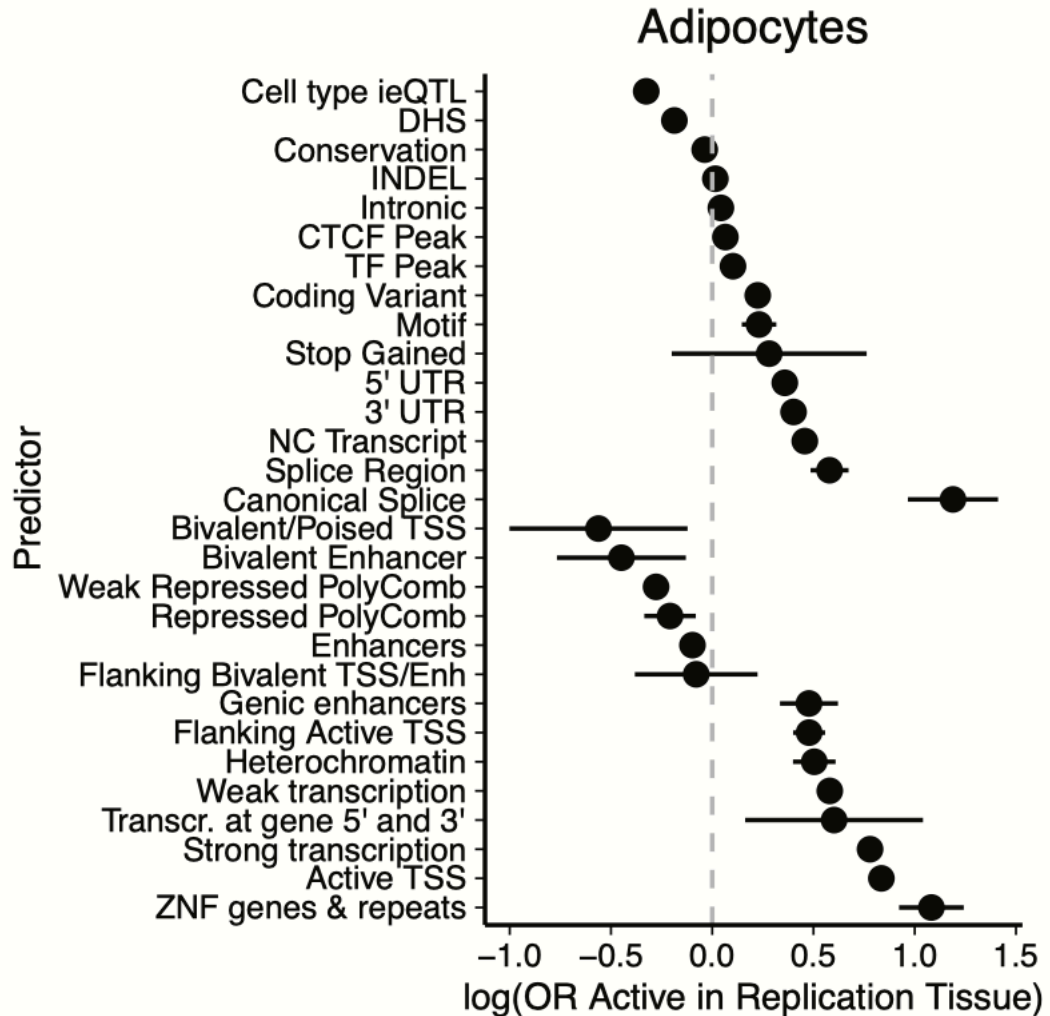
- The coefficients represent the log(odds ratio) that an eQTL is active in a replication tissue given a predictor
- Bars represent the 95% confidence interval



OR: odds ratio

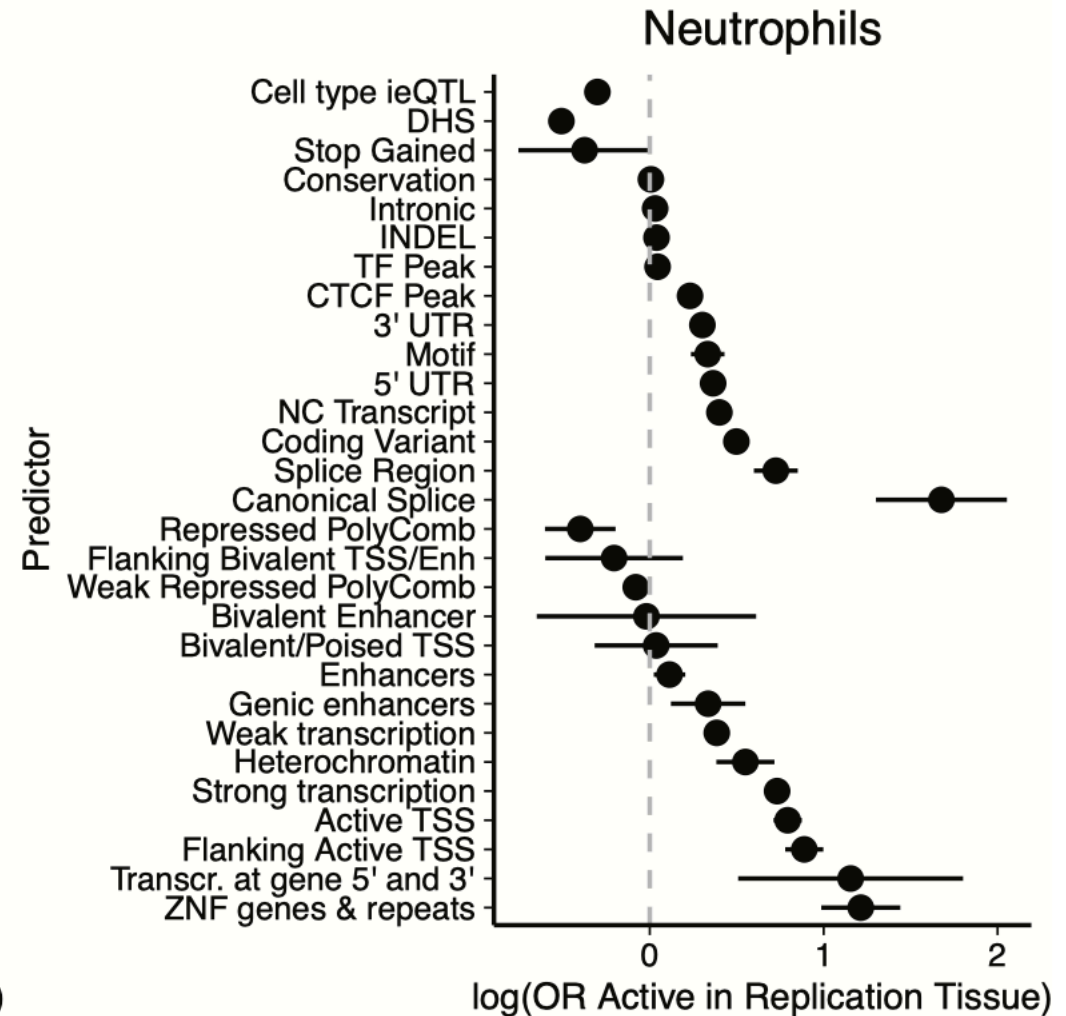
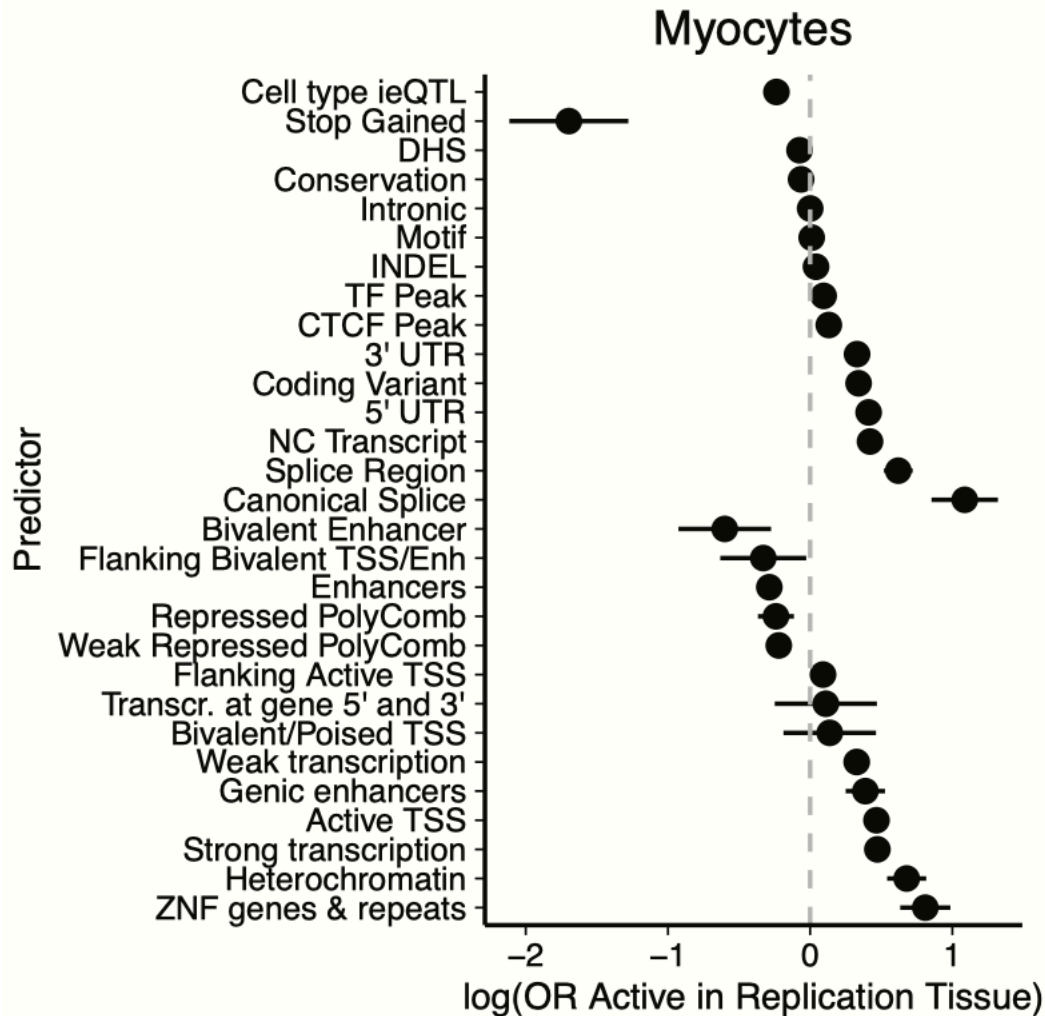
Examples of adipocytes and keratinocytes

- The coefficients represent the log(odds ratio) that an eQTL is active in a replication tissue given a predictor
- Bars represent the 95% confidence interval



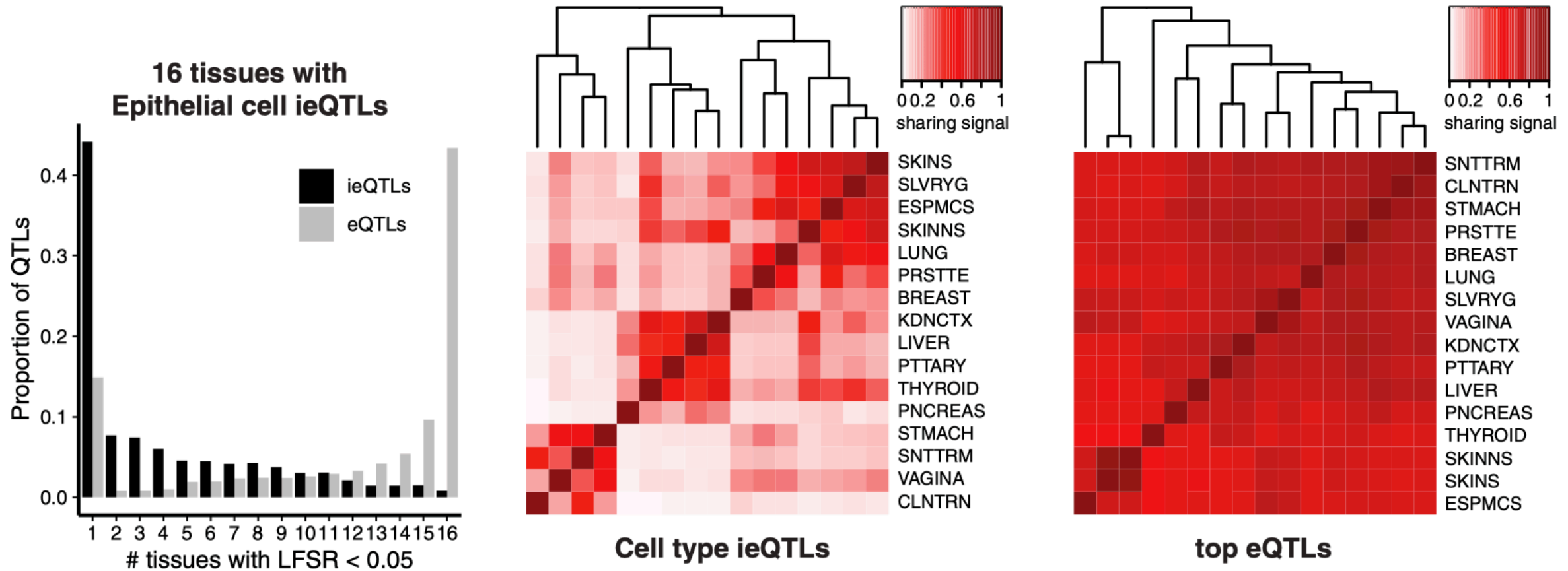
Examples of adipocytes and keratinocytes

- The coefficients represent the $\log(\text{odds ratio})$ that an eQTL is active in a replication tissue given a predictor
- Bars represent the 95% confidence interval



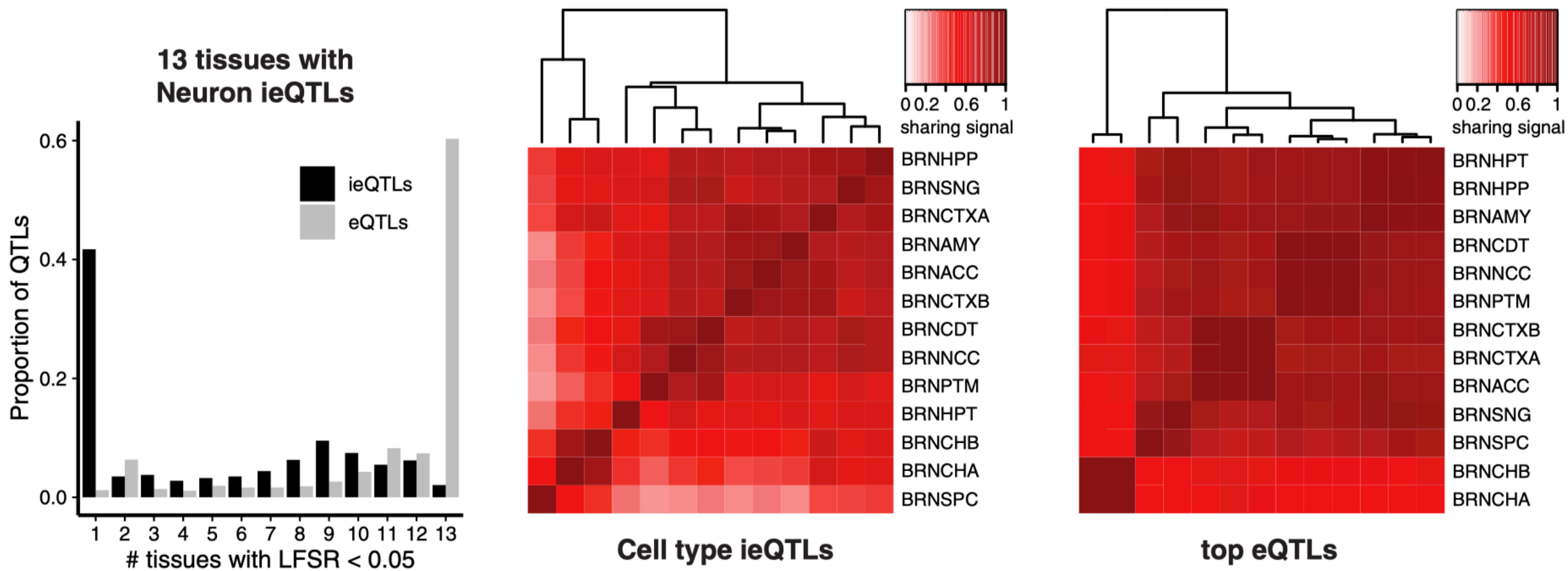
Comparison of tissue specificity of cell type ieQTLs and eQTLs (Epithelial cell)

- The authors used two cell types with ieQTLs mapped in >10 tissues (16 tissues for epithelial cells and 13 for neurons) to examine the sharing patterns of cell type ieQTLs across tissues



Comparison of tissue specificity of cell type ieQTLs and eQTLs (Neuron)

- The authors used two cell types with ieQTLs mapped in >10 tissues (16 tissues for epithelial cells and 13 for neurons) to examine the sharing patterns of cell type ieQTLs across tissues



GWAS and tissue-specific eQTLs and sQTLs

Summary of 87 GWAS traits studies in this work

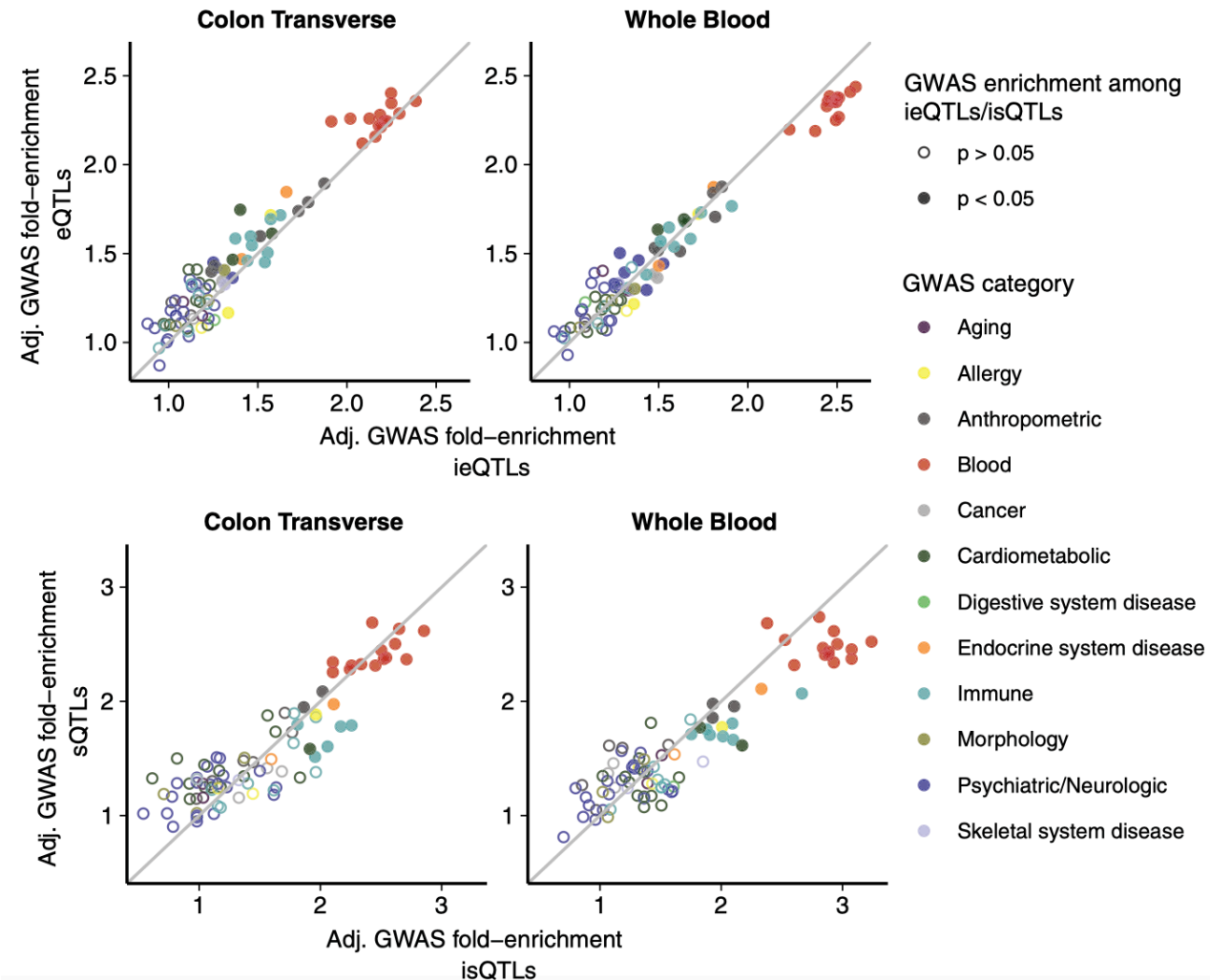
Eosinophil_Count	Birth_Weight	Depressive_Symptoms	Ankylosing_Spondylitis_UKBS	Heart_Attack_UKB
Granulocyte_Count	Intracranial_Volume	Education_Years	Eczema_UKBS	Deep_Venous_Thrombosis_UKB
High_Light_Scatter_Reticulocyte_Count	Bone_Mineral_Density	Sleep_Duration_UKB	Psoriasis_UKBS	Pulmonary_Embolism_UKB
Lymphocyte_Count	Height	Chronotype_UKB	Inflammatory_Bowel_Disease_UKBS	Asthma_UKB
Monocyte_Count	Crohns_Disease	Insomnia_UKB	Crohns_Disease_UKBS	Hayfever_UKB
Myeloid_White_Cell_Count	Inflammatory_Bowel_Disease	Fathers_Age_At_Death_UKB	Ulcerative_Colitis_UKBS	Epilepsy_UKB
Neutrophil_Count	Ulcerative_Colitis	Hypertension_UKBS	Rheumatoid_Arthritis_UKBS	Migraine_UKB
Platelet_Count	Alzheimers_Disease	Deep_Venous_Thrombosis_UKBS	Gout_UKBS	
Red_Blood_Cell_Count	Systemic_Lupus_Erythematosus	Asthma_UKBS	High_Cholesterol_UKBS	
Reticulocyte_Count	Chronotype	Irritable_Bowel_Syndrome_UKBS	Insomnia_UKBS	
Sum_Basophil_Neutrophil_Count	Sleep_Duration	Type_1_Diabetes_UKBS	Fluid_Intelligence_Score_UKB	
Sum_Eosinophil_Basophil_Count	Fasting_Glucose	Type_2_Diabetes_UKBS	Birth_Weight_UKB	
Sum_Neutrophil_Eosinophil_Count	Fasting_Insulin	Hyperthyroidism_UKBS	Neuroticism_UKB	
White_Blood_Cell_Count	CH2DB_NMR	Hypothyroidism_UKBS	BMI_UKB	
ER-negative_Breast_Cancer	HDL_Cholesterol_NMR	Psychological_Problem_UKBS	Body_Fat_Percentage_UKB	
ER-positive_Breast_Cancer	Triglycerides_NMR	Multiple_Sclerosis_UKBS	Balding_Pattern_2_UKB	
Breast_Cancer	LDL_Cholesterol_NMR	Parkinsons_Disease_UKBS	Balding_Pattern_3_UKB	
Coronary_Artery_Disease	Attention_Deficit_Hyperactivity_Disorder	Migraine_UKBS	Balding_Pattern_4_UKB	
Insomnia_In_Both_Sexes	Schizophrenia	Schizophrenia_UKBS	Mothers_Age_At_Death_UKB	
Eczema	Rheumatoid_Arthritis	Osteoporosis_UKBS	Standing_Height_UKB	

Category of 87 GWAS traits

Digestive system disease	Allergy
Endocrine system disease	Anthropometric
Skeletal system disease	Cardiometabolic
Morphology	Immune
Aging	Psychiatric_neurologic
Blood	Cancer

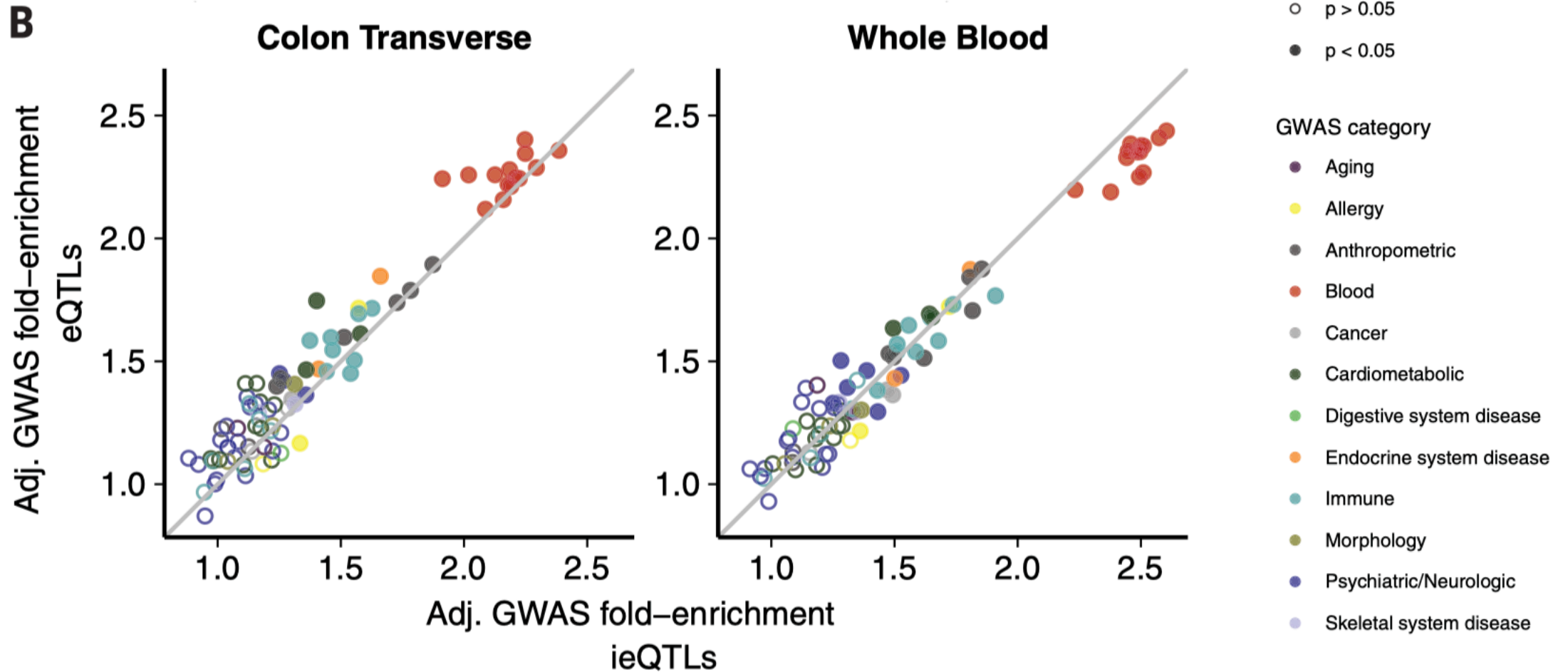
Enrichments of the individual traits for iQTLs of neutrophil in blood and epithelial cell in transverse colon

- Neutrophil iQTLs in blood and epithelial cell iQTLs in transverse colon are the two cell types that had the largest number of ieQTLs



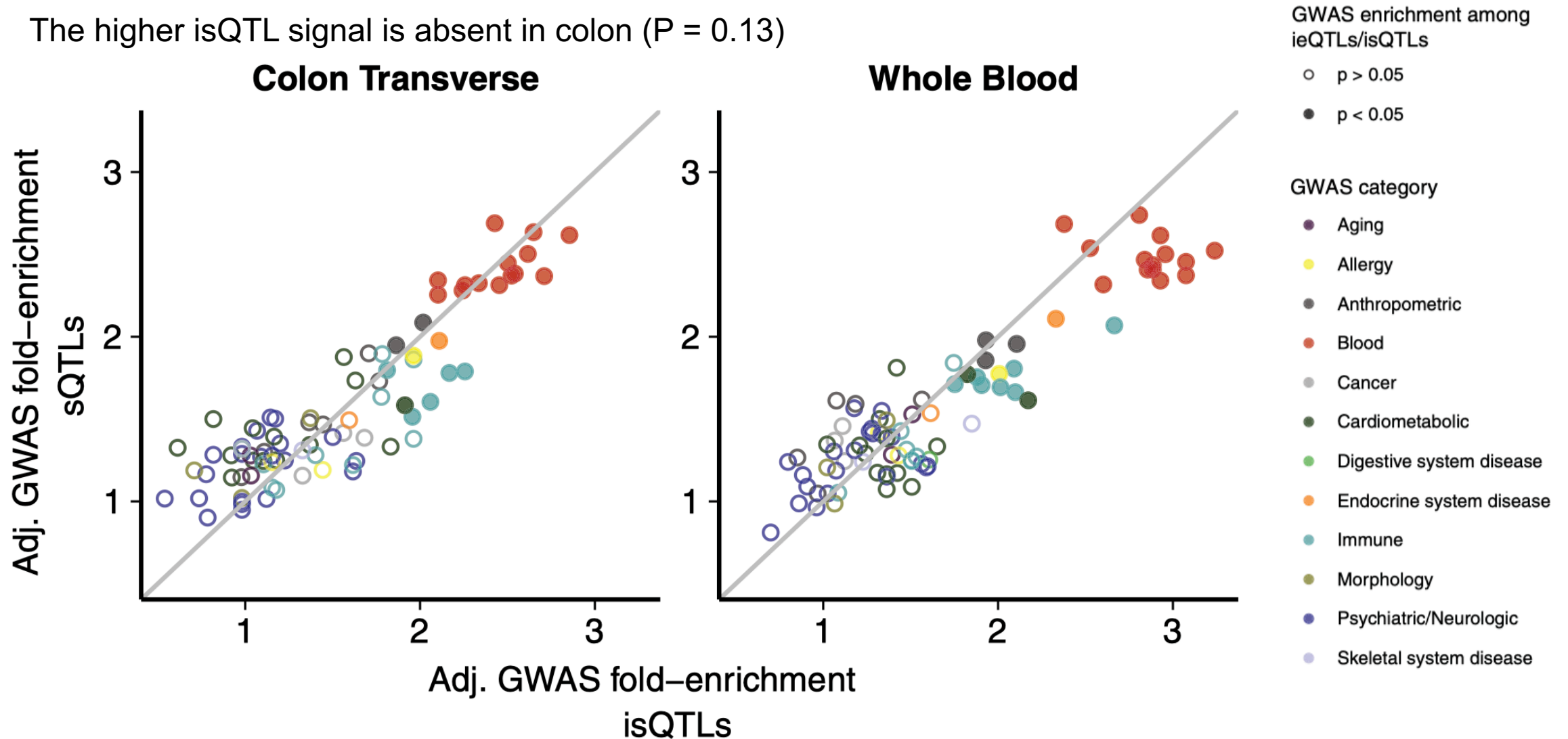
A significant shift toward higher enrichment for ieQTLs in blood

- Filled circles indicate significant GWAS enrichment among ieQTLs at $p < 0.05$ (Bonferroni-corrected)
- One-sided, paired Wilcoxon rank sum test; $P = 0.0026$
- The higher ieQTL signal is absent in colon ($P = 1$)



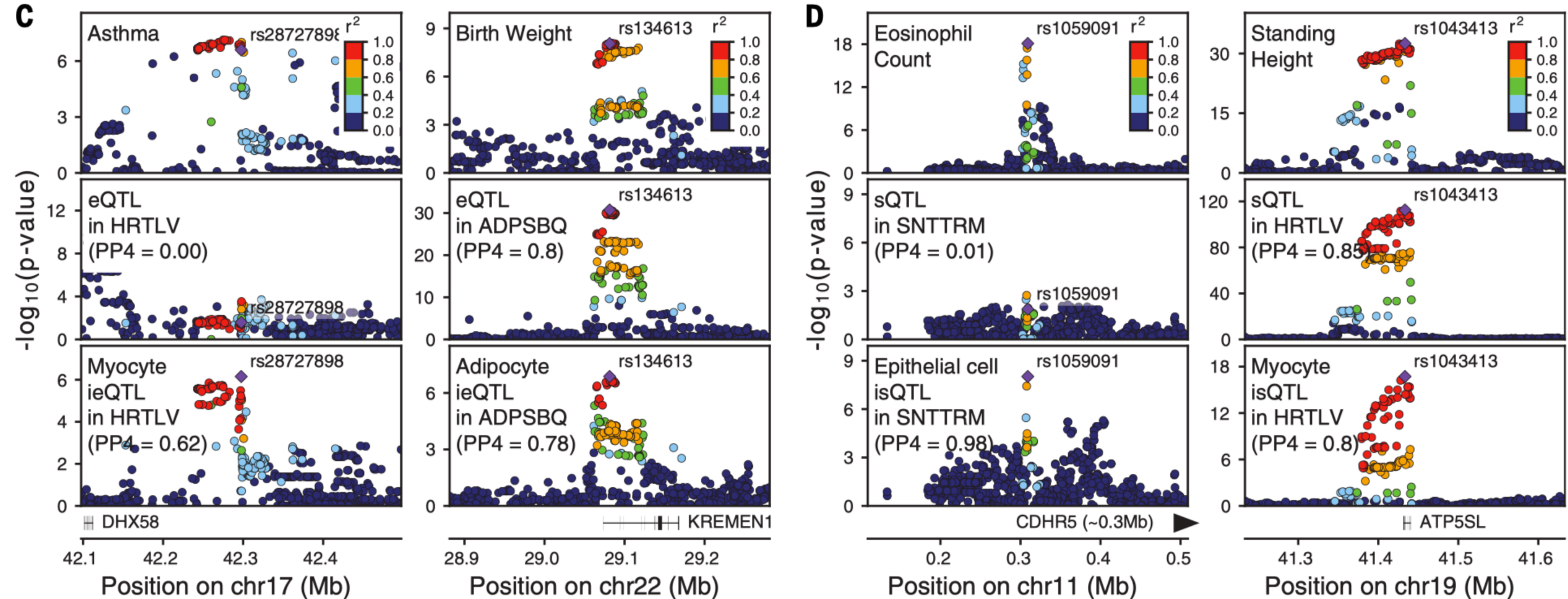
A significant shift toward higher enrichment for isQTLs in blood

- Filled circles indicate significant GWAS enrichment among isQTLs at $p < 0.05$ (Bonferroni-corrected)
- One-sided, paired Wilcoxon rank sum test; $P = 2.8 \times 10^{-05}$
- The higher isQTL signal is absent in colon ($P = 0.13$)



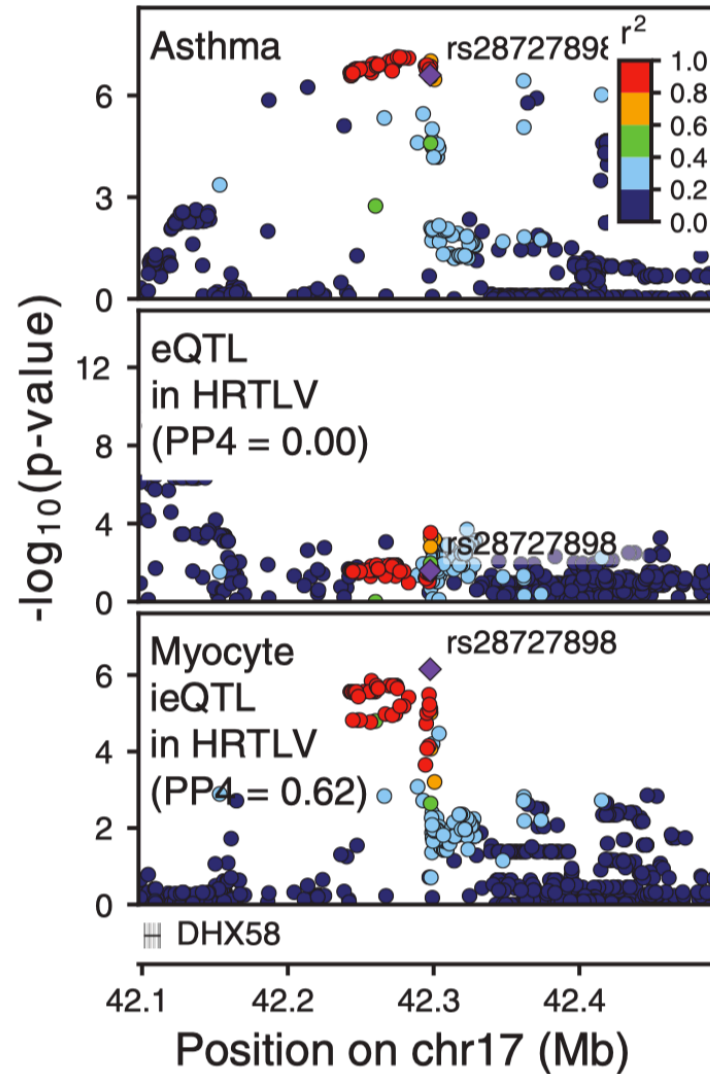
Cell type–interaction QTLs yield new potential target genes for GWAS loci that are missed by standard QTLs

- Cell type specificity plays a role in the GWAS locus



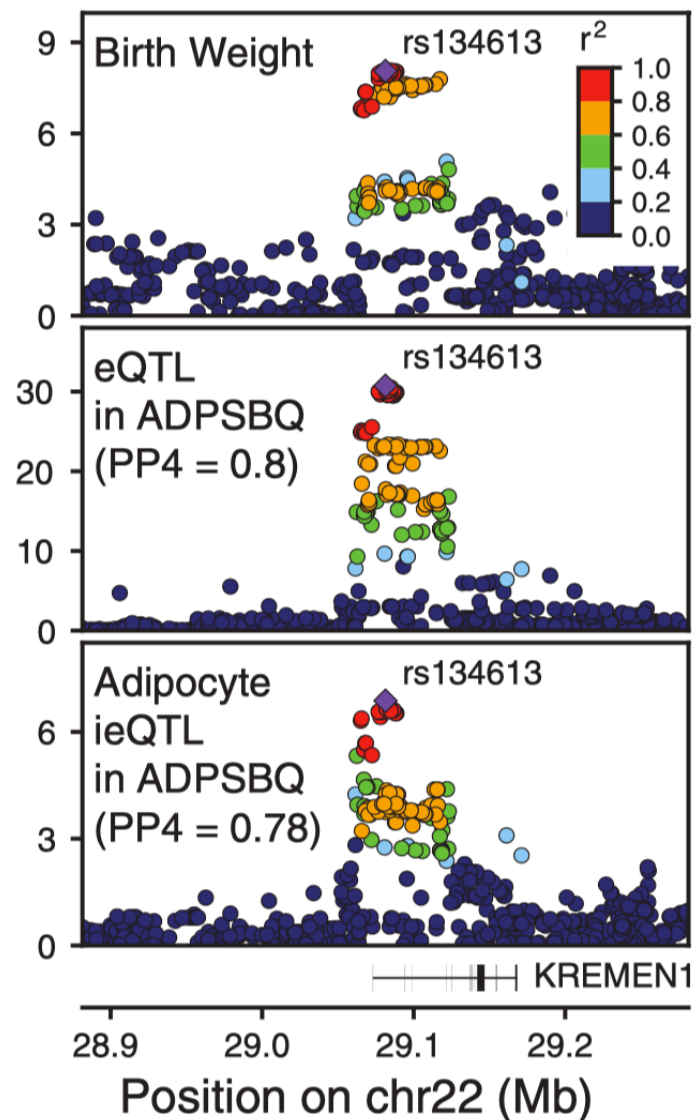
A colocalization between *DHX58* was only identified through the corresponding myocyte but not the standard eQTL

- *DHX58* in the left ventricle of the heart and an asthma GWAS



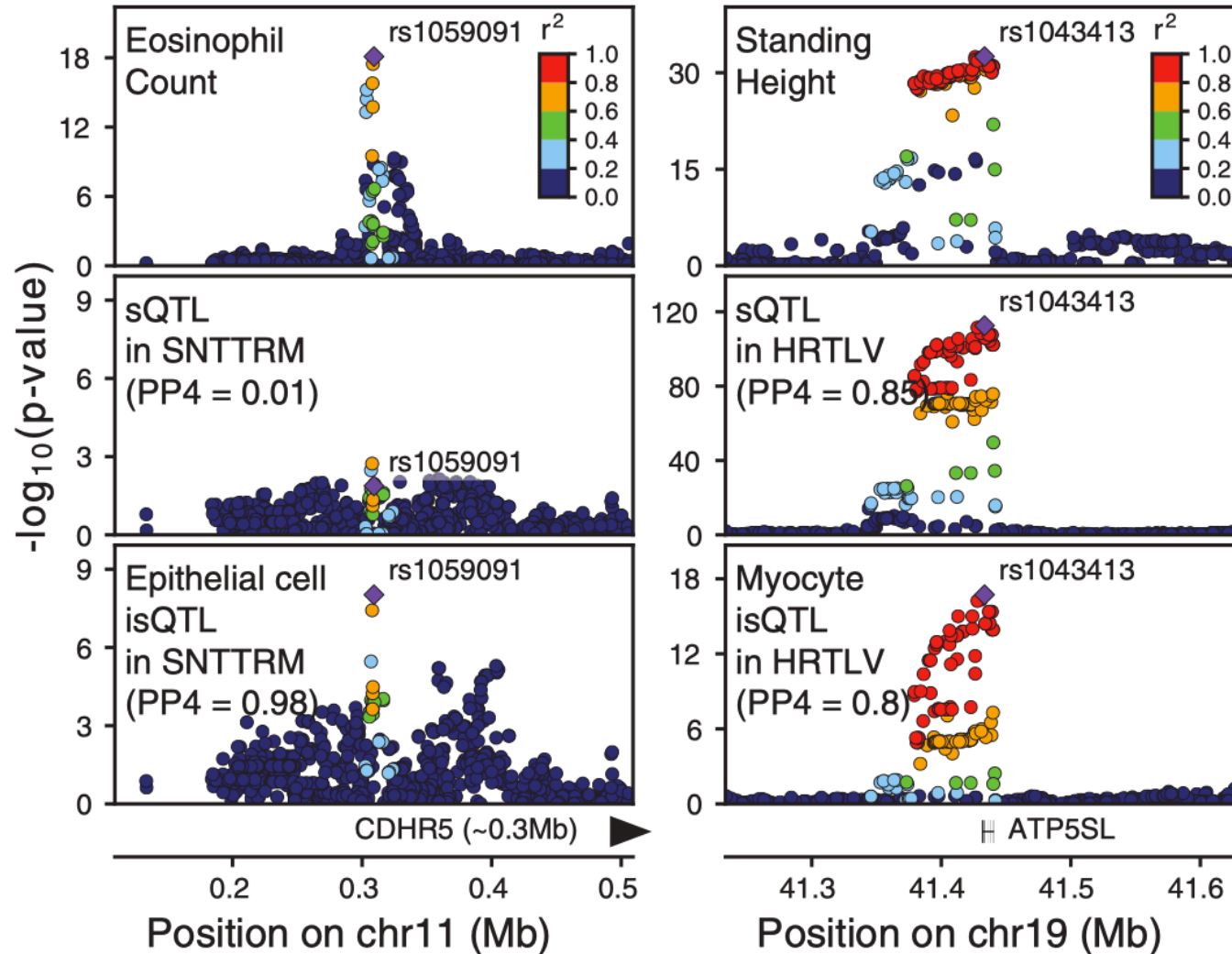
Another example: both the standard eQTL and the cell type ieQTL colocalize with the trait

- *KREMEN1* in adipocytes in subcutaneous adipose tissue and a birth weight GWAS



Two analogous examples for isQTLs

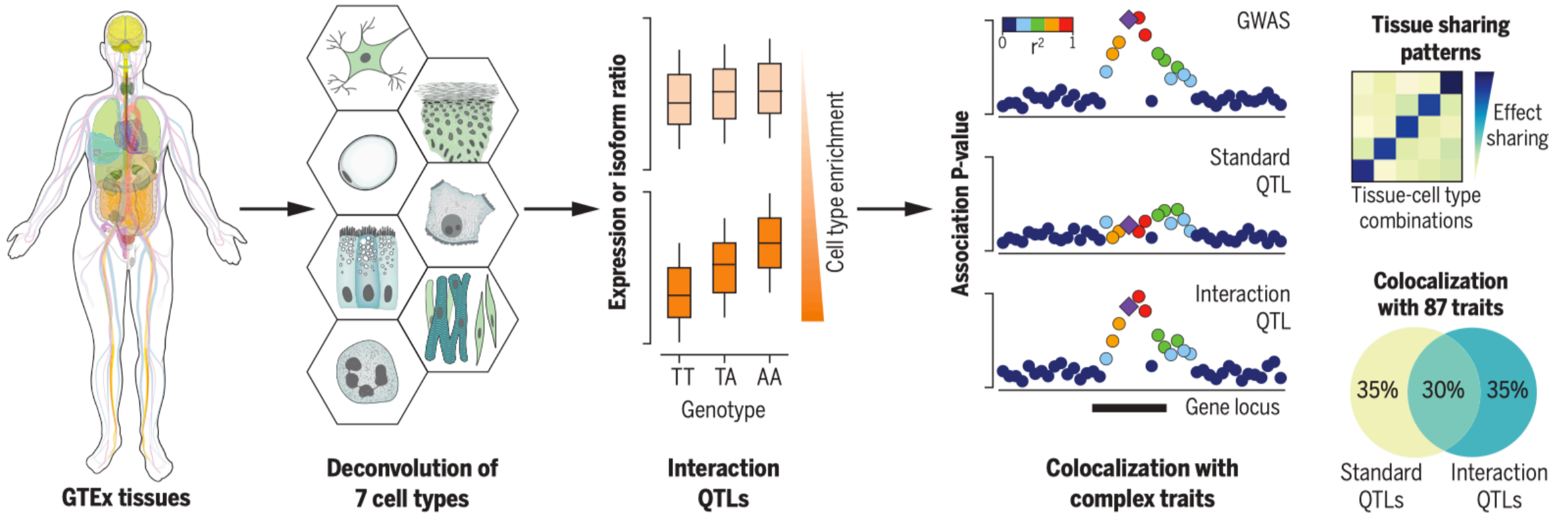
- The epithelial cell isQTL for *CDHR5* in small intestine colocalized with eosinophil counts, whereas the standard sQTL did not
- Both the standard sQTL and myocyte isQTL for *ATP5SL* in the left ventricle of the heart colocalized with standing height



Summary

- The authors used *in silicon* cell type deconvolution methods to identify cell type–interaction QTLs for expression and splicing by testing for interactions between genotype and cell type enrichment
- Cell type iQTLs are strongly enriched for tissue and cellular specificity and provide a finer resolution to tissue specificity than that of bulk cis-QTLs that are highly shared between tissues
- Cell type–interaction QTLs yield new potential target genes for GWAS loci that are missed by standard QTLs, and provide hypotheses for the cellular specificity of regulatory effects underlying complex traits

Cellular specificity of regulatory effects underlying complex traits



Thinking

How to better integrate bulk RNA sequencing data and single-cell RNA sequencing data?

Thank you!



Discussion