



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project**

**Citation for published version:**

The FAANG Consortium 2015, 'Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project' *Genome Biology*, vol. 16, no. 1, 57. DOI: 10.1186/s13059-015-0622-4

**Digital Object Identifier (DOI):**

[10.1186/s13059-015-0622-4](https://doi.org/10.1186/s13059-015-0622-4)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Genome Biology

**Publisher Rights Statement:**

© 2015 Andersson et al.; licensee BioMed Central.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



OPEN LETTER

Open Access

# Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project

The FAANG Consortium, Leif Andersson<sup>1,2</sup>, Alan L Archibald<sup>3</sup>, Cynthia D Bottema<sup>4</sup>, Rudiger Brauning<sup>5</sup>, Shane C Burgess<sup>6</sup>, Dave W Burt<sup>3</sup>, Eduardo Casas<sup>7</sup>, Hans H Cheng<sup>8</sup>, Laura Clarke<sup>9</sup>, Christine Couldrey<sup>10</sup>, Brian P Dalrymple<sup>11</sup>, Christine G Elsik<sup>12</sup>, Sylvain Foissac<sup>13</sup>, Elisabetta Giuffra<sup>14\*</sup>, Martien A Groenen<sup>15</sup>, Ben J Hayes<sup>16,17,18</sup>, LuSheng S Huang<sup>19</sup>, Hassan Khatib<sup>20</sup>, James W Kijas<sup>11</sup>, Hee-bal Kim<sup>21</sup>, Joan K Lunney<sup>22</sup>, Fiona M McCarthy<sup>23</sup>, John C McEwan<sup>24</sup>, Stephen Moore<sup>25</sup>, Bindu Nanduri<sup>26</sup>, Cedric Notredame<sup>27</sup>, Yniv Palti<sup>28</sup>, Graham S Plastow<sup>29</sup>, James M Reecy<sup>30</sup>, Gary A Rohrer<sup>31</sup>, Elena Sarropoulou<sup>32</sup>, Carl J Schmidt<sup>33</sup>, Jeffrey Silverstein<sup>34</sup>, Ross L Tellam<sup>35</sup>, Michele Tixier-Boichard<sup>14</sup>, Gwenola Tosser-Klopp<sup>13</sup>, Christopher K Tuggle<sup>30\*</sup>, Johanna Vilkki<sup>36</sup>, Stephen N White<sup>37,38</sup>, Shuhong Zhao<sup>39</sup> and Huaijun Zhou<sup>40</sup>

## Abstract

We describe the organization of a nascent international effort, the Functional Annotation of Animal Genomes (FAANG) project, whose aim is to produce comprehensive maps of functional elements in the genomes of domesticated animal species.

## Predictive biology: from sequence to consequence

Most phenotypes are complex and quantitative in nature, and a major goal of biological research lies in using genome information to predict such complex outcomes, whether it is the efficacy of a drug, susceptibility to cancer, or the performance of the daughters of an elite dairy bull. Many of the recent advances in biology have been driven by genome sequence information. The capability to sequence and decipher the instructions encoded in complex animal genomes quickly and at modest cost is now well established. The next challenge is to be able to read the subtlety and complexity of these instructions and to predict the resulting phenotypes, that is, to predict the consequences encoded in sequences. While significant progress in functional genome annotation has been made using various human cell types [1], we argue that filling the genotype-to-phenotype gap requires

functional genome annotation of species with substantial phenotype information.

## The unique value of domesticated animal species for accelerating our understanding of genomes and phenomes

Research on domesticated animals has important scientific and socioeconomic impacts, including contributing to medical research, improving the health and welfare of companion animals, and underpinning improvements in the animal sector of agriculture. A key to these impacts is the wealth of genetic and phenotypic diversity among domesticated animals, coupled with research to elucidate the genetic architecture underlying quantitative traits.

## From association to causation: pioneering success in domesticated species

Deep pedigrees with extensive phenotypic records, genetic and phenotypic diversity shaped by natural and artificial selection, and the latest molecular genomics and statistical tools provide an opportunity to understand the relationship between genotype and phenotype in outbred domesticated and farmed animal species [2]. We cite four examples of past successes. First, the identification of a single base-pair change as the causal genetic variant for the complex callipyge muscle hypertrophy phenotype in sheep [3]. Second, the finding that a single nucleotide change in the 3'-untranslated region of the sheep myostatin gene creates a new microRNA binding site that decreases myostatin protein expression [4]. Third,

\* Correspondence: elisabetta.giuffra@jouy.inra.fr; cktuggle@iastate.edu

<sup>14</sup>UMR1313 Génétique Animale et Biologie Intégrative (GABI), French National Institute for Agricultural Research (INRA), F-78352 Jouy-en-Josas, France

<sup>30</sup>Department of Animal Science, Iowa State University, Ames, IA 50011, USA  
Full list of author information is available at the end of the article

the identification of a single nucleotide change in an *IGF2* intron that is the causal mutation for a quantitative trait locus with effects on muscle growth and fat depth in pigs [5]. Finally, the finding that a premature stop codon in the *DMRT3* gene has a major effect on the pattern of locomotion in horses [6]. Much of the genetic variation underlying quantitative traits is likely to be located in regulatory sequences [7], and two of the examples cited above [3,5] demonstrate the importance of epigenetic mechanisms in determining complex phenotypes.

### Evolution, selection, adaptation

The study of genomes of domesticated animals provides insight into evolution, adaptation and genetic selection. Domesticated and farmed animals represent a wide evolutionary spectrum from bees, through shellfish, fish, birds and mammals, and analyses of their genomes have revealed relationships between sequence and function [8-12]. Genome-wide analysis of domesticated species and their putative wild ancestors has shed light on domestication [8,13-15]. Importantly, the footprint of artificial selection can also be detected and provides glimpses of the relationship between sequence and selected phenotypes [16-18].

### Biomedical models

Several domesticated animal species are widely used to model human biology, including the pig, sheep, chicken and dog. However, while coding sequence variants can be major determinants of phenotype as exemplified by many monogenic inherited diseases, attempts to recapitulate the disease phenotype in genetically modified mice often fail [19]. This lack of accurate translation to human biology demonstrates the need for a better understanding of the genotype-to-phenotype relationship [20], potentially through the use of additional species that better approximate human physiology [21].

### Modeling animals as systems: success in phenotypic selection but little mechanistic knowledge

Animals are complex systems in which predicting phenotype from genotype (sequence) is challenging. However, quantitative geneticists and animal breeders have been remarkably successful at developing statistical animal models that are effective predictors of future performance [22]. The accuracy of these models has been increased by using high-density single nucleotide polymorphism genotypes [22,23]. Further improvements can be achieved through the use of genome sequence data [24-26] and by adding knowledge of the likely effects of the sequence variants, whether coding or regulatory [27]. However, while artificial selection acting on the enormous underlying genetic diversity has made improvements in traits of economic importance, there is little understanding of the biological mechanisms underpinning such phenotypes.

### Recent progress in animal genome sequencing provides new opportunities in elucidating the genotype-to-phenotype connection

Coordinated genome-wide identification of functional elements in multiple species would be an invaluable resource for the dissection of genotype-to-phenotype relationships. The evolutionary breadth of the Encyclopedia of DNA Elements (ENCODE) projects has been expanded from humans to classical model species (mouse [28,29], *Drosophila* [30], *Caenorhabditis elegans* [31] and zebrafish [32]). However, transcriptome complexity differs significantly between species [33]; in general, extrapolation of regulatory sequence data across species has not proven useful [34]. In line with previous evidence, the mouse ENCODE project provided multiple lines of evidence that gene expression and its underlying regulatory programs have substantially diverged between the human and mouse lineages, although a subset of core regulatory programs is largely conserved [29]. Thus, additional sampling of species, especially those with deep phenotypic records, is needed to fully understand how these functional elements define the timing, amplitude and response to developmental and environmental cues [35].

A prerequisite for mapping functional elements is a reference genome assembly. Reference genome sequences have been established for a range of important domesticated animals (Additional file 1). However, the annotation of these genome sequences is currently limited to gene models deduced using RNA expression and DNA variation data. Thus, in comparison to human and mouse, the complexity of the transcriptomes in domesticated animals is inadequately characterized. This is exacerbated by the fact that while 70% to 90% of the coding elements can be readily identified, there is little information on noncoding genes, and even less on the regulatory sequences that often underlie complex traits.

The ENCODE and epigenome consortia have already demonstrated that improved functional annotation is most efficiently delivered collaboratively [1,28-32,36]. Thus, in combination with filling the gap in deriving phenotype from genotype described above, this advantage is a strong motivation for an internationally coordinated Functional Annotation of Animal Genomes (FAANG) project as proposed below.

### The FAANG Consortium

In January 2014, a workshop was convened by the Animal Biotechnology Working Group of the EU-US Biotechnology Research Task Force in San Diego, CA, USA. During this workshop, and in subsequent discussions, basic principles were laid out to establish the FAANG Consortium and to outline plans for a FAANG project (see below). The aim of the Consortium is to produce comprehensive maps of functional elements in the genomes of domesticated

animal species based on common standardized protocols and procedures. The FAANG Consortium signatories are committing to work within the FAANG community to define and improve experimental, metadata and bioinformatics standards; ensure that experiments conducted to produce functional annotation adhere to these standards; and release all the experimental and metadata in an open access manner, rapidly and before publication, in accordance with the Toronto Statement [37].

A web portal has been established to consolidate and distribute information on the FAANG Consortium (standardized protocols and pipelines of analysis, data summaries, and publications) and as a means for new participants to join the Consortium [38]. Additional details on the FAANG Consortium, including current membership and goals, can also be found on the web portal.

### **Delivering the FAANG project**

The human ENCODE project cost over \$150 million and involved at least 442 scientists in 32 institutions around the world. Lessons learned from this project and advances in high-throughput technologies have transformed the ease and efficiency with which this type of project can be executed. A coordinated effort to generate data from similar tissues using common core assays to minimize redundancy and leverage existing activity will enable the FAANG project to make significant progress in a cost-effective manner. ENCODE-type data will be generated at a fraction of the original cost and in a distributed way, thanks to the modular nature of experiments.

### **Parallel sample and data collection from species ready to implement FAANG**

A high-quality reference genome assembly is a prerequisite to initiate a functional annotation effort. Consequently, we propose to start by selecting taxonomically diverse species with high-quality genome assemblies. These species need to have the support of their research community and a critical mass of investigators, as demonstrated by expression of interest and willingness to use core assays and a common data-sharing infrastructure. Currently, domesticated animal species that meet this requirement include chicken, pig, cattle and sheep. We note, however, that research on other species (for example, goat, salmon and catfish) is rapidly expanding the range of genomes suited for a FAANG approach (Additional file 1).

The first phase of the FAANG project will focus on sampling biological replicates representing a limited number of specific biological states to maximize comparisons across species. Where possible, animals with minimal genetic diversity within a species will be sampled. For example, highly inbred lines of chicken can be used. While each species' community will decide on a particular breed, genetic line or cross, FAANG members are committed to

collecting, storing and sharing tissues for initial data collection as well as holding them in reserve for future additional assays. Similarly to recent phases of ENCODE and modENCODE [29,39], FAANG will mostly focus on tissue samples. A first core set of tissues directly related to the large number of quantitative phenotypes available in several domesticated species has been defined. This includes skeletal muscle, adipose, liver, and tissues collected from the reproductive, immune and nervous systems. We believe this will allow a more direct connection between genome function and quantitative phenotype than the transformed cell lines used extensively in the first phase of the ENCODE project [39]. Both male and female progeny will be sampled at neonatal and mature stages.

### **FAANG data types**

Both ENCODE and the International Human Epigenome Consortium have defined robust experimental protocols [40]. We will use these standards as a baseline, adapting them where necessary to reflect the complexities of animal breeds and the different tissues available for animal-based experiments. We plan to employ a few specific core assays, which for the most part employ technologies that work across all targeted species (RNA sequencing, chromatin accessibility, and histone marks) as well as have selected laboratories run these assays for the community with standard protocols (Box 1). Additional assays may be performed by individual research groups based upon specific needs and research interests.

### **Common data infrastructure**

Effective coordination, data management and robust quality control (QC) are essential to converting data generated across multiple laboratories into knowledge. The FAANG consortium will promote standardization of experimental protocols and procedures in computational analysis. A sampling coordination task force will promote standards for sampling and storing conditions, including the documentation of animal origin and environmental conditions. A FAANG Data Coordination Centre (DCC) and a Data Analysis Centre (DAC) will be established to ensure high-quality and standardized data generation and analysis, and accessibility of the data to the wider community [41]. The FAANG DCC will work with the Sequence, Variation and Sample archives at European Molecular Biology Laboratory European Bioinformatics Institute and the National Center for Biotechnology Information to ensure the data are deposited, with suitable metadata descriptions, in the appropriate archives. In addition, the FAANG DCC will provide quality-controlled data to resources like Ensembl, so that the improved annotation is available to the broadest audience possible. Appropriate metadata and data quality standards for test samples will be defined, and the DCC will help to collect and QC data generated by

**Box 1****Core assays:**

**Transcribed loci** - Identifying the transcribed elements of the genome is a key starting point for functional annotation. RNA-sequencing data generated from libraries prepared using stranded protocols [43], the species of interest, and a wide range of tissues, cells and states are critical to describe transcript complexity, including alternative splicing and non-coding RNAs [44-46].

**Chromatin accessibility and architecture** - Assay for transposase-accessible chromatin sequencing (ATAC-seq), which is based on direct *in vitro* transposition of sequencing adaptors into native chromatin, represents a rapid and sensitive alternative to DNaseI footprinting [47] for detecting open chromatin [48,49]. Chromatin immunoprecipitation sequencing (ChIP-seq) to identify sites bound by the highly conserved insulator-binding protein CCCTC-binding factor can be useful for bridging the gap between gene expression and nuclear organization [50].

**Histone modification marks** - The presence of modified histones and the characterization of the sequences to which they are bound will be assayed by well-standardized ChIP-seq assays using validated antibodies that work across a broad range of species. We will start with four histone modification marks among those found most informative by the ENCODE projects [1,51]:

- Histone H3 lysine 4 trimethylation (H3K4me3), which correlates with promoters of active genes and transcription start sites;
- Histone H3 lysine 27 trimethylation (H3K27me3), which marks genes that have been silenced through regional modification;
- Histone H3 lysine 27 acetylation (H3K27ac), which marks active regulatory elements, and may distinguish active enhancers and promoters from their inactive counterparts;
- H3 lysine 4 monomethylation (H3K4me1), which marks regulatory elements associated with enhancers and other distal elements, but is also enriched downstream of transcription start sites.

**Additional assays:**

**Methylation** - DNA methylation is a major epigenetic mark and a well-known regulator of gene expression. Genome-wide analysis of 5-methylcytosines can be performed at nucleotide-level resolution by whole-genome or reduced representation bisulfite sequencing [52].

**Transcription factor binding sites** - ChIP-seq assays can be used to identify sequences bound by specific proteins [40]. However, generating and validating the antibodies necessary will be the main challenge in mapping binding sites for the diversity of transcription factors in species of interest.

**Genome conformation** - Methods of chromosome conformation capture allow the study of the genome-wide chromatin interactome. Hi-C is an upgraded method that provides information about distal elements that, while far apart in the primary sequence, are brought together because of chromosome folding [53,54]. Hi-C results can be optimally integrated with those from other assays of chromatin accessibility (for example, ATAC-seq).

FAANG partners. The DCC will help groups to appropriately archive sample data and metadata and provide mechanisms to share and access data [37]. Key tasks such as mapping the primary sequence data to the appropriate reference genome will be performed by the DCC. The FAANG DAC will consist of distributed groups to establish the best bioinformatic pipelines to analyze FAANG consortium data, and will work closely with the DCC to ensure appropriate QC standards are defined.

**Future expansion of covered species and diversity within and between species**

As reference genomes for new species are added across the tree of life, new insights can be obtained through functional analysis of such species. Thus, it will be important to continue to expand the evolutionary diversity of FAANG over time.

It is expected that additional insights will be gained by expanding the genetic diversity within a given species. This fine-scale detail will provide invaluable insight into genetic regulation of phenotypic diversity at a mechanistic level. Furthermore, additional samples and species relevant to specific groups will be collected. New samples may include rumen tissues from ruminant species, mammary tissue from mammals and fiber-producing tissue in animals raised for fiber production. Many aquatic species are able to produce interesting atypical progeny (double haploid and sex-reversed progeny) and both poultry and aquatic species produce very large full-sibling cohorts.

**Impact of FAANG**

Similar to the ENCODE projects, the FAANG functional maps will generate a comprehensive data resource to be used by multiple groups, over a long time, for multiple purposes [42]. Thanks to this organized effort in coordination and standardization, individual research groups will be able to effectively use - and refer to - FAANG datasets, as well as contribute their own datasets from specific genome-to-phenome investigations in different species.

Overall, we predict completing the aims of the FAANG project will enable the application of molecular phenotypes to the prediction of complex phenotypes and further our understanding of additive and non-additive genetic mechanisms such as dominance and epistasis. Such knowledge can be applied to animal production, human and animal health, evolution, adaptation, and understanding the role of animals in their ecosystem. There is also evidence that early developmental influences can affect transiently inherited acquired traits, indicating that epigenetic modifications to the genome may be another important factor in understanding the inheritance of complex traits. FAANG will provide critical basic information, which will be used to improve food production and inform studies of agriculture, biomedical science, evolution and the environment.

## Additional file

**Additional file 1: Table S1.** Reference species and publications of livestock genomes.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

All authors are signatories of the FAANG Consortium; they have contributed to its conception and in drafting of the manuscript. ALA proposed the initial concept and framework for FAANG. ALA, EG, MAG and CKT finalized the manuscript. EG and CKT developed the FAANG web portal and submitted the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

We recognize the role of the EC-US Biotechnology Research Task Force and the Animal Biotechnology Working Group in providing a forum for the initial discussions that led to this paper and the FAANG Consortium.

### Author details

<sup>1</sup>Science for Life Laboratory Uppsala, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala SE 751 23, Sweden. <sup>2</sup>Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala SE-750 07, Sweden. <sup>3</sup>The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh EH25 9RG, UK. <sup>4</sup>School of Animal and Veterinary Sciences, The University of Adelaide, Roseworthy, SA 5371, Australia. <sup>5</sup>Invermay Agricultural Centre, AgResearch Limited, Mosgiel 9053, New Zealand. <sup>6</sup>College of Agriculture and Life Sciences, University of Arizona, Tucson, AZ 85719, USA. <sup>7</sup>National Animal Disease Center, United States Department of Agriculture, Agricultural Research Service, Ames, IA 50010, USA. <sup>8</sup>Avian Disease and Oncology Laboratory, United States Department of Agriculture, Agricultural Research Service, East Lansing, MI 48823, USA. <sup>9</sup>European Molecular Biology Laboratory European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK. <sup>10</sup>Livestock Improvement Corporation, Hamilton 3284, New Zealand. <sup>11</sup>Agriculture Flagship, Commonwealth Scientific and Industrial Research Organisation, St Lucia 4067 Brisbane, Australia. <sup>12</sup>Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA. <sup>13</sup>UMR1388 Génétique, Physiologie et Systèmes d'Élevage (GenPhySE), French National Institute for Agricultural Research (INRA), F-31326 Castanet-Tolosan, France. <sup>14</sup>UMR1313 Génétique Animale et Biologie Intégrative (GABI), French National Institute for Agricultural Research (INRA), F-78352 Jouy-en-Josas, France. <sup>15</sup>Animal Breeding and Genomics Centre, Wageningen University, 6708 PB Wageningen, The Netherlands. <sup>16</sup>Biosciences Research Division, Department of Environment and Primary Industries Victoria, Bundoora 3083, Australia. <sup>17</sup>Dairy Futures Cooperative Research Centre, Bundoora 3083, VIC, Australia. <sup>18</sup>La Trobe University, Bundoora 3086, VIC, Australia. <sup>19</sup>Key Laboratory for Animal Biotechnology of Jiangxi Province and the Ministry of Agriculture of China, Jiangxi Agricultural University, Jiangxi 330029, People's Republic of China. <sup>20</sup>Department of Animal Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA. <sup>21</sup>Department of Agricultural Biotechnology, Seoul National University, Seoul, Republic of Korea. <sup>22</sup>Animal Parasitic Diseases Laboratory, United States Department of Agriculture, Agricultural Research Service, Beltsville, MD 20705, USA. <sup>23</sup>School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ 85721, USA. <sup>24</sup>Animal Productivity Group, AgResearch Limited, Mosgiel 9053, New Zealand. <sup>25</sup>Centre for Animal Science, Queensland Alliance for Agricultural & Food Innovation, University of Queensland, St Lucia, QLD 4067, Australia. <sup>26</sup>Department of Basic Sciences, College of Veterinary Medicine and Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi, MS 39762, USA. <sup>27</sup>Comparative Bioinformatics, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain. <sup>28</sup>National Center for Cool and Cold Water Aquaculture, United States Department of Agriculture, Agricultural Research Service, Kearneysville, WV 25430, USA. <sup>29</sup>Livestock Gentec Centre, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta T6G 2C8, Canada. <sup>30</sup>Department of Animal Science, Iowa State University, Ames, IA 50011, USA. <sup>31</sup>US Meat Animal Research Center, United States Department of Agriculture, Agricultural Research Service, Clay Center, NE 68933, USA. <sup>32</sup>Institute of Marine Biology, Biotechnology and Aquaculture,

Hellenic Centre for Marine Research, 71003 Heraklion, Greece. <sup>33</sup>Department of Animal and Food Sciences, University of Delaware, Newark, DE 19716, USA. <sup>34</sup>Animal Production and Protection, United States Department of Agriculture, Agricultural Research Service Aquaculture, Beltsville, MD 20705, USA. <sup>35</sup>CSIRO Agriculture, Commonwealth Scientific and Industrial Research Organisation, St Lucia 4067, Australia. <sup>36</sup>Green Technology, Natural Resources Institute Finland, 31600 Jokioinen, Finland. <sup>37</sup>Animal Disease Research Unit, United States Department of Agriculture, Agricultural Research Service, Pullman, WA 99164-6630, USA. <sup>38</sup>Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA 99164-7040, USA. <sup>39</sup>Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education of China, Huazhong Agricultural University, Wuhan, Hubei 430070, People's Republic of China. <sup>40</sup>Department of Animal Science, University of California, Davis, CA 95616, USA.

Published online: 25 March 2015

### References

1. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
2. Andersson L. Molecular consequences of animal breeding. *Curr Opin Genet Dev*. 2013;23:295–301.
3. Freking BA, Murphy SK, Wylie AA, Rhodes SJ, Keele JW, Leymaster KA, et al. Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res*. 2002;12:1496–506.
4. Clop A, Marcq F, Takeda H, Pirotin D, Tordo X, Bibé B, et al. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet*. 2006;38:813–8.
5. Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, et al. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature*. 2003;425:832–6.
6. Andersson LS, Larhammar M, Memic F, Wootz H, Schwochow D, Rubin CJ, et al. Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature*. 2012;488:642–6.
7. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res*. 2012;22:1748–59.
8. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. 2004;432:695–716.
9. Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*. 2012;491:393–8.
10. Elsik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, Weinstock GM, et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*. 2009;324:522–8.
11. Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*. 2014;344:1168–73.
12. Brawand D, Wagner CE, Li Yi, Malinsky M, Keller I, Fan S, et al. The genomic substrate for adaptive radiation in African cichlid fish. *Nature*. 2014;513:375–81.
13. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*. 2010;464:587–91.
14. Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alföldi J, Barrio AM, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science*. 2014;345:1074–9.
15. Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, et al. Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet*. 2014;10:e1004016.
16. Larkin DM, Daetwyler HD, Hernandez AG, Wright CL, Hetrick LA, Boucek L, et al. Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle. *Proc Natl Acad Sci U S A*. 2012;109:7693–8.
17. Rubin CJ, Megens HJ, Martinez Barrio A, Maqbool K, Sayyab S, Schwochow D, et al. Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci U S A*. 2012;109:19529–36.

18. Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci U S A*. 2014;111:E5661–9.
19. Guilbault C, Saeed Z, Downey GP, Radziuch D. Cystic fibrosis mouse models. *Am J Respir Cell Mol Biol*. 2007;36:1–7.
20. Devoy A, Bunton-Stasyshyn RK, Tybulewicz VL, Smith AJ, Fisher EM. Genomically humanized mice: technologies and promises. *Nat Rev Genet*. 2012;13:14–20.
21. Walters EM, Wolf E, Whyte JJ, Mao J, Renner S, Nagashima H. Completion of the swine genome will simplify the production of swine as a large animal biomedical model. *BMC Med Genomics*. 2012;5:55.
22. Hill WG. Applications of population genetics to animal breeding, from Wright, Fisher and Lush to genomic prediction. *Genetics*. 2014;196:1–16.
23. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157:1819–29.
24. Meuwissen T, Goddard M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics*. 2010;185:623–31.
25. Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brondum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet*. 2014;46:858–65.
26. MacLeod IM, Hayes BJ, Goddard ME. The effects of demography and long-term selection on the accuracy of genomic prediction with sequence data. *Genetics*. 2014;198:1671–84.
27. Koufariotis L, Chen YP, Bolormaa S, Hayes BJ. Regulatory and coding genome regions are enriched for trait associated variants in dairy and beef cattle. *BMC Genomics*. 2014;15:436.
28. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. A map of the cis-regulatory sequences in the mouse genome. *Nature*. 2012;488:116–20.
29. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature*. 2014;515:355–64.
30. Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, et al. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science*. 2010;330:1787–97.
31. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, et al. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science*. 2010;330:1775–87.
32. Sivasubbu S, Sachidanandan C, Scaria V. Time for the zebrafish ENCODE. *J Genet*. 2013;92:695–701.
33. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, et al. The evolutionary landscape of alternative splicing in vertebrate species. *Science*. 2012;338:1587–93.
34. Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*. 2010;328:1036–40.
35. Tagu D, Colbourne JK, Nègre N. Genomic data integration for ecological and evolutionary traits in non-model organisms. *BMC Genomics*. 2014;15:490.
36. Bae JB. Perspectives of international human epigenome consortium. *Genomics Inform*. 2013;11:7–14.
37. Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, Rogers J, et al. Prepublication data sharing. *Nature*. 2009;461:168–70.
38. The FAANG Consortium. <http://www.faanng.org>
39. Stamatoyannopoulos JA. What does our genome encode? *Genome Res*. 2012;22:1602–11.
40. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*. 2012;22:1813–31.
41. Birney E. The making of ENCODE: lessons for big-data projects. *Nature*. 2012;489:49–51.
42. Eddy SR. The ENCODE project: missteps overshadowing a success. *Curr Biol*. 2013;23:R259–61.
43. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitch S, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res*. 2009;37:e123.
44. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22:1775–89.
45. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012;22:1760–74.
46. Mudge JM, Frankish A, Harrow J. Functional transcriptomics in the post-ENCODE era. *Genome Res*. 2013;23:1961–73.
47. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012;489:75–82.
48. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013;10:1213–8.
49. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol*. 2015;109:21.29.1–9.
50. Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*. 2014;15:234–46.
51. Ho JW, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, et al. Comparative analysis of metazoan chromatin organization. *Nature*. 2014;512:449–52.
52. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res*. 2005;33:5868–77.
53. van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, et al. Hi-C: a method to study the three-dimensional architecture of genomes. *J Vis Exp* 2010;39:pii:1869.
54. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159:1665–80.