ROSLIN

Improving and adding value to the pig genome sequence

Alan Archibald
The Roslin Institute and R(D)SVS
University of Edinburgh

THE UNIVERSITY *of* EDINBURGH

BBSRC
bioscience for the future

Acknowledgements

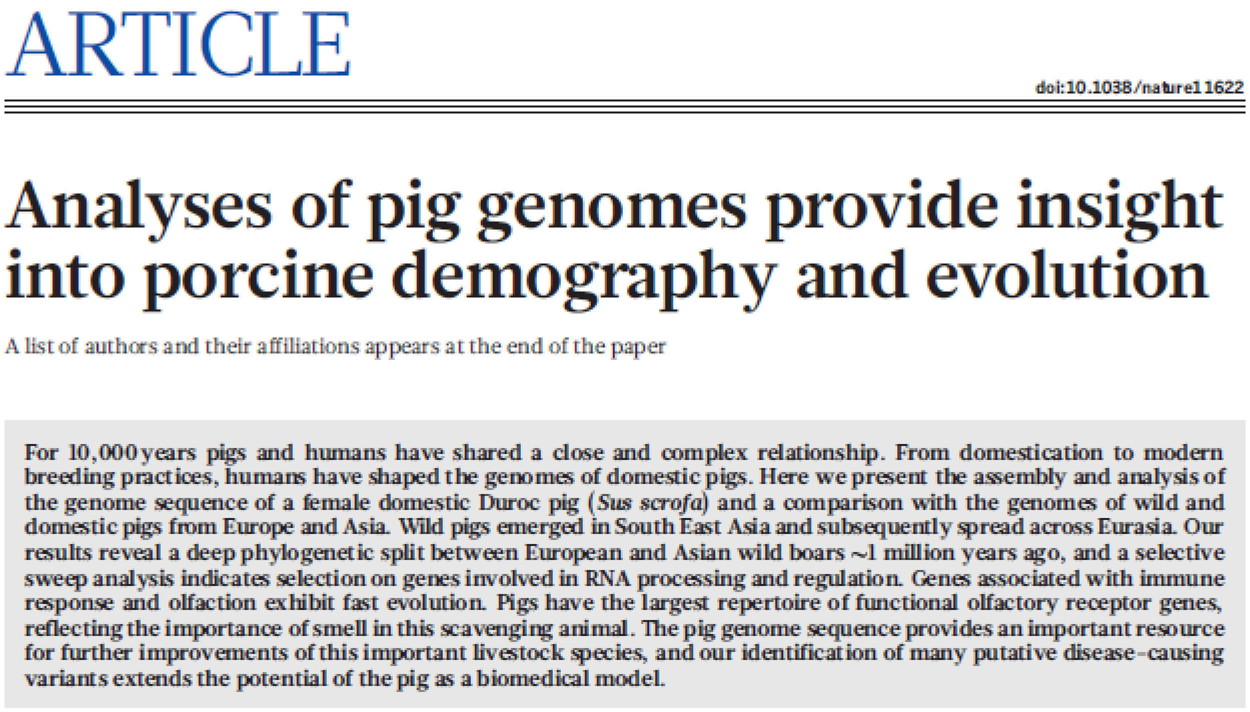
- Lel Eory (Roslin)
- Paul Flicek (EBI)
- Tim Hubbard, Steve Searle (Sanger)
- Martien Groenen, Larry Schook, SGSC

- many others
- multiple funding organisations
  - (see Nature paper)

A sequenced genome is a requirement for 21st Century biological research

Pig genome sequencing project

Swine Genome Sequencing Consortium


ARTICLE

doi:10.1038/nature11622

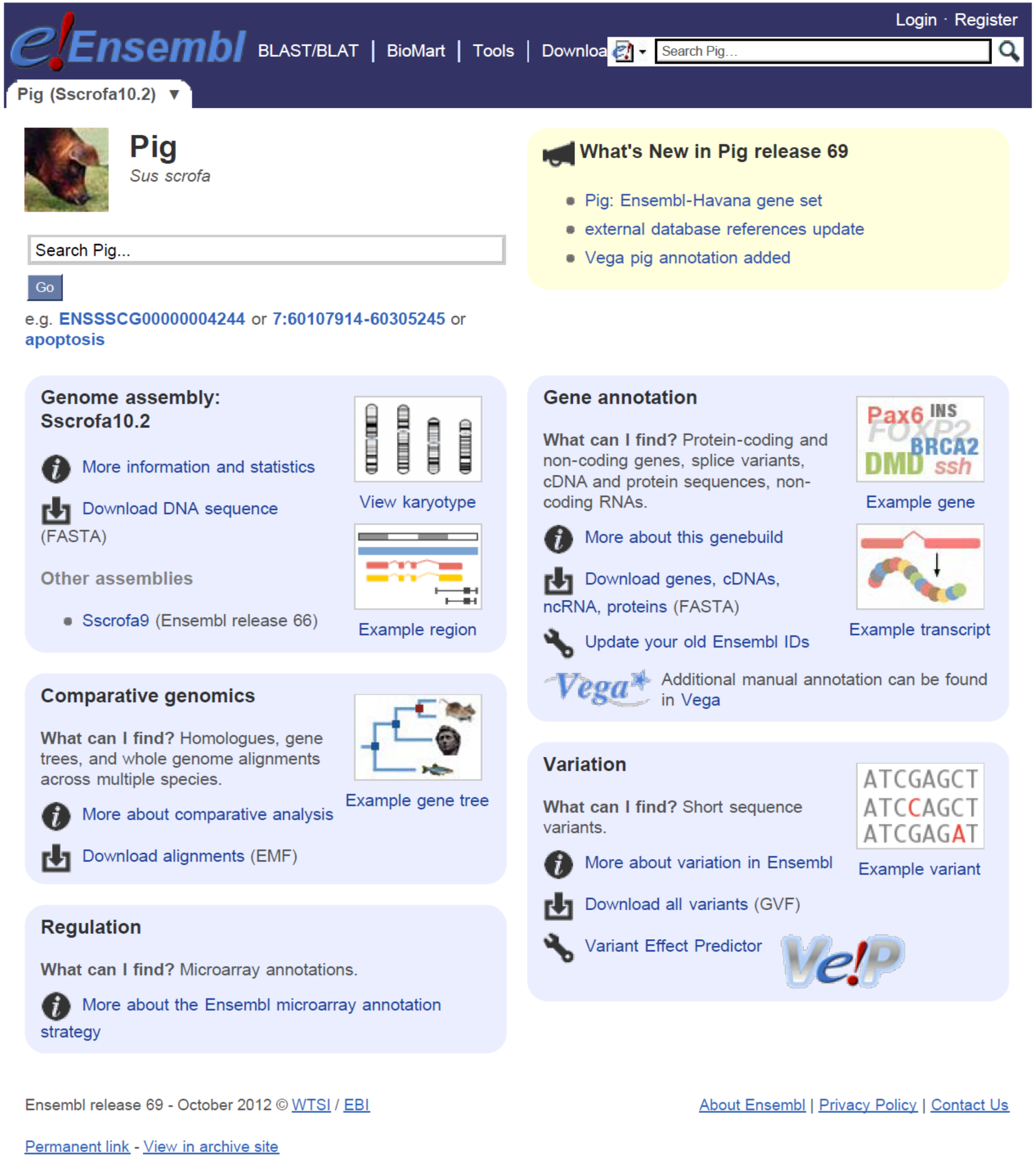
Analyses of pig genomes provide insight into porcine demography and evolution

A list of authors and their affiliations appears at the end of the paper

For 10,000 years pigs and humans have shared a close and complex relationship. From domestication to modern breeding practices, humans have shaped the genomes of domestic pigs. Here we present the assembly and analysis of the genome sequence of a female domestic Duroc pig (*Sus scrofa*) and a comparison with the genomes of wild and domestic pigs from Europe and Asia. Wild pigs emerged in South East Asia and subsequently spread across Eurasia. Our results reveal a deep phylogenetic split between European and Asian wild boars ~1 million years ago, and a selective sweep analysis indicates selection on genes involved in RNA processing and regulation. Genes associated with immune response and olfaction exhibit fast evolution. Pigs have the largest repertoire of functional olfactory receptor genes, reflecting the importance of smell in this scavenging animal. The pig genome sequence provides an important resource for further improvements of this important livestock species, and our identification of many putative disease-causing variants extends the potential of the pig as a biomedical model.

Pig genomes provide insight into porcine demography, domestication and evolution

Martien A.M. Groenen[1], Alan L. Archibald[2*], Hirohide Uenishi[3], Christopher K. Tuggle[4], Yasu Takeuchi[5], Max F. Rothschild[4], Claire Rogel-Gaillard[6], Chankyu Park[7], Denis Milan[8], Hendrik-Jan Megens[1], Shengting Li[9], Denis Larkin[10], Heebal Kim[11], Laurent A. F. Frantz[1], Mario Caccamo[12], Hyeonju Ahn[11], Bronwen L. Aken[13], Anna Anselmo[14], Christian Anthon[15], Loretta Auvil[16], Bouabid Badaoui[14], Craig W. Beattie[17], Christian Bendixen[18], Daniel Berman[19], Frank Blecha[20], Jonas Blomberg[21], Lars Bolund[9], Mirte Bosse[1], Sara Botti[14], Zhan Bujie[18], Megan Bystrom[4], Boris Capitanu[16], Denise Carvalho-Silva[22], Patrick Chardon[6], Celine Chen[24], Ryan Cheng[4], Sang-Haeng Choi[25], William Chow[13], Richard C. Clark[13], Christopher Clee[13], Richard P.M.A. Crooijmans[1], Harry D. Dawson[24], Patrice Dehais[8], Floravante De Sapio[2], Bert Dibbits[1], Nizar Drou[12], Zhi-Qiang Du[4], Kellye Eversole[26], João Fadista[18], Susan Fairley[13], Thomas Faraut[8], Geoffrey J. Faulkner[2], Katie E. Fowler[27], Merete Fredholm[15], Eric Fritz[4], James G.R. Gilbert[13], Elisabetta Giuffra[14], Jan Gorodkin[15], Darren K. Griffin[27], Jennifer L. Harrow[13], Alexander Hayward[28], Kerstin Howe[13], Zhi-Liang Hu[4], Sean J. Humphray[13], Toby Hunt[13], Henrik H. Jensen[18], Patric Jern[28], Matthew Jones[13], Jerzy Jurka[29], Hiroyuki Kanamori[30], Ronan Kapetanovic[2], Jaebum Kim[31,23], Jae-Hwan Kim[32], Kyu-Won Kim[33], Tae-Hun Kim[34], Greger Larson[35], Kyooyeol Lee[7], Kyung-Tai Lee[34], Richard Leggett[12], Harris A. Lewin[36], Yingrui Li[9], Wansheng Liu[37], Jane E. Loveland[13], Yao Lu[9], Joan K. Lunney[19], Jian Ma[38], Ole Madsen[1], Katherine Mann[19], Lucy Matthews[13], Stuart McLaren[13], Takeya Morozumi[30], Michael Murtaugh[39], Jitendra Narayan[10], Dinh Truong Nguyen[7], Peixiang Ni[9], Song-Jung Oh[40], Suneel Onteru[4], Frank Panitz[18], Eung-Woo Park[34], Hong-Seog Park[25], Geraldine Pascal[41], Yogesh Paudel[1], Miguel Perez-Enciso[42], Ricardo Ramirez-Gonzalez[12], James M. Reecy[4], Sandra Rodriguez-Zas[43], Gary A. Rohrer[44], Lauretta Rund[43], Yongming Sang[20], Kyle Schachtschneider[43], Joshua Schraiber[45], John Schwartz[39], Linda Scobie[46], Carol Scott[13], Stephen Searle[13], Bertrand Servin[8], Bruce R. Southey[43], Goran Sperber[47], Peter Stadler[48], Jonathan Sweedler[49], Hakim Tafer[48], Bo Thomsen[18], Rashmi Wali[46], Jian Wang[9], Jun Wang[9], Simon White[13], Xun Xu[9], Martine Yerle[8], Jianguo Zhang[9], Guojie Zhang[9], Jie Zhang[50], Shuhong Zhao[50], Jane Rogers[12], Carol Churcher[13] and Lawrence B. Schook[51].*



Pig

Sus scrofa

e.g. **ENSSSCG00000004244** or **7:60107914-60305245** or **apoptosis**

What's New in Pig release 69

- Pig: Ensembl-Havana gene set
- external database references update
- Vega pig annotation added

Genome assembly: Sscrofa10.2

More information and statistics

Download DNA sequence (FASTA)

View karyotype

Example region

Other assemblies

- Sscrofa9 (Ensembl release 66)

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

Example gene tree

More about comparative analysis

Download alignments (EMF)

Regulation

What can I find? Microarray annotations.

More about the Ensembl microarray annotation strategy

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

Example gene

More about this genebuild

Download genes, cDNAs, ncRNA, proteins (FASTA)

Example transcript

Update your old Ensembl IDs

Additional manual annotation can be found in Vega

Variation

What can I find? Short sequence variants.

Example variant

More about variation in Ensembl

Download all variants (GVF)

Variant Effect Predictor

Ensembl release 69 - October 2012 © WTSI / EBI

About Ensembl | Privacy Policy | Contact Us

Permanent link - View in archive site

http://www.ensembl.org/Sus_scrofa/Info/Index

Adding value – companion papers

BioMed Central
The Open Access Publisher

BMC series blog

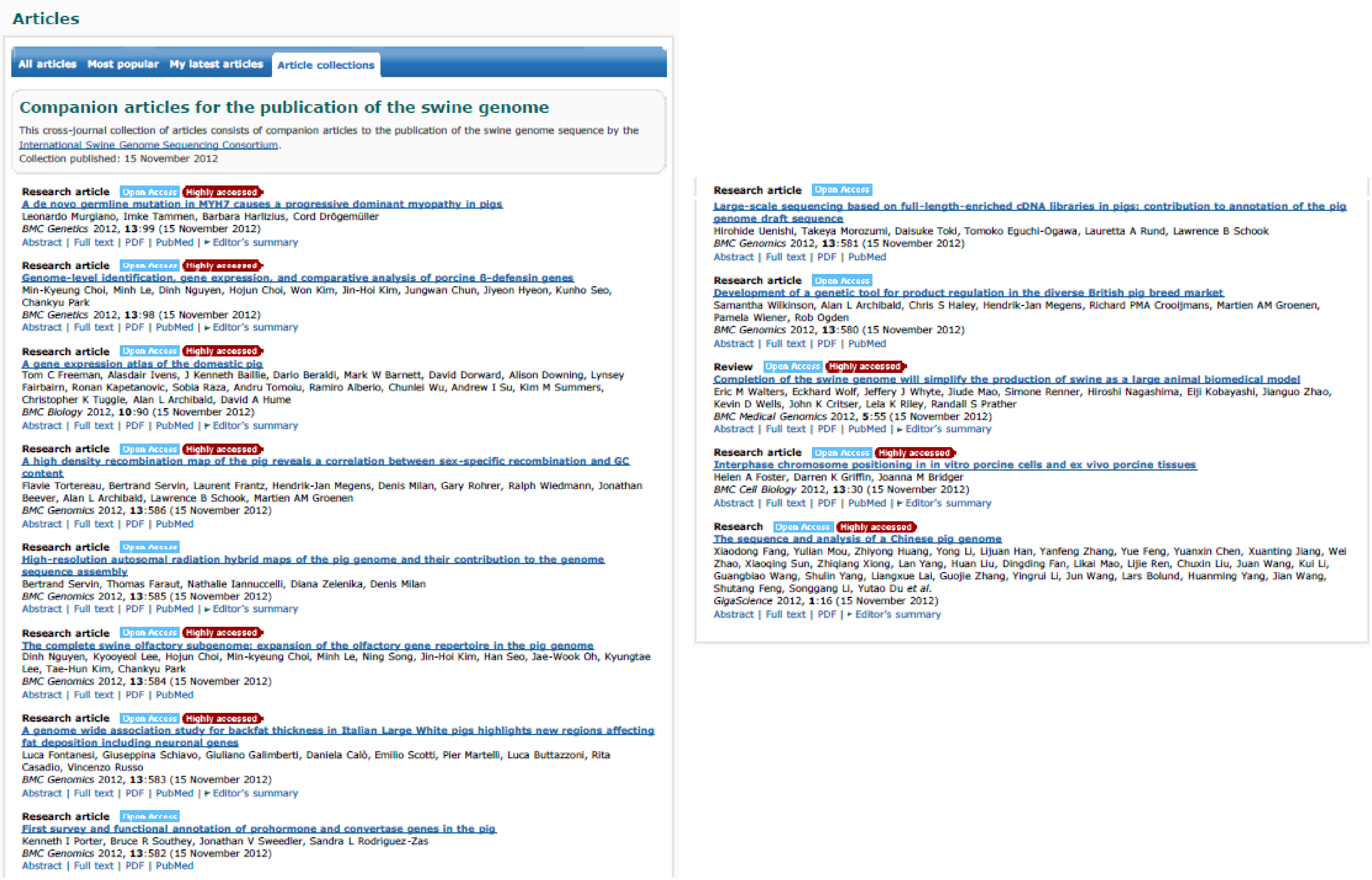
BMC series publishes companion papers to the pig genome sequence

Tim Sands on November 15, 2012 at 3:02 pm - 0 Comments

The BMC series is delighted to host a series of companion articles to the publication of the swine genome.

The articles are published in a number of BMC series journals and the journal *GigaScience* are gathered on a cross-journal article series page. The papers cover a broad spectrum of subjects related to the pig genome: more details of the content is given in our blog, "Companion articles to the swine genome sequence".

Image credit: David Merrett on Flickr, CC BY 2.0



Companion articles for the publication of the swine genome

This cross-journal collection of articles consists of companion articles to the publication of the swine genome sequence by the International Swine Genome Sequencing Consortium.

Collection published: 15 November 2012

Research article Open Access Highly accessed
A de novo germline mutation in MYH7 causes a progressive dominant myopathy in pigs
Leonardo Murgiano, Imke Tammen, Barbara Harlizius, Cord Drögemüller
BMC Genetics 2012, **13**:99 (15 November 2012)
Abstract | Full text | PDF | PubMed | ▸ Editor's summary

Research article Open Access Highly accessed
Genome-level identification, gene expression, and comparative analysis of porcine ß-defensin genes
Min-Kyeung Choi, Minh Le, Dinh Nguyen, Hojun Choi, Won Kim, Jin-Hoi Kim, Jungwan Chun, Jiyeon Hyeon, Kunho Seo, Chankyu Park
BMC Genetics 2012, **13**:98 (15 November 2012)
Abstract | Full text | PDF | PubMed | ▸ Editor's summary

Research article Open Access Highly accessed
A gene expression atlas of the domestic pig
Tom C Freeman, Alasdair Ivens, J Kenneth Baillie, Dario Beraldi, Mark W Barnett, David Dorward, Alison Downing, Lynsey Fairbairn, Ronan Kapetanovic, Sobia Raza, Andru Tomoiu, Ramiro Alberio, Chunlei Wu, Andrew I Su, Kim M Summers, Christopher K Tuggle, Alan L Archibald, David A Hume
BMC Biology 2012, **10**:90 (15 November 2012)
Abstract | Full text | PDF | PubMed | ▸ Editor's summary

Research article Open Access Highly accessed
A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content
Flavie Tortereau, Bertrand Servin, Laurent Frantz, Hendrik-Jan Megens, Denis Milan, Gary Rohrer, Ralph Wiedmann, Jonathan Beever, Alan L Archibald, Lawrence B Schook, Martien AM Groenen
BMC Genomics 2012, **13**:586 (15 November 2012)
Abstract | Full text | PDF | PubMed

Research article Open Access
High-resolution autosomal radiation hybrid maps of the pig genome and their contribution to the genome sequence assembly
Bertrand Servin, Thomas Faraut, Nathalie Iannuccelli, Diana Zelenika, Denis Milan
BMC Genomics 2012, **13**:585 (15 November 2012)
Abstract | Full text | PDF | PubMed | ▸ Editor's summary

Research article Open Access Highly accessed
The complete swine olfactory subgenome: expansion of the olfactory gene repertoire in the pig genome
Dinh Nguyen, Kyooyeol Lee, Hojun Choi, Min-kyeung Choi, Minh Le, Ning Song, Jin-Hoi Kim, Han Seo, Jae-Wook Oh, Kyungtae Lee, Tae-Hun Kim, Chankyu Park
BMC Genomics 2012, **13**:584 (15 November 2012)
Abstract | Full text | PDF | PubMed

Research article Open Access Highly accessed
A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes
Luca Fontanesi, Giuseppina Schiavo, Giuliano Galimberti, Daniela Calò, Emilio Scotti, Pier Martelli, Luca Buttazzoni, Rita Casadio, Vincenzo Russo
BMC Genomics 2012, **13**:583 (15 November 2012)
Abstract | Full text | PDF | PubMed | ▸ Editor's summary

Research article Open Access
First survey and functional annotation of prohormone and convertase genes in the pig
Kenneth I Porter, Bruce R Southey, Jonathan V Sweedler, Sandra L Rodriguez-Zas
BMC Genomics 2012, **13**:582 (15 November 2012)
Abstract | Full text | PDF | PubMed

Research article Open Access
Large-scale sequencing based on full-length-enriched cDNA libraries in pigs: contribution to annotation of the pig genome draft sequence
Hirohide Uenishi, Takeya Morozumi, Daisuke Toki, Tomoko Eguchi-Ogawa, Lauretta A Rund, Lawrence B Schook
BMC Genomics 2012, **13**:581 (15 November 2012)
Abstract | Full text | PDF | PubMed

Research article Open Access
Development of a genetic tool for product regulation in the diverse British pig breed market
Samantha Wilkinson, Alan L Archibald, Chris S Haley, Hendrik-Jan Megens, Richard PMA Crooijmans, Martien AM Groenen, Pamela Wiener, Rob Ogden
BMC Genomics 2012, **13**:580 (15 November 2012)
Abstract | Full text | PDF | PubMed

Review Open Access Highly accessed
Completion of the swine genome will simplify the production of swine as a large animal biomedical model
Eric M Walters, Eckhard Wolf, Jeffery J Whyte, Jiude Mao, Simone Renner, Hiroshi Nagashima, Eiji Kobayashi, Jianguo Zhao, Kevin D Wells, John K Critser, Lela K Riley, Randall S Prather
BMC Medical Genomics 2012, **5**:55 (15 November 2012)
Abstract | Full text | PDF | PubMed | ▸ Editor's summary

Research article Open Access Highly accessed
Interphase chromosome positioning in in vitro porcine cells and ex vivo porcine tissues
Helen A Foster, Darren K Griffin, Joanna M Bridger
BMC Cell Biology 2012, **13**:30 (15 November 2012)
Abstract | Full text | PDF | PubMed | ▸ Editor's summary

Research Open Access Highly accessed
The sequence and analysis of a Chinese pig genome
Xiaodong Fang, Yulian Mou, Zhiyong Huang, Yong Li, Lijuan Han, Yanfeng Zhang, Yue Feng, Yuanxin Chen, Xuanting Jiang, Wei Zhao, Xiaoqing Sun, Zhiqiang Xiong, Lan Yang, Huan Liu, Dingding Fan, Likai Mao, Lijie Ren, Chuxin Liu, Juan Wang, Kui Li, Guangbiao Wang, Shulin Yang, Liangxue Lai, Guojie Zhang, Yingrui Li, Jun Wang, Lars Bolund, Huanming Yang, Jian Wang, Shutang Feng, Songgang Li, Yutao Du *et al.*
GigaScience 2012, **1**:16 (15 November 2012)
Abstract | Full text | PDF | ▸ Editor's summary

PNAS

INAUGURAL ARTICLE

Strong signatures of selection in the domestic pig genome

Carl-Johan Rubin[a,1], Hendrik-Jan Megens[b,1], Alvaro Martinez Barrio[a], Khurram Maqbool[c], Shumaila Sayyab[c], Doreen Schwochow[c], Chao Wang[a], Örjan Carlborg[d], Patric Jern[a], Claus B. Jørgensen[e], Alan L. Archibald[f], Merete Fredholm[e], Martien A. M. Groenen[b], and Leif Andersson[a,c,2]

[a]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, SE-751 23 Uppsala, Sweden; [b]Animal Breeding and Genomics Centre, Wageningen University, 6708 WD, Wageningen, The Netherlands; Departments of [c]Animal Breeding and Genetics and [d]Clinical Sciences, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden; [e]Department of Veterinary Clinical and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, DK-1165 Copenhagen, Denmark; and [f]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian EH25 9RG, United Kingdom

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2012.

Structural and functional annotation of the porcine immunome.
Dawson, H.D., Loveland, J.E., Pascal, G., Gilbert, J.G., Uenishi, H., Mann, K.M., Sang, Y., Zhang, J., Carvalho-Silva, D., Hunt, T., Hardy, M., Hu, Z., Zhao, S.-H., Anselmo, A., Shinkai, H., Chen, C., Badaoui, B., Berman, D., Amid, C., Kay, M., Lloyd, D., Snow, C., Morozumi, T., Cheng, R.P., Bystrom, M., Kapetanovic, R., Schwartz, J.C., Kataria, R., Astley, M., Fritz, E., Steward, C., Thomas, M., Wilming, L., Toki, D., Archibald, A.L., Bed'hom, B., Beraldi, D., Huang, T.H., Ait-Ali, T., Blecha, F., Botti, S., Freeman, T.C., Giuffra, E., Hume, D.A., Lunney, J.K., Murtaugh, M.P., Reecy, J.M., Harrow, J.L., Rogel-Gaillard, C. and Tuggle, C.K., 2013
BMC Genomics **14**: 332

Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*.
Frantz, L.A.F., Schraiber, J.G., Madsen, O., Megens, H.-J., Bosse, M., Paudel, Y., Semiadi, G., Meijaard, E., Li, N., Crooijmans, R.P.M.A., Archibald, A.L., Slatkin, M., Schook, L.B., Larson, G. and Groenen, M.A.M., *Genome Biology* in press

Adding value – sequence improvement

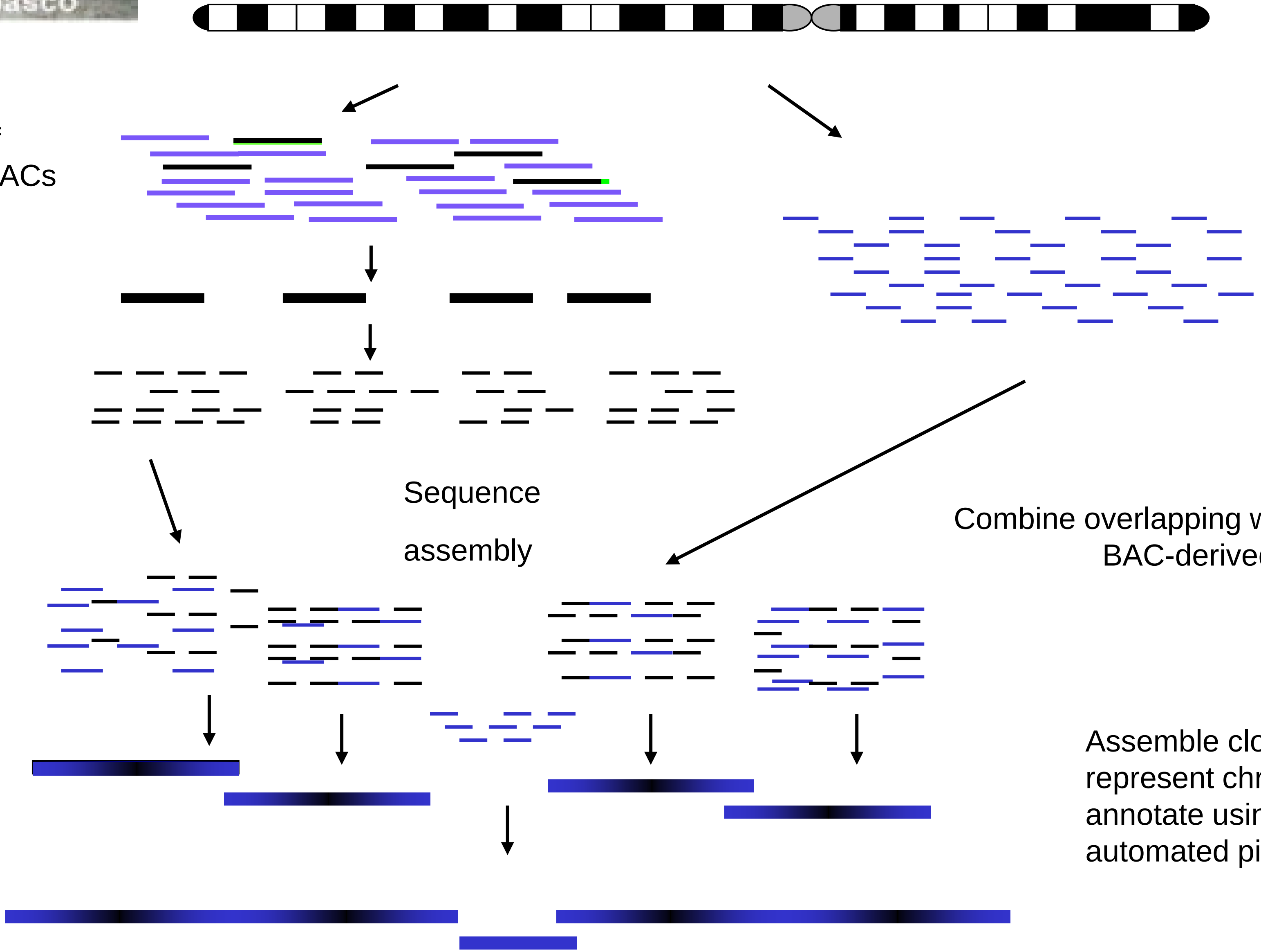
Sscrofa10.2 – assembly, genes

Assembly			**Annotation***	
	Placed	**Unplaced**		
Total length	**2,596,639,456**	**211,869,922**	**Protein-coding genes:**	**21,627**
Ungapped length	**2,323,671,356**	**195,490,322**	**Pseudogenes:**	**380**
Scaffolds	**5,343**	**4,562**	**ncRNA genes**:**	**2,965**
Contigs	**73,524**	**168,358**	**Gene exons:**	**197,675**
Scaffold N50	**637,332**	**98,022**	**Gene transcripts:**	**26,487**
Contig N50	**80,720**	**2,423**		

Sequence improvement

	Sequence coverage	Sequence accuracy	Number of sequence contigs	Average contig length bp	Genes with incorrect structures
4x draft sequence	*97% (~80Mb missing)*	*99.99% accurate (1 error in10kbp)*	*160,000 (7k / chr)*	*17,000*	*30%*
Sscrofa10.2 (placed)	***92%?***	***?***	**73,524**	**31,604**	***?***
Improved draft sequence	*99% (~32Mb missing)*	*99.99% accurate (1 error in10kbp)*	*65,000 (3k / chr)*	*42,000*	*5%*
Gold standard finished sequence	*99.9% (~5Mb missing)*	*99.999% (1 error in 100kbp)*	*<200*	*14,000,000*	*0%*

Hybrid Shotgun Sequencing Strategy



Minimal set of overlapping BACs selected from physical map

Whole- genome shotgun reads

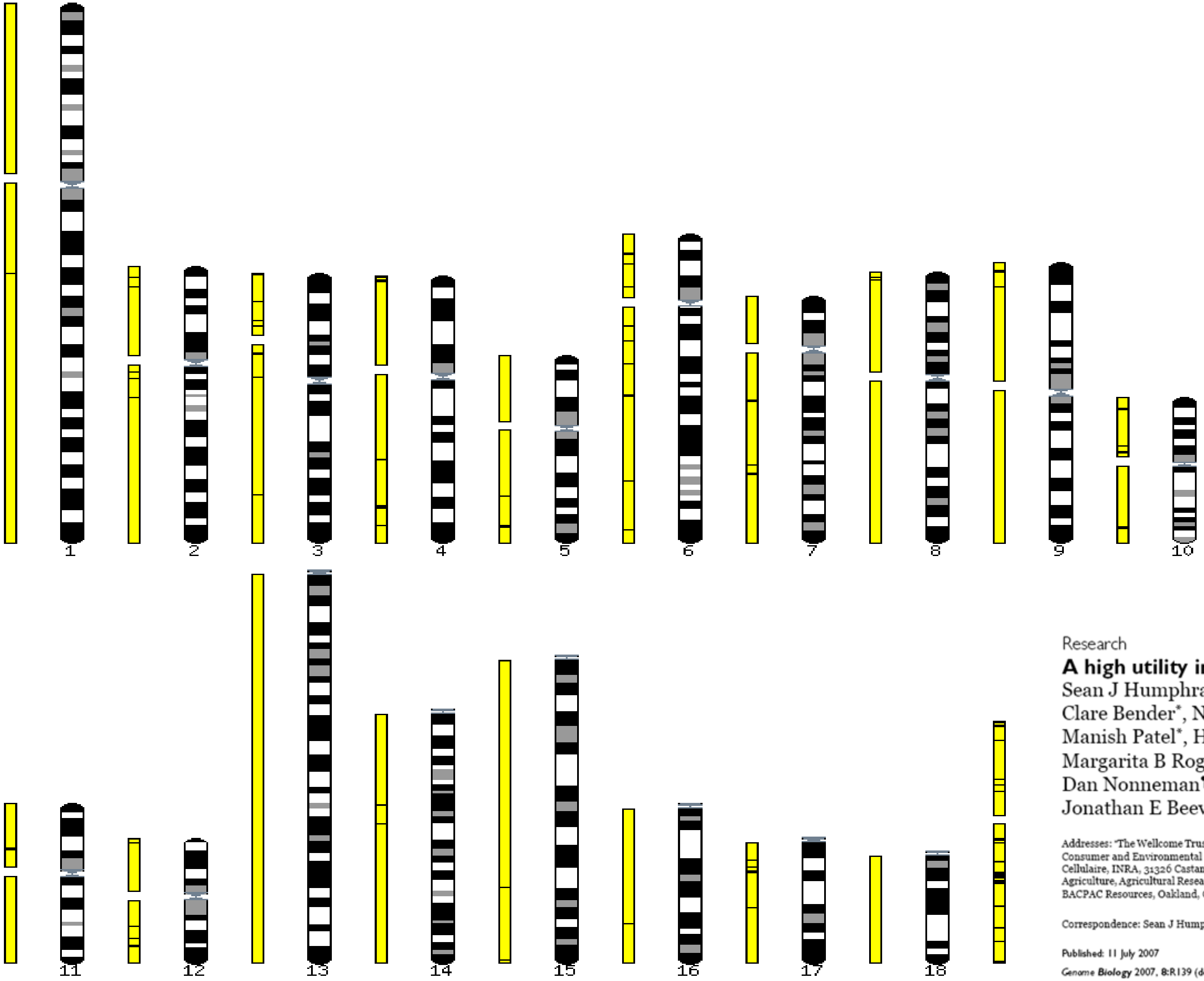
BAC shotgun reads

Sequence assembly

Combine overlapping whole-genome and BAC-derived reads

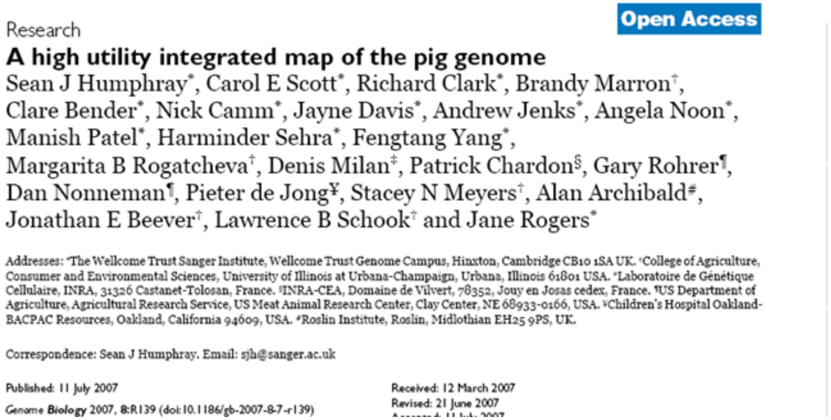
Assemble clone sequences to represent chromosomes and annotate using Ensembl automated pipeline

The best clone-based physcial map of a mammal



- 172 placed contigs
- average length 15 Mb
- covering 2.58Gb
- >98% of euchromatin

Humphray *et al.*, 2007. *Genome Biology*


Research

Open Access

A high utility integrated map of the pig genome

Sean J Humphray[*], Carol E Scott[*], Richard Clark[*], Brandy Marron[†], Clare Bender[*], Nick Camm[*], Jayne Davis[*], Andrew Jenks[*], Angela Noon[*], Manish Patel[*], Harminder Sehra[*], Fengtang Yang[*], Margarita B Rogatcheva[†], Denis Milan[‡], Patrick Chardon[§], Gary Rohrer[¶], Dan Nonneman[¶], Pieter de Jong[¥], Stacey N Meyers[†], Alan Archibald[#], Jonathan E Beever[†], Lawrence B Schook[†] and Jane Rogers[*]

Addresses: [*]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA UK. [†]College of Agriculture, Consumer and Environmental Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801 USA. [‡]Laboratoire de Génétique Cellulaire, INRA, 31326 Castanet-Tolosan, France. [§]INRA-CEA, Domaine de Vilvert, 78352, Jouy en Josas cedex, France. [¶]US Department of Agriculture, Agricultural Research Service, US Meat Animal Research Center, Clay Center, NE 68933-0166, USA. [¥]Children's Hospital Oakland-BACPAC Resources, Oakland, California 94609, USA. [#]Roslin Institute, Roslin, Midlothian EH25 9PS, UK.

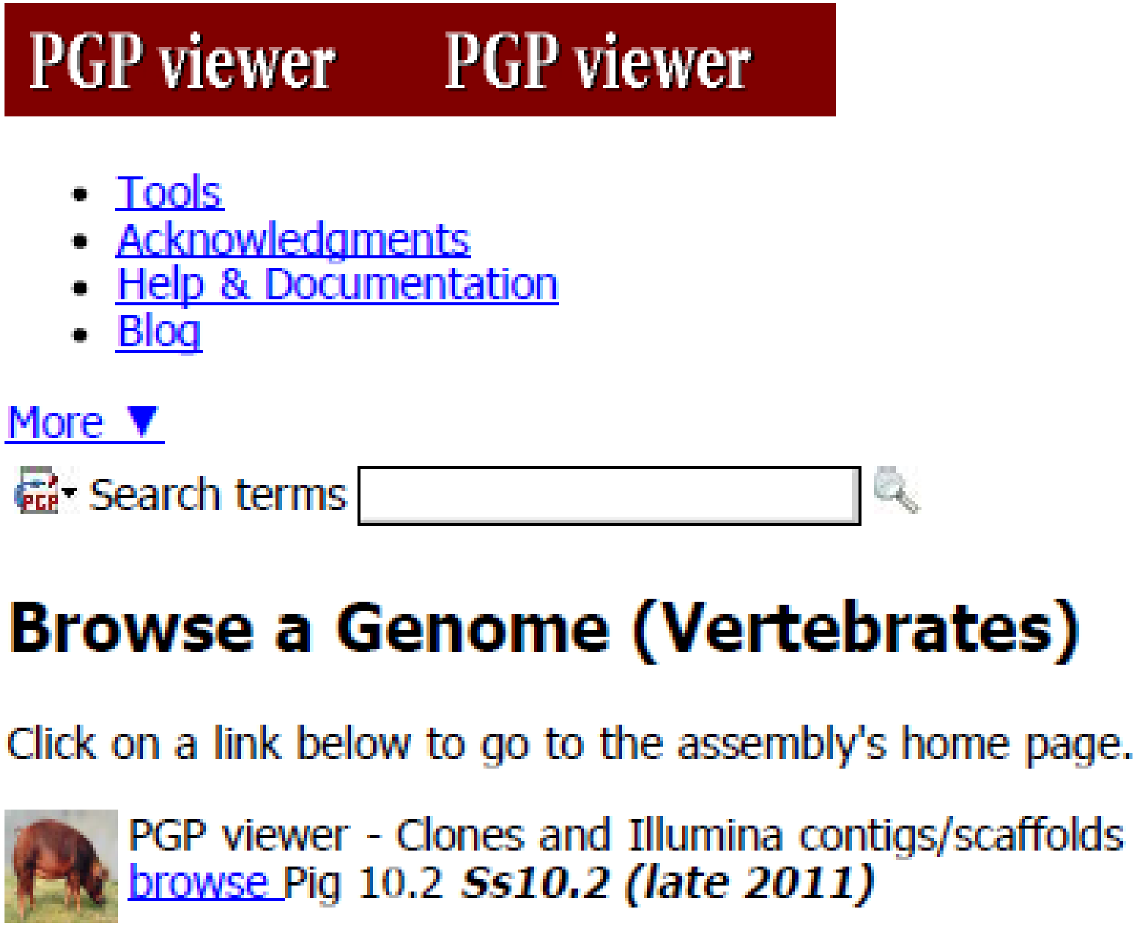
Correspondence: Sean J Humphray. Email: sjh@sanger.ac.uk

Published: 11 July 2007
*Genome **Biology*** 2007, 8:R139 (doi:10.1186/gb-2007-8-7-r139)

Received: 12 March 2007
Revised: 21 June 2007
Accepted: 11 July 2007

Improving the reference genome

- Additional data
  - SSCX, SSCY
  - ~300 CH242 BACs
    - Illumina reads (pools, individual clones)
  - Sanger WGS paired end reads
  - Illumina WGS paired end reads
  - Mate pair reads
    - CH242 BAC clones (~177 kbp)
    - Fosmid clones (~45 kbp)
    - Illumina (~6 kbp)



PGP viewer PGP viewer

- Tools
- Acknowledgments
- Help & Documentation
- Blog

More ▼

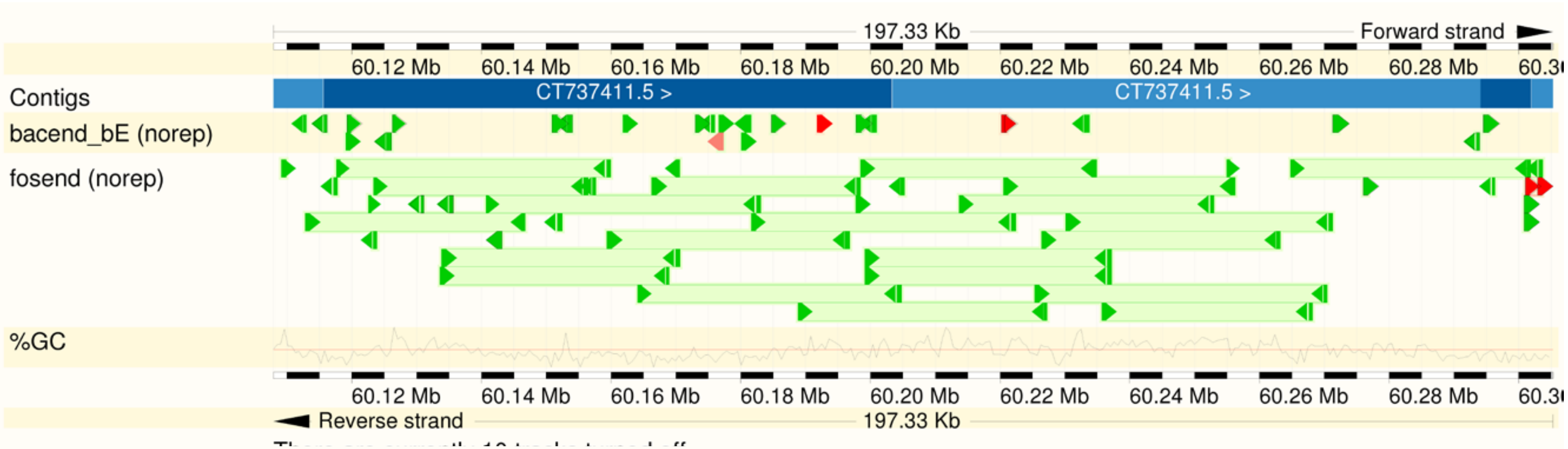
Search terms

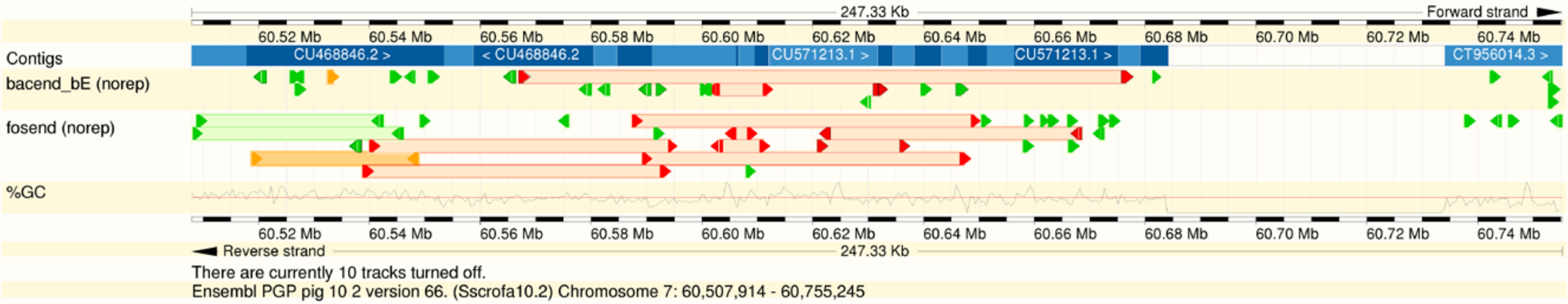
Browse a Genome (Vertebrates)

Click on a link below to go to the assembly's home page.

PGP viewer - Clones and Illumina contigs/scaffolds
browse Pig 10.2 ***Ss10.2 (late 2011)***

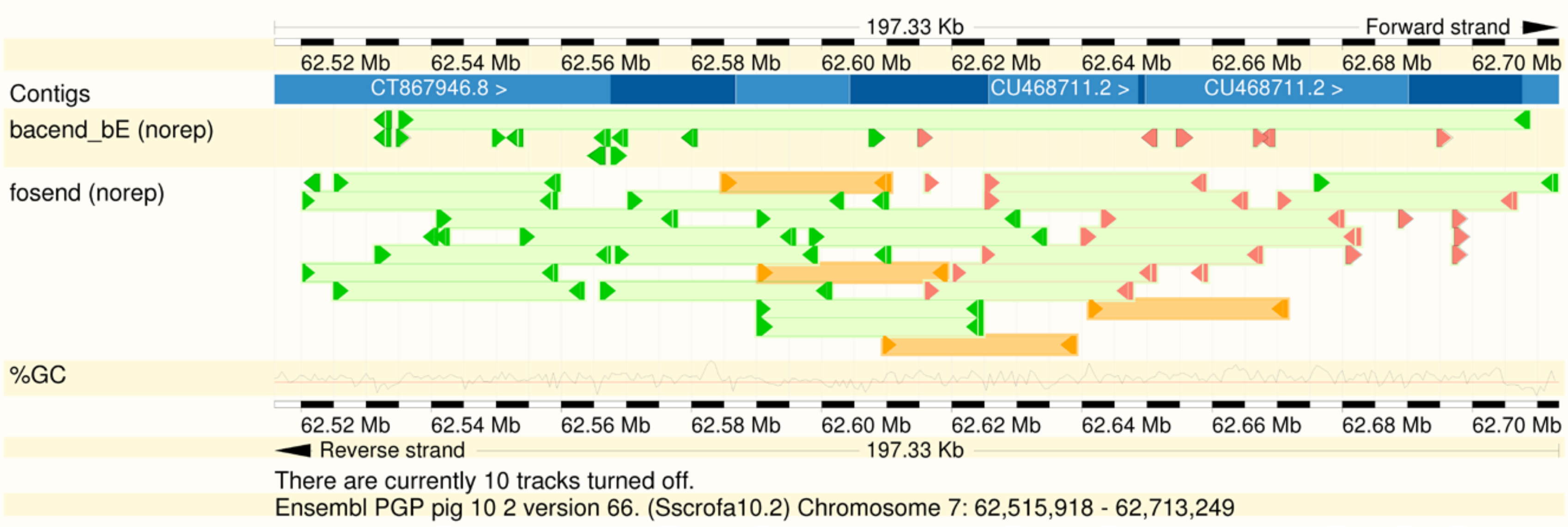
http://pgpviewer.ensembl.org





There are currently 10 tracks turned off.

Ensembl PGP pig 10 2 version 66. (Sscrofa10.2) Chromosome 7: 60,507,914 - 60,755,245

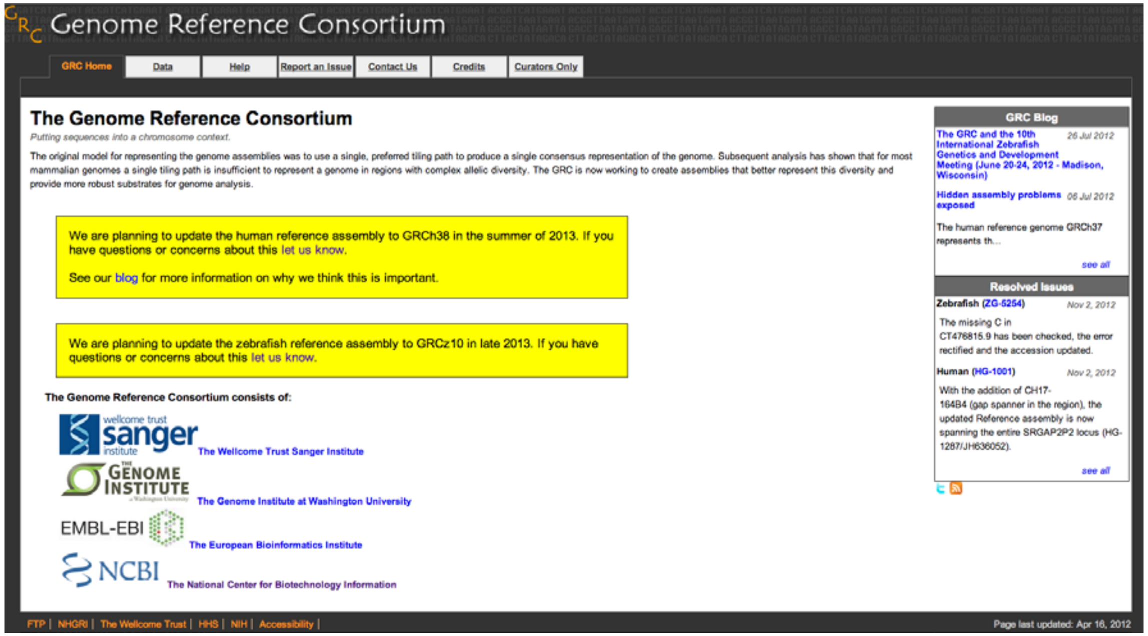


There are currently 10 tracks turned off.

Ensembl PGP pig 10 2 version 66. (Sscrofa10.2) Chromosome 7: 62,515,918 - 62,713,249

Improving a reference genome sequence

- Genome Reference Consortium
- Patches
  - Fix
  - Novel



The Genome Reference Consortium

Putting sequences into a chromosome context.

The original model for representing the genome assemblies was to use a single, prefered tiling path to produce a single consensus representation of the genome. Subsequent analysis has shown that for most mammalian genomes a single tiling path is insufficient to represent a genome in regions with complex allelic diversity. The GRC is now working to create assemblies that better represent this diversity and provide more robust substrates for genome analysis.

We are planning to update the human reference assembly to GRCh38 in the summer of 2013. If you have questions or concerns about this let us know.

See our blog for more information on why we think this is important.

We are planning to update the zebrafish reference assembly to GRCz10 in late 2013. If you have questions or concerns about this let us know.

The Genome Reference Consortium consists of:

- The Wellcome Trust Sanger Institute
- The Genome Institute at Washington University
- The European Bioinformatics Institute
- The National Center for Biotechnology Information

The GRC and the 10th International Zebrafish Genetics and Development Meeting (June 20-24, 2012 - Madison, Wisconsin) *26 Jul 2012*

Hidden assembly problems exposed *06 Jul 2012*

The human reference genome GRCh37 represents th...

Zebrafish (ZG-5254) *Nov 2, 2012*

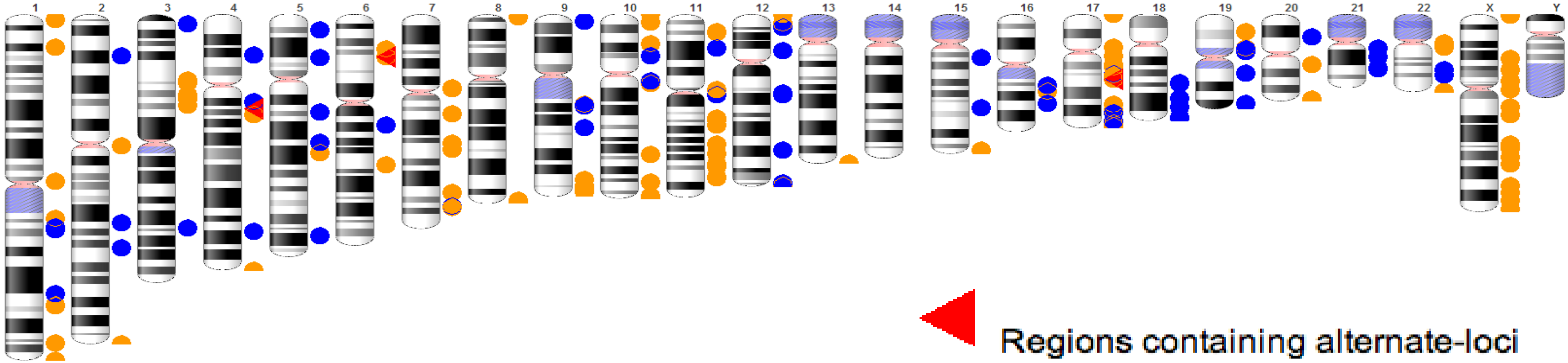
The missing C in CT476815.9 has been checked, the error rectified and the accession updated.

Human (HG-1001) *Nov 2, 2012*

With the addition of CH17-164B4 (gap spanner in the region), the updated Reference assembly is now spanning the entire SRGAP2P2 locus (HG-1287/JH636052).

Page last updated: Apr 16, 2012

http://genomereference.org



◀ Regions containing alternate-loci

● Regions containing fix patches

● Regions containing novel patches

GRCh37.p10
(160 regions: 2.89% of chromosomes)

111 Fix PATCHES: Chromosome update in GRCh38

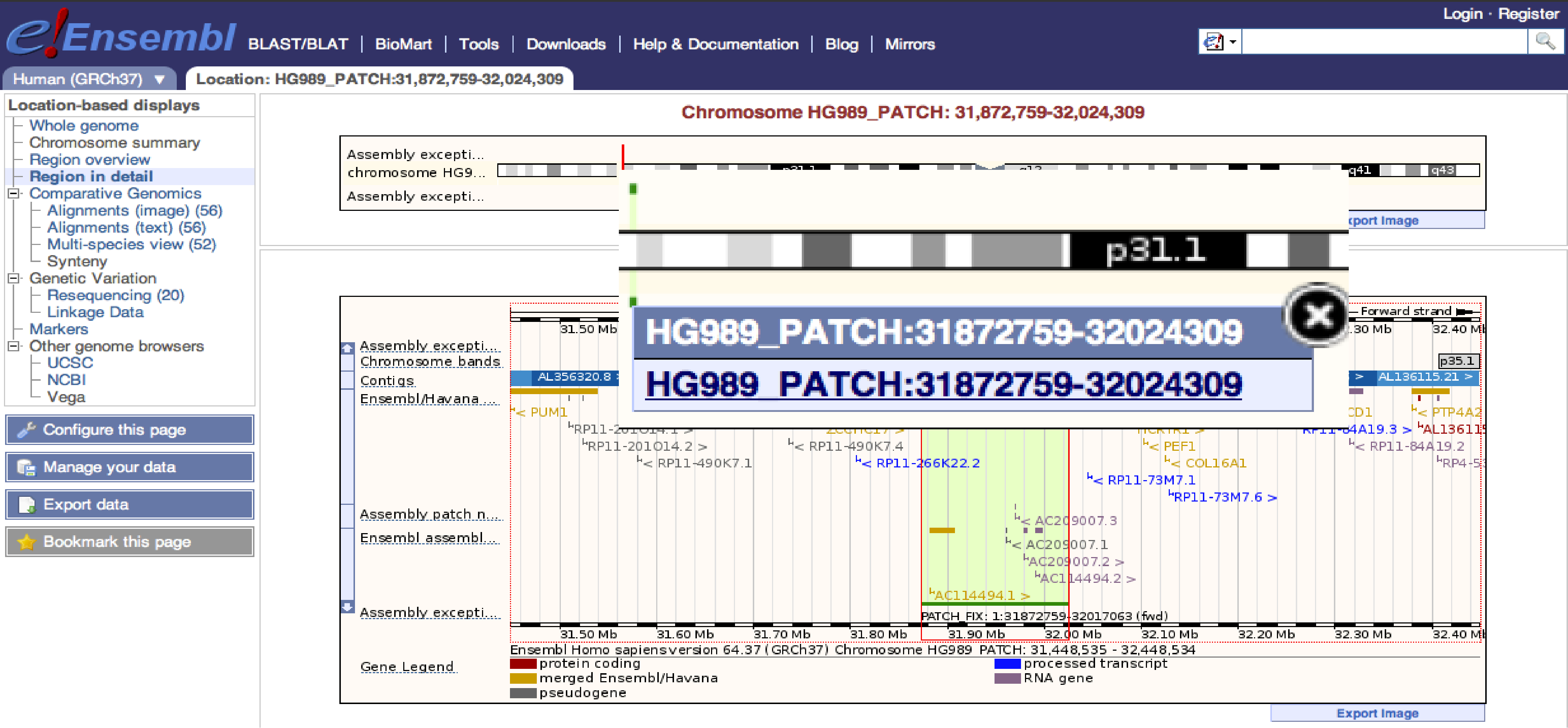
(adds >5 Mb of novel sequence to the assembly)

71 Novel PATCHES: Additional sequence added

(adds >800K of novel sequence to the assembly)

Releasing patches quarterly

Finding the data



Location: HG989_PATCH:31,872,759-32,024,309

Chromosome HG989_PATCH: 31,872,759-32,024,309

HG989_PATCH:31872759-32024309

HG989_PATCH:31872759-32024309

PATCH_FIX: 1:31872759-32017063 (fwd)

Ensembl Homo sapiens version 64.37 (GRCh37) Chromosome HG989_PATCH: 31,448,535 - 32,448,534

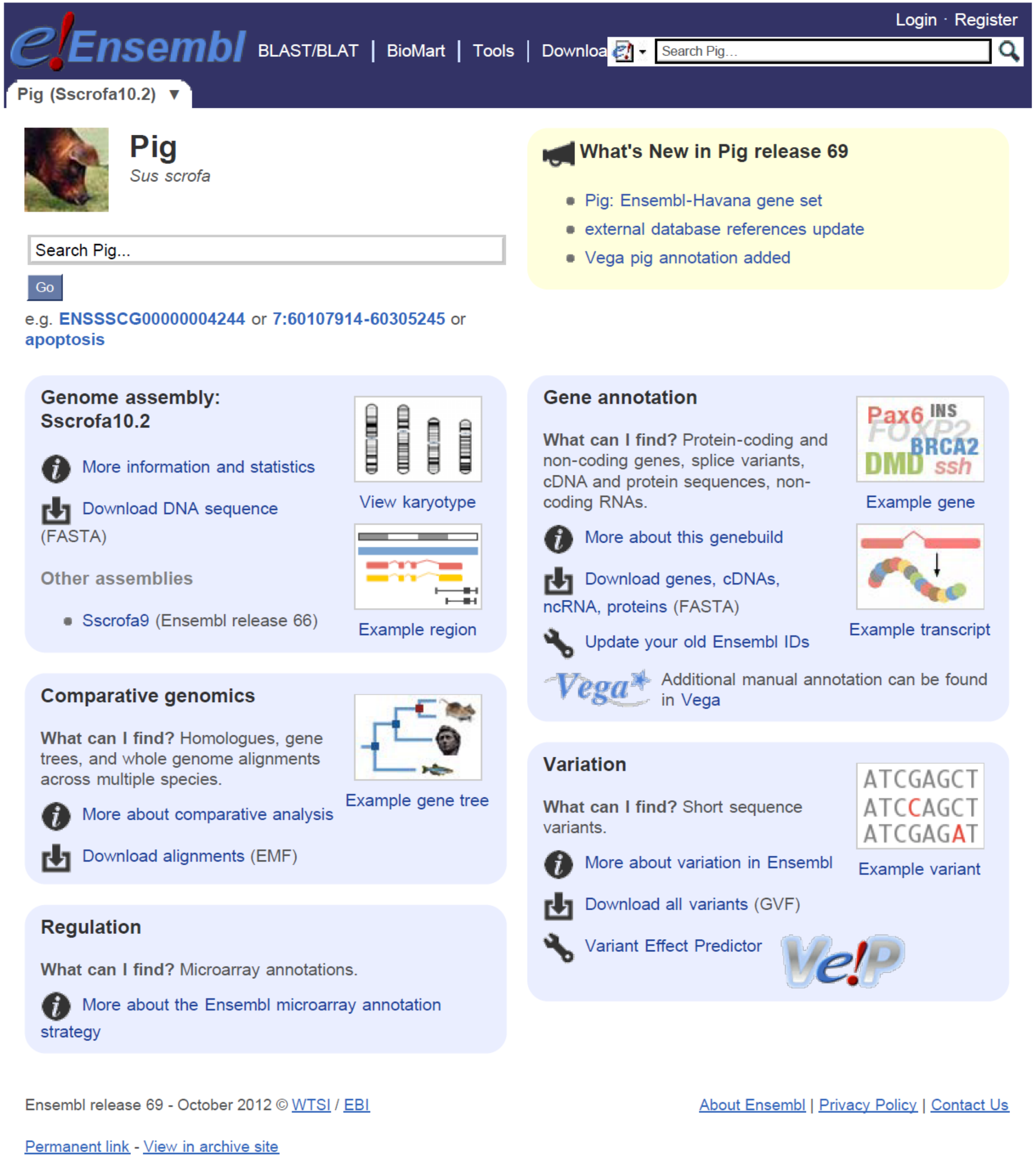
Improving the reference genome

- Improving the assembly
- Sscrofa10.5
  - Replace SSCX, SSCY assemblies
  - Add ‘novel’ and ‘fix’ patches
    - e.g. MHC, LRC
- Sscrofa11
  - iterative exploitation of existing data
  - global and targeted approaches
  - community engagement

Adding value – improved annotation

Improving genome annotation

- Gene models
  - manual annotation
- ncRNAs
- Gene expression
- Genome variation
  - multiple genomes
- Functional sequences
  - cf. ENCODE



Pig

Sus scrofa

Search Pig...

Go

e.g. **ENSSSCG00000004244** or **7:60107914-60305245** or **apoptosis**

What's New in Pig release 69

- Pig: Ensembl-Havana gene set
- external database references update
- Vega pig annotation added

Genome assembly: Sscrofa10.2

More information and statistics

Download DNA sequence (FASTA)

View karyotype

Example region

Other assemblies

- Sscrofa9 (Ensembl release 66)

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

Example gene

More about this genebuild

Download genes, cDNAs, ncRNA, proteins (FASTA)

Example transcript

Update your old Ensembl IDs

Vega Additional manual annotation can be found in Vega

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

Example gene tree

More about comparative analysis

Download alignments (EMF)

Variation

What can I find? Short sequence variants.

ATCGAGCT
ATCCAGCT
ATCGAGAT

Example variant

More about variation in Ensembl

Download all variants (GVF)

Variant Effect Predictor VEP

Regulation

What can I find? Microarray annotations.

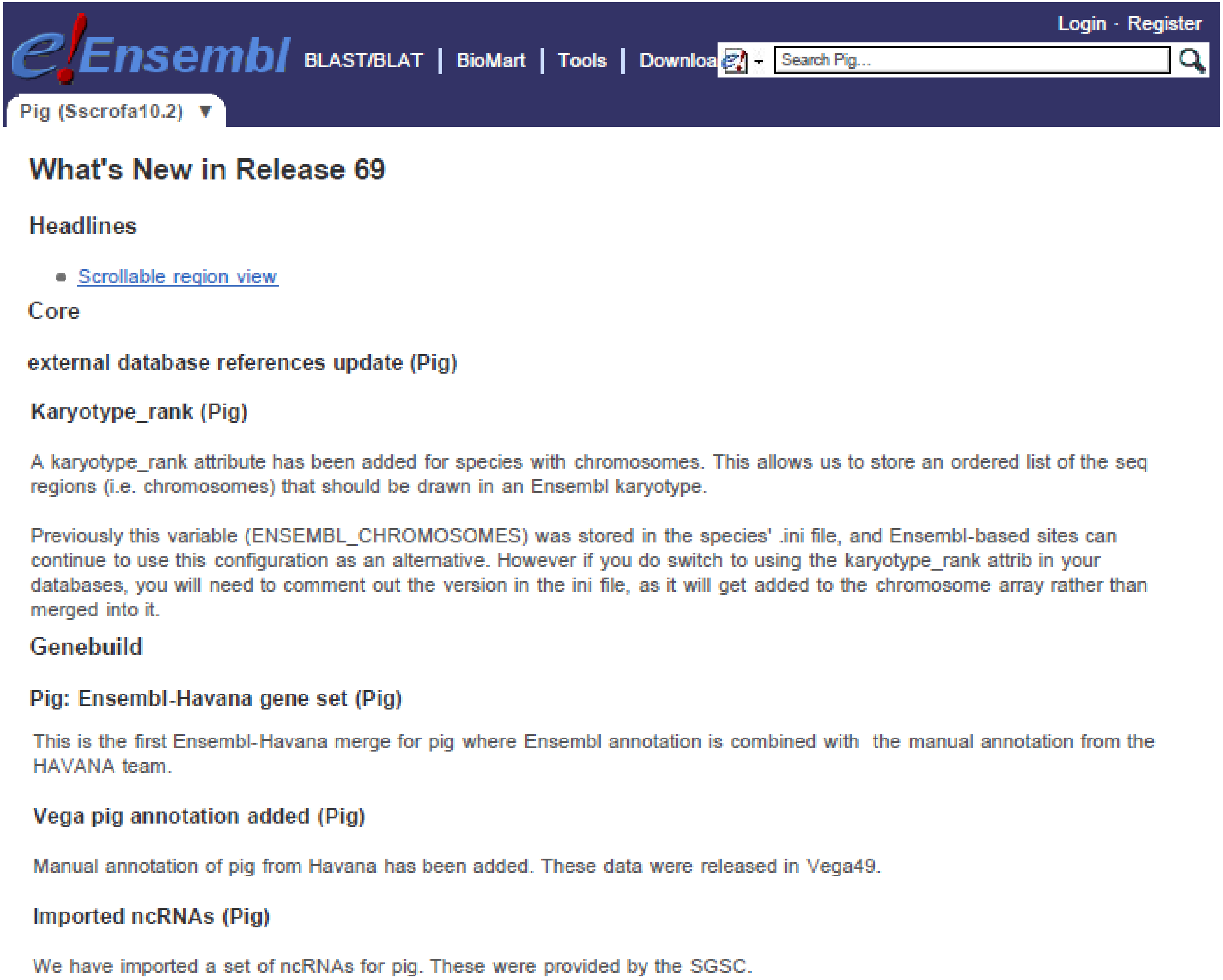
More about the Ensembl microarray annotation strategy

Ensembl release 69 - October 2012 © WTSI / EBI

About Ensembl | Privacy Policy | Contact Us

Permanent link - View in archive site

http://www.ensembl.org/Sus_scrofa/Info/Index



What's New in Release 69

Headlines

- Scrollable region view

Core

external database references update (Pig)

Karyotype_rank (Pig)

A karyotype_rank attribute has been added for species with chromosomes. This allows us to store an ordered list of the seq regions (i.e. chromosomes) that should be drawn in an Ensembl karyotype.

Previously this variable (ENSEMBL_CHROMOSOMES) was stored in the species' .ini file, and Ensembl-based sites can continue to use this configuration as an alternative. However if you do switch to using the karyotype_rank attrib in your databases, you will need to comment out the version in the ini file, as it will get added to the chromosome array rather than merged into it.

Genebuild

Pig: Ensembl-Havana gene set (Pig)

This is the first Ensembl-Havana merge for pig where Ensembl annotation is combined with the manual annotation from the HAVANA team.

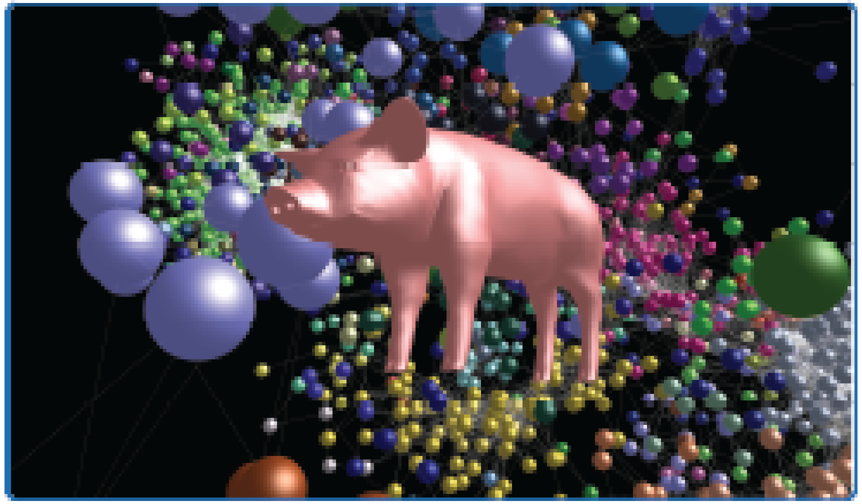
Vega pig annotation added (Pig)

Manual annotation of pig from Havana has been added. These data were released in Vega49.

Imported ncRNAs (Pig)

We have imported a set of ncRNAs for pig. These were provided by the SGSC.

BMC Biology
incorporating Journal of Biology



A gene expression atlas of the domestic pig

Freeman *et al.*

BioMed Central

Freeman et al. BMC Biology 2012, 10:90
http://www.biomedcentral.com/1741-7007/10/90 ([illegible] November 2012)

Freeman *et al. BMC Biology* 2012, **10**:90
http://www.biomedcentral.com/1741-7007/10/90

BMC Biology
incorporating Journal of Biology

RESEARCH ARTICLE **Open Access**

A gene expression atlas of the domestic pig

Tom C Freeman[1*], Alasdair Ivens[2,6], J Kenneth Baillie[1], Dario Beraldi[1,7], Mark W Barnett[1], David Dorward[1], Alison Downing[1], Lynsey Fairbairn[1], Ronan Kapetanovic[1], Sobia Raza[1], Andru Tomoiu[1], Ramiro Alberio[3], Chunlei Wu[4], Andrew I Su[4], Kim M Summers[1], Christopher K Tuggle[5], Alan L Archibald[1*] and David A Hume[1*]

Abstract

Background: This work describes the first genome-wide analysis of the transcriptional landscape of the pig. A new porcine Affymetrix expression array was designed in order to provide comprehensive coverage of the known pig transcriptome. The new array was used to generate a genome-wide expression atlas of pig tissues derived from 62 tissue/cell types. These data were subjected to network correlation analysis and clustering.

Results: The analysis presented here provides a detailed functional clustering of the pig transcriptome where transcripts are grouped according to their expression pattern, so one can infer the function of an uncharacterized gene from the company it keeps and the locations in which it is expressed. We describe the overall transcriptional signatures present in the tissue atlas, where possible assigning those signatures to specific cell populations or pathways. In particular, we discuss the expression signatures associated with the gastrointestinal tract, an organ that was sampled at 15 sites along its length and whose biology in the pig is similar to human. We identify sets of genes that define specialized cellular compartments and region-specific digestive functions. Finally, we performed a network analysis of the transcription factors expressed in the gastrointestinal tract and demonstrate how they subdivide into functional groups that may control cellular gastrointestinal development.

Conclusions: As an important livestock animal with a physiology that is more similar than mouse to man, we provide a major new resource for understanding gene expression with respect to the known physiology of mammalian tissues and cells. The data and analyses are available on the websites http://biogps.org and http://www.macrophages.com/pig-atlas.

Keywords: pig, porcine, *Sus scrofa*, microarray, transcriptome, transcription network, pathway, gastrointestinal tract

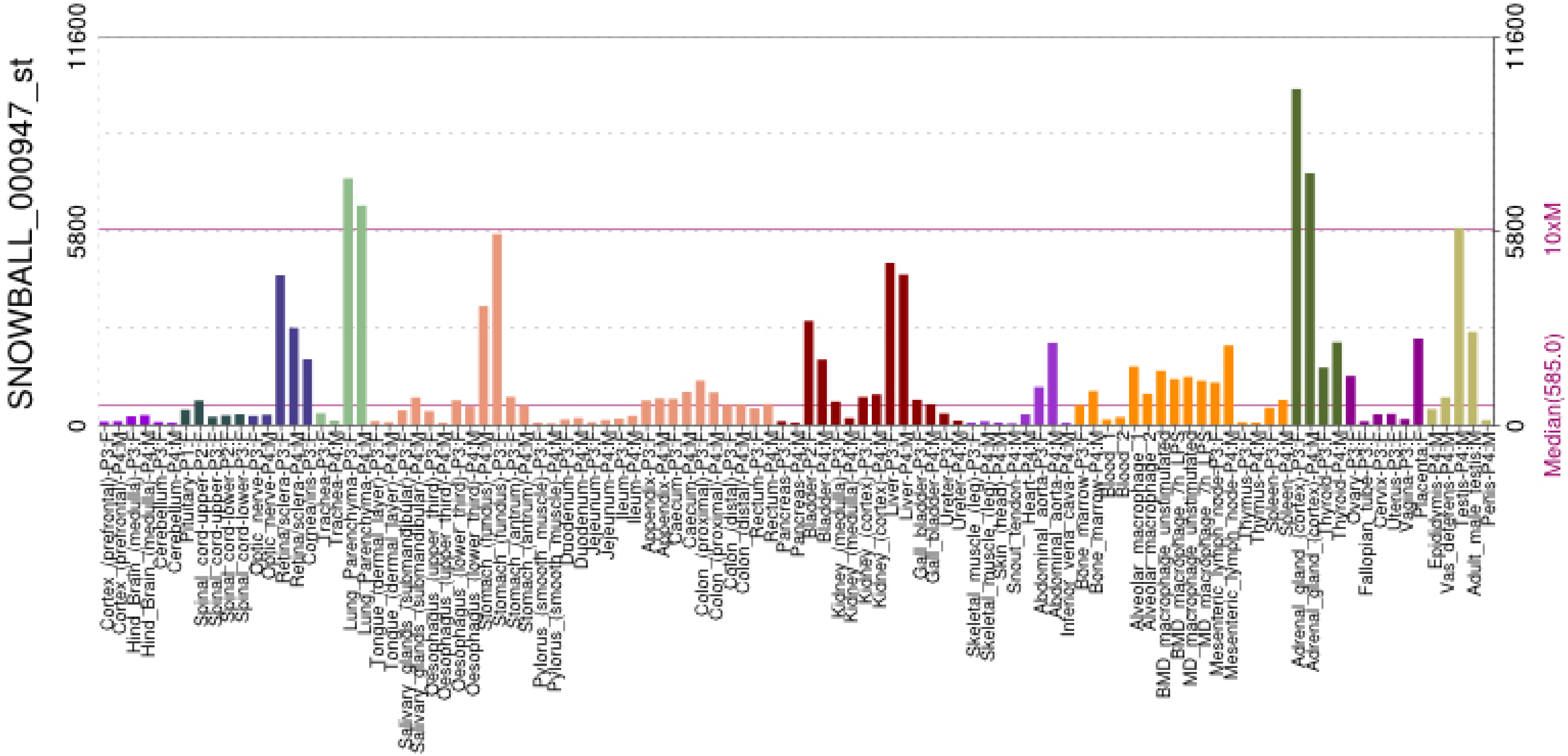
Expression array, atlas

- Tool for monitoring gene expression
- Inferring function of unknowns
  - Inform genome annotation
- Comparative functional genomics
  - Is pig kidney more/less like human kidney than mouse kidney?
- Microarray-based atlas
- RNAseq atlas in progress

Affymetrix Porcine Snowball Array content

- 123 Affy controls
- 35 virus genomes (tiled 17 bp spacing)
- 1,857 miRNA probes
- 37 MT-mRNA
- 45,927 mRNA
  - 37,842 with annotation
    - 6,767 LOC annotations
    - 16,626 unique genes with official symbol/description

Expression profiles



http://biogps.org

Genome variation

- Multiple genomes
  - ~300 domestic and wild pigs genomes (Groenen et al.)
  - ~60 Korean pig genomes
  - mini-pig genome (China)
  - ? Chinese pig genomes
  - breed pooled DNA sequences (Rubin et al)
- Ensembl variation
  - SNPs displayed in Ensembl
  - Variant Effect Predictor implemented in Ensembl

Enabling reading of pig genome sequences

- Annotation of functional sequences
  - Protein coding
  - Non-coding RNA sequences
  - Regulatory sequences
- cf. human ENCODE project
  - Encyclopedia of DNA elements in the human genome

ENCODE for farmed / companion animals

- Genetic variation underlying trait variation
  - Coding sequence
    - RYR1, DGAT1, ..
    - human putative LoF mutations comment
  - Regulatory sequence
    - IGF2, callipyge, …
    - likely to be more important, common
- Current annotation limited
  - cDNA / EST based gene models
  - SNP variation

ARTICLE

doi:10.1038/nature11247

An integrated encyclopedia of DNA elements in the human genome

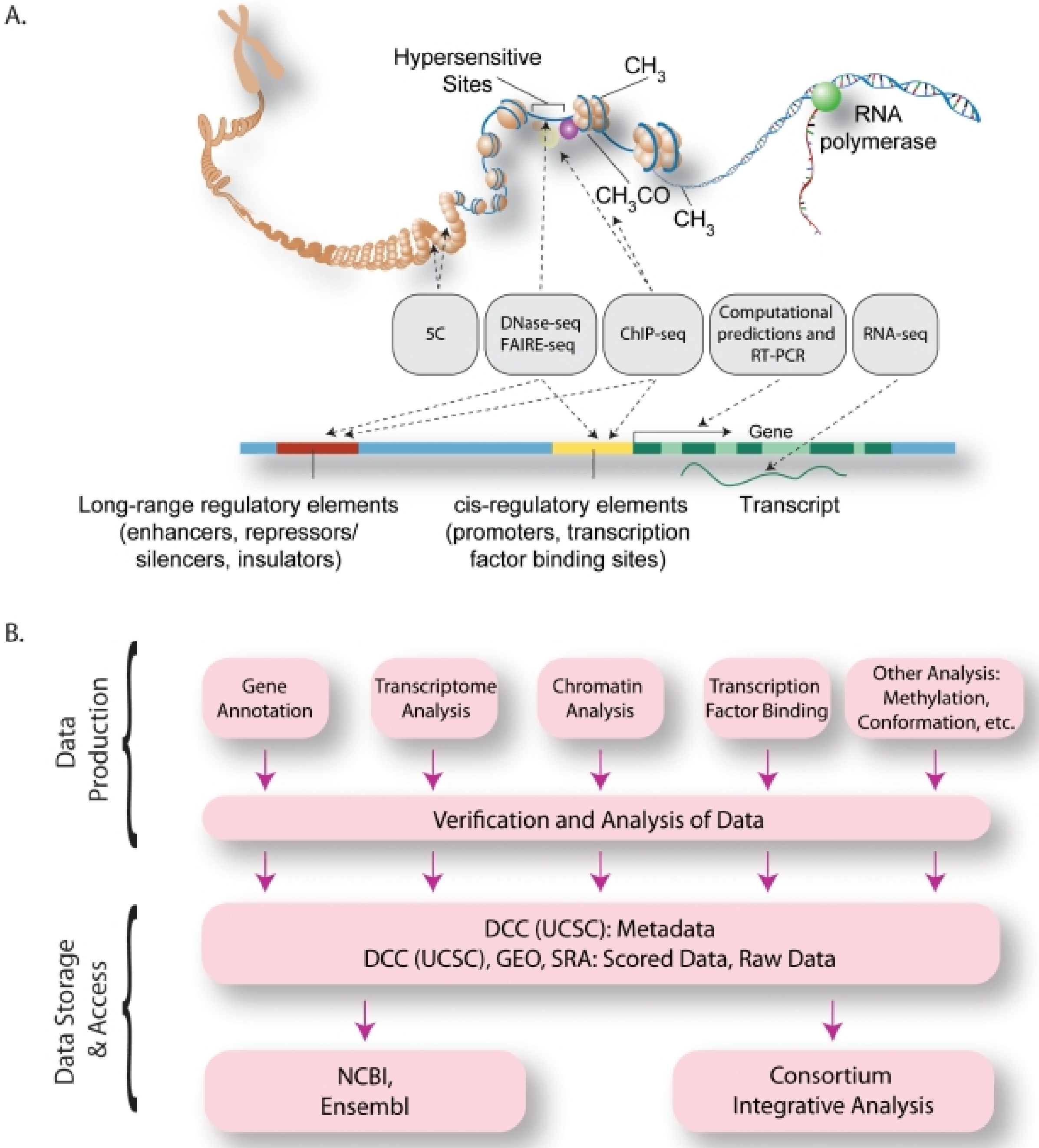
The ENCODE Project Consortium*

The human genome encodes the blueprint of life, but the function of the vast majority of its nearly three billion bases is unknown. The Encyclopedia of DNA Elements (ENCODE) project has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification. These data enabled us to assign biochemical functions for 80% of the genome, in particular outside of the well-studied protein-coding regions. Many discovered candidate regulatory elements are physically associated with one another and with expressed genes, providing new insights into the mechanisms of gene regulation. The newly identified elements also show a statistical correspondence to sequence variants linked to human disease, and can thereby guide interpretation of this variation. Overall, the project provides new insights into the organization and regulation of our genes and genome, and is an expansive resource of functional annotations for biomedical research.

The human genome sequence provides the underlying code for human biology. Despite intensive study, especially in identifying protein-coding genes, our understanding of the genome is far from complete, particularly with

ENCODE
Encyclopedia of DNA Elements
nature.com/encode

95% of the genome lies within 8 kilobases (kb) of a DNA–protein interaction (as assayed by bound ChIP-seq motifs or DNase I footprints), and 99% is within 1.7 kb of at least one of the biochemical events measured by ENCODE.


A.
Hypersensitive Sites
CH_3
RNA polymerase
CH_3CO
CH_3
5C
DNase-seq FAIRE-seq
ChIP-seq
Computational predictions and RT-PCR
RNA-seq
Gene
Long-range regulatory elements (enhancers, repressors/ silencers, insulators)
cis-regulatory elements (promoters, transcription factor binding sites)
Transcript
B.
Data Production
Gene Annotation
Transcriptome Analysis
Chromatin Analysis
Transcription Factor Binding
Other Analysis: Methylation, Conformation, etc.
Verification and Analysis of Data
Data Storage & Access
DCC (UCSC): Metadata
DCC (UCSC), GEO, SRA: Scored Data, Raw Data
NCBI, Ensembl
Consortium Integrative Analysis

ENCODE for farmed / companion animals

- By-product of biology-led research
  - development, differentiation, responses to perturbation
- Focus on target tissues
  - musco-skeletal
  - immune tissues
- Limited assays
  - DNaseI, FAIREseq
  - histone marks (promoters, enhancers)
  - methylation
  - RNAseq **(stranded)**, CAGE

nature COMMUNICATIONS

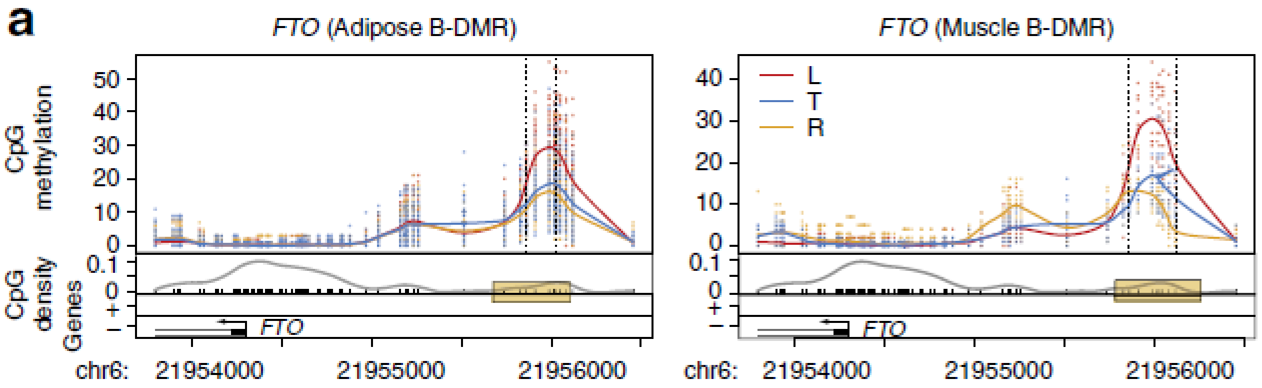
ARTICLE

Received 7 Oct 2011 | Accepted 19 Apr 2012 | Published 22 May 2012

DOI: 10.1038/ncomms1854

An atlas of DNA methylomes in porcine adipose and muscle tissues

Mingzhou Li[1,*], Honglong Wu[2,*], Zonggang Luo[1,3], Yudong Xia[2], Jiuqiang Guan[1], Tao Wang[1], Yiren Gu[4], Lei Chen[5], Kai Zhang[1,†], Jideng Ma[1], Yingkai Liu[1], Zhijun Zhong[1], Jing Nie[1], Shuling Zhou[1], Zhiping Mu[1], Xiaoyan Wang[1], Jingjing Qu[1], Long Jing[1], Huiyu Wang[1], Shujia Huang[2], Na Yi[2], Zhe Wang[2], Dongxing Xi[2], Juan Wang[2], Guangliang Yin[2], Li Wang[2], Ning Li[2], Zhi Jiang[2], Qiulei Lang[6], Huasheng Xiao[7], Anan Jiang[1], Li Zhu[1], Yanzhi Jiang[1], Guoqing Tang[1], Miaomiao Mai[1], Surong Shuai[1], Ning Li[8], Kui Li[9], Jinyong Wang[5], Xiuqing Zhang[2], Yingrui Li[2], Haosi Chen[10], Xiaolian Gao[10], Graham S. Plastow[11], Stephen Beck[12], Huanming Yang[2], Jian Wang[2], Jun Wang[2], Xuewei Li[1] & Ruiqiang Li[2,†]



Cell

Cell PRESS

Volume 149, Issue 6, 8 June 2012, Pages 1381–1392

Resource

Comparative Epigenomic Annotation of Regulatory DNA

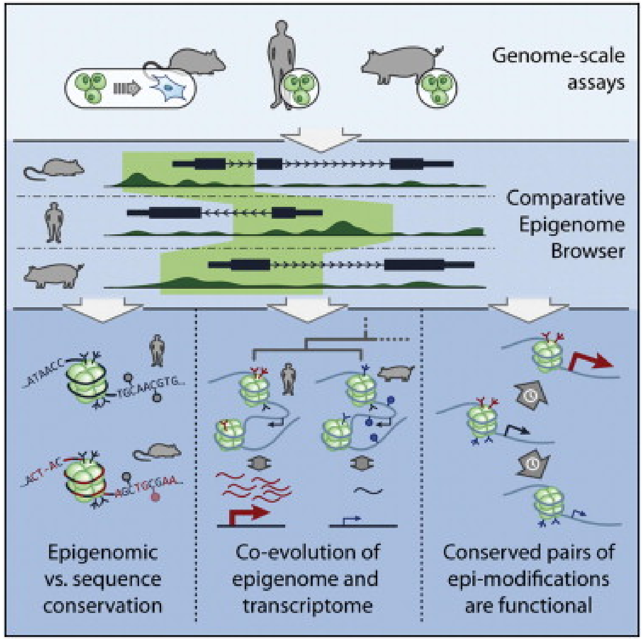
Shu Xiao[1,2,6], Dan Xie[1,2,6], Xiaoyi Cao[2,3,6], Pengfei Yu[2,3,6], Xiaoyun Xing[5], Chieh-Chun Chen[1,2], Meagan Musselman[1], Mingchao Xie[5], Franklin D. West[4], Harris A. Lewin[2], Ting Wang[5], Sheng Zhong[1,2,3,]

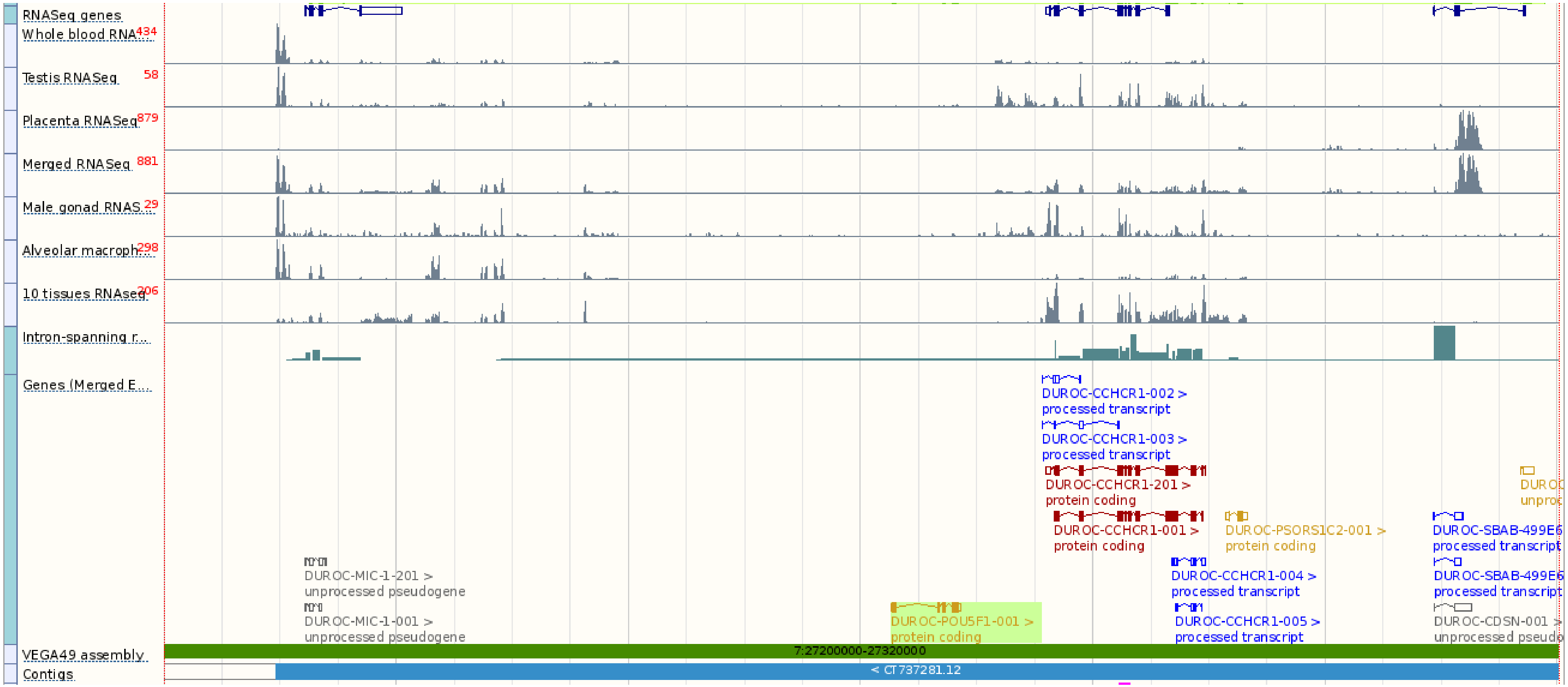
[1] Department of Bioengineering, University of Illinois at Urbana-Champaign, 1304 West Springfield Avenue, Urbana, IL 61801, USA

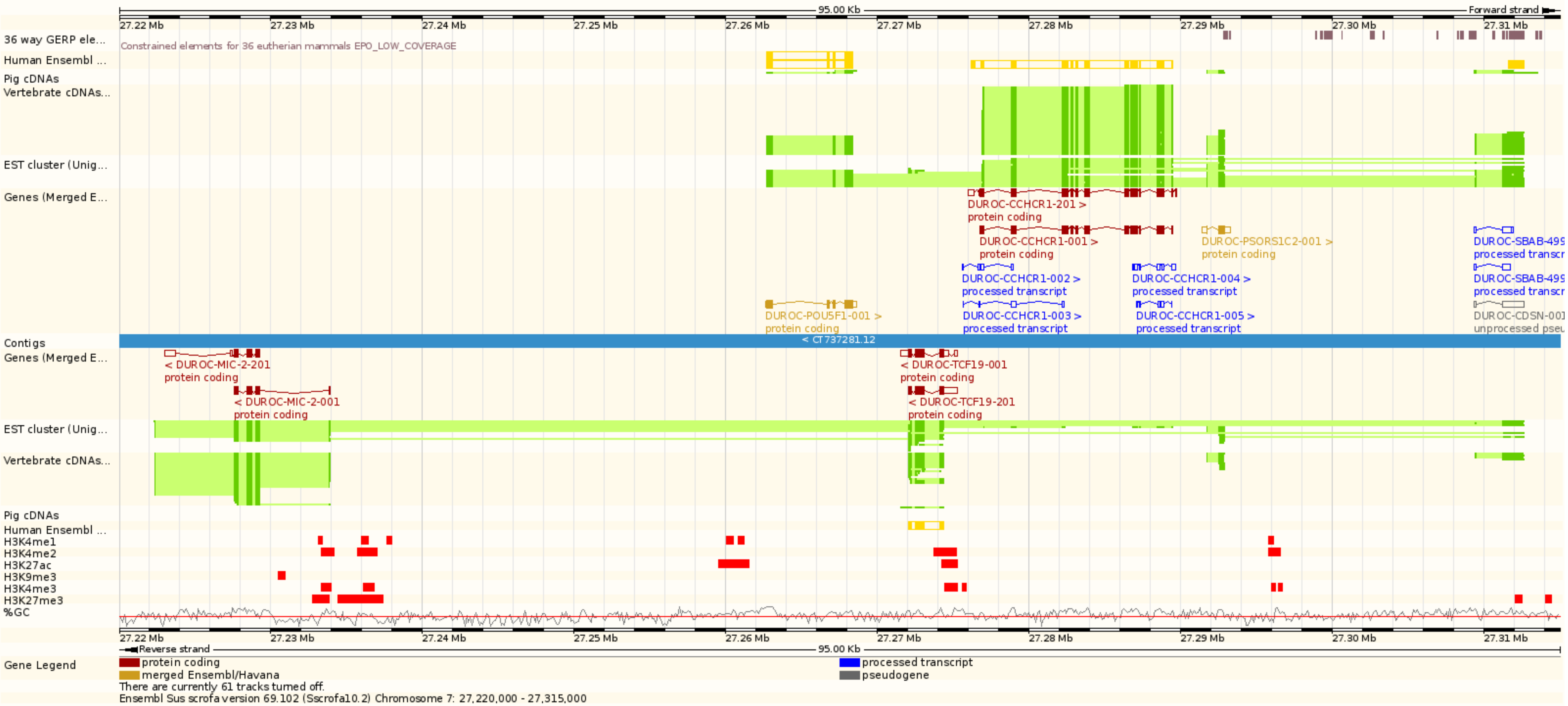
[2] Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1304 West Springfield Avenue, Urbana, IL 61801, USA

[3] Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, 1304 West Springfield Avenue, Urbana, IL 61801, USA

[4] Department of Animal and Dairy Science, University of Georgia, 425 River Road, Athens, GA 30602, USA




RNASeq genes
Whole blood RNA 434
Testis RNASeq 58
Placenta RNASeq 879
Merged RNASeq 881
Male gonad RNAS 29
Alveolar macroph 298
10 tissues RNAseq 206
Intron-spanning r...
Genes (Merged E...
DUROC-CCHCR1-002 >
processed transcript
DUROC-CCHCR1-003 >
processed transcript
DUROC-CCHCR1-201 >
protein coding
DUROC-CCHCR1-001 >
protein coding
DUROC-PSORS1C2-001 >
protein coding
DUROC-SBAB-499E6
processed transcript
DUROC-SBAB-499E6
processed transcript
DUROC
unproc
DUROC-MIC-1-201 >
unprocessed pseudogene
DUROC-MIC-1-001 >
unprocessed pseudogene
DUROC-POU5F1-001 >
protein coding
DUROC-CCHCR1-004 >
processed transcript
DUROC-CCHCR1-005 >
processed transcript
DUROC-CDSN-001 >
unprocessed pseudo
VEGA49 assembly
7:27200000-27320000
Contigs
< CT737281.12



There are currently 61 tracks turned off.

Ensembl Sus scrofa version 69.102 (Sscrofa10.2) Chromosome 7: 27,220,000 - 27,315,000

ENCODE for farmed / companion animals

- Species
  - single consortium / one per species / species groups
- Cells
  - transformed cells / primary cells / iPS cells
  - sharing
- Data management, publication
  - across groups
  - wider community
  - Toronto Statement
- Coordination

alan.archibald@roslin.ed.ac.uk