

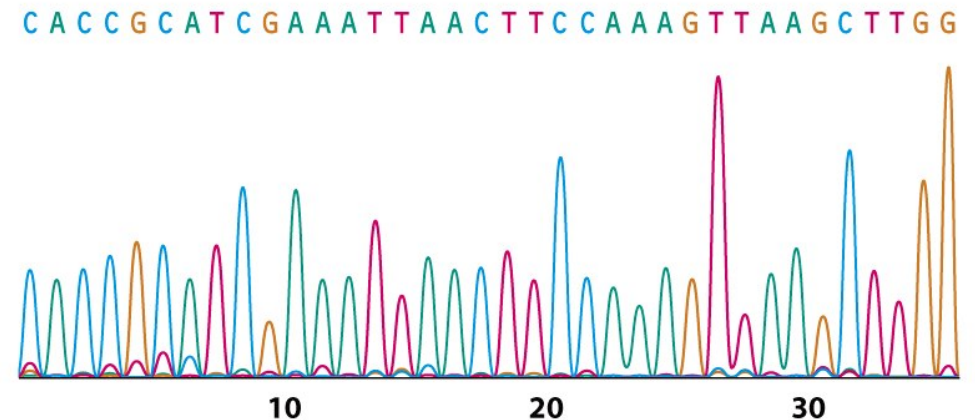


# Il sequenziamento del genoma equino e le sue principali applicazioni The horse genome sequence and its main applications

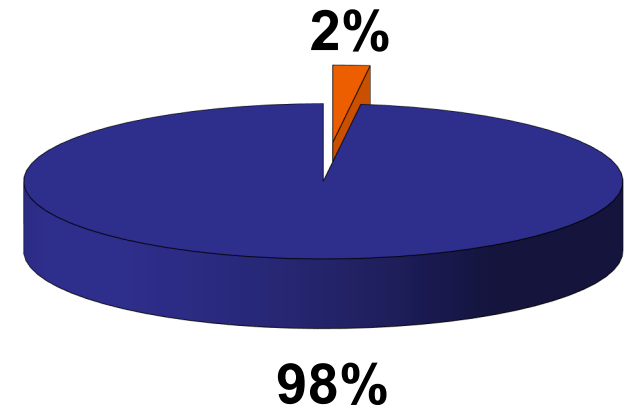


**Elena Giulotto**

**Dipartimento di Biologia e Biotechnologie,  
Università di Pavia**



**The genome is the totality of the DNA of an organism and contains all its genetic information**



**In mammals, protein coding genes: less than 2% of genome. The genome is mainly non-coding: tandem repeats, interspersed repeats, unique non-coding. But 90% of genome is transcribed (miRNA, lncRNA...)**

## First sequence of the human genome, 2001



About  $3 \times 10^9$  base pairs in the human genome, about 20 000 protein coding genes

The status of horse genome sequencing was very preliminary: only a map



<https://horsegenomeworkshop.com>

## **1995, First International Equine Gene Mapping Workshop.**

**The Horse Genome Workshop is a cooperative international effort by more than 100 scientists from 25 countries to define the genome of the domestic horse.**

**Meetings have been sponsored every 2 years by the Havemeyer Foundation (Pavia, 2018; Ithaca, NY, 2022; Paris, 2024).**

**Collaboration between Horse Genome Project and Broad Institute to sequence the horse genome**



## Publication of the first reference genome for the horse



### Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse

C. M. Wade,<sup>1,2,3\*</sup> E. Giulotto,<sup>4</sup> S. Sigurdsson,<sup>1,5</sup> M. Zoli,<sup>6</sup> S. Gnerre,<sup>1</sup> F. Imsland,<sup>5</sup> T. L. Lear,<sup>7</sup> D. L. Adelson,<sup>8</sup> E. Bailey,<sup>7</sup> R. R. Bellone,<sup>9</sup> H. Blöcker,<sup>10</sup> O. Distl,<sup>11</sup> R. C. Edgar,<sup>12</sup> M. Garber,<sup>1</sup> T. Leeb,<sup>11,13</sup> E. Mauceli,<sup>1</sup> J. N. MacLeod,<sup>7</sup> M. C. T. Penedo,<sup>14</sup> J. M. Raison,<sup>8</sup> T. Sharpe,<sup>1</sup> J. Vogel,<sup>15</sup> L. Andersson,<sup>5</sup> D. F. Antczak,<sup>16</sup> T. Biagi,<sup>1</sup> M. M. Binns,<sup>17</sup> B. P. Chowdhary,<sup>8</sup> S. J. Coleman,<sup>7</sup> G. Della Valle,<sup>6</sup> S. Fryc,<sup>1</sup> G. Guérin,<sup>19</sup> T. Hasegawa,<sup>20</sup> E. W. Hill,<sup>21</sup> J. Jurka,<sup>22</sup> A. Kiialainen,<sup>23</sup> G. Lindgren,<sup>24</sup> J. Liu,<sup>25</sup> E. Magnani,<sup>4</sup> J. R. Mickelson,<sup>26</sup> J. Murray,<sup>27</sup> S. G. Nergadze,<sup>4</sup> R. Onofrio,<sup>1</sup> S. Pedroni,<sup>14</sup> M. F. Piras,<sup>4</sup> T. Raudsepp,<sup>8</sup> M. Rocchi,<sup>28</sup> K. H. Røed,<sup>9</sup> O. A. Ryder,<sup>30</sup> S. Searle,<sup>15</sup> L. Skow,<sup>18</sup> J. E. Swinburne,<sup>31</sup> A. C. Syvänen,<sup>23</sup> T. Tozaki,<sup>32</sup> S. J. Valberg,<sup>26</sup> M. Vaudin,<sup>31</sup> J. R. White,<sup>1</sup> M. C. Zody,<sup>1,5</sup> Broad Institute Genome Sequencing Platform,<sup>1</sup> Broad Institute Whole Genome Assembly Team,<sup>1</sup> E. S. Lander,<sup>1,33</sup> K. Lindblad-Toh<sup>1,5\*</sup>

**SCIENCE 2009**



Twilight

### A Horse Is a Horse, of Course

The history of horse domestication is closely tied to the history of the human society. Wade *et al.* (p. 865) report on the sequencing and provide a single nucleotide polymorphism map of the horse (*Equus caballus*) genome. Horses are a member of the order perissodactyla (odd-toed animals with hooves). The analysis reveals an evolutionarily new centromere on equine chromosome 11 that displays properties of an immature but fully functioning centromere and is devoid of centromeric satellite sequence. The findings clarify the nature of genetic diversity within and across horse breeds and suggest that the horse was domesticated from a relatively large number of females, but few males.

## THIS WEEK IN Science

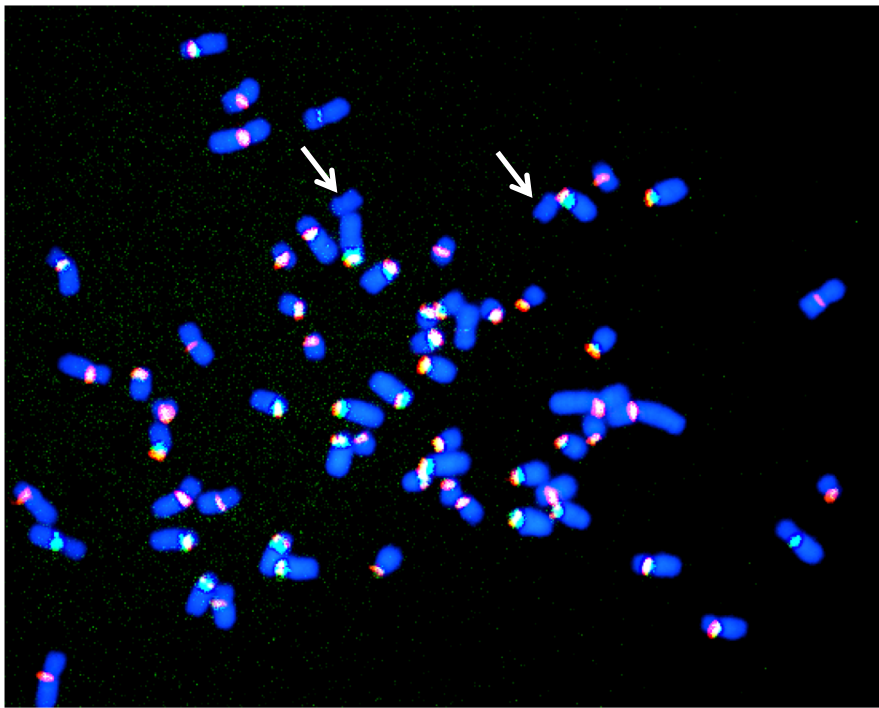
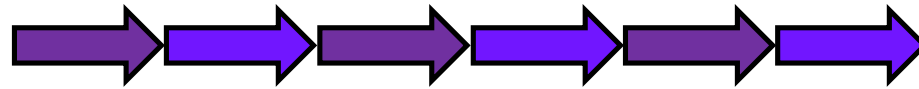
**Genome size  $2.7 \times 10^9$  bp, slightly smaller than the human genome. About 20 000 genes very similar in number and function to human genes.**

**Our work on the satellite-less centromere on chromosome 11 was the highlight of the paper.**

**This finding opened the way to our still ongoing project on organization, function and evolution of equid centromeres**

**Satellite DNA, long series of tandem repeats, is the typical DNA component of centromeric chromatin**

**It **was** believed to be essential for mammalian centromere function and stability**



**In the horse all centromeres contain satellite DNA **except the one of chromosome 11.**  
**A revolutionary finding.****

Wade et al Science 2009  
Piras et al PLoS Genet 2010  
Nergadze et al Genome Res 2018



**Nature Italy**

**RESEARCH HIGHLIGHT**

29 September 2022

## **Lessons from the DNA of zebras**

When cells divide, centromeres make sure everything goes well. A study on the chromosomes of equids investigated the mechanisms.



**Commentary on our recent paper  
Cappelletti et al, Molecular Biology and Evolution,  
August 2022**





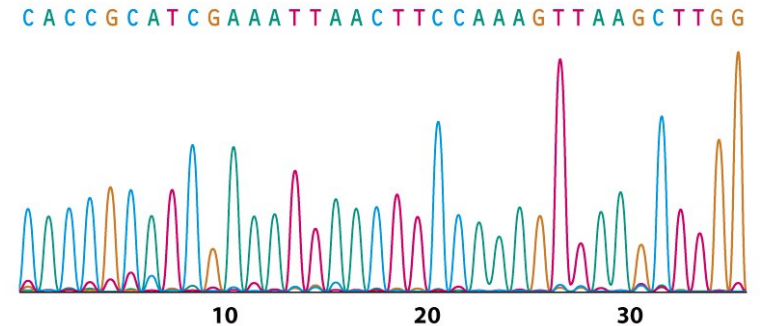
## Main phases of horse genome sequencing and analysis

- 2009, Publication of the first horse genome sequence **EquCab2** (Sanger method)
- 2018, Refinement of the horse reference genome **EquCab3** (Next Generation Sequencing methods)
- 2009..., Post-genomic annotation and analysis
- 2015..., Functional Annotation of Animal Genomes (**FAANG**), Horse FAANG consortium (Epigenetic regulation of genome function)

# DNA sequencing methods and costs

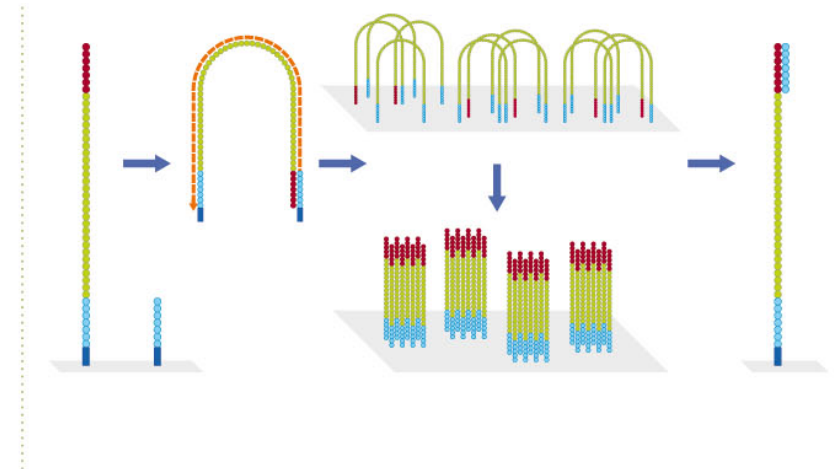
**The first reference genome of Twilight costed >7 M \$**

**With traditional Sanger method 700-800 bp  
in a single read. Cloning**



**Now a new draft genome costs <1000 \$**

**With Next Generation Sequencing methods  
hundreds of millions reads from one DNA  
sample.  
No cloning**



**More than 5000 horse genome sequences available in 2022**

**Thousands of functional sequencing data sets**

**The importance of bioinformatic tools**

# **From Genomics to Post-Genomics**



**POST-GENOMICS is the interpretation of a genome sequence**

**To identify:**

**All genes and their function**

**All regulatory elements**

**The function of extra-genic DNA sequences**

**Thanks to genome-wide technologies more and more genes relevant  
for the horse industry are identified**

**Disease**

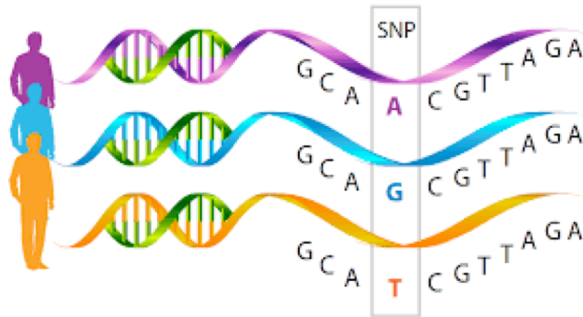
**Performance**

**Coat color**

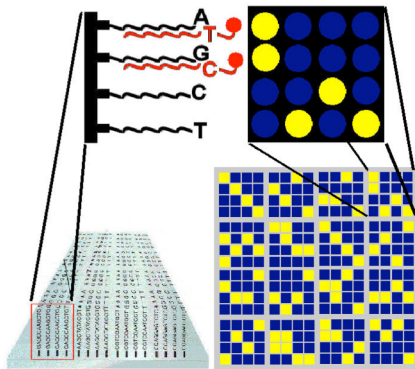
**Morphology traits**

**Complex traits**

# SNP arrays



**SNP: Single Nucleotide Polymorphism. Is the variation of a single nucleotide with a frequency of >1% in the population**



**SNP array: an array of thousands SNP markers to hybridize with DNA samples from many individuals in search of linkage with trait under study**

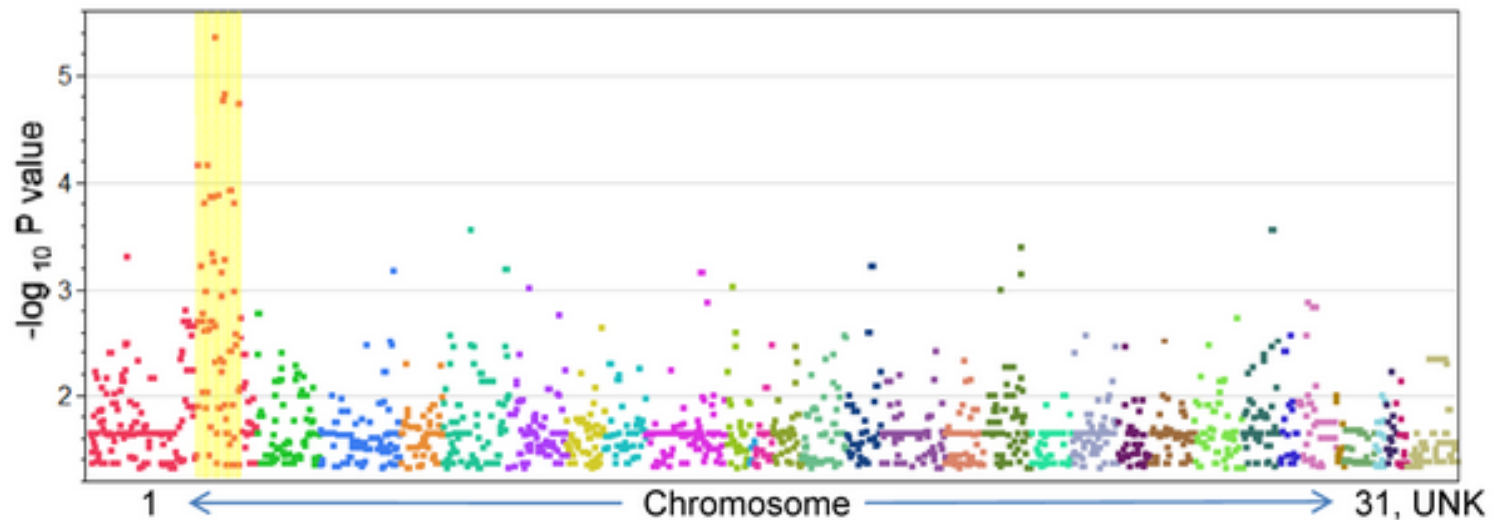
**Over 600 000 SNP markers available now for genome-wide association studies**

## Example of a successful application of the SNP array method

### Whole-Genome SNP Association in the Horse: Identification of a Deletion in Myosin Va Responsible for Lavender Foal Syndrome

Samantha A. Brooks et al PLoS Genet 2010

Genotypic association tests using six affected foals and their 30 healthy relatives revealed a single region on chromosome 1 (ECA1) with statistical significance above that of the rest of the genome. Development of a diagnostic test for the LFS recessive allele.





## **Several genes identified with the SNP array approach**

**Lavander Foal Syndrome (Brooks et al 2010)**

**Muscular Dystrophy with Hydrocephalus (Ducro et al 2015)**

**Ocular Squamous Cell Carcinoma (Bellone et al 2017)**

**Naked Foal Syndrome (Bauer et al 2017)**

**Coarly Coat (Thomer 2018)**

**...and many more**

**For a comprehensive list: Table 2 in the review from Raudsepp et al, Animal Genetics 2019**

## Some genetic diseases with a simple test available

Disease	Major breed(s)
Hyperkalemic periodic paralysis (HYPP) (Rudolph <i>et al.</i> 1992a,b; Cannon <i>et al.</i> 1995; Naylor <i>et al.</i> 1999)	Quarter Horse Paint QH-related
Polysaccharide storage myopathy (PSSM) (McCue <i>et al.</i> 2008a,b, 2009)	Quarter Horse QH-related Warmbloods Draughts
Malignant hyperthermia (Aleman <i>et al.</i> 2005, 2009)	Quarter Horse
Glycogen branching enzyme deficiency (GBED) (Valberg <i>et al.</i> 2001; Ward <i>et al.</i> 2003, 2004)	Quarter Horse
Severe combined immunodeficiency (SCID) (Wiler <i>et al.</i> 1995; Bailey <i>et al.</i> 1997; Shin <i>et al.</i> 1997)	Arabian
Junctional epidermolysis bullosa (JEB) (Lieto and Cothran 2003; Graves <i>et al.</i> 2008)	American Saddlebred
Junctional epidermolysis bullosa (JEB) (Spirito <i>et al.</i> 2002; Milenkovic <i>et al.</i> 2003)	Belgian French draughts
Hereditary equine regional dermal asthenia (HERDA) (White <i>et al.</i> 2004, 2007; Tryon <i>et al.</i> 2005, 2007)	Quarter Horse
Overo lethal white syndrome (OLWS), Ileocolonic aganglionosis (Metallinos <i>et al.</i> 1998; Yang <i>et al.</i> 1998; Santschi <i>et al.</i> 2001)	Paint
Grey horse melanoma (Pielberg <i>et al.</i> 2008)	Many
Lavender foal syndrome (Brooks <i>et al.</i> 2010)	Arabian

**Some coat colors with a simple  
genetic test available**

Red/Black Factor

Agouti (Bay/Black)

Cream Dilution

Pearl Dilution

Champagne Dilution

Grey

Frame Overo (LWO)

Splashed White Overo (SW1, SW2, SW3)

Sabino 1

Tobiano

Dominant White

Appaloosa Coat Pattern (LP)

Dun Coat Colour

## **Multigene complex traits under study: candidate genes from GWAS**

**Osteochondrosis**

**Metabolic syndrome**

**Laminitis**

**Equine asthma (RAO)**

**Insect bite hypersensitivity**

**Infertility**

**Size**

**Athletic performance**

**For a comprehensive list: Table 4 in the review from Raudsepp et al, Animal Genetics 2019**



## **Horse athletic performance**

**More than 100 genes in human athletic performance.  
Some are under study in the horse**

**Strength and composition of muscle fibers (myostatin)**

**Muscle metabolism and tolerance to exercise**

**Hemodynamics and aerobic metabolism**

**Physiology of tendons and ligaments**

**Attitude and motivation**

**Horses are models for human athletes**

# MYOSTATIN

**Hormone controlling (reducing) muscle development**

**Defects in myostatin gene cause muscular hypertrophy**



**Double muscle: Piemontese, Belgian Blue**

## In the horse

**Insertion of a transposable element ERE1 in myostatin promoter reduces gene expression (Santagostino et al BMC Genet 2015)**

**Very frequent in quarter horses**

45% two variant alleles

25% one variant allele

**Frequent in thoroughbreds**

24% two variant alleles

39% one variant allele

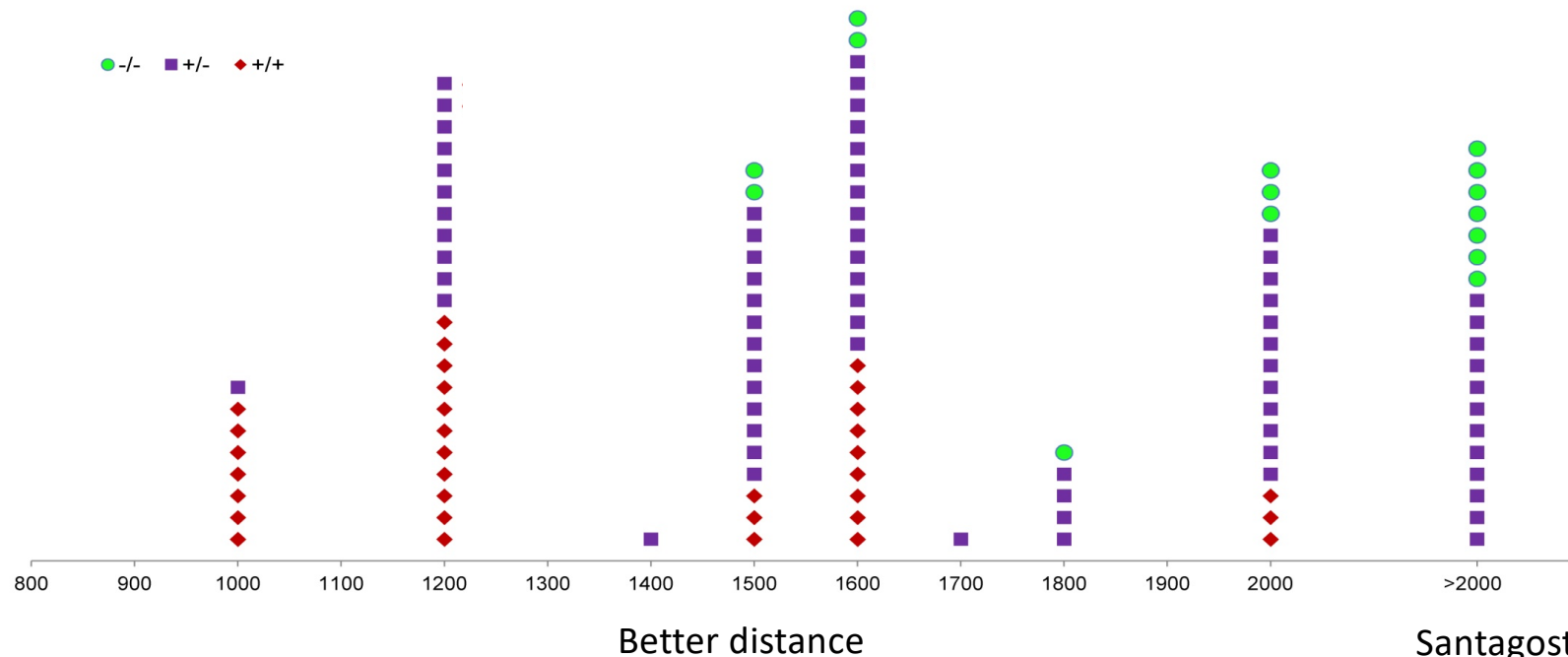
**Absent in trotters, show jumpers, lipizzaner...**

Elite Thoroughbreds



Short distance 2 variant alleles  
Medium distance 1 variant allele  
Long distance 2 normal alleles

The myostatin variant has been  
artificially selected in sprint horses



Santagostino et al 2015

**The availability of many horse genome sequences is also providing massive information on domestication and its genetic cost, history of horse breeds and population genetics.**

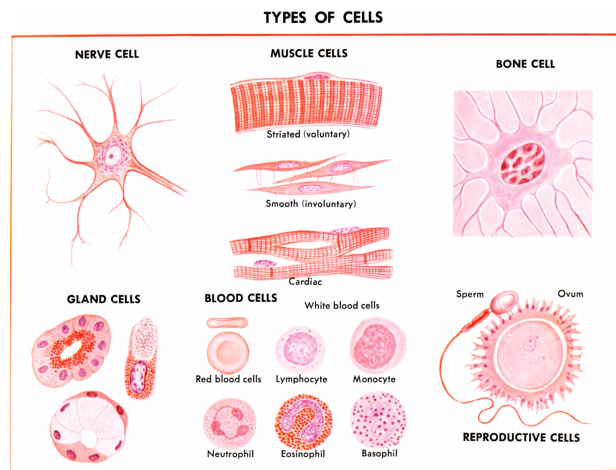
**Comparative analyses with the genomes of other equids and more evolutionarily distant mammals is providing exciting information on the evolution of the genus *Equus* and on the mechanisms of speciation.**



# **From Genomics to Post-Genomics to Epigenomics**

# EPIGENETICS

The study of changes in gene function that are mitotically and/or meiotically *heritable* and are not due to a change in DNA sequence.



How can gene function be altered in a way that is both *heritable* and conservative of DNA sequence?

By “marking” the DNA.

Two primary mechanisms :

- Covalent modification of DNA: DNA methylation
- Covalent modification of key DNA-associated proteins: histone post-translational modifications

**Thousands of functional sequencing data sets:**

**RNA-seq (transcriptome)**

**ChIP-seq (DNA sequences interacting with specific protein)**

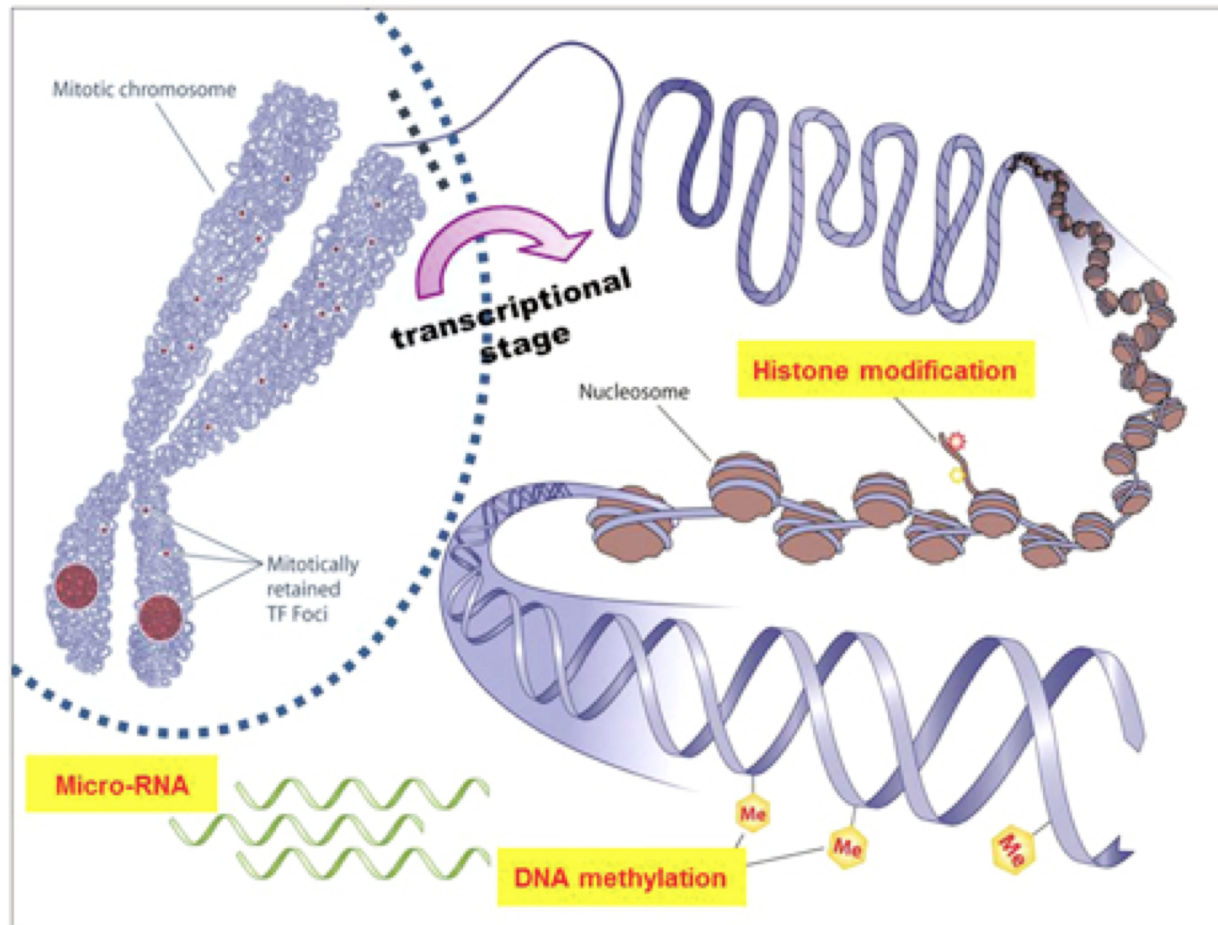
**ATAC-seq (chromatin accessibility)**

**DNA methylation (post-replication modification of DNA)**

**The importance of bioinformatic tools**



# Gene expression regulation and chromatin epigenetic markers



The Encyclopedia of DNA Elements started in 2003. ENCODE is a public research project which aims to identify functional elements in the human genome.

A worldwide consortium to understand the most critical factors involved in genome function.

Example: cancer

# The goals of ENCODE applied to farmed animals



A coordinated international action to  
accelerate genome to phenome

<http://www.faang.org>

@faangomics

faang@iastate.edu

**Since 2015, the FAANG consortium is working to discover basic functional knowledge of genome function to decipher the genotype-to-phenotype (G2P) link in farmed animals.**

**Over 500 scientists in the FAANG consortium**

# Core Assays

- RNA-Seq
  - Stranded
- Chromatin Accessibility
  - ATAC-Seq
- Histone Modification
  - H3K4me3
  - H3K27me3
  - H3K27ac
  - H3K4me1



# The Horse FAANG





## **The Horse FAANG**

**2 Healthy Thoroughbred mares**

**2 Healthy Thoroughbred stallions**

**>80 tissues collected**

- Whole-genome sequencing**

- From prioritized tissues:**

  - RNA-seq (mRNA, smallRNA)**

  - ChIP-seq with several histone markers**

  - DNA methylation**

  - epigenetic markers at centromeres**

# Special thanks to Twilight

The thouroughbred mare whose genome was sequenced



**Twilight in July 2022 with Francesca Piras, Eleonora Cappelletti and myself at the 13th International Havemeyer Horse Genome Workshop, Cornell University, Ithaca, NY**



University of Pavia  
**Solomon Nergadze**  
**Eleonora Cappelletti**  
**Francesca Piras**  
**Elena Raimondi**  
**Wasma Amin**  
**Lorenzo Sola**  
**Marco Santagostino**  
**Francesco Lescai**

University of Nebraska  
Jessica Petersen  
  
UCDavis  
Rebecca Bellone  
  
University of Minnesota  
Carrie Finno

University of Kentucky  
Ernie Bailey,  
Ted Kalbfleish

Cornell University, USA  
**Doug Antczac**  
**Don Miller**

Galway University, Ireland  
**Kevin Sullivan**

University of Barcelona  
**Aurora Ruiz-Herrera**

University of Munich  
**Irina Solovei**

**The Horse Genome Project**

**The Horse FAANG Consortium**





# Suggested reading

**Raudsepp et al , Ten years of the horse reference genome: insights into equine domestication and population dynamics in the post-genome era***Animal Genetics* 50, 569. 2019

**Bellone and Avila, Genetic testing in the horse.** *Vet Clin Equine* 36, 221. 2020

**Petersen and Coleman, Next-generation sequencing in equine genomics.** *Vet Clinics of North America: equine practice* 36, 195. 2020

**Brook, Genomics in the horse industry: discovering new questions at every turn.** *J Equine Vet Science* 100. 2021

**Thank you for your attention**

