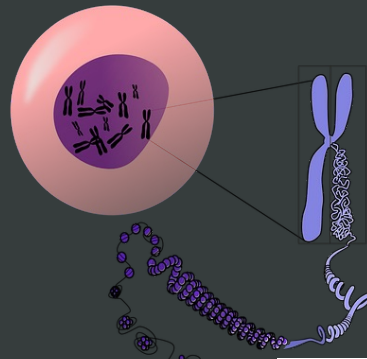


Equine FAANG: From Starting Gate to Winner's Circle

Ted Kalbfleisch, Sichong Peng, Alexa M. Barber, Rebecca R. Bellone, Nicole M. Kingsley, Eleonara Cappaletti, Francesca M. Pires, Elena Giulotto, Carrie J. Finno, Jessica L. Petersen



Generation of a Biobank From Two Adult Thoroughbred Stallions for the Functional Annotation of Animal Genomes Initiative

Callum G Donnelly¹, Rebecca R Bellone^{1,2}, Erin N Hales³, Annee Nguyen¹, Scott A Katzman⁴, Ghislaine A Dujovne¹, Kelly E Knickelbein⁴, Felipe Avila², Ted S Kalbfleisch⁵, Elena Giulotto⁶, Nicole B Kingsley², Jocelyn Tanaka², Elizabeth Esdaile², Sichong Peng¹, Anna Dahlgren¹, Anna Fuller⁷, Michael J Mienaltowski⁸, Terje Raudsepp⁹, Verena K Affolter¹⁰, Jessica L Petersen⁷, Carrie J Finno¹

UCD/UNL Biobank



Original Article | [Free Access](#)

Generation of an equine biobank to be used for Functional Annotation of Animal Genomes project

E. N. Burns, M. H. Bordbari, M. J. Mienaltowski, V. K. Affolter, M. V. Barro, F. Gianino, G. Gianino, E. Giulotto, T. S. Kalbfleisch, S. A. Katzman, M. Lassaline, T. Leeb, M. Mack, E. J. Müller ... [See all authors](#)

First published: 11 October 2018 | <https://doi.org/10.1111/age.12717> | Citations: 16

- 4 healthy Thoroughbreds (2 F, 2 M; 3-5 years)
- Full clinical phenotyping
 - > 80 tissues, 4 fluid types and 9 microbiome sites collected



Core FAANG Studies

8 tissues: adipose, cerebral cortex, heart, liver, lung, laminae, ovary, skeletal muscle



Data complete
Uploaded to EMBL-ENL





Sequencing
Complete






Samples obtained
Funding secured

Core FAANG Studies – 4 TBs

- Whole-genome sequencing (20x) 
- RNA-seq (mRNA and small RNA) 



Core FAANG Studies – 4 TBs

- Whole-genome sequencing (20x) 
- RNA-seq (mRNA and small RNA) 
- ChIP-seq: 
 - H3K4me1
 - H3K4me3
 - H3K27me3
 - H3K27ac









genes



Article

Functionally Annotating Regulatory Elements in the Equine Genome Using Histone Mark ChIP-Seq

N. B. Kingsley ^{1,2} , Colin Kern ³, Catherine Creppe ⁴, Erin N. Hales ² , Huaijun Zhou ³ , T. S. Kalbfleisch ⁵, James N. MacLeod ⁵, Jessica L. Petersen ⁶ , Carrie J. Finno ²  and Rebecca R. Bellone ^{1,2,*} 

frontiers
in Genetics






DATA REPORT
published: 26 March 2021
doi: 10.3389/fgene.2021.649959



“Adopt-a-Tissue” Initiative Advances Efforts to Identify Tissue-Specific Histone Marks in the Mare







N. B. Kingsley ^{1,2}, Natasha A. Hamilton ³, Gabriella Lindgren ^{4,5}, Ludovic Orlando ⁶, Ernie Bailey ⁷, Samantha Brooks ⁸, Molly McCue ⁹, T. S. Kalbfleisch ⁷, James N. MacLeod ⁷, Jessica L. Petersen ¹⁰, Carrie J. Finno ² and Rebecca R. Bellone ^{1,2*}

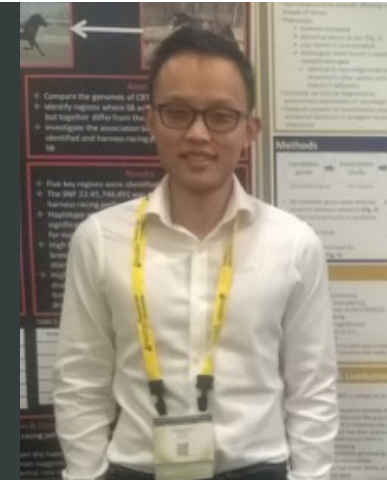
Core FAANG Studies – 4 TBs

- Whole-genome sequencing (20x) 
- RNA-seq (mRNA and small RNA) 
- ChIP-seq: 
 - H3K4me1
 - H3K4me3
 - H3K27me3
 - H3K27ac
- RRBS (National Research Institute, Poland)  

Mares Stallions







Core FAANG Studies – 4 TBs

- Whole-genome sequencing (20x) 
- RNA-seq (mRNA and small RNA) 
- ChIP-seq: 
 - H3K4me1
 - H3K4me3
 - H3K27me3
 - H3K27ac
- RRBS (National Research Institute, Poland)  
- CTCF-seq 



Mares Stallions

Core FAANG Studies – 4 TBs

- Whole-genome sequencing (20x) 
- RNA-seq (mRNA and small RNA) 
- ChIP-seq: 
 - H3K4me1
 - H3K4me3
 - H3K27me3
 - H3K27ac
- RRBS (National Research Institute, Poland) 
- CTCF-seq 
- ATAC-seq 



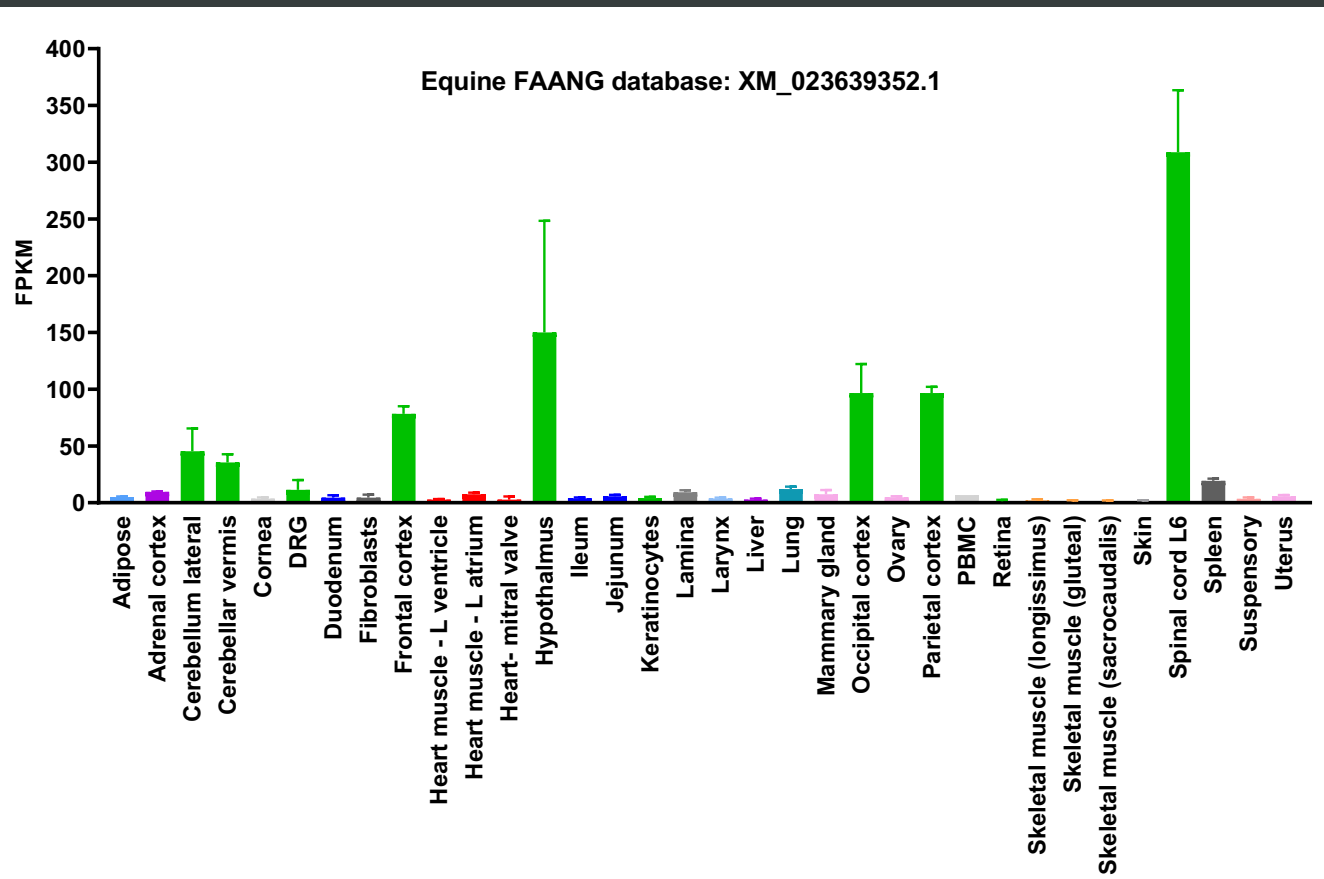
Mares Stallions



Adopt-A-Tissue Collaborators



Adopt-A-Tissue – RNA-seq



34 tissues from 2 horses

Adopt-A-Tissue – ChIP-seq

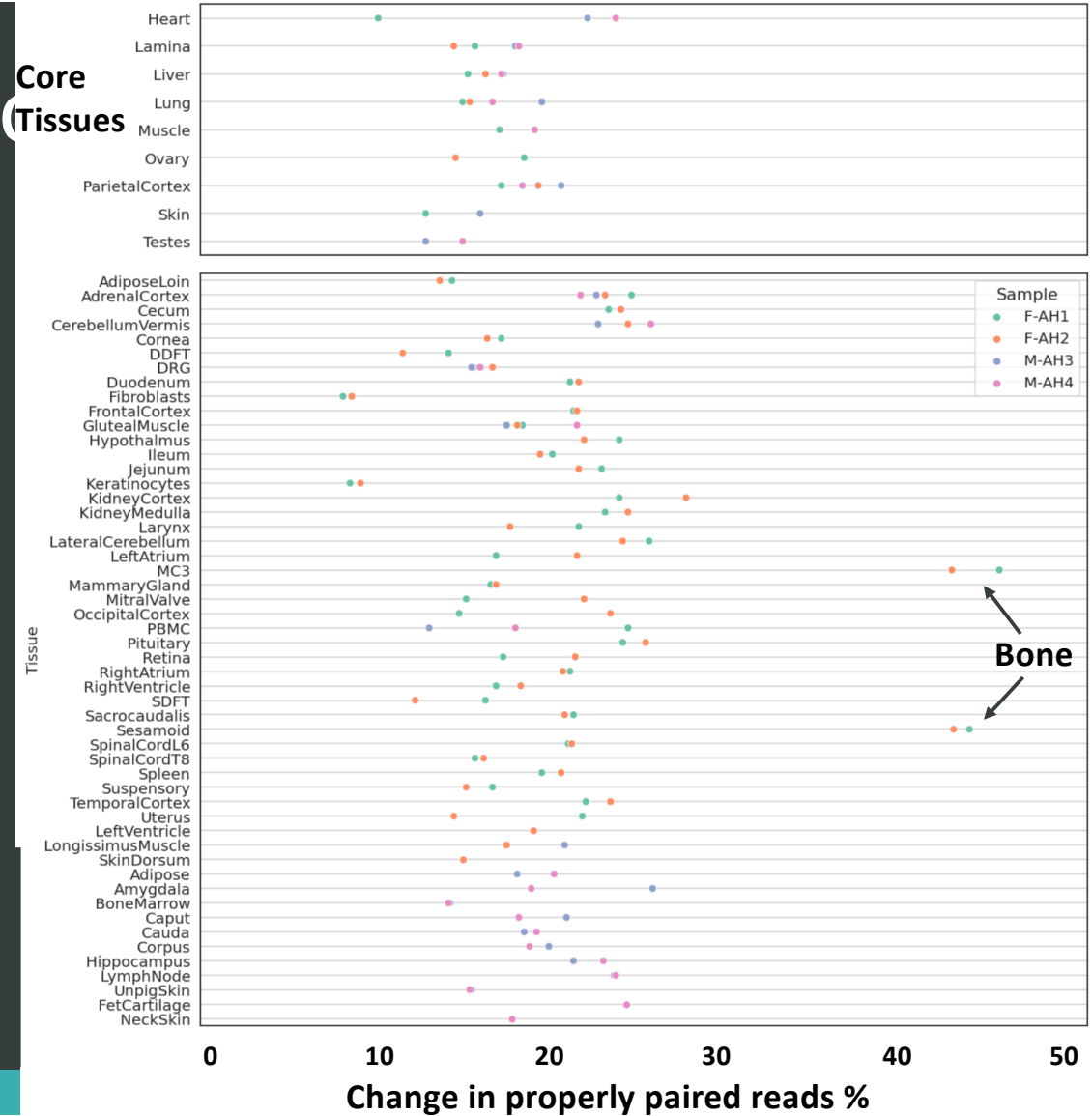


- 4 additional tissues
 - Spleen
 - Bone (MC III)
 - Bone (sesamoid)
 - Skin



Iso-seq | Core Tissues

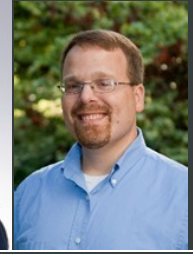
- 8 core tissues (mares) ✓
- Integrated FAANG transcriptome (RefSeq, Ensembl, FAANG)
- 8 core tissues (stallions) ✓





Additional FAANG Studies

- Karyotyping ✓
- Keratinocyte cell culture ✓
- Fibroblast cell culture ✓





Additional FAANG Studies

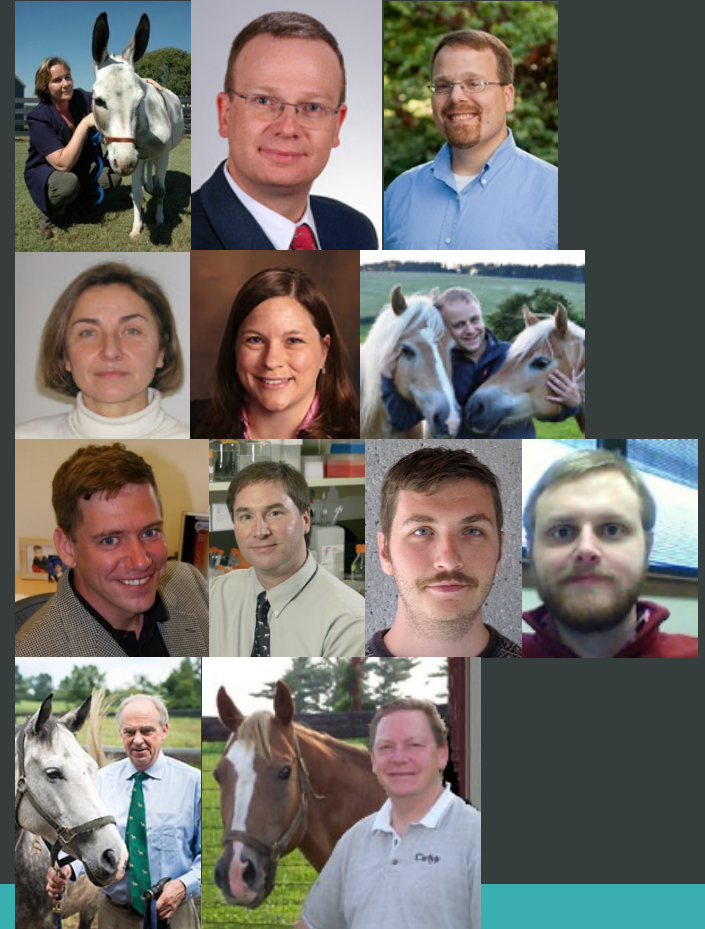
- Karyotyping ✓
- Keratinocyte cell culture ✓
- Fibroblast cell culture ✓
- Centromere Mapping ✓
- Microbiome Analysis ✓
- Tissue-specific mRNA/smRNA-seq comparison ✓













Additional FAANG Studies

- Karyotyping ✓
- Keratinocyte cell culture ✓
- Fibroblast cell culture ✓
- Centromere Mapping ✓
- Microbiome Analysis ✓
- Tissue-specific mRNA/smRNA-seq comparison ✓
- Further annotation of EquCab3.0 ✓
- ChRO-seq ✓





Additional FAANG Studies

- Karyotyping 
- Keratinocyte cell culture 
- Fibroblast cell culture 
- Centromere Mapping 
- Microbiome Analysis 
- Tissue-specific mRNA/smRNA-seq comparison 
- Further annotation of EquCab3.0 
- ChRO-seq 

COMMUNICATIONS BIOLOGY

ARTICLE

DOI: 10.1038/s42003-018-0199-z [OPEN](#)

Improved reference genome for the domestic horse increases assembly contiguity and composition

Theodore S. Kalbfleisch¹, Edward S. Rice², Michael S. DePriest Jr.¹, Brian P. Walenz³, Matthew S. Hestand⁴, Joris R. Vermeesch⁴, Brendan L. O'Connell^{2,16}, Ian T. Fiddes^{2,5}, Alisa O. Vershinina⁴, Nedda F. Sarem², Jessica L. Petersen⁷, Carrie J. Finno⁸, Rebecca R. Bellone^{8,9}, Molly E. McCue¹⁰, Samantha A. Brooks¹¹, Ernest Bailey¹², Ludovic Orlando^{13,14}, Richard E. Green², Donald C. Miller¹⁵, Douglas F. Antczak¹⁵ & James N. MacLeod¹²



HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2022 September 10.

Published in final edited form as:

Nat Genet. 2022 March ; 54(3): 295–305. doi:10.1038/s41588-022-01026-x.

Prediction of histone post-translational modification patterns based on nascent transcription data

Zhong Wang^{1,2*}, Alexandra G. Chivu^{1,3,*}, Lauren A. Choate¹, Edward J. Rice¹, Donald C. Miller¹, Tinyi Chu¹, Shao-Pei Chou¹, Nicole B. Kingsley², Jessica L. Petersen⁸, Carrie J. Finno⁷, Rebecca R. Bellone⁹, Douglas F. Antczak¹, John T. Lis³, Charles G. Danko^{1,4,*}

¹Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA.



Data Availability

UCSC genome browser track hub for equine FAANG dataset

The screenshot shows the GitHub interface for the repository 'FinnoLab / FAANGtracks'. The repository is public and has 2 issues, 2 branches, and 0 tags. The main branch is selected. The repository contains several files and folders:

File/Folder	Description	Last Commit
data	add raw data	20 days ago
hubDirectory	turn off auto scaling for RNA tracks	4 days ago
scripts	INteraction file	4 days ago
tools	fix: human readable gene names now showing up	last month
LICENSE	Initial commit	last month
README.md	Update README.md	4 minutes ago

The README.md file is displayed below, containing the following text:

FAANGtracks

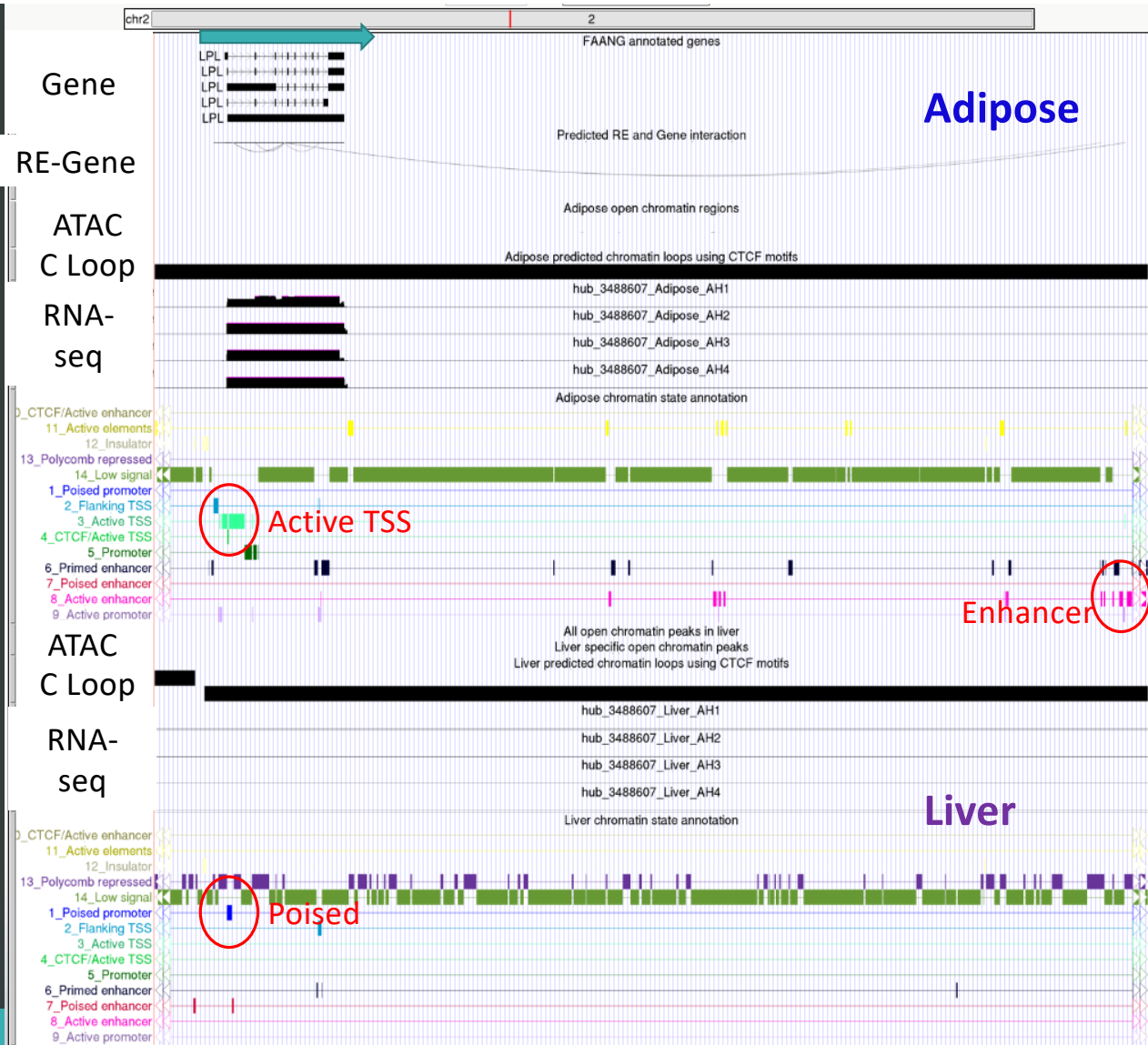
A collection of UCSC tracks based on data generated by the equine FAANG project

Use this link to view tracks in UCSC genome browser:
<https://genome.ucsc.edu/s/cjfinno/equCab3>

Data Availability

Lipoprotein lipase (LPL)

- Expressed adipose
- Poised promoter liver

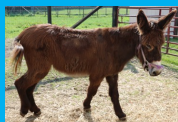


Where are we going from here?

- Equine Pan genome
- Equid T2T project
- Continued incremental funding and contributions
- USDA Animal Genome Blueprint project

Pan Genome Efforts

Donkey
sire



Tbred
mare

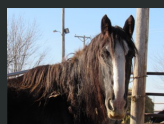


Mule F1

Arabian



Shire



F1

Tbred
sire



Donkey
mare



Hinny
F1

Grevy's
Zebra sire



Quarter
horse mare



Zorse F1

Scaffold N50s

Donkey:	92,846,730	Arabian:	80,722,027
Thoroughbred:	84,665,427	Shire:	83,119,518

Scaffold N50s

Donkey:	96,591,109	Quarter horse:	?
Thoroughbred:	87,715,332	Zebra:	?

T2T Effort

Donkey
sire



Tbred
mare



Mule F1

Data In hand:

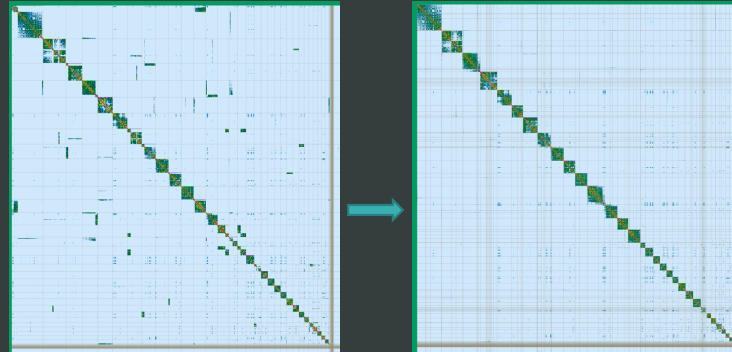
Mule HiFi: 46.94X
Mule Illumina: 64.12X
Arima HiC: 39.97X
Dam Illumina: 50.02X
Sire Illumina: 50.28X

Assembly Statistics:

Donkey Haplotype
ContigN50: 31.35MB

Assembly Statistics:

Thoroughbred Haplotype
ContigN50: 40.80MB
Salsa2 ScaffoldN50: 50.633Mb Salsa2 ScaffoldN50: 66.42Mb



Currently In Process
ONT long read data:
Adam Phillippy,
Alice Young,
and Shelise Brooks NHGRI

Data re-use: Containerized Workflows

- Snakemake workflows in a singularity container
 - Can run within slurm
- Short read mapping and variant calling
- RNA-Seq mapping and quantification

Solution: Real time access to processed data sets

Many online tools can use links to indexed genomic data (bam,gtf,vcf, BigWig files directly)

Animal	Tissue	Data Type	Bam URL	Bam Index
AH1	AH1 whole tissue	WGS	bamFile	bamIndex
AH2	AH2 whole tissue	WGS	bamFile	bamIndex
AH3	AH3 whole tissue	WGS	bamFile	bamIndex
AH4	AH4 whole tissue	WGS	bamFile	bamIndex

Command line tools like Samtools

IGV

UCSC
Genome
Browser

NCBI
Genome
Browser

Phenotype Data

- Grayson Jockey Club: Thoroughbreds
- American Quarter Horse Association: Quarter Horses
- Disease Phenotypes:
 - Idiopathic abortions
 - Various conditions identified by horse owners, veterinarians, and breed associations
 - Megaesophagus
 - Parrot Mouth
 - Bone Striation

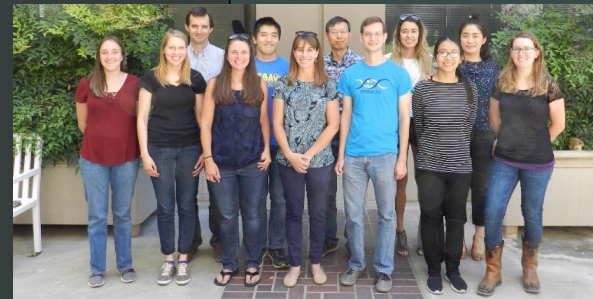
Hands-on Demo of Containerized
Workflows and a Data Repository
to Support Efficient,
Standardized, Interoperable
Genomic Analysis: An Equine
Example

Pacific H-1 (2nd Floor)
Monday 6:20-8:30

Acknowledgements



- USDA A1201 Grant (2019-67015-29340, Project Accession 1018854)
- Grayson Jockey Club Foundation
- USDA NRSP-8 Species Coordinator Funds
- UC Davis Center for Equine Health Funding



Acknowledgements

