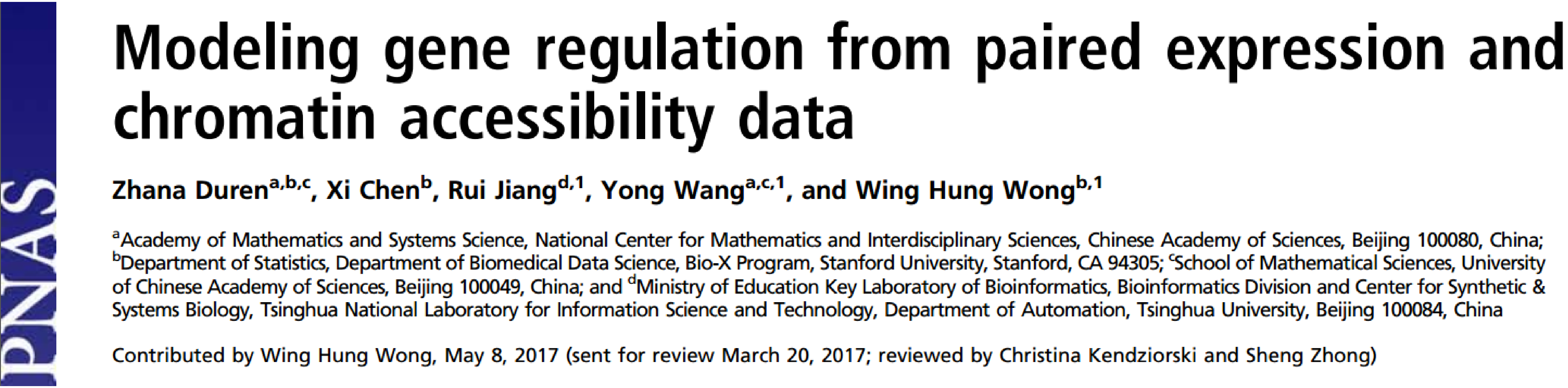


Modeling gene regulation from paired expression and chromatin accessibility data

Zhana Duren[a,b,c], Xi Chen[b], Rui Jiang[d,1], Yong Wang[a,c,1], and Wing Hung Wong[b,1]

[a]Academy of Mathematics and Systems Science, National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, Beijing 100080, China; [b]Department of Statistics, Department of Biomedical Data Science, Bio-X Program, Stanford University, Stanford, CA 94305; [c]School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China; and [d]Ministry of Education Key Laboratory of Bioinformatics, Bioinformatics Division and Center for Synthetic & Systems Biology, Tsinghua National Laboratory for Information Science and Technology, Department of Automation, Tsinghua University, Beijing 100084, China

Contributed by Wing Hung Wong, May 8, 2017 (sent for review March 20, 2017; reviewed by Christina Kendziorski and Sheng Zhong)

文献阅读

姚红琳
2017.12.1

1.背景

- 关键词

TF：transcription factor

REs：cis-regulatory elements

CRs：chromatin regulators

TGs：target genes

PECA：paired expression and chromatin accessibility

- 数据来源

训练集数据(expression and accessibility data)来自mouse ENCODE project

Protein-protein 互作数据来自 BIOGRID database

TFs motif 数据来自JASPAR, TRANSFAC, UniPROBE, and Taipale

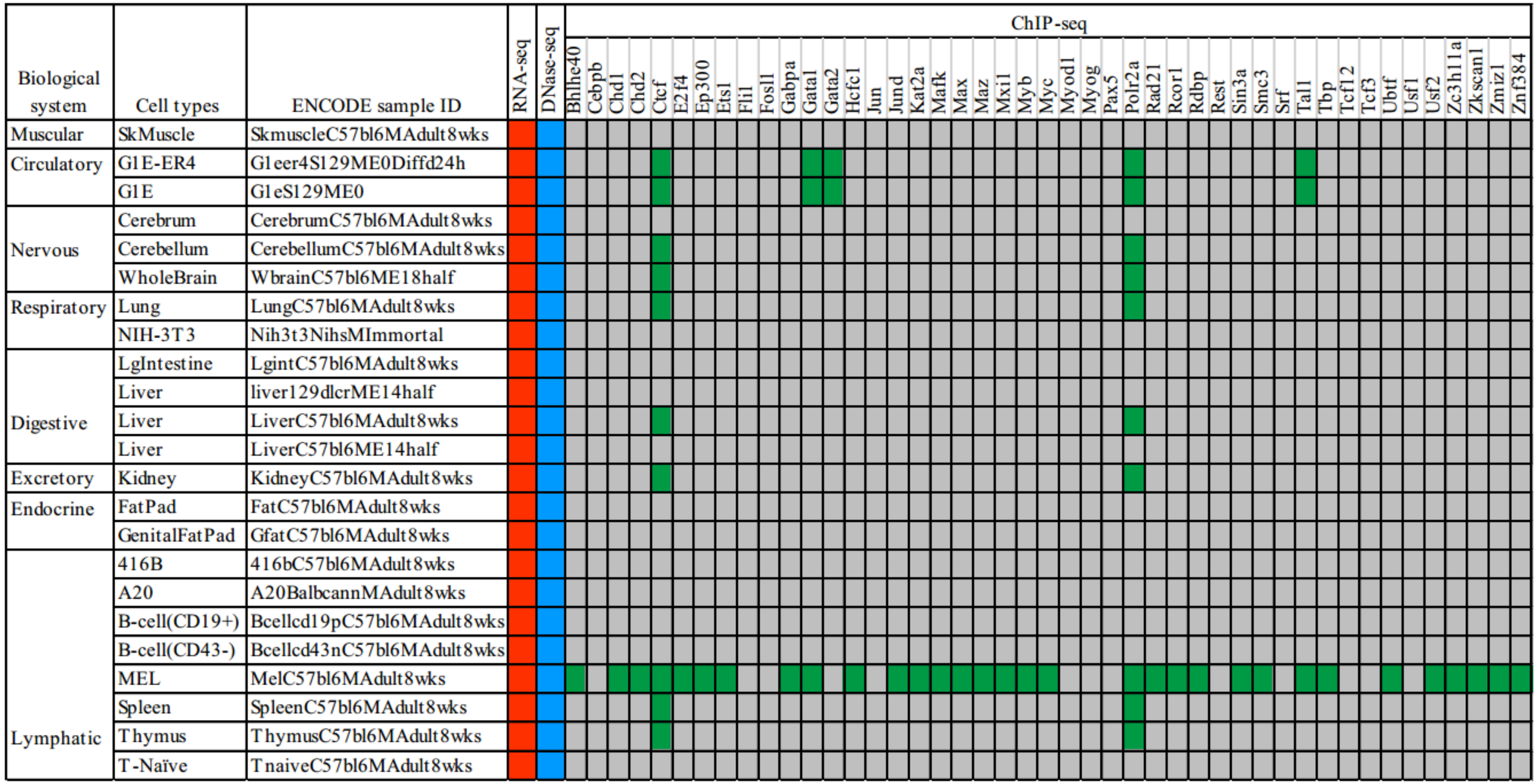
CRs 来自GO注释

1.背景

- ➢ RNA-seq：对于了解转录机制只能提供少部分信息（无转录因子的结合，染色质修饰等信息）

- ➢ ChIP-seq：检测到特定转录因子结合位点，一些遗传学标记，但是one by one

- ➢ Dnase-seq or ATAC-seq：检测到染色质的开放状态

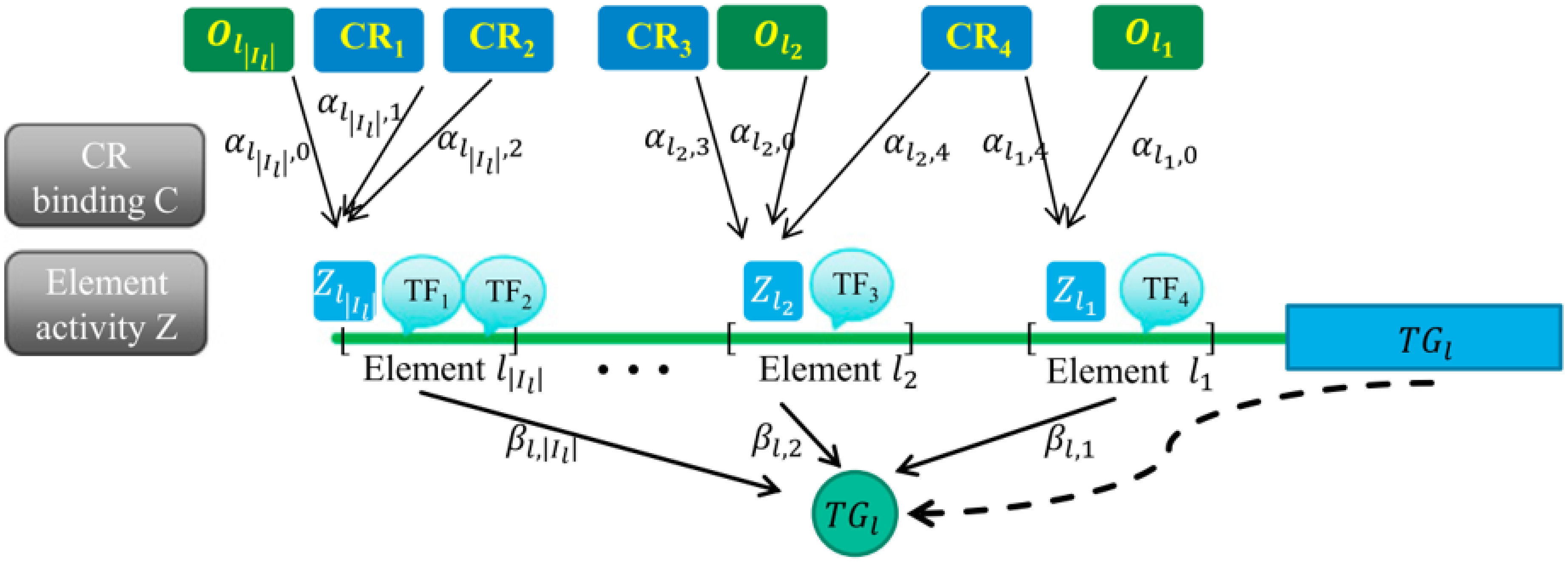


Biological system	Cell types	ENCODE sample ID	RNA-seq	DNase-seq	ChIP-seq																																													
					Bhlhe40	Cebpb	Chd1	Chd2	Ctcf	E2f4	Ep300	Ets1	Fli1	Fosl1	Gabpa	Gata1	Gata2	Hcfc1	Jun	Jund	Kat2a	Mafk	Max	Maz	Mxi1	Myb	Myc	Myod1	Myog	Pax5	Polr2a	Rad21	Rcor1	Rdbp	Rest	Sin3a	Smc3	Srf	Tal1	Tbp	Tcf12	Tcf3	Ubtf	Usf1	Usf2	Zc3h11a	Zkscan1	Zmiz1	Znf384
Muscular	SkMuscle	SkmuscleC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
Circulatory	G1E-ER4	G1eer4S129ME0Diffd24h	🟥	🟦	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	🟩	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	G1E	G1eS129ME0	🟥	🟦	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	🟩	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
Nervous	Cerebrum	CerebrumC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	Cerebellum	CerebellumC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	WholeBrain	WbrainC57bl6ME18half	🟥	🟦	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
Respiratory	Lung	LungC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	NIH-3T3	Nih3t3NihsMImmortal	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
Digestive	LgIntestine	LgintC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	Liver	liver129dlcrME14half	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	Liver	LiverC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	Liver	LiverC57bl6ME14half	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
Excretory	Kidney	KidneyC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
Endocrine	FatPad	FatC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	GenitalFatPad	GfatC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
Lymphatic	416B	416bC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	A20	A20BalbcannMAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	B-cell(CD19+)	Bcellcd19pC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	B-cell(CD43-)	Bcellcd43nC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	MEL	MelC57bl6MAdult8wks	🟥	🟦	🟩	⬜	🟩	🟩	🟩	🟩	🟩	🟩	⬜	⬜	🟩	🟩	⬜	🟩	⬜	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	⬜	⬜	⬜	🟩	🟩	🟩	🟩	⬜	🟩	🟩	⬜	🟩	🟩	⬜	🟩	⬜	🟩	🟩	🟩	🟩	🟩	🟩
	Spleen	SpleenC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	Thymus	ThymusC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟩	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜
	T-Naïve	TnaiveC57bl6MAdult8wks	🟥	🟦	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜

form mouse ENCODE project

2.方法

➢ PECA model



Input of PECA：the expression of TF genes, CR genes, and TGs

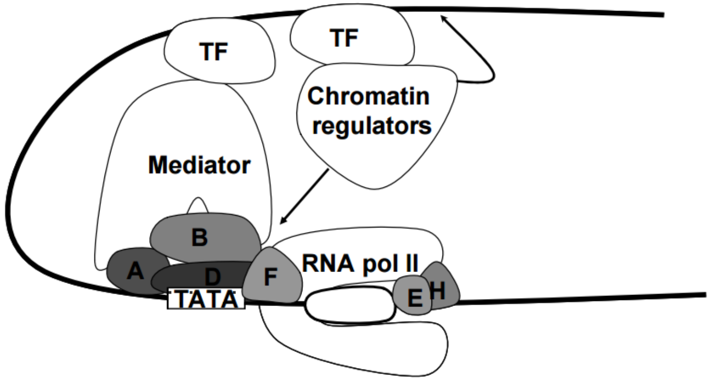
the openness of REs

the motif binding in the elements for TFs

Protein – protein interactions (PPI) among CRs and TFs

2.方法

- PECA model
  - Model of CR binding to REs



$$\log\frac{P(C_{i,j}=1|TF,O_i)}{1-P(C_{i,j}=1|TF,O_i)}=\eta_{l,0}+\eta_{l,1}\sum_{k\in S_{i,j}}(TF_kTFS_kB_{i,k}O_i)^{\frac{1}{4}}$$

$$P(C_{i,j}=1|TF,O_i)=\frac{\exp\left(\eta_{l,0}+\eta_{l,1}\sum_{k\in S_{i,j}}(TF_kTFS_kB_{i,k}O_i)^{\frac{1}{4}}\right)}{1+\exp\left(\eta_{l,0}+\eta_{l,1}\sum_{k\in S_{i,j}}(TF_kTFS_kB_{i,k}O_i)^{\frac{1}{4}}\right)},$$

$C_{i,j}$ ：recruitment status of the jth CRs on the ith RE

TF_k：TF expression

TFS_k：TF specificity expression score

B_i：TF motif-binding strength on RE

$O_{i,k}$：openness of RE

2.方法

- PECA model

  - Model of RE activity

$$\log\left(\frac{P(Z_i=1|O_i,CR,C_i)}{1-P(Z_i=1|O_i,CR,C_i)}\right)=\alpha_{i,-1}+\alpha_{i,0}O_i+\sum_{j=1}^{J}\alpha_{ij}C_{i,j}CR_j$$

$$P(Z_i=1|O_i,CR,C_i)=\frac{\exp\left(\alpha_{i,-1}+\alpha_{i,0}O_i+\sum_{j=1}^{J}\alpha_{ij}C_{i,j}CR_j\right)}{1+\exp\left(\alpha_{i,-1}+\alpha_{i,0}O_i+\sum_{j=1}^{J}\alpha_{ij}C_{i,j}CR_j\right)},$$

$C_{i,j}$ ：recruitment status of the jth CRs on the ith RE

CR_j ：the expressions of binding CRs

O_i ：openness of RE

Z_i： activation status of the ith RE

2.方法

- ➢ PECA model

  - ➢ Model of TG expression

$$TG_l|TF, Z \sim N\left(\beta_{l,0} + \sum_{i\in I_l} \beta_{l,i} Z_i \left(\sum_{k\in MB_i} \gamma_{l,k} B_{i,k} TF_k \right), \sigma_l^2 \right); l \in \{1,2, \ldots L\}.$$

TG_l :TG expression

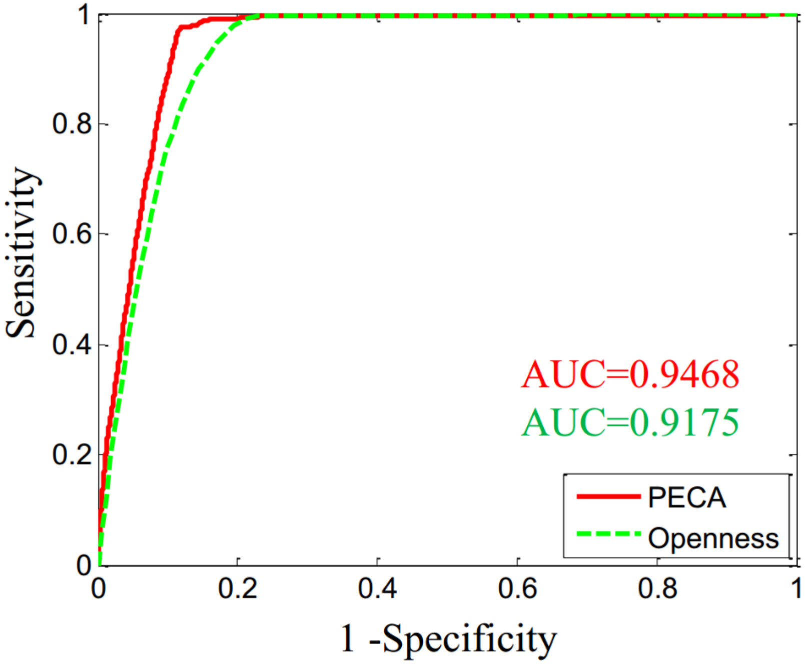
Z_i : activation status of the ith RE

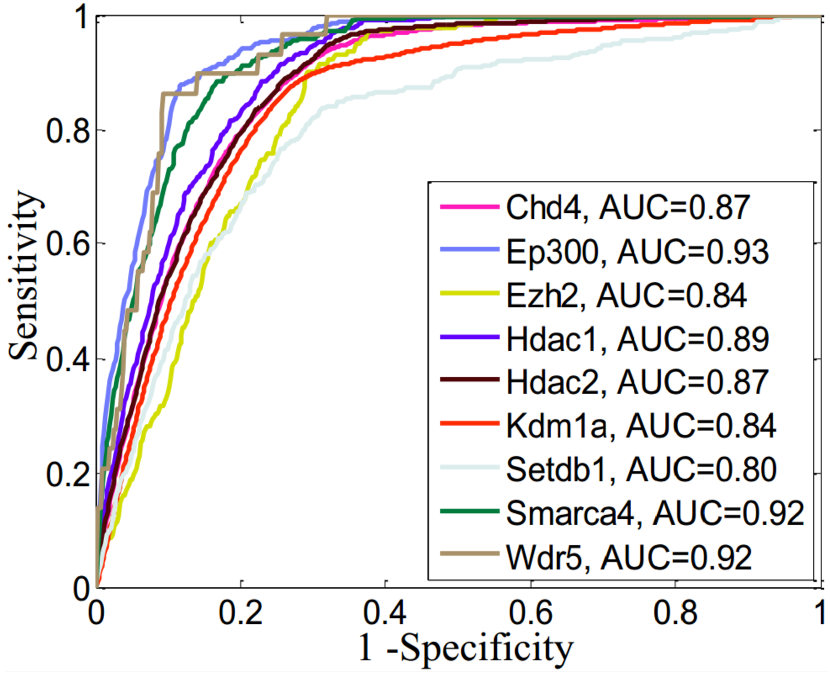
B_i : TF motif-binding strength on RE

TF_k : TF expression

3.结果

- Inference of the Recruitment Status of Chromatin Regulators





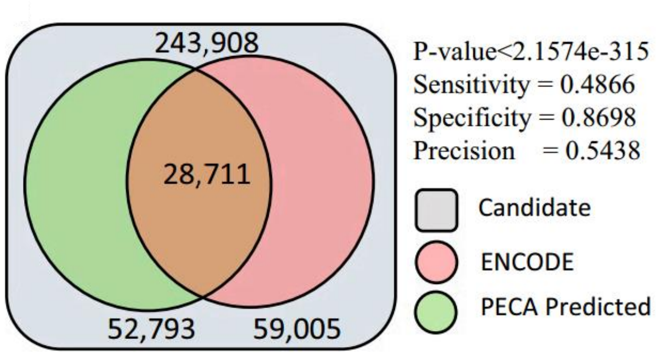
1. 基于PECA预测与基于RE可及性预测的ROC曲线比较

2.基于PECA预测不同CR招募水平的RC曲线

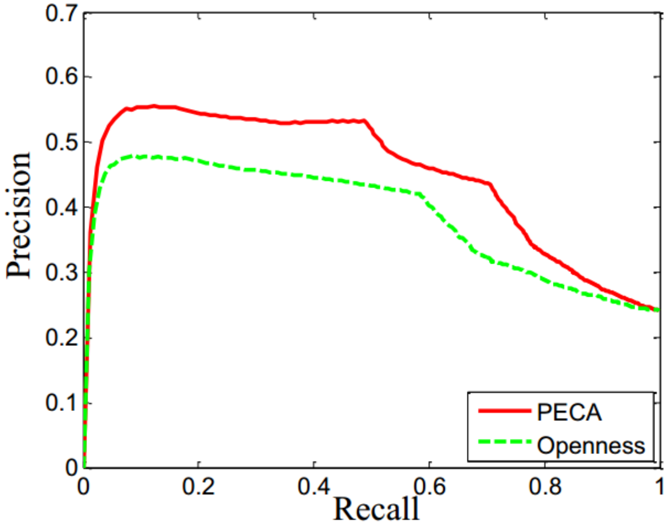
- PECA预测CR招募水平的结果较好，好于只用openness数据预测

3.结果

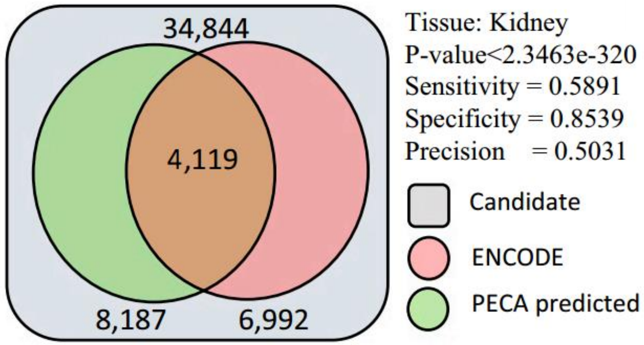
- ➢ Prediction of the Activation Status of Regulatory Elements



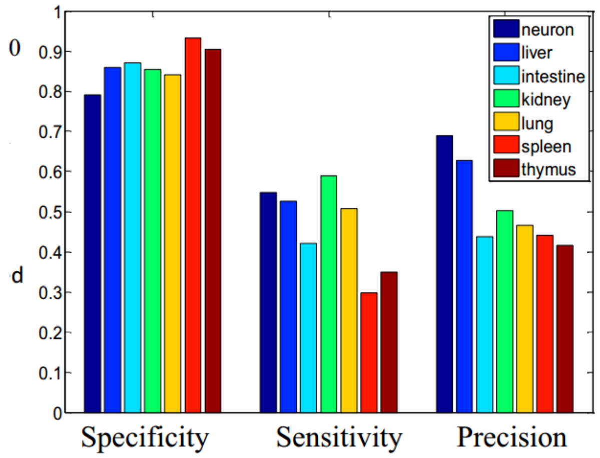
1.(7个组织)PECA预测RE活性与ENCODE注释比较



2.PECA预测RE活性与基于开放性预测比较



3.(肾)PECA预测RE活性与基于开放性预测比较

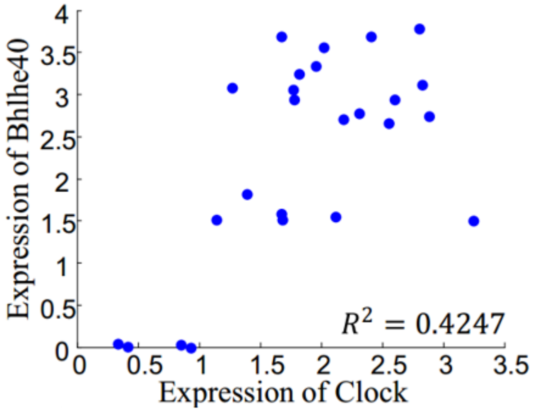


4.各组织PECA预测RE活性与ENCODE注释比较

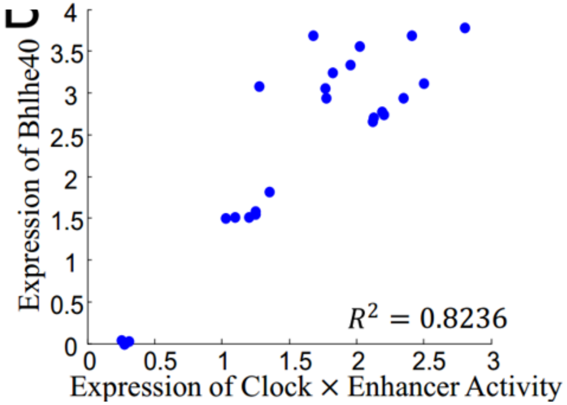
- ➢ PECA能够较准确的预测RE活性

3.结果

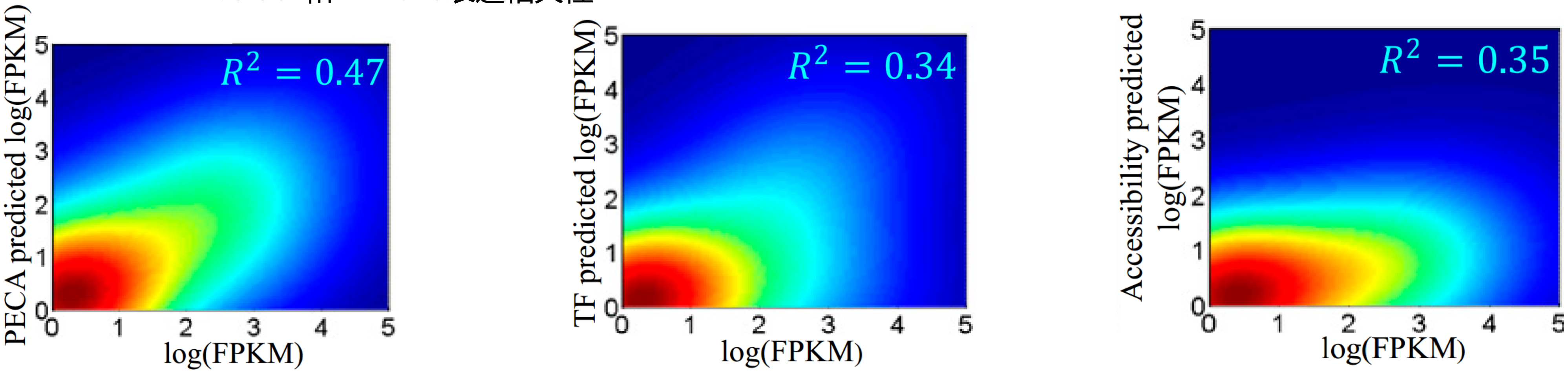
- Prediction of Gene Expression



1.Clock和Bhlhe40表达相关性



2. (Clock表达与enhancer活性)和Bhlhe40表达相关性

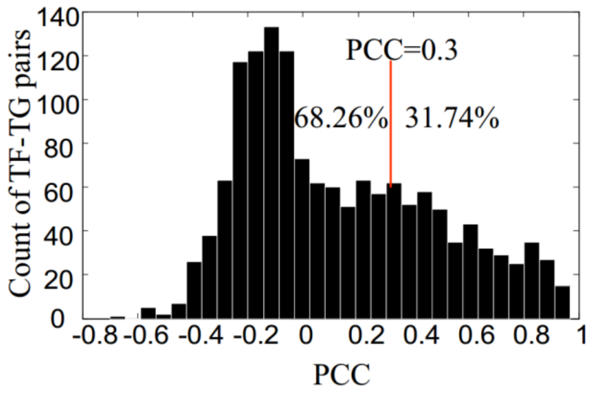


3.新的细胞背景（RA处理）基于PECA预测、基于TF表达量预测、基于RE可及性预测与TG表达量之间的相关性比较

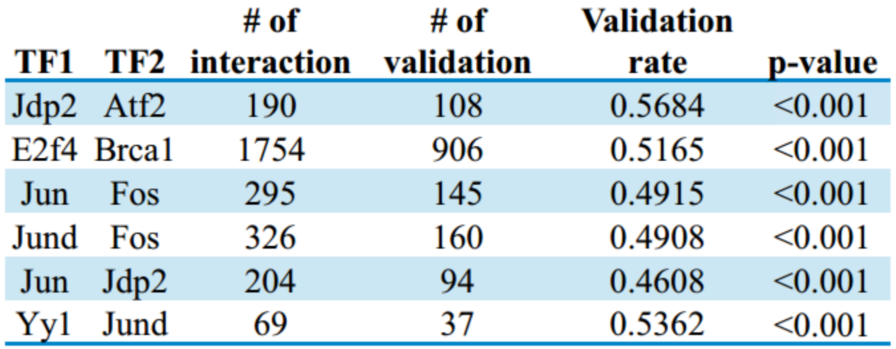
- 结合TF表达和RE活性显示出与TG表达量之间更高的相关性

3.结果

- ➢ Extraction of Regulatory Relations

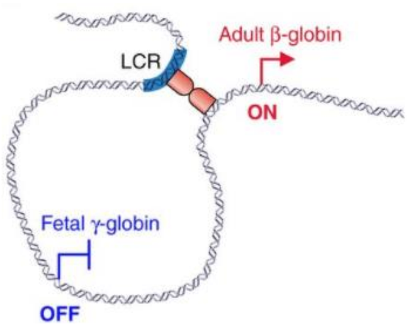


1.PECA 中TF-TG pairs PCC分布

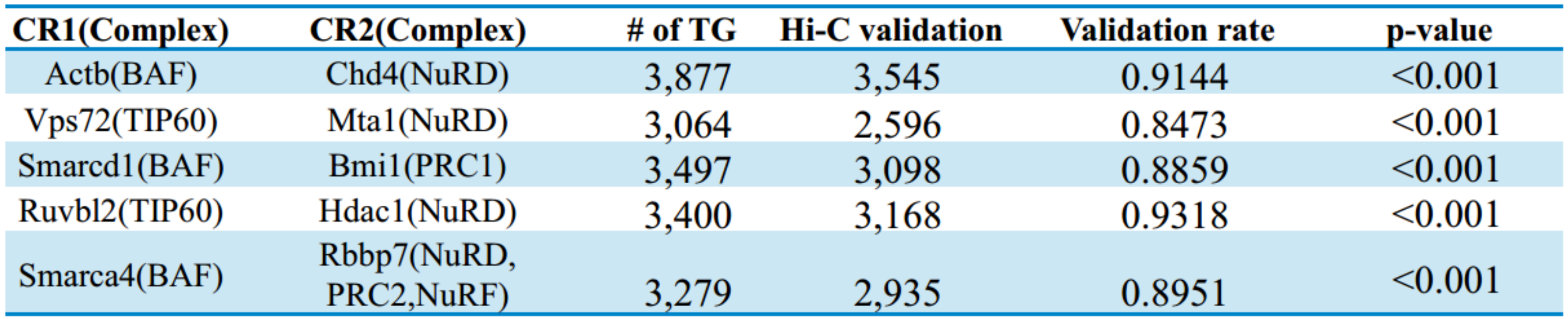


TF1	TF2	# of interaction	# of validation	Validation rate	p-value
Jdp2	Atf2	190	108	0.5684	<0.001
E2f4	Brca1	1754	906	0.5165	<0.001
Jun	Fos	295	145	0.4915	<0.001
Jund	Fos	326	160	0.4908	<0.001
Jun	Jdp2	204	94	0.4608	<0.001
Yy1	Jund	69	37	0.5362	<0.001

2.Hi-C验证TF-TF pairs

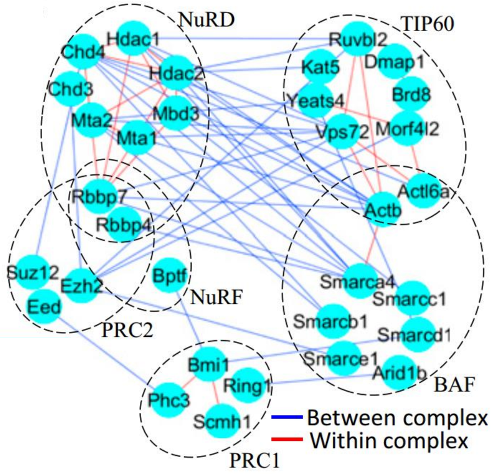


chromatin looping



CR1(Complex)	CR2(Complex)	# of TG	Hi-C validation	Validation rate	p-value
Actb(BAF)	Chd4(NuRD)	3,877	3,545	0.9144	<0.001
Vps72(TIP60)	Mta1(NuRD)	3,064	2,596	0.8473	<0.001
Smarcd1(BAF)	Bmi1(PRC1)	3,497	3,098	0.8859	<0.001
Ruvbl2(TIP60)	Hdac1(NuRD)	3,400	3,168	0.9318	<0.001
Smarca4(BAF)	Rbbp7(NuRD, PRC2,NuRF)	3,279	2,935	0.8951	<0.001

4. Hi-C验证CR–CR pairs

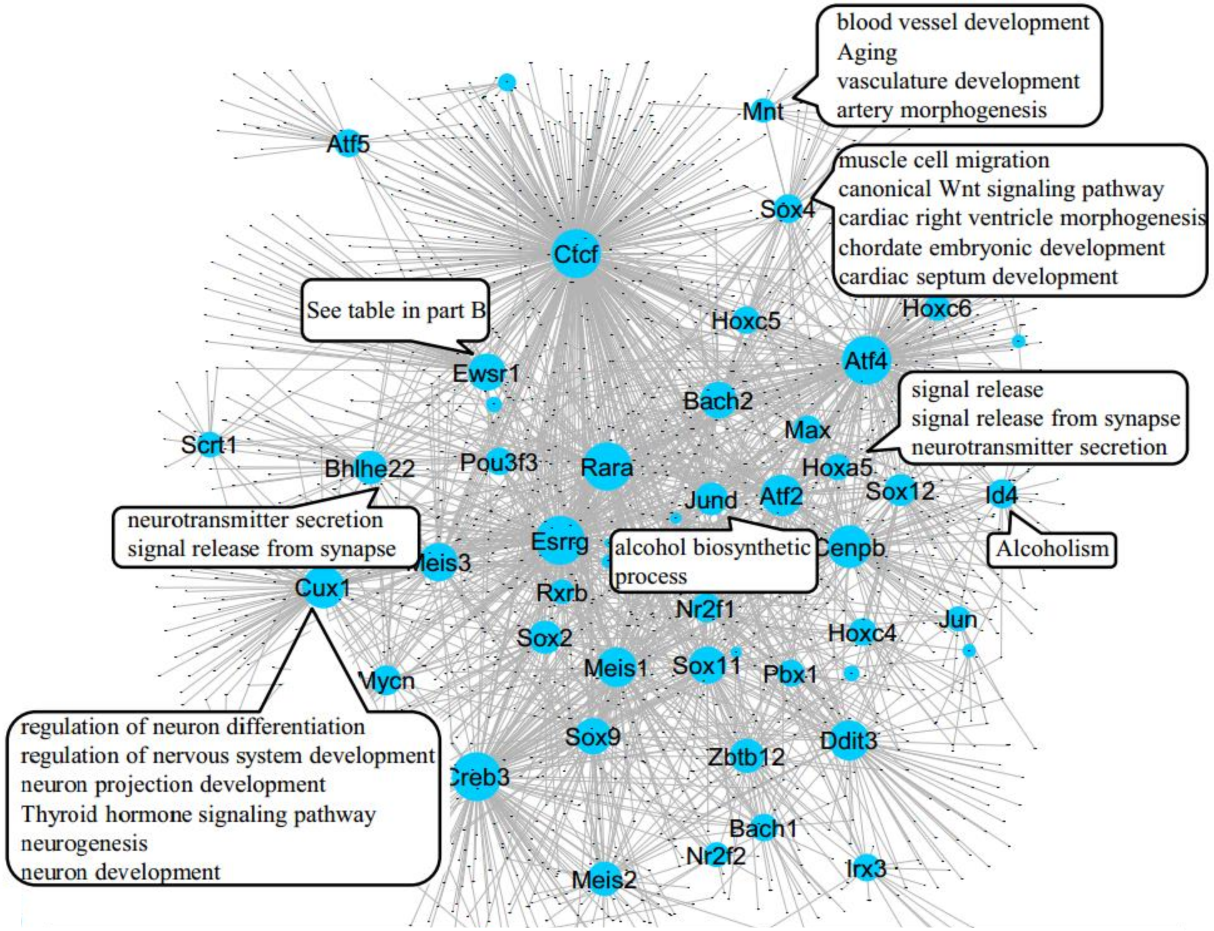


3. Cooperating CR–CR pairs

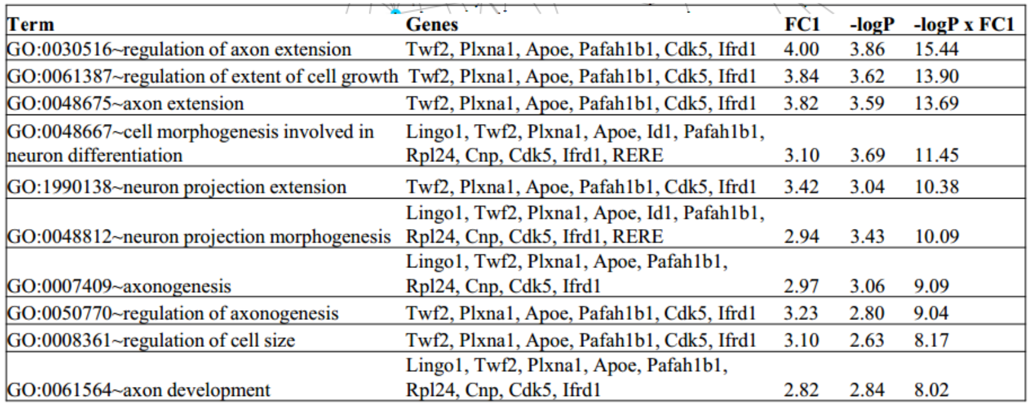
- ➢ PECA能够检测到低PCC的TF-TG pairs
- ➢ 经Hi-C实验验证，most TF-TF pairs和CR-CR pairs都是通过chromatin looping互作
- ➢ Most CR-CR pairs来自不同的CR complex

3.结果

- ➢ Inference of Context-Specific Regulatory Network


Atf5
Mnt
blood vessel development
Aging
vasculature development
artery morphogenesis
Sox4
muscle cell migration
canonical Wnt signaling pathway
cardiac right ventricle morphogenesis
chordate embryonic development
cardiac septum development
Ctcf
See table in part B
Hoxc5
Hoxc6
Ewsr1
Atf4
Bach2
signal release
signal release from synapse
neurotransmitter secretion
Max
Scrt1
Bhlhe22
Pou3f3
Rara
Hoxa5
Jund
Atf2
Sox12
Id4
neurotransmitter secretion
signal release from synapse
Meis3
Esrrg
alcohol biosynthetic process
Cenpb
Alcoholism
Cux1
Rxrb
Nr2f1
Sox2
Hoxc4
Jun
Mycn
Meis1
Sox11
Pbx1
regulation of neuron differentiation
regulation of nervous system development
neuron projection development
Thyroid hormone signaling pathway
neurogenesis
neuron development
Sox9
Ddit3
Creb3
Zbtb12
Bach1
Nr2f2
Irx3
Meis2

1.结果中部分背景特异的TFs；对应TGs的GO富集分析



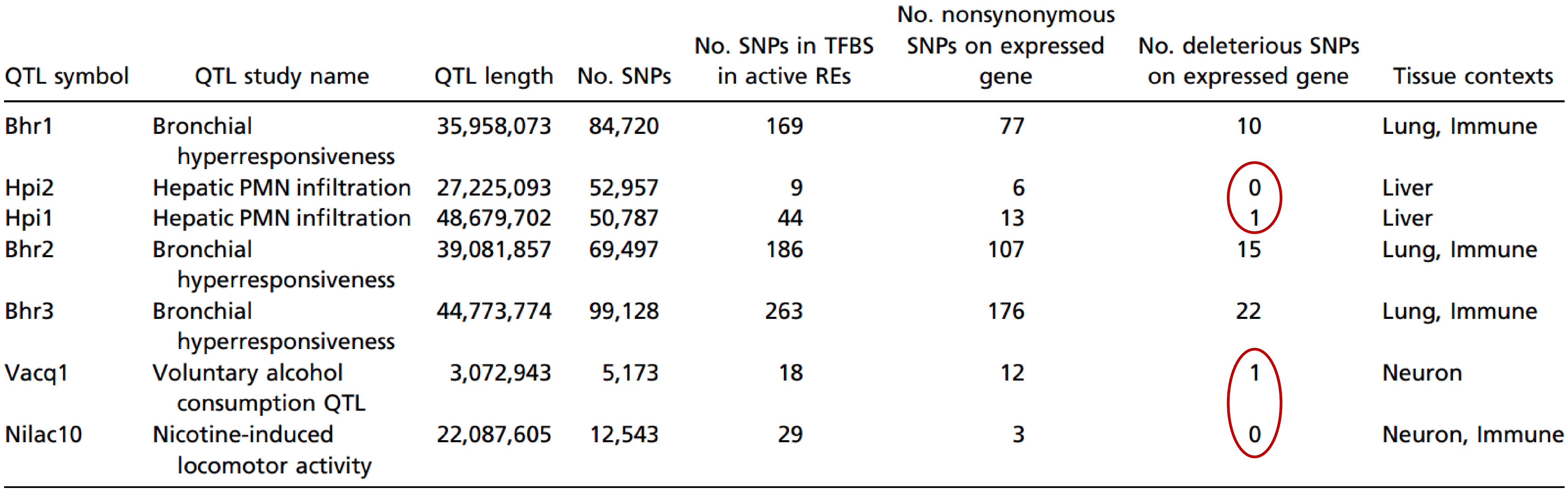
Term	Genes	FC1	-logP	-logP x FC1
GO:0030516~regulation of axon extension	Twf2, Plxna1, Apoe, Pafah1b1, Cdk5, Ifrd1	4.00	3.86	15.44
GO:0061387~regulation of extent of cell growth	Twf2, Plxna1, Apoe, Pafah1b1, Cdk5, Ifrd1	3.84	3.62	13.90
GO:0048675~axon extension	Twf2, Plxna1, Apoe, Pafah1b1, Cdk5, Ifrd1	3.82	3.59	13.69
GO:0048667~cell morphogenesis involved in neuron differentiation	Lingo1, Twf2, Plxna1, Apoe, Id1, Pafah1b1, Rpl24, Cnp, Cdk5, Ifrd1, RERE	3.10	3.69	11.45
GO:1990138~neuron projection extension	Twf2, Plxna1, Apoe, Pafah1b1, Cdk5, Ifrd1	3.42	3.04	10.38
GO:0048812~neuron projection morphogenesis	Lingo1, Twf2, Plxna1, Apoe, Id1, Pafah1b1, Rpl24, Cnp, Cdk5, Ifrd1, RERE	2.94	3.43	10.09
GO:0007409~axonogenesis	Lingo1, Twf2, Plxna1, Apoe, Pafah1b1, Rpl24, Cnp, Cdk5, Ifrd1	2.97	3.06	9.09
GO:0050770~regulation of axonogenesis	Twf2, Plxna1, Apoe, Pafah1b1, Cdk5, Ifrd1	3.23	2.80	9.04
GO:0008361~regulation of cell size	Twf2, Plxna1, Apoe, Pafah1b1, Cdk5, Ifrd1	3.10	2.63	8.17
GO:0061564~axon development	Lingo1, Twf2, Plxna1, Apoe, Pafah1b1, Rpl24, Cnp, Cdk5, Ifrd1	2.82	2.84	8.02

2. Ewsr1 的TGs 具体GO富集结果

- ➢ 用训练集构建模型，表达数据和可及性数据来自RA处理6d的mESC
- ➢ 选择活性REs，此背景下特异表达的TFs和TGs构建网络

3.结果

- ➢ Interpretation of Genetic Variants Relevant to Traits and Diseases



QTL symbol	QTL study name	QTL length	No. SNPs	No. SNPs in TFBS in active REs	No. nonsynonymous SNPs on expressed gene	No. deleterious SNPs on expressed gene	Tissue contexts
Bhr1	Bronchial hyperresponsiveness	35,958,073	84,720	169	77	10	Lung, Immune
Hpi2	Hepatic PMN infiltration	27,225,093	52,957	9	6	0	Liver
Hpi1	Hepatic PMN infiltration	48,679,702	50,787	44	13	1	Liver
Bhr2	Bronchial hyperresponsiveness	39,081,857	69,497	186	107	15	Lung, Immune
Bhr3	Bronchial hyperresponsiveness	44,773,774	99,128	263	176	22	Lung, Immune
Vacq1	Voluntary alcohol consumption QTL	3,072,943	5,173	18	12	1	Neuron
Nilac10	Nicotine-induced locomotor activity	22,087,605	12,543	29	3	0	Neuron, Immune

1. 选取7个QTL区段(有明确的相关组织背景)的统计

- ➢ 99% 品种特异的SNP(定位到QTL区段)处于非编码区段
- ➢ Hpi1，Hpi2，Vacq1，Nilac10表达基因中基本没有有害SNP，说明位于非编码区的变异发挥重要作用，体现了该模型的重要性

4.归纳总结

- 总结：利用基因表达量数据和染色质可及性数据构建基因调控模型。创新点在于该研究结合了多种信息构建模型。
- 启发： 生物技术的发展(如ATAC-seq、Hi-C等）会使生物数据的种类更加多样，而这些数据会帮助我们进一步解答生物学问题。
- 存在的问题：

模型过于复杂，各参数对TG表达的影响权重不易确定，该研究用了各参数的几何平均数，合理性未知。

THANKS