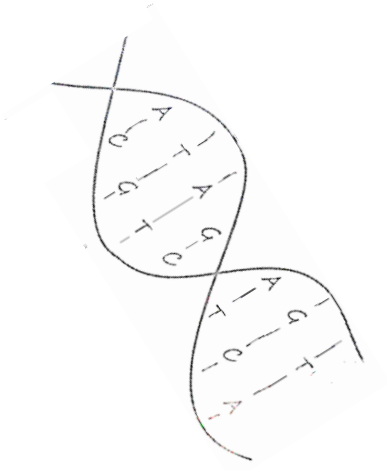
浙江大学 医学中心
ZHEJIANG UNIVERSITY MEDICAL CENTER

单细胞多组学在解析慢性肾脏病的遗传学致病机制的应用



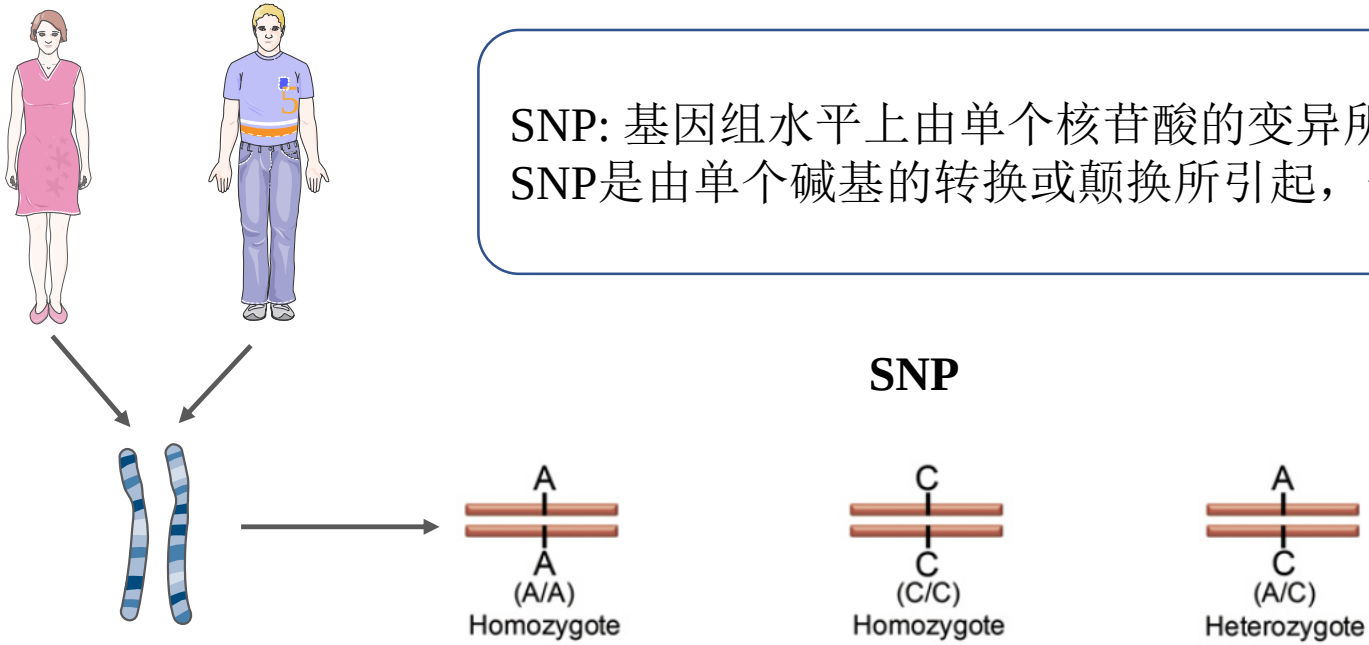
盛 欣
浙江大学医学中心

内容大纲

- **基本概念**
- 慢性肾脏病及其并发症
- 多组学数据整合分析方法
- 肾脏具有高度细胞异质性
- 从单细胞水平解析**SNP-基因-CKD机制**
- 总结

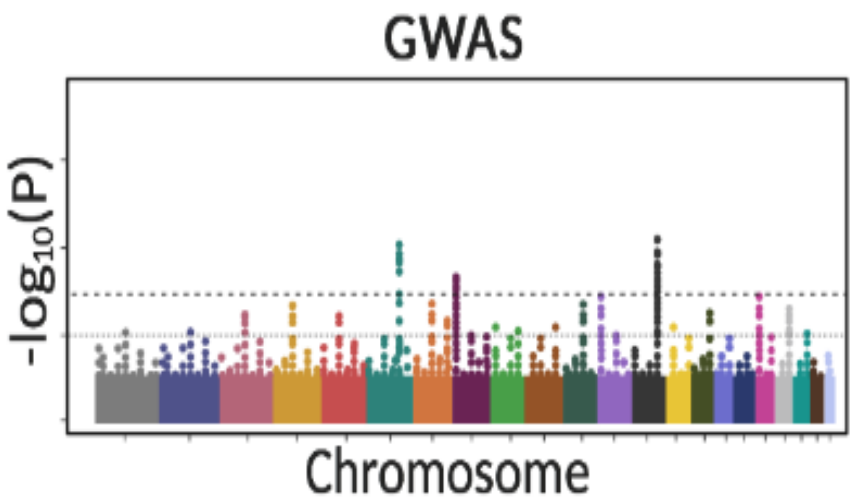
单核苷酸多态（SNP）

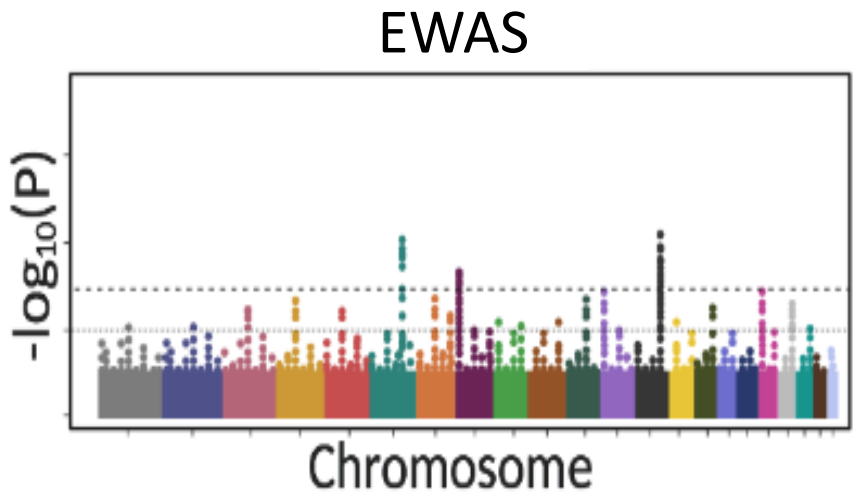
人与人的基因序列中，有99.9%以上的序列都是相同的，仅有0.1%不同。



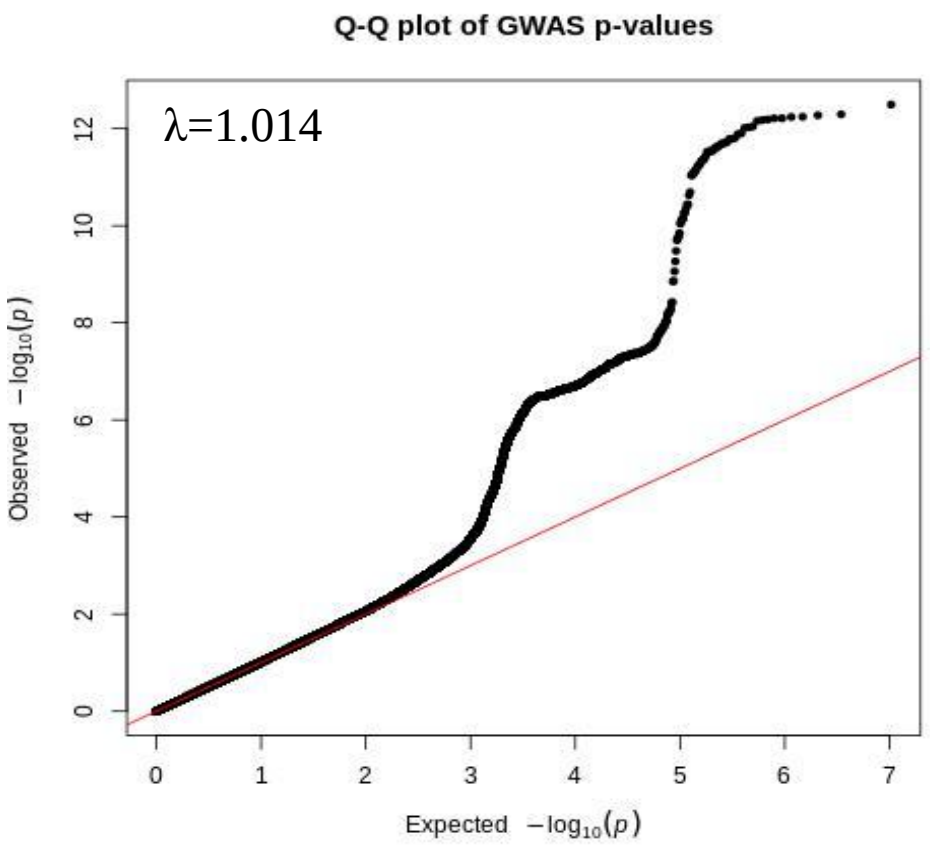
SNP: 基因组水平上由单个核苷酸的变异所引起的DNA序列多态性。
SNP是由单个碱基的转换或颠换所引起，也可由碱基的插入或缺失所致。

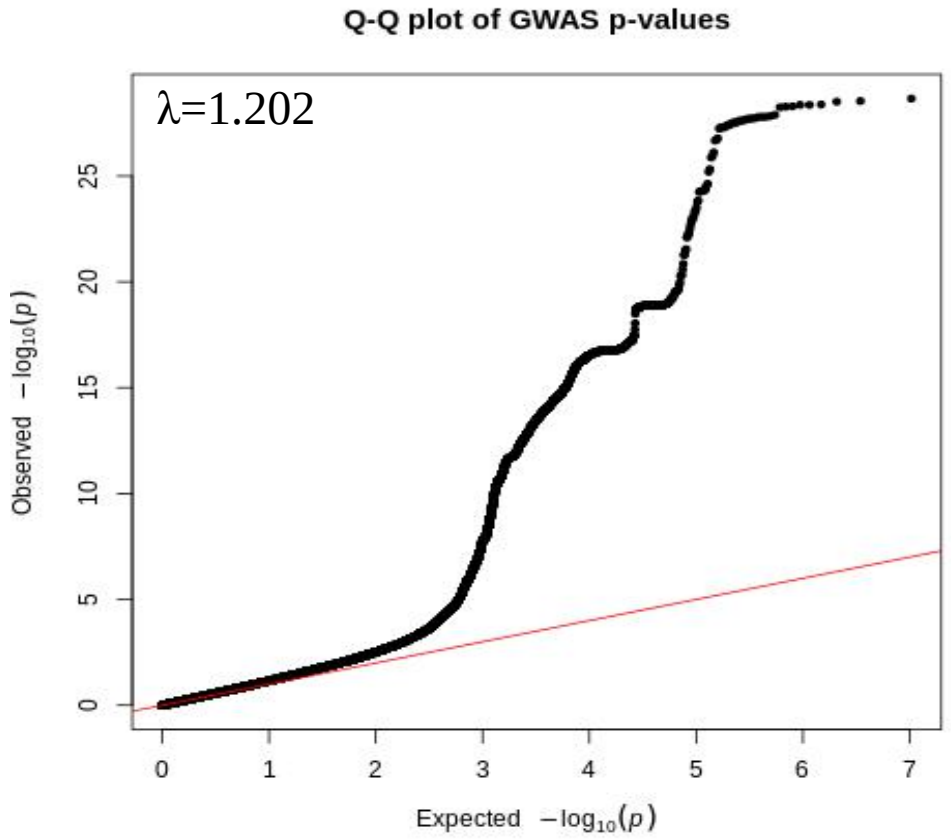
全X组关联分析（XWAS）





膨胀系数：不止是线性回归





GWAS信息能有效指导新药研发

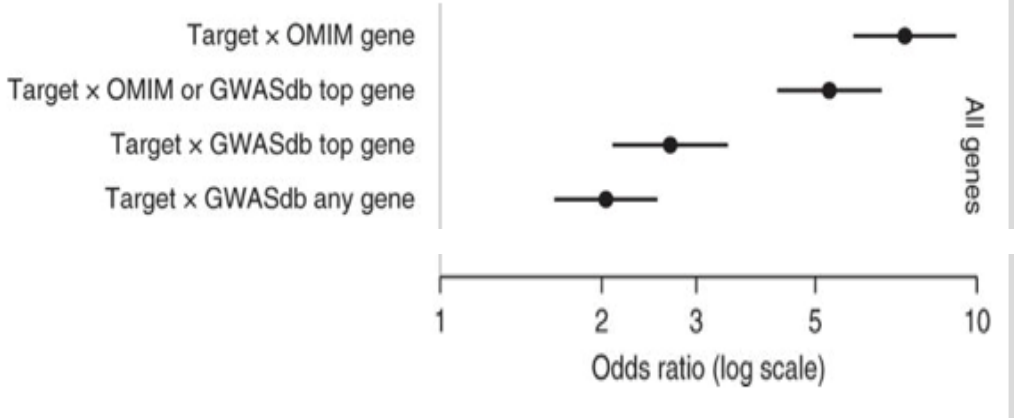
ANALYSIS

nature genetics

The support of human genetic evidence for approved drug indications

Matthew R Nelson[1], Hannah Tipney[2], Jeffery L Painter[1], Judong Shen[1], Paola Nicoletti[3], Yufeng Shen[3,4], Aris Floratos[3,4], Pak Chung Sham[5,6], Mulin Jun Li[6,7], Junwen Wang[6,7], Lon R Cardon[8], John C Whittaker[2] & Philippe Sanseau[2]

Figure 2: Enrichment of target genes for drugs approved in the United States or the European Union.



但有GWAS还远远不够…

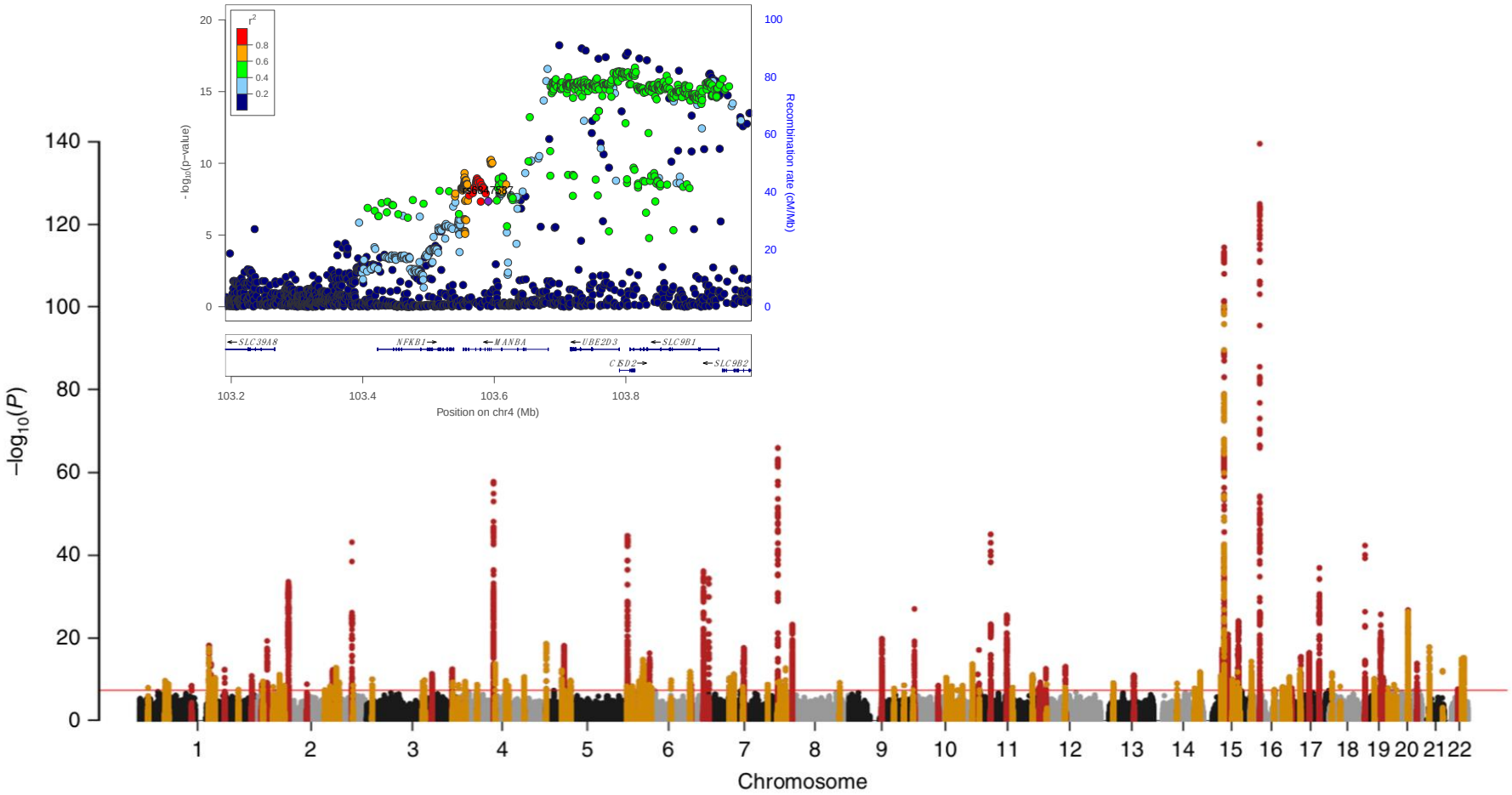


Fig. 1 Manhattan plot summarizing transethnic discovery meta-analysis of eGFR. The *y* axis shows the -log10 *P*-values and the *x* axis shows the chromosomal positions. The horizontal red line represents the thresholds of *P*-value = 5×10^{-8} for genome-wide significance. SNPs in red are in previously-identified kidney function loci, whereas SNPs in orange are in novel loci

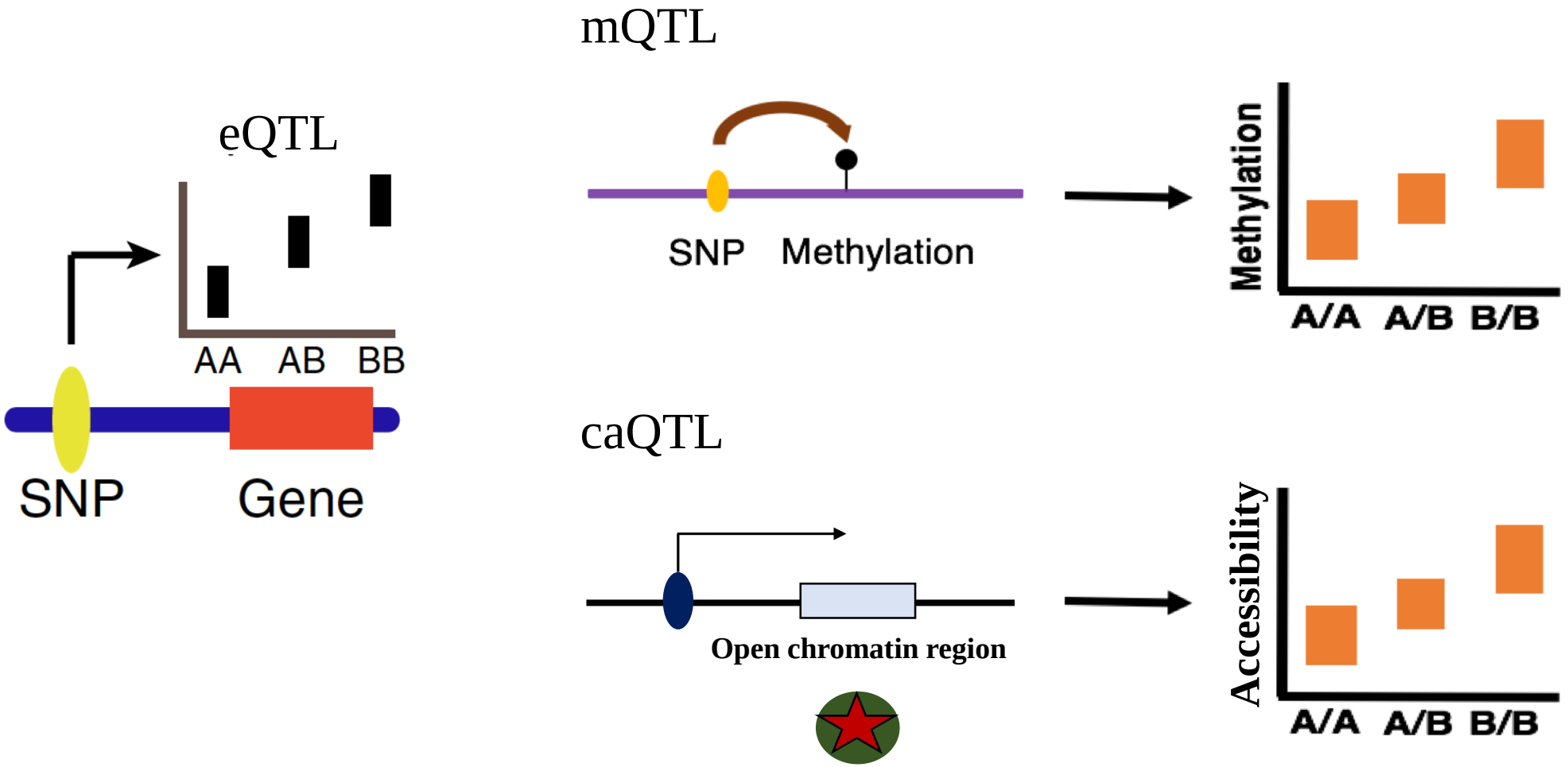
Gene

Cell type

Mechanism

Hellwege et al, *Nature Communications 2019*

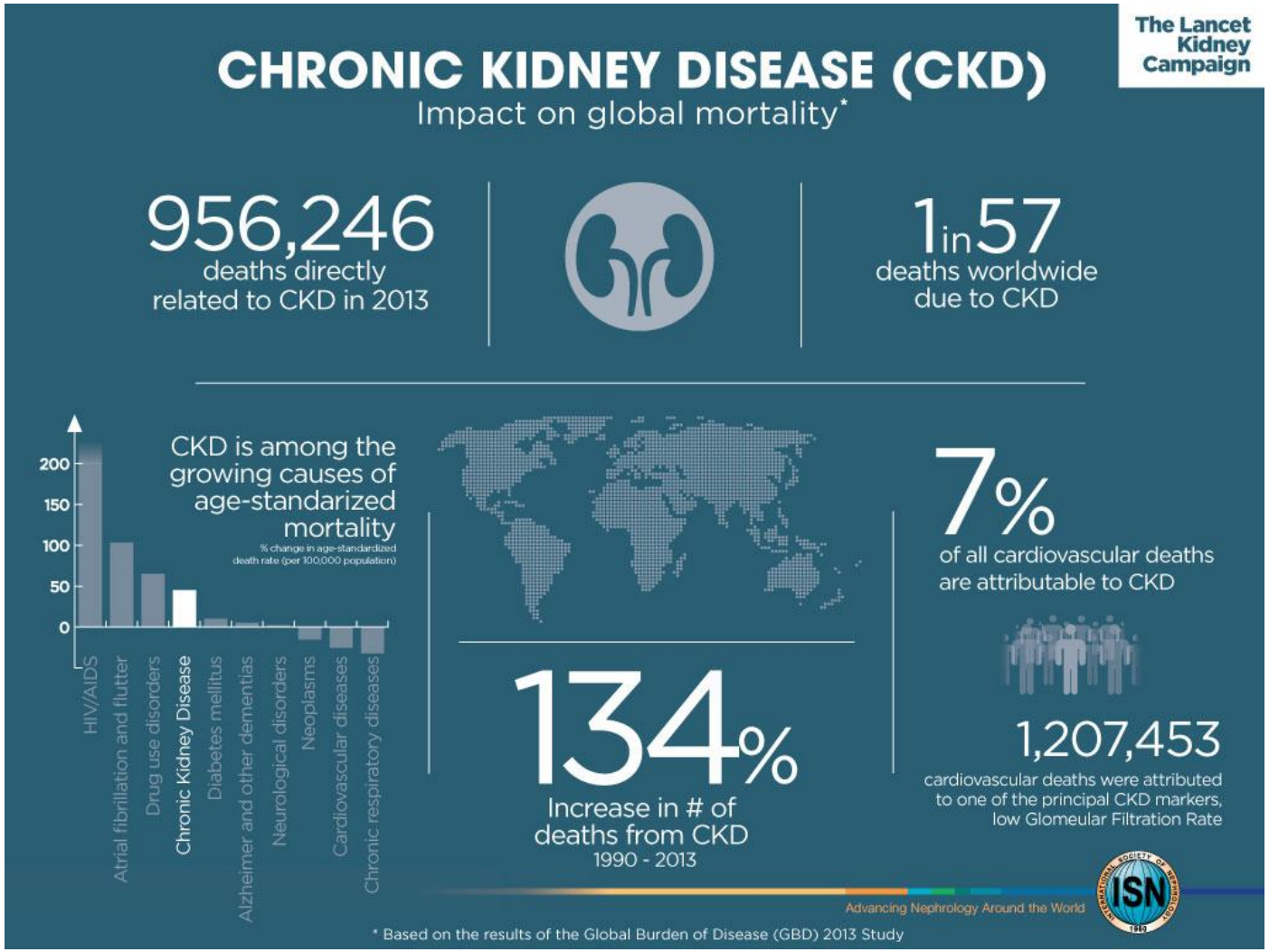
数量性状位点（xQTL）


eQTL
AA
AB
BB
SNP
Gene
mQTL
SNP
Methylation
Methylation
A/A
A/B
B/B
caQTL
Open chromatin region
Accessibility
A/A
A/B
B/B

内容大纲

- 基本概念
- **慢性肾脏病及其并发症**
- 多组学数据整合分析方法
- 肾脏具有高度细胞异质性
- 从单细胞水平解析SNP-基因-CKD机制
- 总结

慢性肾脏病（Chronic Kidney Disease, CKD）



回顾经典：糖尿病控制及并发症研究（DCCT）

初级预防队列N=726

- 1-5年糖尿病病程
- 无视网膜病变
- 尿白蛋白肌酐比< 40 mg/24h

二级干预防队列N=715

- 1-15年糖尿病病程
- 轻到中度非增殖期视网膜病变
- 尿白蛋白肌酐比< 200 mg/24h

随机分组

强化治疗：胰岛素泵或每天≥3次/天胰岛素注射，N=348

标准治疗：1-2次/天胰岛素注射，N=378

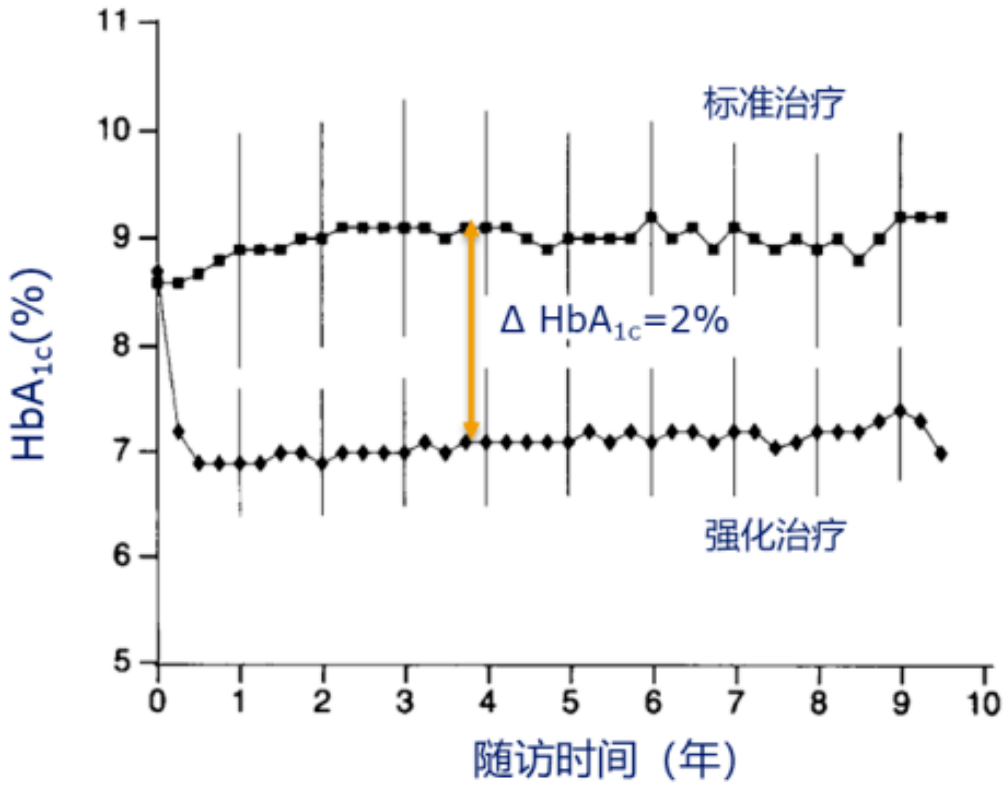
强化治疗：胰岛素泵或每天≥3次/天胰岛素注射，N=363

标准治疗：1-2次/天胰岛素注射，N=352

1983年 → **1993年**

1441名T1DM患者：胰岛素依赖性、年龄13-39岁、无高血压、高胆固醇血症、糖尿病严重并发症或内科疾病

回顾经典：糖尿病控制及并发症研究（DCCT）



微血管并发症	相对风险减少，%（95%置信区间）
视网膜病变进展	63（52-71）
视网膜黄斑变性	26（-8-50）
严重非增殖期/增殖期视网膜病变	47（15-67）
视网膜病变激光治疗	51（21-70）
尿白蛋白肌酐比（mg/24h）	
◆ ≥40	39（21-52）
◆ ≥300	54（19-74）
临床神经病变（随访第5年）	60（38-74）

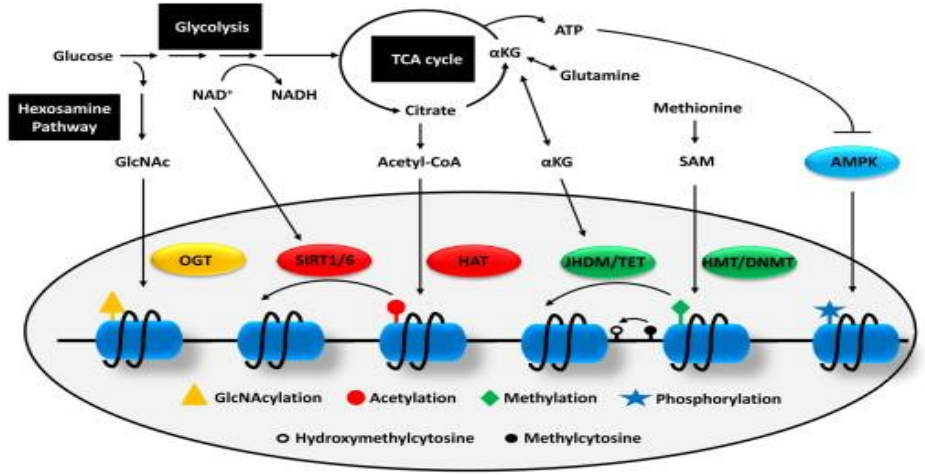
环境所留下的足迹：表观基因组修饰

- Metabolic factors play critical roles in CKD development.

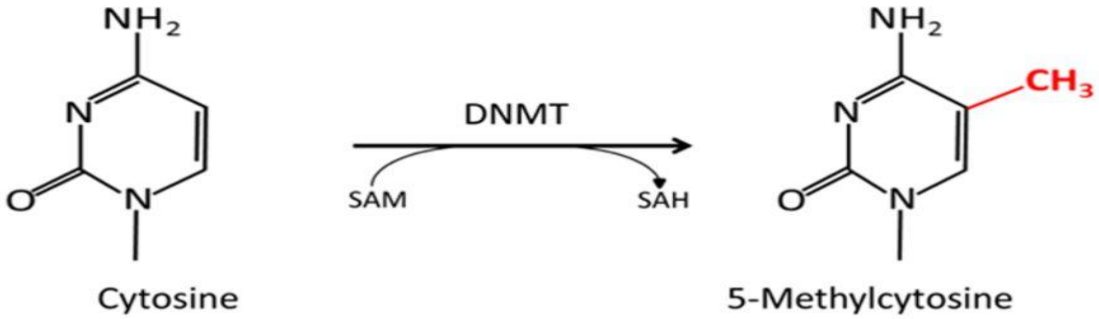
- Metabolic memory or programming:

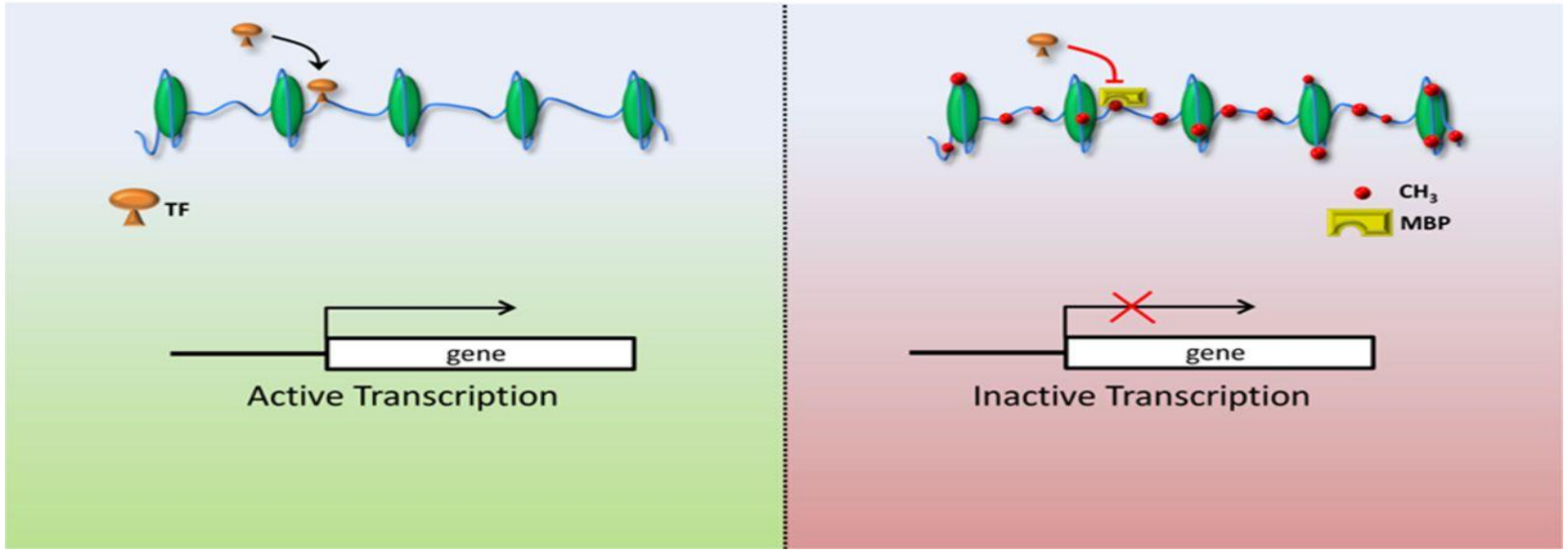
  Intrauterine nutritional deprivation or periods of hyperglycemia will increase kidney disease risk, even after several decades of good metabolic control.

- Epigenetic modifications are maintained during cell division.



环境所留下的足迹：DNA甲基化





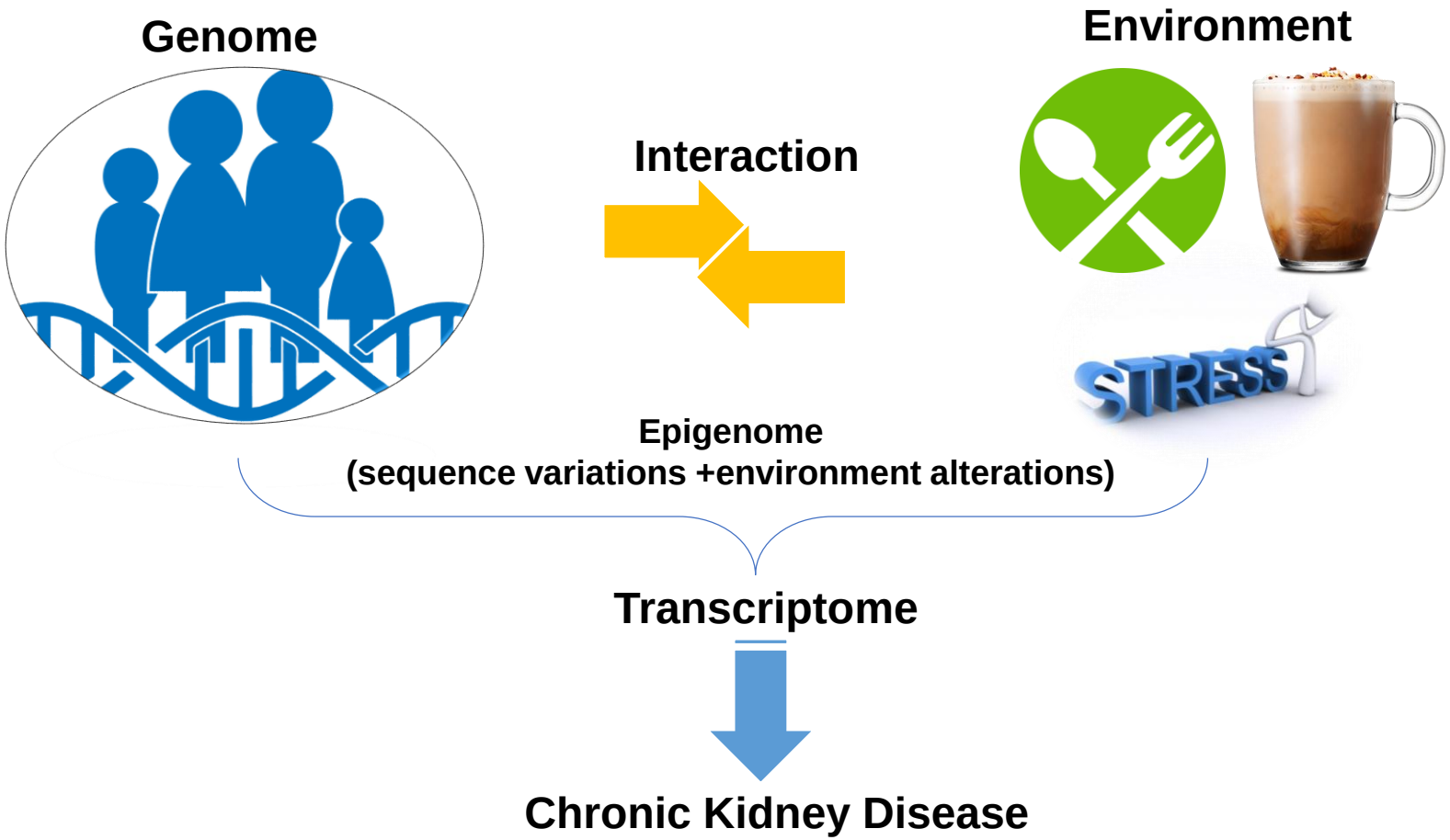
MBP: methyl CpG-binding protein

DNMT: DNA methyltransferase

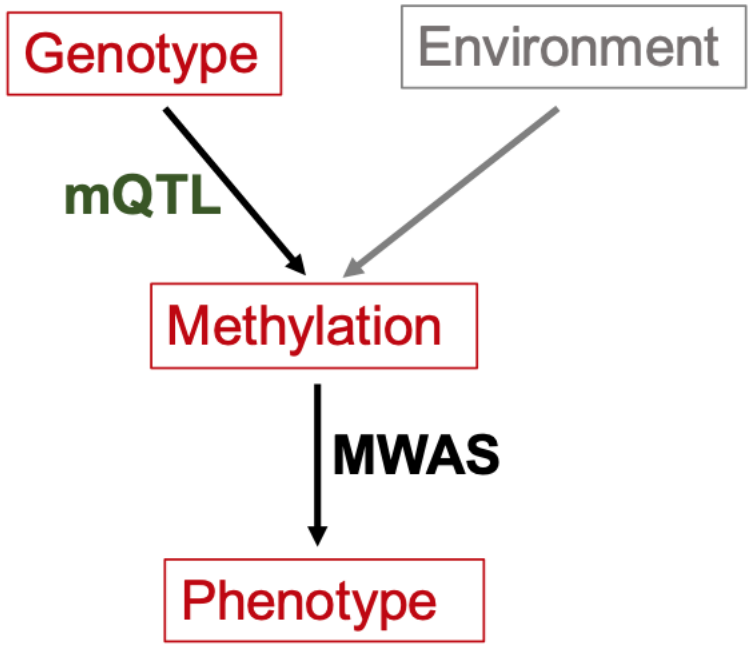
SAM: S-adenosylmethionine

SAH: S-adenosylhomocysteine

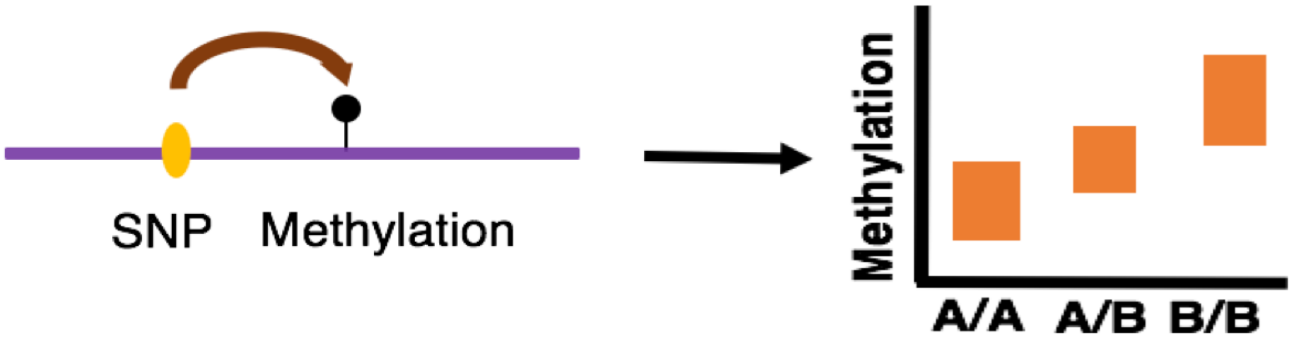
遗传-环境相互作用→糖尿病肾病（DKD）



遗传-环境相互作用→糖尿病肾病（DKD）



Methylation quantitative trait (mQTL)



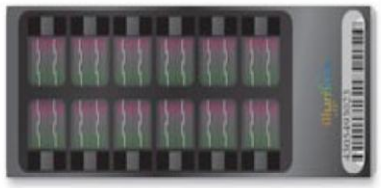
Sheng et al. PNAS, 2020

全甲基化组关联分析（MWAS）

We defined genome-wide methylation changes associated with four DKD phenotypes: Glycemia, albuminuria, eGFR, and eGFR decline.

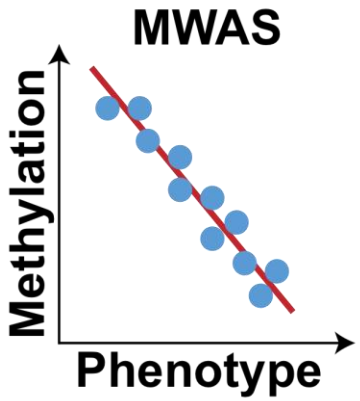
CRIC Study
~500 DKD

~8×10^5 CpCs
Methylation

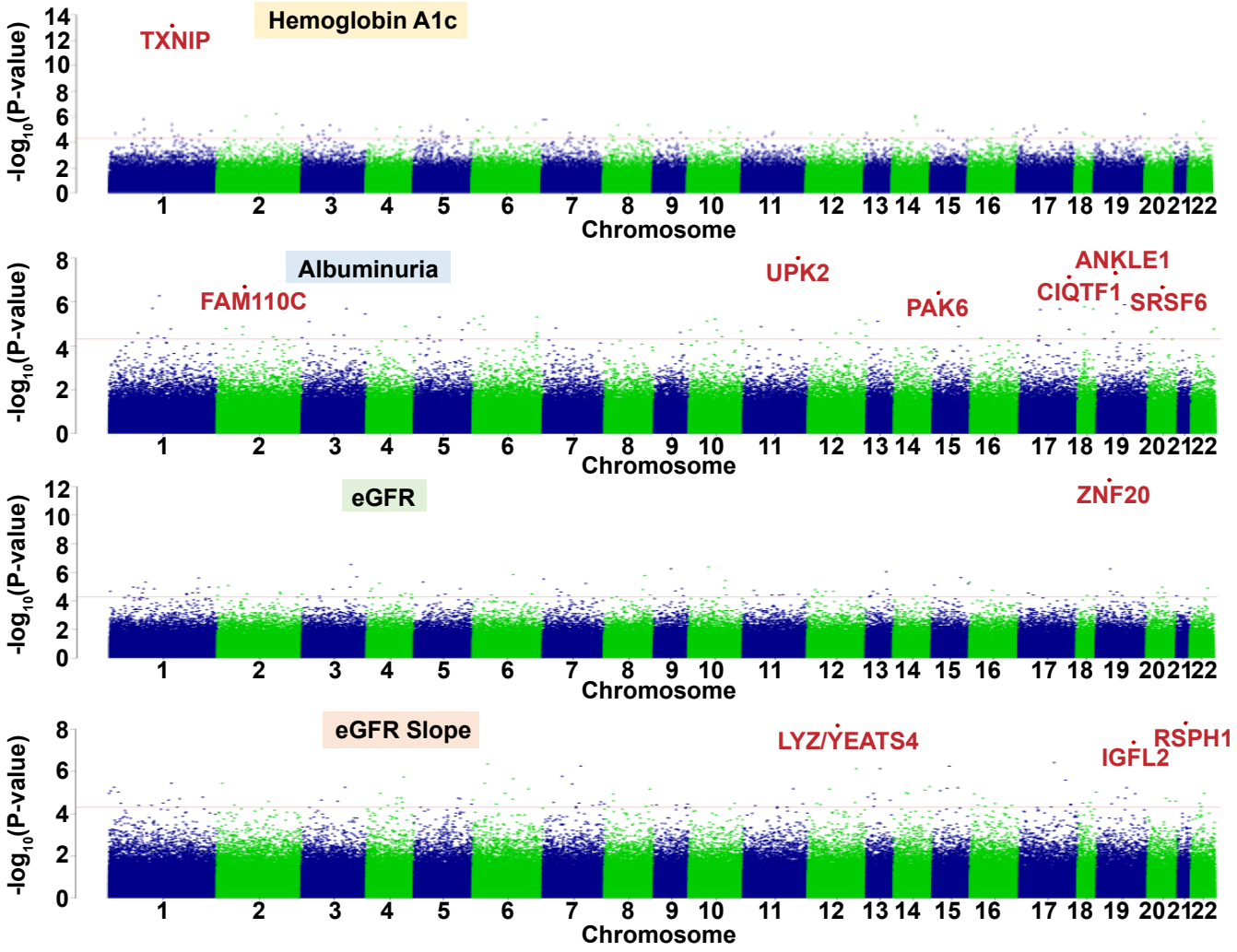


Phenotype

Glycemia (HgbA1c)
Albuminuria
Kidney function (eGFR)
Kidney function decline (eGFR slope)



与糖尿病肾病相关的甲基化位点



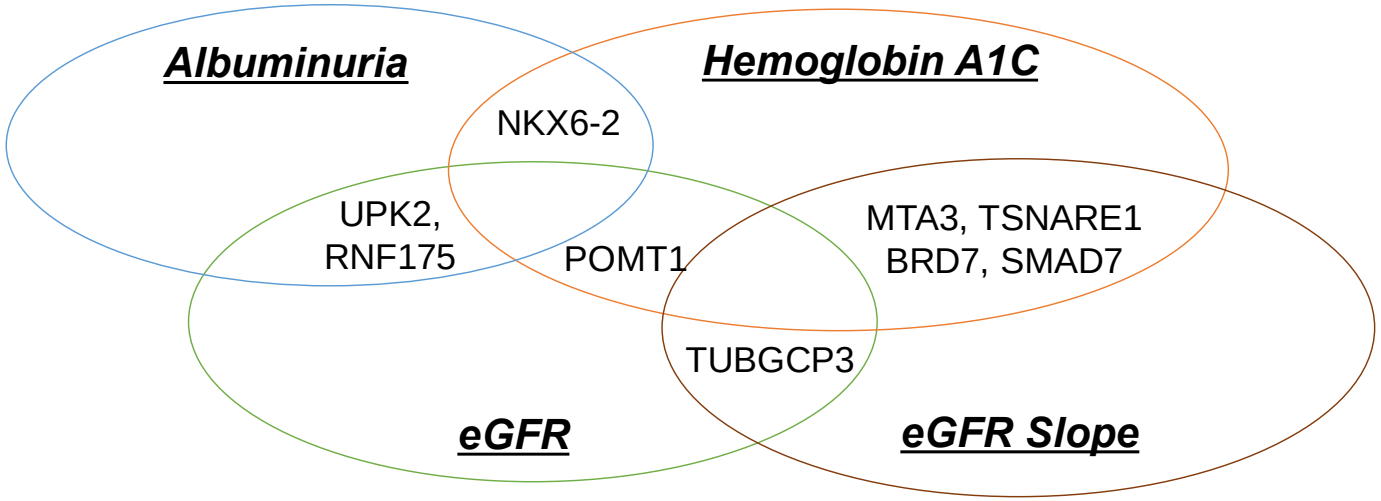
TXNIP (thioredoxin-interacting protein) plays an important role in redox homeostasis and a physiologic regulator of peripheral glucose uptake into fat and muscle in human.

Uroplakins cover urothelial apical surfaces. Mice with null mutation of Upk2 are often born with congenital kidney disease.

Cg17944885 (ZNF20) could be validated in multiple studies that analyzed blood or kidney samples.

高血糖与DKD的其他表型密切相关

- Methylation changes were mostly specific to the analyzed phenotypes.
- Glycemic control showed the greatest overlap with other DKD phenotypes.



Hemoglobin A1C

single organismal cell-cell adhesion
response to drug
histone H2B ubiquitination
cell chemotaxis
response to glucose
fibroblast migration
very long-chain fatty acid metabolism

Albuminuria

wound healing
GTPase activity
viral process
receptor activity
plasma membrane lactate transport
respiratory electron transport chain

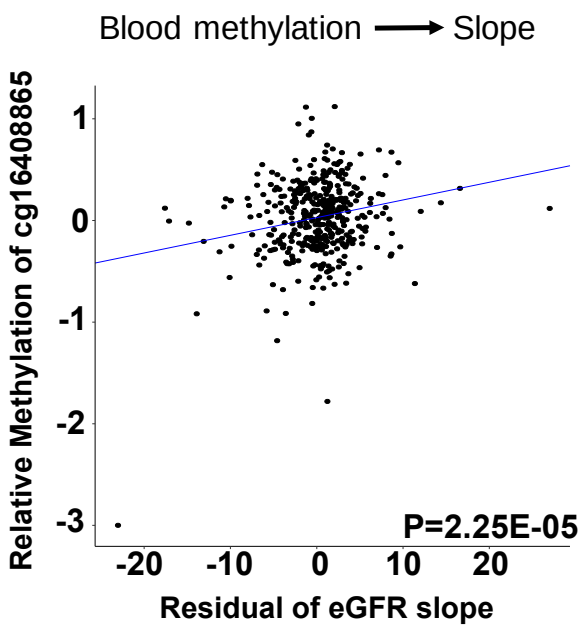
eGFR

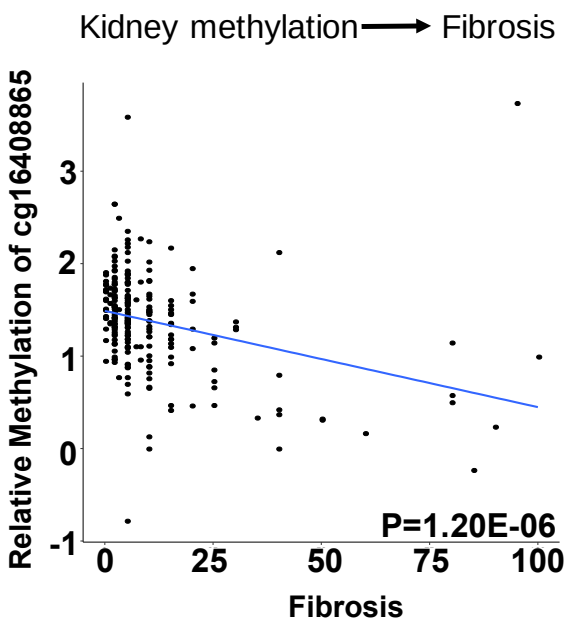
transcription
morphogenesis
canonical Wnt signaling
chromatin assembly
protein sumoylation
mesonephros development
NIK/NF-kappaB signaling

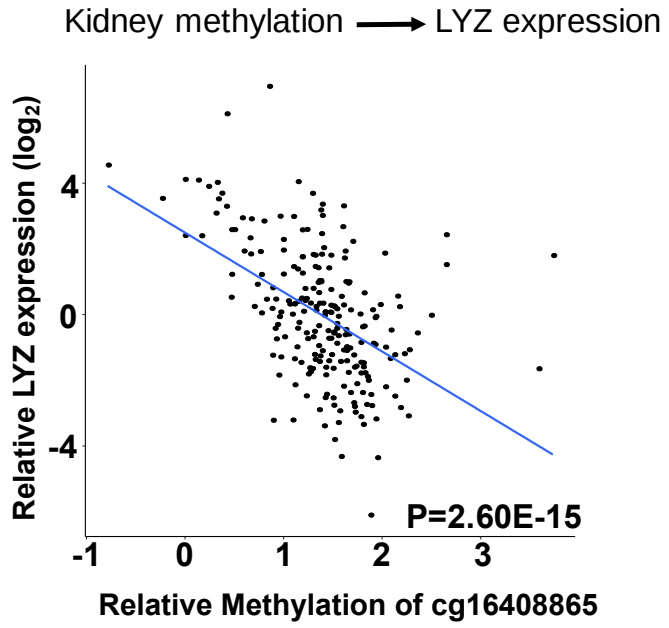
eGFR Slope

cAMP metabolic process
cardiac muscle cell apoptosis
MAPK cascade
transcription
protein autophosphorylation
GMP metabolic process
piRNA metabolic process
regulation of JNK cascade

血液中与肾纤维化相关的甲基化位点

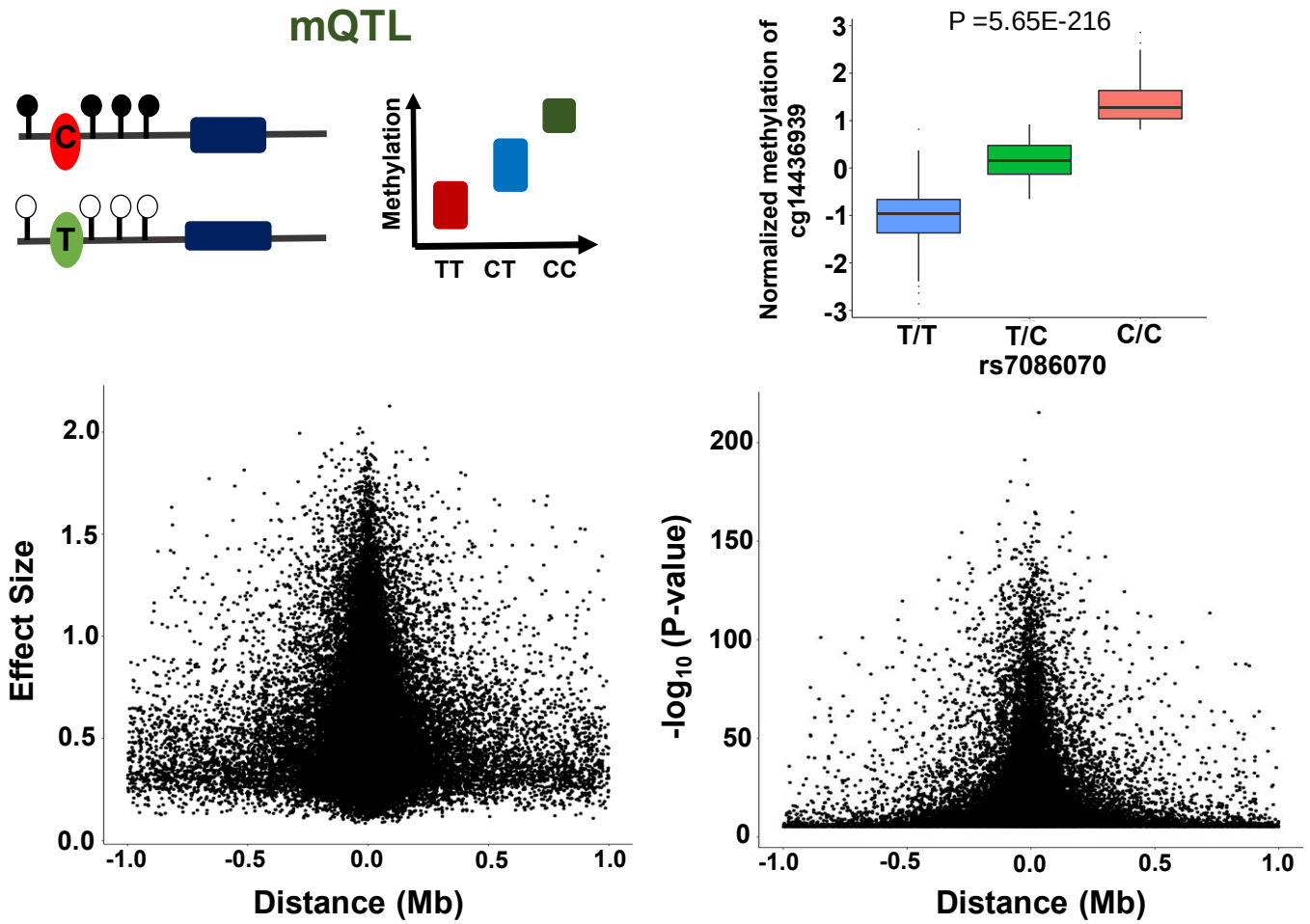




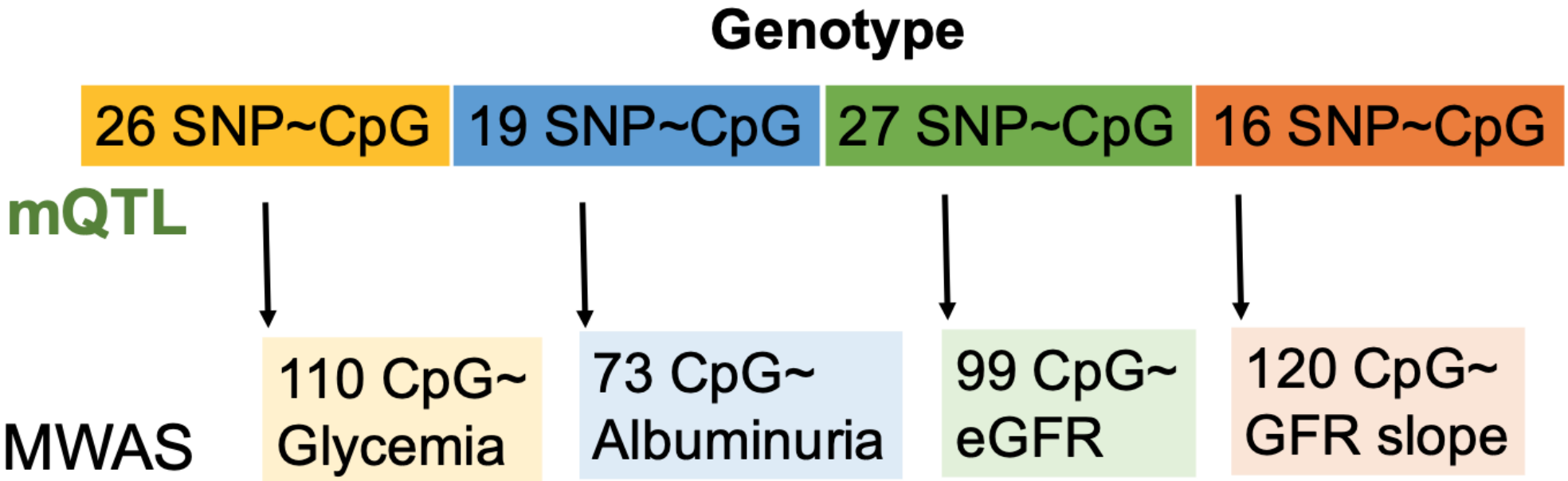


遗传因素→DNA甲基化

- 171,732 CpG sites were identified as significant mCpGs (CpG site that regulated by at least one SNP)



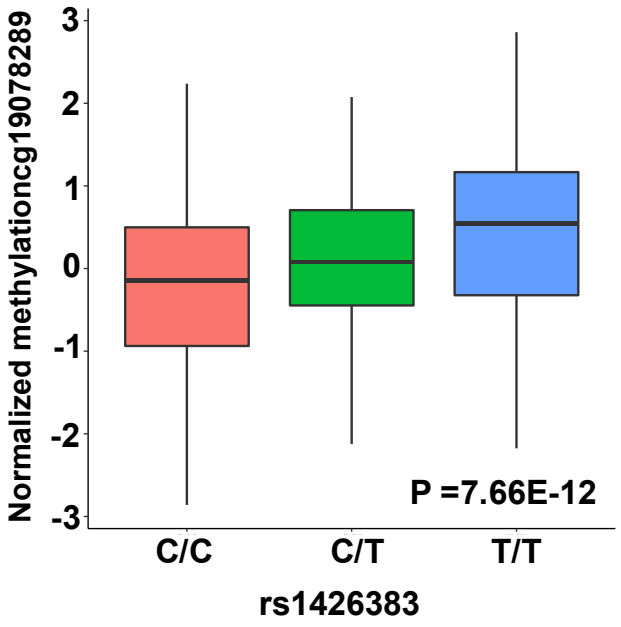
遗传因素→DNA甲基化→疾病表型



MBNL1: the muscleblind-like protein 1 locus

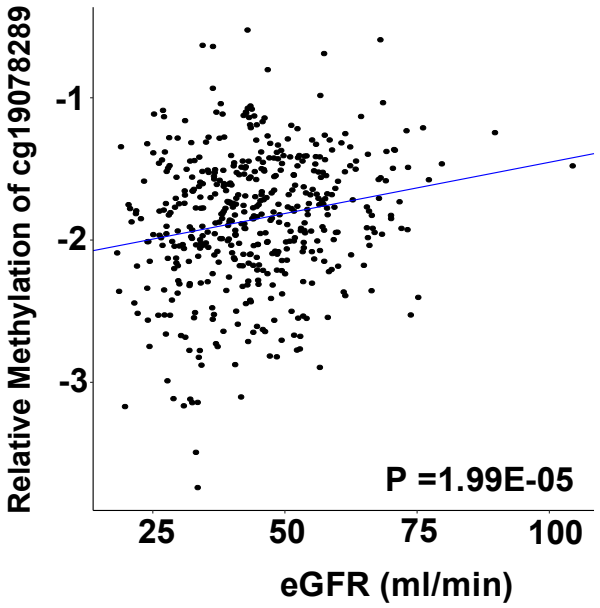
Genotype rs1426383 → Methylation

mQTL

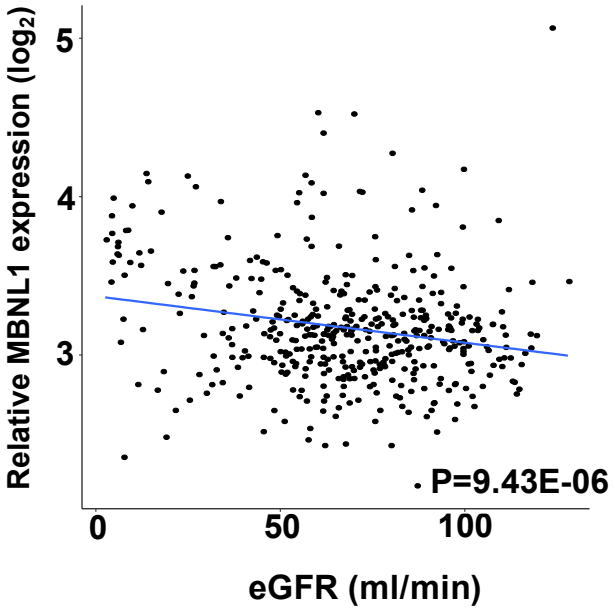


Methylation → Phenotype

MWAS



Expression → Kidney function



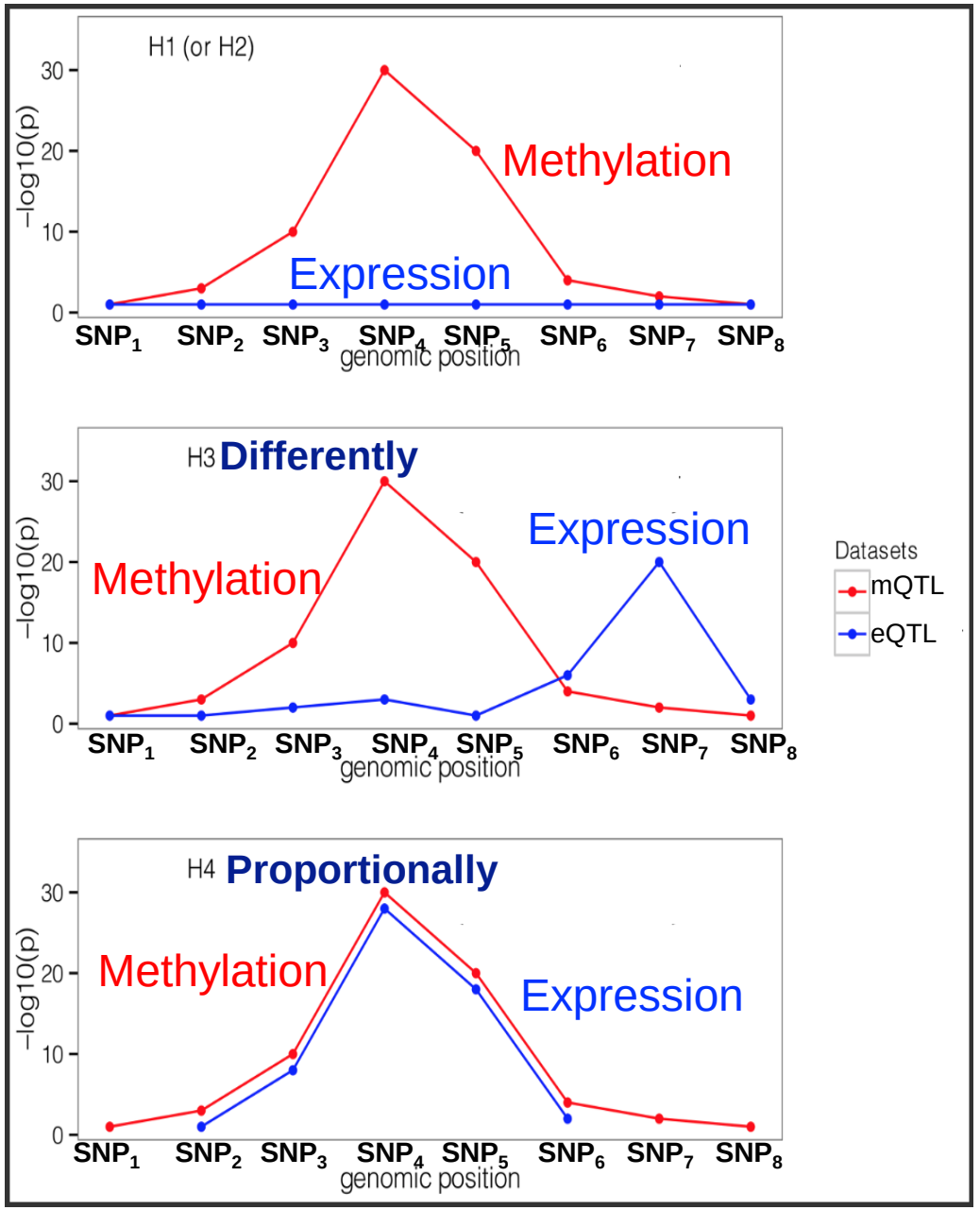
内容大纲

- 基本概念
- 慢性肾脏病及其并发症
- **多组学数据整合分析方法**
- 肾脏具有高度细胞异质性
- 从单细胞水平解析SNP-基因-CKD机制
- 总结

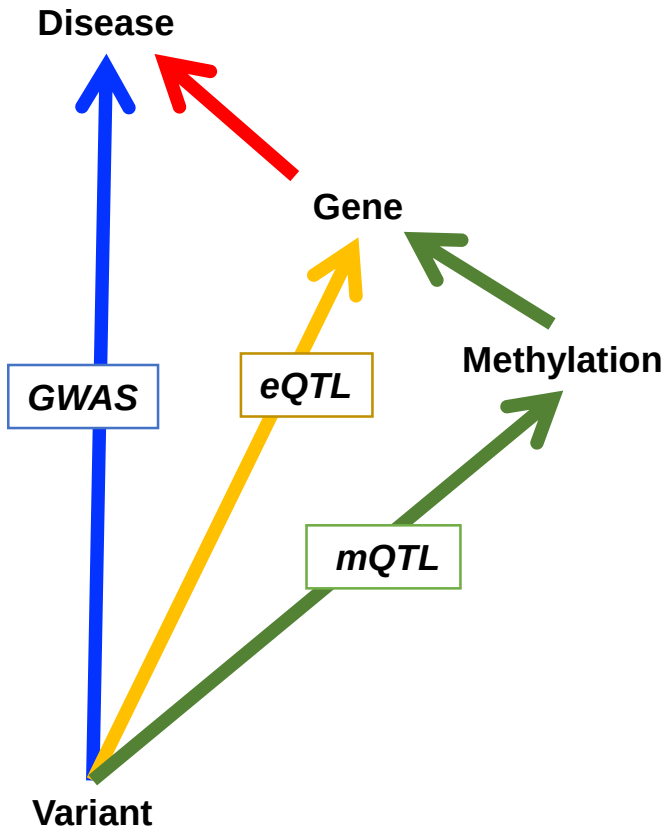
贝叶斯共定位分析 (coloc)

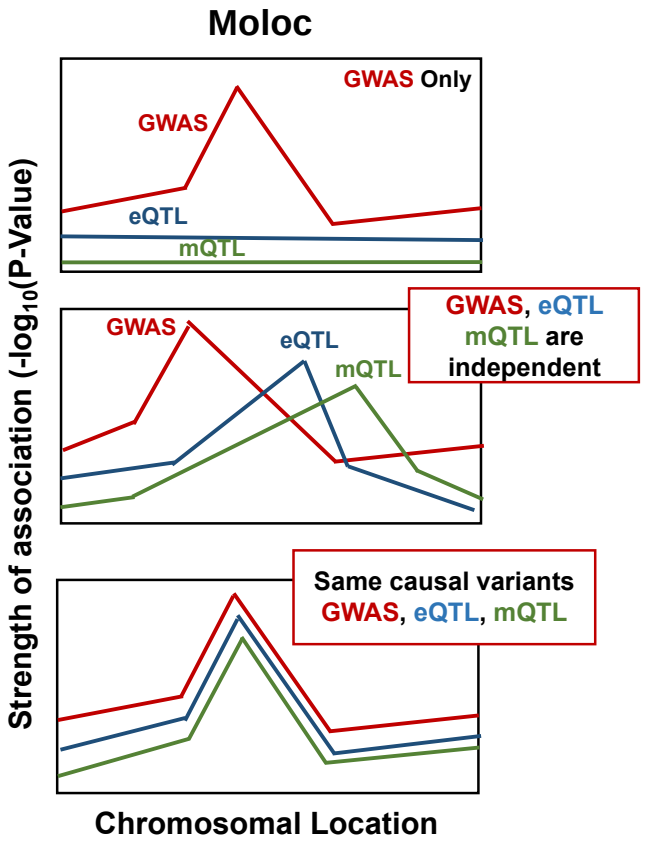
*The **posterior probability** of each possible configuration can be calculated:*
H0: neither trait has a genetic association in the region
H1: only trait 1 has a genetic association in the region
H2: only trait 2 has a genetic association in the region
H3: both traits are associated, but with different causal variants
H4: both traits are associated and share a single causal variant

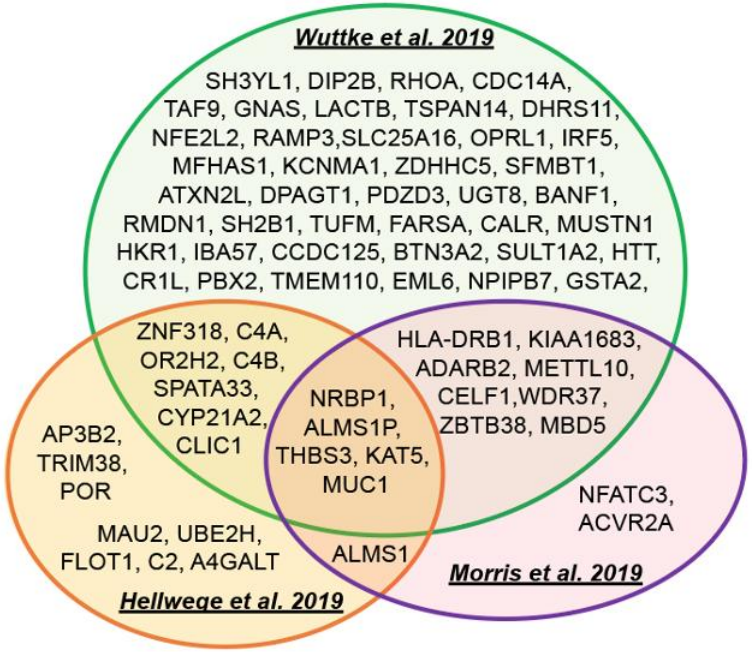
Usually use PP4>0.8 as cutoff to define co-localization.



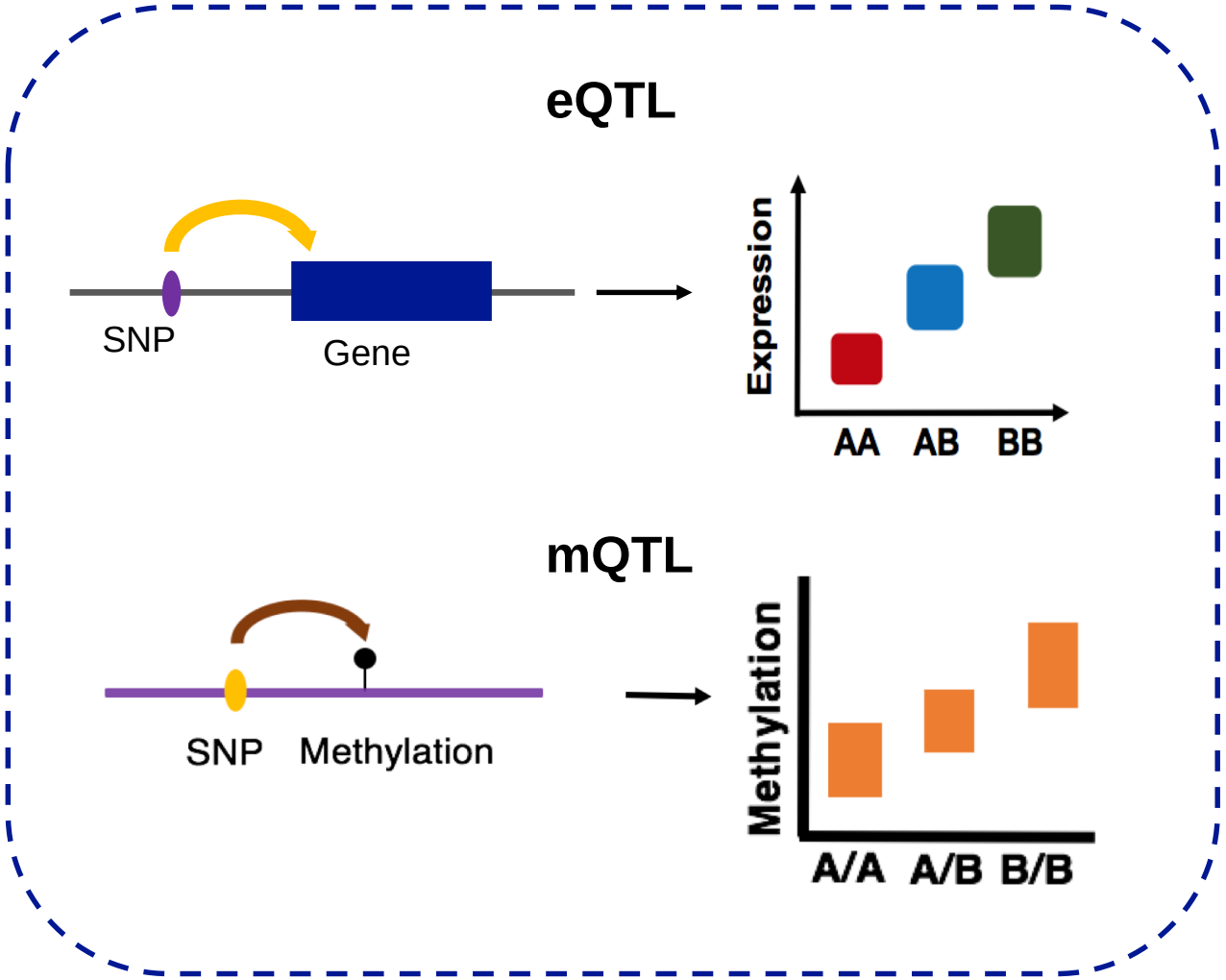
贝叶斯多性状共定位分析（moloc）



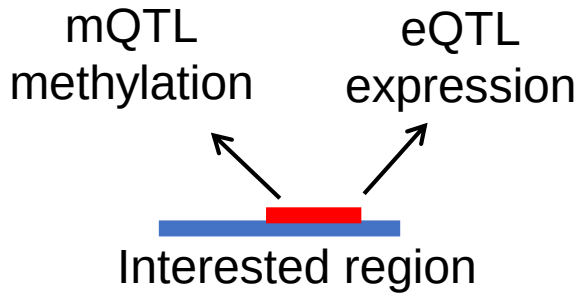




调控因果关系推断？



Are methylation and gene expression linked?



Genetic variants can influence gene expression (eQTL) and DNA methylation (mQTL).

因果推断：孟德尔随机化（MR）

Summary data based Mendelian Randomization

Genotype

↓

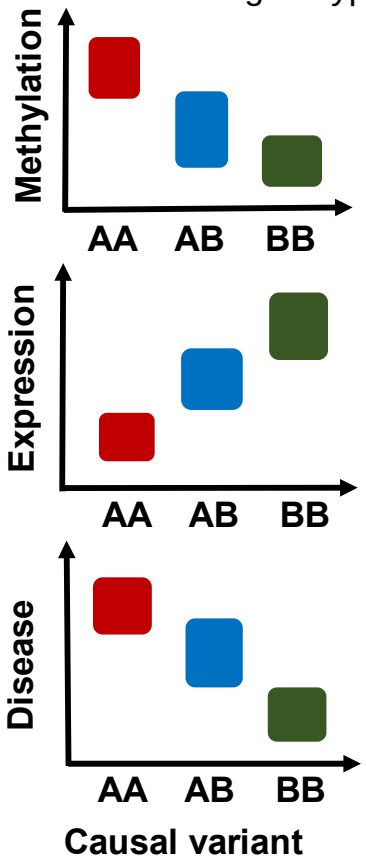
Methylation **mQTL**

↓

Expression **eQTL**

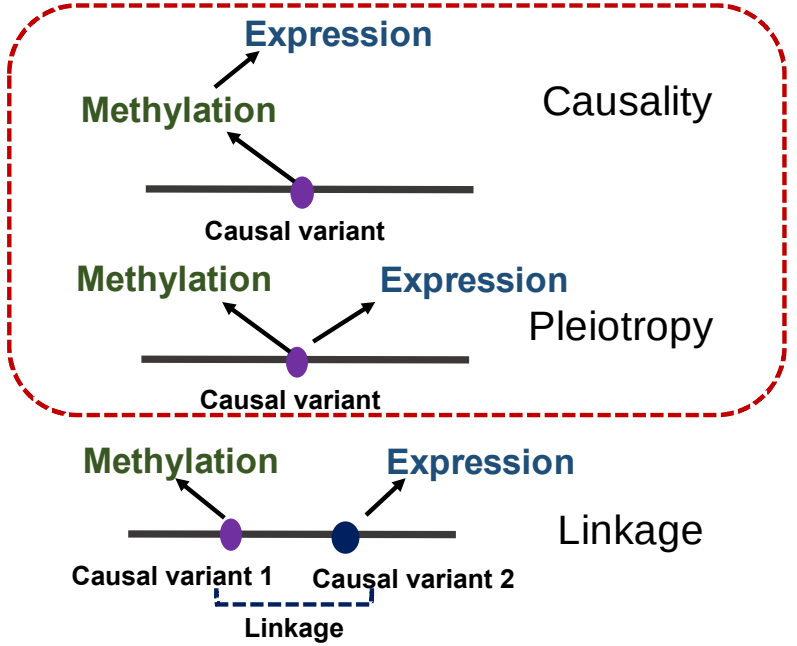
↓

Disease **GWAS**



Pleiotropic association

Identify methylation and gene expression that mediate the effect of genotypic variation on phenotype development



Mendelian Randomization

Basic rules:

Here Z is an instrumental variable

Exp SNP

$$X = \alpha_1 + \beta_1 Z + \gamma_1 U$$

If Z associated with X, and X has a causal relationship with Y (the effect of X on Y is causal)

Trait Exp

$$\begin{aligned} Y &= \alpha_2 + \beta_2 X + \gamma_2 U \\ &= \alpha_2 + \beta_2(\alpha_1 + \beta_1 Z + \gamma_1 U) + \gamma_2 U \\ &= \alpha_2 + \beta_2\alpha_1 + \beta_2\beta_1 Z + (\beta_2\gamma_1 + \gamma_2) U \\ &= \alpha_3 + \beta_3 Z + \gamma_3 U \end{aligned}$$

lm(Exp~SNP)

Z should associated with Y, if the effect of X on Y is causal.

↓

The relationship between Z and Y should be indirect. That is Z affects Y through X.
Z is the instrumental variable and Z represent the genetic variant or a combination of genetic variants. Because the fact that the alleles of a genetic variant are inherited randomly from parents to offspring, so that the relation of a genetic variant with a phenotype should not be confounded.

Beta1=bzx
Beta2=bxy
Beta3=bzy
Bzy=bxy*bzx

The slope between Z and Y should be

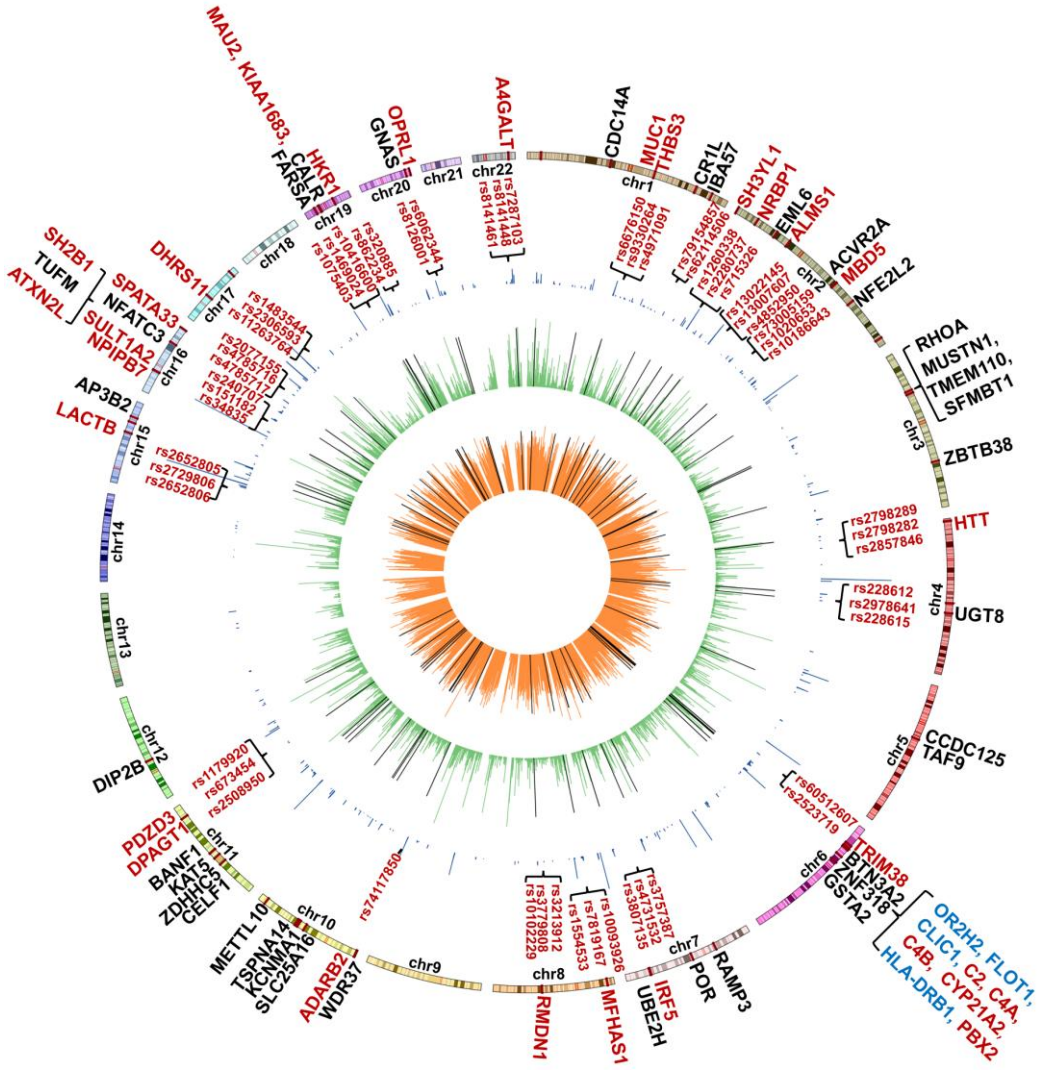
$$\beta_3 = \beta_2\beta_1$$

Is there slope significant? (the real relationship significantly existing?)

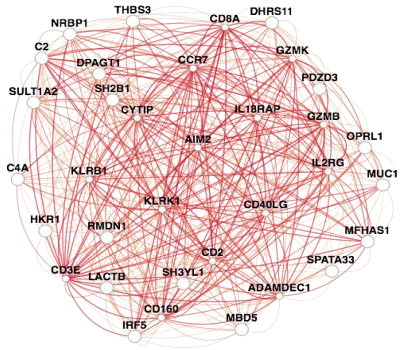
↓

Hence, the effect of X on Y is causal

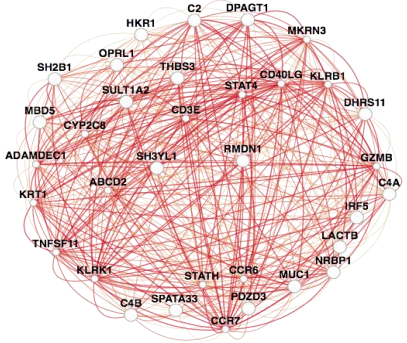
DKD致病基因鉴定



Natural Killer cell



CD8+ T-cell



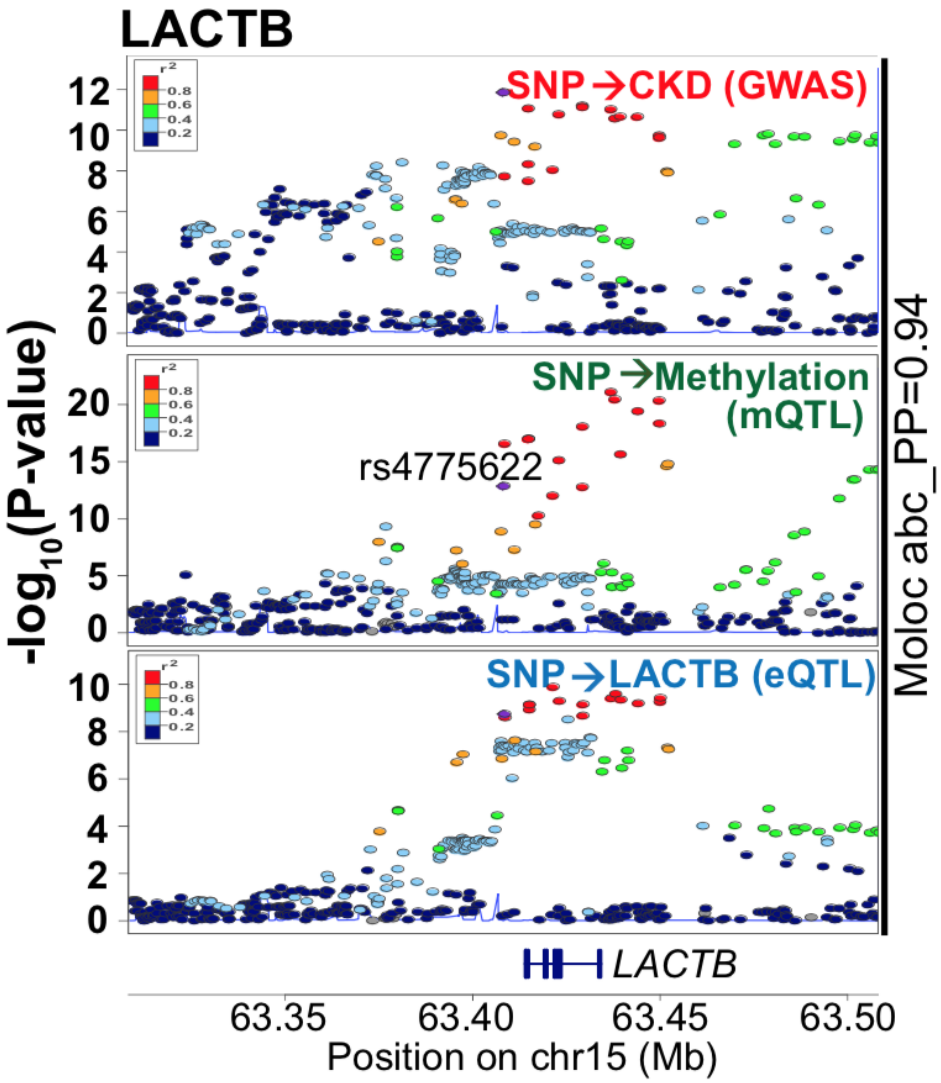
Interaction confidence:

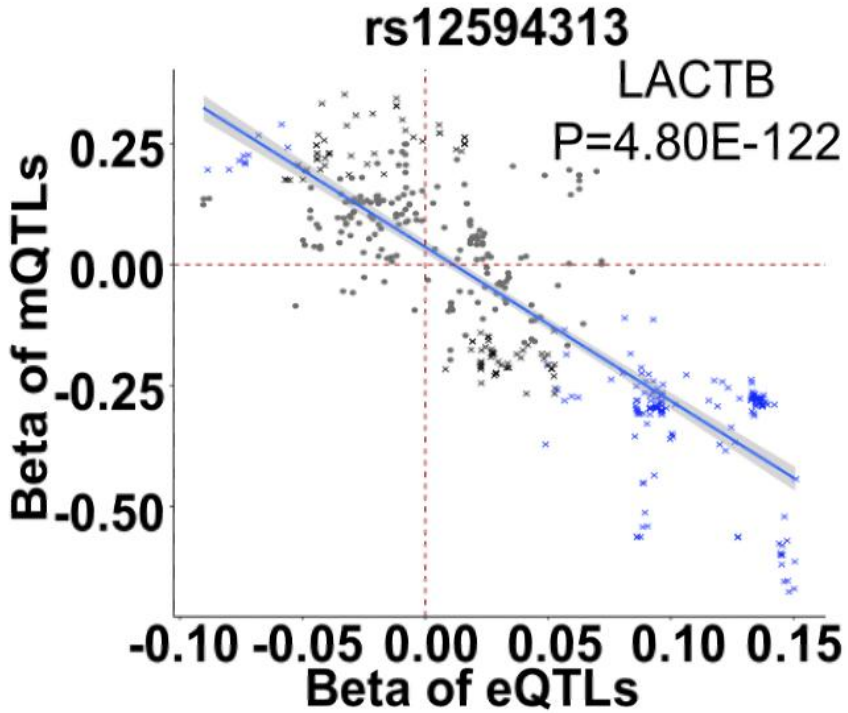
0.0 0.5 1.0

Enriched GO terms (BP):

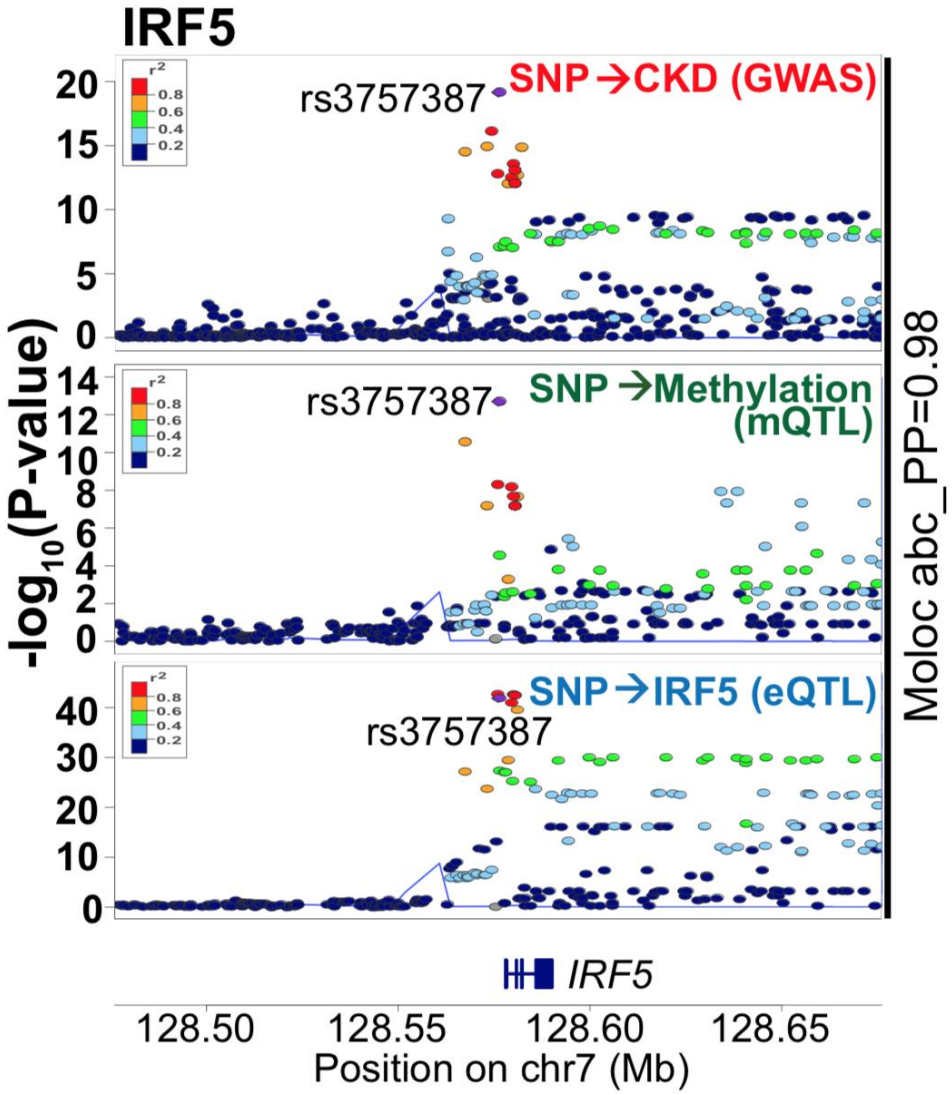
Apoptotic cell clearance, complement deficiency, immune effector process etc.

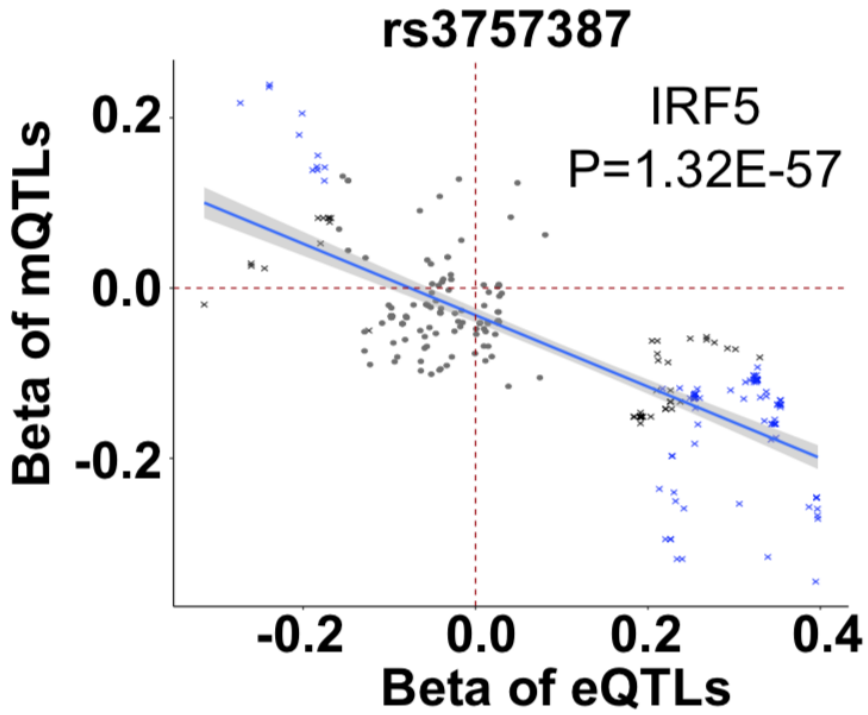
LACTB: the Serine β-lactamase-like protein



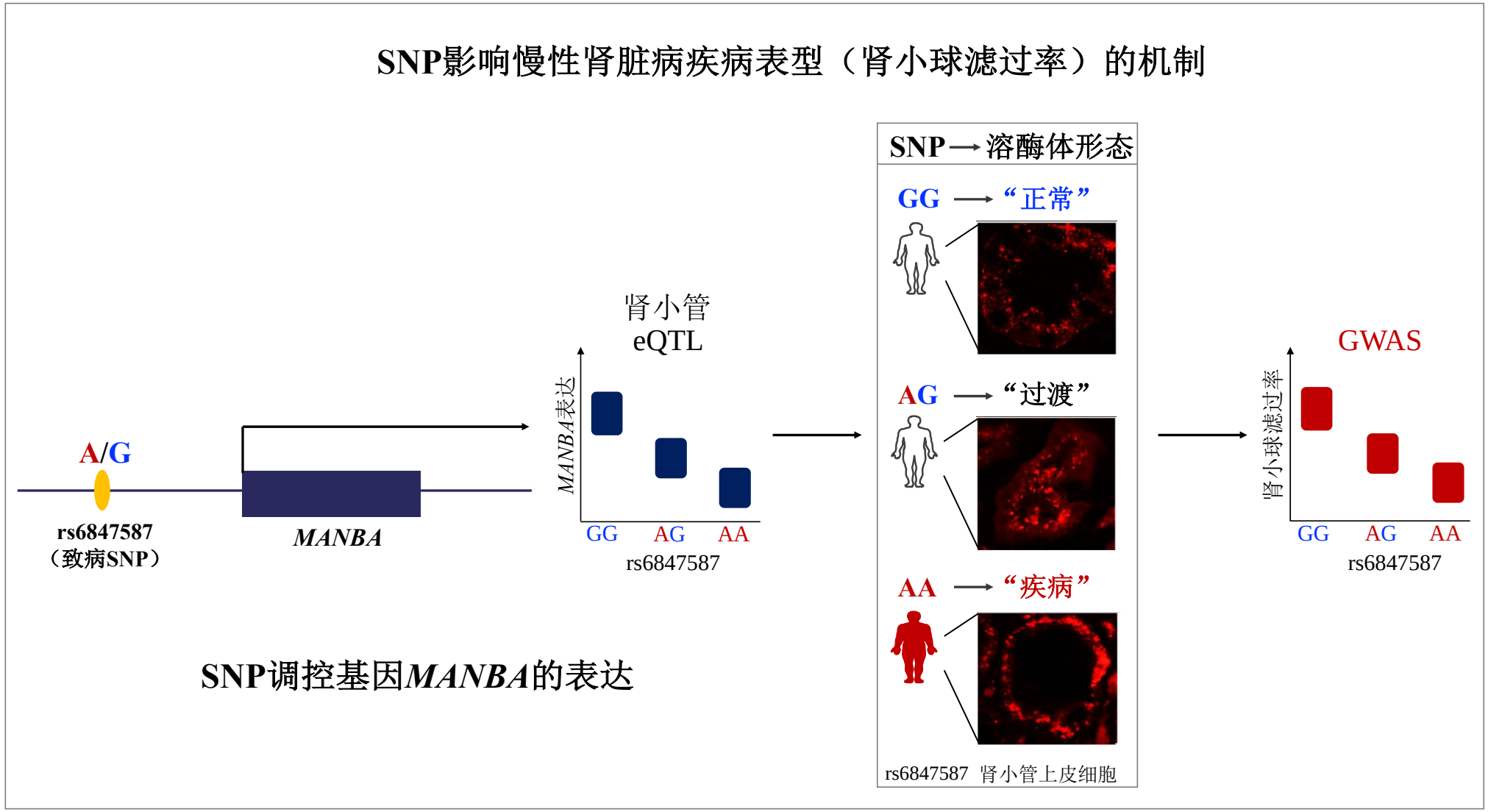


IRF5: an IFN-responsive transcription factor





SNP→基因表达→疾病表型



Gu#, Young#, Sheng# et al., Science Translational Medicine, 2021

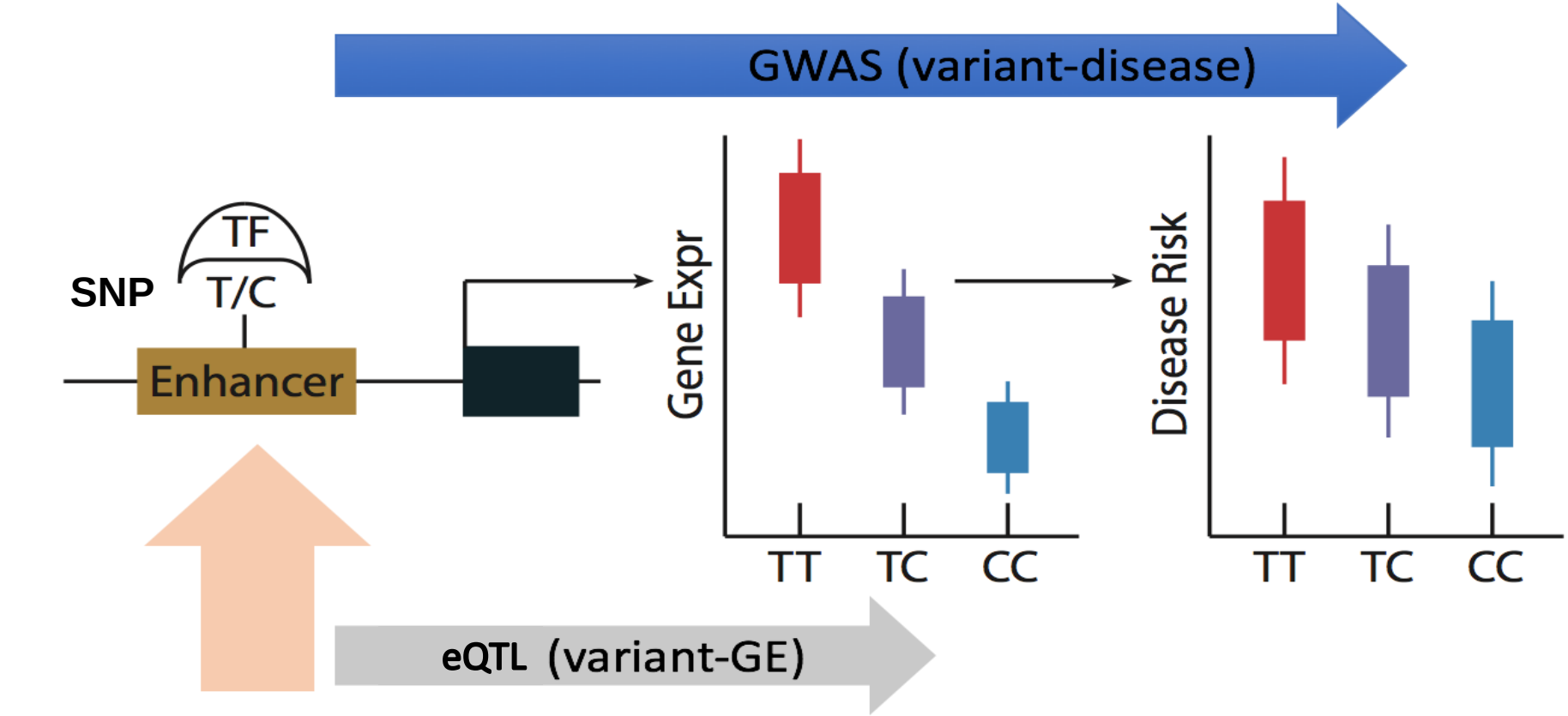
小结

- 基因组上单核苷酸位点的遗传变异（SNP）可能会引起其邻近区域的甲基化水平、染色质开放程度，以及基因表达等数量性状的变化。
- 以GWAS数据为指导，对数量性状位点（xQTL）的多组学整合分析，将有利于识别疾病的潜在致病基因，并解析疾病的发病机制。
- 贝叶斯共定位分析、孟德尔随机化方法是整合多组学分析的常用方法。

内容大纲

- 基本概念
- 慢性肾脏病及其并发症
- 多组学数据整合分析方法
- **肾脏具有高度细胞异质性**
- 从单细胞水平解析SNP-基因-CKD机制
- 总结

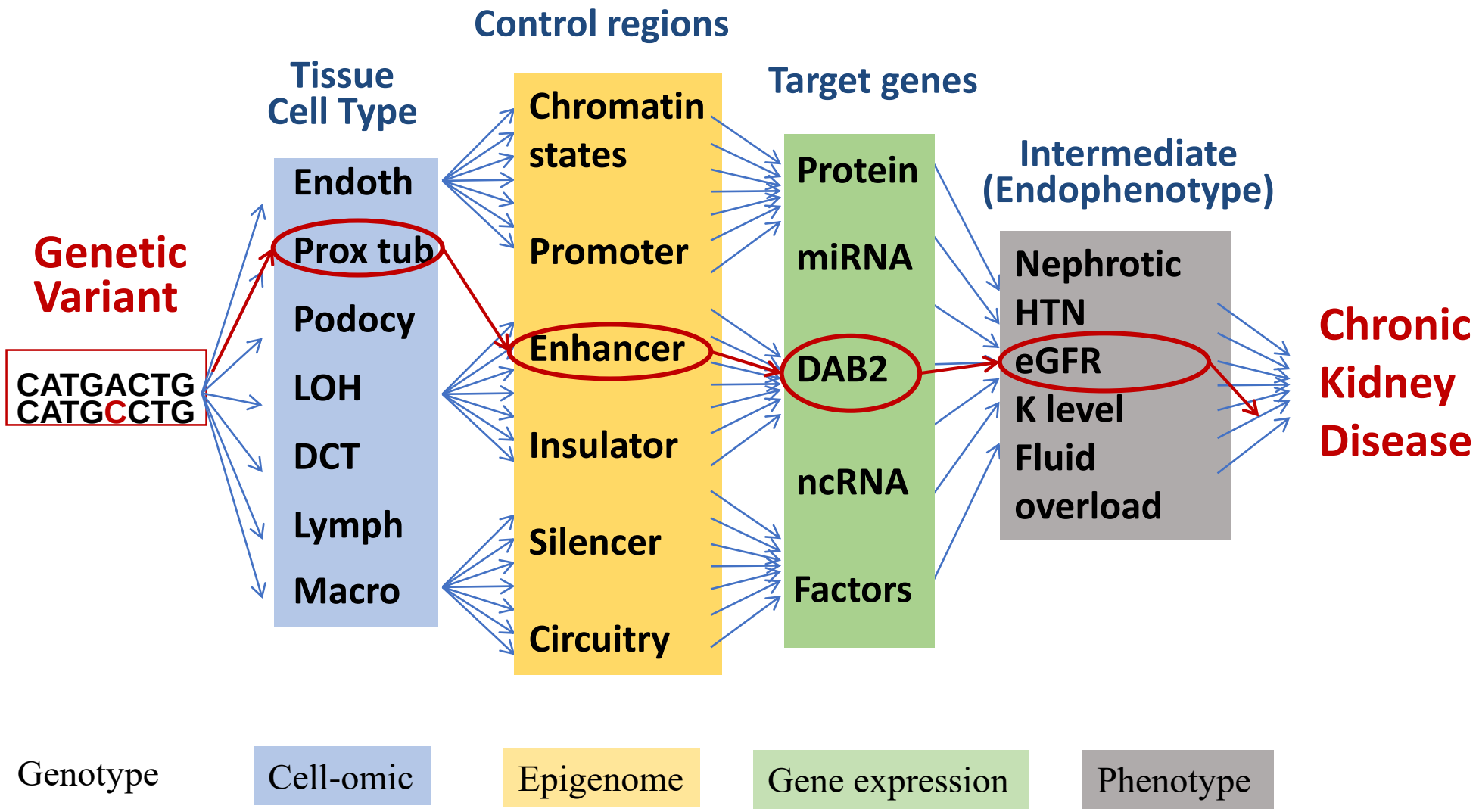
多组学整合分析：解析遗传变异的调控机制



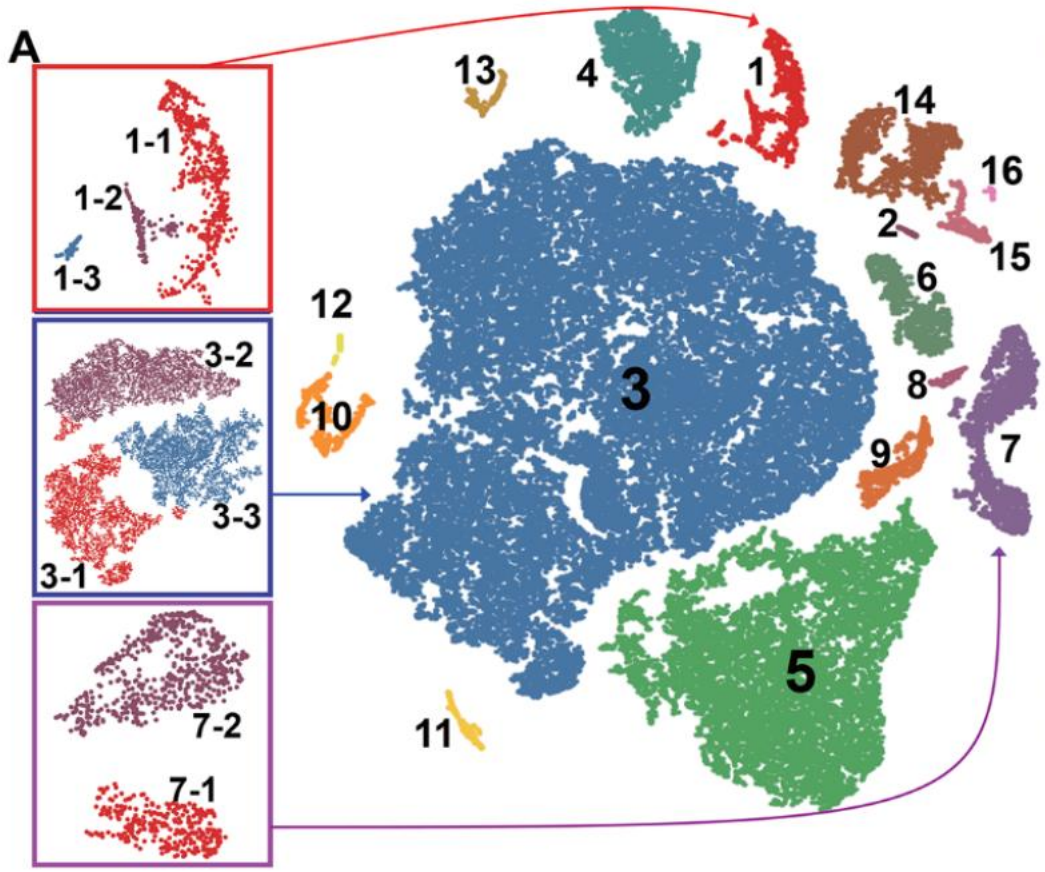
Defining regulatory regions
Chromatin immunoprecipitation

Mechanism

多组学整合分析：精准定位遗传因素导致CKD的具体细胞



肾脏中发现新的细胞类型

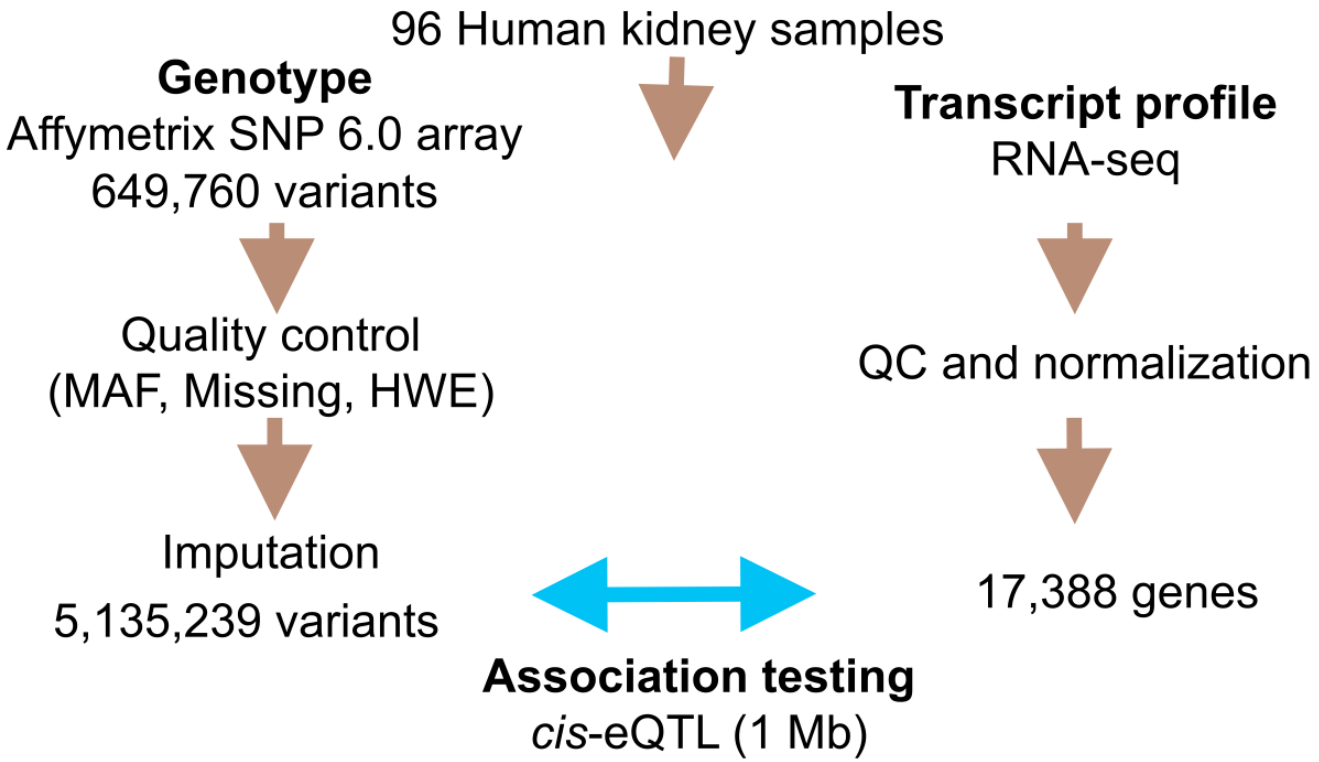


cluster	nCells	%	
1	1,001	2.29	Endo
2	78	0.18	Podo
3	26,482	60.54	PT
4	1,581	3.61	LOH
5	8,544	19.53	DCT
6	870	1.99	CD-PC
7	1,729	3.95	CD-IC
8	110	0.25	CD-Trans
9	601	1.37	Novel1
10	549	1.26	Fib
11	228	0.52	Macro
12	74	0.17	Neutro
13	235	0.54	B lymph
14	1,308	2.99	T lymph
15	313	0.72	NK
16	42	0.1	Novel2

Park et al, *Science 2018*

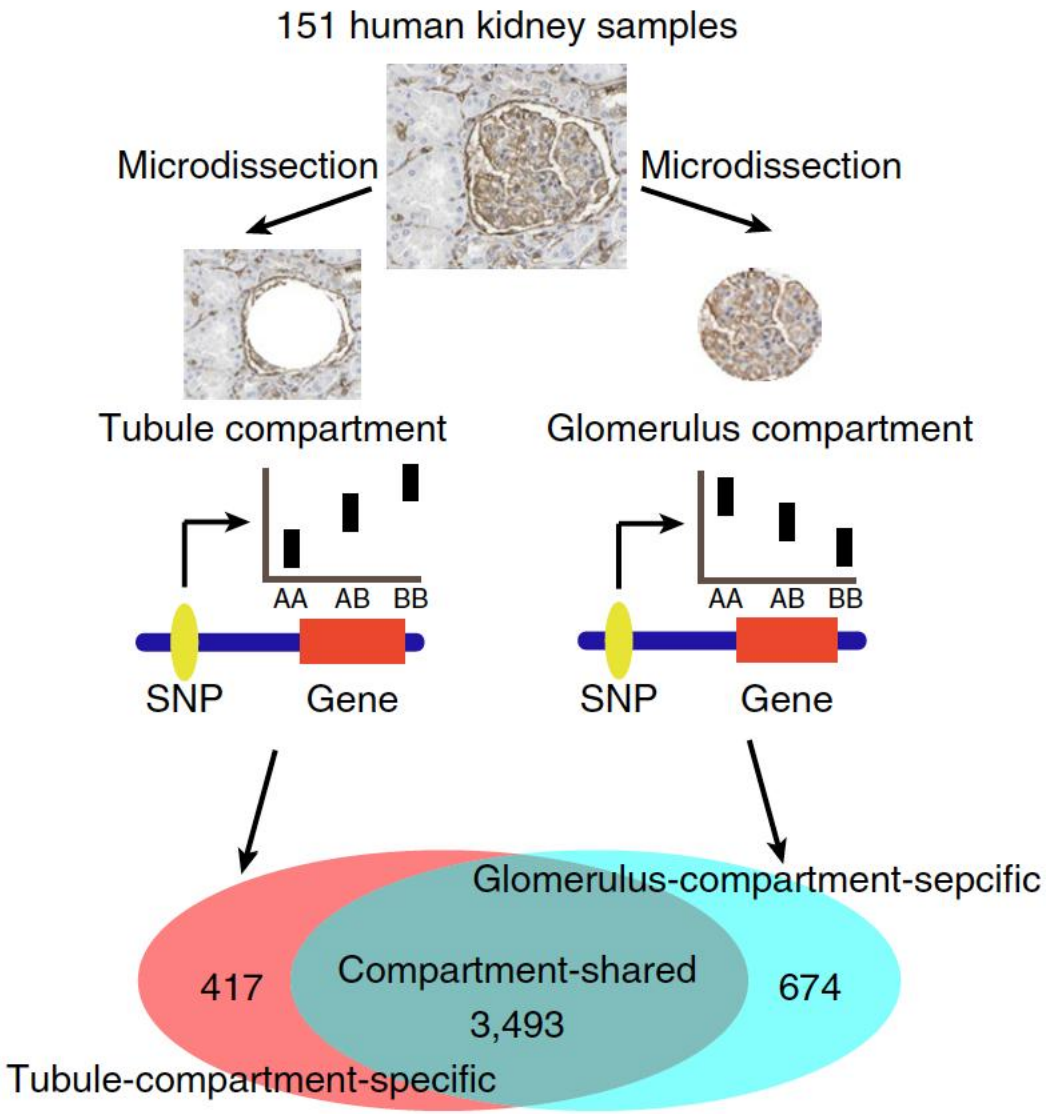
整个肾脏（Whole Kidney）

Only 9 putative causal genes were identified.



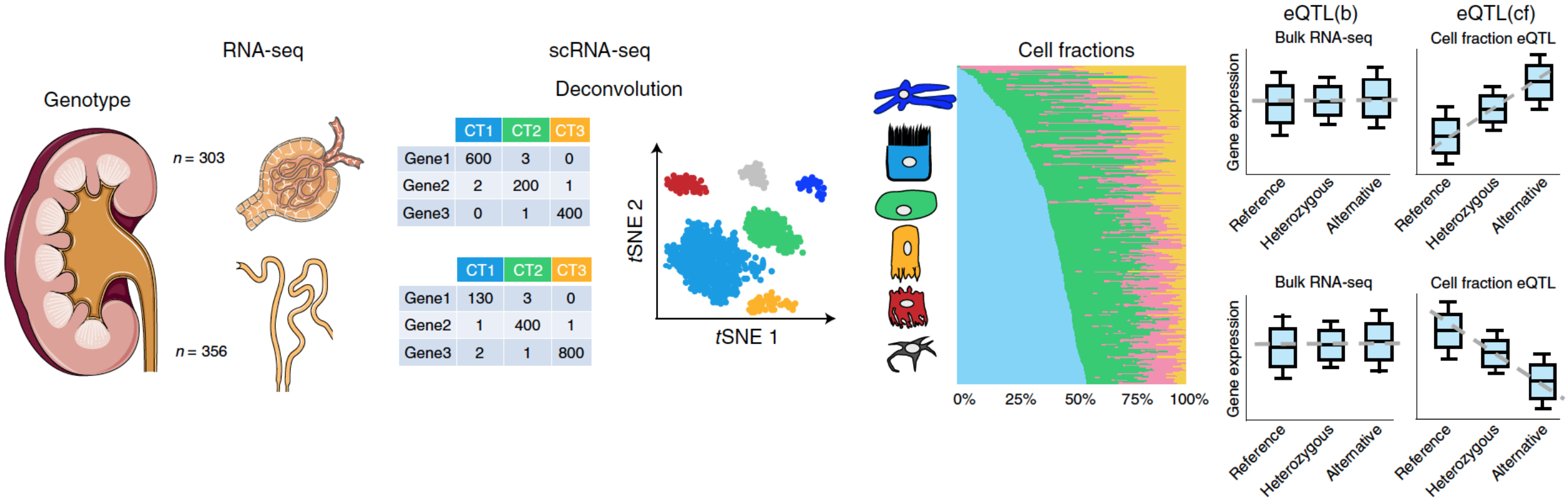
Ko et al, *AJHG 2017*

显微切割：肾小球与肾小管(Glom and Tubule)



Qiu et al, *Nature Medicine 2018*

考虑个体细胞组分的差异



eQTL的计算方式

（a）常用计算方法：

线性回归：lm (Gene Exp ～ SNP dosage + Confounders)

Confounders: age, gender, genetic PCs, and PEER factors (estimated hidden confounders)

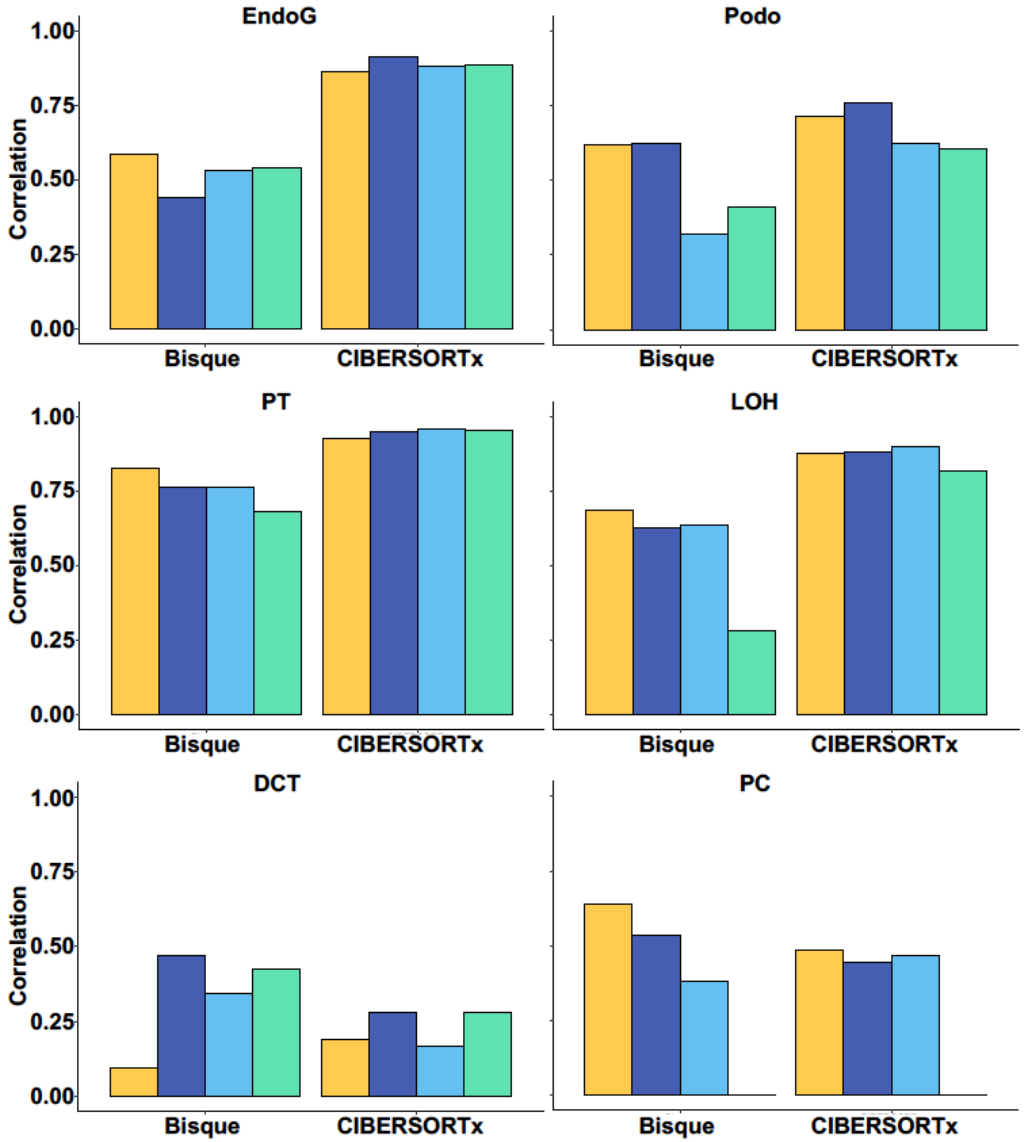
（b）Cell fraction eQTLs:

线性回归：lm (Gene Exp ～ SNP dosage + Confounders)

Confounders: age, gender, genetic PCs, Cell fractions, and PEER factors (estimated hidden confounders)

可能存在的问题：相似细胞之间的共线性问题，细胞组分较少的细胞类型

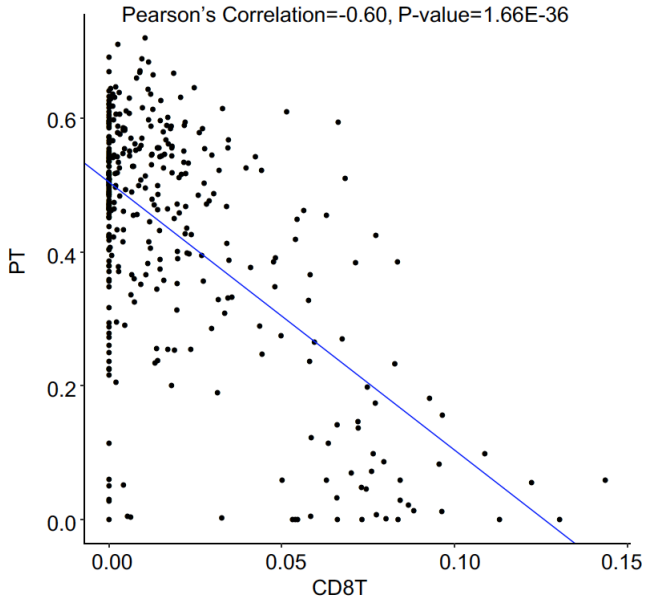
反卷积算法比较

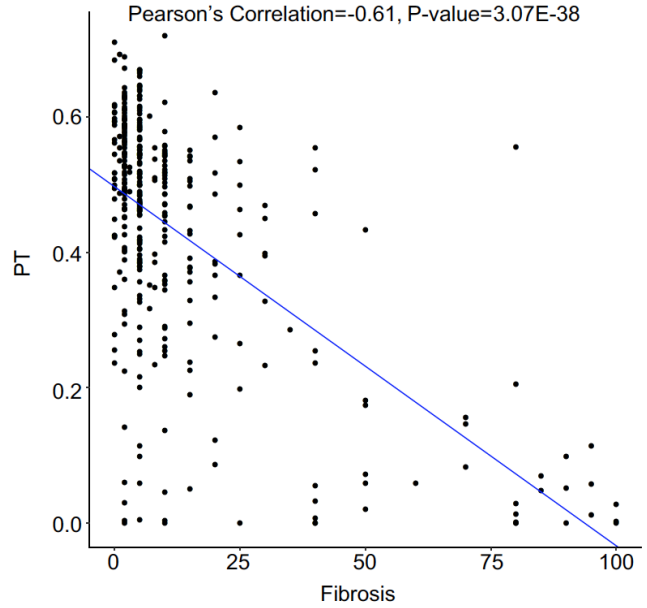

EndoG
Podo
PT
LOH
DCT
PC
Correlation
1.00
0.75
0.50
0.25
0.00
Bisque
CIBERSORTx
Our Study
Park et al.
Wilson et al.
Young et al.

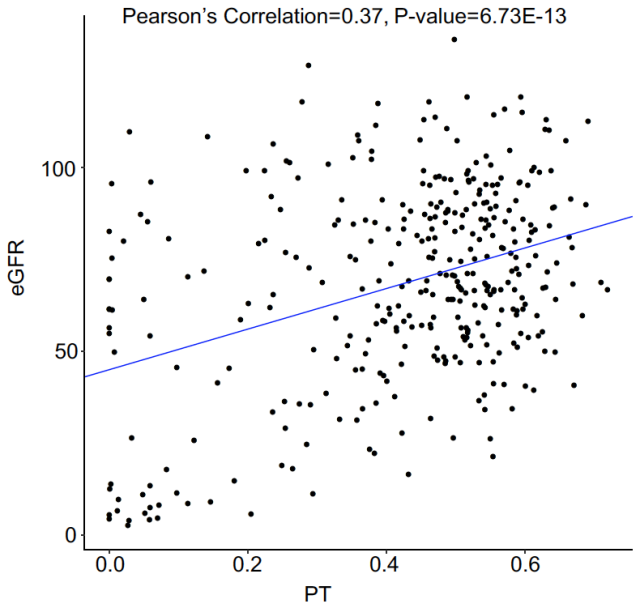
个体细胞组分随疾病状态改变

Cell Type	Pearson's Correlation	Two-sided P-value
CD8T	-0.60	1.66E-36
Macro	-0.68	1.25E-48
Treg	-0.38	2.62E-13
Mono	-0.33	1.19E-10
B	-0.41	9.57E-16

Two-sided P-value was calculated by t-test (df=354).

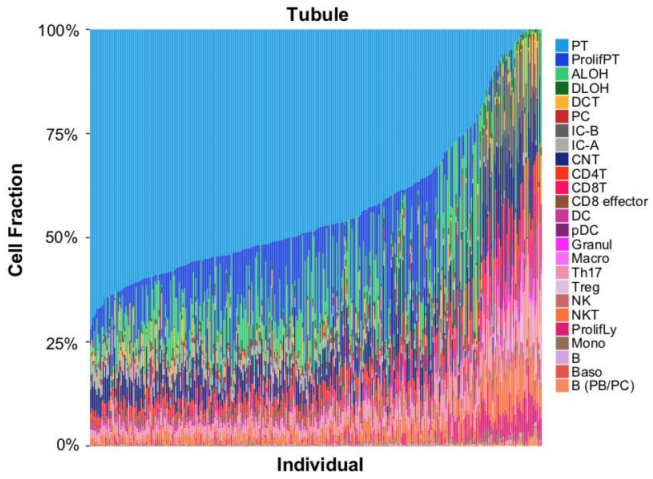


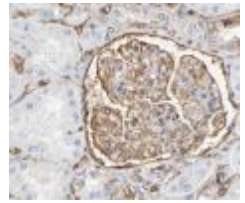




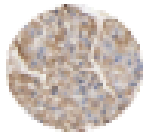
考虑个体细胞组分的差异

359 human kidneys

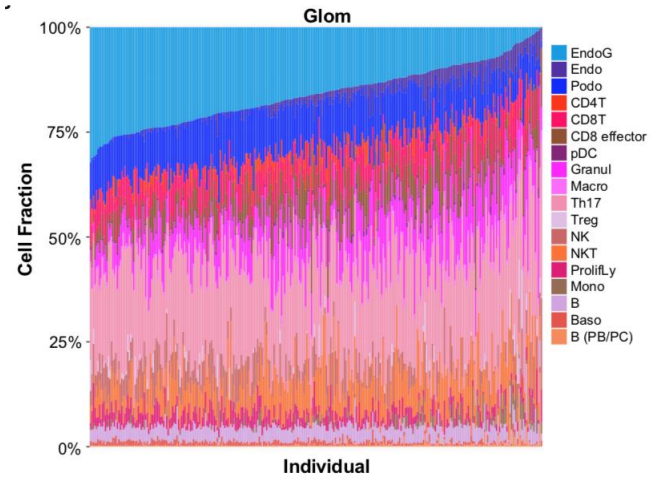


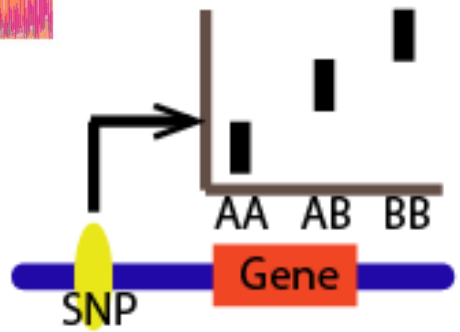


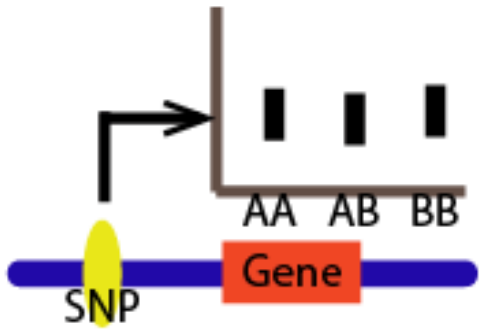
tubule



glomerulus







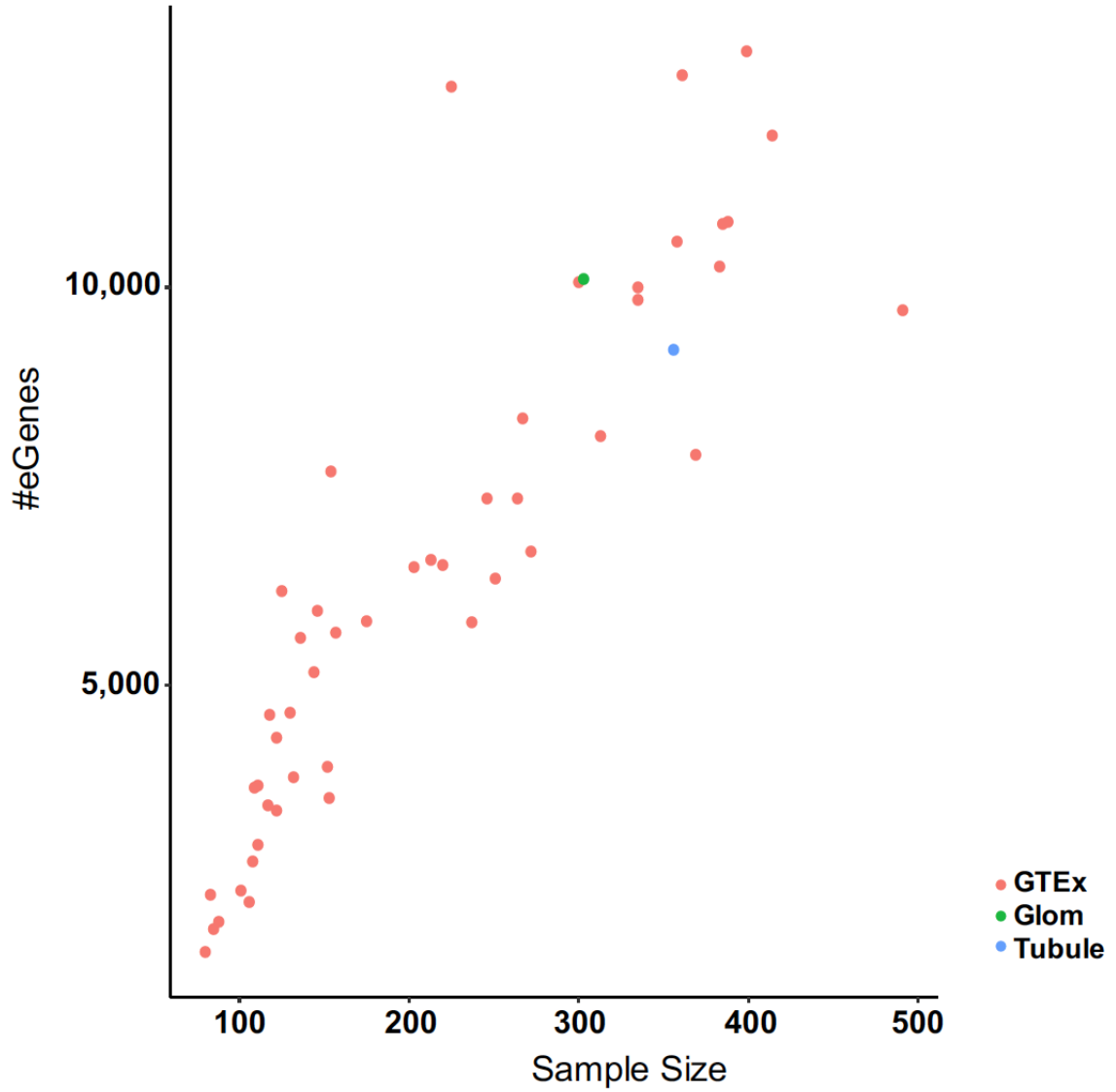
9,723 eGenes

9,209 eGenes

(eGene: gene contains at least one eQTL variant)

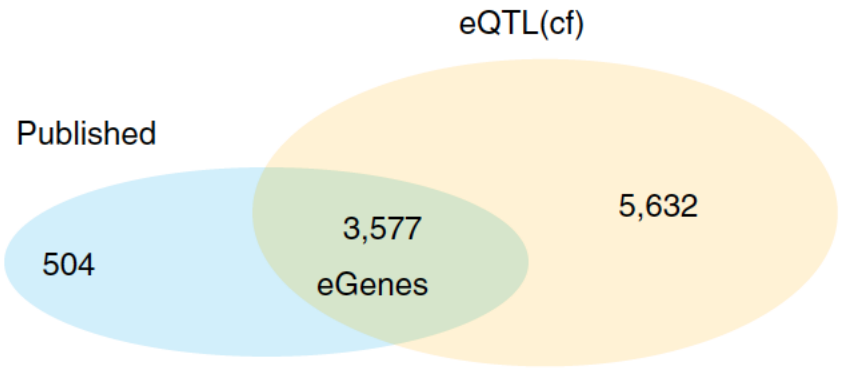
Sheng et al, *Nature Genetics 2021*

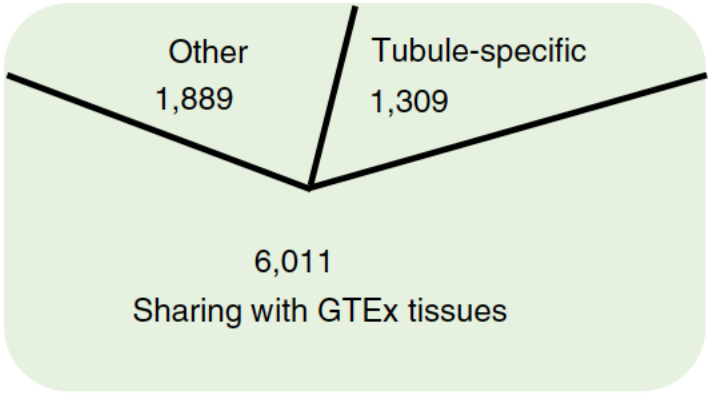
样本数量与eGene个数的线性关系



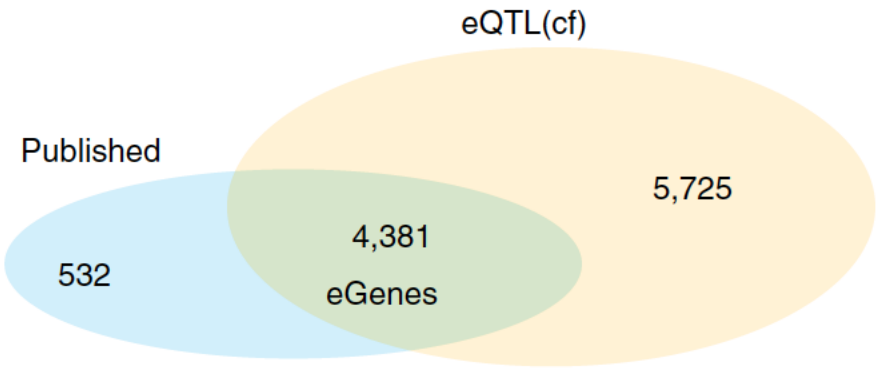
新鉴定的SNP-Gene调控

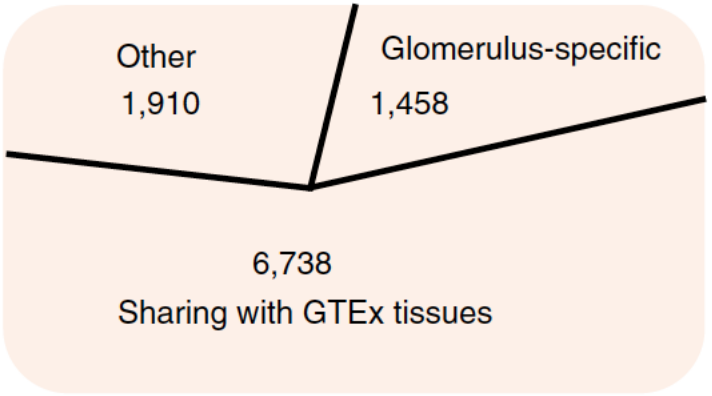
Tubule



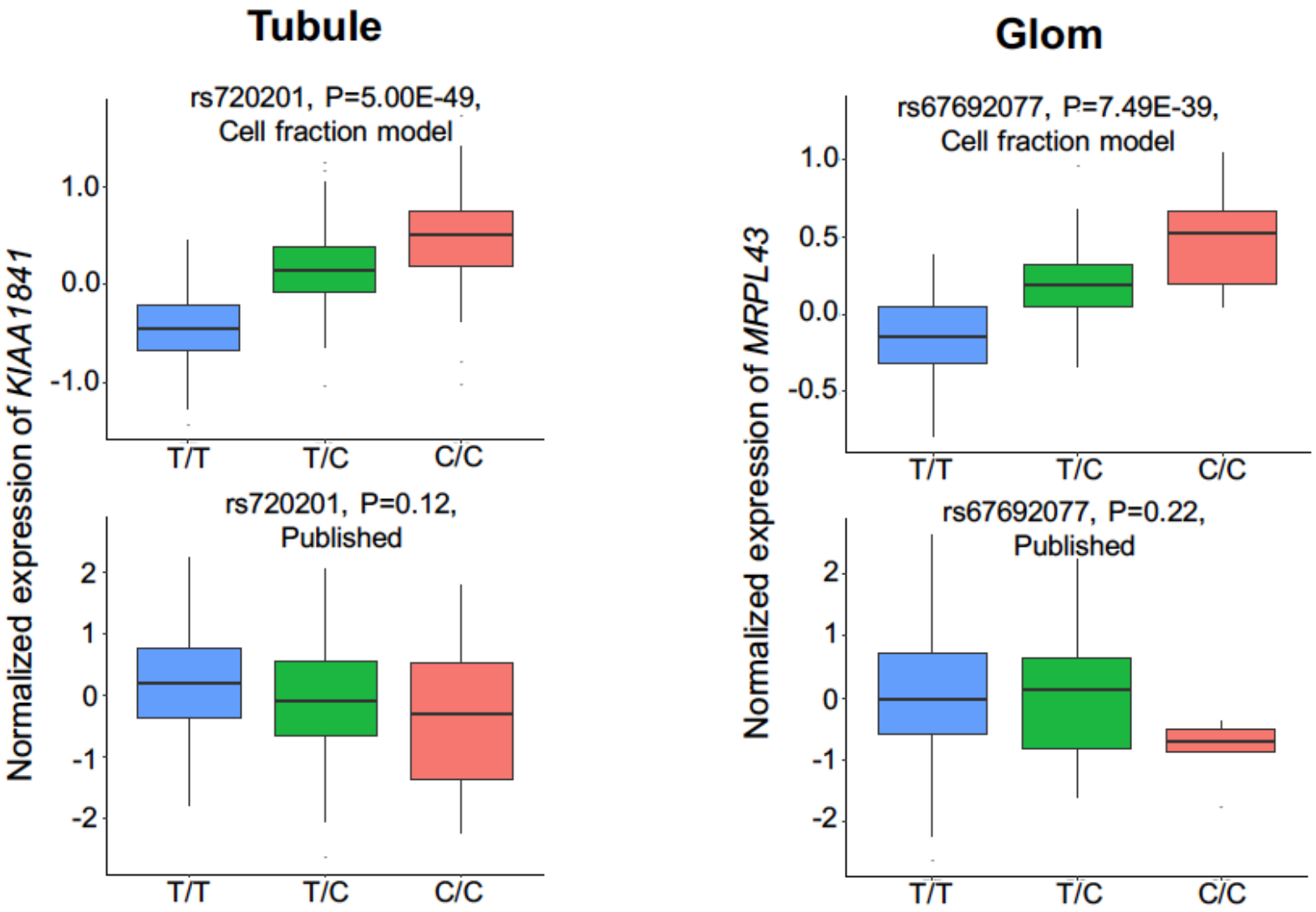


Glom

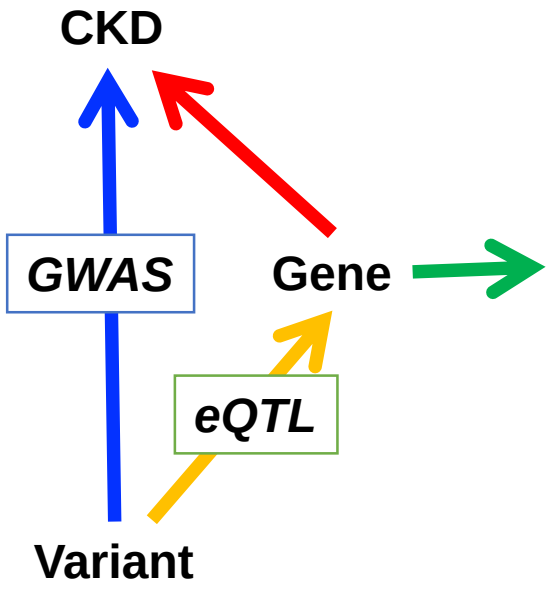


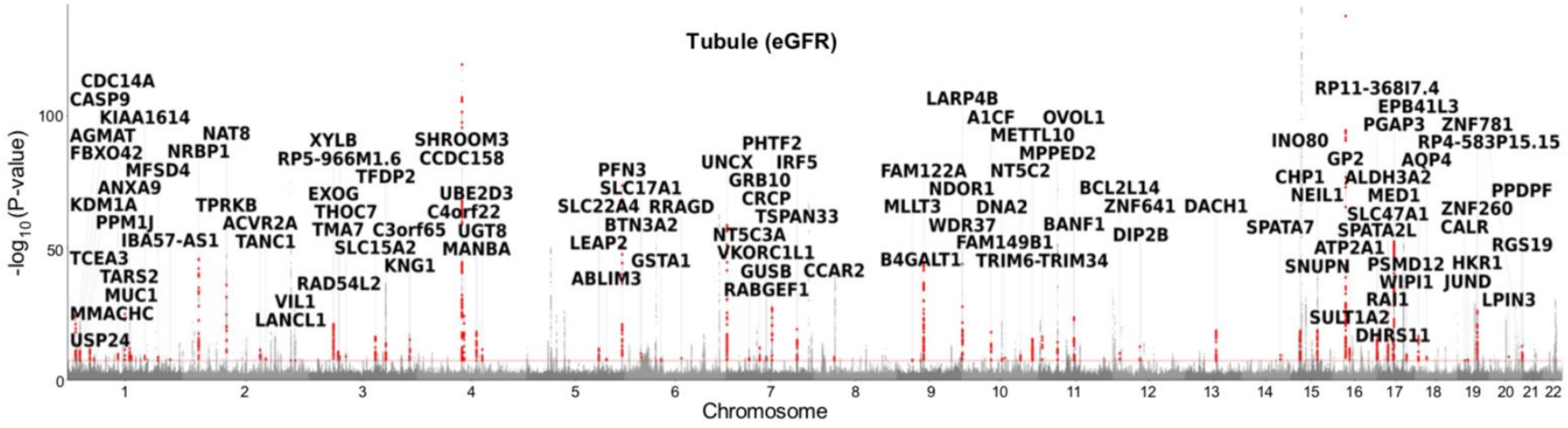


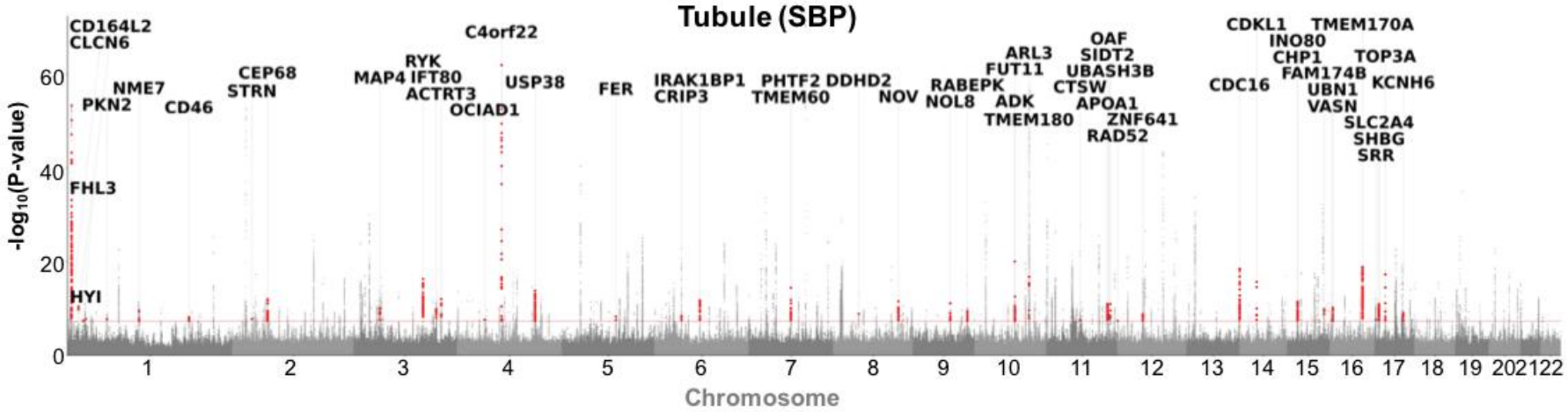
Cell fraction eQTLs



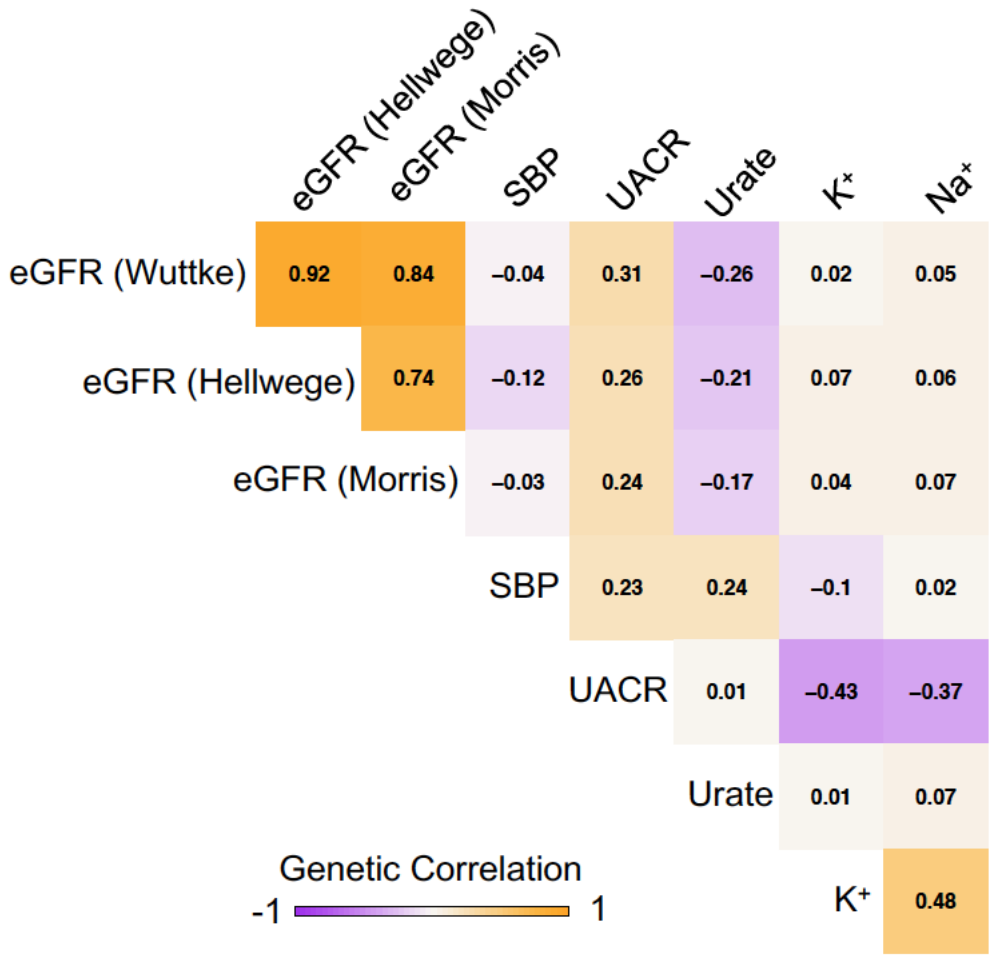
CKD和高血压的潜在致病基因







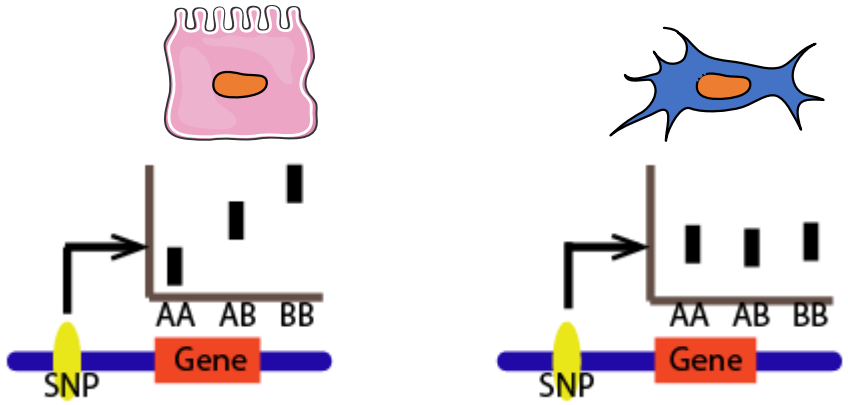
CKD疾病表型的相关性

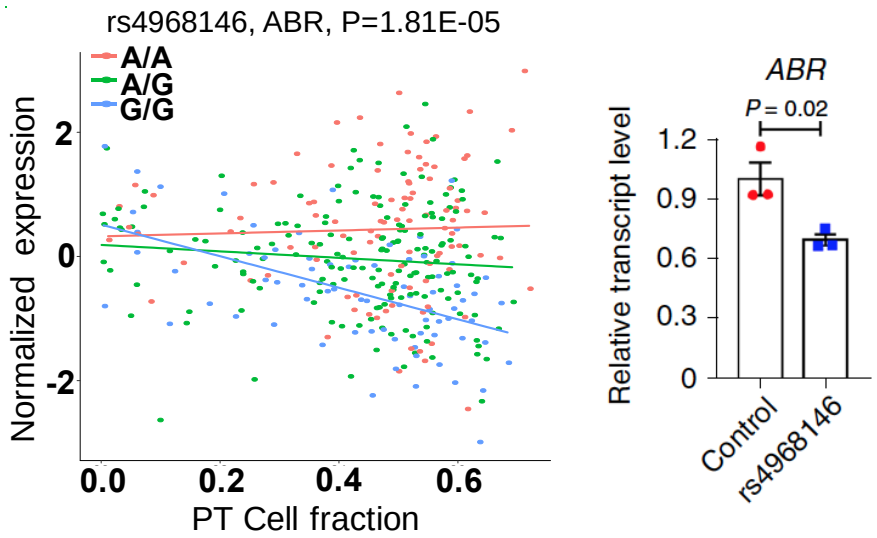


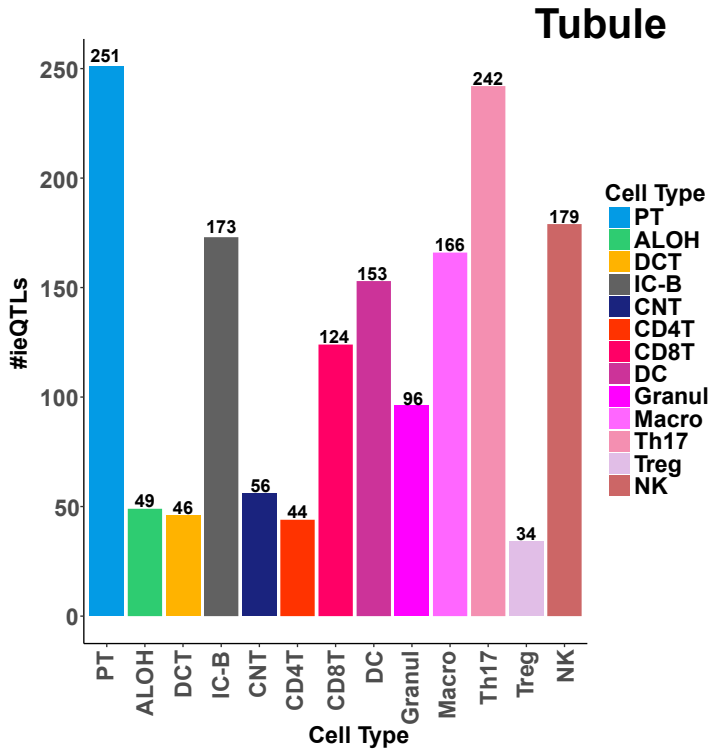
内容大纲

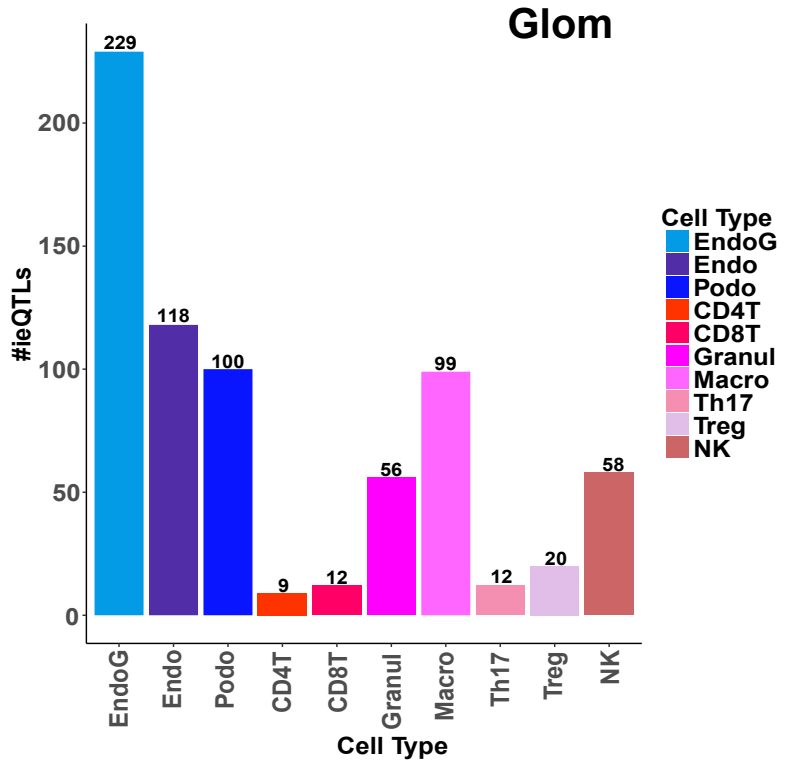
- 基本概念
- 慢性肾脏病及其并发症
- 多组学数据整合分析方法
- 肾脏具有高度细胞异质性
- **从单细胞水平解析SNP-基因-CKD机制**
- 总结

与细胞“互作”的：SNP-Gene调控







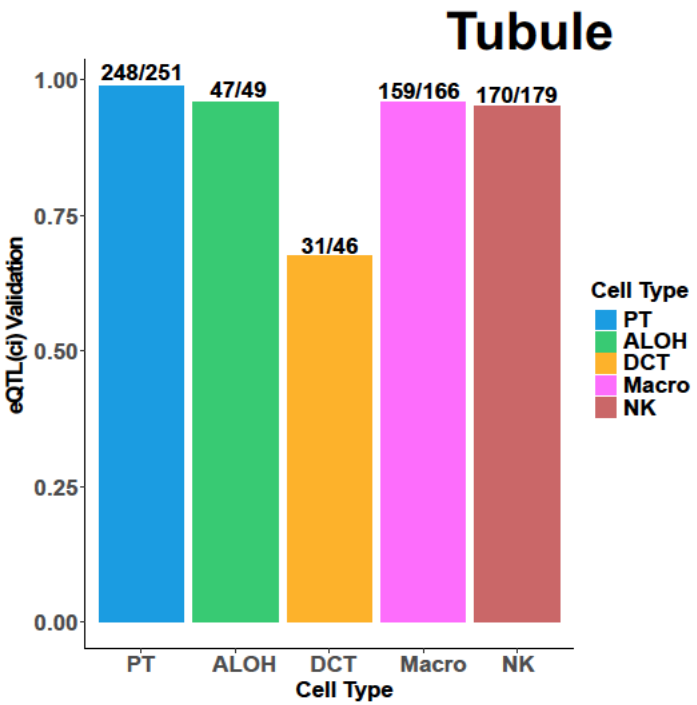


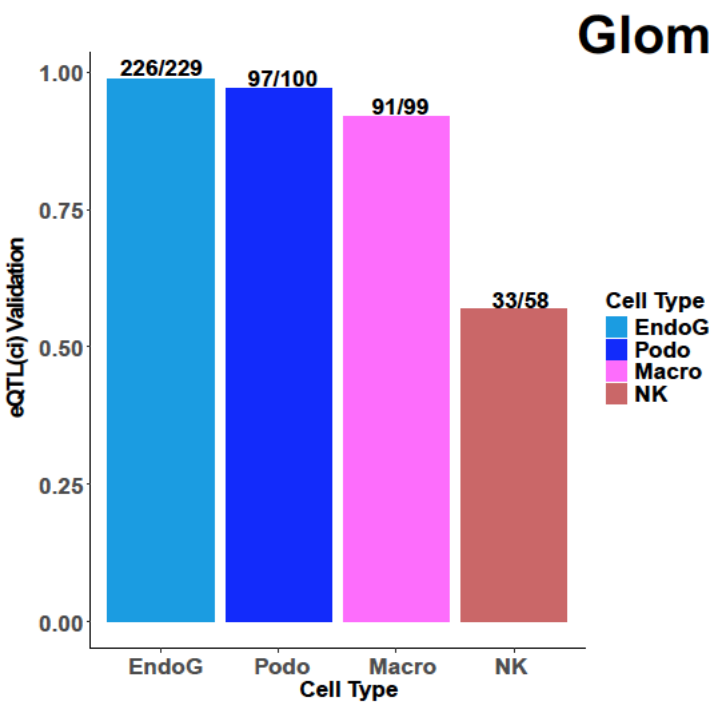
To identify eQTL(ci):

H_0: INT transformed gene expression~Beta2×SNP dosage + Confounders (1)

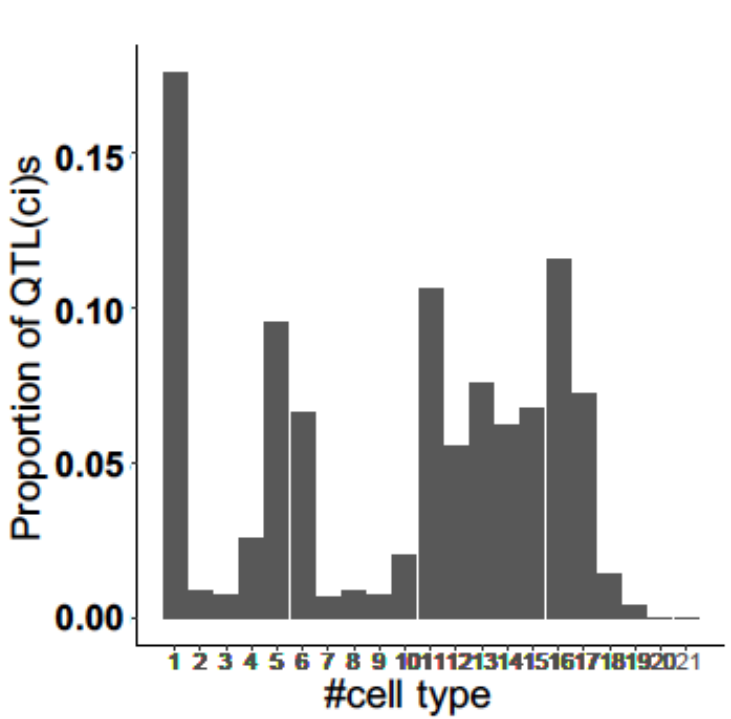
H_1: INT transformed gene expression~Beta1×**SNP dosage:cell fraction of interested cell type**+Beta2×SNP dosage + Confounders (2)

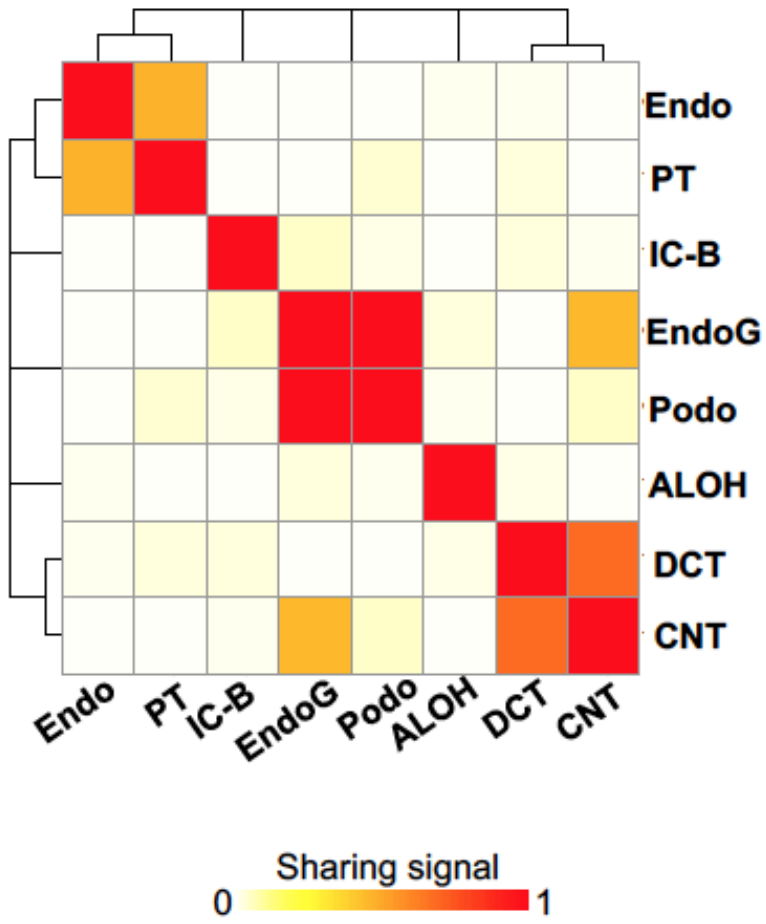
证明算法的可重复性

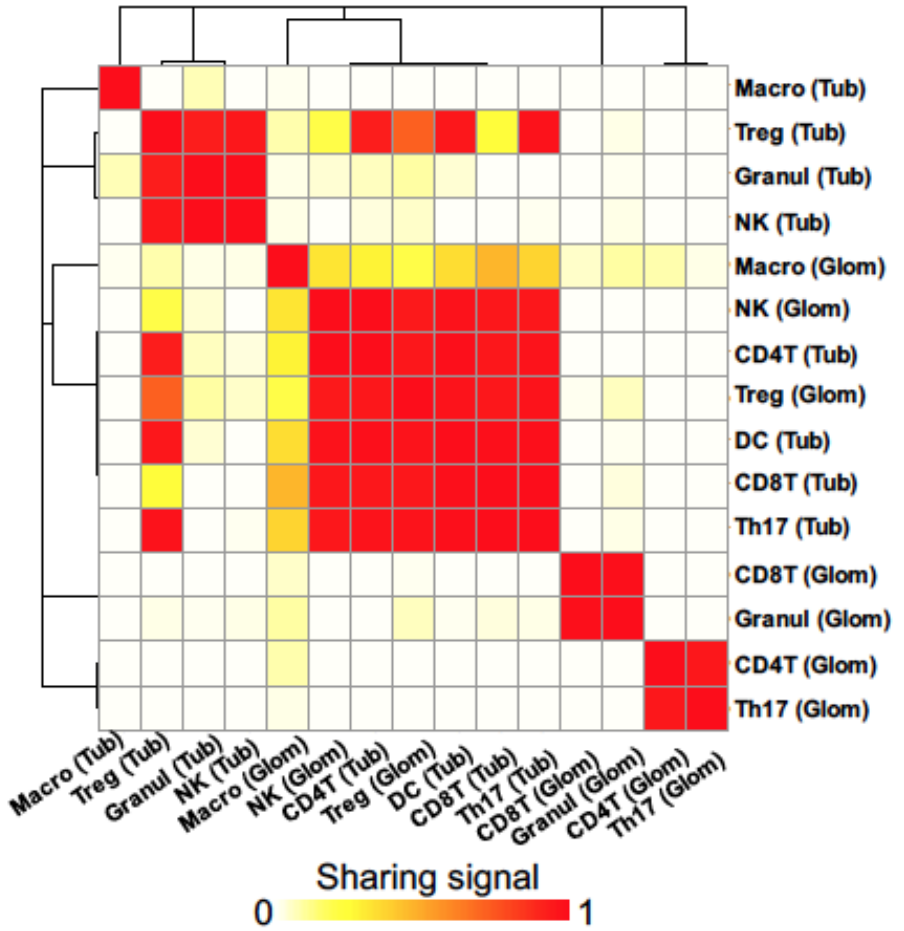

Tubule
eQTL(ci) Validation
1.00
0.75
0.50
0.25
0.00
248/251
47/49
31/46
159/166
170/179
PT
ALOH
DCT
Macro
NK
Cell Type
Cell Type
PT
ALOH
DCT
Macro
NK


Glom
eQTL(ci) Validation
1.00
0.75
0.50
0.25
0.00
226/229
97/100
91/99
33/58
EndoG
Podo
Macro
NK
Cell Type
Cell Type
EndoG
Podo
Macro
NK

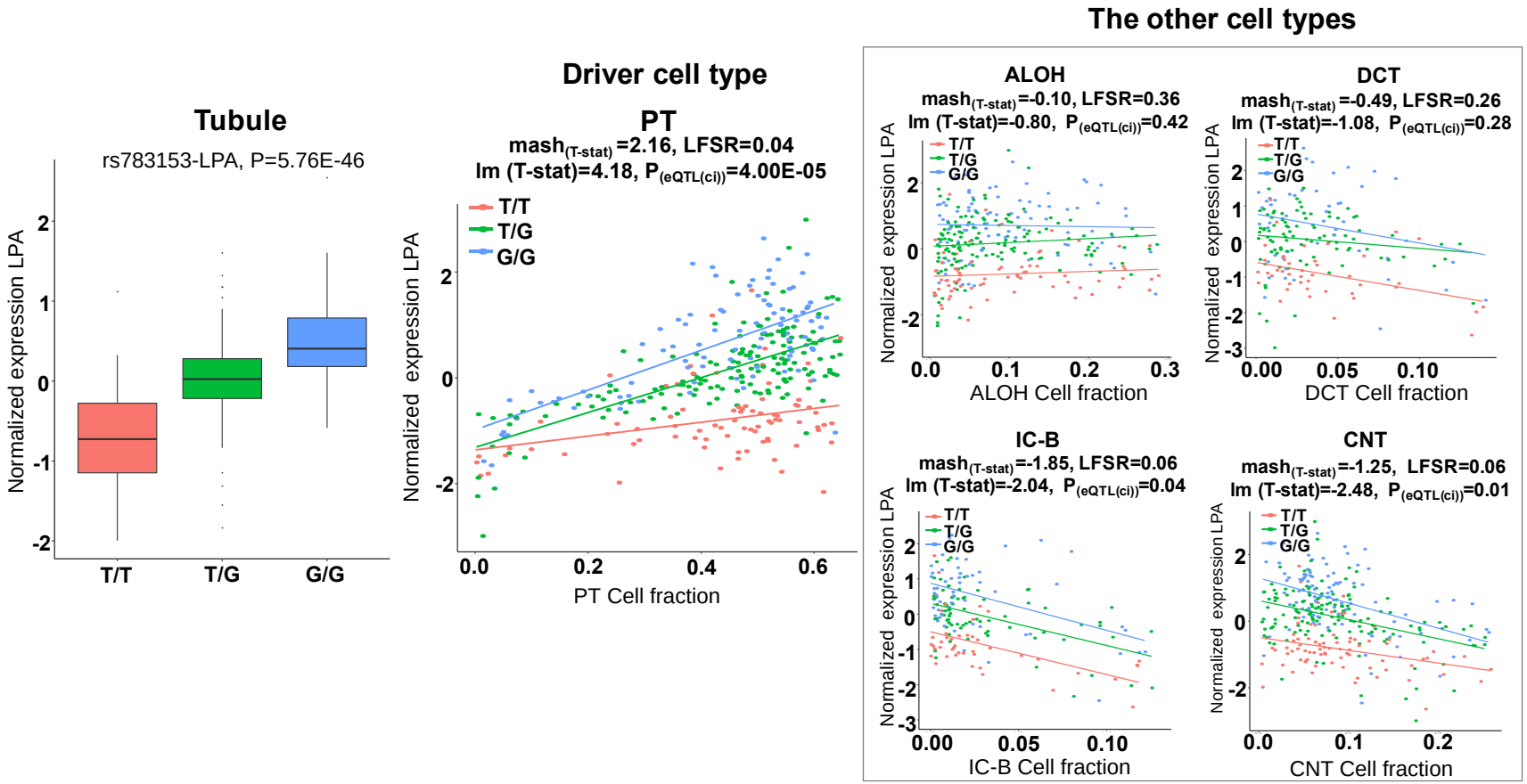
细胞“互作”≠细胞特异性


Proportion of QTL(ci)s
0.15
0.10
0.05
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
#cell type

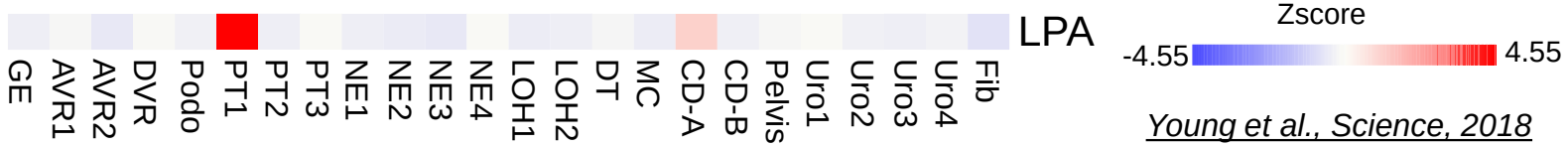

Endo
PT
IC-B
EndoG
Podo
ALOH
DCT
CNT
Sharing signal
0
1


Macro (Tub)
Treg (Tub)
Granul (Tub)
NK (Tub)
Macro (Glom)
NK (Glom)
CD4T (Tub)
Treg (Glom)
DC (Tub)
CD8T (Tub)
Th17 (Tub)
CD8T (Glom)
Granul (Glom)
CD4T (Glom)
Th17 (Glom)
Sharing signal
0
1

细胞特异的eQTL调控

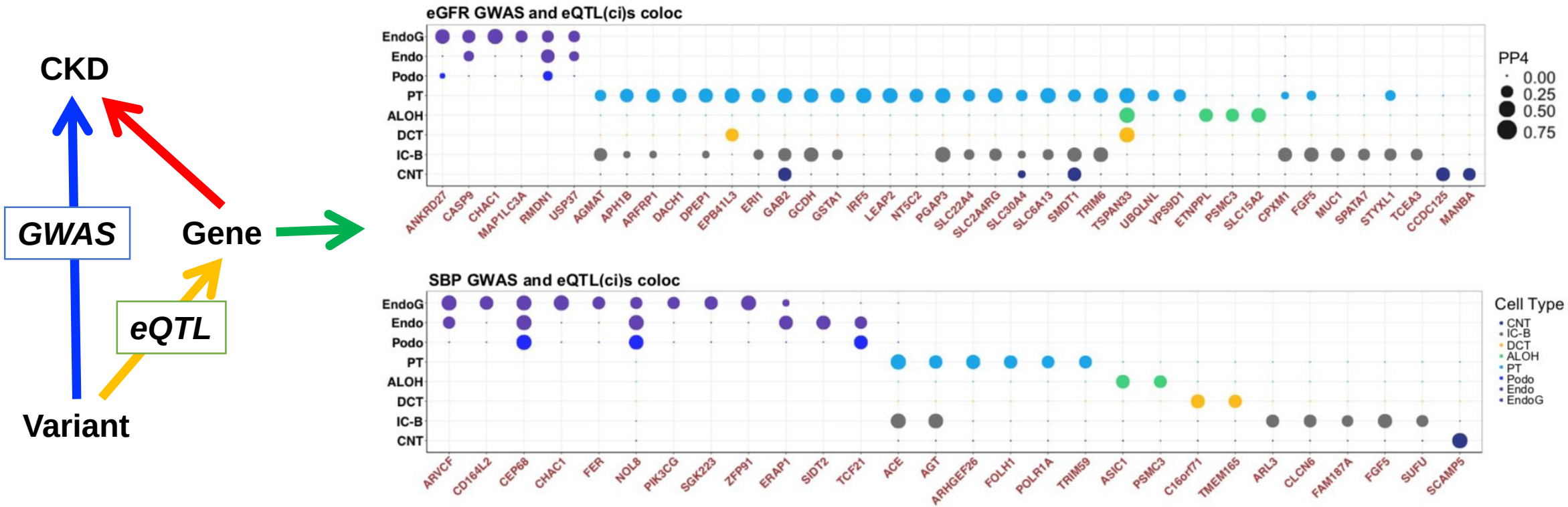


PT-specific expression in human kidney scRNA-seq data



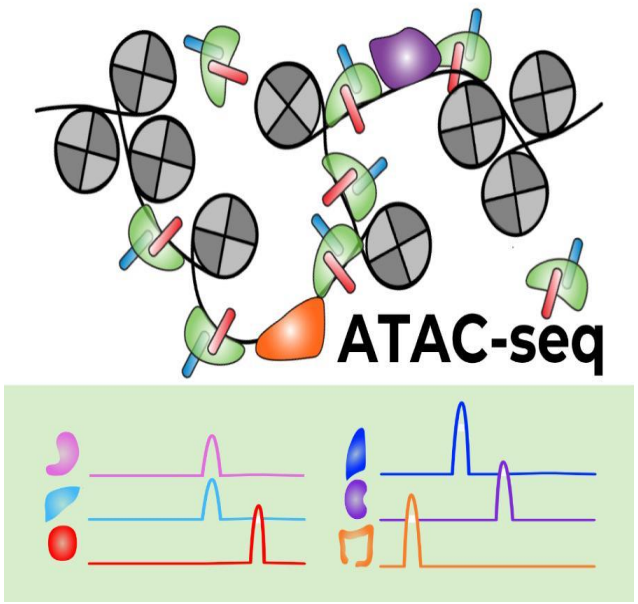
Young et al., Science, 2018

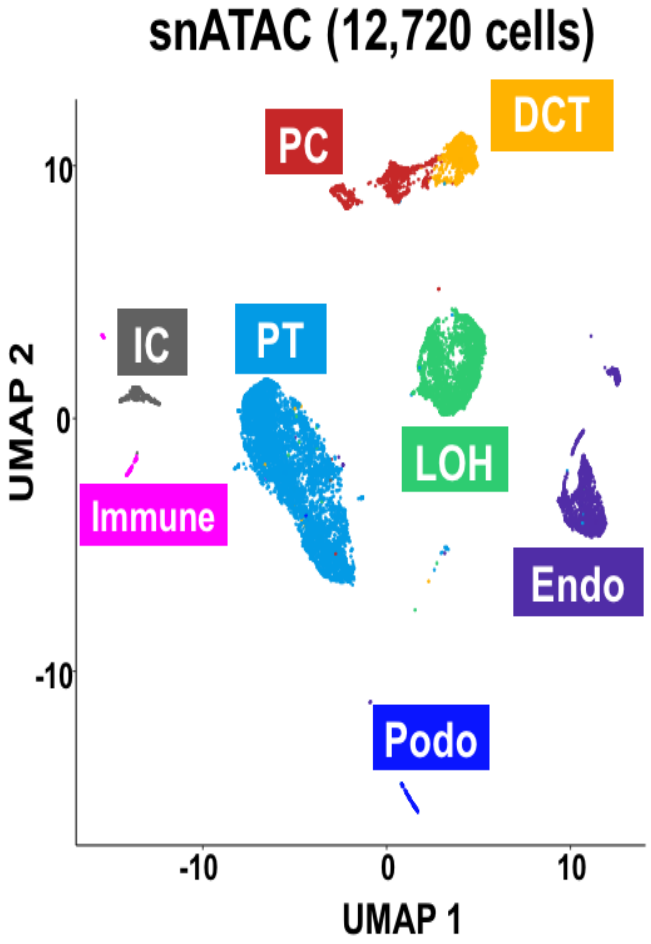
依赖特定细胞的：CKD和高血压的潜在致病基因

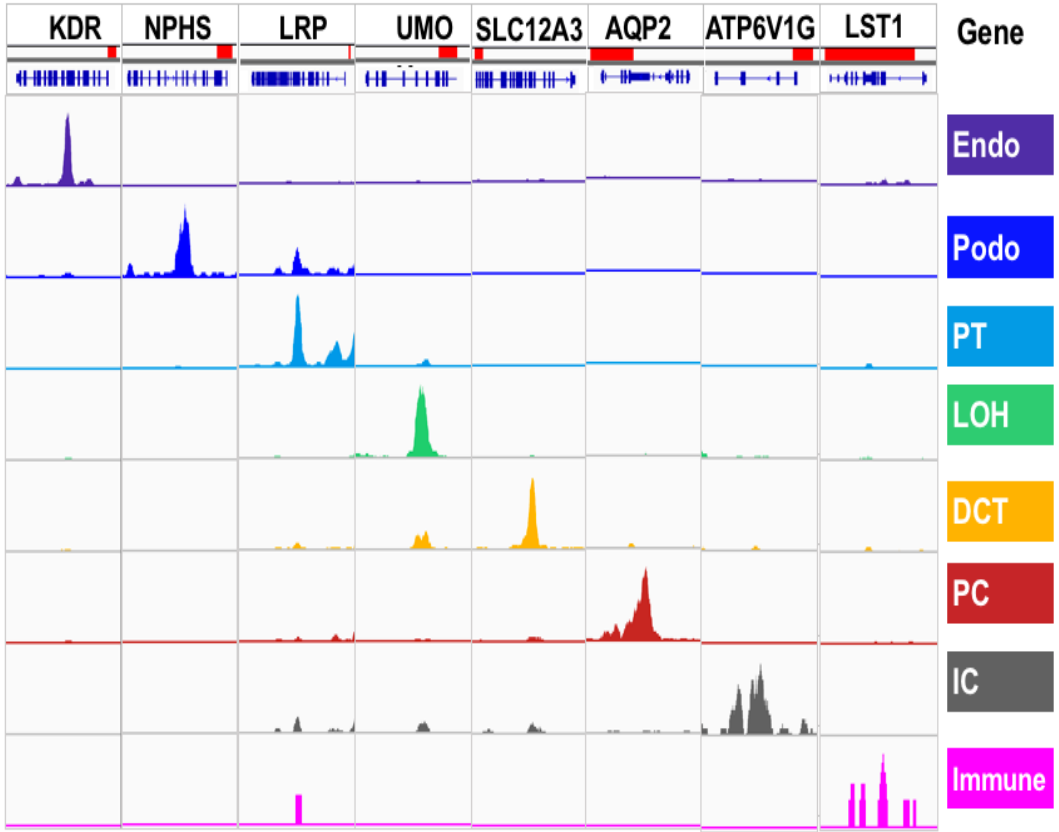


肾脏：单细胞水平染色质可及性

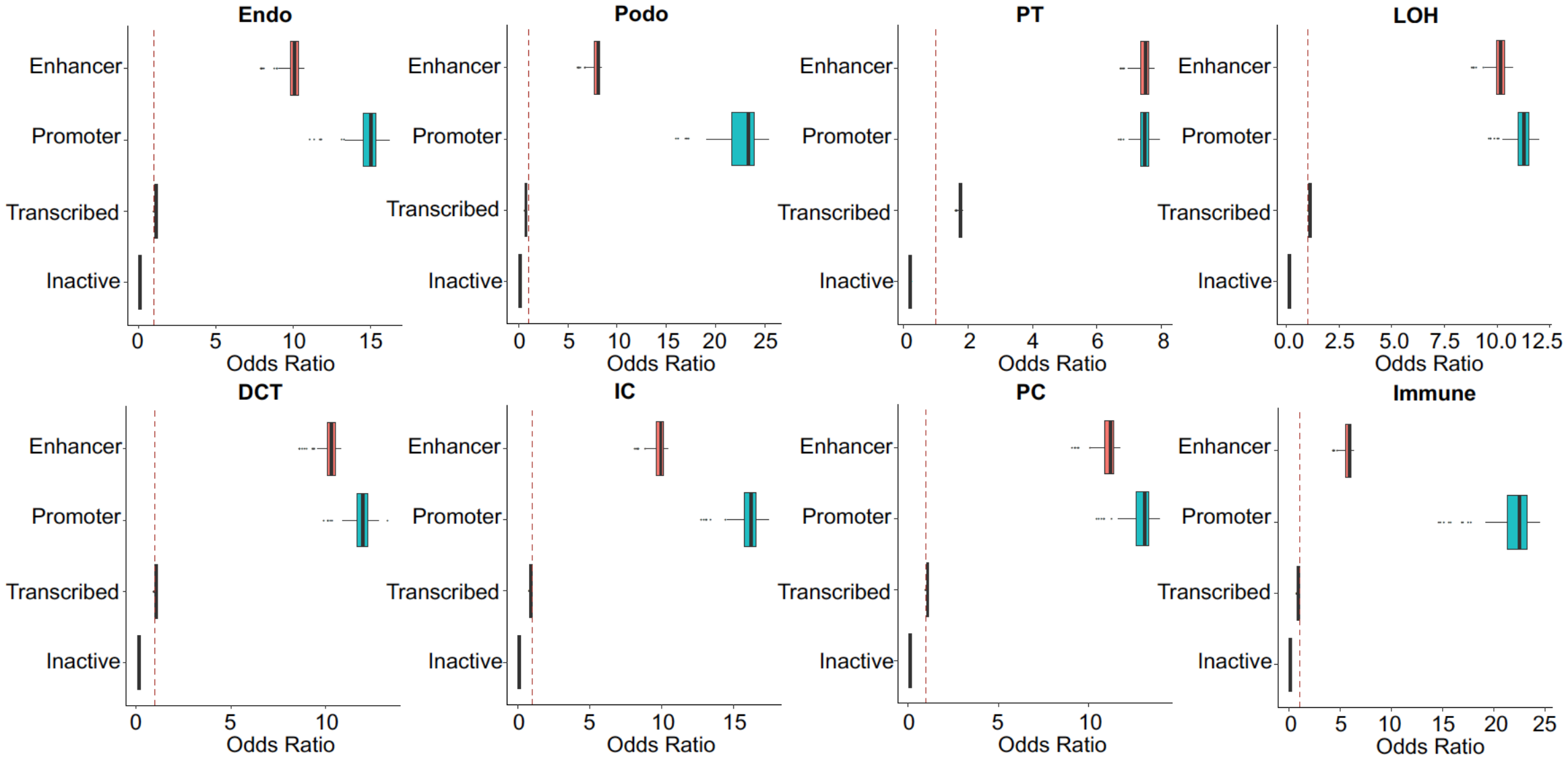
Single nuclear ATAC-seq



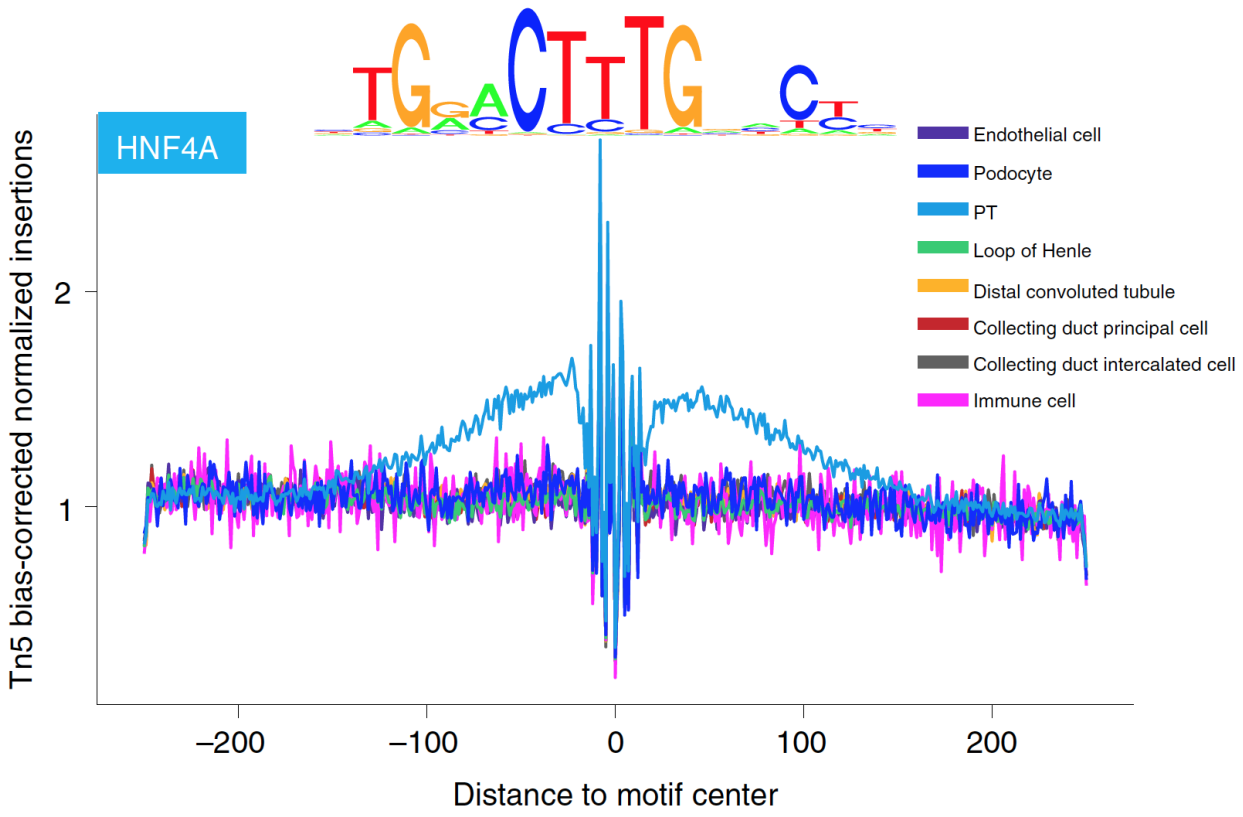


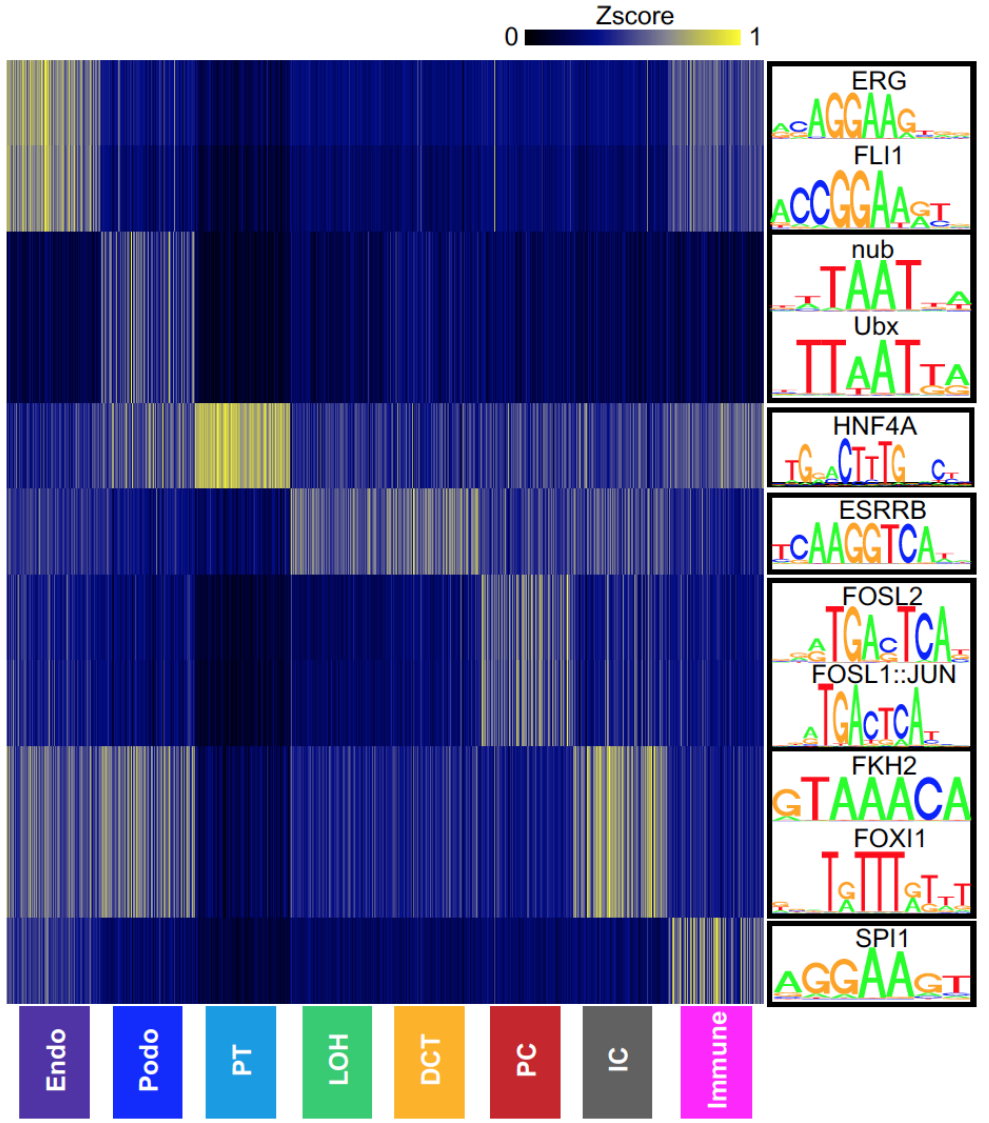


肾脏：单细胞水平开放染色质区域

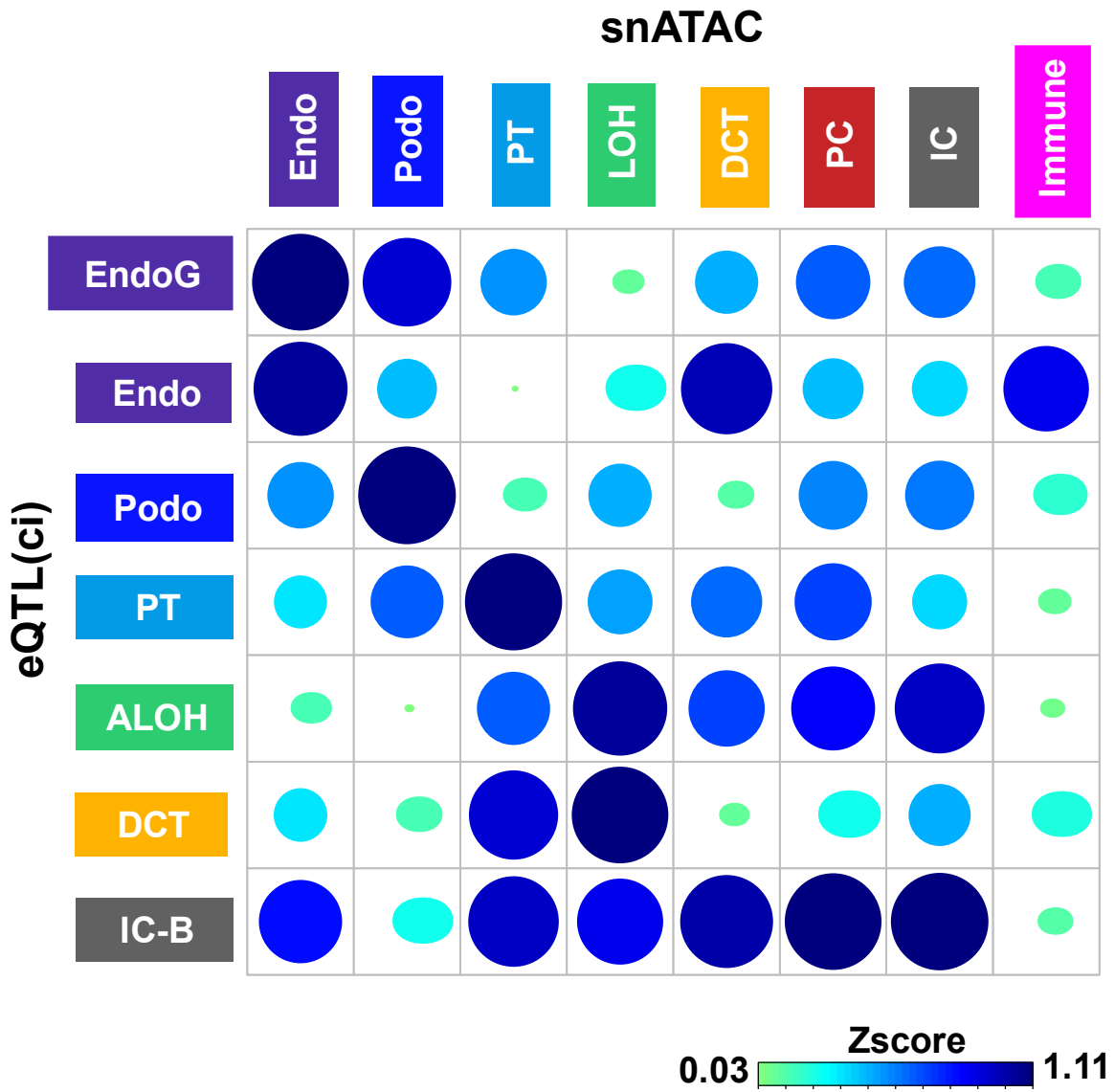


细胞特异的转录因子

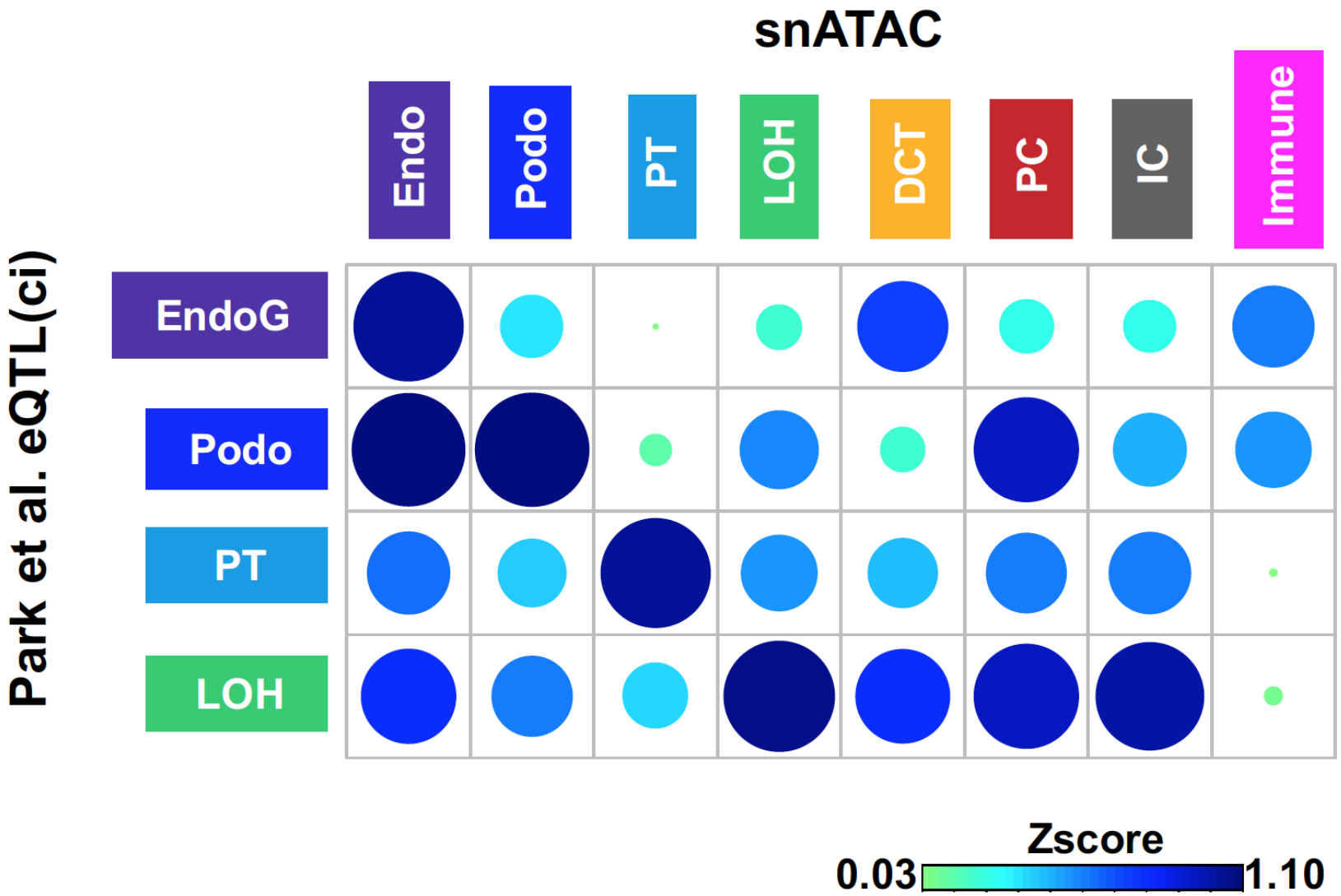

HNF4A
Tn5 bias-corrected normalized insertions
2
1
−200
−100
0
100
200
Distance to motif center
Endothelial cell
Podocyte
PT
Loop of Henle
Distal convoluted tubule
Collecting duct principal cell
Collecting duct intercalated cell
Immune cell


Zscore
0
1
ERG
FLI1
nub
Ubx
HNF4A
ESRRB
FOSL2
FOSL1::JUN
FKH2
FOXI1
SPI1
Endo
Podo
PT
LOH
DCT
PC
IC
Immune

eQTL(ci) SNPs富集在相应细胞的调控区域

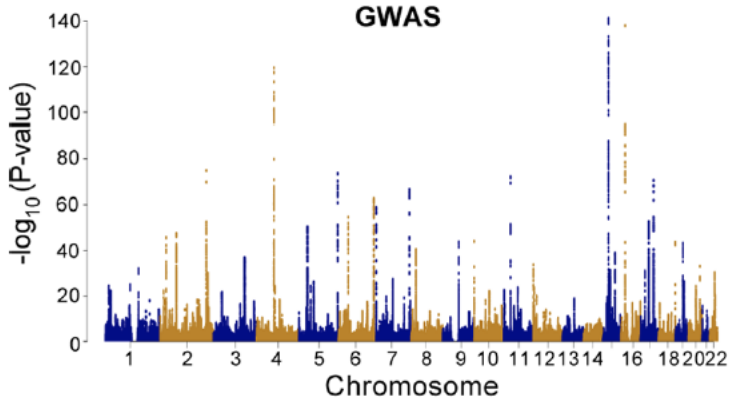


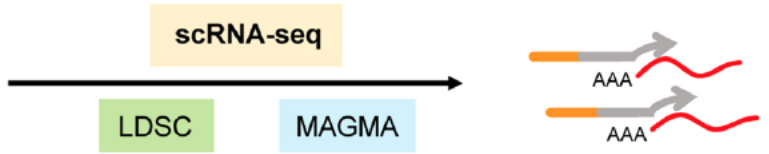
研究结论的可重复性

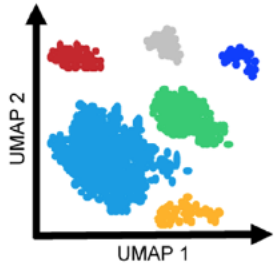


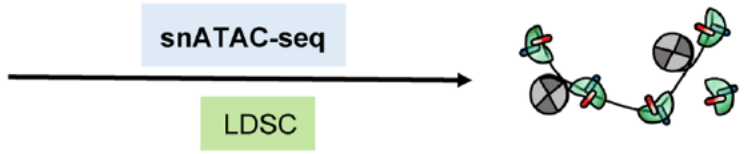
Park et al, *Science 2018*

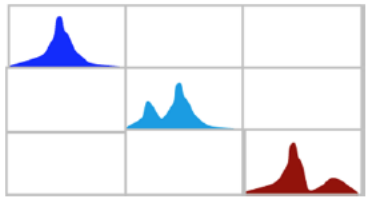
LDSC: 注释肾脏各类细胞所承担的生物学功能

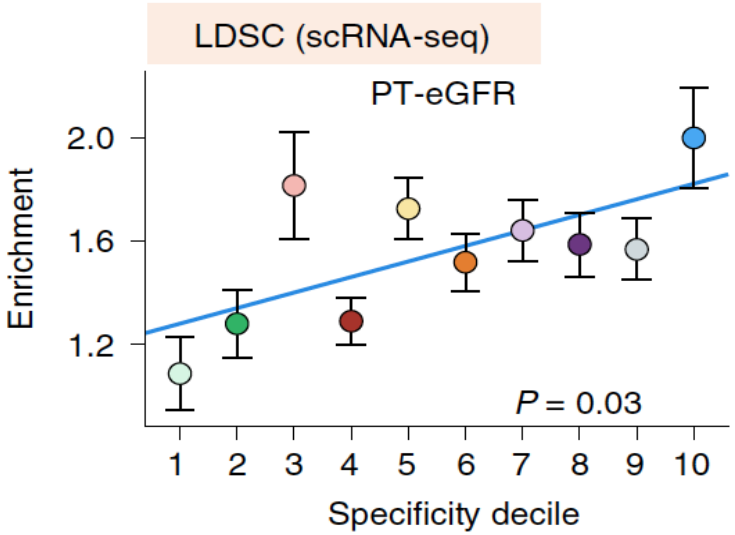


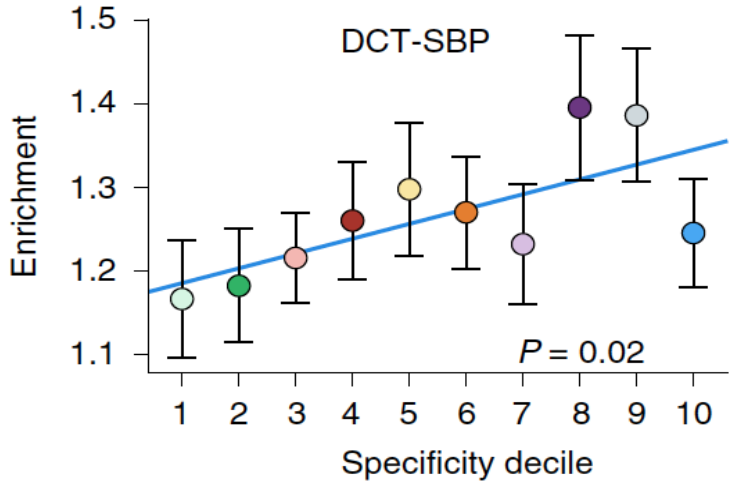


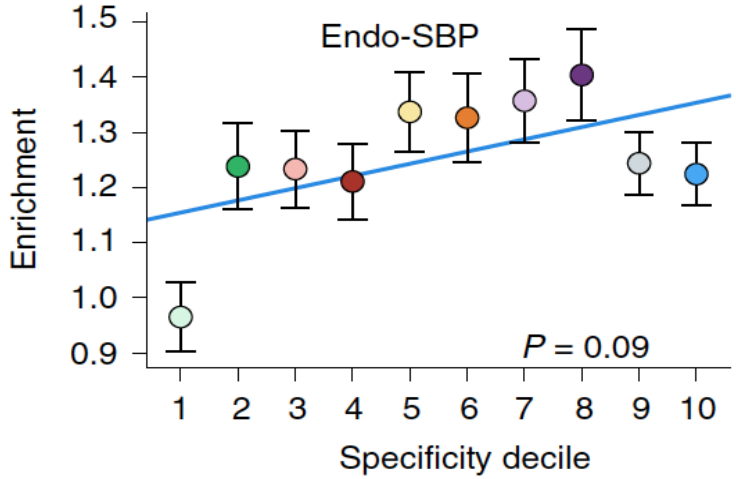




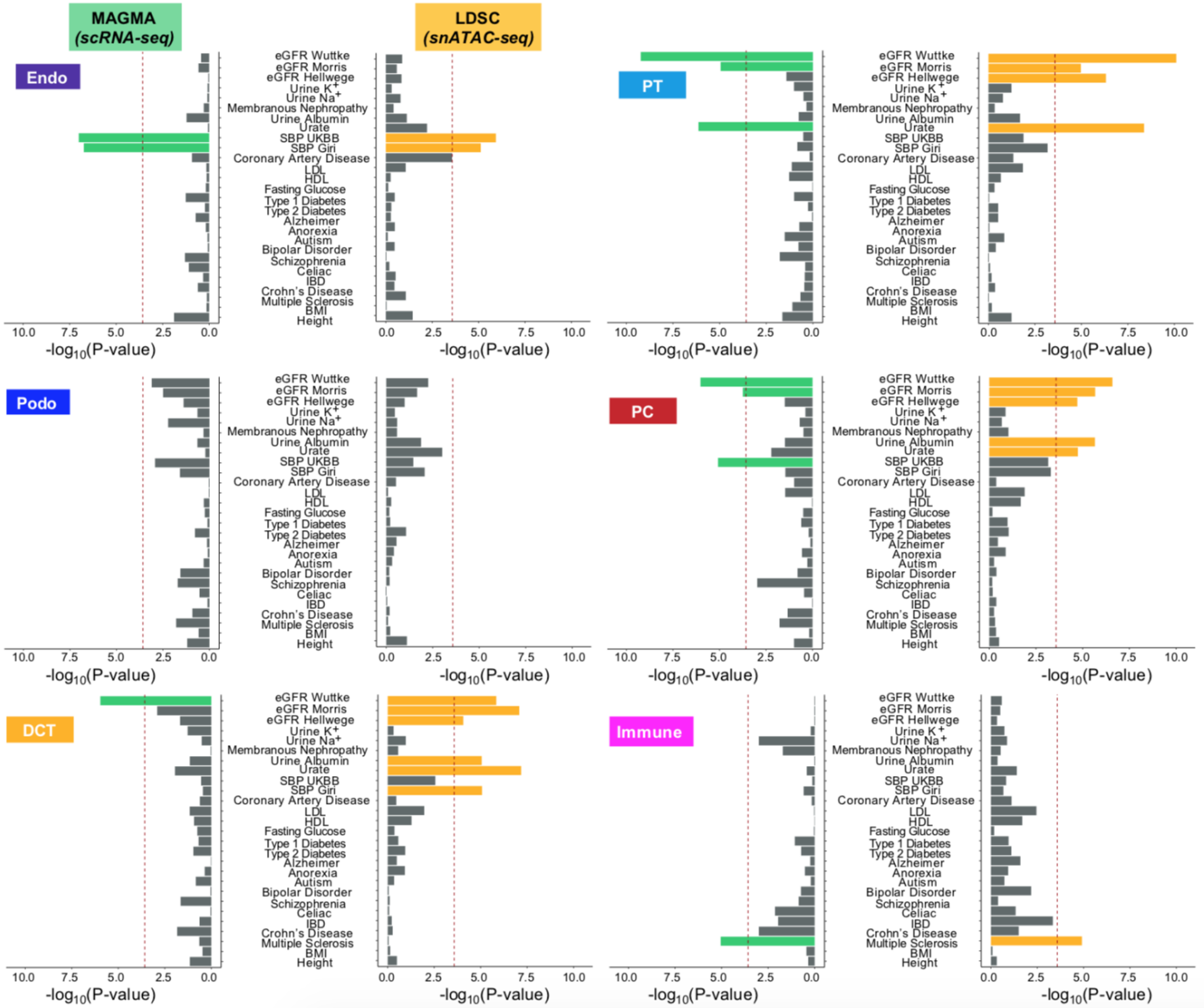






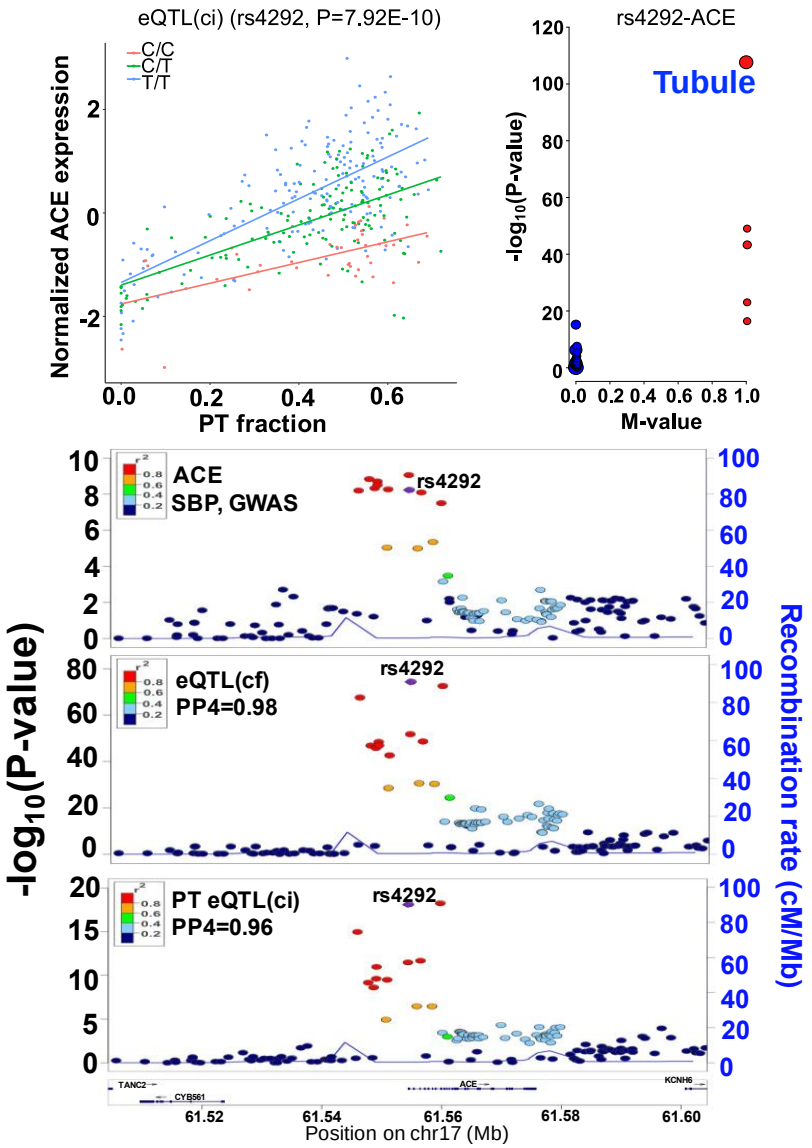


MAGMA和LDSC: 注释肾脏各类细胞所承担的生物学功能

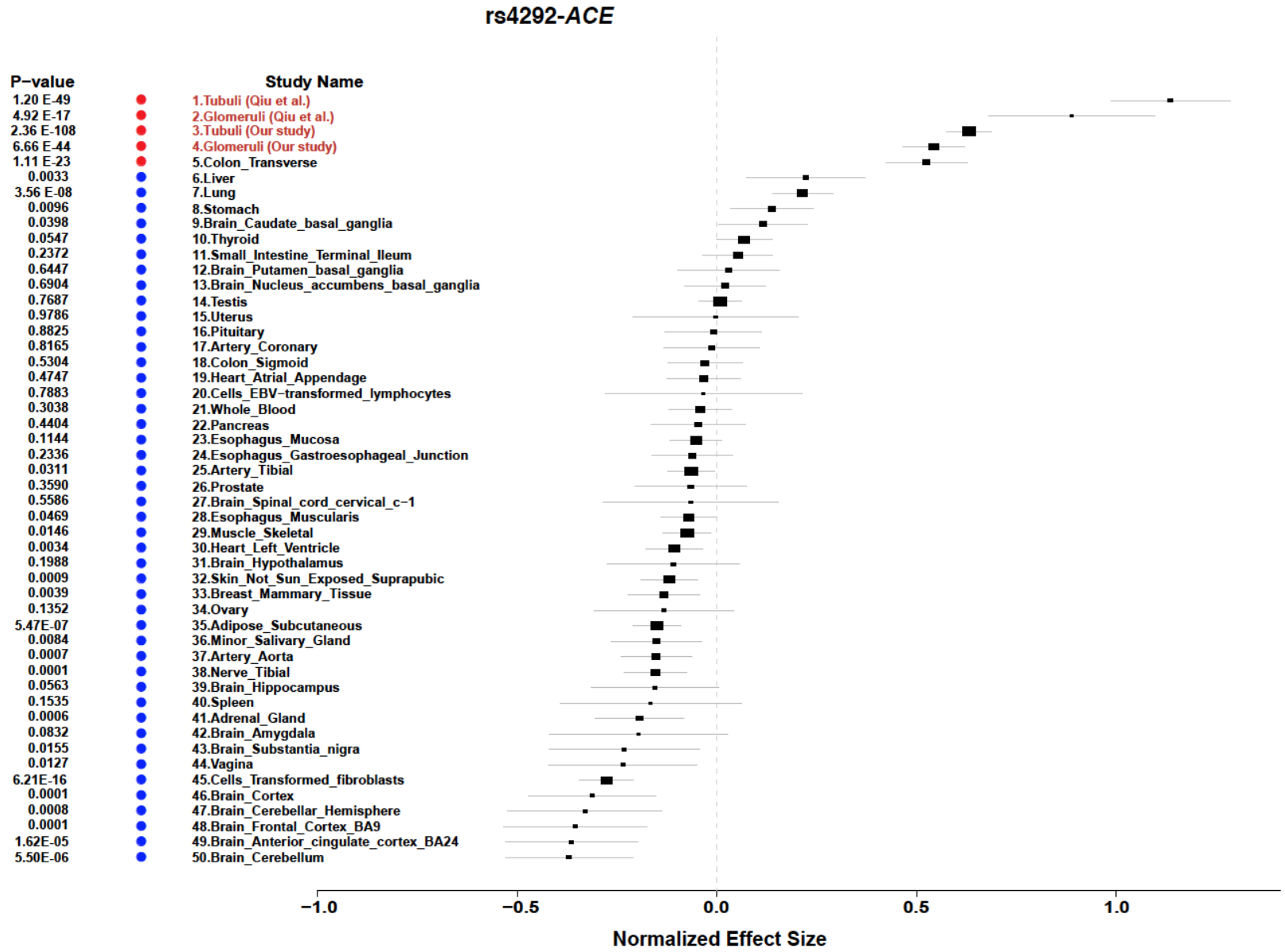


肾素-血管紧张素系统（ACEi）

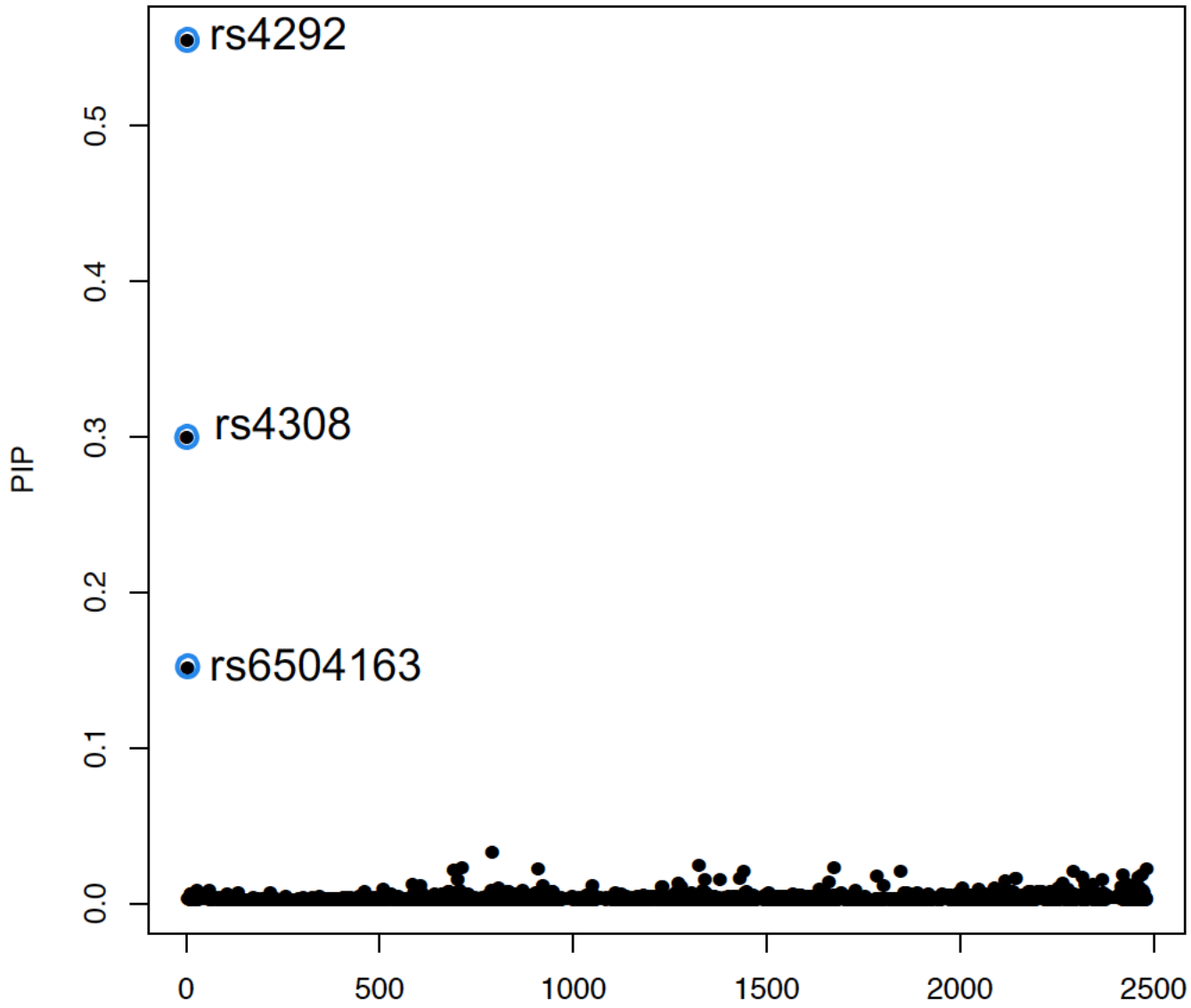
rs4292-ACE



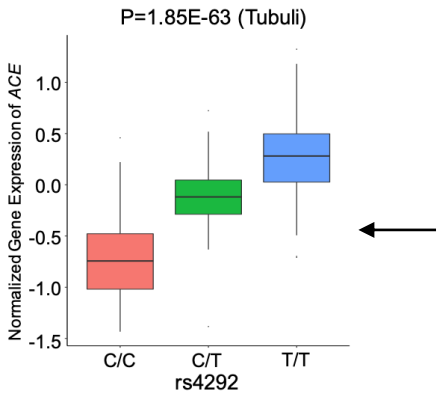
肾脏特异的调控作用：rs4292-*ACE*

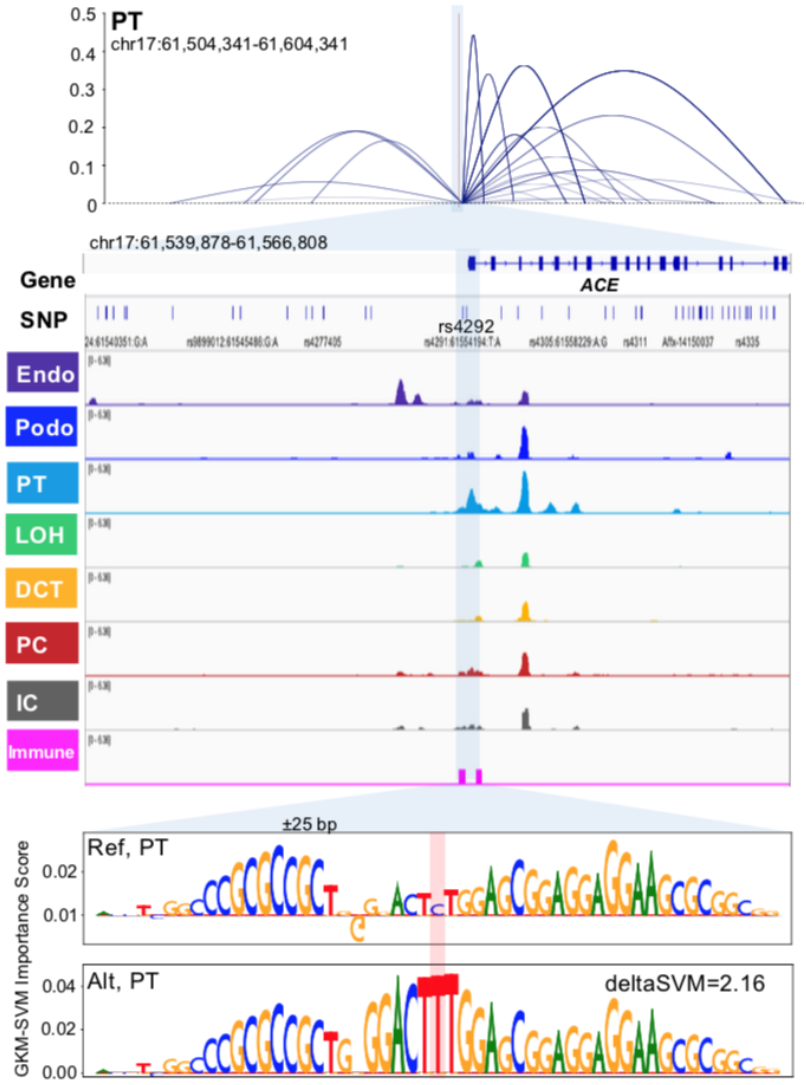


Fine mapping: 定位致病SNP



肾素-血管紧张素系统（ACEi）





内容大纲

- 基本概念
- 慢性肾脏病及其并发症
- 多组学数据整合分析方法
- 肾脏具有高度细胞异质性
- 从单细胞水平解析SNP-基因-CKD机制
- 总结

总结

- 慢性肾脏病（CKD）与高血压
- 肾脏具有高度细胞异质性
- 鉴定细胞“互作”的eQTL调控的新算法
- 组织的细胞异质性对复杂疾病的潜在致病基因鉴定十分重要