

Integrative Alignments of DNA Elements for Transcriptional Regulation in Swine Epigenome

Jinahua Cao, Ph.D

Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction (Huazhong Agricultural University), Ministry of Education, China

fitset all a fits

Outlines

Introduction

• Data presentation

• Results(data quality; main output; examples)

Conclusion



Histone modification

"An integrated encyclopedia of DNA elements in the human genome", Nature, 2012



ENCODE/ Roadmap



Function

- * Gene transcription regulation
- ✤ DNA replication
- * Genome translocation
- DNA repair



ChIP/RNA-Seq data of porcine cells/tissue



H3K4me1 ✓ ✓ ✓✓ ✓✓ H3K4me3 ✓ ✓✓ ✓✓ ✓✓ ✓✓ H3K4me3 ✓ ✓✓ ✓✓ ✓✓ ✓✓ H3K9me3 ✓ ✓✓ ✓✓ ✓✓ ✓✓ H3K9me3 ✓ ✓✓ ✓✓ ✓✓ ✓✓ H3K27ac ✓✓ ✓✓ ✓✓ ✓✓ ✓✓ H3K27me3 ✓✓ ✓✓ ✓✓ ✓✓ ✓✓ H3K36me3 ✓ ✓ ✓ ✓ ✓	ом	LDM	PK-15	PAM	3D4/21	
H3K4me3✓✓✓✓✓H3K9me3✓✓✓✓H3K27ac✓✓✓✓✓✓✓H3K27me3✓✓✓✓✓✓✓✓H3K36me3✓✓✓✓	/	~	~~	~	~	H3K4me1
H3K9me3 ✓ ✓ ✓ H3K27ac ✓ ✓✓ ✓✓ ✓✓ H3K27me3 ✓✓ ✓✓ ✓✓ ✓✓ H3K36me3 ✓ ✓✓ ✓✓ ✓✓	' ~	~~	~	~~	~	H3K4me3
H3K27ac ✓ ✓✓ ✓✓ ✓✓ H3K27me3 ✓✓ ✓✓ ✓✓ ✓✓ H3K36me3 ✓ ✓✓ ✓✓ ✓✓	/	~	~	~	~	H3K9me3
H3K27me3 VV VV V V H3K36me3 V V V V V	/	~	~~	~~	~	H3K27ac
H3K36me3 🗸 🖌	/	~	~~	~~	$\checkmark\checkmark$	H3K27me3
	/	~	~	~	~	H3K36me3
CTCF V V V	/	~	~	~	~	CTCF
POL2 VV V V V	' ~	~~	~	~	$\checkmark\checkmark$	POL2
INPUT 🗸 🔨 🔨 🗸	/	~	~	~	V	INPUT

- Over 15,000 peaks for narrow peak
- ~10,000
 enriched region
 for broad peak
- PAMs have better performance on H3K27me3.

Data quality assessments: RNA-Seq Reproducibility and ChIP-Seq Correlations



RNA-Seq data characterization on 3 types of cells

Condition

8

6

4

2

0

ΡK

RPL35

GAPDH

UBA52

ACTG1

ACTB

SF3B5

MRFAP1

CENPB

NACA2

LSM5

ENY2 PCBP1

YKT6

H1FX

GPX1

HMGN2

TMSB4X

RPLP2

TMSB10

RPLP0

APOE

TNF

ITGB2

CCL3L1

EID1

C1QB

C1QA

CTSL

TLR2

TYROBP

CD163

SRGN

DOK3

CD209

MAFB

DUOXA1

CXCR4

CCR1

FCN2

TMSB4Y

RPL5





Cell-type specific transcription (Top20)

- \geq 21,874 genes were investigated (Genome assembly: ssc11.1, annotation: Ensembl release 90)
- \geq Reads normalized using TPM
- 2 rep. per sample \geq
- #COM: 13,778 \geq
- #SP: PK<PAM<3D4 \triangleright





structural constituent of ribosome	18	9.24e-
structural molecule activity	20	7.15e-
RNA binding	15	2.62e-
transport	21	0.0000
structural constituent of cytoskeleton	2	0.012
mRNA binding	2	0.0189
double-stranded DNA binding	2	0.0403
hydrolase activity, hydrolyzing N-glycosyl compour	1	0.0461
structure-specific DNA binding	2	0.0943
oxidoreductase activity, acting on the aldehyde or ox	1	0.096

Gene networks and GO functional annotations

- Over 75% genes are common transcriptional.
- COM genes are responsible for cell structure, metabolism,...
- SP genes are more dependents on cell types





- Assigned 15 chromatin states based on 6 histone markers.
- Markers within a state have more correlation.
- Primary PAM cells have lower correlations between markers.







-0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 -1 -0.9

Overview of ChIP-Seq and RNA-Seq data

- MALAT1 and NEAT1 sites
- RNAPII and CTCF factor work well in swine genome
- 6 Histone markers are also work well.







Integrative Alignments of DNA Elements for Transcriptional Regulation in Swine Epigenome

- RNAPII (POL2), transcriptional factor
- CTCF, insulator for 3D genomics
- PAM with unique characters at H3K27ac/me3 profiling

		Colorkey				Colorkey				Colorkey		
				_								
-5	-2	0	4	8 -	-3 -2	. 0	4 1	9 -4	-2	0	5 9	∞ RNA.3D4
-2000	TSS	HNA.3D4	TES	2000 -2	000 TSS	HNA.PAM	TES 24	00 -2000	TSS	HNA.PK	TES 2000	RNA
-2000	- TSS	1 1 33% 60% H3K4me3.3D4	TES	2000 -2	2000 TSS	10C2.PAM 33% 60% H3K4me3.PAM	TES 2	000 -2000	TSS	1 1 33% 60% H3K4me3.PK	TES 2000	POL2
-2000	T _{TSS}	1 I 33% cm 66%	TES	2000 -4	2000 TSS	1 1	TES 2	1000 -200	o TSS	33% 60%	- TES 2000	-2000 TSS 33% 66% TES 2000 Genomic Region (5' -> 3)
		H3K27ac.3D4	TEP			H3K27ac.PAM			F00	H3K27ac.PK	TE9 999	H3K27ac
	133 Tee	H3K27me3.3D4	TES	2000 -2		H3K27me3.PAM				H3K27me3.PK		H3K27me3
		H3K4me1.3D4				H3K4me1.PAM				H3K4me1.PK		H3K4me1
-2000	TSS	1 1 33% 68% H3K9me3.3D4	TES	2000 -	2000 TSS	1 1 33% 66% H3K9me3.PAM	TES 2	2000 -200	o TSS	33% 68% H3K9me3.PK	TES 2000	H3K9me3
-2000	TSS	1 1 33% 66% H3K36me3.3D4	TES	2000 -	-2000 TSS	1 1 33% 86% H3K36me3.PA	TES	2000 -20	x TSS	1 1 33% 68% H3K36me3.P	TES 2000	H3K36mo3
-2000	TSS	1 1 33% 66% CTCF.3D4	TES	2000	-2000 TSS	s 33% 66% CTCF.PAM	6 TES	2000 -24	000 TSS	33% 66 CTCF.PK	5% TES 2000	กงกงขาแยง
2000	ISS	33% 66%	TES	2000	-2000 TSS		TES	2000 -20		33% 66	5% TES 2000	CTCF

				76,800 kb 76,900 kb
Colorkey	Colorkey	Colorkey	SOX17	CT0000006591 SDX17-291
-5 -2 0 3 5	-3 -2 0 3 7	-4 -2 0 3 5	DK_15	
RNA.3D4	RNA.PAM	RNA.PK	F IX-13	15 10 14 15 15 14 15 15 14 15 15 14 15 15 14 15 15 14 15 15 14 15 15 14 15 15 14 15 15 14 15 15 14 15 15 14 15 15 15 14 15 15 15 15 15 15 15 15 15 15 15 15 15
			RNA	15 14 14 13 14 15 14 15 14 14 13 13 13 14 14 14 13 10 15 2 0 2 0 14 13 13 13 14 13 13 13
-2000 TSS 33% 66% TES 2000 POL2.3D4	-2000 TSS 33% 66% TES 2000 - POL2.PAM	-2000 TSS 33% 66% TES 2001 POL2.PK)	i
H3K4me3.3D4	H3K4me3.PAM	H3K4me3.PK	POL2	
			H3K4me3	k sł
H3K27ac.3D4	H3K27ac.PAM	H3K27ac.PK	H3K27ac	
H3K27me3.3D4	H3K27me3.PAM H3K36me3.PAM	H3K27me3.PK	H3K27me3	- U.
H3K9me3.3D4	H3K9me3.PAM	H3K9me3.PK	H3K4me1	
			H3K9me3	
CTCE 904			H3K36me3	and and the second contract of the second
	UTO-PAM		CTCF	and a second data and a data data and a data data
-2000 TSS 33% 66% TES 2000	-2000 TSS 33% 66% TES 2000 -	-2000 TSS 33% 66% TES 20	T IOC	

المساهدة فريف فالتلا مانية مقدمة والمعني ويرون والمتعادية ومرجب المعادية والمتعاد المعادية والمتعاد المالي

_____ 391 kb _____

Constantly high expressed genes

- > Two categories:
 - Chromatin state free
 - Chromatin state related
- Extension discussion:
 - Conserved "transcription factory"
 - "loop" genes





GAPDH

ACTB



Primary Porcine Alveolar Macrophages Specific expressed genes

- > APOE: CTCF binding motif
- C1QA/C1QB: long-range interactions or enhancers



APOE	C1QA/C1QB
63 kb 63 kb	
╫→╞═╸┉ <mark>╂┼┼┼→┼→──╫</mark> ╞╸┝╴┽┼╊╸	i+ja s <mark>i→jaa</mark> s→tjaa
TOMM40-201 APDE-201 →→→→→↓↓ TOMM40-202	C10A-201 C10C-201 C10B-201
A las sumake	ii
2 6 3 4 9 6 9	13 14 9 13 14 13 13 9 14 13
	· · · · · · · · · · · · · · · · · · ·
-	
	and all the second second
· · · · · · · · · · · · · · · · · · ·	
	h







3D4/21 cell line specific highly expressed

CTCF binding motif CTCF TAD POL2 transcription factory formed

SCG2 gene transcription

Summary and perspectives

• 8 epigenomic markers on 3 cell types were surveyed chromatin states to explore DNA elements modifications in swine epigenomes.

 Integrative analysis was used to elucidate the relationship between gene transcription and proximal modifications by using RNA-Seq and ChIP-Seq data.

• Purposefully using RNAPII and CTCF factor in this study to peep the **3D** chromatin conformation organizations and dynamics in swine genome.







Dr. Shuhong's Group in **HZAU** Jianhua Cao Xingyun Li Mengjin Zhu Changchun Li Xiaolei Liu Shen Wang Remin Ren Ying Huang Yan Sun Liangliang Fu Changzhi Zhao





Thanks!