

Review

Decoding the Equine Genome: Lessons from ENCODE

Sichong Peng ¹, Jessica L. Petersen ² , Rebecca R. Bellone ^{1,3} , Ted Kalbfleisch ⁴, N. B. Kingsley ^{1,3} ,
Alexa M. Barber ², Eleonora Cappelletti ⁵, Elena Giulotto ⁵  and Carrie J. Finno ^{1,*}

- ¹ Department of Population Health and Reproduction, School of Veterinary Medicine, University of California-Davis, Davis, CA 95616, USA; scpeng@ucdavis.edu (S.P.); rbellone@ucdavis.edu (R.R.B.); nbkingsley@ucdavis.edu (N.B.K.)
- ² Department of Animal Science, University of Nebraska, Lincoln, NE 68583-0908, USA; jessica.petersen@unl.edu (J.L.P.); alexa.barber@huskers.unl.edu (A.M.B.)
- ³ Veterinary Genetics Laboratory, School of Veterinary Medicine, University of California, Davis, CA 95616, USA
- ⁴ Department of Veterinary Science, Gluck Equine Research Center, University of Kentucky, Lexington, KY 40503, USA; ted.kalbfleisch@uky.edu
- ⁵ Department of Biology and Biotechnology "L. Spallanzani", University of Pavia, 27100 Pavia, Italy; eleonora.cappelletti@unipv.it (E.C.); elena.giulotto@unipv.it (E.G.)
- * Correspondence: cjfinno@gmail.com

Abstract: The horse reference genome assemblies, EquCab2.0 and EquCab3.0, have enabled great advancements in the equine genomics field, from tools to novel discoveries. However, significant gaps of knowledge regarding genome function remain, hindering the study of complex traits in horses. In an effort to address these gaps and with inspiration from the Encyclopedia of DNA Elements (ENCODE) project, the equine Functional Annotation of Animal Genome (FAANG) initiative was proposed to bridge the gap between genome and gene expression, providing further insights into functional regulation within the horse genome. Three years after launching the initiative, the equine FAANG group has generated data from more than 400 experiments using over 50 tissues, targeting a variety of regulatory features of the equine genome. In this review, we examine how valuable lessons learned from the ENCODE project informed our decisions in the equine FAANG project. We report the current state of the equine FAANG project and discuss how FAANG can serve as a template for future expansion of functional annotation in the equine genome and be used as a reference for studies of complex traits in horse. A well-annotated reference functional atlas will also help advance equine genetics in the pan-genome and precision medicine era.



Citation: Peng, S.; Petersen, J.L.; Bellone, R.R.; Kalbfleisch, T.; Kingsley, N.B.; Barber, A.M.; Cappelletti, E.; Giulotto, E.; Finno, C.J. Decoding the Equine Genome: Lessons from ENCODE. *Genes* **2021**, *12*, 1707. <https://doi.org/10.3390/genes12111707>

Academic Editor: Eric Barrey

Received: 30 September 2021

Accepted: 26 October 2021

Published: 27 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: FAANG; gene regulation; horse; functional annotation; transcriptome; epigenetics; welfare; health

1. The Horse Genome

The horse reference genomes (EquCab2.0 [1] and EquCab3.0 [2]) are based on a Thoroughbred mare *Twilight* and remain the only high-quality genome assemblies for equids. EquCab2.0 has 42,304 gaps comprising 55 Mb (2.2% of the genome) in total, with a scaffold N50 of 46 Mb. In comparison, EquCab3.0 contains 3771 gaps comprising 9 Mb (0.34% of the genome) with a scaffold N50 of 86 Mb. It has 99.7% mammalian Benchmarking Universal Single-Copy Orthologs (BUSCO) (5 fragmented and 7 missing out of 4104 mammalian universal orthologs), compared to that of 99.0% (4064 complete orthologs) in EquCab2.0 [2]. Owing to the availability of a high-quality reference genome sequence, researchers have been able to utilize a wide variety of high-throughput tools to interrogate genetic etiologies for various equine traits. Recently, Raudsepp et al. provided a comprehensive review of major discoveries using combinations of recent technologies including genome-wide association studies (GWAS), whole-genome sequencing (WGS), and RNA-seq [3].

Using these tools, successful identification of the genetic variants responsible for simple Mendelian traits have been identified, including a novel variant in glutamate metabotropic receptor 6 (*GRM6*) associated with congenital stationary night blindness [4] and a nonsense variant in rap guanine nucleotide exchange factor 5 (*RAPGEF5*) associated with equine familial isolated hypoparathyroidism [5]. However, many GWA studies conducted in horses have identified significant regions of association that do not contain any known genes. In humans, it was estimated that 88% of trait/disease associated single nucleotide polymorphisms (SNPs) identified from GWAS were either intergenic or intronic [6]. These SNPs would later be recognized as enriched in various functional elements [7]. Since then, numerous studies have examined different mechanisms by which noncoding variants may affect phenotype. Variants near these significantly associated SNPs have been found to create transcription factor (TF) binding sites [8], disrupt binding motifs [9], or alter TF binding affinities [10,11].

These findings support the notion that many noncoding regions of DNA have important regulatory functions that affect gene expression. With a comprehensive registry of 926,535 human regulatory elements [12], it is now common to include functional annotation in the fine mapping of traits post-GWAS [13]. However, no such resources are available for most animal species, including horses. To address this critical gap in knowledge, FAANG was proposed as an effort to identify important regulatory elements in the major livestock species [14].

2. Functional Annotation of Animal Genomes

The ENCODE initiative was proposed in 2003 as an ambitious effort to “identify all functional elements in the human genome sequence” [15]. In 2017, ENCODE concluded its third phase, delivering an integrated set of DNA transcription, regulation, and epigenetic modifications from a total of 7495 experiments in more than 500 cell types and tissues [12].

After almost two decades, ENCODE improved our understanding of gene regulation and delivered a wide range of computational tools, as well as a rich deposit of well-documented, publicly available experimental datasets [12]. Inspired by its phenomenal success, an international group of researchers proposed a similar, coordinated effort to systematically annotate animal genomes, providing vital resources to animal genetics research communities, termed Functional Annotation of Animal Genomes (FAANG) [14]. As part of the FAANG initiative, the equine FAANG group has been actively working with the larger FAANG community and ENCODE researchers to lead the annotation efforts for the horse genome.

The first stage of the equine FAANG initiative was to generate a biobank of reference tissues from comprehensively phenotyped animals. Burns et al. [16] and Donnelly et al. [17] detailed the phenotyping of four selected reference animals (UCD_AH1 – UCD_AH4) and a collection of over 80 tissues from each individual. These healthy animals were selected from the same breed (Thoroughbred) as *Twilight*, the horse used to construct the equine reference genome. When considering selection for the FAANG horses, the priority was placed on representing healthy Thoroughbred horses. Because *Twilight* was selected for the equine reference sequence based on homozygosity across the equine leukocyte antigen (ELA) region [1], the decision was made to include three unrelated Thoroughbreds and one (AH4) half-sibling of *Twilight* to achieve this goal while still aligning well with the reference sequence. A unique aspect of this biobank is that horses were extensively phenotyped, both antemortem by experienced veterinarians and postmortem by veterinary pathologists. This not only ensured that there was no evidence of clinical or subclinical disease in these animals, but it also provided insight into the cellular composition of the tissues selected for assays. These tissues are stored at -80°C in a biobank at UC Davis and are available to all equine FAANG researchers.

Here, we briefly discuss some of the most relevant findings from ENCODE and their implications for functionally annotating the equine genome.

3. Transcriptome

The transcriptome is the collection of all transcripts in an organism. It includes protein-coding mRNAs as well as noncoding RNAs. During the second phase of ENCODE, 62% of the human genome was found to be transcribed with 31% of transcribed bases located in intergenic regions [18]. Many of these transcripts have been recognized as noncoding RNAs with important regulatory roles [19–23]. Additionally, in any cell line, 39% of the genome was transcribed on average. Up to 56.7% of transcriptome was detected in at least one of fifteen studied cell lines. Interestingly, only 7% of protein-coding genes were cell-line specific, while 53% were constitutive. In comparison, long-noncoding RNAs (lncRNAs) appeared to contribute more to cell-line specificity, with 29% of lncRNAs detected in only one of the fifteen studied cell lines and 10% expressed in all cell lines [18]. These results highlighted the necessity of characterizing transcriptome in a cell-specific manner.

As part of ENCODE, GENCODE was initially founded to provide high-quality reference gene annotation for the human genome and subsequently expanded into a long-running partnership between several groups and institutes. In its most recent release based on GRCh38, a total of 60,649 genes have been identified in the human genome, of which 19,955 are protein coding, with an average isoform-to-gene ratio of 3.9 [24]. It was also demonstrated that genes tend to express many isoforms simultaneously, with a dominant isoform comprising 30% or more of its corresponding gene expression. Isoforms also appeared to contribute to cell type specificity, with over 75% of protein-coding genes having different dominant isoforms in different cell lines [18].

In addition to protein-coding transcripts, the transcriptome also consists of many noncoding RNA species, including both small and long noncoding RNAs. The functions of these RNAs have been extensively examined and implicated in important biological pathways [25–28]. The small noncoding RNAs present a unique opportunity to new therapeutic approaches [29]. Extensive efforts have been put into cataloguing noncoding RNAs in the human and mouse genome [30,31]. These efforts have further detailed the extent of noncoding RNA regulatory network and the diversity of noncoding RNA species and their functions.

Taken together, these findings from ENCODE demonstrated the importance of noncoding RNAs and of alternative splicing in cell-specific expression and regulation. Both Ensembl [32] and RefSeq [33] provide noncoding RNA and isoform annotation for EquCab3.0 by utilizing the high-quality annotation of the human genome as well as publicly available horse RNA-seq data. RefSeq annotation for EquCab3.0 consists of 30,022 genes, of which 21,129 are protein coding, with an average isoform-to-gene ratio of 2.6 [34]. The Ensembl annotation of the equine genome contains 30,371 genes (20,955 protein coding) with an average isoform-to-gene ratio of 1.9 [35]. Assuming the human and equine genomes have a similar number of genes and consistent isoform-to-gene ratio, the current horse gene annotation likely lacks many noncoding RNAs and alternate isoforms.

The FAANG initiative proposed RNA-seq assays for both mRNA and smRNA to identify and quantify these transcripts in a tissue-specific manner [14]. These assays have been performed for eight prioritized tissues (liver, lamina, heart, parietal cortex, adipose, skeletal muscle, ovary/testis, and lung) (Table 1).

To facilitate data generation for the remaining biobanked tissues, we proposed a unique “Adopt-A-Tissue” model for mRNA-seq. Researchers were invited to “adopt” a tissue or tissues fitting their research interests, which meant they would cover the assay and sequencing costs. All library preparations and sequencing were performed at the same two locations (female samples at UC Davis, male samples at University of Nebraska-Lincoln) to minimize variability. This approach allowed the community to contribute to the initiative together while still being able to limit technical variations across laboratories during library constructions [36]. Owing to this unique strategy, the equine community has sequenced over 40 tissues, and the data have been made publicly available (Table 1).

Table 1. Overview of Available Data and Assay Details.

Project Accession	Assay	Samples	Tissues	Instrument	Library Layout	Number of Experiments
PRJEB26698	WGS	Two females	1	HiSeq 2500 (San Diego, CA, USA)	2 × 250 bp	2
PRJEB42407	WGS	Two males	1	NovaSeq 6000 (San Diego, CA, USA)	2 × 150 bp	2
PRJEB26787	RNA-seq	Two females	30	HiSeq 2500 (San Diego, CA, USA)	2 × 250 bp	60
PRJEB32645	RRBS	Two females	10	HiScanSQ (San Diego, CA, USA)	1 × 50 bp	20
PRJEB35307	Histone ChIP-seq	Two females	8	HiSeq 4000 (San Diego, CA, USA)	1 × 50 bp	80
PRJEB42315	Histone ChIP-seq	Two females	4	HiSeq 4000 (San Diego, CA, USA)	1 × 50 bp	38
PRJEB41079	CTCF ChIP-seq	Two females	8	HiSeq 4000 (San Diego, CA, USA)	1 × 50 bp	28
PRJEB41317	ATAC-seq pilot	Two females	2	HiSeq 4000/NextSeq 500 (San Diego, CA, USA)	2 × 75 bp/2 × 42 bp	16

WGS: whole-genome sequencing; RNA-seq: mRNA sequencing; RRBS: reduced-representation bisulfite sequencing; Histone ChIP-seq: chromatin immunoprecipitation using sequencing for the four major histone marks; CTCF ChIP-seq: chromatin immunoprecipitation using sequencing for CTCF protein; ATAC-seq pilot: assay for transposase accessibility using sequencing.

More recently, long-read sequencing assays such as PacBio Isoform sequencing (Iso-seq) have emerged as powerful tools to determine the splicing patterns of transcripts. To address the poor isoform annotations currently available for the horse genome, Iso-seq assays are being performed in 8 tissues (liver, lung, lamina, heart, ovary, testis, muscle, skin, and parietal cortex) across eight PacBio Sequel 8M SMRT cells. By combining a wide variety of assays, the equine FAANG initiative aims to deliver a comprehensively annotated transcriptome for the horse genome.

4. Chromatin Accessibility

In mammalian cells, DNA molecules are packed by histone proteins to form nucleosomes and are subsequently compacted into chromatin [37,38]. Compact chromatin restricts access to DNA molecules by transcription factors and serves as a way to regulate gene expression [39]. For example, nucleosomes are densely arranged in facultative and constitutive heterochromatin while depleted in active regions such as active enhancers, insulators, and transcribed gene bodies [40,41]. Using DNase-seq, a DNase I assay quantifying susceptibility of chromatin to DNase I, Boyle et al. identified 94,925 DNase I hypersensitive sites (DHS) covering 2.1% of the human genome [42]. It was also found that only 13% of DHS were located within promoters, while up to 78% were in intergenic or intronic regions. Remarkably, DHS were found in or near the transcription start sites (TSS) of nearly all highly expressed genes. However, while DNase I hypersensitivity appeared to be necessary for gene expression, it was not sufficient as DHS were also observed in unexpressed genes [42]. The association between accessible chromatin and active elements present a unique opportunity to study tissue- and cell-specific gene regulation [43–47].

Echoing their strong functional implications, accessible chromatin was also shown to be associated with noncoding variants identified in GWAS studies of common traits. Maurano et al. examined 5654 noncoding variants identified in the GWAS studies of 207 diseases and 447 quantitative traits and found 76.6% of these variants lie either within a DHS or in complete linkage disequilibrium (LD) with another SNP in DHS [48]. The data further demonstrated that many of these DHS were strongly correlated with the promoter of a distal gene target [48]. Gusev et al. analyzed the heritability of 11 common diseases and found that SNPs contained within DHS explained up to 79% of heritability [49]. The strong association between accessible chromatin and functional elements warranted efforts to establish a catalog of tissue-specific DHS to facilitate discoveries of functionally relevant variants [47].

Although DNase-seq has proven successful in identifying accessible chromatin, its laborious protocol, slow turn-around time, and large sample size requirements severely limit large-scale applications [50,51]. Buenrostro et al. developed Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq), which greatly reduced both time and labor costs while requiring lower nuclei input [51]. Owing to its simple protocol and comparable output [52], ATAC-seq has been widely adopted as a state-of-the-art method for interrogating genome-wide chromatin accessibility; further, several variations in methodology have been developed to apply ATAC-seq to frozen tissues [53], cryopreserved nuclei [54], or to improve sensitivity in low-input materials [55].

Using ATAC-seq on cryopreserved nuclei from eight tissues across pig, cattle, and mouse, Halstead et al. showed a lack of conservation of sequence and accessibility in accessible sites across evolutionary distance, with 20% shared sites between pig and cattle and only 10% between mouse and ungulates [56]. Therefore, it is necessary to establish a tissue-specific catalog of accessible sites specifically for the horse genome. A pilot study was recently carried out to evaluate the suitability of frozen equine tissue derived nuclei for ATAC-seq [57]. Following protocols established by this study, additional ATAC-seq experiments are underway to expand this assay to eight prioritized tissues for the equine FAANG project.

5. Histone Modifications

Histone proteins form the basic building blocks of hierarchical chromatin structures and have been recognized to play an important role in modulating gene expression through post-transcriptional modifications [58–61]. A nucleosome core is formed by two copies of each of the four major types of histone proteins: H2A, H2B, H3, and H4 [62]. Since Allfrey first suggested the potential role of histone acetylation in regulating gene expression in 1964 [58], extensive research has been carried out to understand the roles, mechanisms, and implications of different histone modifications. Histone 3 lysine 4 monomethylation (H3K4me1), H3K4me3, H3K27me3, and H3K27ac are among some of the most studied and best understood modifications. Hyun et al. provided a detailed review of molecular mechanisms associated with histone lysine modifications and their regulatory functions [63]. Here, we briefly discuss ENCODE findings regarding histone marks and how they can be integrated to provide a more comprehensive view of regulatory activities.

Barski et al. first comprehensively assayed histone modifications across the human genome using high-throughput sequencing [64]. Consistent with previous studies, H3K4 methylation marks were enriched in promoter regions. A significant drop in signal between –200 bp and +50 bp of TSS was observed for H3K4me3 with major peaks at –300 bp and +100 bp [64]. This was consistent with observations that H3K4me3 was primarily associated with promoter regions [65] and that nucleosomes were depleted near active TSS [40]. On the other hand, H3K4me1 showed a distinct bimodal signal with peaks around –900 bp and +1000 bp of TSS [64], in agreement with previous observations that H3K4me1 was enriched in enhancer regions [66]. Similarly, H3K27me3 was observed at a higher level around the TSS of silent genes than those around active genes, supporting correlation between H3K27me3 and gene repression [67]. Conversely, H3K27ac was observed around active elements and associated with higher expression level [68].

Taken together, the four histone modifications discussed in this manuscript represent major regulatory elements and can provide valuable information regarding tissue-specific regulatory activities in the horse genome. Using genome-wide chromatin immunoprecipitation sequencing (ChIP-seq) for these four marks in eight prioritized tissues in the two female FAANG horses, Kingsley et al. reported over one million putative regulatory sites [69]. The utility of these data were demonstrated when a 16 kb intergenic deletion associated with an ocular condition in horses, namely distichiasis, was discovered and FAANG ChIP-seq data showed that this region harbors a tissue specific active enhancer [70]. Undoubtedly, these data will continue to aid in the understanding of other structural variants causing or associated with disease in the horse as additional tissues are evaluated.

Following the success of the mRNA Adopt-A-Tissue initiative, similar efforts have facilitated characterization of histone marks in four tissues important to equine health and traits of economic impact (spleen, metacarpal 3, sesamoid, and skin) [71]. Furthermore, additional Adopt-A-Tissue efforts are currently ongoing to facilitate histone ChIP-seq assays for the remaining FAANG tissues.

6. CTCF Binding

CCCTC-binding factor (CTCF) is a well-studied zinc finger protein that serves a central role in the formation of chromatin topology and remodeling. It was first discovered as a repressive transcription factor in chicken for *c-MYC* [72] as well as *LYZ* [73]. It was later shown that CTCF may also serve as an activator for the Amyloid β -Protein Precursor gene (*APP*) [74]. In 1999, Bell et al. reported a CTCF binding site at the core of an insulator element at the 5' end of the chicken β -globin gene *HBB* [75]. Insulators are genomic regions that separate genes from cis-regulatory elements [76]. This site also sits at a boundary between active and inactive chromatin [77], a typical feature of an insulator element [78,79].

Many seemingly contradictory functions of CTCF have attracted extensive efforts to understand the mechanisms of its multivalent roles. CTCF is highly conserved across species [80,81] and embryonically lethal when knocked out in mice [82]. The binding motif of CTCF consists of a ~20 bp core consensus sequence and less conserved peripheral sequences, comprising ~50 bp [83,84]. ChIP assays targeting CTCF revealed several unique patterns. First, CTCF binding sites were observed across the genome, with over 40% within intergenic regions [64,83,85]. Consistent with the insulator activity of CTCF, two distinct types of loci with opposing CTCF binding patterns were observed. Loci depleted of CTCF binding sites tend to include clusters of related gene families and transcriptionally coregulated genes, while loci enriched in CTCF binding sites tend to have genes with alternative promoters [83]. Furthermore, CTCF was shown to be crucial for chromatin loop formation at the mouse β -globin locus [86]. Similarly, Hou et al. described an alternative loop formation by inserting a CTCF binding insulator HS5 between the β -globin locus and its upstream locus control region [87]. Additionally, cohesin has been functionally associated with CTCF in mediating chromatin loops [88,89]. These results suggested a potential mechanism via which CTCF mediates regulation of chromatin conformation and gene expression.

The introduction of Hi-C technology that enabled genome-wide interrogation of long-range interactions [90] quickly brought about new insights into the mechanisms of CTCF function. Refining the resolution of the Hi-C interaction maps to kilobases, Rao et al. observed that the majority of chromatin loops were associated with convergent pairs of CTCF motifs, as well as colocalizing with cohesin proteins [91]. The orientation of CTCF motifs was also shown to determine the directionality of the CTCF mediated interactions [92]. Finally, the significance of such directionality was functionally demonstrated by inverting CTCF sites with CRISPR to alter genome topology as well as promoter function [93].

These findings led to a proposed extrusion model [94,95], where a chromatin loop is pulled through an extrusion complex consisting of cohesin and CTCF and is stabilized by a CTCF dimer. This model explains the convergence of a CTCF pair surrounding a chromatin loop, as well as the many regulatory functions of CTCF observed in early studies. More evidence is emerging in support of this model. Based on this model, Fudenberg et al. used simulation to reproduce topologically associated domains (TADs) and contact frequencies observed in Hi-C studies as well as to recapitulate experimental results where TADs were observed to spread upon depletion of CTCF binding sites [96]. Haarhuis et al. showed that cohesin release factor WAPL could restrict chromatin loop extrusion by releasing cohesin from DNA and that knocking out *WAPL* results in enlarged chromatin loops between incorrectly orientated CTCF motifs [97]. Allahyar et al., employing a multi-contact 4C technology, showed that such enlarged loops in *WAPL* knockout cells are a result of aggregated CTCF loop anchors, or a “cohesin traffic jam” [98].

Given its central role in chromatin loop formation, CTCF binding sites can be considered an intermediate between the 1D genomic sequence and 3D chromatin topology. Although there is no simple rule to determine the functional outcome of a disrupted CTCF binding site, as it largely depends on its interaction with surrounding regulatory elements, there is no doubt that a catalog of CTCF binding sites in a given cellular context can provide valuable information when decoding the functional implications of DNA variants.

Following the practices established by the FAANG community, characterization of CTCF binding sites using ChIP-seq is being performed on eight prioritized tissues for both sexes. Analyses to identify both tissue and sex-specific CTCF binding and integrate all of the FAANG ChIP-seq data into chromatin state annotations are currently underway.

7. Chromatin States

While the associations between individual histone marks and regulatory activities are noteworthy, combinations of histone marks have proven to be more reliable in the fine-scale predictions of regulatory elements. For example, Creighton et al. observed that the H3K27ac mark could distinguish active enhancers from inactive/poised enhancers, which are both marked by H3K4me1 [68]. Bernstein et al. similarly identified a bivalent signal with both H3K4 methylation and H3K27 methylation, suggesting a poised regulatory element [99]. These findings prompted hypotheses that various regulatory functions of noncoding DNA could be explained by either additive properties [100] or unique combinations of histone modifications [101]. New unsupervised computational approaches were subsequently developed to classify histone modification patterns and partition them into different chromatin states [102,103]. Ernst et al. identified 11 promoter states, all marked by H3K4me3 and varying presence and levels of several other marks, as well as 4 enhancer-associated states, all marked by H3K4me1 and varying frequencies of acetylation marks [103]. These findings suggest that some histone modifications (H3K4me1, H3K4me3) designate unique regulatory elements while other modifications (acetylation marks including H3K27ac) enhance regulatory activity in an additive fashion.

The recognition of chromatin states and introduction of computation tools such as ChromHMM [104] provided a way to systematically profile the regulatory landscape in any given cellular context. Taking advantage of this development and the availability of ChIP-seq data from the four major histone marks and CTCF, efforts to compose an integrated tissue-specific chromatin state map are currently underway for the equine genome.

8. Unique Aspects of the Horse Genome

Centromeres are enigmatic structures because, contrary to other genetic loci, their function is not determined by the underlying DNA sequence but depends on epigenetic factors. The Centromere Protein A (CENP-A) is a centromere-specific variant of histone H3 that epigenetically identifies, maintains, and propagates centromere function [105]. The characteristics of its binding domain have been elusive to investigators due to its typical association with tandemly repeated DNA (satellite DNA). In this context, a turning point was the discovery that the centromere of horse chromosome 11 (ECA11) was completely devoid of satellite DNA, demonstrating for the first time that a natural mammalian centromere, fixed in a species, can exist without satellite sequences [1]. Owing to the lack of satellite repeats at the centromere of ECA11 and the availability of the horse reference genome, the genomic position of the corresponding CENP-A binding domain could be precisely identified by ChIP-on-chip with an anti-CENP-A antibody [1]. Later, several satellite-less centromeres were identified by ChIP-seq in the donkey genome [106]. These peculiar centromeres found in equid species represent an immature stage of “centromerization”, being the result of centromere repositioning, which is the movement of the centromeric function without detectable chromosomal rearrangements. This event was exceptionally frequent during the rapid evolution of the genus *Equus* [107–109]. Such centromeres, being uncoupled from satellite DNA, provide a unique model for dissecting the molecular structure of the centromere [110].

The position of the ECA11 satellite-less centromere, identified as the CENP-A binding domain, is not fixed in the horse population but slides within an about the 500 kb region, giving rise to different positional alleles or “epialleles” [106,111,112]. The analysis of these epialleles carried out on families composed by horses, donkeys, and their hybrid offspring (mule/hinny) revealed that they are inherited as Mendelian traits, but their position can slide in one generation [106]. Conversely, the position of the centromere is stable during mitotic propagation of cultured cells grown for several population doublings, suggesting that the sliding may presumably take place during meiosis or early embryogenesis [106].

The absence of satellite DNA at these centromeres also provides a unique opportunity to understand whether some typical features of mammalian centromeres depend on the presence of satellite DNA. In particular, it was possible to demonstrate that satellite DNA was not necessary for segregation fidelity of the centromere [113] and was not implicated in the suppression of meiotic recombination, which is typically exerted by the centromere [112].

The rich repository of tissues from different developmental origins available through the FAANG project will allow us to answer other important questions on centromere biology using the ECA11 centromere as model system. We will test whether the centromere position is conserved during development or if it can slide during tissue differentiation. In addition, thanks to the large amount of data regarding the functional annotation of the horse genome, generated within the FAANG effort, we will be able to map the epigenetic marks available through the consortium in the ECA11 centromeric region. The results will indicate whether chromatin markers and transcriptional activity at ECA11 centromere vary across tissues and individuals, and with respect to centromere position. Furthermore, CENP-A has been shown to bind at TF binding sites and promoters, suggesting potential regulatory activities [114]. Therefore, utilizing FAANG data, we will be able to identify the regulatory activities of CENP-A and any roles centromeres may play during tissue differentiation

9. Summary and Future Perspectives

Just three years after starting the tissue and data collection for the equine FAANG initiative, the community has completed over 400 experiments from more than 50 tissues using a variety of assays targeting different features of the horse regulatory landscape (Table 2). Data are being made available to the public as they are generated and evaluated for passing quality control measures; these data have been and continue to be utilized in unrelated research projects [5,70,115]. Integrated analysis is currently ongoing to provide a systematic annotation of major functional elements in the horse genome available, as a central hub hosted on UCSC genome browser to the research community.

With over 80 tissues collected from four healthy and comprehensively phenotyped animals, we will be able to generate a map of gene expression and regulation throughout the horse body, providing unique opportunities to investigate tissue-specific gene expression and gene networks. However, this tissue collection presents a serious challenge for data analyses. Heterogeneity both within tissues as a result of cell-type differentiation and across tissues as a result of tissue infiltration or contamination during collection, can confound analysis of tissue-specific expression and regulation. The prevalence of this issue was recently reported by Sturm et al. [116]. To mitigate this issue, careful histological assessment was performed during the tissue collection phase to minimize the possibility of tissue infiltration or contamination. However, caution should be taken to assess the extent of tissue heterogeneity during data analysis. Additionally, single-cell based technologies have proven useful to profile cell types from complex tissues [117–120], and the adoption of these technologies to equine FAANG data are being discussed within the community and will likely be integrated in the next steps of the multi-phased approach of this project.

While the equine FAANG biobank represents a wide variety of tissue types, the four horses these tissues were collected from represent only a narrow subset of the horse population, as well as developmental stages. These horses were intentionally selected to be of the same breed as the reference genome assembly in order to better annotate the reference genome assembly. However, caution should be taken with interpretation and extrapolation

of these data to other breeds or developmental stages. Regardless, this initiative will serve as a template and reference point for the future expansion of the transcriptome and epigenome of equids.

FAANG represents a notable international collaborative effort in the equine community that has brought together equine researchers and practitioners from around the globe. Most importantly, FAANG collaborators have been vocal proponents of open science and broad data accessibility within the equine community. The growing number of publicly available datasets is accelerating discoveries and powering large-scale analyses. Well-annotated and carefully documented FAANG data with accompanying comprehensive metadata will serve as a reference point for many future discoveries in horse.

Table 2. Overview of Completed Assays.

Assay	Animals	Tissue Types	Total Experiments
WGS	AH1-AH4	Blood	4
mRNA-seq	AH1	47	140
	AH2	46	
	AH3	23	
	AH4	24	
Iso-seq	AH1-AH4	12	48
ChIP-seq-H3K4me1	AH1-AH2	12	40
	AH3-AH4	8	
ChIP-seq-H3K4me3	AH1-AH2	12	40
	AH3-AH4	8	
ChIP-seq-H3K27ac	AH1-AH2	12	40
	AH3-AH4	8	
ChIP-seq-H3K27me3	AH1-AH2	12	40
	AH3-AH4	8	
ChIP-seq-CTCF	AH1-AH2	8	32
	AH3-AH4	8	
ATAC-seq	AH1-AH4	10	40
RRBS	AH1-AH2	10	20
smRNA-seq	AH1-AH2	48	96
Total		48	444

Author Contributions: Conceptualization, S.P. and C.J.F.; writing—original draft preparation, S.P., E.G. and C.J.F.; writing—review and editing, J.L.P., R.R.B., T.K., N.B.K., A.M.B. and E.C.; supervision, C.J.F., J.L.P., R.R.B., T.K. and E.G. All authors have read and agreed to the published version of the manuscript.

Funding: Portions of this work were supported by Animal Breeding and Functional Annotation of Genomes (A1201) Grant 2019-67015-29340/Project Accession 1018854 from the USDA National Institute of Food and Agriculture, the Grayson Jockey Club Foundation, USDA NRSP-8 and the UC Davis Center for Equine Health, Italian Ministry of Education, University and Research (MIUR) [Dipartimenti di Eccellenza Program (2018–2022)—Dept. of Biology and Biotechnology “L. Spalanzani”, University of Pavia]. Support for C.J.F was provided by the National Institutes of Health (NIH) (L40 TR001136). None of the funding agencies had any role in the design of the study, analysis, interpretation of the data, or writing of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All FAANG data discussed in this manuscript can be accessed from Sequence Read Archive (SRA), European Nucleotide Archive (ENA), or faang.org/dataset using accession numbers listed in Table 1.

Acknowledgments: We would like to acknowledge UC Davis FAANG group for their input in experimental designs and data analyses as well as three anonymous reviewers for their suggestions and comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wade, C.M.; Giulotto, E.; Sigurdsson, S.; Zoli, M.; Gnerre, S.; Inslund, F.; Lear, T.L.; Adelson, D.L.; Bailey, E.; Bellone, R.R.; et al. Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. *Science* **2009**, *326*, 865–867. [[CrossRef](#)]
2. Kalbfleisch, T.S.; Rice, E.S.; DePriest, M.S.; Walenz, B.P.; Hestand, M.S.; Vermeesch, J.R.; O’Connell, B.L.; Fiddes, I.T.; Vershinina, A.O.; Saremi, N.F.; et al. Improved Reference Genome for the Domestic Horse Increases Assembly Contiguity and Composition. *Commun. Biol.* **2018**, *1*, 197. [[CrossRef](#)]
3. Raudsepp, T.; Finno, C.J.; Bellone, R.R.; Petersen, J.L. Ten Years of the Horse Reference Genome: Insights into Equine Biology, Domestication and Population Dynamics in the Post-genome Era. *Anim. Genet.* **2019**, *50*, 569–597. [[CrossRef](#)] [[PubMed](#)]
4. Hack, Y.L.; Crabtree, E.E.; Avila, F.; Sutton, R.B.; Grahn, R.; Oh, A.; Gilger, B.; Bellone, R.R. Whole-genome Sequencing Identifies Missense Mutation in *GRM6* as the Likely Cause of Congenital Stationary Night Blindness in a Tennessee Walking Horse. *Equine Vet. J.* **2021**, *53*, 316–323. [[CrossRef](#)]
5. Rivas, V.N.; Magdesian, K.G.; Fagan, S.; Slovis, N.M.; Luethy, D.; Javsicas, L.H.; Caserto, B.G.; Miller, A.D.; Dahlgren, A.R.; Peterson, J.; et al. A Nonsense Variant in Rap Guanine Nucleotide Exchange Factor 5 (RAPGEF5) Is Associated with Equine Familial Isolated Hypoparathyroidism in Thoroughbred Foals. *PLoS Genet.* **2020**, *16*, e1009028. [[CrossRef](#)] [[PubMed](#)]
6. Hindorff, L.A.; Sethupathy, P.; Junkins, H.A.; Ramos, E.M.; Mehta, J.P.; Collins, F.S.; Manolio, T.A. Potential Etiologic and Functional Implications of Genome-Wide Association Loci for Human Diseases and Traits. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9362–9367. [[CrossRef](#)]
7. The ENCODE Project Consortium An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **2012**, *489*, 57–74. [[CrossRef](#)]
8. Musunuru, K.; Strong, A.; Frank-Kamenetsky, M.; Lee, N.E.; Ahfeldt, T.; Sachs, K.V.; Li, X.; Li, H.; Kuperwasser, N.; Ruda, V.M.; et al. From Noncoding Variant to Phenotype via SORT1 at the 1p13 Cholesterol Locus. *Nature* **2010**, *466*, 714–719. [[CrossRef](#)]
9. Bauer, D.E.; Kamran, S.C.; Lessard, S.; Xu, J.; Fujiwara, Y.; Lin, C.; Shao, Z.; Canver, M.C.; Smith, E.C.; Pinello, L.; et al. An Erythroid Enhancer of BCL11A Subject to Genetic Variation Determines Fetal Hemoglobin Level. *Science* **2013**, *342*, 253–257. [[CrossRef](#)]
10. Tuupanen, S.; Turunen, M.; Lehtonen, R.; Hallikas, O.; Vanharanta, S.; Kivioja, T.; Björklund, M.; Wei, G.; Yan, J.; Niittymäki, I.; et al. The Common Colorectal Cancer Predisposition SNP Rs6983267 at Chromosome 8q24 Confers Potential to Enhanced Wnt Signaling. *Nat. Genet.* **2009**, *41*, 885–890. [[CrossRef](#)] [[PubMed](#)]
11. Wright, J.B.; Brown, S.J.; Cole, M.D. Upregulation of C- MYC in Cis through a Large Chromatin Loop Linked to a Cancer Risk-Associated Single-Nucleotide Polymorphism in Colorectal Cancer Cells. *Mol. Cell Biol.* **2010**, *30*, 1411–1420. [[CrossRef](#)] [[PubMed](#)]
12. The ENCODE Project Consortium; Moore, J.E.; Purcaro, M.J.; Pratt, H.E.; Epstein, C.B.; Shores, N.; Adrian, J.; Kawli, T.; Davis, C.A.; Dobin, A.; et al. Expanded Encyclopaedias of DNA Elements in the Human and Mouse Genomes. *Nature* **2020**, *583*, 699–710. [[CrossRef](#)]
13. Edwards, S.L.; Beesley, J.; French, J.D.; Dunning, A.M. Beyond GWASs: Illuminating the Dark Road from Association to Function. *Am. J. Hum. Genet.* **2013**, *93*, 779–797. [[CrossRef](#)] [[PubMed](#)]
14. Andersson, L.; Archibald, A.L.; Bottema, C.D.; Brauning, R.; Burgess, S.C.; Burt, D.W.; Casas, E.; Cheng, H.H.; Clarke, L.; Couldrey, C.; et al. Coordinated International Action to Accelerate Genome-to-Phenome with FAANG, the Functional Annotation of Animal Genomes Project. *Genome Biol.* **2015**, *16*, 57. [[CrossRef](#)] [[PubMed](#)]
15. Consortium, T.E.P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **2004**, *306*, 636–640. [[CrossRef](#)]
16. Burns, E.N.; Bordbari, M.H.; Mienaltowski, M.J.; Affolter, V.K.; Barro, M.V.; Gianino, F.; Gianino, G.; Giulotto, E.; Kalbfleisch, T.S.; Katzman, S.A.; et al. Generation of an Equine Biobank to Be Used for Functional Annotation of Animal Genomes Project. *Anim. Genet.* **2018**, *49*, 564–570. [[CrossRef](#)] [[PubMed](#)]
17. Donnelly, C.G.; Bellone, R.R.; Hales, E.N.; Nguyen, A.; Katzman, S.A.; Dujovne, G.A.; Knickelbein, K.E.; Avila, F.; Kalbfleisch, T.S.; Giulotto, E.; et al. Generation of a Biobank From Two Adult Thoroughbred Stallions for the Functional Annotation of Animal Genomes Initiative. *Front. Genet.* **2021**, *12*, 650305. [[CrossRef](#)]
18. Djebali, S.; Davis, C.A.; Merkel, A.; Dobin, A.; Lassmann, T.; Mortazavi, A.; Tanzer, A.; Lagarde, J.; Lin, W.; Schlesinger, F.; et al. Landscape of Transcription in Human Cells. *Nature* **2012**, *489*, 101–108. [[CrossRef](#)]
19. Schmitz, S.U.; Grote, P.; Herrmann, B.G. Mechanisms of Long Noncoding RNA Function in Development and Disease. *Cell. Mol. Life Sci.* **2016**, *73*, 2491–2509. [[CrossRef](#)] [[PubMed](#)]
20. Turner, M.; Galloway, A.; Vigorito, E. Noncoding RNA and Its Associated Proteins as Regulatory Elements of the Immune System. *Nat. Immunol.* **2014**, *15*, 484–491. [[CrossRef](#)] [[PubMed](#)]

21. Lin, N.; Chang, K.-Y.; Li, Z.; Gates, K.; Rana, Z.A.; Dang, J.; Zhang, D.; Han, T.; Yang, C.-S.; Cunningham, T.J.; et al. An Evolutionarily Conserved Long Noncoding RNA TUNA Controls Pluripotency and Neural Lineage Commitment. *Mol. Cell* **2014**, *53*, 1005–1019. [[CrossRef](#)] [[PubMed](#)]
22. Long, J.; Badal, S.S.; Ye, Z.; Wang, Y.; Ayanga, B.A.; Galvan, D.L.; Green, N.H.; Chang, B.H.; Overbeek, P.A.; Danesh, F.R. Long Noncoding RNA Tug1 Regulates Mitochondrial Bioenergetics in Diabetic Nephropathy. *J. Clin. Investig.* **2016**, *126*, 4205–4218. [[CrossRef](#)]
23. St. Laurent, G.; Wahlestedt, C.; Kapranov, P. The Landscape of Long Noncoding RNA Classification. *Trends Genet.* **2015**, *31*, 239–251. [[CrossRef](#)]
24. Frankish, A.; Diekhans, M.; Ferreira, A.-M.; Johnson, R.; Jungreis, I.; Loveland, J.; Mudge, J.M.; Sisu, C.; Wright, J.; Armstrong, J.; et al. GENCODE Reference Annotation for the Human and Mouse Genomes. *Nucleic Acids Res.* **2019**, *47*, D766–D773. [[CrossRef](#)] [[PubMed](#)]
25. Meller, V.H.; Joshi, S.S.; Deshpande, N. Modulation of Chromatin by Noncoding RNA. *Annu. Rev. Genet.* **2015**, *49*, 673–695. [[CrossRef](#)]
26. Shen, Y.; Huang, Z.; Yang, R.; Chen, Y.; Wang, Q.; Gao, L. Insights into Enhancer RNAs: Biogenesis and Emerging Role in Brain Diseases. *Neuroscientist* **2021**, 107385842110468. [[CrossRef](#)]
27. Moazzendizaji, S.; Sevbitov, A.; Ezzatifar, F.; Jalili, H.R.; Aalii, M.; Hemmatzadeh, M.; Aslani, S.; Gholizadeh Navashenaq, J.; Safari, R.; Hosseinzadeh, R.; et al. MicroRNAs: Small Molecules with a Large Impact on Colorectal Cancer. *Biotechnol. Appl. Biochem.* **2021**, bab.2255. [[CrossRef](#)] [[PubMed](#)]
28. Wen, Z.-J.; Xin, H.; Wang, Y.-C.; Liu, H.-W.; Gao, Y.-Y.; Zhang, Y.-F. Emerging Roles of CircRNAs in the Pathological Process of Myocardial Infarction. *Mol. Ther. Nucleic Acids* **2021**, S2162253121002456. [[CrossRef](#)]
29. Winkle, M.; El-Daly, S.M.; Fabbri, M.; Calin, G.A. Noncoding RNA Therapeutics—Challenges and Potential Solutions. *Nat. Rev. Drug. Discov.* **2021**, *20*, 629–651. [[CrossRef](#)]
30. He, P.; Williams, B.A.; Trout, D.; Marinov, G.K.; Amrhein, H.; Berghella, L.; Goh, S.-T.; Plajzer-Frick, I.; Afzal, V.; Pennacchio, L.A.; et al. The Changing Mouse Embryo Transcriptome at Whole Tissue and Single-Cell Resolution. *Nature* **2020**, *583*, 760–767. [[CrossRef](#)]
31. Lorenzi, L.; Chiu, H.-S.; Avila Cobos, F.; Gross, S.; Volders, P.-J.; Cannoodt, R.; Nuytens, J.; Vanderheyden, K.; Anckaert, J.; Lefever, S.; et al. The RNA Atlas Expands the Catalog of Human Non-Coding RNAs. *Nat. Biotechnol.* **2021**. [[CrossRef](#)]
32. Howe, K.L.; Achuthan, P.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; Bhai, J.; et al. Ensembl 2021. *Nucleic Acids Res.* **2021**, *49*, D884–D891. [[CrossRef](#)]
33. Pruitt, K.D.; Brown, G.R.; Hiatt, S.M.; Thibaud-Nissen, F.; Astashyn, A.; Ermolaeva, O.; Farrell, C.M.; Hart, J.; Landrum, M.J.; McGarvey, K.M.; et al. RefSeq: An Update on Mammalian Reference Sequences. *Nucl. Acids Res.* **2014**, *42*, D756–D763. [[CrossRef](#)] [[PubMed](#)]
34. Equus Caballus RefSeq Annotation Release 103. Available online: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Equus_caballus/103/ (accessed on 10 September 2021).
35. Equus Caballus Ensembl Annotation Release 104. Available online: https://uswest.ensembl.org/Equus_caballus/Info/Annotation (accessed on 10 September 2021).
36. McIntyre, L.M.; Lopiano, K.K.; Morse, A.M.; Amin, V.; Oberg, A.L.; Young, L.J.; Nuzhdin, S.V. RNA-Seq: Technical Variability and Sampling. *BMC Genom.* **2011**, *12*, 293. [[CrossRef](#)]
37. Kornberg, R.D. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **1974**, *184*, 868–871. [[CrossRef](#)] [[PubMed](#)]
38. Olins, D.E.; Olins, A.L. Chromatin History: Our View from the Bridge. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 809–814. [[CrossRef](#)] [[PubMed](#)]
39. Lorch, Y.; LaPointe, J.W.; Kornberg, R.D. Nucleosomes Inhibit the Initiation of Transcription but Allow Chain Elongation with the Displacement of Histones. *Cell* **1987**, *49*, 203–210. [[CrossRef](#)]
40. Lee, C.-K.; Shibata, Y.; Rao, B.; Strahl, B.D.; Lieb, J.D. Evidence for Nucleosome Depletion at Active Regulatory Regions Genome-Wide. *Nat. Genet.* **2004**, *36*, 900–905. [[CrossRef](#)]
41. Gaszner, M.; Felsenfeld, G. Insulators: Exploiting Transcriptional and Epigenetic Mechanisms. *Nat. Rev. Genet.* **2006**, *7*, 703–713. [[CrossRef](#)] [[PubMed](#)]
42. Boyle, A.P.; Davis, S.; Shulha, H.P.; Meltzer, P.; Margulies, E.H.; Weng, Z.; Furey, T.S.; Crawford, G.E. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* **2008**, *132*, 311–322. [[CrossRef](#)]
43. Stergachis, A.B.; Neph, S.; Reynolds, A.; Humbert, R.; Miller, B.; Paige, S.L.; Vernot, B.; Cheng, J.B.; Thurman, R.E.; Sandstrom, R.; et al. Developmental Fate and Cellular Maturity Encoded in Human Regulatory DNA Landscapes. *Cell* **2013**, *154*, 888–903. [[CrossRef](#)]
44. Song, L.; Zhang, Z.; Grasfeder, L.L.; Boyle, A.P.; Giresi, P.G.; Lee, B.-K.; Sheffield, N.C.; Gräf, S.; Huss, M.; Keefe, D.; et al. Open Chromatin Defined by DNaseI and FAIRE Identifies Regulatory Elements That Shape Cell-Type Identity. *Genome Res.* **2011**, *21*, 1757–1767. [[CrossRef](#)]
45. Natarajan, A.; Yardimci, G.G.; Sheffield, N.C.; Crawford, G.E.; Ohler, U. Predicting Cell-Type-Specific Gene Expression from Regions of Open Chromatin. *Genome Res.* **2012**, *22*, 1711–1722. [[CrossRef](#)]
46. Thurman, R.E.; Rynes, E.; Humbert, R.; Vierstra, J.; Maurano, M.T.; Haugen, E.; Sheffield, N.C.; Stergachis, A.B.; Wang, H.; Vernot, B.; et al. The Accessible Chromatin Landscape of the Human Genome. *Nature* **2012**, *489*, 75–82. [[CrossRef](#)] [[PubMed](#)]

47. Meuleman, W.; Muratov, A.; Rynes, E.; Halow, J.; Lee, K.; Bates, D.; Diegel, M.; Dunn, D.; Neri, F.; Teodosiadis, A.; et al. Index and Biological Spectrum of Human DNase I Hypersensitive Sites. *Nature* **2020**, *584*, 244–251. [[CrossRef](#)]
48. Maurano, M.T.; Humbert, R.; Rynes, E.; Thurman, R.E.; Haugen, E.; Wang, H.; Reynolds, A.P.; Sandstrom, R.; Qu, H.; Brody, J.; et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **2012**, *337*, 1190–1195. [[CrossRef](#)]
49. Gusev, A.; Lee, S.H.; Trynka, G.; Finucane, H.; Vilhjálmsson, B.J.; Xu, H.; Zang, C.; Ripke, S.; Bulik-Sullivan, B.; Stahl, E.; et al. Partitioning Heritability of Regulatory and Cell-Type-Specific Variants across 11 Common Diseases. *Am. J. Hum. Genet.* **2014**, *95*, 535–552. [[CrossRef](#)]
50. Crawford, G.E. Genome-Wide Mapping of DNase Hypersensitive Sites Using Massively Parallel Signature Sequencing (MPSS). *Genome Res.* **2005**, *16*, 123–131. [[CrossRef](#)] [[PubMed](#)]
51. Buenrostro, J.D.; Giresi, P.G.; Zaba, L.C.; Chang, H.Y.; Greenleaf, W.J. Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position. *Nat. Methods* **2013**, *10*, 1213–1218. [[CrossRef](#)]
52. Buenrostro, J.D.; Wu, B.; Chang, H.Y.; Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **2015**, *109*. [[CrossRef](#)] [[PubMed](#)]
53. Corces, M.R.; Trevino, A.E.; Hamilton, E.G.; Greenside, P.G.; Sinnott-Armstrong, N.A.; Vesuna, S.; Satpathy, A.T.; Rubin, A.J.; Montine, K.S.; Wu, B.; et al. An Improved ATAC-Seq Protocol Reduces Background and Enables Interrogation of Frozen Tissues. *Nat. Methods* **2017**, *14*, 959–962. [[CrossRef](#)] [[PubMed](#)]
54. Halstead, M.M.; Kern, C.; Saelao, P.; Chanthavixay, G.; Wang, Y.; Delany, M.E.; Zhou, H.; Ross, P.J. Systematic Alteration of ATAC-Seq for Profiling Open Chromatin in Cryopreserved Nuclei Preparations from Livestock Tissues. *Sci. Rep.* **2020**, *10*, 5230. [[CrossRef](#)] [[PubMed](#)]
55. Sos, B.C.; Fung, H.-L.; Gao, D.R.; Osothprarop, T.F.; Kia, A.; He, M.M.; Zhang, K. Characterization of Chromatin Accessibility with a Transposome Hypersensitive Sites Sequencing (THS-Seq) Assay. *Genome Biol.* **2016**, *17*, 20. [[CrossRef](#)] [[PubMed](#)]
56. Halstead, M.M.; Kern, C.; Saelao, P.; Wang, Y.; Chanthavixay, G.; Medrano, J.F.; Van Eenennaam, A.L.; Korf, I.; Tuggle, C.K.; Ernst, C.W.; et al. A Comparative Analysis of Chromatin Accessibility in Cattle, Pig, and Mouse Tissues. *BMC Genom.* **2020**, *21*, 698. [[CrossRef](#)] [[PubMed](#)]
57. Peng, S.; Bellone, R.; Petersen, J.L.; Kalbfleisch, T.S.; Finno, C.J. Successful ATAC-Seq From Snap-Frozen Equine Tissues. *Front. Genet.* **2021**, *12*, 641788. [[CrossRef](#)]
58. Allfrey, V.G.; Faulkner, R.; Mirsky, A.E. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. USA* **1964**, *51*, 786–794. [[CrossRef](#)]
59. Yang, X.-J.; Seto, E. HATs and HDACs: From Structure, Function and Regulation to Novel Strategies for Therapy and Prevention. *Oncogene* **2007**, *26*, 5310–5318. [[CrossRef](#)]
60. Oki, M.; Aihara, H.; Ito, T. Role of Histone Phosphorylation in Chromatin Dynamics and Its Implications in Diseases. *Subcell Biochem.* **2007**, *41*, 319–336. [[PubMed](#)]
61. Bedford, M.T.; Clarke, S.G. Protein Arginine Methylation in Mammals: Who, What, and Why. *Mol. Cell* **2009**, *33*, 1–13. [[CrossRef](#)]
62. Luger, K.; Mäder, A.W.; Richmond, R.K.; Sargent, D.F.; Richmond, T.J. Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution. *Nature* **1997**, *389*, 251–260. [[CrossRef](#)]
63. Hyun, K.; Jeon, J.; Park, K.; Kim, J. Writing, Erasing and Reading Histone Lysine Methylations. *Exp. Mol. Med.* **2017**, *49*, e324. [[CrossRef](#)]
64. Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.-Y.; Schones, D.E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **2007**, *129*, 823–837. [[CrossRef](#)]
65. Liang, G.; Lin, J.C.Y.; Wei, V.; Yoo, C.; Cheng, J.C.; Nguyen, C.T.; Weisenberger, D.J.; Egger, G.; Takai, D.; Gonzales, F.A.; et al. Distinct Localization of Histone H3 Acetylation and H3-K4 Methylation to the Transcription Start Sites in the Human Genome. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7357–7362. [[CrossRef](#)]
66. Heintzman, N.D.; Stuart, R.K.; Hon, G.; Fu, Y.; Ching, C.W.; Hawkins, R.D.; Barrera, L.O.; Van Calcar, S.; Qu, C.; Ching, K.A.; et al. Distinct and Predictive Chromatin Signatures of Transcriptional Promoters and Enhancers in the Human Genome. *Nat. Genet.* **2007**, *39*, 311–318. [[CrossRef](#)] [[PubMed](#)]
67. Boyer, L.A.; Plath, K.; Zeitlinger, J.; Brambrink, T.; Medeiros, L.A.; Lee, T.I.; Levine, S.S.; Wernig, M.; Tajonar, A.; Ray, M.K.; et al. Polycomb Complexes Repress Developmental Regulators in Murine Embryonic Stem Cells. *Nature* **2006**, *441*, 349–353. [[CrossRef](#)] [[PubMed](#)]
68. Creyghton, M.P.; Cheng, A.W.; Welstead, G.G.; Kooistra, T.; Carey, B.W.; Steine, E.J.; Hanna, J.; Lodato, M.A.; Frampton, G.M.; Sharp, P.A.; et al. Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21931–21936. [[CrossRef](#)] [[PubMed](#)]
69. Kingsley, N.B.; Kern, C.; Creppe, C.; Hales, E.N.; Zhou, H.; Kalbfleisch, T.S.; MacLeod, J.N.; Petersen, J.L.; Finno, C.J.; Bellone, R.R. Functionally Annotating Regulatory Elements in the Equine Genome Using Histone Mark ChIP-Seq. *Genes* **2019**, *11*, 3. [[CrossRef](#)]
70. Hisey, E.A.; Hermans, H.; Lounsbury, Z.T.; Avila, F.; Grahn, R.A.; Knickelbein, K.E.; Duward-Akhurst, S.A.; McCue, M.E.; Kalbfleisch, T.S.; Lassaline, M.E.; et al. Whole Genome Sequencing Identified a 16 Kilobase Deletion on ECA13 Associated with Distichiasis in Friesian Horses. *BMC Genom.* **2020**, *21*, 848. [[CrossRef](#)] [[PubMed](#)]

71. Kingsley, N.B.; Hamilton, N.A.; Lindgren, G.; Orlando, L.; Bailey, E.; Brooks, S.; McCue, M.; Kalbfleisch, T.S.; MacLeod, J.N.; Petersen, J.L.; et al. “Adopt-a-Tissue” Initiative Advances Efforts to Identify Tissue-Specific Histone Marks in the Mare. *Front. Genet.* **2021**, *12*, 649959. [[CrossRef](#)]
72. Lobanenkov, V.V.; Nicolas, R.H.; Adler, V.V.; Paterson, H.; Klenova, E.M.; Polotskaja, A.V.; Goodwin, G.H. A Novel Sequence-Specific DNA Binding Protein Which Interacts with Three Regularly Spaced Direct Repeats of the CCCTC-Motif in the 5'-Flanking Sequence of the Chicken c-Myc Gene. *Oncogene* **1990**, *5*, 1743–1753. [[PubMed](#)]
73. Baniahmad, A.; Steiner, C.; Köhne, A.C.; Renkawitz, R. Modular Structure of a Chicken Lysozyme Silencer: Involvement of an Unusual Thyroid Hormone Receptor Binding Site. *Cell* **1990**, *61*, 505–514. [[CrossRef](#)]
74. Vostrov, A.A.; Quitschke, W.W. The Zinc Finger Protein CTCF Binds to the APB β Domain of the Amyloid β -Protein Precursor Promoter. *J. Biol. Chem.* **1997**, *272*, 33353–33359. [[CrossRef](#)] [[PubMed](#)]
75. Bell, A.C.; West, A.G.; Felsenfeld, G. The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. *Cell* **1999**, *98*, 387–396. [[CrossRef](#)]
76. Dorsett, D. Distance-Independent Inactivation of an Enhancer by the Suppressor of Hairy-Wing DNA-Binding Protein of *Drosophila*. *Genetics* **1993**, *134*, 1135–1144. [[CrossRef](#)] [[PubMed](#)]
77. Hebbes, T.R.; Clayton, A.L.; Thorne, A.W.; Crane-Robinson, C. Core Histone Hyperacetylation Co-Maps with Generalized DNase I Sensitivity in the Chicken Beta-Globin Chromosomal Domain. *EMBO J.* **1994**, *13*, 1823–1830. [[CrossRef](#)] [[PubMed](#)]
78. Kellum, R.; Schedl, P. A Group of Scs Elements Function as Domain Boundaries in an Enhancer-Blocking Assay. *Mol. Cell Biol.* **1992**, *12*, 2424–2431. [[CrossRef](#)]
79. Udvardy, A.; Maine, E.; Schedl, P. The 87A7 Chromomere. Identification of Novel Chromatin Structures Flanking the Heat Shock Locus That May Define the Boundaries of Higher Order Domains. *J. Mol. Biol.* **1985**, *185*, 341–358. [[CrossRef](#)]
80. Filippova, G.N.; Fagerlie, S.; Klenova, E.M.; Myers, C.; Dehner, Y.; Goodwin, G.; Neiman, P.E.; Collins, S.J.; Lobanenkov, V.V. An Exceptionally Conserved Transcriptional Repressor, CTCF, Employs Different Combinations of Zinc Fingers to Bind Diverged Promoter Sequences of Avian and Mammalian c-Myc Oncogenes. *Mol. Cell Biol.* **1996**, *16*, 2802–2813. [[CrossRef](#)]
81. Ohlsson, R.; Renkawitz, R.; Lobanenkov, V. CTCF Is a Uniquely Versatile Transcription Regulator Linked to Epigenetics and Disease. *Trends Genet.* **2001**, *17*, 520–527. [[CrossRef](#)]
82. Heath, H.; de Almeida, C.R.; Sleutels, F.; Dingjan, G.; van de Nobelen, S.; Jonkers, I.; Ling, K.-W.; Gribnau, J.; Renkawitz, R.; Grosveld, F.; et al. CTCF Regulates Cell Cycle Progression of $\alpha\beta$ T Cells in the Thymus. *EMBO J.* **2008**, *27*, 2839–2850. [[CrossRef](#)]
83. Kim, T.H.; Abdullaev, Z.K.; Smith, A.D.; Ching, K.A.; Loukinov, D.I.; Green, R.D.; Zhang, M.Q.; Lobanenkov, V.V.; Ren, B. Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. *Cell* **2007**, *128*, 1231–1245. [[CrossRef](#)]
84. Nakahashi, H.; Kwon, K.-R.K.; Resch, W.; Vian, L.; Dose, M.; Stavreva, D.; Hakim, O.; Pruett, N.; Nelson, S.; Yamane, A.; et al. A Genome-Wide Map of CTCF Multivalency Redefines the CTCF Code. *Cell Rep.* **2013**, *3*, 1678–1689. [[CrossRef](#)]
85. Jothi, R.; Cuddapah, S.; Barski, A.; Cui, K.; Zhao, K. Genome-Wide Identification of in Vivo Protein-DNA Binding Sites from ChIP-Seq Data. *Nucleic Acids Res.* **2008**, *36*, 5221–5231. [[CrossRef](#)] [[PubMed](#)]
86. Splinter, E.; Heath, H.; Kooren, J.; Palstra, R.-J.; Klous, P.; Grosveld, F.; Galjart, N.; de Laat, W. CTCF Mediates Long-Range Chromatin Looping and Local Histone Modification in the Beta-Globin Locus. *Genes Dev.* **2006**, *20*, 2349–2354. [[CrossRef](#)] [[PubMed](#)]
87. Hou, C.; Zhao, H.; Tanimoto, K.; Dean, A. CTCF-Dependent Enhancer-Blocking by Alternative Chromatin Loop Formation. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20398–20403. [[CrossRef](#)]
88. Wendt, K.S.; Yoshida, K.; Itoh, T.; Bando, M.; Koch, B.; Schirghuber, E.; Tsutsumi, S.; Nagae, G.; Ishihara, K.; Mishiro, T.; et al. Cohesin Mediates Transcriptional Insulation by CCCTC-Binding Factor. *Nature* **2008**, *451*, 796–801. [[CrossRef](#)]
89. Parelho, V.; Hadjir, S.; Spivakov, M.; Leleu, M.; Sauer, S.; Gregson, H.C.; Jarmuz, A.; Canzonetta, C.; Webster, Z.; Nesterova, T.; et al. Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell* **2008**, *132*, 422–433. [[CrossRef](#)]
90. Lieberman-Aiden, E.; Berkum, N.L.; van Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.; et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **2009**, *326*, 289–293. [[CrossRef](#)] [[PubMed](#)]
91. Rao, S.S.P.; Huntley, M.H.; Durand, N.C.; Stamenova, E.K.; Bochkov, I.D.; Robinson, J.T.; Sanborn, A.L.; Machol, I.; Omer, A.D.; Lander, E.S.; et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **2014**, *159*, 1665–1680. [[CrossRef](#)]
92. Vietri Rudan, M.; Barrington, C.; Henderson, S.; Ernst, C.; Odom, D.T.; Tanay, A.; Hadjir, S. Comparative Hi-C Reveals That CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Rep.* **2015**, *10*, 1297–1309. [[CrossRef](#)] [[PubMed](#)]
93. Guo, Y.; Xu, Q.; Canzio, D.; Shou, J.; Li, J.; Gorkin, D.U.; Jung, I.; Wu, H.; Zhai, Y.; Tang, Y.; et al. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **2015**, *162*, 900–910. [[CrossRef](#)]
94. Nichols, M.H.; Corces, V.G. A CTCF Code for 3D Genome Architecture. *Cell* **2015**, *162*, 703–705. [[CrossRef](#)]
95. Sanborn, A.L.; Rao, S.S.P.; Huang, S.-C.; Durand, N.C.; Huntley, M.H.; Jewett, A.I.; Bochkov, I.D.; Chinnappan, D.; Cutkosky, A.; Li, J.; et al. Chromatin Extrusion Explains Key Features of Loop and Domain Formation in Wild-Type and Engineered Genomes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E6456–E6465. [[CrossRef](#)] [[PubMed](#)]
96. Fudenberg, G.; Imakaev, M.; Lu, C.; Goloborodko, A.; Abdennur, N.; Mirny, L.A. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **2016**, *15*, 2038–2049. [[CrossRef](#)] [[PubMed](#)]

97. Haarhuis, J.H.I.; van der Weide, R.H.; Blomen, V.A.; Yáñez-Cuna, J.O.; Amendola, M.; van Ruiten, M.S.; Krijger, P.H.L.; Teunissen, H.; Medema, R.H.; van Steensel, B.; et al. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **2017**, *169*, 693–707.e14. [[CrossRef](#)]
98. Allahyar, A.; Vermeulen, C.; Bouwman, B.A.M.; Krijger, P.H.L.; Versteegen, M.J.A.M.; Geeven, G.; van Kranenburg, M.; Pieterse, M.; Straver, R.; Haarhuis, J.H.I.; et al. Enhancer Hubs and Loop Collisions Identified from Single-Allele Topologies. *Nat. Genet.* **2018**, *50*, 1151–1160. [[CrossRef](#)] [[PubMed](#)]
99. Bernstein, B.E.; Mikkelsen, T.S.; Xie, X.; Kamal, M.; Huebert, D.J.; Cuff, J.; Fry, B.; Meissner, A.; Wernig, M.; Plath, K.; et al. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **2006**, *125*, 315–326. [[CrossRef](#)] [[PubMed](#)]
100. Schreiber, S.L.; Bernstein, B.E. Signaling Network Model of Chromatin. *Cell* **2002**, *111*, 771–778. [[CrossRef](#)]
101. Strahl, B.D.; Allis, C.D. The Language of Covalent Histone Modifications. *Nature* **2000**, *403*, 41–45. [[CrossRef](#)]
102. Hon, G.; Ren, B.; Wang, W. ChromaSig: A Probabilistic Approach to Finding Common Chromatin Signatures in the Human Genome. *PLoS Comput. Biol.* **2008**, *4*, e1000201. [[CrossRef](#)]
103. Ernst, J.; Kellis, M. Discovery and Characterization of Chromatin States for Systematic Annotation of the Human Genome. *Nat. Biotechnol.* **2010**, *28*, 817–825. [[CrossRef](#)]
104. Ernst, J.; Kellis, M. ChromHMM: Automating Chromatin-State Discovery and Characterization. *Nat. Methods* **2012**, *9*, 215–216. [[CrossRef](#)] [[PubMed](#)]
105. Fachinetti, D.; Diego Folco, H.; Nechemia-Arbely, Y.; Valente, L.P.; Nguyen, K.; Wong, A.J.; Zhu, Q.; Holland, A.J.; Desai, A.; Jansen, L.E.T.; et al. A Two-Step Mechanism for Epigenetic Specification of Centromere Identity and Function. *Nat. Cell Biol.* **2013**, *15*, 1056–1066. [[CrossRef](#)] [[PubMed](#)]
106. Nergadze, S.G.; Piras, F.M.; Gamba, R.; Corbo, M.; Cerutti, F.; McCarter, J.G.W.; Cappelletti, E.; Gozzo, F.; Harman, R.M.; Antczak, D.F.; et al. Birth, Evolution, and Transmission of Satellite-Free Mammalian Centromeric Domains. *Genome Res.* **2018**, *28*, 789–799. [[CrossRef](#)] [[PubMed](#)]
107. Carbone, L.; Nergadze, S.G.; Magnani, E.; Misceo, D.; Francesca Cardone, M.; Roberto, R.; Bertoni, L.; Attolini, C.; Francesca Piras, M.; de Jong, P.; et al. Evolutionary Movement of Centromeres in Horse, Donkey, and Zebra. *Genomics* **2006**, *87*, 777–782. [[CrossRef](#)]
108. Piras, F.M.; Nergadze, S.G.; Poletto, V.; Cerutti, F.; Ryder, O.A.; Leeb, T.; Raimondi, E.; Giulotto, E. Phylogeny of Horse Chromosome 5q in the Genus *Equus* and Centromere Repositioning. *Cytogenet. Genome Res.* **2009**, *126*, 165–172. [[CrossRef](#)]
109. Piras, F.M.; Nergadze, S.G.; Magnani, E.; Bertoni, L.; Attolini, C.; Khoraiuli, L.; Raimondi, E.; Giulotto, E. Uncoupling of Satellite DNA and Centromeric Function in the Genus *Equus*. *PLoS Genet.* **2010**, *6*, e1000845. [[CrossRef](#)]
110. Giulotto, E.; Raimondi, E.; Sullivan, K.F. The Unique DNA Sequences Underlying Equine Centromeres. In *Centromeres and Kinetochores*; Black, B.E., Ed.; Progress in Molecular and Subcellular Biology; Springer: Cham, Switzerland, 2017; Volume 56, pp. 337–354. ISBN 978-3-319-58591-8.
111. Purgato, S.; Belloni, E.; Piras, F.M.; Zoli, M.; Badiale, C.; Cerutti, F.; Mazzagatti, A.; Perini, G.; Della Valle, G.; Nergadze, S.G.; et al. Centromere Sliding on a Mammalian Chromosome. *Chromosoma* **2015**, *124*, 277–287. [[CrossRef](#)]
112. Cappelletti, E.; Piras, F.M.; Badiale, C.; Bambi, M.; Santagostino, M.; Vara, C.; Masterson, T.A.; Sullivan, K.F.; Nergadze, S.G.; Ruiz-Herrera, A.; et al. CENP-A Binding Domains and Recombination Patterns in Horse Spermatocytes. *Sci. Rep.* **2019**, *9*, 15800. [[CrossRef](#)] [[PubMed](#)]
113. Roberti, A.; Bensi, M.; Mazzagatti, A.; Piras, F.M.; Nergadze, S.G.; Giulotto, E.; Raimondi, E. Satellite DNA at the Centromere Is Dispensable for Segregation Fidelity. *Genes* **2019**, *10*, 469. [[CrossRef](#)]
114. Athwal, R.K.; Walkiewicz, M.P.; Baek, S.; Fu, S.; Bui, M.; Camps, J.; Ried, T.; Sung, M.-H.; Dalal, Y. CENP-A Nucleosomes Localize to Transcription Factor Hotspots and Subtelomeric Sites in Human Cancer Cells. *Epigenetics Chromatin* **2015**, *8*, 2. [[CrossRef](#)] [[PubMed](#)]
115. Gao, S.; Nanaei, H.A.; Wei, B.; Wang, Y.; Wang, X.; Li, Z.; Dai, X.; Wang, Z.; Jiang, Y.; Shao, J. Comparative Transcriptome Profiling Analysis Uncovers Novel Heterosis-Related Candidate Genes Associated with Muscular Endurance in Mules. *Animals* **2020**, *10*, 980. [[CrossRef](#)]
116. Sturm, G.; List, M.; Zhang, J.D. Tissue Heterogeneity Is Prevalent in Gene Expression Studies. *NAR Genom. Bioinform.* **2021**, *3*, lqab077. [[CrossRef](#)] [[PubMed](#)]
117. Nagano, T.; Lubling, Y.; Stevens, T.J.; Schoenfelder, S.; Yaffe, E.; Dean, W.; Laue, E.D.; Tanay, A.; Fraser, P. Single-Cell Hi-C Reveals Cell-to-Cell Variability in Chromosome Structure. *Nature* **2013**, *502*, 59–64. [[CrossRef](#)]
118. Patel, A.P.; Tirosch, I.; Trombetta, J.J.; Shalek, A.K.; Gillespie, S.M.; Wakimoto, H.; Cahill, D.P.; Nahed, B.V.; Curry, W.T.; Martuza, R.L.; et al. Single-Cell RNA-Seq Highlights Intratumoral Heterogeneity in Primary Glioblastoma. *Science* **2014**, *344*, 1396–1401. [[CrossRef](#)] [[PubMed](#)]
119. Pott, S.; Lieb, J.D. Single-Cell ATAC-Seq: Strength in Numbers. *Genome Biol.* **2015**, *16*, 172. [[CrossRef](#)]
120. Rotem, A.; Ram, O.; Shores, N.; Sperling, R.A.; Goren, A.; Weitz, D.A.; Bernstein, B.E. Single-Cell ChIP-Seq Reveals Cell Subpopulations Defined by Chromatin State. *Nat. Biotechnol.* **2015**, *33*, 1165–1172. [[CrossRef](#)]