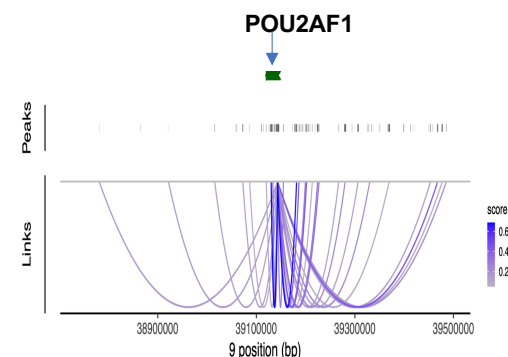# Final report on USDA-NIFA Pig FAANG Resource Project: Functional Annotation of the Porcine Genome

Christopher Tuggle
Iowa State University

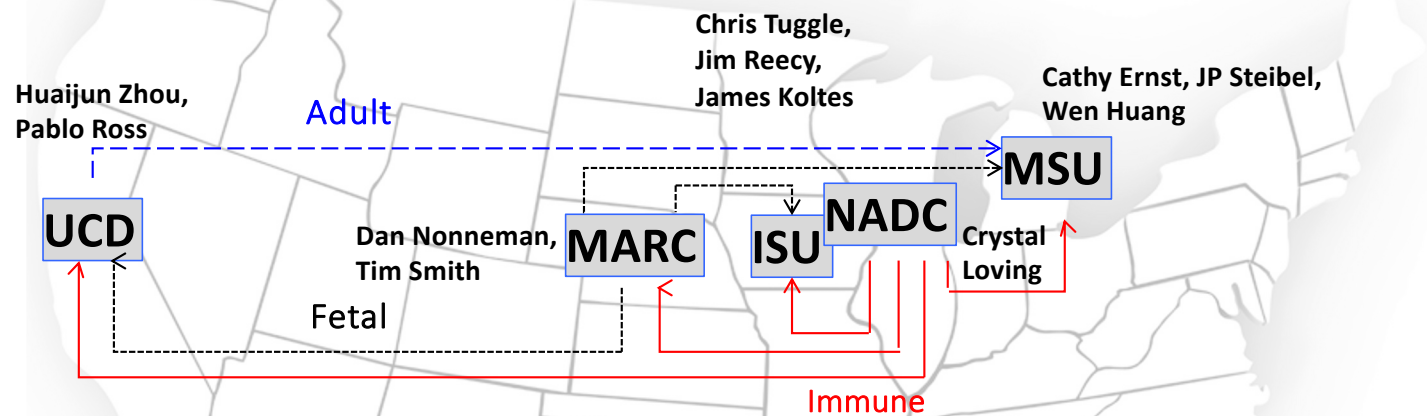FAANG Workshop 2024

NIFA Project 2018-67015-2701

Pig FAANG: Community Sharing of Samples and Expertise

Huaijun Zhou, Pablo Ross

Chris Tuggle, Jim Reecy, James Koltes

Cathy Ernst, JP Steibel, Wen Huang

Adult

UCD

Dan Nonneman, Tim Smith

MARC

ISU

NADC

MSU

Crystal Loving

Fetal

Immune

Location	Tissue collection	Sample Analyses
ISU	None	Histone ChIP, ATAC-Seq
MARC	Fetal*	RNAseq, IsoSeq
MSU	None#	DNA Methylation (WG-BS)
NADC	Immune	None
UCD	Adult#	Histone ChIP, ATAC-Seq, RAMPAGE
*Fetal Tissue already collected		
# Some adult tissues to be collected at MSU through 2016 AFRI project to UCD		

FAANG
Functional Annotation of Animal Genomes

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Functional Annotation of the Porcine Genome

Overall goal: *catalog functional elements in the porcine genome for >40 biological states, create Chromatin state map*

Aim 1 Adult tissues

- extend FAANG pilot of 8 -> 25 tissues

Aim 2 Fetal tissues

- Identify and correlate allele-specific expression and AS chromatin modification
- Parent-of-origin effects (reciprocal crosses Meishan x WhiteCross)

Aim 3 Immune System

- Stimulated macrophages
- Single cell analysis of blood PMBC and immune tissues
- Epigenetics of circulating Blood Cells- flow-sorted into functional types
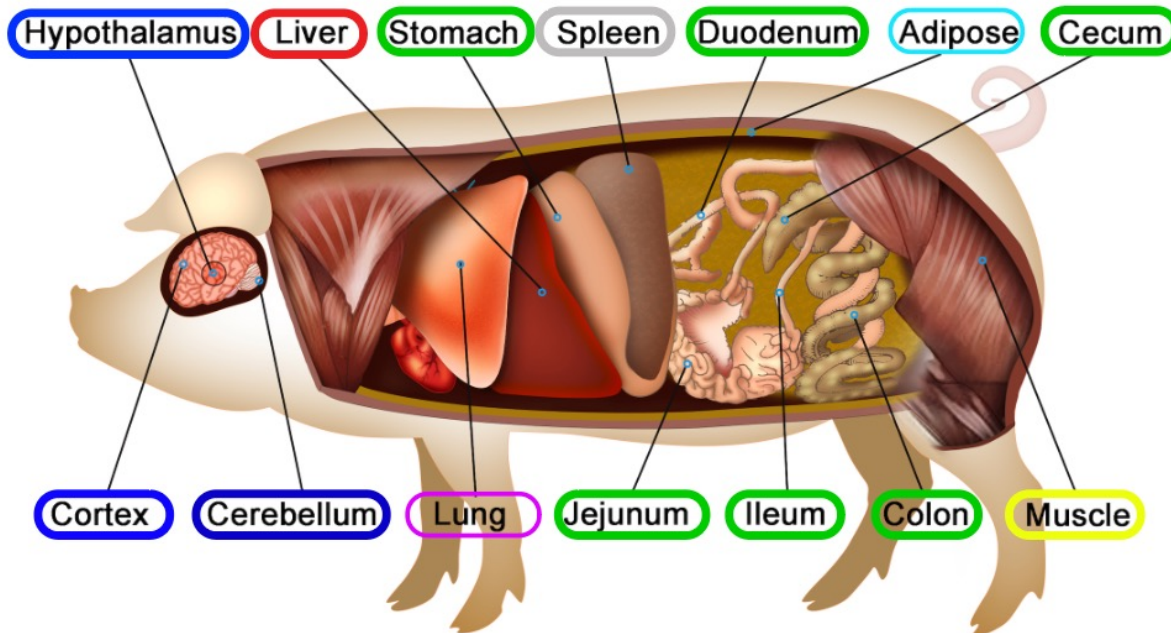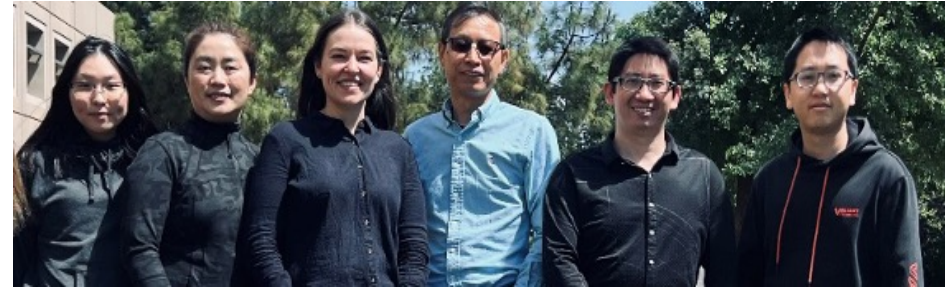
Aim 4 Data Integration

- Integrate all public and project data to develop a higher-order regulatory understanding of the porcine genome, including a predicted chromatin state map.

Aim 1: Adult tissues *(UC-Davis)*
- extend FAANG pilot of 8 -> 25 tissues

Liqi An, Ying WANG, Claire PROWSE-WILKINS, Huaijun ZHOU, Zhangyuan PAN, Dailu GUAN
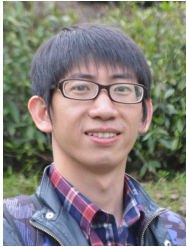
Finished

Core tissues (8): cortex, cerebellum, hypothalamus, liver, lung, spleen, and adipose muscle (Kern, et. al., Nat Commun 2021).

Gut-associated tissues (6): stomach, duodenum, jejunum, ileum, cecum, colon (Pan Z. et al. Nat Commun 2021).
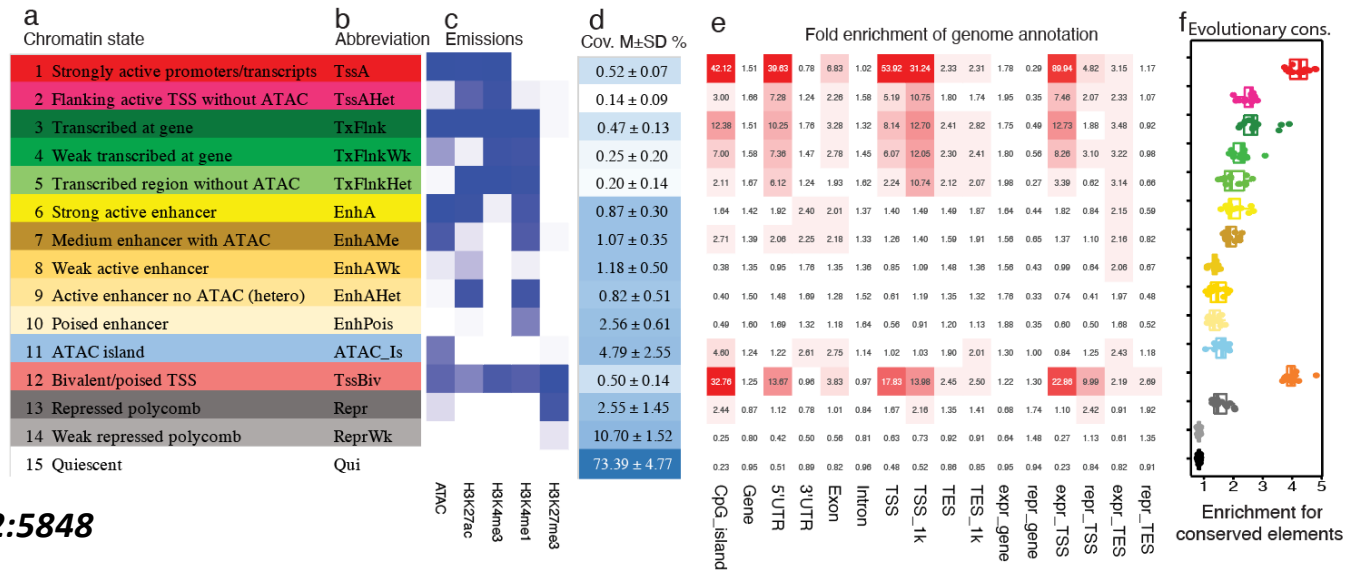
Hypothalamus · Liver · Stomach · Spleen · Duodenum · Adipose · Cecum

Cortex · Cerebellum · Lung · Jejunum · Ileum · Colon · Muscle

Functional Annotation of the Porcine Genome- Adult Tissues

Zhangyuan Pan/H. Zhou et al.,
UC Davis
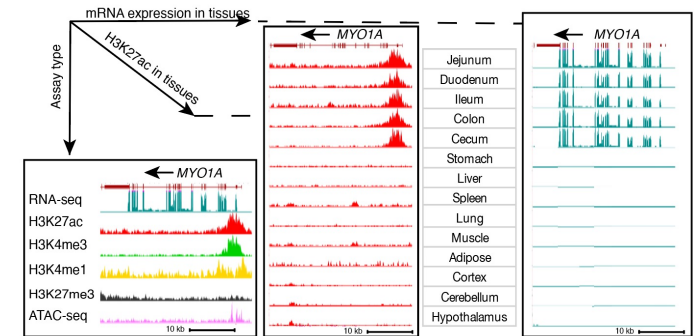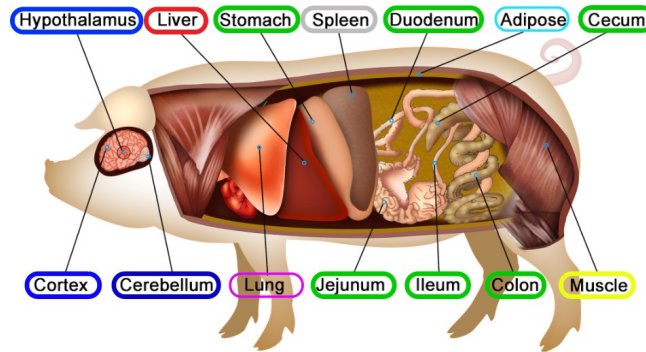
- Most comprehensive catalogue to date of pig regulatory elements

- Integrated 223 epigenomic and transcriptomic data sets, representing 14 biologically important tissues.
- Annotated 15 different chromatin states and defining their tissue-specific regulatory activities.
- **Identified 2,097,958 regulatory elements** (excluding Qui)

Hypothalamus · Liver · Stomach · Spleen · Duodenum · Adipose · Cecum
Cortex · Cerebellum · Lung · Jejunum · Ileum · Colon · Muscle

mRNA expression in tissues
H3K27ac in tissues
Assay type

MYO1A

RNA-seq
H3K27ac
H3K4me3
H3K4me1
H3K27me3
ATAC-seq

Jejunum, Duodenum, Ileum, Colon, Cecum, Stomach, Liver, Spleen, Lung, Muscle, Adipose, Cortex, Cerebellum, Hypothalamus

a Chromatin state / b Abbreviation / c Emissions / d Cov. M±SD %

#	Chromatin state	Abbreviation	Cov. M±SD %
1	Strongly active promoters/transcripts	TssA	0.52 ± 0.07
2	Flanking active TSS without ATAC	TssAHet	0.14 ± 0.09
3	Transcribed at gene	TxFlnk	0.47 ± 0.13
4	Weak transcribed at gene	TxFlnkWk	0.25 ± 0.20
5	Transcribed region without ATAC	TxFlnkHet	0.20 ± 0.14
6	Strong active enhancer	EnhA	0.87 ± 0.30
7	Medium enhancer with ATAC	EnhAMe	1.07 ± 0.35
8	Weak active enhancer	EnhAWk	1.18 ± 0.50
9	Active enhancer no ATAC (hetero)	EnhAHet	0.82 ± 0.51
10	Poised enhancer	EnhPois	2.56 ± 0.61
11	ATAC island	ATAC_Is	4.79 ± 2.55
12	Bivalent/poised TSS	TssBiv	0.50 ± 0.14
13	Repressed polycomb	Repr	2.55 ± 1.45
14	Weak repressed polycomb	ReprWk	10.70 ± 1.52
15	Quiescent	Qui	73.39 ± 4.77

Emissions columns: ATAC, H3K27ac, H3K4me3, H3K4me1, H3K27me3

e Fold enrichment of genome annotation

CpG_island	Gene	5'UTR	3'UTR	Exon	Intron	TSS	TSS_1k	TES	TES_1k	expr_gene	repr_gene	expr_TSS	repr_TSS	expr_TES	repr_TES
42.12	1.51	39.63	0.78	6.83	1.02	53.92	31.24	2.33	1.78	0.29	89.94	4.82	3.15	1.17	
3.00	1.66	7.28	1.24	2.26	1.58	5.19	10.75	1.60	1.74	1.95	0.35	7.46	2.07	2.33	1.07
12.38	1.51	10.25	1.78	3.28	1.32	8.14	12.70	2.41	2.82	1.75	0.49	12.73	1.88	3.48	0.92
7.00	1.58	7.36	1.47	2.78	1.45	6.07	12.05	2.30	2.41	1.80	0.56	8.26	3.10	3.22	0.98
2.11	1.67	6.12	1.24	1.93	1.62	2.24	10.74	2.12	2.07	1.98	0.27	3.39	0.82	3.14	0.66
1.64	1.42	1.92	2.40	2.01	1.37	1.40	1.49	1.40	1.87	1.64	0.44	1.82	0.84	2.15	0.59
2.71	1.39	2.06	2.25	2.18	1.33	1.26	1.40	1.59	1.91	1.56	0.65	1.37	1.10	2.16	0.67
0.38	1.35	0.95	1.76	1.35	1.36	0.85	1.09	1.48	1.38	1.56	0.43	0.99	0.84	2.06	0.67
0.40	1.50	1.48	1.69	1.28	1.52	0.61	1.19	1.35	1.32	1.76	0.33	0.74	0.41	1.97	0.48
0.49	1.60	1.69	1.32	1.18	1.64	0.56	0.91	1.20	1.13	1.88	0.35	0.60	0.50	1.68	0.52
4.60	1.24	1.22	2.61	2.75	1.14	1.02	1.03	1.90	2.01	1.30	1.00	0.84	1.25	2.43	1.18
32.76	1.25	13.67	0.96	3.83	0.97	17.83	13.98	2.45	2.50	1.22	1.30	22.86	9.99	2.19	2.69
2.44	0.87	1.12	0.78	1.01	0.84	1.67	2.16	1.35	1.41	0.68	1.74	1.10	2.42	0.91	1.92
0.25	0.80	0.42	0.50	0.56	0.81	0.63	0.73	0.92	0.91	0.64	1.48	0.27	1.13	0.81	1.35
0.23	0.95	0.51	0.89	0.82	0.96	0.48	0.52	0.86	0.85	0.95	0.94	0.23	0.84	0.82	0.91

f Evolutionary cons.

Enrichment for conserved elements

Aim 1 current processing (12 additional adult tissues)

2024

Tissue	Animal	H3K4me3	H3K27me3	H3K27ac	H3K4me1	ATAC-seq	RNA-seq
Bone Marrow	P348						
Bone Marrow	P350				▨		
Bladder	P348						
Bladder	P350	▨	▨	▨	▨		
Heart	P348						
Heart	P350						
Kidney	P348		▨				
Kidney	P350						
Mammarygland	6796						
Mammarygland	6797						
Ovary	6796	▩	▩	▩	▩		
Ovary	6797		▧		▩		
Skin	P348						
Skin	P350						
Testis	6798						
Testis	6800	▨			▨		
Thyroidgland	P348		▧	▩	▧		
Thyroidgland	P350		▧	▩	▧		
Thymus	P348			▩	▧		
Thymus	P350						
Trachea	P348	▨	▨	▨	▨		
Trachea	P350	▨	▨	▨	▨		
Uterus	6796	▨	▨	▨	▨		
Uterus	6797	▨	▨	▨	▨		

Legend:
- Completed
- Repeated and pass QC
- Waiting for sequencing results
- Repeated and failed

Summary

Pig: P348, P350 and 6796, 6797 (for reproductive tissues)

Tissue and assay: 12 tissues and 4 ChIP-seq assays

Library: 168 libraries (138 completed, 24 pending and 6 failed for 9 tissues)
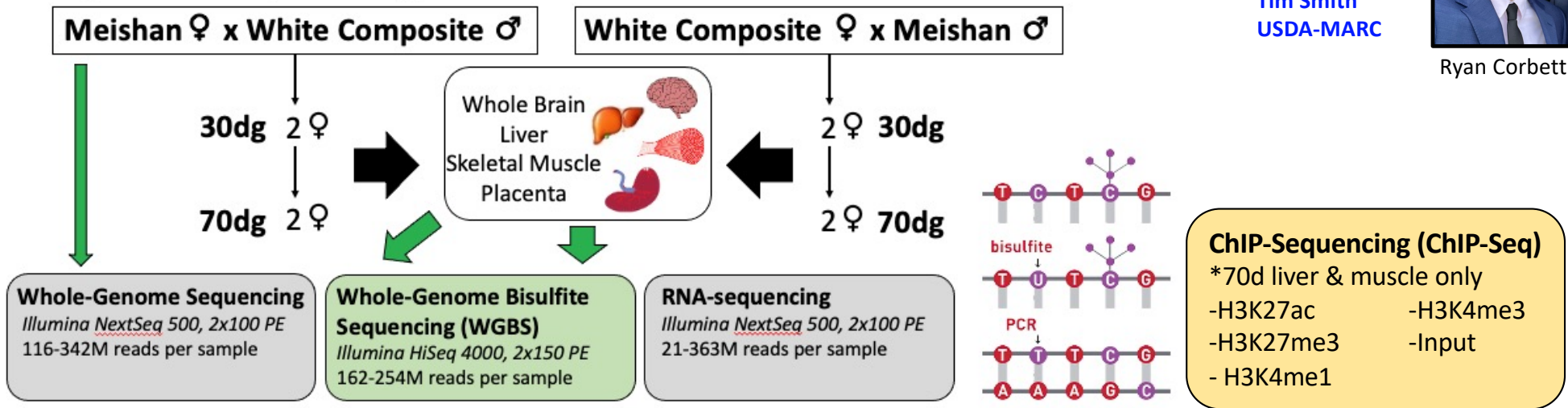
Next steps- Aim 1

- 12 tissues worked on currently:
  - 3 tissues completed (heart, mammary gland and skin).
  - The sequencing results for the remaining 24 repeated samples for 9 tissues are queued for QC and will be processed soon.
- Libraries will be considered completed when passing all QC parameters and getting the expected sequencing depth.
- All sequencing raw data will be submitted to ENA upon completion.

FAANG
Functional Annotation of Animal Genomes

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Aim 2: Characterization of allele-biased expression and epigenetic modifications in porcine fetal tissues

Collaboration:
Dan Nonneman,
Tim Smith
USDA-MARC

Ryan Corbett

Meishan ♀ x White Composite ♂

White Composite ♀ x Meishan ♂

30dg 2♀

2♀ 30dg

70dg 2♀

2♀ 70dg

Whole Brain
Liver
Skeletal Muscle
Placenta

bisulfite

PCR

ChIP-Sequencing (ChIP-Seq)
*70d liver & muscle only
-H3K27ac   -H3K4me3
-H3K27me3   -Input
- H3K4me1

Whole-Genome Sequencing
Illumina NextSeq 500, 2x100 PE
116-342M reads per sample

Whole-Genome Bisulfite Sequencing (WGBS)
Illumina HiSeq 4000, 2x150 PE
162-254M reads per sample

RNA-sequencing
Illumina NextSeq 500, 2x100 PE
21-363M reads per sample

Negative binomial regression model:

$$E = \mu + AG + PO + ID + \varepsilon$$

E = allele count
m = population mean
AG = allelic genotype
PO = parent of origin
ID = animal ID

	WC♀ X MS♂		MS♀ X WC♂	
	WC♀	MS♂	MS♀	WC♂
AG	1	0	0	1
PO	0	1	0	1

Breed allele-biased expression is widespread and impacts unique biological pathways

← WC Bias   MS Bias →

brain–30dg

ASE: 13   ABE: 274   ABE: 193   ASE: 8

GABRA6
CCT2
ADAM10
CRYAB   C11orf1
PTRM1
ELOVL6
SPOCK3A1

liver–30dg

ASE: 21   ABE: 171   ABE: 161   ASE: 9

MT1A
MT1A
LY75
MCFD2
DCT2
METTL8
KDM5B
GCH1   SPOCK3
COL19A

muscle–30dg

ASE: 20   ABE: 121   ABE: 89   ASE: 8

CCT2
ADAM10
METTL8
CRYAB
ELOVL6
SPOCK3

placenta–30dg

ASE: 31   ABE: 230   ABE: 189   ASE: 7

CCT2
ADAM10
METTL8
SPAG16
SLC4FB21
COL19A1
SPOCK3

-log10(FDR)

Meishan allele frequency

White Composite-biased genes

leukocyte_degranulation
myeloid_cell_activation_involved_in_immune_response
androgen_catabolic_process
positive_regulation_of_DNA_biosynthetic_process
positive_regulation_of_telomere_maintenance_via_telomerase
glycogen_catabolic_process
cellular_carbohydrate_catabolic_process
lipid_catabolic_process
antibacterial_humoral_response
vacuolar_transmembrane_transport
regulated_exocytosis
cholesterol_import
metallopeptidase_activity
carboxylic_acid_transport
response_to_nutrient
hormone_metabolic_process
protein_targeting_to_ER
lipoxin_A4_metabolic_process
small_molecule_metabolic_process

Brain30 Brain70 Liver30 Liver70 Muscle30 Muscle70 Placenta30 Placenta70

Meishan-biased genes

translational_elongation
negative_regulation_of_low-density_lipoprotein_receptor_activity
lipoxin_A4_metabolic_process
regulation_of_cell_communication_by_electrical_coupling
regulation_of_relaxation_of_muscle
translation
type_I_interferon_signaling_pathway
lipid_catabolic_process
regulation_of_steroid_metabolic_process
interferon-gamma-mediated_signaling_pathway
negative_regulation_of_neurotransmitter_uptake
small_molecule_metabolic_process
fatty_acid_metabolic_process
lipid_metabolic_process

Brain30 Brain70 Liver30 Liver70 Muscle30 Muscle70 Placenta30 Placenta70

Parent-of-origin ABE is observed in human-validated imprinted genes

brain–30dg

ASE: 19   ABE: 61   ABE: 31   ASE: 7

liver–30dg

ASE: 13   ABE: 32   ABE: 29   ASE: 7

muscle–30dg

ASE: 16   ABE: 30   ABE: 20   ASE: 7

placenta–30dg

ASE: 20   ABE: 18   ABE: 64   ASE: 11

Maternal allele frequency

← Pat Bias   Mat Bias →

- Validated paternally-expressed genes are enriched in PO ABE genes

- PO ASE genes provide evidence for putative novel imprinted loci

ABE
Biallelic
Not Covered

Pig–Validated
Human/Mouse–Validated
Novel
Novel, ICR–Proximal

Allele biases in Chromatin Modifications: H3K27ac and H3K4me3 ChIP peak strength biases are prevalent in fetal liver and muscle and <u>associated with ABE</u>

- Parent-of-origin allele biased histone modifications (ABHM) are associated with validated imprinted genes
- Biases in H3K27ac, H3K4me3 peak allele counts are associated with genes exhibiting ABE
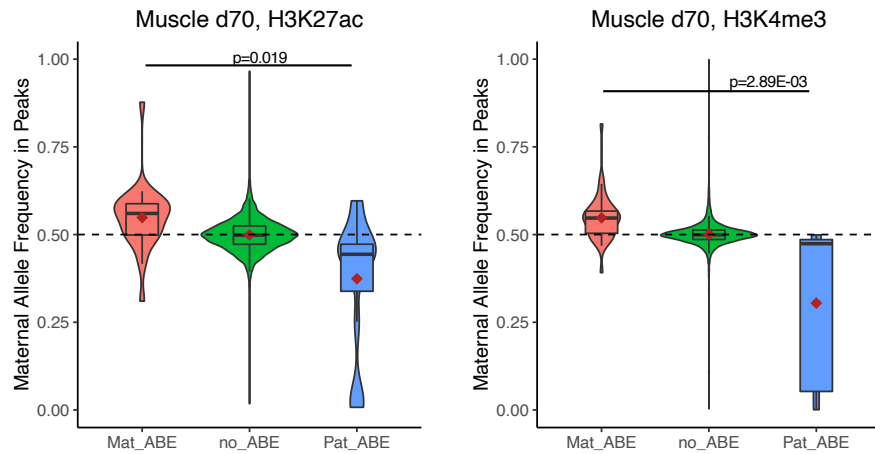
Frequency of maternal alleles in H3K27ac & H3K4me3 peaks overlapping promoters of genes exhibiting maternal, paternal, and no ABE in d70 muscle.
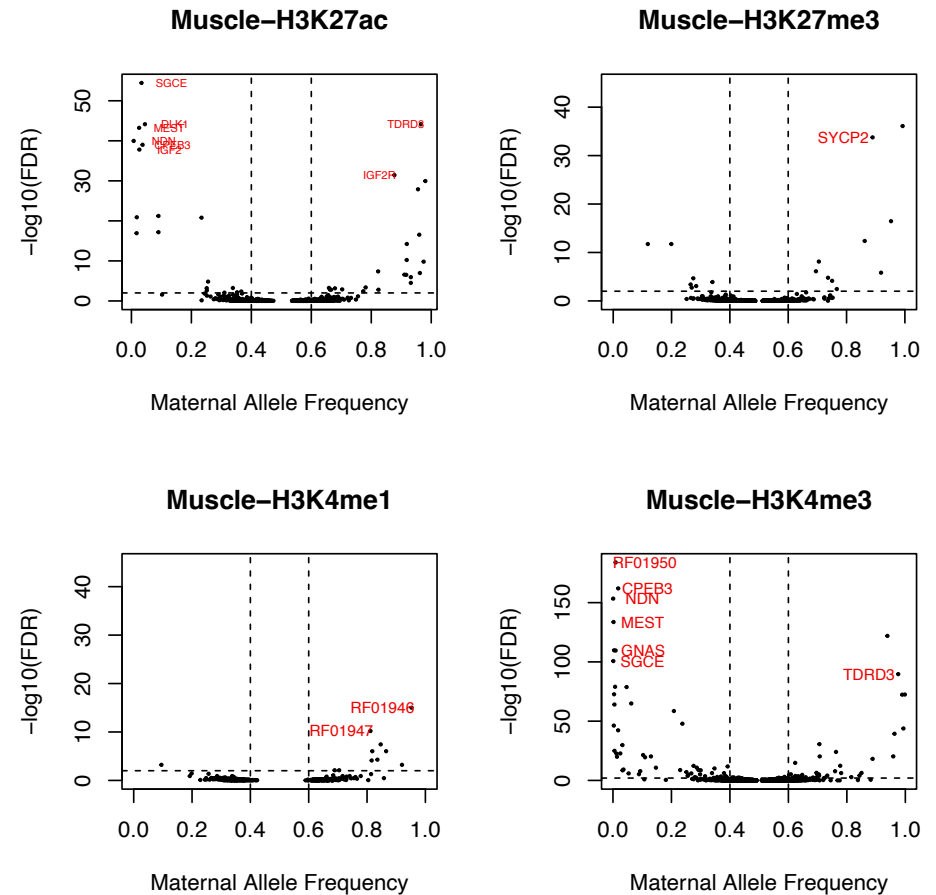
Next steps- Aim 2

- Day 70 stage RNA, WGBS complete. HM data – 50% of tissues completed. All data submitted. Data analysis nearly complete. Manuscript to be completed in 2024.

- Day 30 stage data- only RNA and methylation completed, data submitted.
  - Investigating Cut &RUN to see if we can use small amounts of tissue for histone assays.

Aim 3: Functional Annotation of Circulating Porcine Immune Cell Populations

Published:
Herrera-Uribe et al. 2020
Herrera-Uribe & Wiarda 2021
Corbett et al. 2022
Herrera-Uribe et al. 2023
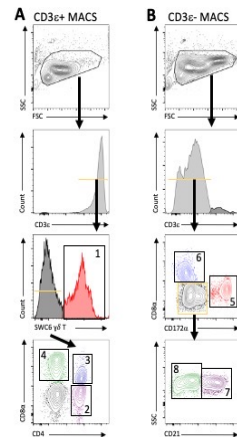
Isolation of eight leukocyte pops from pig PBMCs.
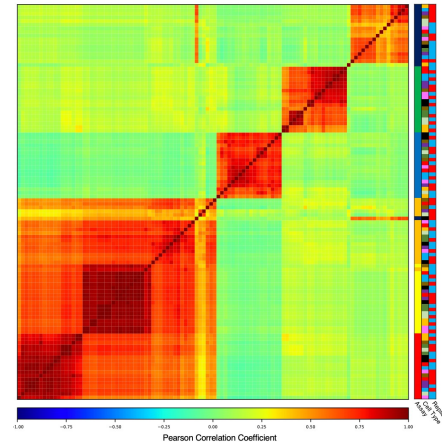
Global correlation among assays, tissues, and replicates
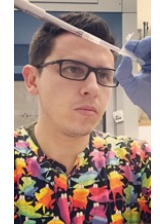
Ryan Corbett　　**Juber Herrera Uribe**

Collaboration:
Lingzhao Fang, et al. Aarhus U
Crystal Loving, USDA-NADC
Cathy Ernst, Michigan State

Overview of immune cell samples and epigenomic marks assayed

Distribution of marks around transcription start sites (TSS)

Genome-wide integration of epigenomic data can predict immune cell chromatin states for each location in the genome for each cell type

mark prob. (0 – 1)

Mark columns (heatmap): CTCF, ATAC, H3K4me3, H3K27ac, H3K4me1, H3K27me3

Chromatin state	Cov (%)	CpGIslands	Genes	TSS1k	TSS	X5prime_UTR	Exons	Introns	X3prime_UTR	TES	TES1k
Repressed Polycomb (Repr)	8.47 (+5.60)	1.21	0.95	1.62	1.06	1.13	0.94	1.01	0.93	0.98	1.1
Poised Enhancer (EnhPois)	0.52 (+0.15)	18.16	1.15	8.61	8.87	4.21	1.01	2.99	1.63	2.11	2.12
Poised TSS (TssPois)	0.24 (+0.06)	50.73	1.28	17.97	33.94	11.51	0.95	5.94	1.21	2.23	1.97
Flanking TSS (TssFlnk)	0.35 (+0.08)	5.79	1.52	14.41	4.7	5.21	1.46	2.98	1.63	2.53	2.37
Active TSS (TssA)	0.37 (+0.09)	26.85	1.33	24.15	31.64	10.05	1.12	4.64	1.77	2.61	2.96
Active Enhancer (EnhA1)	0.54 (+0.30)	0.64	1.3	2.02	1.69	1.59	1.31	1.36	1.55	1.33	1.4
ATAC Island (ATACisl)	0.90 (+0.89)	0.38	0.62	0.74	0.88	0.68	0.63	0.58	0.53	0.44	0.6
Strong Active TSS (TssStrA)	0.24 (+0.05)	55.35	1.47	30.48	106.55	22.29	1.02	8.22	0.96	1.6	2.16
Downstream TSS (TssDwn)	0.08 (+0.11)	10.32	1.43	6.85	5.83	5.15	1.37	2.72	1.33	1.54	1.28
Quiescent (Qui)	83.3 (+6.18)	0.5	0.98	0.56	0.43	0.79	0.98	0.91	0.97	0.94	0.94
Active Region (Act)	0.38 (+0.25)	0.73	0.94	0.58	0.5	0.75	0.95	0.91	1	1.28	1.06
Active Enhancer (EnhA2)	0.18 (+0.11)	1.1	1.44	2.96	1.23	1.98	1.43	1.81	1.99	2.17	1.85
H3K4me1 (H3K4me1)	4.01 (+0.82)	1.09	1.47	2.39	1.11	1.88	1.47	1.68	1.67	1.84	1.68
Enhancer w/ CTCF (EnhCTCF)	0.13 (+0.06)	4.17	1.09	3.23	2.09	2.03	1	2.42	2.68	3.46	3.16
Insulator (Ins)	0.30 (+0.16)	0.76	1.01	0.98	0.64	1.02	0.99	1.41	1.7	1.92	1.95

Feature Enrichment

Genome-wide integration of epigenomic data can predict immune cell chromatin states for each location in the genome for each cell type

Chromatin state	TSS_repr	TSS_exprQ1	TSS_exprQ2	TSS_exprQ3	TSS_exprQ4
Repressed Polycomb (Repr)	2.31	0.52	0.13	0.08	0.09
Poised Enhancer (EnhPois)	21.90	6.70	1.75	0.93	0.99
Poised TSS (TssPois)	66.92	68.69	19.38	10.43	6.91
Flanking TSS (TssFlnk)	2.07	6.24	6.55	7.03	11.12
Active TSS (TssA)	10.73	59.08	61.80	56.60	51.96
Active Enhancer (EnhA1)	1.25	2.55	1.51	1.23	1.59
ATAC Island (ATACisl)	1.14	0.89	0.66	0.36	0.66
Strong Active TSS (TssStrA)	12.12	122.42	223.87	250.59	247.66
Downstream TSS (TssDwn)	2.27	5.56	6.06	8.88	15.67
Quiescent (Qui)	0.55	0.21	0.13	0.11	0.10
Active Region (Act)	0.56	0.19	0.14	0.16	0.13
Active Enhancer (EnhA2)	0.65	0.97	0.83	1.35	2.92
H3K4me1 (H3K4me1)	0.82	0.96	0.97	1.10	1.52
Enhancer w/ CTCF (EnhCTCF)	2.19	2.92	1.48	1.18	2.10
Insulator (Ins)	0.85	0.30	0.21	0.14	0.13

Integration of chromatin state prediction and gene expression in different cell types
Example: CD8A,B expression/regulation in CD8T, NK cells

Cell-type specific Regulatory Elements (csRE) are:
a) active enhancers and ATACseq islands (open chromatin)
b) targets of csRE are enriched in relevant cell pathways

Cell-type specific Regulatory Elements:
Transcription factor prediction and integration with Tissue RE

Enriched TF binding motifs among cell- and lineage-specific EnhA1 states

Enrichment of pig tissue enhancers among immune-cell specific enhancers
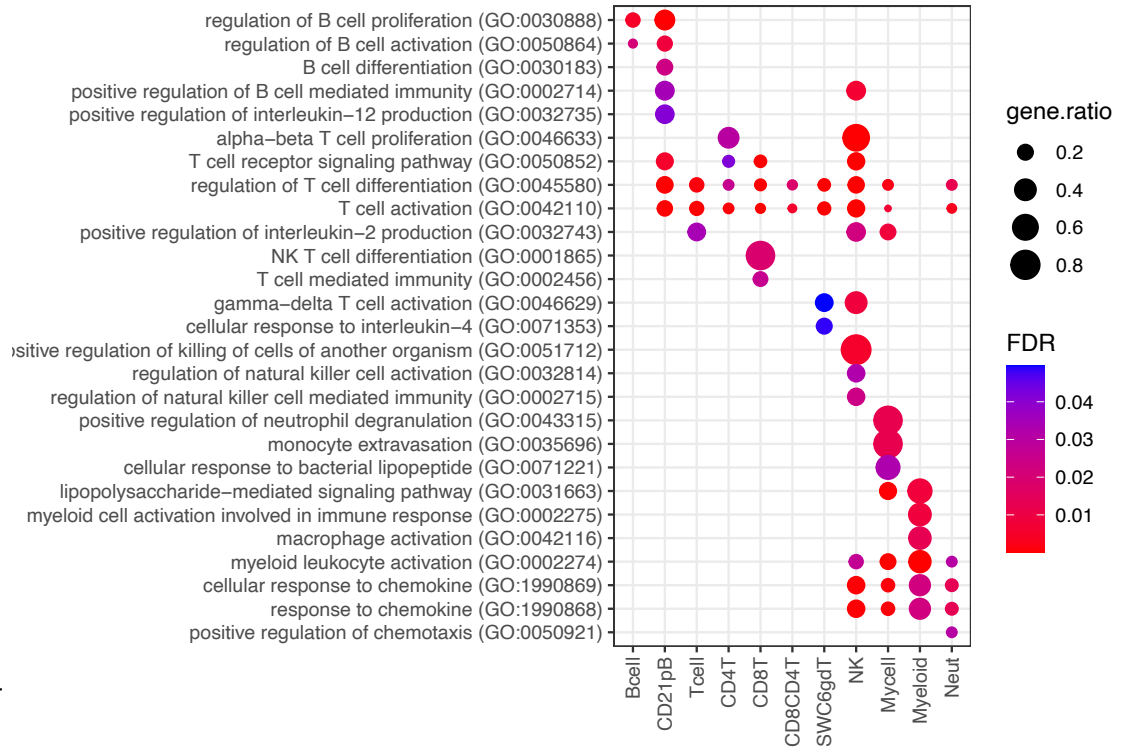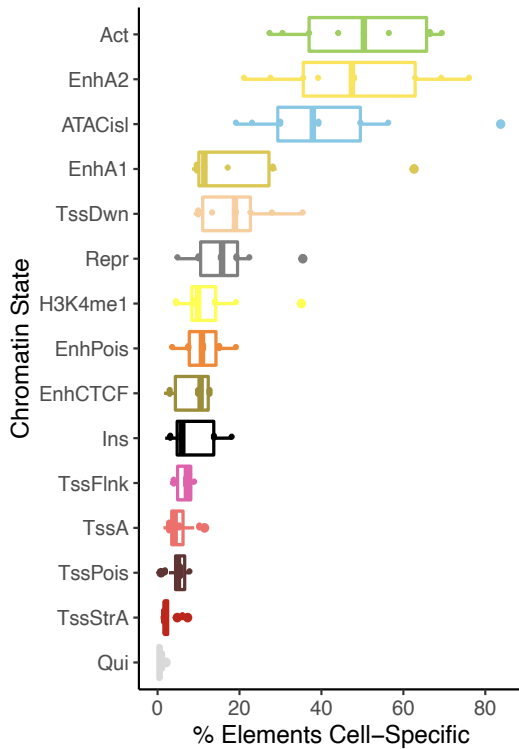
Example of confirming and extending Tissue Epigenomics data: BANK1 B cell specific enhancer
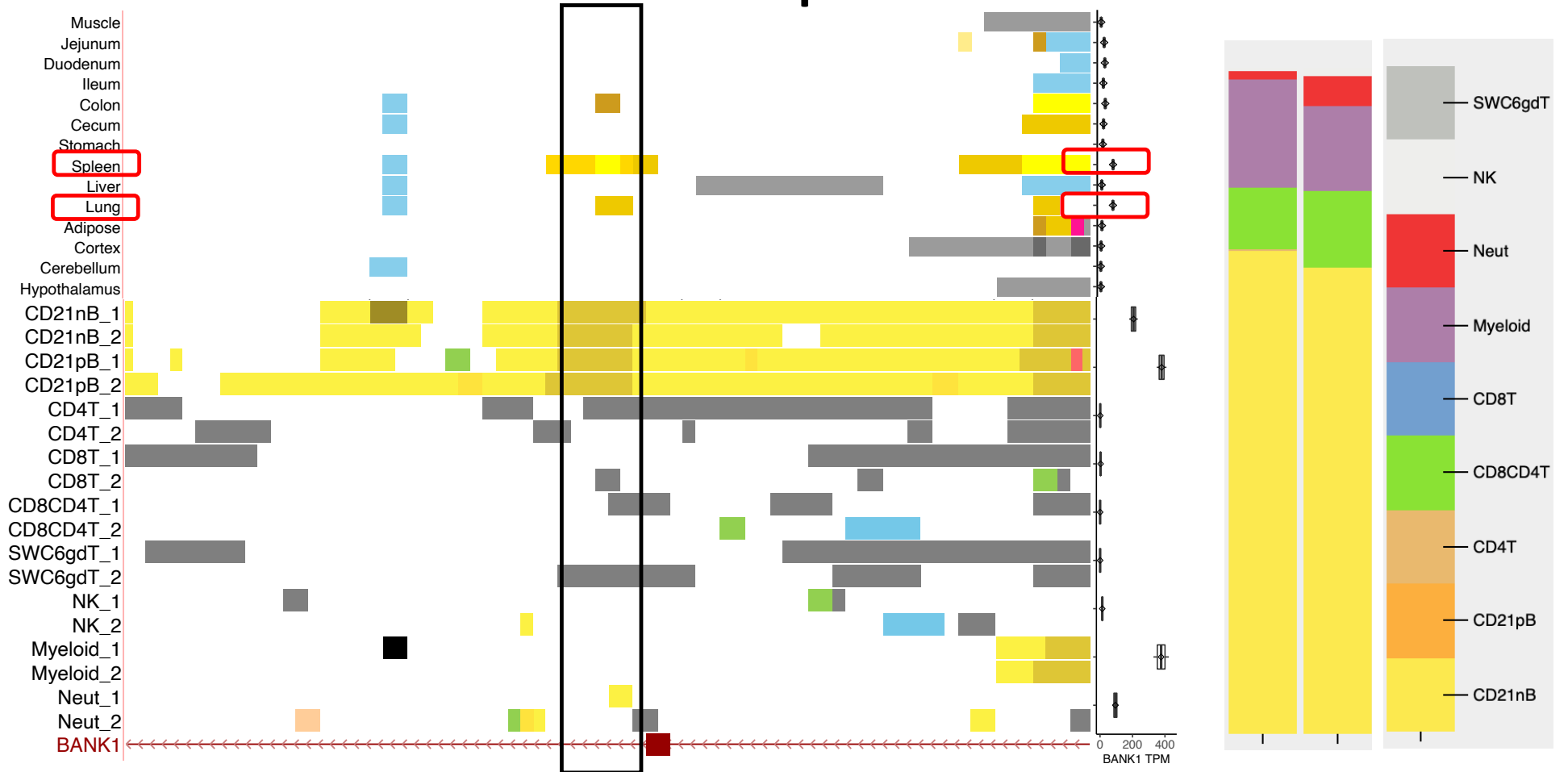
Super-enhancers: unique regulatory features

Enhancer clusters ranked by relative H3K27ac signal in each cell type: very high H3K27 signal = SE
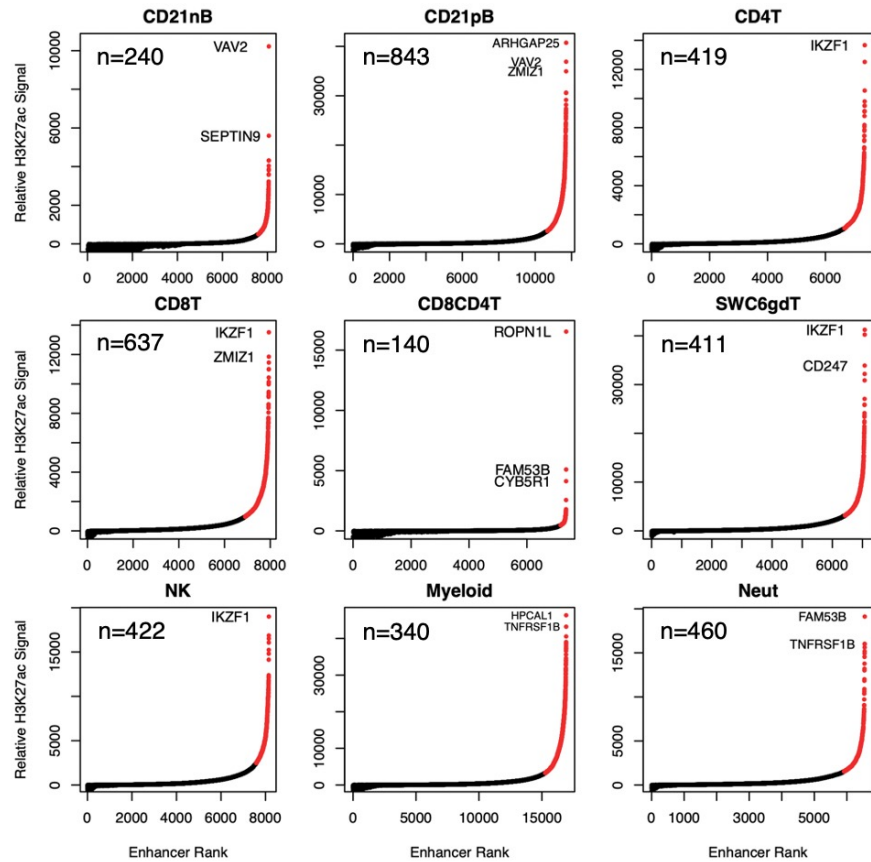
Transcript abundances across cell populations of genes possessing cell-specific SEs (csSEs) (red=sig diff)
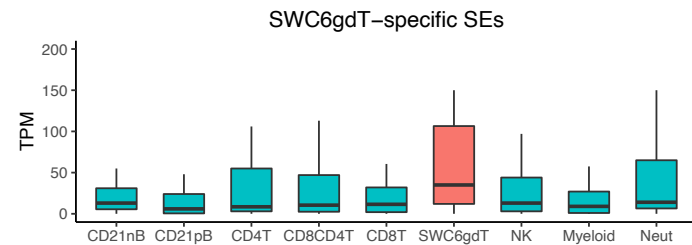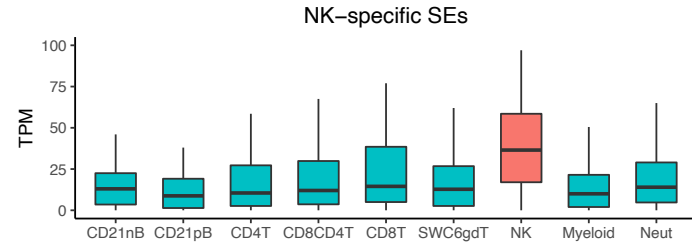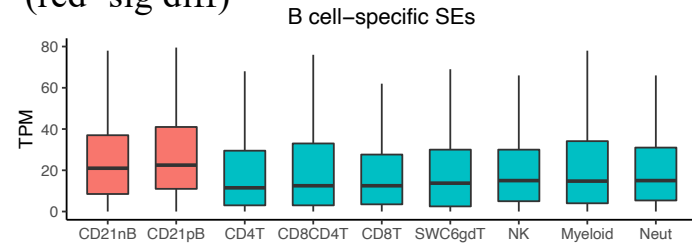
Super-enhancers: gene examples

H3K27ac signal and chromatin states within a *SPI1* (myeloid cells) or RUNX3 SE (in CD8T cells), and relationship between SE presence and *SPI1 or RUNX3* expression across cell types
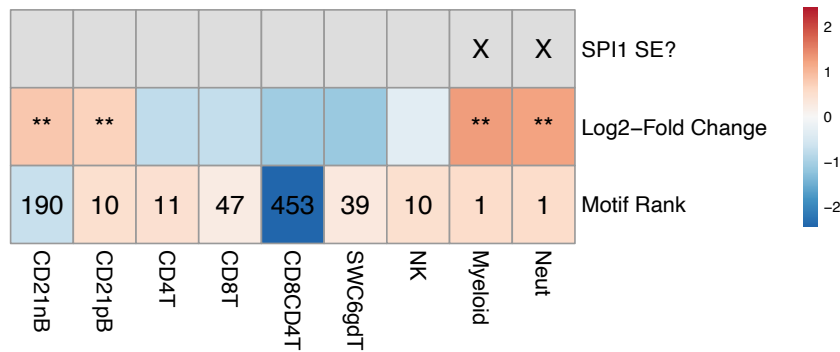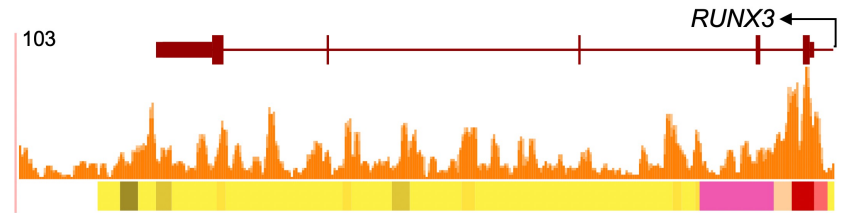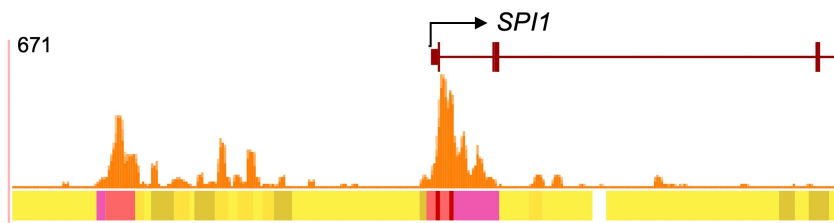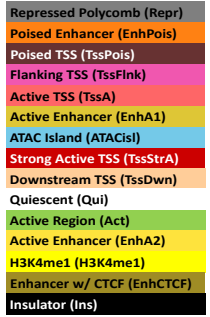
Spi1 motif enrichment across csSEs
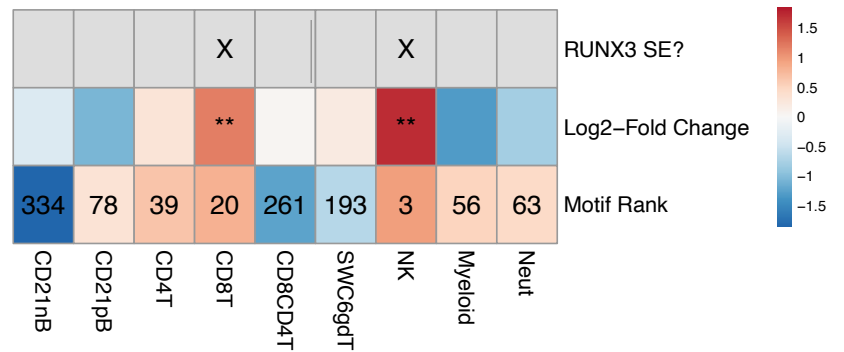
Runx3 motif enrichment across csSEs

Active chromatin states are enriched for Pig GTEx eQTL in immune tissues
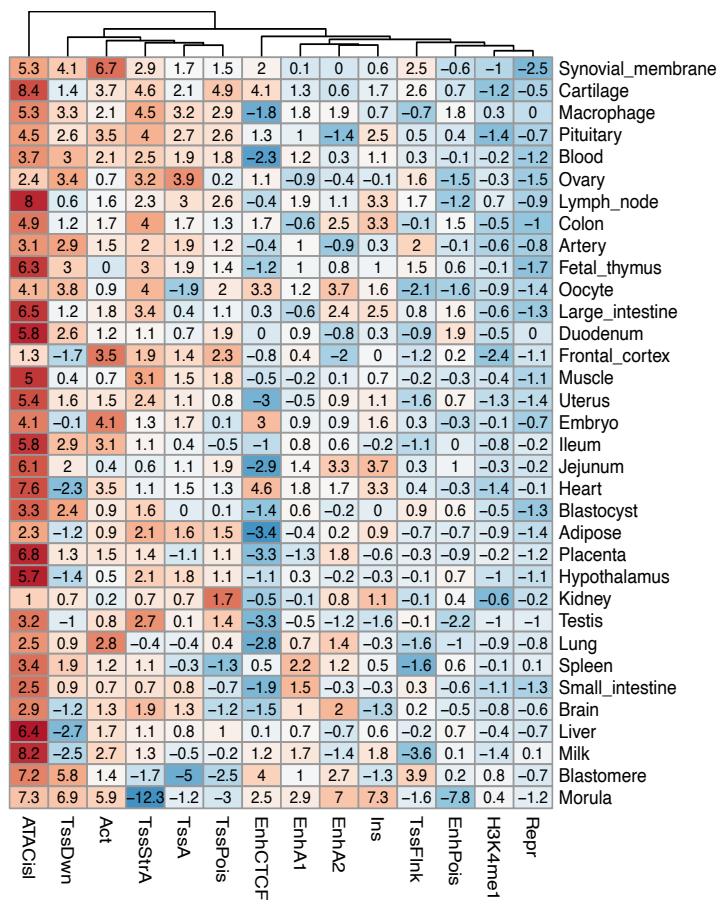
Macrophage *IGSF8* eQTL peak SNPs co-localize with a myeloid-specific ATAC island/active enhancer
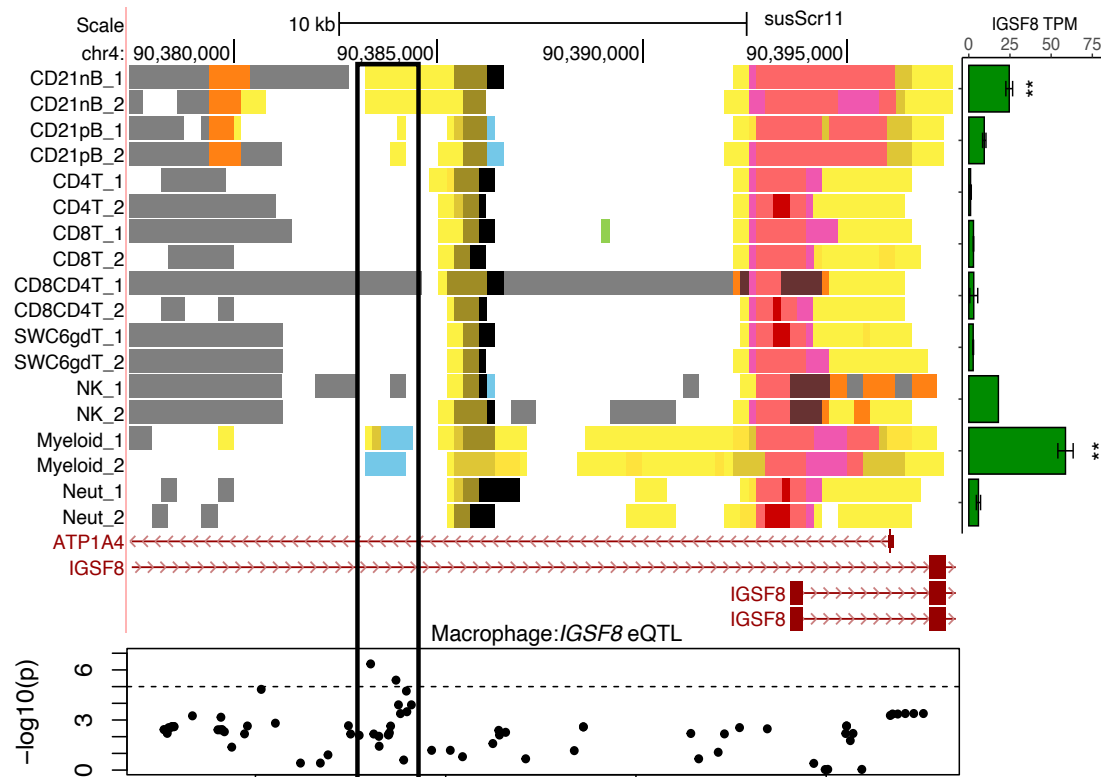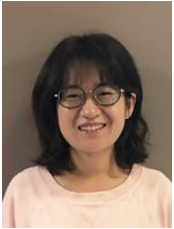
FAANG Project addition: snATAC-seq porcine PBMC chromatin accessibility is powerful for identifying regulatory elements

Pengxin Yang
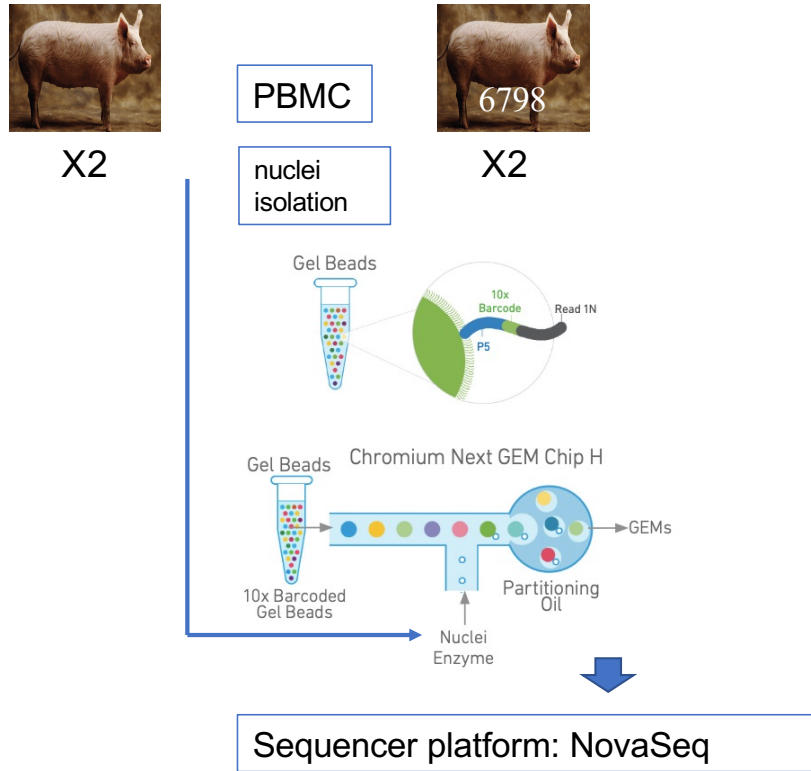
snATAC-seq and scRNA-seq have similar power to distinguish cell types
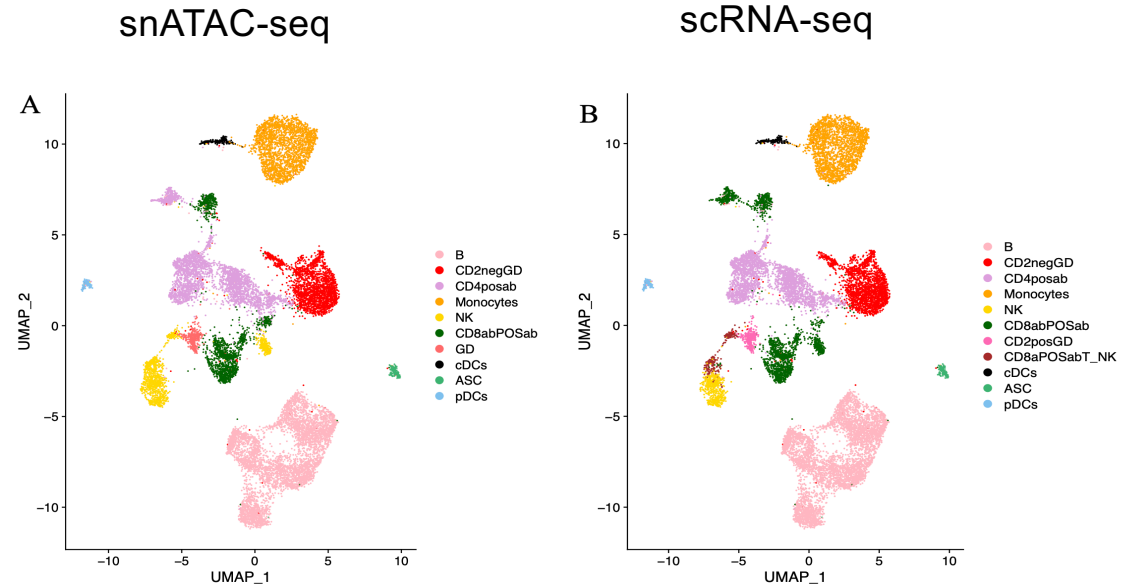
Cell type specific differentially accessible peaks (DAPs): clues to immune regulatory control

14,092 unique cell type-specific DAPs

Cell type	# DAPs	# closest genes
cDCs	1481	1178
Monocytes	2524	1943
CD4posab	557	429
CD2negGD	572	435
B	1114	855
ASC	1455	1109
CD2posGD	602	456
CD8abPOSab	342	256
NK	1157	909
pDCs	2402	1890
CD8aPOSabT_NK	1220	968

DAP analysis identifies a cis-element within PAX5 for B cells, which specifically express PAX5
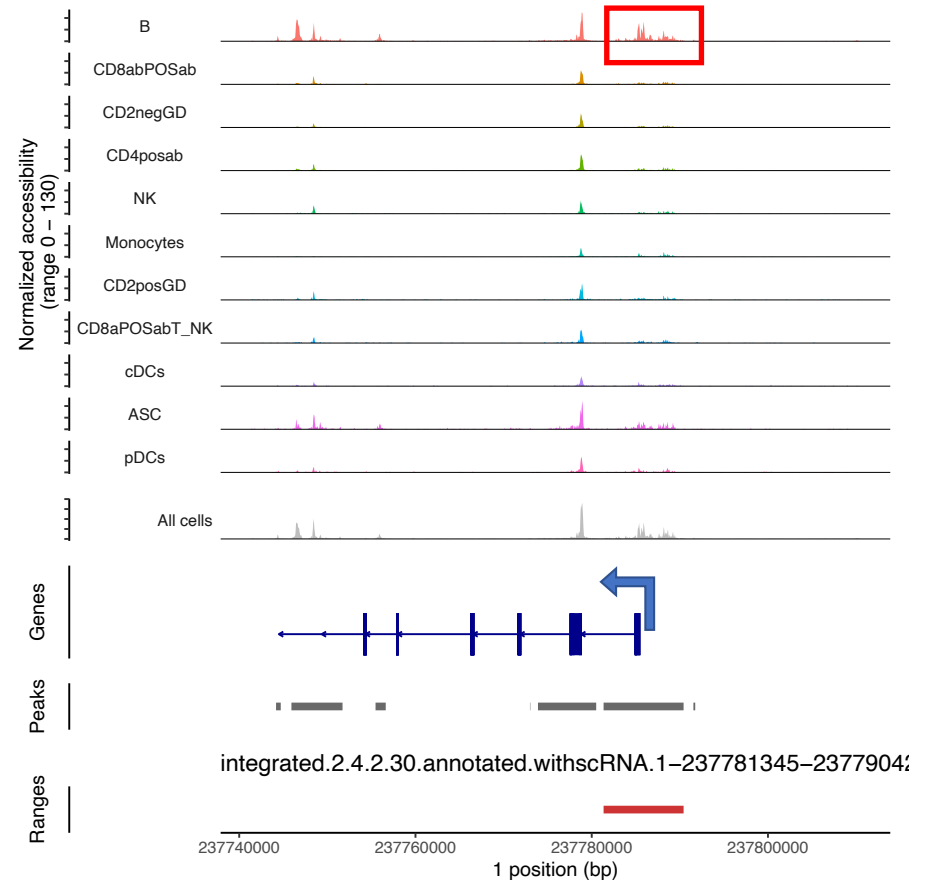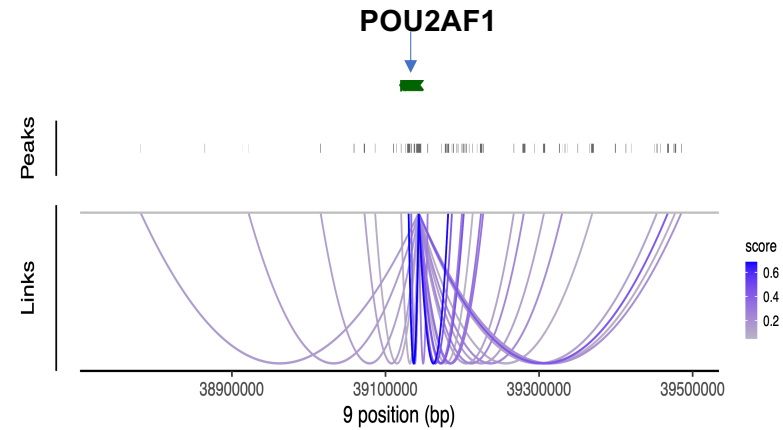
TFBM analysis on the cell type specific open chromatin region identifies relevant TF for cell type regulatory networks

TFs

cell type
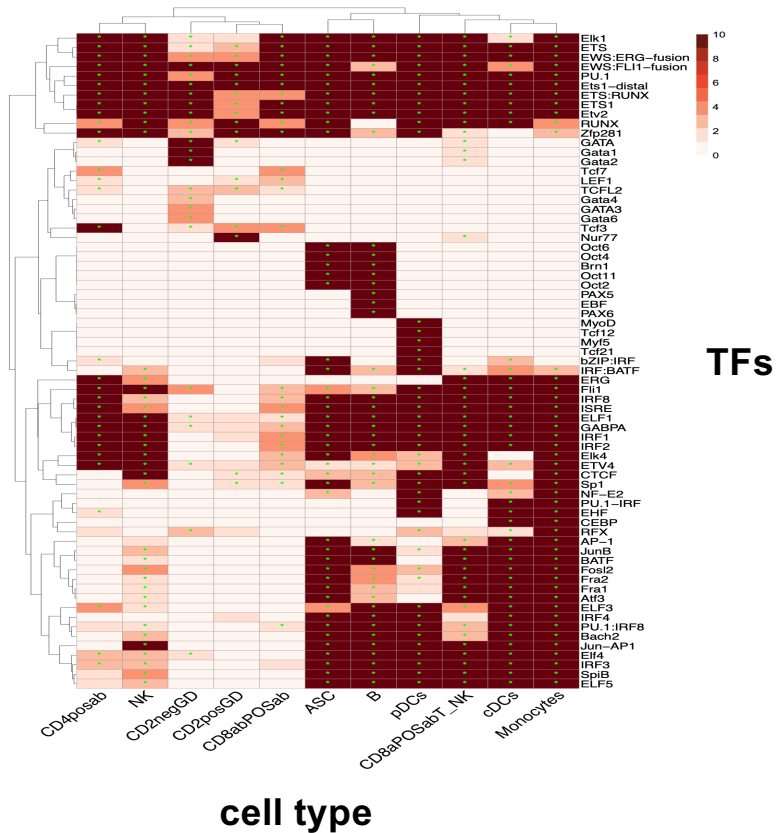
Linking regulatory regions with target genes: cis co-accessibility network (CCAN) at a DEG

Pengxin Yang et al.
Definition of regulatory elements and transcription factors controlling immune cell gene expression at single cell resolution using single nucleus ATAC-seq

Submitted to Genomics, under review

Data status summary and next steps: Aim 3

Alveolar Mac +/- stim: → published

Neutrophils: → published

Sorted cells:

    RNA seq   → published

    Methyl-seq   → published

    Histone/CTCF   → data complete + sub'd, data analysis complete

    ATAC-seq   → data complete + sub'd, data analysis complete

Identification of cell-type specific regulatory elements in all major circulating cell types. Paper is in final stages of writing, to be submitted "soon".

Single cells:

    PBMC: scRNAseq   → published

    PBMC: scATACseq   → data complete/sub'd, data analysis complete, manuscript submitted

First report of single nucleus ATACseq analysis of PBMC, open chromatin regions within each cell type predicting regulatory regions, TF and targets
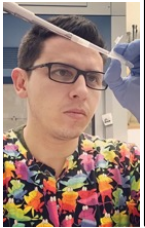
    Tissues: scRNAseq   → data complete, initial analysis ongoing,

Data in submission process, data interpretation ongoing

Acknowledgments- PIGGI@iastate.edu group
(that's Pig ImmunoGenomics and Genetics in Iowa!)

Christopher Tuggle

Juber Herrera Uribe

Ryan Corbett

Pengxin Yang

Muskan Kapoor

Carrie Meeks

Mehak Kapoor

Crystal Loving

Jayne Wiarda

Kristen Byrne

Sathesh Sivasankaran

James Koltes

Bruna Petry

Kyu-sang Lim

Jack Dekkers

Luke Kramer

IOWA STATE UNIVERSITY

USDA Agricultural Research Service

We thank all pig FAANG collaborators!

UC DAVIS UNIVERSITY OF CALIFORNIA

MICHIGAN STATE UNIVERSITY

USDA Agricultural Research Service

USDA Agricultural Research Service

Houcheng Li Lingzhao Fang
Aarhus University

Huaijun Zhou, et al.

Cathy Ernst, et al.

Tim Smith, Dan Nonneman MARC

Joan Lunney BARC

Jinyan Teng Zhe Zhang
So. China Ag University

USDA United States Department of Agriculture National Institute of Food and Agriculture

FAANG Functional Annotation of Animal Genomes

27