

Topic 14: GTEx

Group4: 黄甫保钱; 程心怡; 陈婉桐; 廖阔; 申雪纯

2021-11-25

47.728 > [Science](#). 2020 Sep 11;369(6509):1318-1330. doi: 10.1126/science.aaz1776.

HUMAN GENOMICS

The GTEx Consortium atlas of genetic regulatory effects across human tissues

The GTEx Consortium*

12 groups:

Laboratory and Data Analysis Coordinating Center (LDACC)
Analysis Working Group Funded by GTEx Project Grants
Analysis Working Group Not Funded by GTEx Project Grants
Leidos Biomedical Project Management
Biospecimen Collection Source Sites
Biospecimen Core Resource
Brain Bank Repository
Pathology
ELSI Study
Genome Browser Data Integration and Visualization
eGTEx Groups
NIH Program Management



第一作者

François Aguet

- 瑞士计算生物学家
- 人类遗传变异
- 细胞类型特异性基因表达调控

研究背景

基因型-组织表达 (Genotype-Tissue Expression, GTEx) 计划

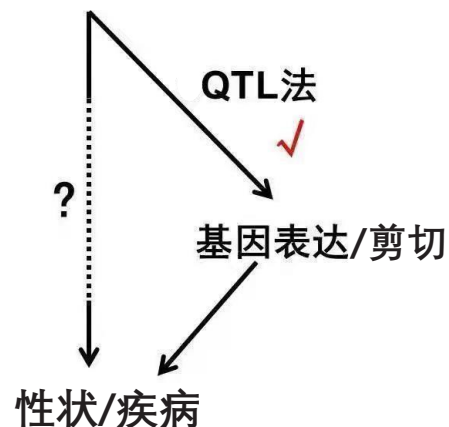
启动：2010年

目的：评估基因变异对人体不同组织的转录水平影响；

将相应的调控机制与疾病、性状关联起来。

成就：GTEx数据库，持续更新（当前为V8）

基因上碱基的改变

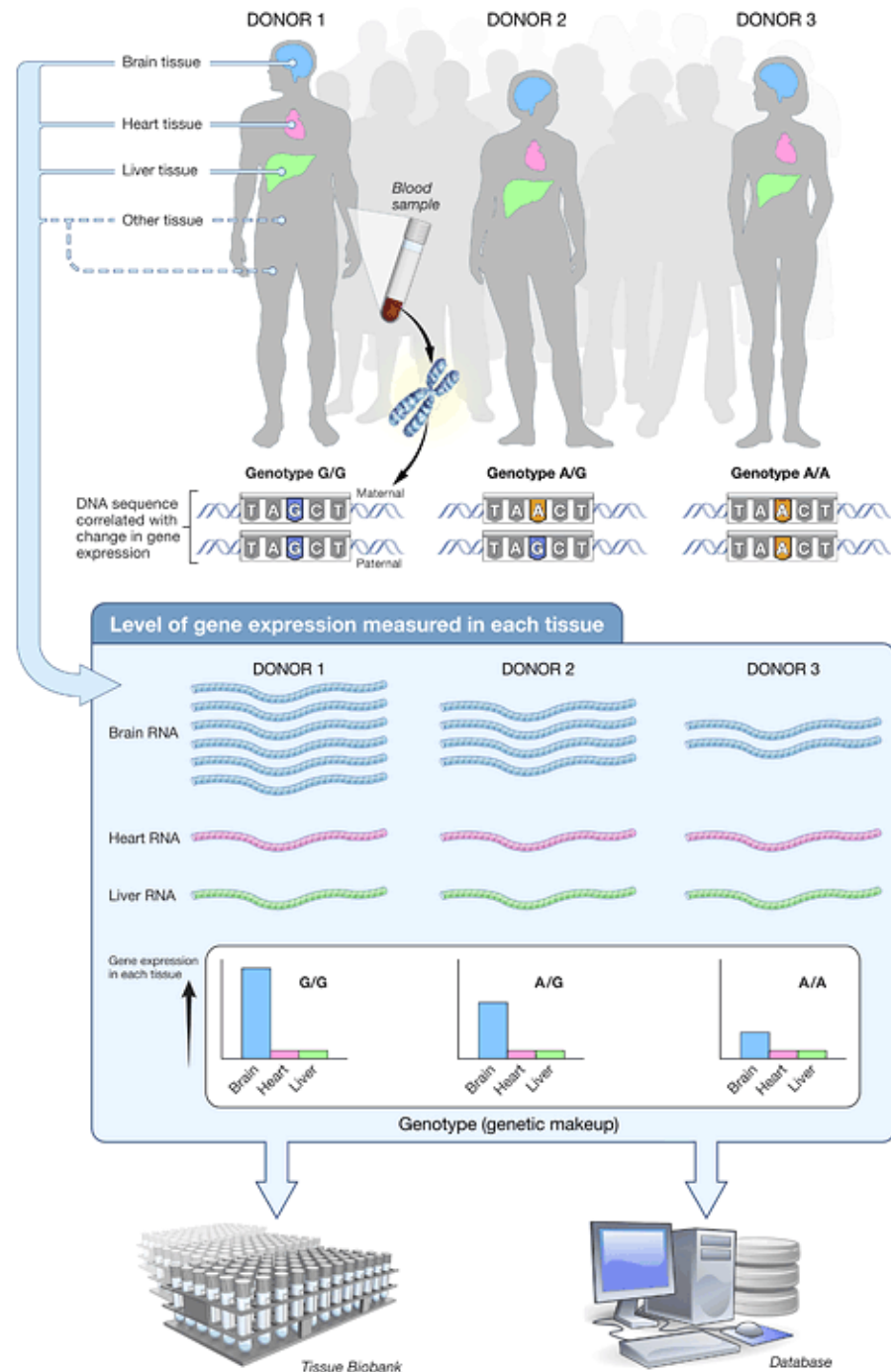


- 54 种人体健康组织样本；
- 948 名个体；
- WGS, WES, RNA-seq等数据

 GTExPortal

[Home](#) [Datasets](#) [Expression](#) [QTLs & Browsers](#) [Sample Data](#) [Documentation](#)

<https://gtexpportal.org/home/documentationPage>



GTEx Consortium Publications

2020

The GTEx Consortium atlas of genetic regulatory effects across human tissues

The GTEx Consortium.

Science. 369 (1318-1330), 10 Sep 2020. doi:10.1126/science.aaz1776

Cell type specific genetic regulation of gene expression across human tissues

Kim-Hellmuth* S, Aguet* F, Oliva M, Muñoz-Aguirre M, Kasela S, et al.

Science. 369 (eaaz8528), 10 Sep 2020. doi:10.1126/science.aaz8528

Transcriptomic signatures across human tissues identify functional rare genetic variation

Ferraro* NM, Strober* BJ, Einson J, Abell NS, Aguet F, et al.

Science. 369 (aaz5900), 10 Sep 2020. doi:10.1126/science.aaz5900

Determinants of telomere length across human tissues

Demanelis K, Jasmine F, Chen LS, Chernoff M, Tong L, et al.

Science. 369 (aaz6876), 10 Sep 2020. doi:10.1126/science.aaz6876

The impact of sex on gene expression across human tissues

Oliva* M, Muñoz-Aguirre* M, Kim-Hellmuth* S, Wucher V, Gewirtz ADH, et al.

Science. 369 (aba3066), 10 Sep 2020. doi:10.1126/science.aba3066

Tissue-specific genetic features inform prediction of drug side effects in clinical trials

Duffy A, Verbanck M, Dobbyn A, Won H-H, Rein JL, et al.

Science Advances. 6(37), eabb6242, 10 Sep 2020. doi:10.1126/sciadv.abb6242

PhenomeXcan: Mapping the genome to the phenome through the transcriptome

Pividori M, Rajagopal PS, Barbeira A, Liang Y, Melia O, et al.

Science Advances. 6(37), eaba2083, 10 Sep 2020. doi:10.1126/sciadv.aba2083

PTWAS: investigating tissue-relevant causal molecular mechanisms of complex traits using probabilistic TWAS analysis

Zhang Y, Quick C, Yu K, Barbeira A, The GTEx Consortium, et al.

Genome Biology. 21(232), 10 Sep 2020. doi:10.1186/s13059-020-02026-y

2020年以GTEx V8为基础发表了
15篇文章。

研究背景

重要概念

- 数量性状 (quantitative trait) : 可以使用数值量化表征的性状, 比如身高和体重等。
- 数量性状基因座 (quantitative trait locus, QTL) : 可以调控数量性状的一段DNA序列或某个突变位点, 例如 SNPs、SSRs和RFLPs等。

1. 表达数量性状基因座 (expression QTL, eQTL) :

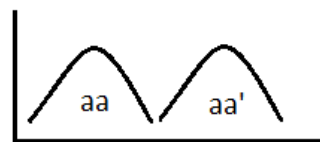
将某个基因 (eGene) 的**表达水平**视为数量性状, 可调控该表达水平的QTL (eVariant) 。

2. 剪接数量性状基因座 (splicing QTL, sQTL) :

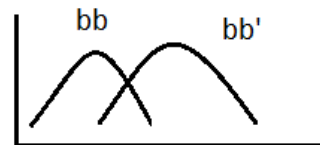
将某个基因 (sGene) 的**剪接比率**视为数量性状, 可调控该剪接比率的QTL (sVariant) 。

<https://www.zhihu.com/question/27695566>

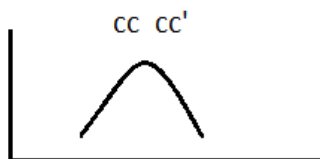
Phenotypic Distributions for Segregating RFLP Alleles at locus "a" "b" and "c"



locus "a" : RFLP tightly linked to QTL



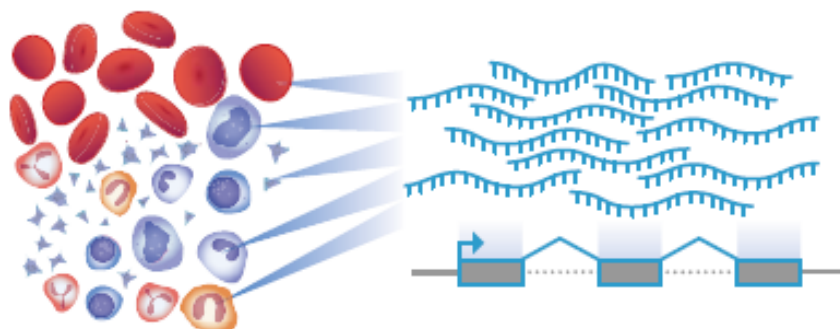
locus "b" : RFLP closely linked to QTL



locus "c" : RFLP independent from QTL

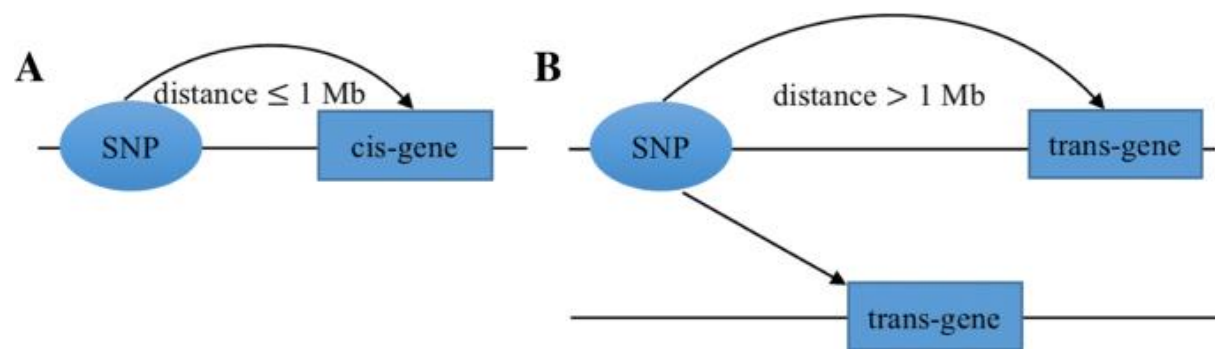
Cell type composition
in tissues

Gene expression
and splicing



研究背景

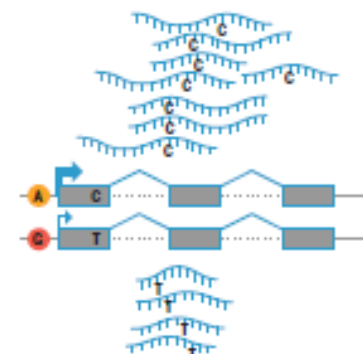
- 顺式-表达/剪接数量性状基因座 (cis-e/sQTL)：与目标基因相距较近 (一般 $\leq 1\text{Mb}$) 的QTL;
- 反式-表达/剪接数量性状基因座 (trans-e/sQTL)：与目标基因相距较远 ($> 1\text{Mb}$) 或位于不同染色体的QTL。



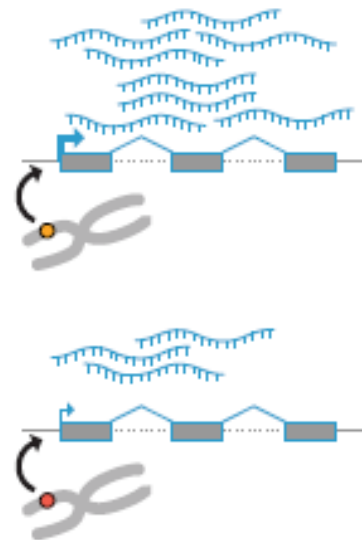
doi: 10.1186/s12859-019-2651-6.

Expression quantitative trait loci (eQTLs)

cis-eQTLs

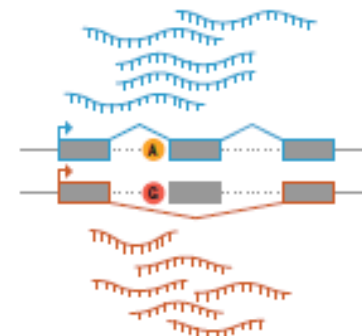


trans-eQTLs

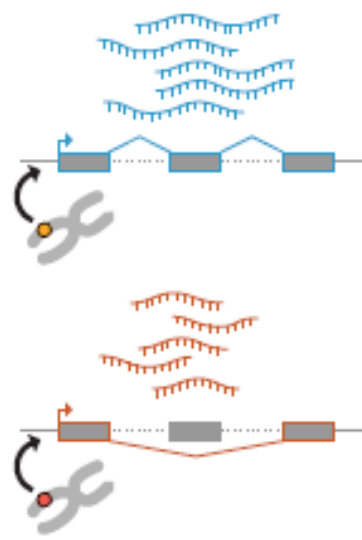


Splicing quantitative trait loci (sQTLs)

cis-sQTLs



trans-sQTLs

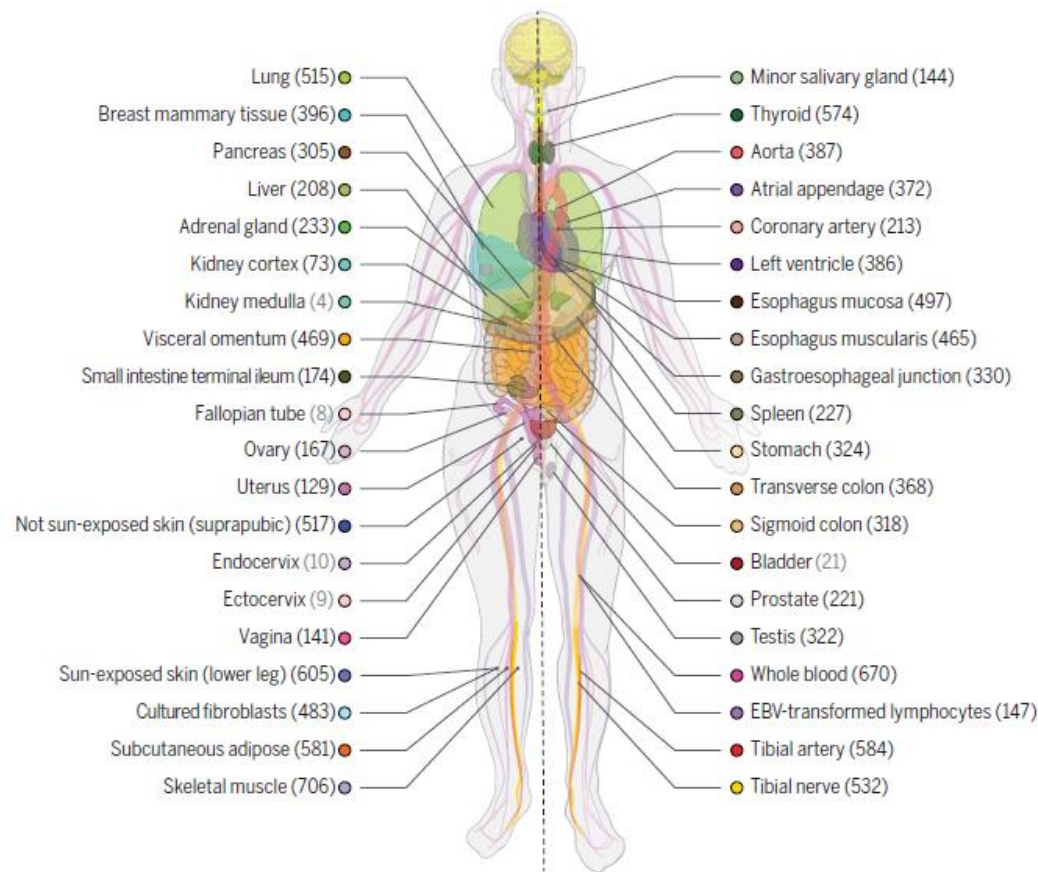
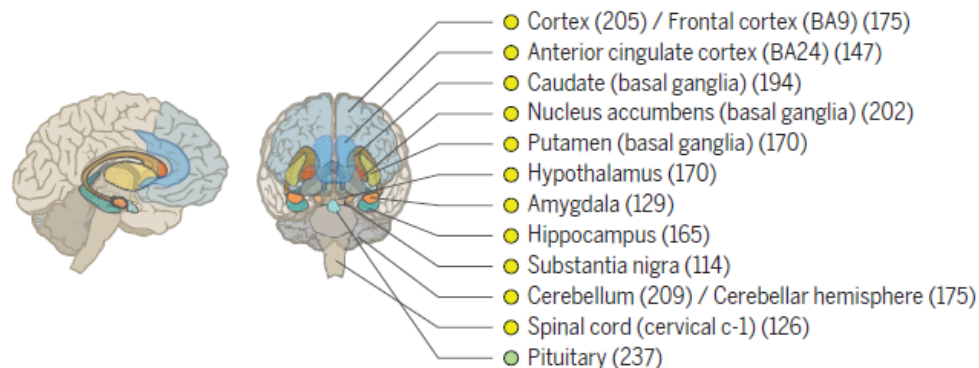


摘要

数据：本研究基于GTEx V8，分析了来自838个遗体、49种组织的15,201个RNA-Seq样本。

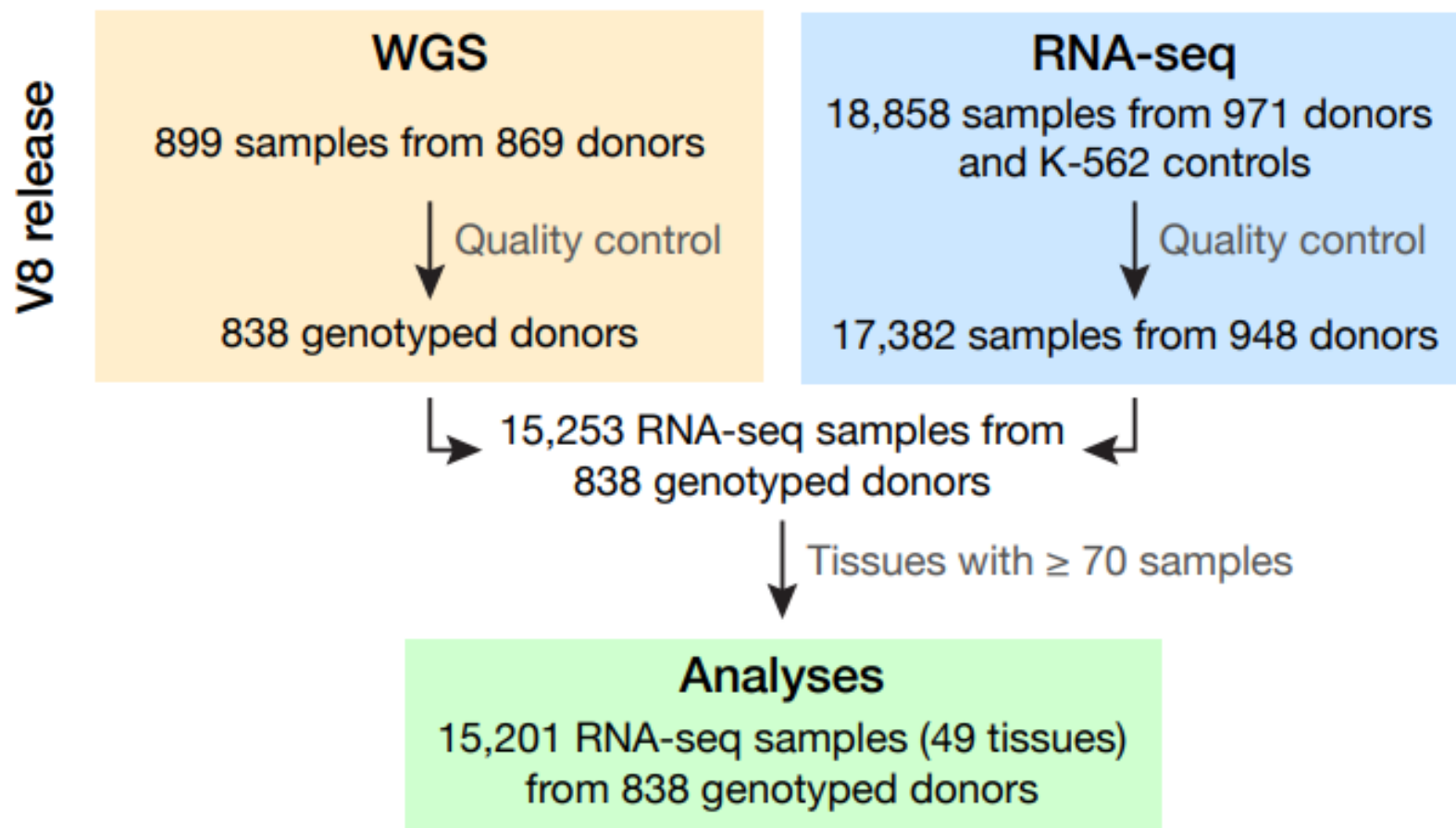
工作：鉴定了顺式和反式的 eQTL、sQTL 的遗传关联，并主要从组织特异性和细胞组成两个方面进行分析。

结论：几乎所有基因都存在调控关联，而且发现了一些潜在分子机制，这些机制对于复杂性状的等位基因异质性和基因多效性具有重要作用。



数据筛选流程

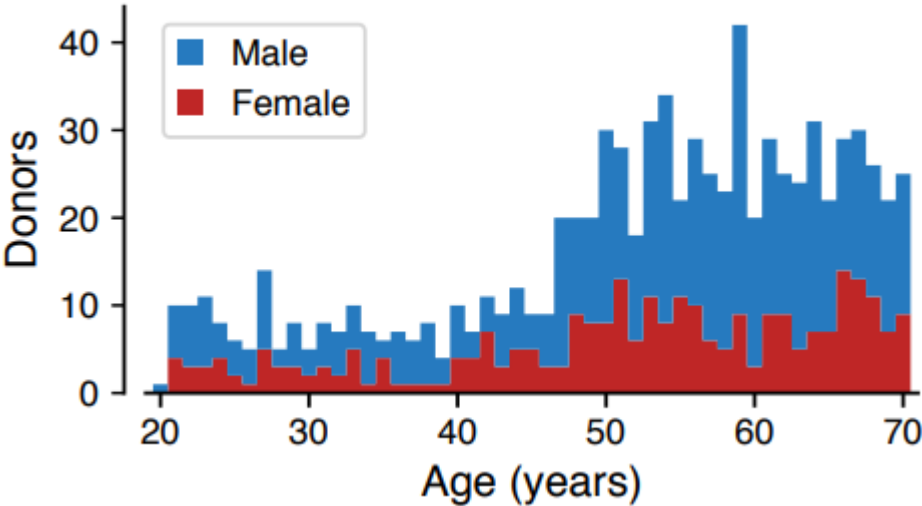
至少70例个体同时拥有 RNA-seq 和 WGS基因型数据。



数据筛选流程

年龄性别分布：20-70岁。中老年和男性居多。

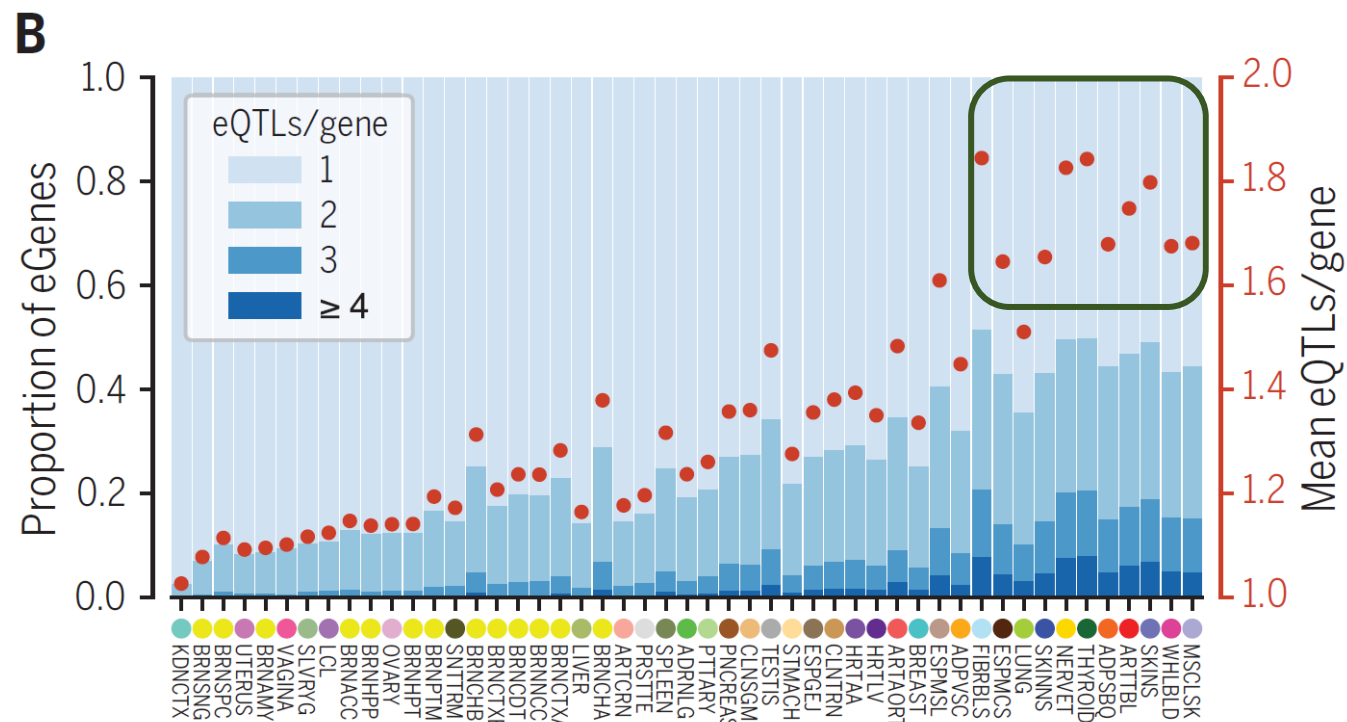
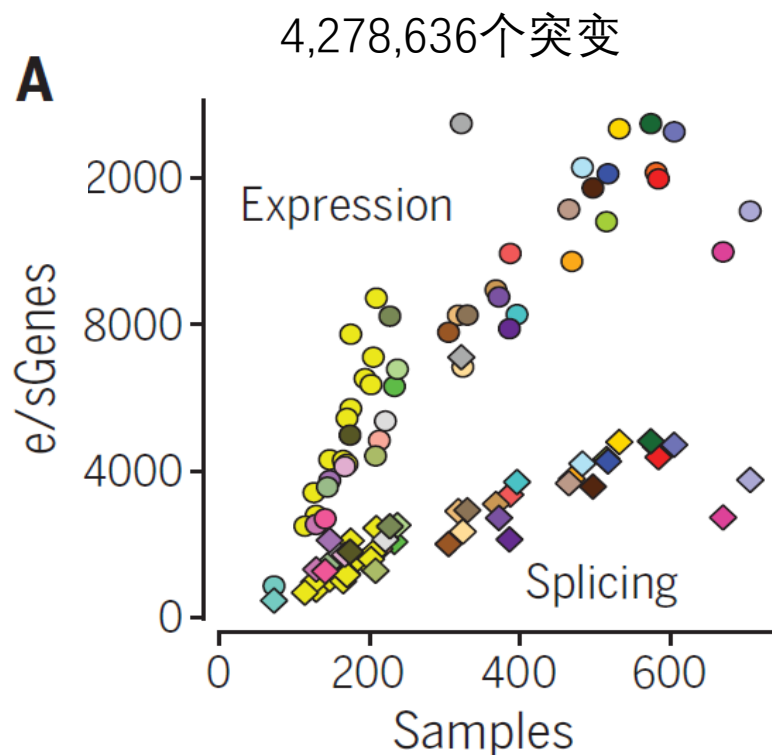
人种分布：主要为欧裔美国人，亦包含亚裔，非洲裔，拉丁裔，印第安人。



<div><div></div><div>Ethnicity</div></div> <div>Race</div>	White	Black or African American	Asian	American Indian or Alaska Native	Unknown	All
Not-Hispanic or Latino	220 119	45 25	4 4	1 0	0 0	270 148
Unknown	224 113	23 7	2 1	1 0	0 0	250 121
Not reported	21 9	2 0	1 0	0 0	0 0	24 9
Hispanic or Latino	7 2	1 0	0 0	0 0	5 1	13 3
All	472 243	71 32	7 5	2 0	5 1	557 281

Male|Female

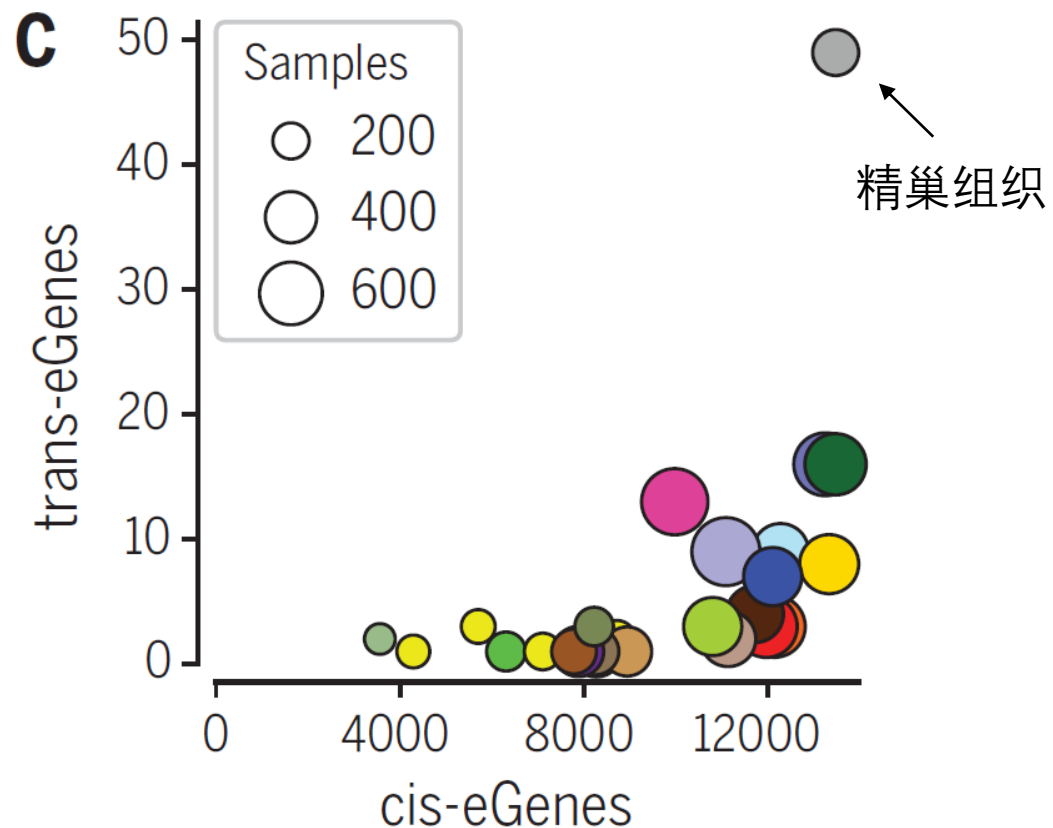
样本量与cis-QTLs数量相关



1. 样本量 ↑, 发现的cis-e/sQTL相关基因 ↑;
 2. 样本量 ↑, 每个基因所关联的cis-eQTL ↑;
 3. 样本量较大的组织中, 几乎每个eGene平均具有>1.6 个eQTL。
- } 增大样本量有利于促进更多发现

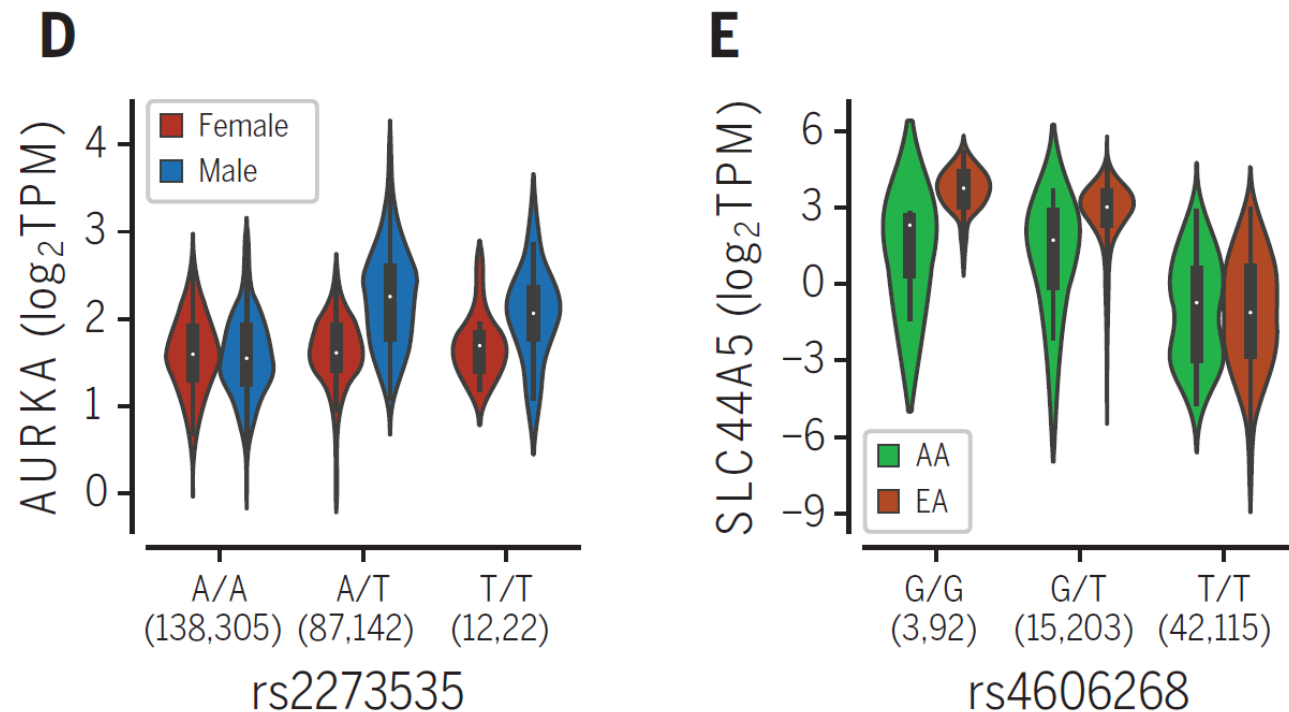
等位基因异质性: 指相同基因座上的基因发生不同突变所引起的相同或者相似的表型。

样本量和cis-QTL与trans-QTLs数量相关



1. 样本量和cis-eQTL相关基因 ↑, trans-eQTL相关基因 ↑
2. 精巢的trans-eQTL相关基因远远多于其他组织。

性别/群体偏向性cis-QTLs



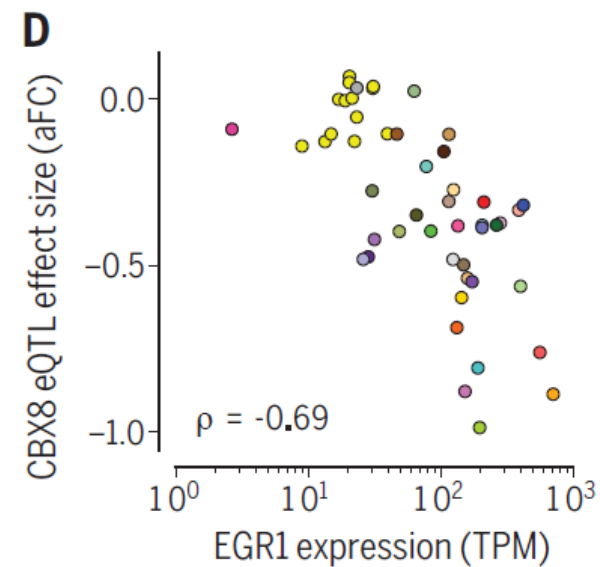
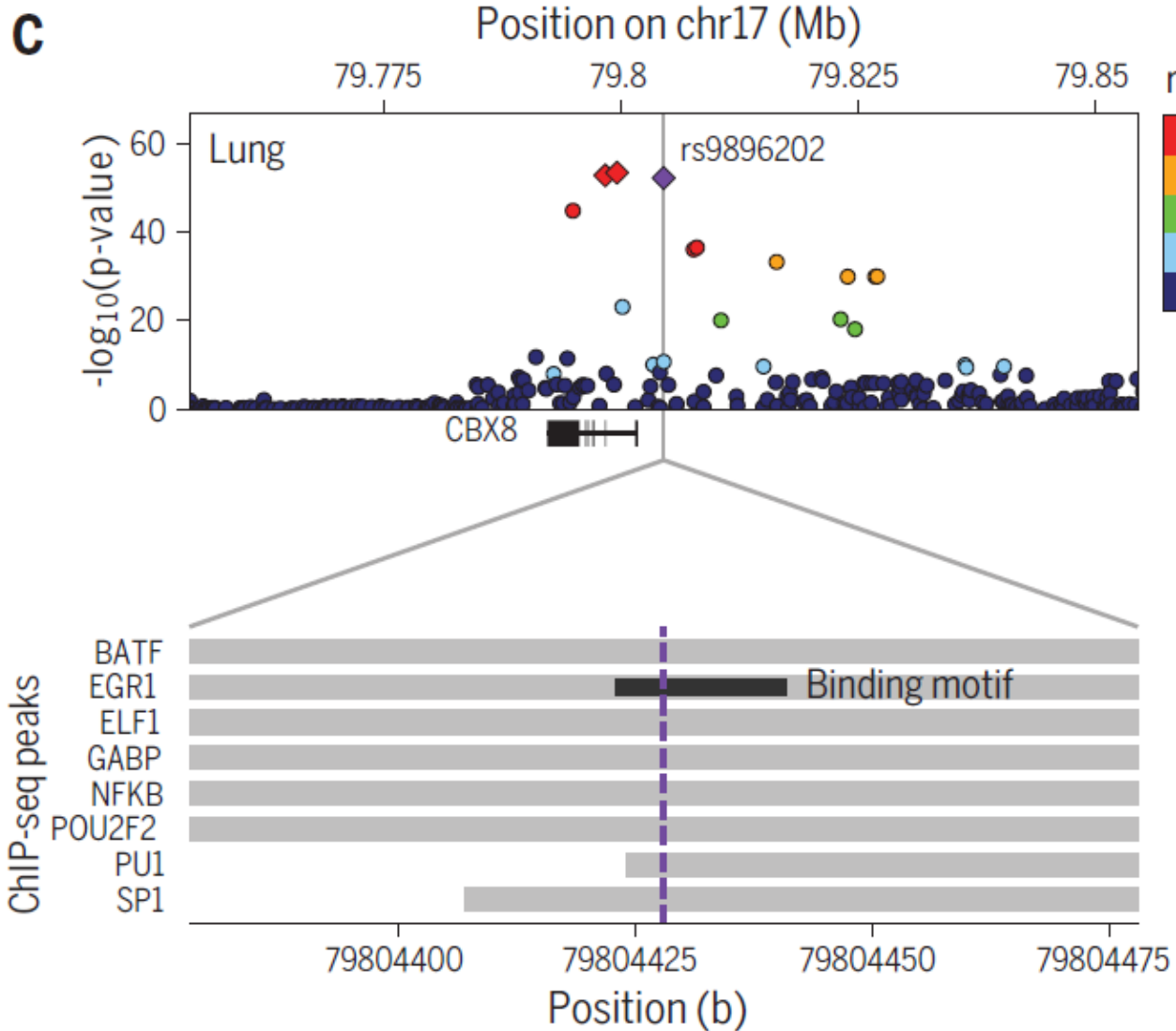
两个cis-QTLs的实例展示

1. 在骨骼肌内，rs2273535变异可以顺式关联*AURKA*基因的性别偏向性表达；
2. rs4606268变异可以顺式关联*SLC44A5*基因的群体偏向性表达
AA: 非裔美国人
EA: 欧裔美国人

* 具有主要调控能力的变异不多，往后需要更大的数据量来做挖掘。

在FDR-pvalue < 25%的宽松条件下，只能找出极少量的性别/群体偏向性 cis-eQTLs（百分之几的发现率）。

eVariant 起调控作用的实例



基因 *CBX8* 的cis-eQTL 信号 (p-value)

↓

找到显著的 fine-mapped 变异rs9896202

↓

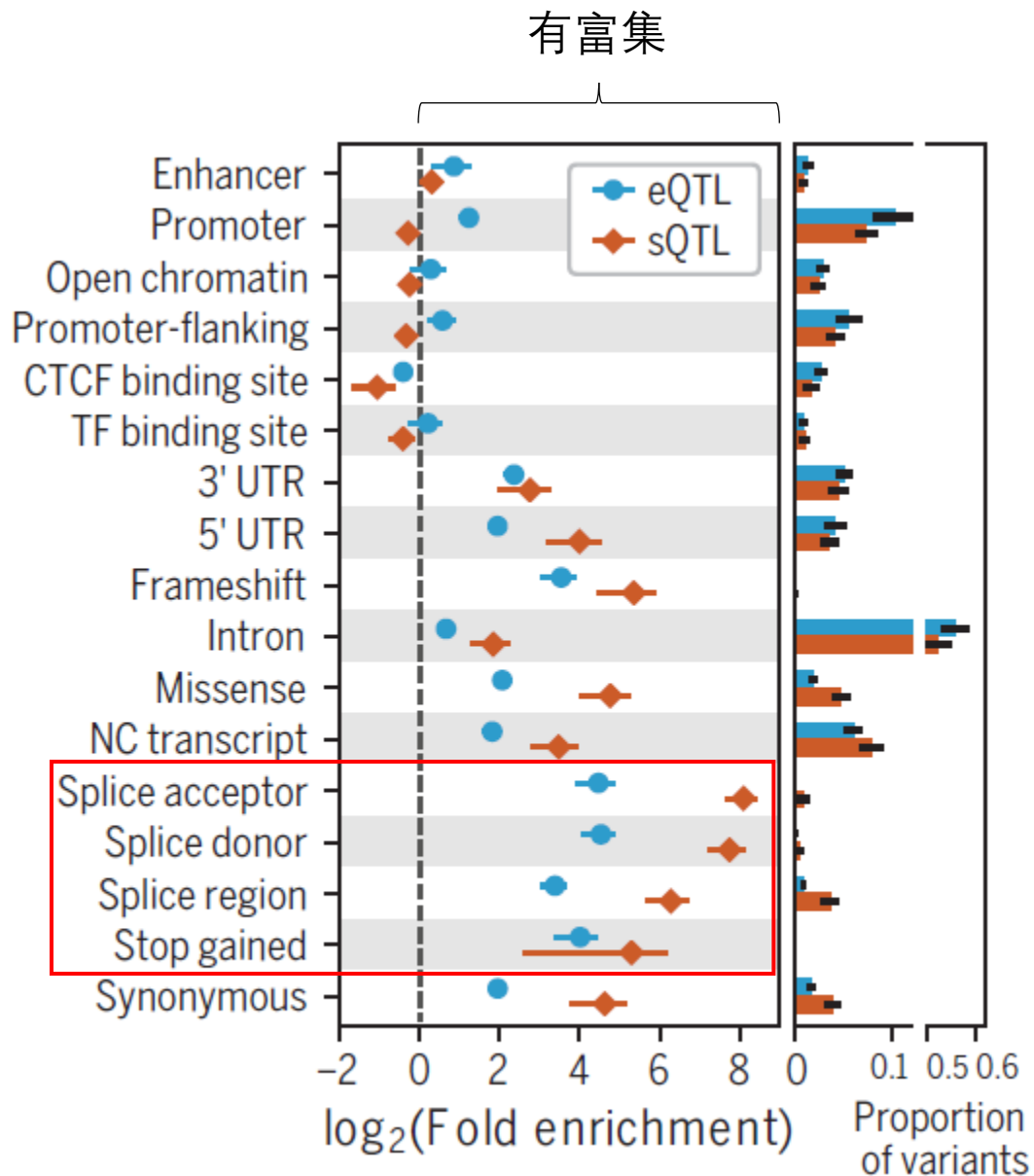
rs9896202干扰转录因子EGR1的结合

↓

EGR1表达量与*CBX8*的eQTL有效量负相关

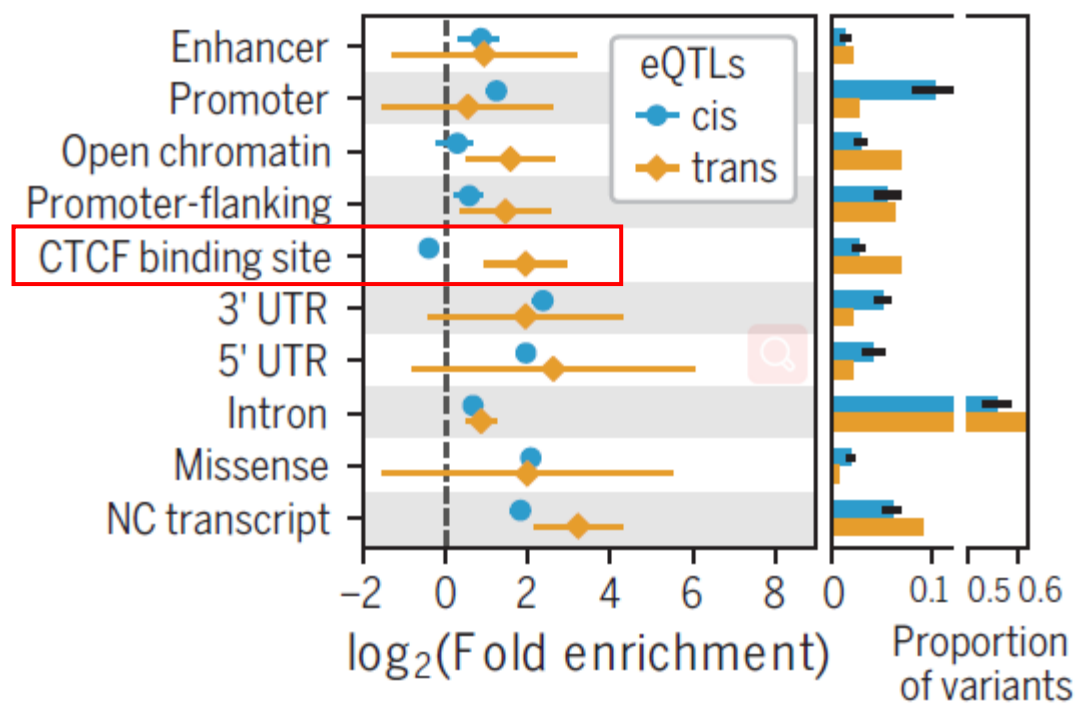
因果性确实存在!

cis-QTLs在基因组的富集情况



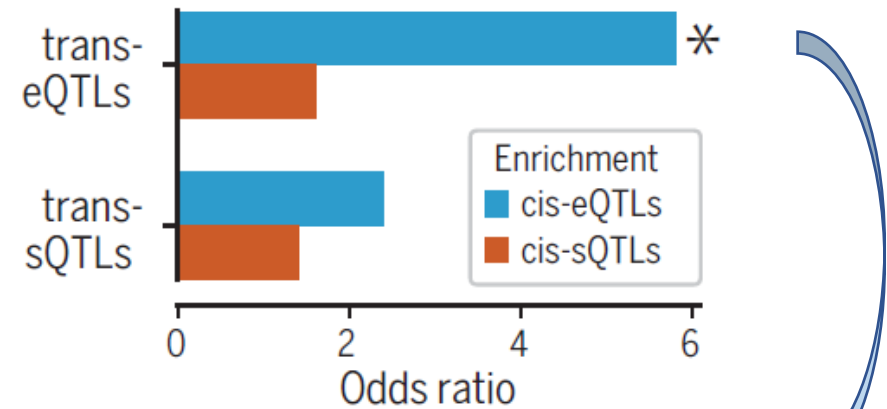
1. cis-eQTLs 与 cis-sQTLs 在功能元件中富集，尤其是剪接相关的元件；
2. cis-eQTLs 在转录区、转录调控区 (enhancer, promoter等) 均有富集，而cis-sQTLs 几乎只富集于转录区；

trans-QTLs在基因组的富集情况

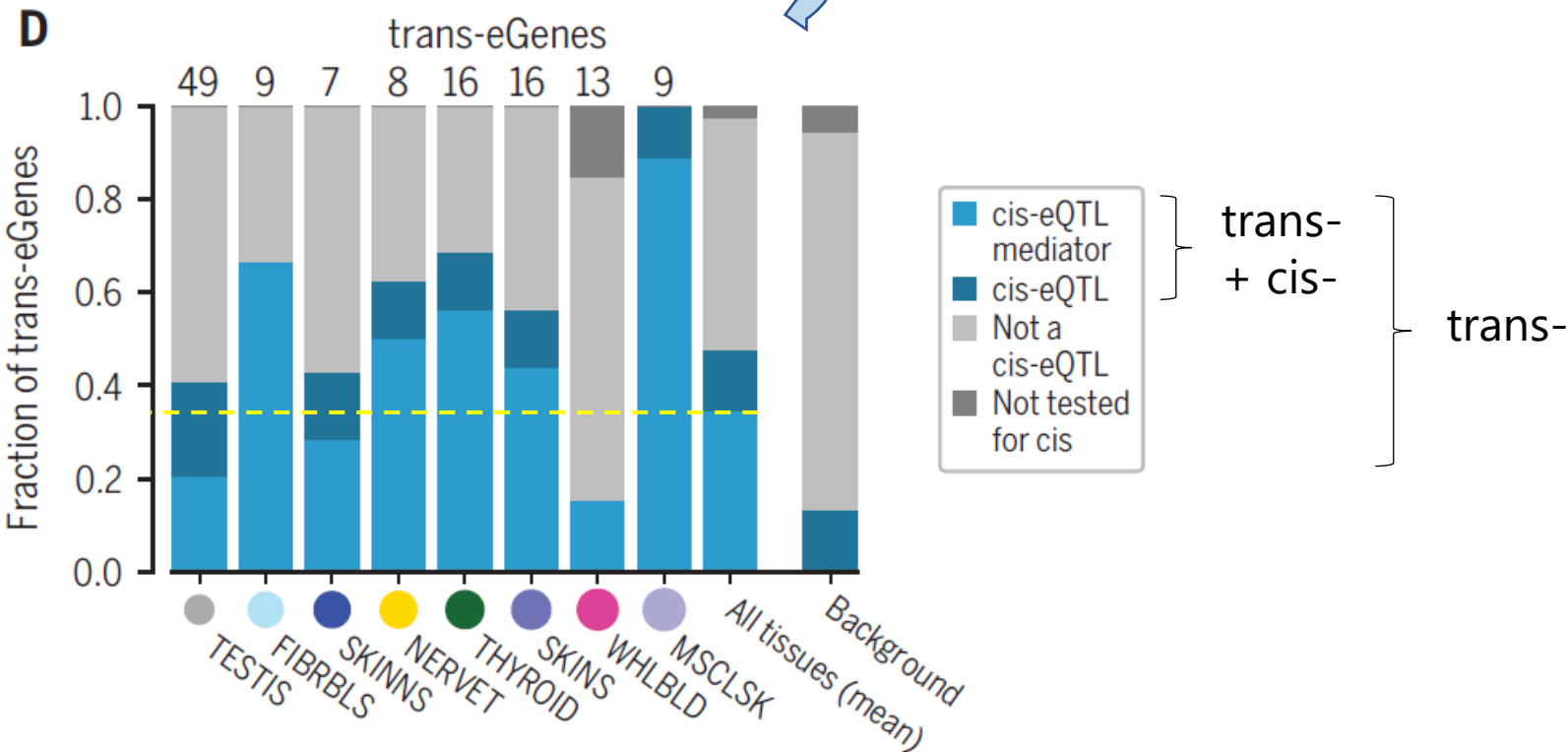


1. trans-eQTLs 富集于多种转录调控元件;
2. 尤其富集于CTCF binding site, 与cis-eQTLs形成鲜明对比。该元件可能与跨染色体的交互作用有关。

trans-与cis-QTLs的关联

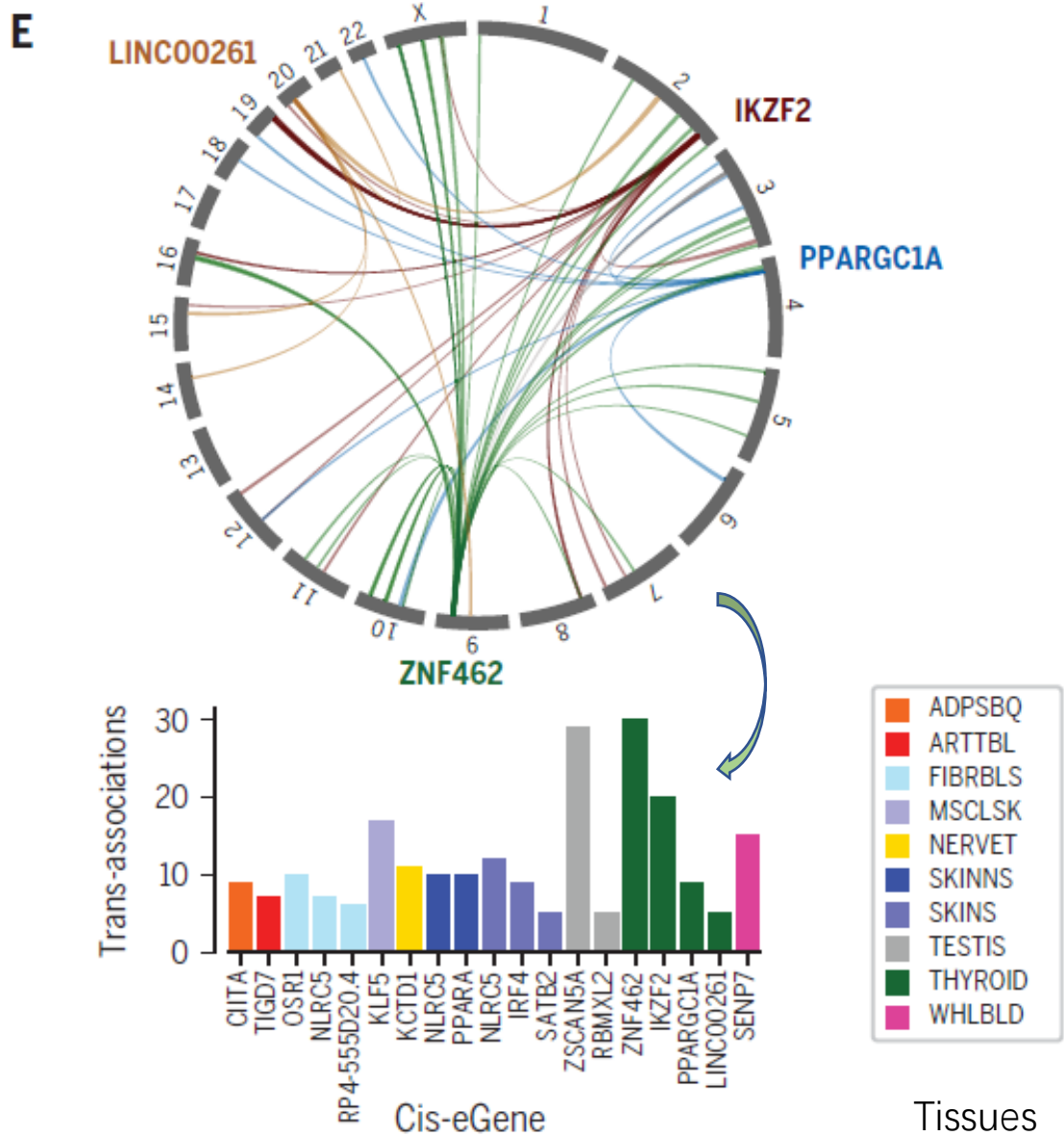


1. trans-eQTLs与cis-eQTLs的富集具有相关性;



77%的 trans- + cis-eVariants (相当于31.6%的trans-eVariants) 受到 cis-eQTL的介导。

trans-与cis-QTLs的关联



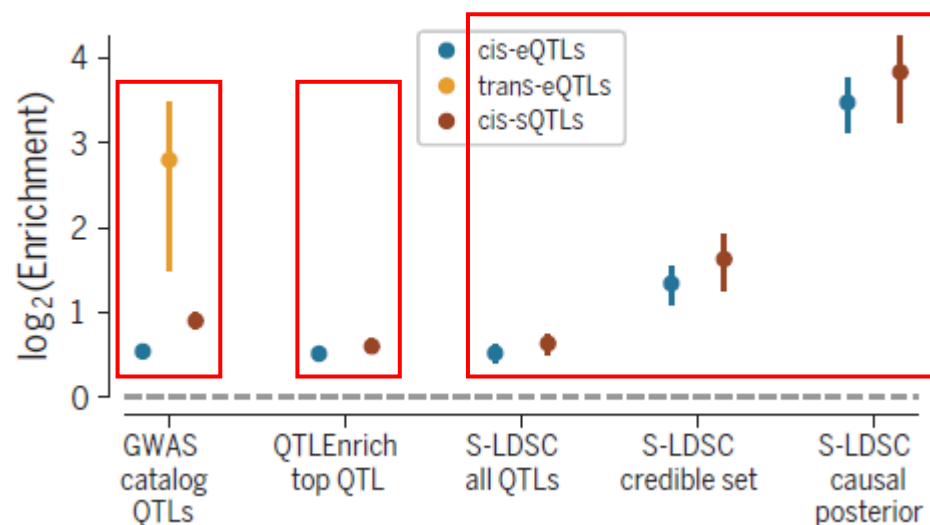
2. cis-eQTLs 与 trans-eQTLs 的关联广泛存在, 且具有组织特异性。

◀ 四种甲状腺 cis-eQTLs 的 trans-关联

◀ 关联了 ≥ 5 个 trans-eQTLs 的 cis-eGenes

探究QTLs与人体性状的相关性

A



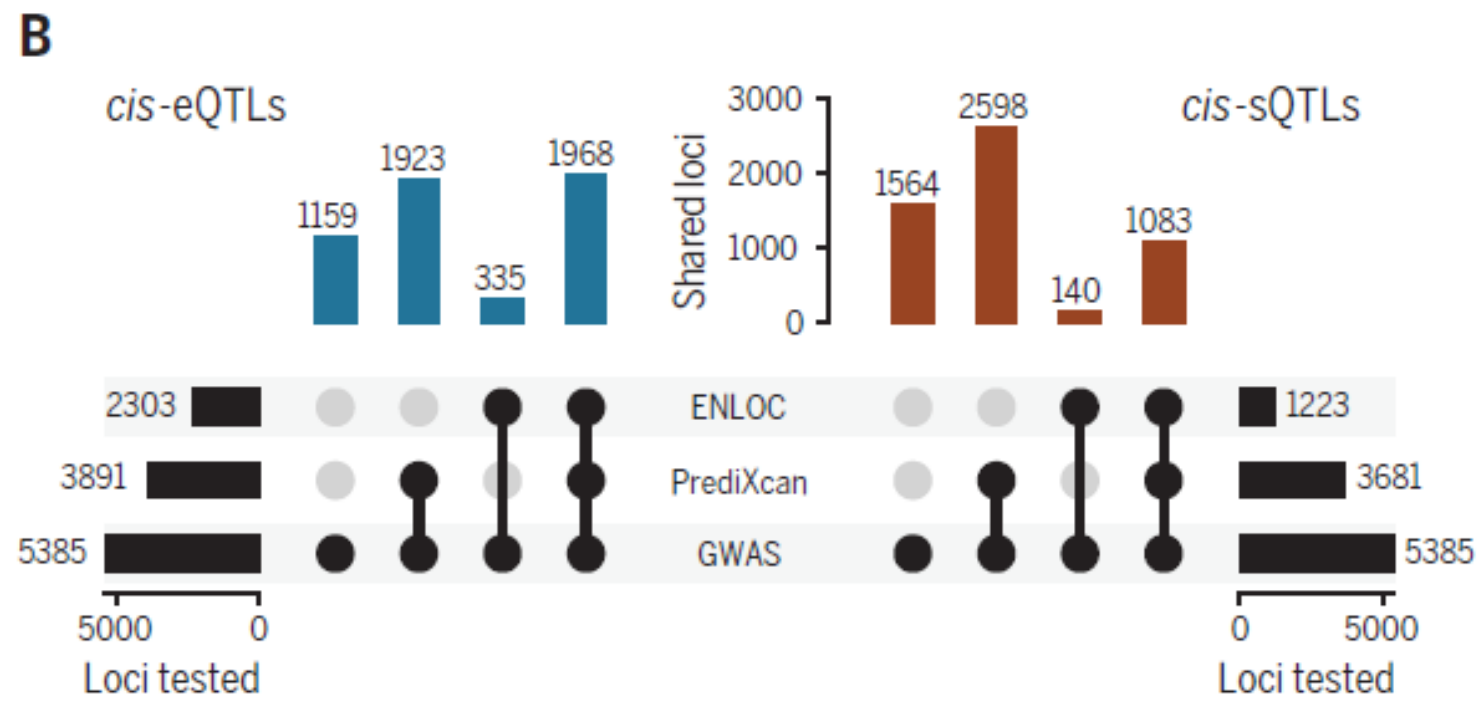
分析QTLs与生物体复杂性状的关联的两种互补方法。

- QTLEnrich：排除混淆因子（比如cis-eQTL的等位基因频率、TSS距离、LD水平）
- S-LDSC：确认结果的稳健性，评估因果变异对GWAS的影响

GWAS找到的variants是否可以富集于显著的QTLs？

1. GWAS的结果在三类QTLs，尤其是trans-eQTLs中显著富集；
2. 用QTLEnrich方法进行验证，发现结果基本不变，说明GWAS分析可靠；
3. 用S-LDSC方法进行验证，发现在 fine mapped cis-QTLs 的富集度显著升高，说明 cis-QTL与GWAS的强关联，以及fine mapping的重要性。

探究QTLs与人体性状的相关性

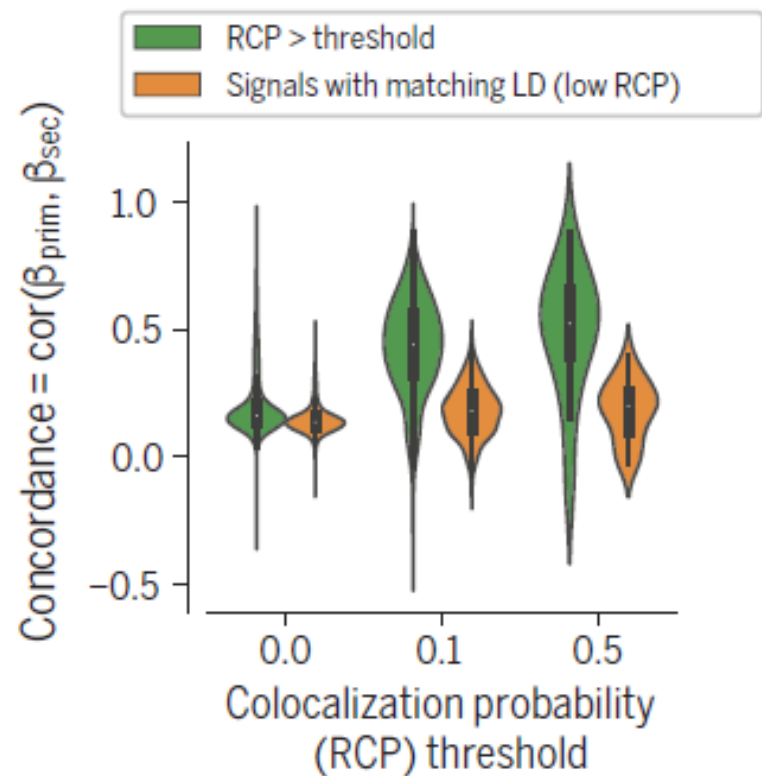


- ENLOC：共定位功能注释
- PrediXcan：基因表达和剪切关联

使用两种数据集进一步验证与GWAS富集关联QTL的可靠性。

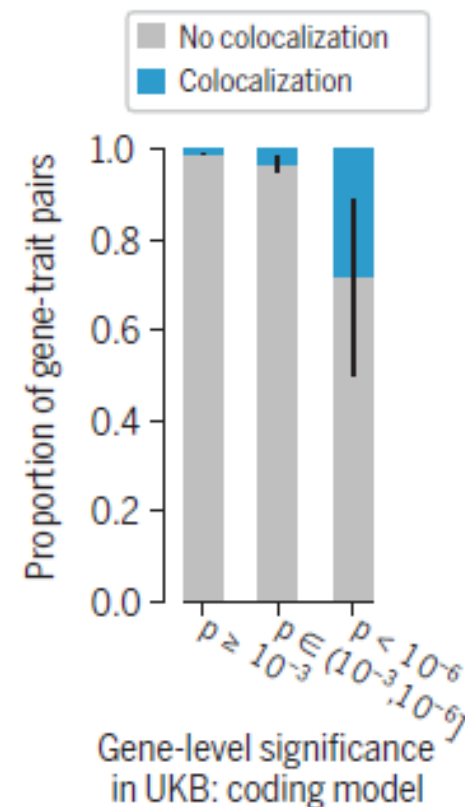
探究QTLs与人体性状的相关性

C



- 共定位无LD eGene基因集比非共定位LD基因集的初级（GWAS）和次级（转录组）cis-eQTL信号的一致性更高。说明多个**cis-eQTL**可能调控同一个基因，从而调控同一个复杂性状。

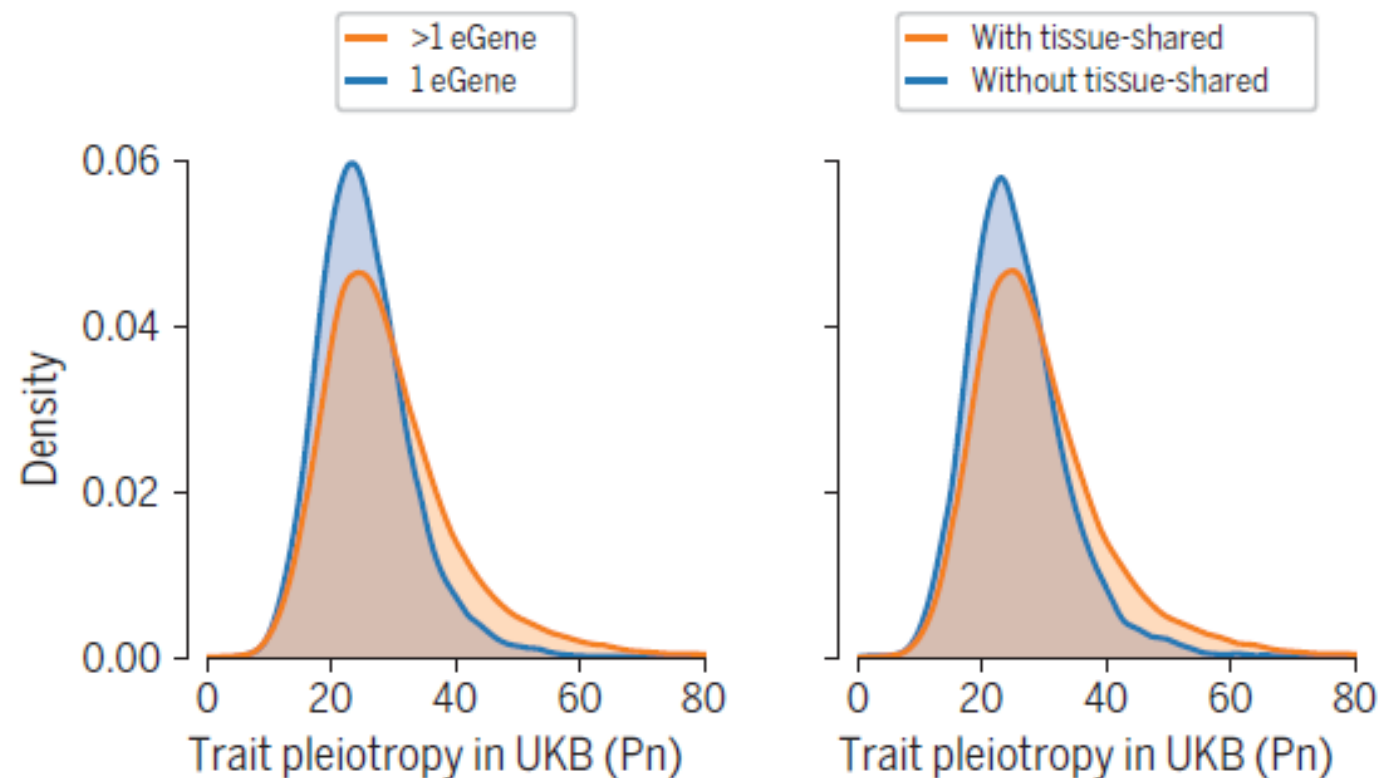
D



- 带有罕见突变的基因，显著有大量共定位的cis-eQTL，证明罕见与常见的cis-eQTL具有相似的调控机制。

探究QTLs与人体性状的相关性

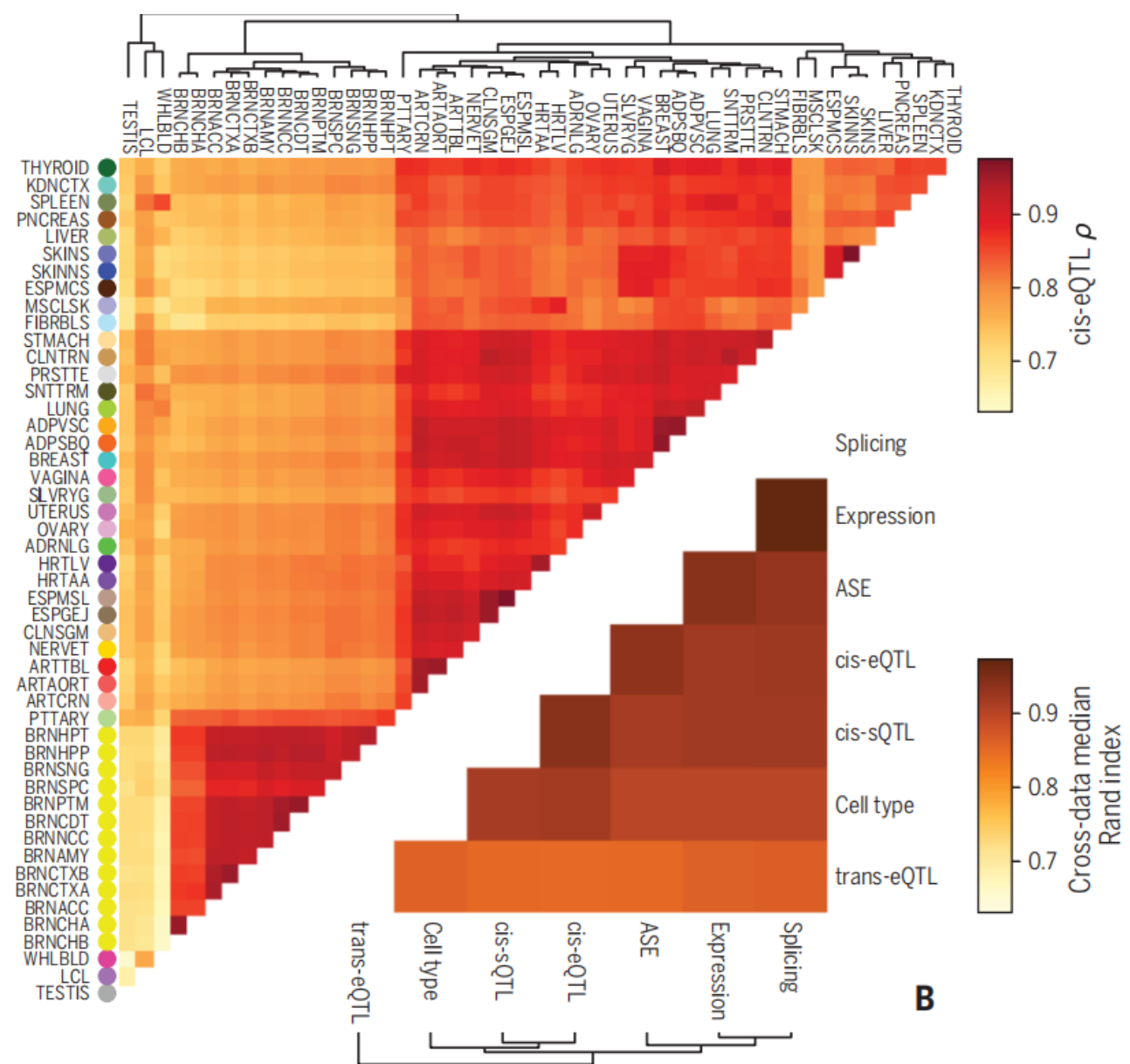
E



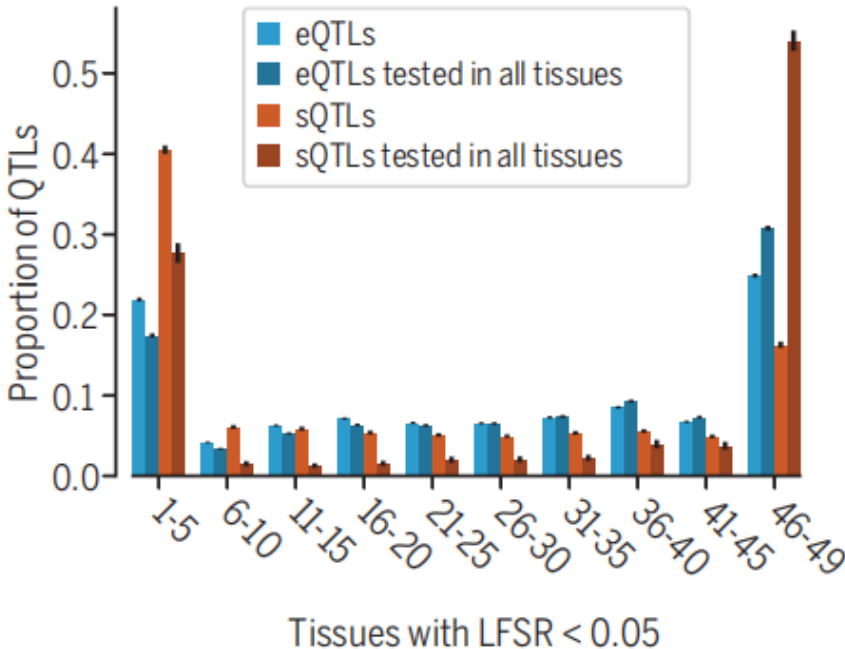
1. 调控多个基因的cis-eQTL比调控单个基因的cis-eQTL更加具有性状基因多态性，即前者可调控更多性状。
2. 在组织共有的cis-eQTL比非组织共有的cis-eQTL更加具有性状基因多效性，说明影响多个组织的相同调控效应可能会产生不同的生物性状。

基因多效性：同一个位点与多个生物性状有关

组织特异性



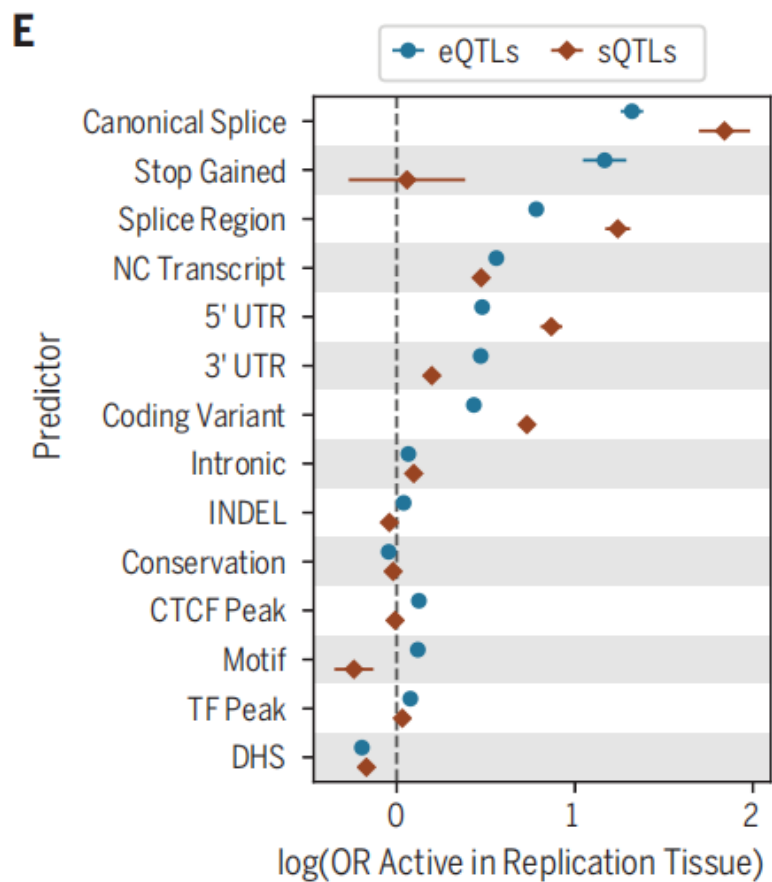
c



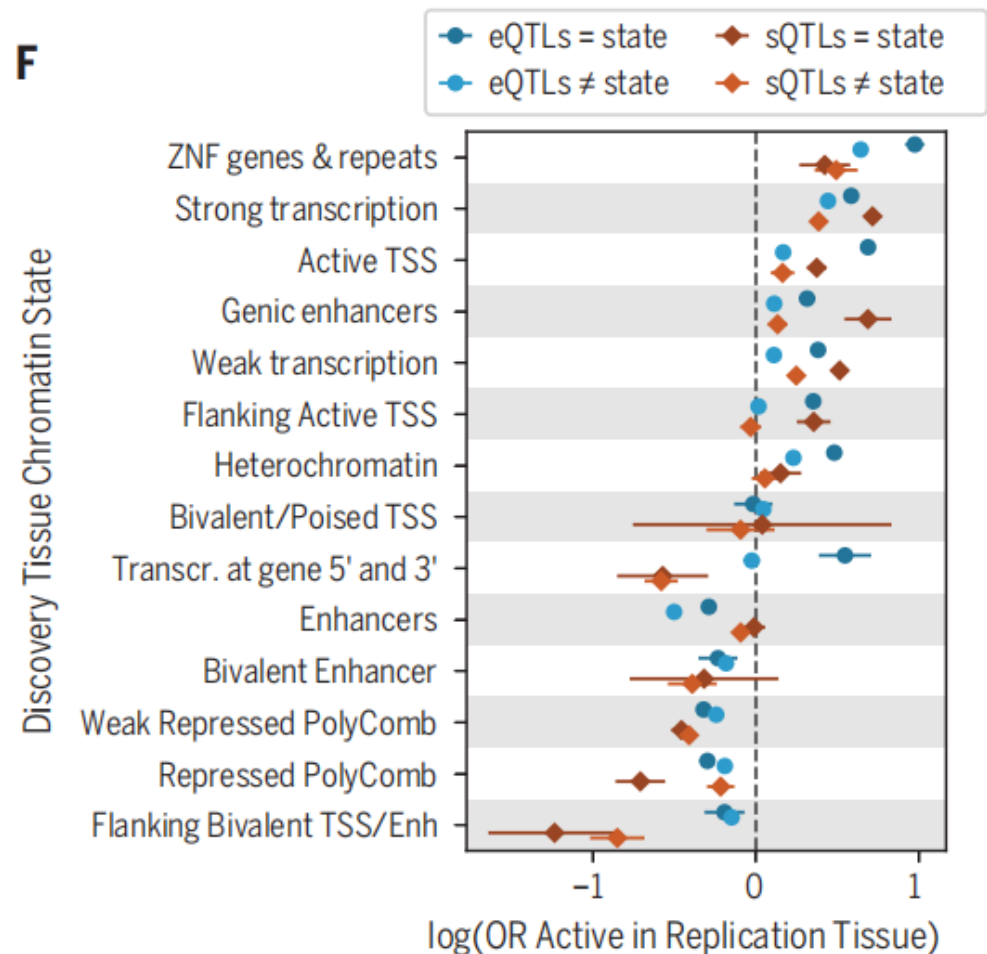
1. 大部分组织共享相似的遗传效应，但脑组织、精巢组织和血液等存在特异性。
2. 剪切、表达、ASE和细胞类型与各种QTL存在相关性。

组织特异性

形成组织特异性的可能机制



- 转录后或共转录遗传调控

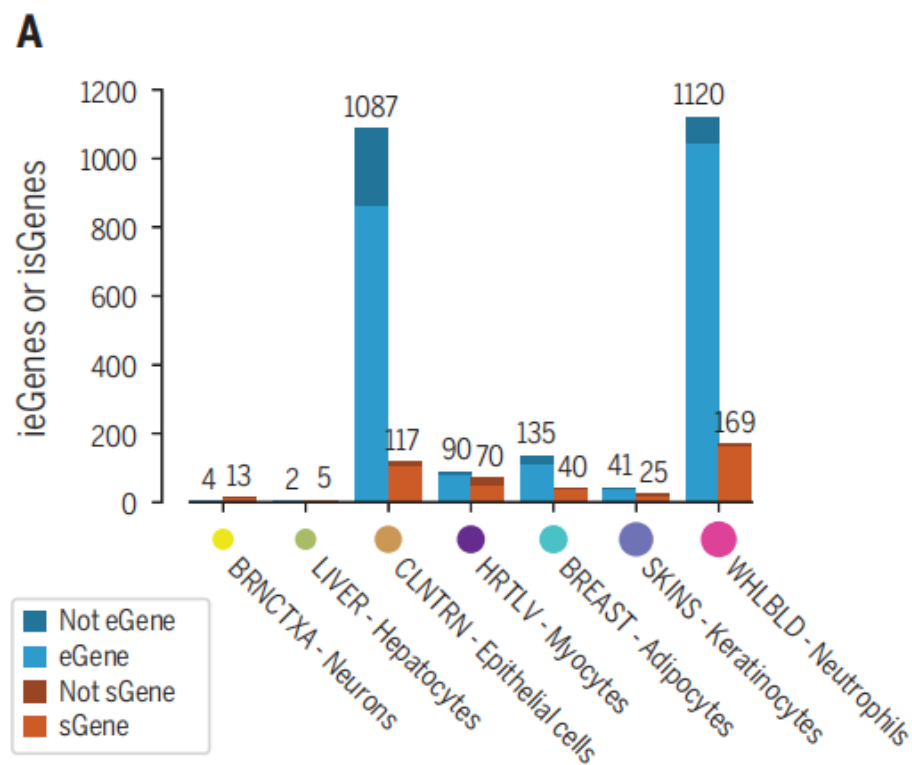


- 染色体状态

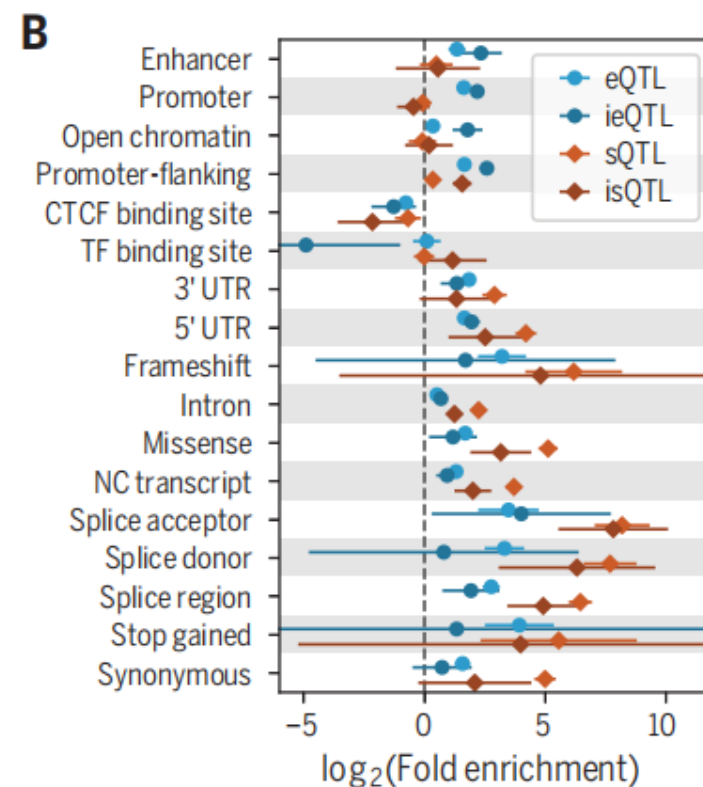
细胞类型特异性

方法: xCell, 解卷积计算样本中细胞类型的富集程度

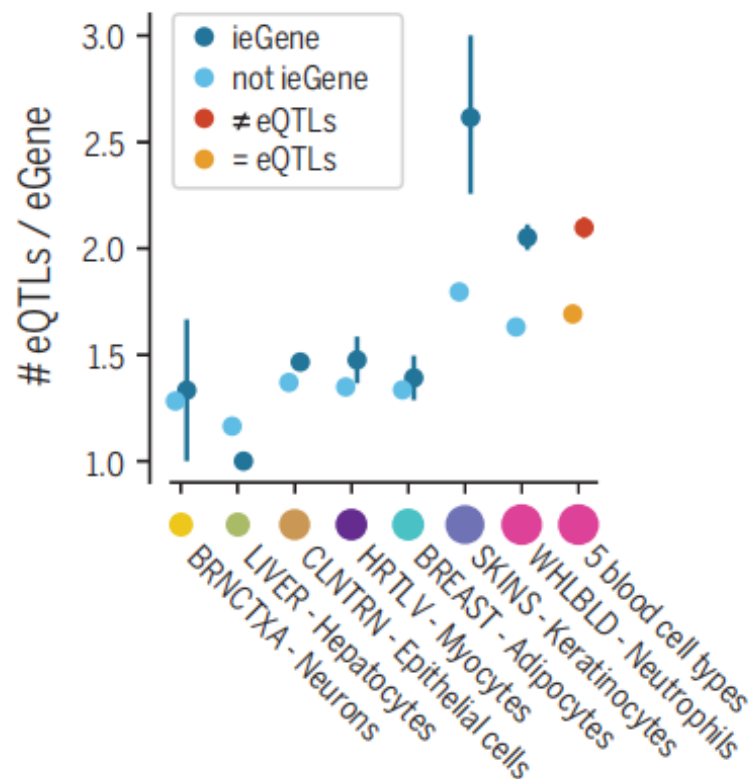
iQTL: 表征细胞类型与基因型之间的存在互作关系的QTL



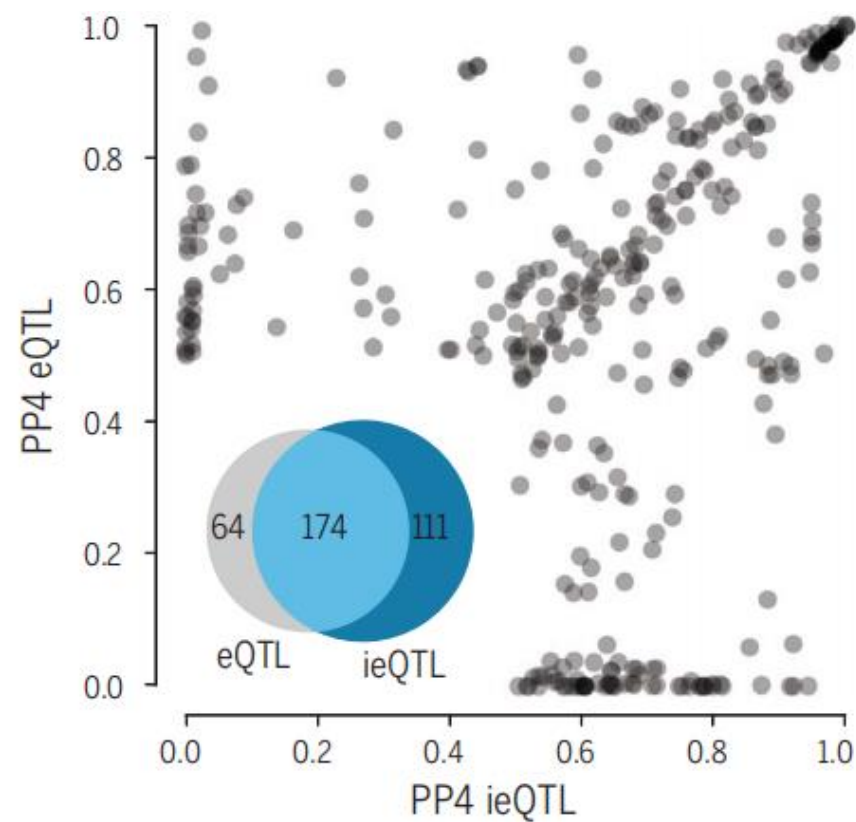
1. 存在不少区别于标准 eGenes/sGenes 的 ieGenes/isGenes
2. 它们的功能富集结果几乎没有差别



细胞类型特异性



- ieGene比eGene具有更多的eQTL;
- 不共享的eQTLs与更多基因关联



- 与GWAS共定位的eQTL与ieQTL大部分相同
- 存在不少特异的ieQTL

QTL从组织的水平到细胞类型的水平

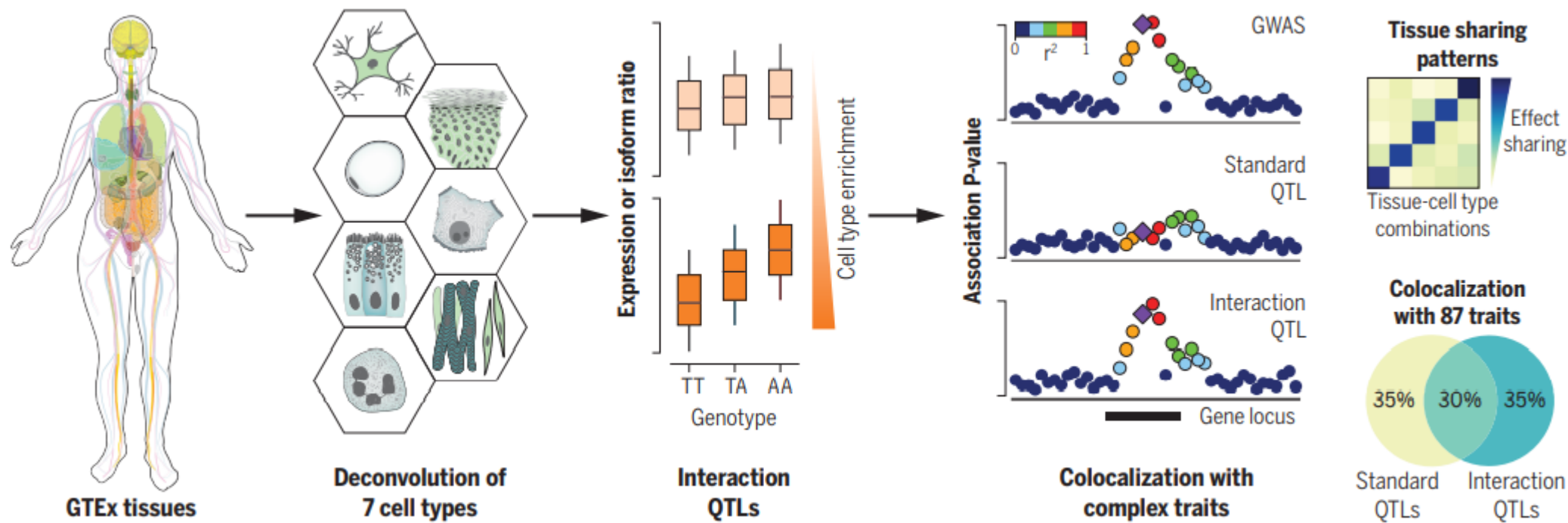
47.728 > Science. 2020 Sep 11;369(6509):eaaz8528. doi: 10.1126/science.aaz8528.

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},

团队的通讯Tuuli Lappalainen是Geuvadis项目RNA测序工作包分析小组的负责人，也是GTEx项目和1000 Genomes项目分析小组的成员。

CONCLUSION



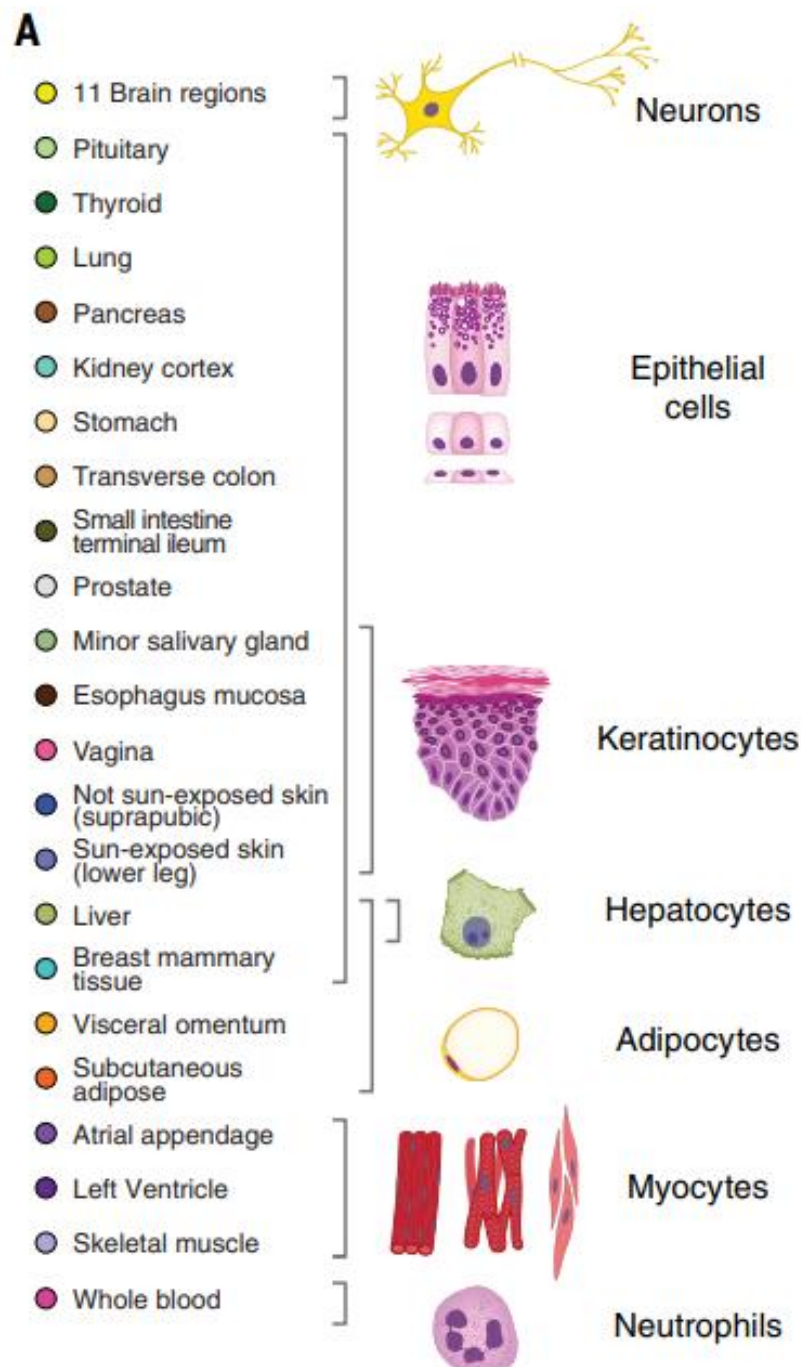
- 1, 鉴定了七种细胞类型的细胞类型相互作用QTL (iQTL)
- 2, 表明了ieQTLs比大量组织cis-eQTLs提供了更好的组织特异性分辨率
- 3, 对87个复杂性状进行的遗传关联分析 (GWAS) 显示了iQTL的作用
- 4, 发现了大量组织测序中数百个以前未被鉴定的共定位基因座, 并且预示了将有更多的iQTL等待发掘

METHOD

◆ **Data:** GTEx项目的35个组织中的7种细胞类型(脂肪细胞、上皮细胞、肝细胞、角质形成细胞、肌细胞、神经元和中性粒细胞), 总共**42对**细胞-组织的对应组合。

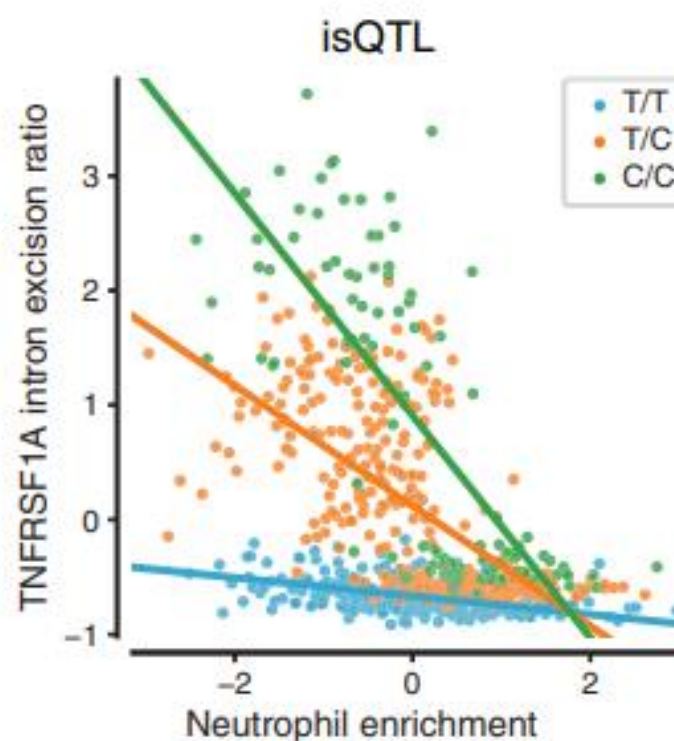
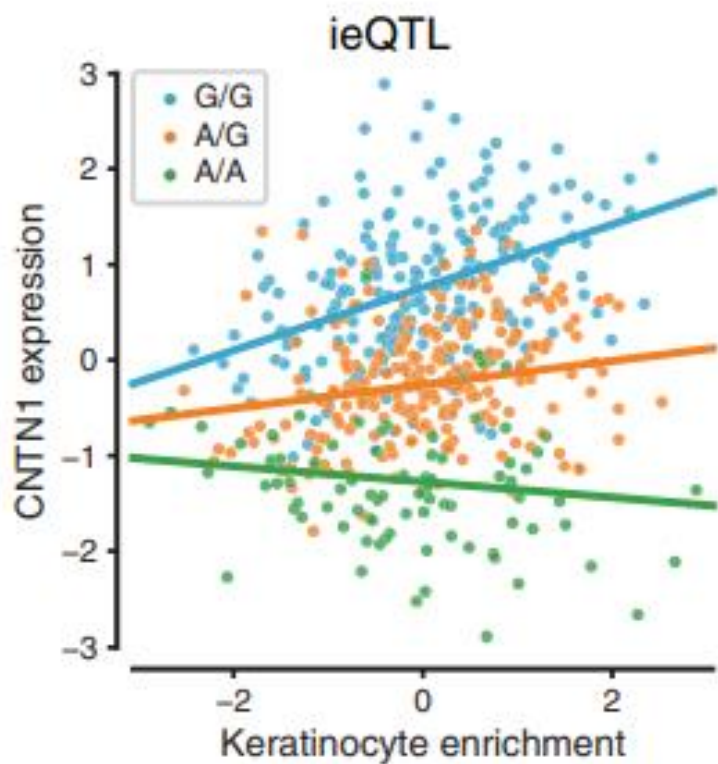
◆ **Method:** 通过**xCell**卷积出细胞块bulk RNA的细胞类型, 然后用**tensorQTL**寻找QTL, 最后GWAS。

? 那为何不直接用近些年很火的单细胞测序呢



RESULTS

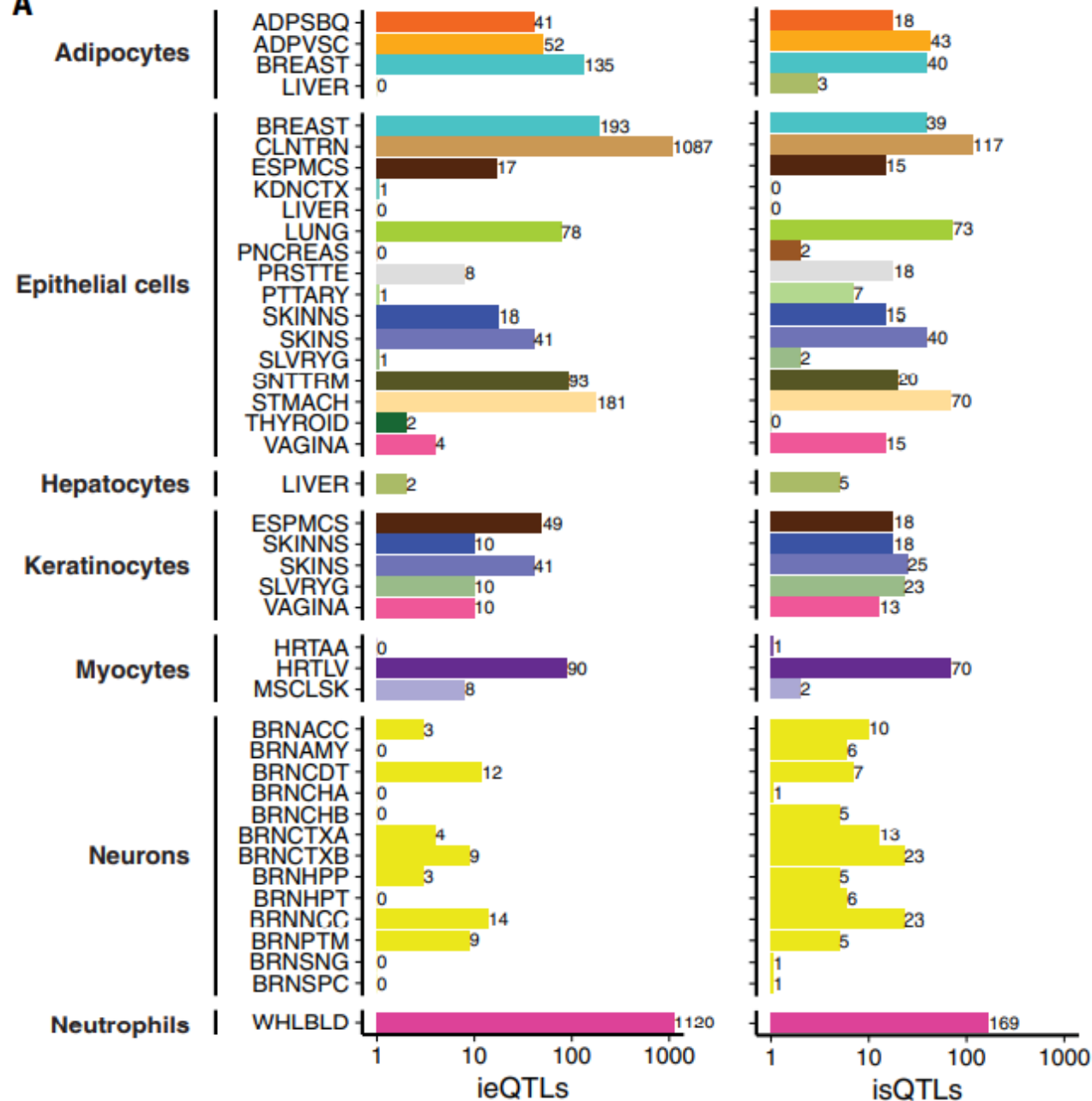
证明ieQTLs和isQTLs的作用是随着细胞类型的富集而变化的



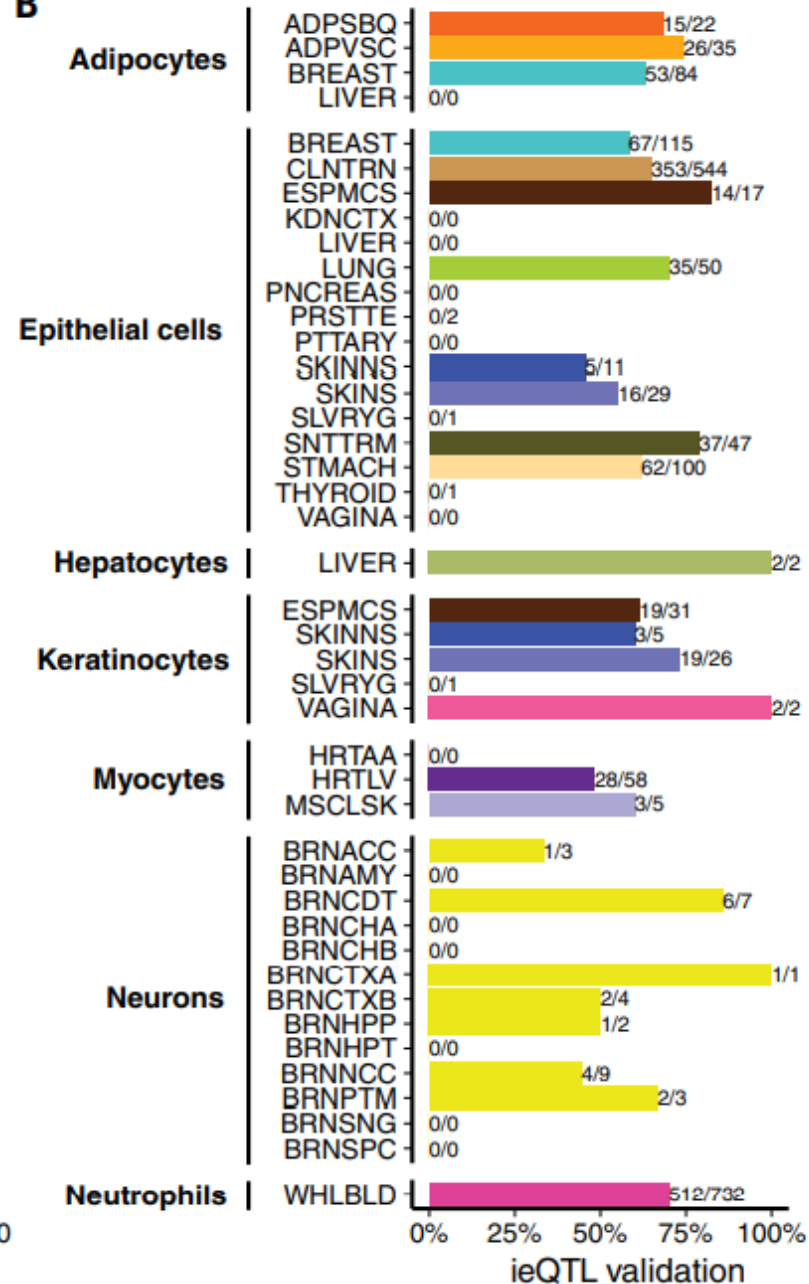
- 分别用角化细胞（keratinocyte）的CNTN1和嗜中性粒细胞（Neutrophil）TNFRSF1A的ieQTL做了线性回归的验证。

RESULT

A

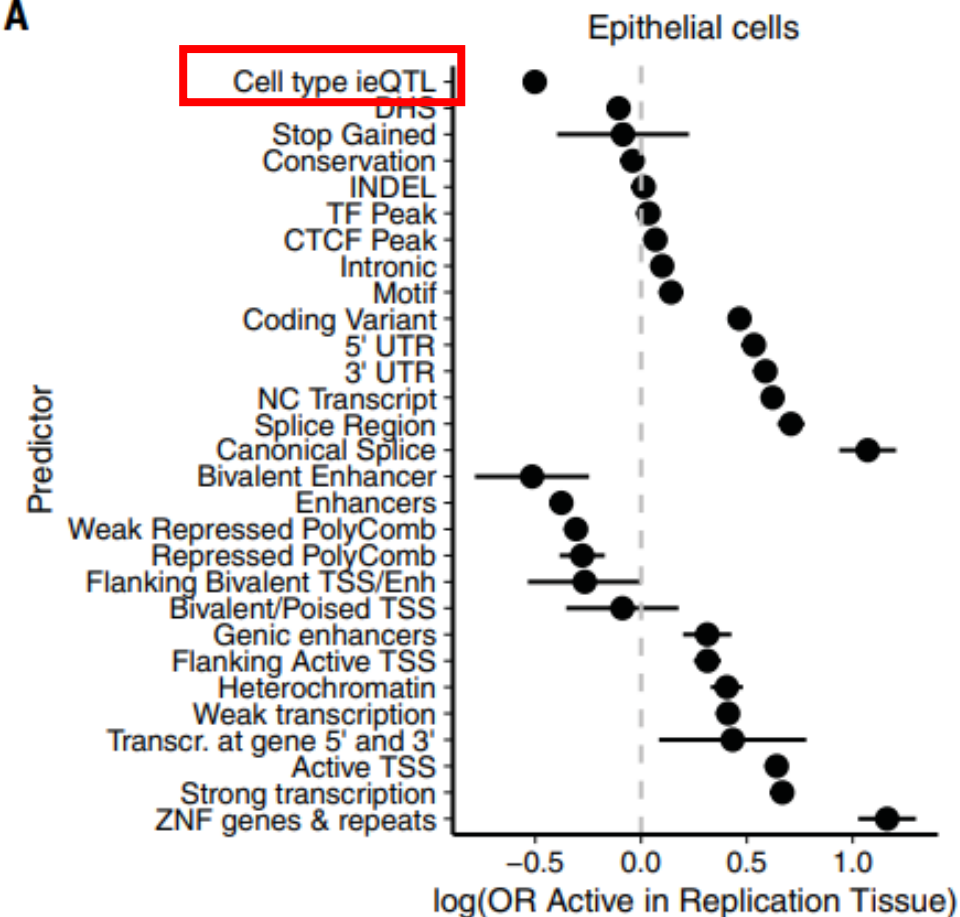


B

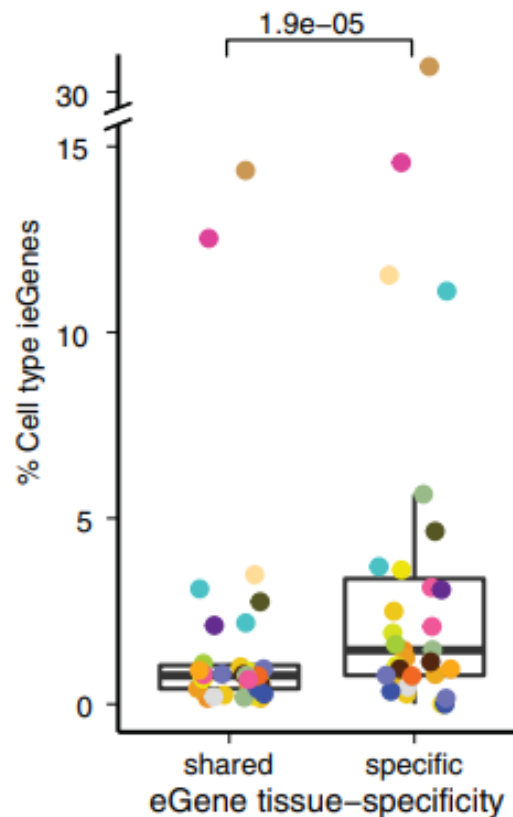


RESULT

A

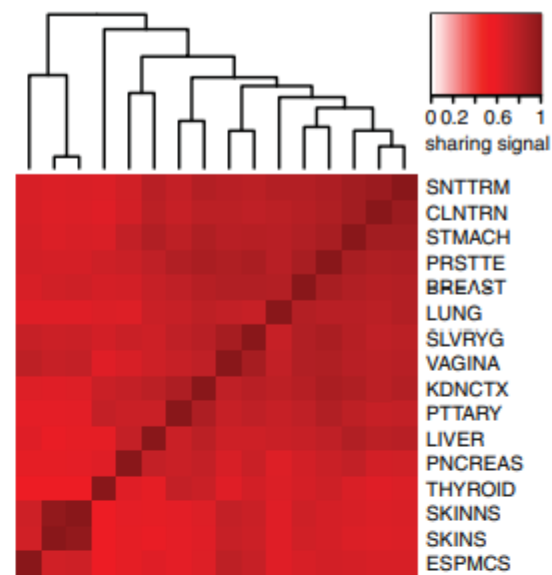
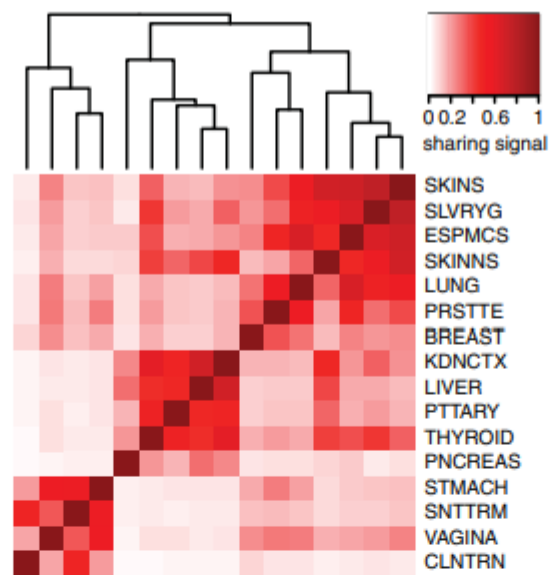
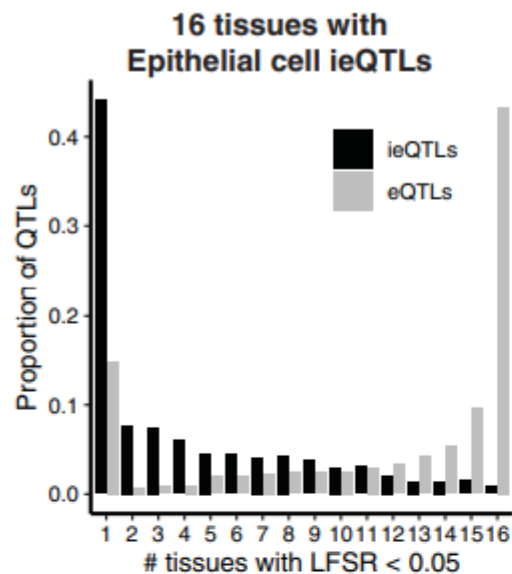


B



RESULT

C



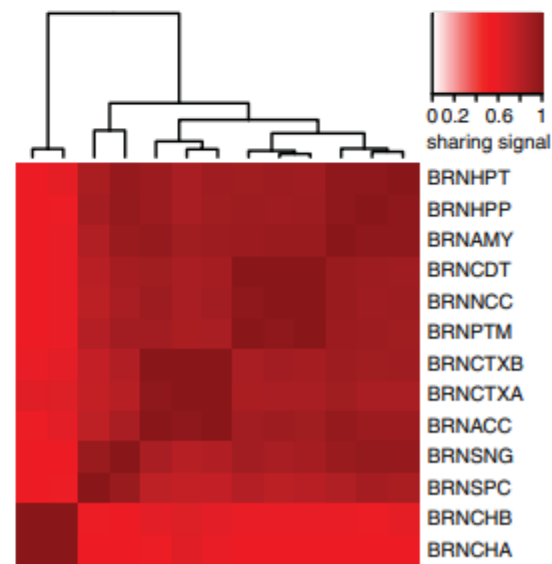
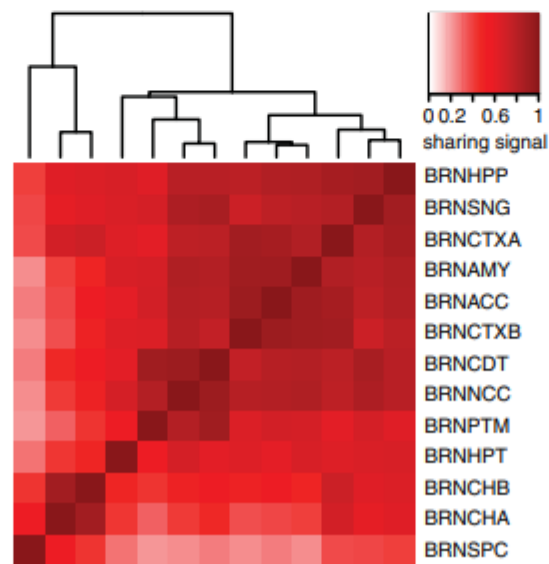
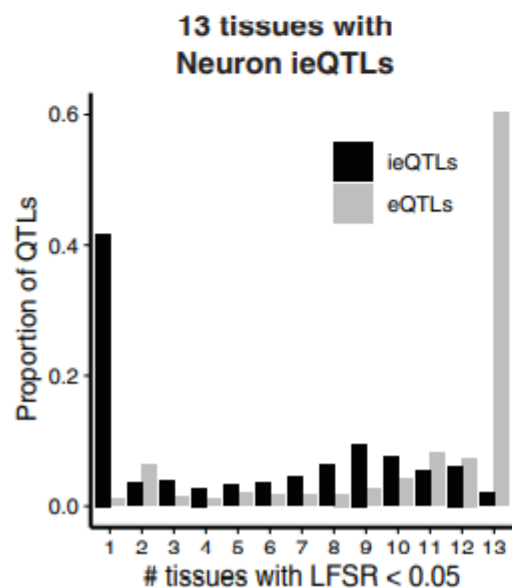
Cell type ieQTLs

top eQTLs

通过ieQTL来区分组织的相似性

(以上皮细胞和神经细胞为例)

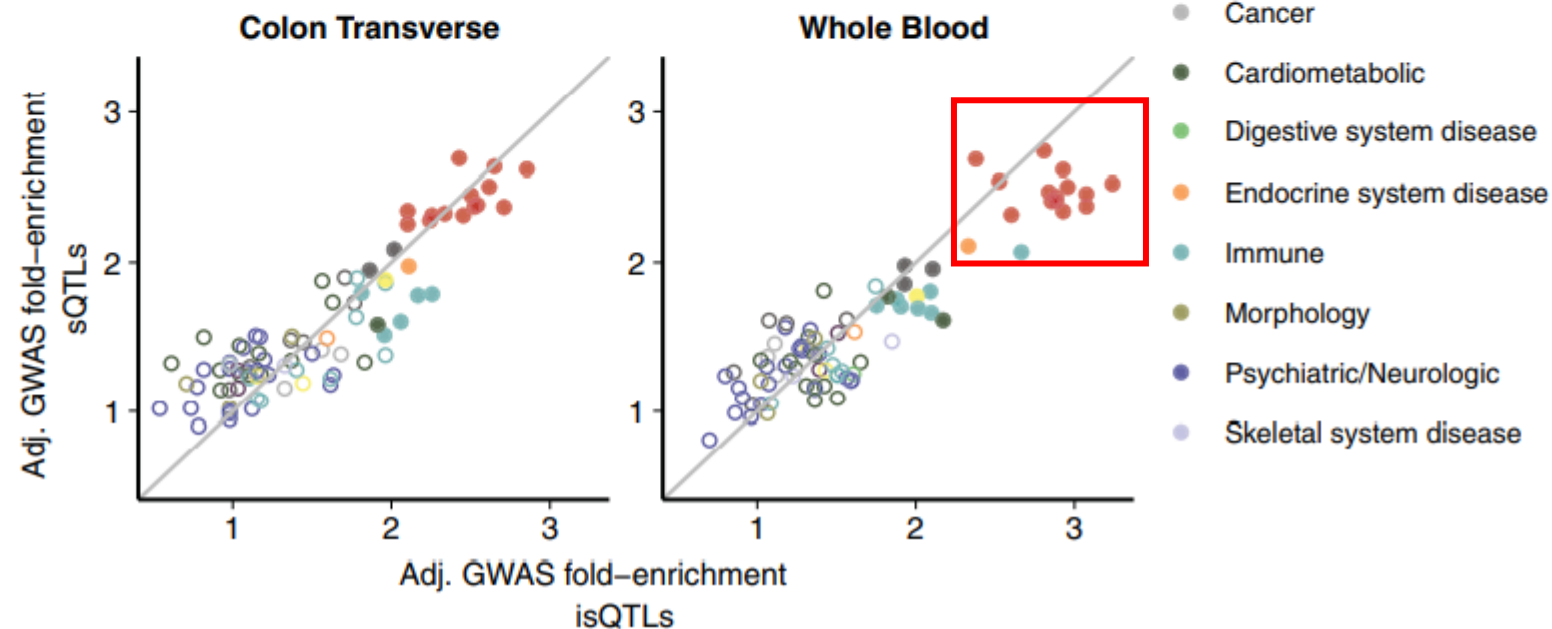
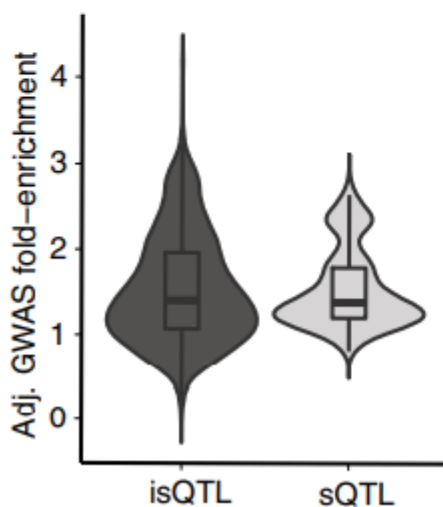
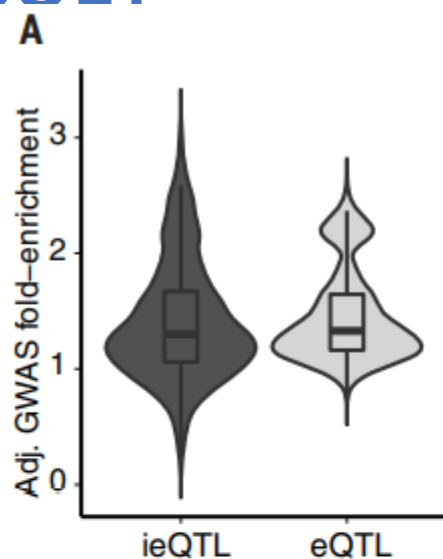
D



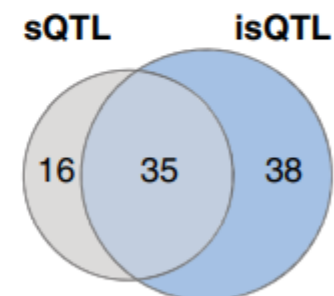
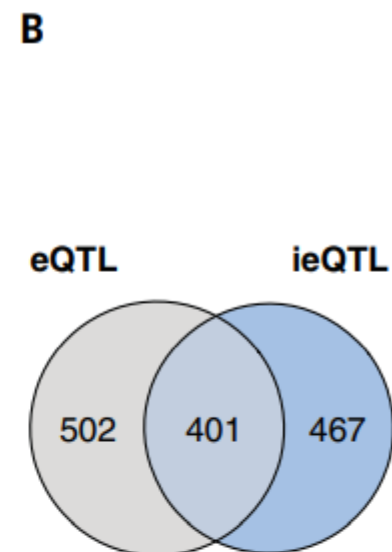
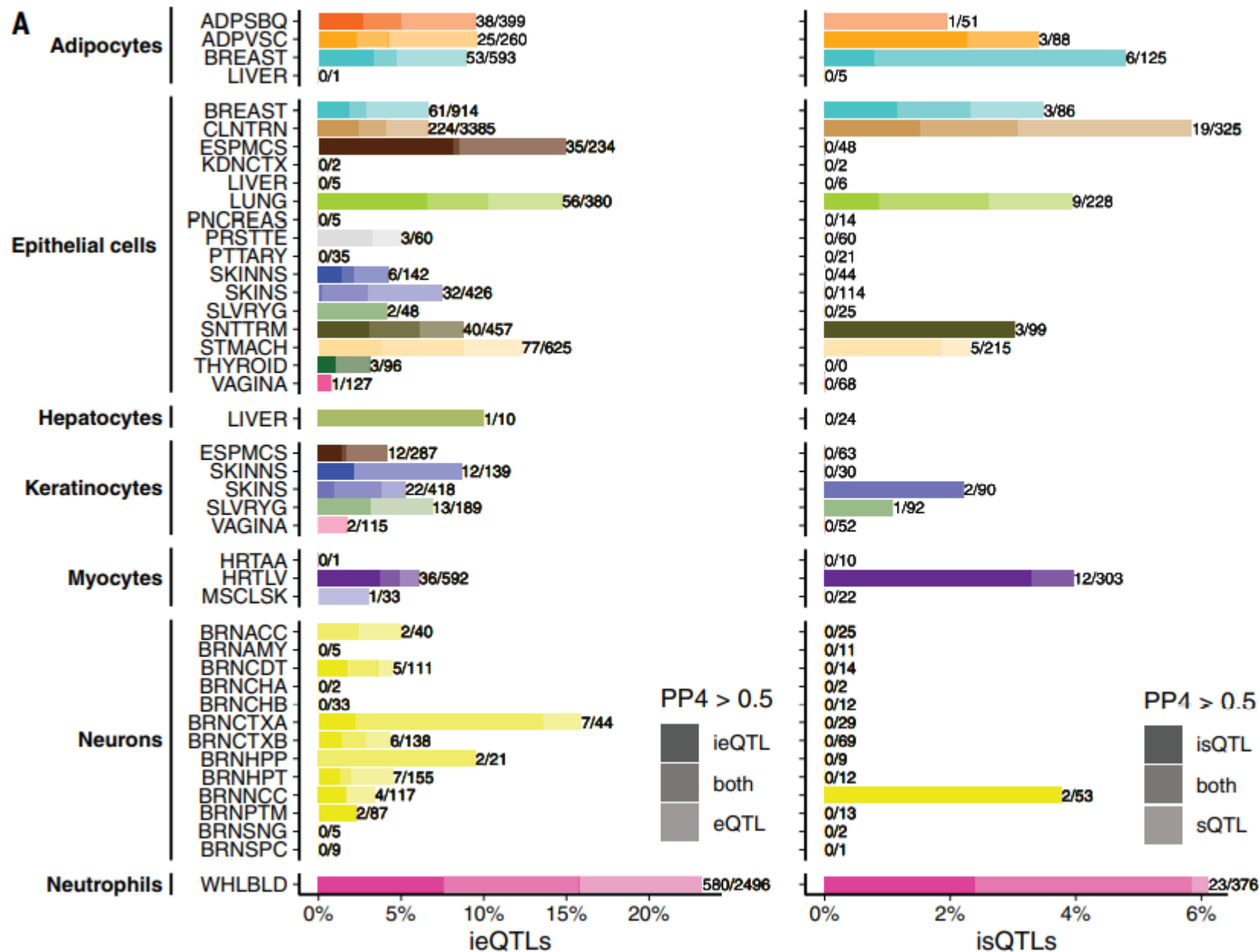
Cell type ieQTLs

top eQTLs

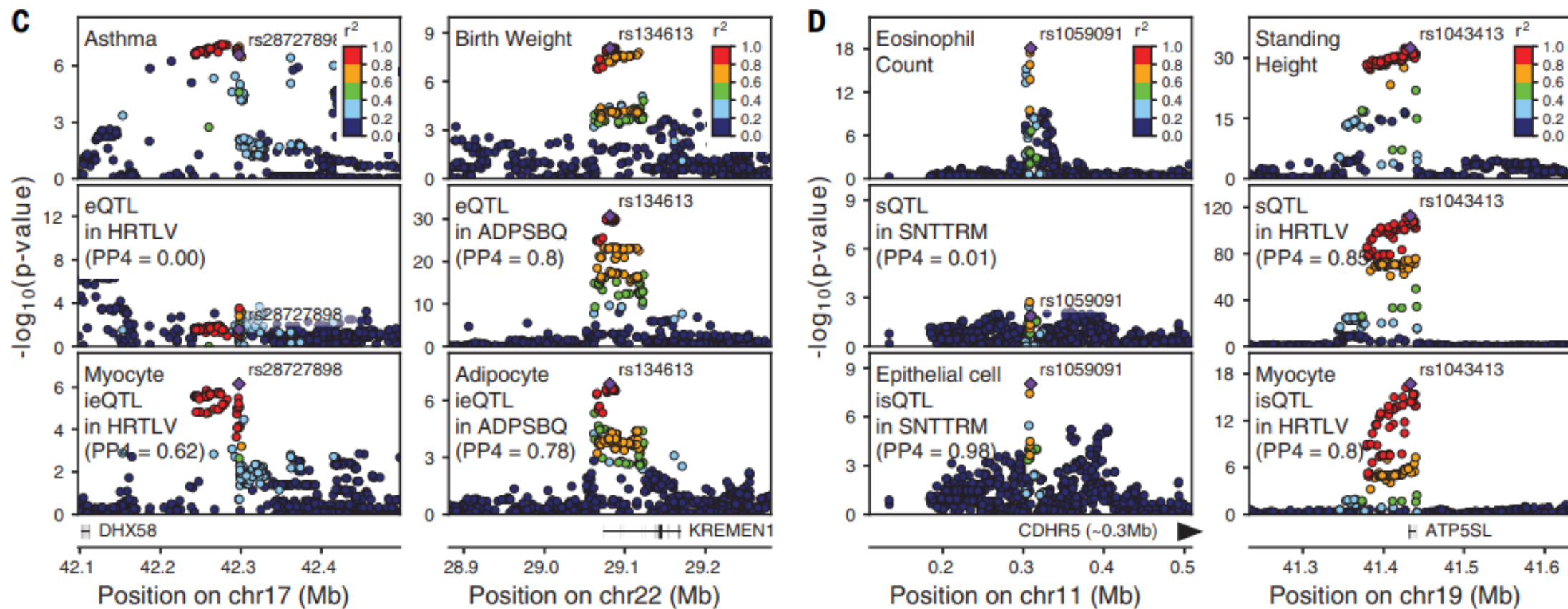
RESULT



RESULT



RESULT



- 在一些细胞类型中，利用共有的iQTLs与GWAS基因座发现新的潜在靶基因，并为复杂性状潜在的调控作用的细胞特异性暗示可能潜在的机制，而标准QTLs却不能。

Thank you

Any question is welcome !