BovReg
Understanding cattle genomes

Newly annotated genomic features for biology-driven genome selection: the BovReg contribution

LIÈGE université
GIGA

Gabriel Costa Monteiro Moreira, Lijing Tang, Sébastien Dupont, Meenu Bhati, Hubert Pausch, Doreen Becker, Mazdak Salavati, Richard Clark, Emily L. Clark, Graham Plastow, Christa Kühn and Carole Charlier on behalf of the BovReg consortium

FAANG
Functional Annotation of Animal Genomes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668

Disclaimer: the sole responsibility of this presentation lies with the authors. The Research Executive Agency is not responsible for any use that may be made of the information contained therein.

Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

- ➢ BovReg consortium brings together scientists from twenty leading laboratories from different countries (EU, Switzerland, international organizations, Canada and Australia)
- ➢ Aims to improve our understanding of the cattle genome; also provide a comprehensive map of functionally active genomic features in cattle
- ➢ Critical for biological discovery and in linking genotype to phenotype; key knowledge for biology-driven genomic prediction needed by scientific and industry livestock communities.

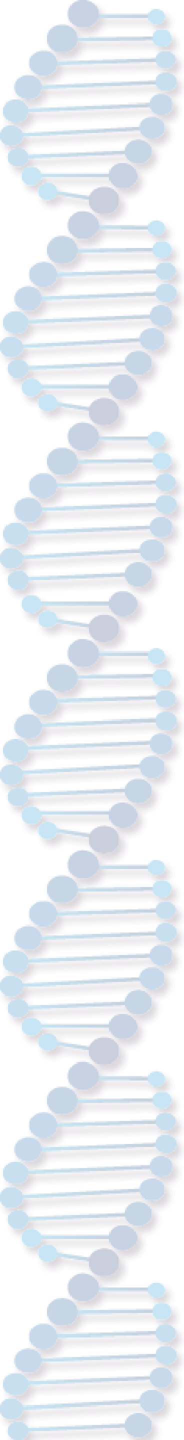
www.bovreg.eu

@BovReg.H2020

@BovReg

BovReg

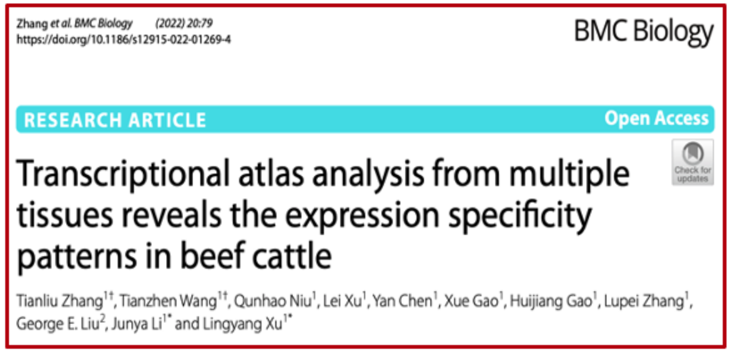
BovReg project



Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

➢ **Recent efforts from the FAANG consortium have contributed to the functional annotation of the bovine genome but, with a limited number of tissues, individuals and populations analyzed**

Chinease Simmental cattle



Zhang *et al. BMC Biology* (2022) 20:79
https://doi.org/10.1186/s12915-022-01269-4

BMC Biology

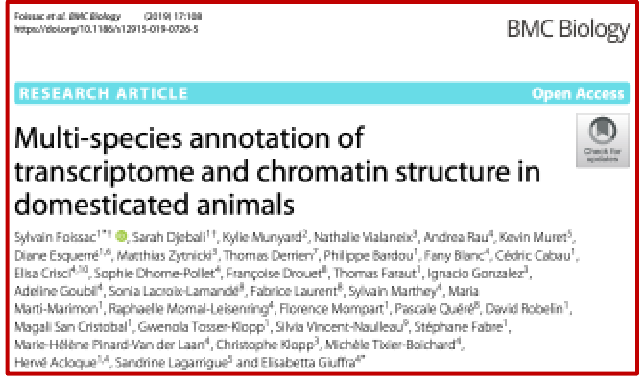
RESEARCH ARTICLE Open Access

Transcriptional atlas analysis from multiple tissues reveals the expression specificity patterns in beef cattle

Tianliu Zhang[1†], Tianzhen Wang[1†], Qunhao Niu[1], Lei Xu[1], Yan Chen[1], Xue Gao[1], Huijiang Gao[1], Lupei Zhang[1], George E. Liu[2], Junya Li[1*] and Lingyang Xu[1*]

- 3 animals; same age
- 51 tissue types
- Only PolyA+ data

Holstein



Foissac *et al. BMC Biology* (2019) 17:108
https://doi.org/10.1186/s12915-019-0726-5

BMC Biology

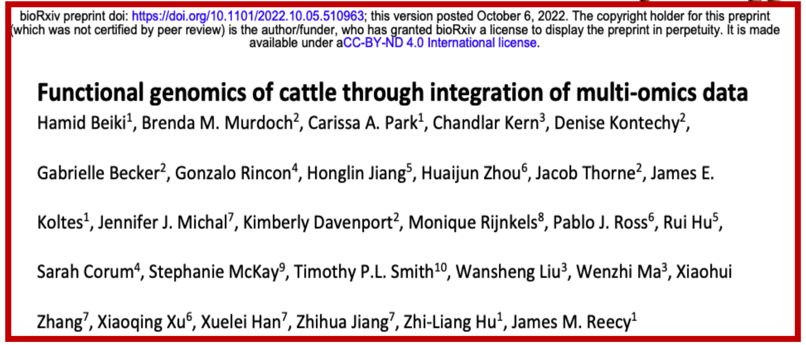
RESEARCH ARTICLE Open Access

Multi-species annotation of transcriptome and chromatin structure in domesticated animals

Sylvain Foissac[1*†], Sarah Djebali[1†], Kylie Munyard[2], Nathalie Vialaneix[3], Andrea Rau[4], Kevin Muret[5], Diane Esquerré[1,6], Matthias Zytnicki[3], Thomas Derrien[7], Philippe Bardou[1], Fany Blanc[4], Cédric Cabau[1], Elisa Crisci[4,10], Sophie Dhorne-Pollet[4], Françoise Drouet[8], Thomas Faraut[1], Ignacio Gonzalez[3], Adeline Goubil[4], Sonia Lacroix-Lamandé[8], Fabrice Laurent[8], Sylvain Marthey[4], Maria Marti-Marimon[1], Raphaelle Momal-Leisenring[4], Florence Mompart[1], Pascale Quéré[8], David Robelin[1], Magali San Cristobal[1], Gwenola Tosser-Klopp[1], Silvia Vincent-Naulleau[9], Stéphane Fabre[1], Marie-Hélène Pinard-Van der Laan[4], Christophe Klopp[3], Michèle Tixier-Boichard[4], Hervé Acloque[1,4], Sandrine Lagarrigue[5] and Elisabetta Giuffra[4*]

- Only Liver. CD4 and CD8 cells
- 4 adult animals (Holstein)
- PolyA+. ATAC and Hi-C

Line 1 Herefords



bioRxiv preprint doi: https://doi.org/10.1101/2022.10.05.510963; this version posted October 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Functional genomics of cattle through integration of multi-omics data

Hamid Beiki[1], Brenda M. Murdoch[2], Carissa A. Park[1], Chandlar Kern[3], Denise Kontechy[2], Gabrielle Becker[2], Gonzalo Rincon[4], Honglin Jiang[5], Huaijun Zhou[6], Jacob Thorne[2], James E. Koltes[1], Jennifer J. Michal[7], Kimberly Davenport[2], Monique Rijnkels[8], Pablo J. Ross[6], Rui Hu[5], Sarah Corum[4], Stephanie McKay[9], Timothy P.L. Smith[10], Wansheng Liu[3], Wenzhi Ma[3], Xiaohui Zhang[7], Xiaoqing Xu[6], Xuelei Han[7], Zhihua Jiang[7], Zhi-Liang Hu[1], James M. Reecy[1]

- 47 tissues from 4 Herefords (male and female) (14 months old)

Line 1 Herefords

nature COMMUNICATIONS

Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research

Colin Kern[1], Ying Wang[1], Xiaoqin Xu[1], Zhangyuan Pan[1], Michelle Halstead[1], Ganrea Chanthavixay[1], Perot Saelao[1], Susan Waters[1], Ruidong Xiang[2,3], Amanda Chamberlain[3], Ian Korf[4], Mary E. Delany[1], Hans H. Cheng[5], Juan F. Medrano[1], Alison L. Van Eenennaam[1], Chris K. Tuggle[6], Catherine Ernst[7], Paul Flicek[8], Gerald Quon[9], Pablo Ross[1✉] & Huaijun Zhou[1✉]

- 8 tissues from 2 Herefords male cattle (14 months old)

nature genetics

Check for updates

comment

The missing diversity in human epigenomic studies

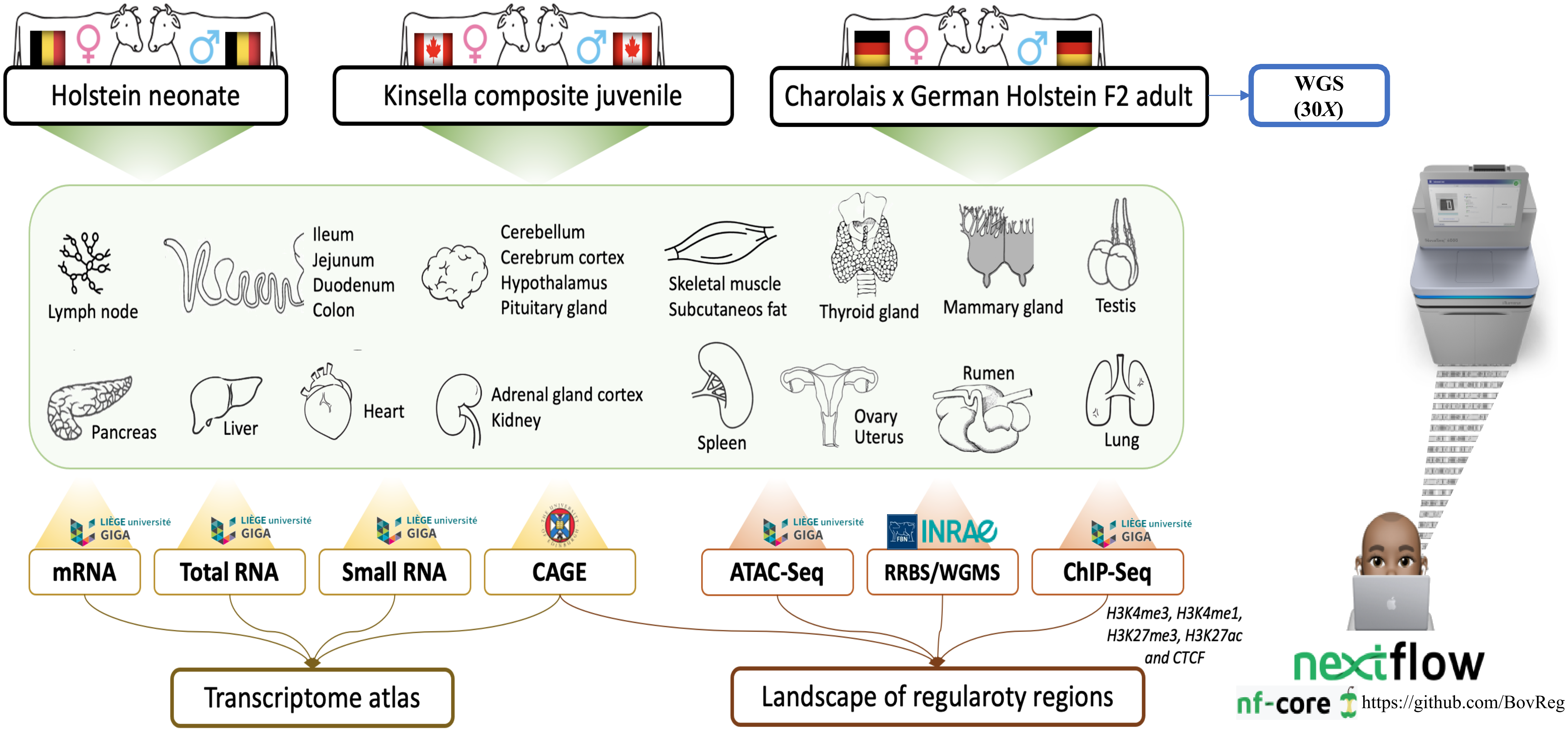
Charles E. Breeze, Stephan Beck, Sonja I. Berndt and Nora Franceschini

BovReg Understanding cattle genomes

- Several assays performed in the same animal/tissue
- Different developmental stages
- Different environments
- Both sexes
- Different breeds

Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

➢ **BovReg utilizes a diverse catalogue of tissues for functional annotation.**



Holstein neonate

Kinsella composite juvenile

Charolais x German Holstein F2 adult

WGS (30*X*)

Lymph node

Ileum
Jejunum
Duodenum
Colon

Cerebellum
Cerebrum cortex
Hypothalamus
Pituitary gland

Skeletal muscle
Subcutaneos fat

Thyroid gland

Mammary gland

Testis

Pancreas

Liver

Heart

Adrenal gland cortex
Kidney

Spleen

Ovary
Uterus

Rumen

Lung

mRNA

Total RNA

Small RNA

CAGE

ATAC-Seq

RRBS/WGMS

ChIP-Seq

H3K4me3, H3K4me1, H3K27me3, H3K27ac and CTCF

Transcriptome atlas

Landscape of regularoty regions

nextflow

nf-core

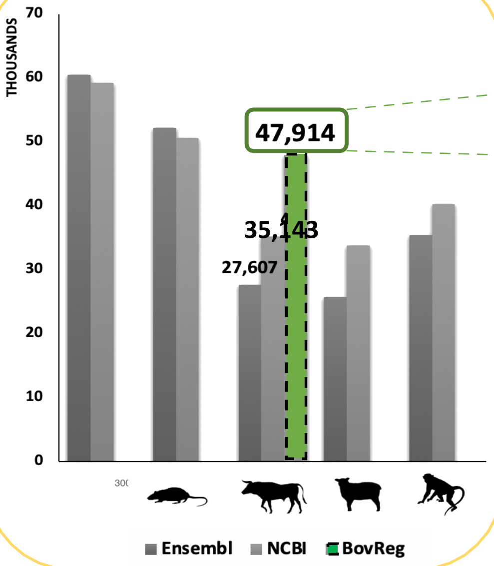
https://github.com/BovReg

Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

Transcriptome atlas

Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

- ➢ **BovReg expanded the catalog of bovine non-coding RNAs by including non-polyadenylated transcripts.**
- ➢ **48k genes models including ≥15k potentially novel transcripts!**



290,707 transcripts

43,979
Ensembl Biotypes

\+

47,167 (CodAn; ORF)
CDS, start- and stop-codon

\+

199,561
Non-coding transcripts

BovReg *de novo* transcriptome atlas

PolyA$^+$

Total RNA

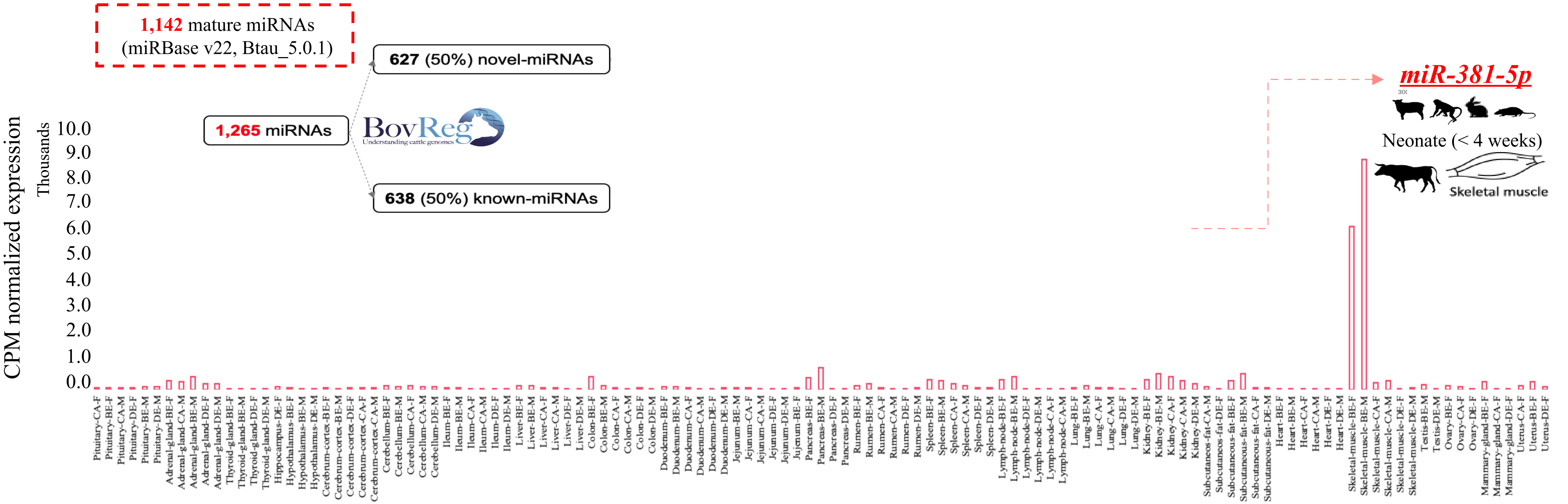
Class code[1]	**Features**[1]	**Compared with bovine annotation**	
		Ensembl r.105	**NCBI r.106**
=	*complete transcripts with exact match of intron chain*	42873 (14.75%)	36660 (12.61%)
c	*transcripts contained in the reference (intron compatible)*	4 (0.00%)	3082 (1.06%)
k	*containment of reference (reverse containment)*	25403 (8.74%)	24814 (8.54%)
m	*transcripts with retained introns (all introns matched or retained)*	4555 (1.57%)	6321 (2.17%)
n	*transcripts with retained introns (not all introns matched or retained)*	13812 (4.75%)	14760 (5.08%)
j	*multi-exon transcripts with at least one junction match*	142987 (49.19%)	150476 (51.76%)
e	*single exon transcript partially covering an intron (possible pre-mRNA fragment)*	0 (0.00%)	170 (0.06%)
o	*other same strand transcripts overlap with reference exons*	6422 (2.21%)	9458 (3.25%)
s	*transcipts with intron match on the opposite strand (likely a mapping error)*	63 (0.02%)	81 (0.03%)
x	*exonic transcripts overlap on the opposite strand (like o or e but on the opposite strand)*	10569 (3.64%)	8817 (3.03%)
i	*transcritps fully contained within a reference intron*	10961 (3.77%)	12419 (4.27%)
y	*transcripts with a reference transcript within its intron*	1871 (0.64%)	999 (0.34%)
p	*possible polymerase run-on (no actual overlap)*	357 (0.12%)	342 (0.12%)
r	*repeat (al least 50% bases soft-masked)*	0 (0.00%)	0 (0.00%)
u	*none of the above (unknown, intergenic)*	30830 (10.61%)	22308 (7.67%)
	Total number of transcripts		290,707

[1] Transcript classification code from gffcompare v.0.12.2[10].

Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

➢ BovReg expanded the repertoire of mature miRNAs annotated in bovine.

Tissue-specific miRNA



1,142 mature miRNAs
(miRBase v22, Btau_5.0.1)

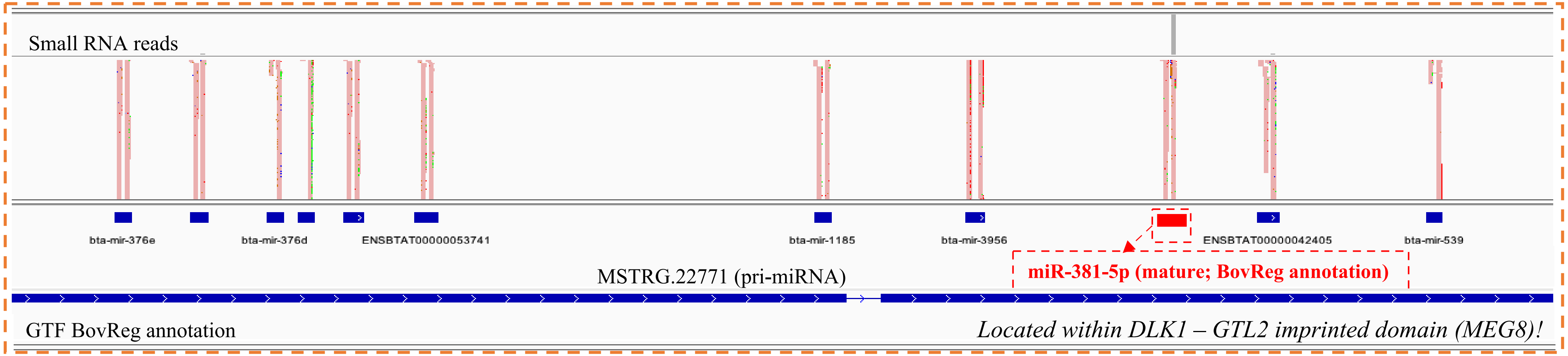
1,265 miRNAs

627 (50%) novel-miRNAs

638 (50%) known-miRNAs

miR-381-5p

Neonate (< 4 weeks)

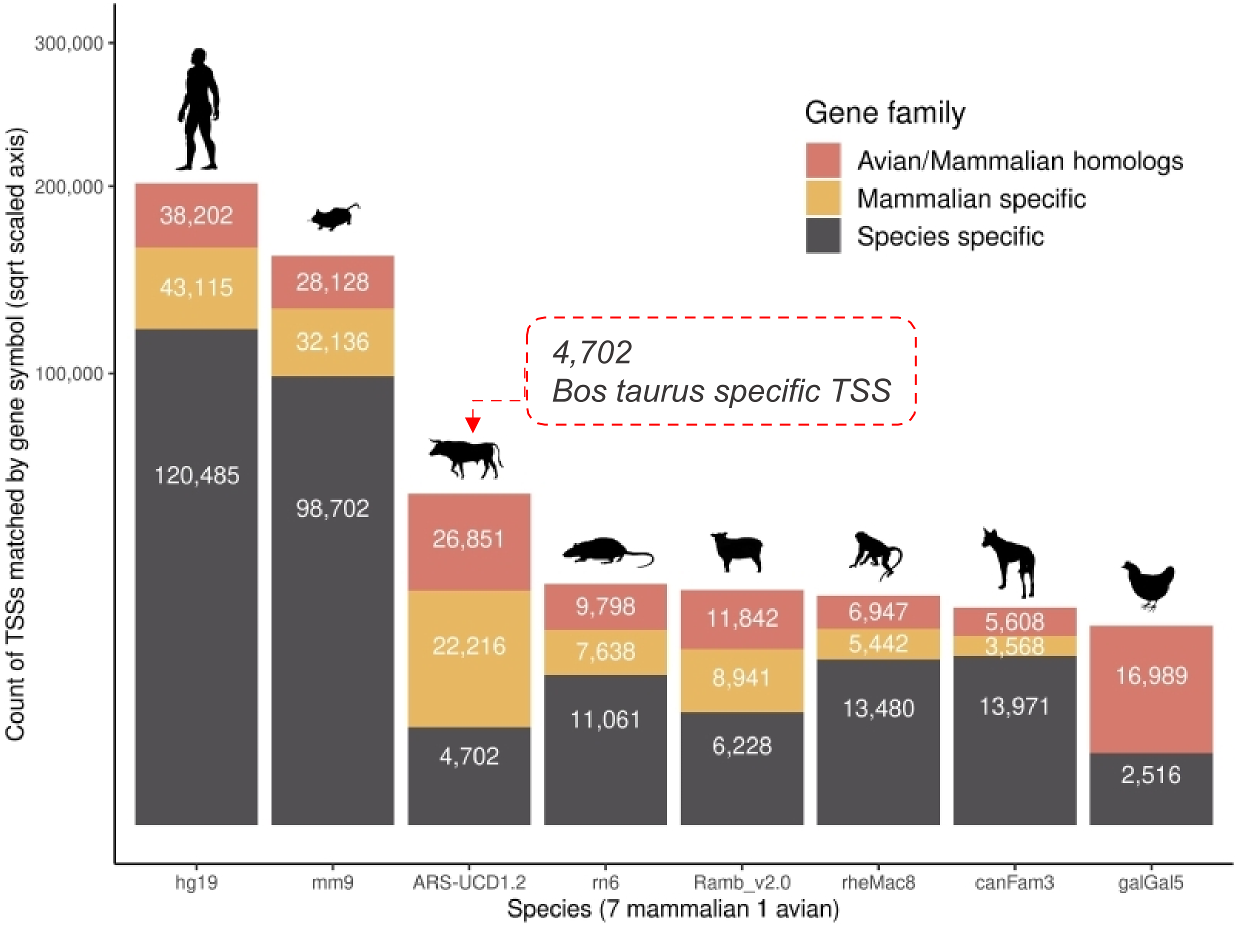


miR-381-5p (mature; BovReg annotation)

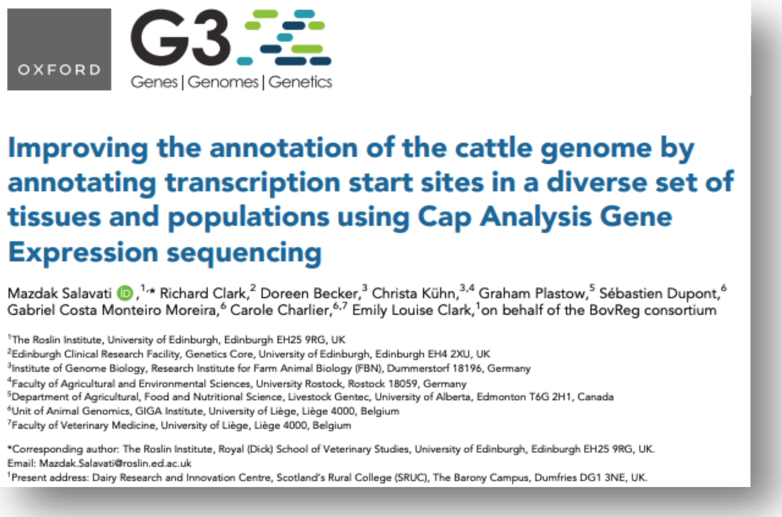
Located within DLK1 – GTL2 imprinted domain (MEG8)!

Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

- **51,295 transcription start sites (TSS)**
- **2,328 TSS-Enhancer regions shared across the three populations**



4,702
Bos taurus specific TSS

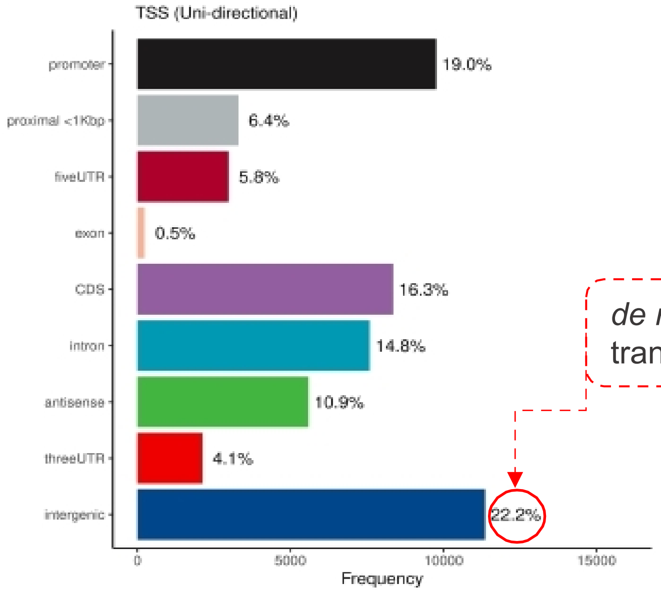


Improving the annotation of the cattle genome by annotating transcription start sites in a diverse set of tissues and populations using Cap Analysis Gene Expression sequencing

Mazdak Salavati,[1,*] Richard Clark,[2] Doreen Becker,[3] Christa Kühn,[3,4] Graham Plastow,[5] Sébastien Dupont,[6] Gabriel Costa Monteiro Moreira,[6] Carole Charlier,[6,7] Emily Louise Clark,[1] on behalf of the BovReg consortium

[1]The Roslin Institute, University of Edinburgh, Edinburgh EH25 9RG, UK
[2]Edinburgh Clinical Research Facility, Genetics Core, University of Edinburgh, Edinburgh EH4 2XU, UK
[3]Institute of Genome Biology, Research Institute for Farm Animal Biology (FBN), Dummerstorf 18196, Germany
[4]Faculty of Agricultural and Environmental Sciences, University Rostock, Rostock 18059, Germany
[5]Department of Agricultural, Food and Nutritional Science, Livestock Gentec, University of Alberta, Edmonton T6G 2H1, Canada
[6]Unit of Animal Genomics, GIGA Institute, University of Liège, Liège 4000, Belgium
[7]Faculty of Veterinary Medicine, University of Liège, Liège 4000, Belgium

*Corresponding author: The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh EH25 9RG, UK. Email: Mazdak.Salavati@roslin.ed.ac.uk
[1]Present address: Dairy Research and Innovation Centre, Scotland's Rural College (SRUC), The Barony Campus, Dumfries DG1 3NE, UK.



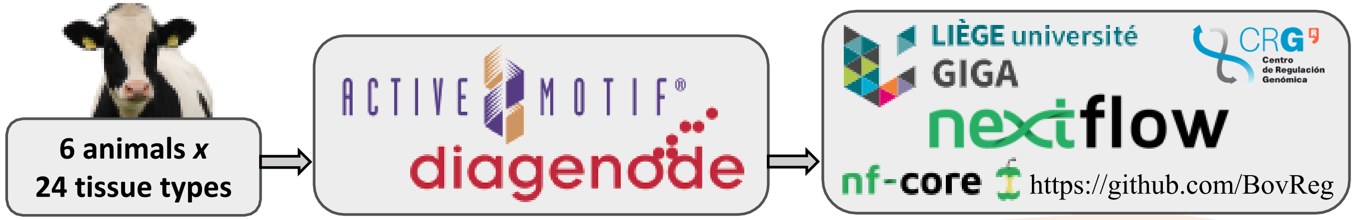
de novo
transcriptome atlas

Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

Regulatory regions

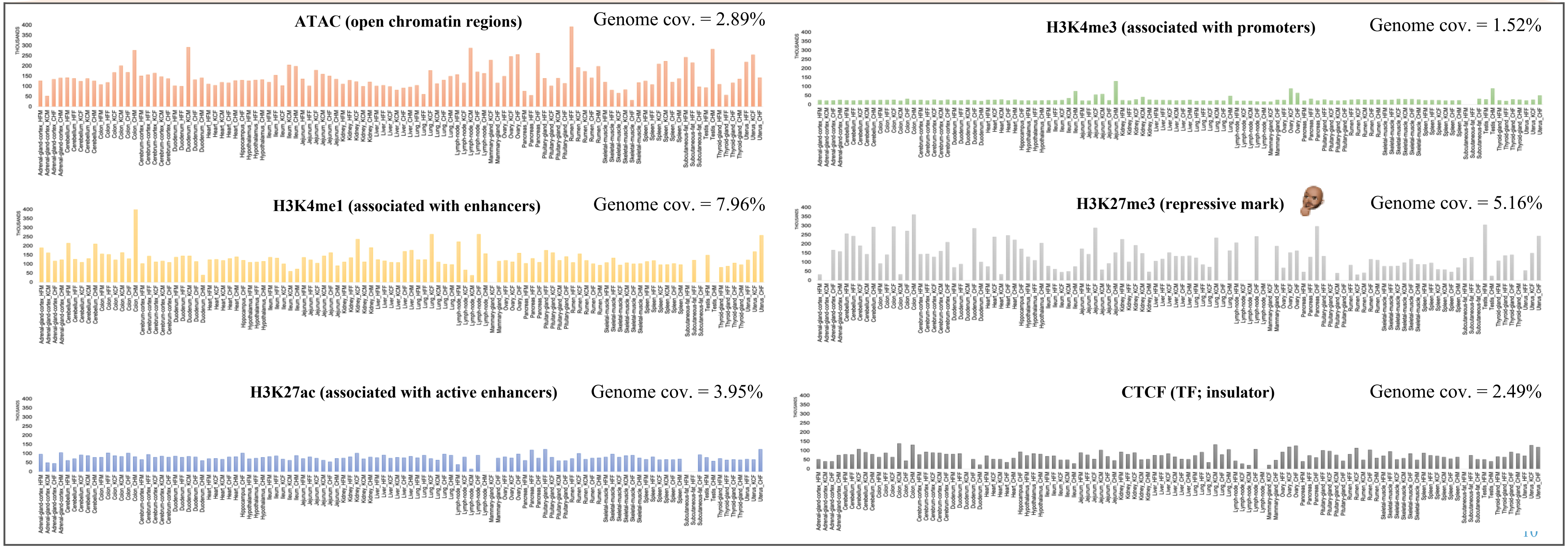
Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

➢ **We identified thousands of peaks for the different marks in our catalogue of tissues!**



6 animals *x* 24 tissue types

https://github.com/BovReg



ATAC (open chromatin regions) Genome cov. = 2.89%

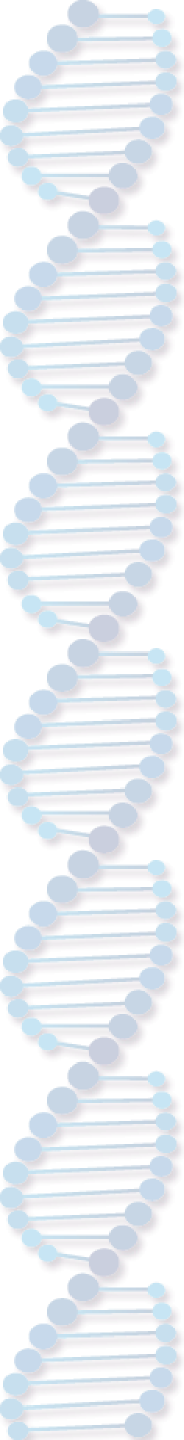
H3K4me3 (associated with promoters) Genome cov. = 1.52%

H3K4me1 (associated with enhancers) Genome cov. = 7.96%

H3K27me3 (repressive mark) Genome cov. = 5.16%

H3K27ac (associated with active enhancers) Genome cov. = 3.95%

CTCF (TF; insulator) Genome cov. = 2.49%



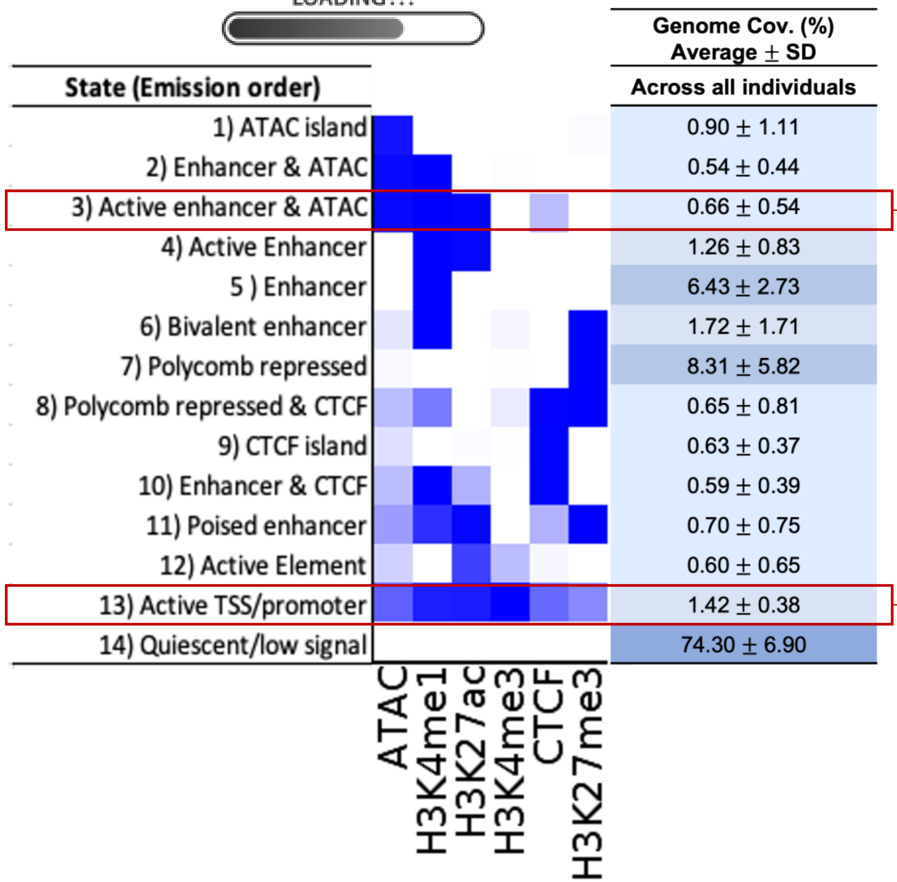
Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

➢ **Around 25% of the genome was covered by different chromatin states (except quiescent)**

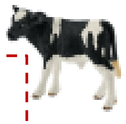
ATAC-Seq

ChIP-Seq

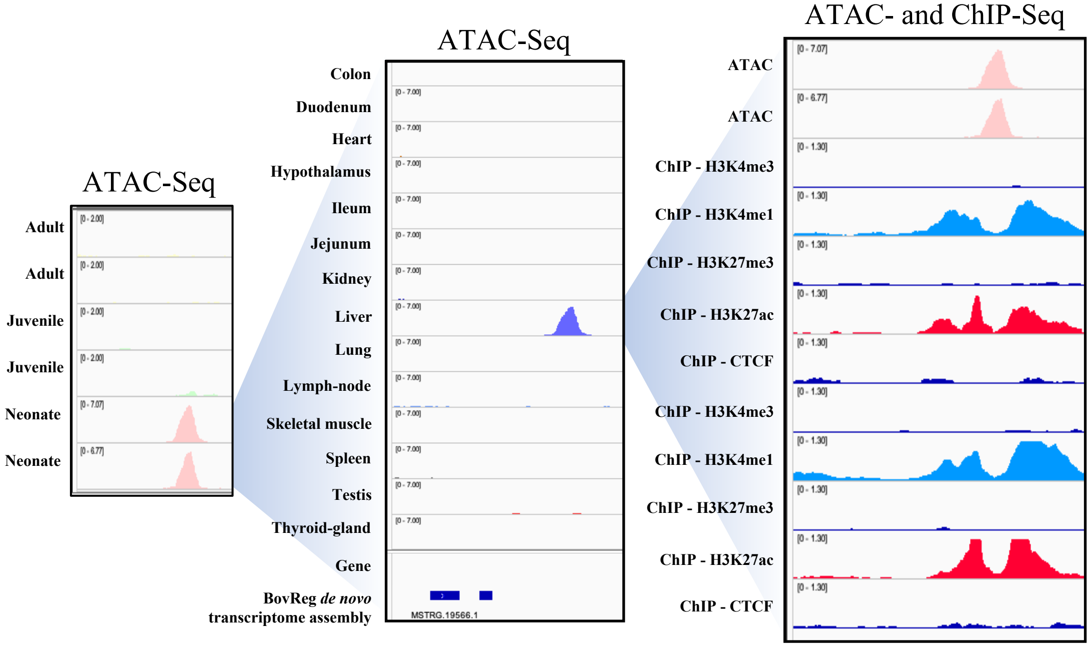
LOADING…



State (Emission order)	Genome Cov. (%) Average ± SD Across all individuals
1) ATAC island	0.90 ± 1.11
2) Enhancer & ATAC	0.54 ± 0.44
3) Active enhancer & ATAC	0.66 ± 0.54
4) Active Enhancer	1.26 ± 0.83
5) Enhancer	6.43 ± 2.73
6) Bivalent enhancer	1.72 ± 1.71
7) Polycomb repressed	8.31 ± 5.82
8) Polycomb repressed & CTCF	0.65 ± 0.81
9) CTCF island	0.63 ± 0.37
10) Enhancer & CTCF	0.59 ± 0.39
11) Poised enhancer	0.70 ± 0.75
12) Active Element	0.60 ± 0.65
13) Active TSS/promoter	1.42 ± 0.38
14) Quiescent/low signal	74.30 ± 6.90



Active enhancer (liver specific) only detected in neonate animals!!



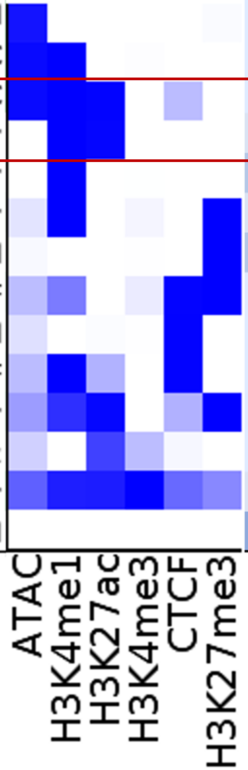
79% TSS annotated on RNA-Seq were validated by chromatin states

Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

➢ We recaptured known enhancers and annotated new ones!

ATAC-Seq **ChIP-Seq**

LOADING…



State (Emission order)	Genome Cov. (%) Average ± SD Across all individuals
1) ATAC island	0.90 ± 1.11
2) Enhancer & ATAC	0.54 ± 0.44
3) Active enhancer & ATAC	0.66 ± 0.54
4) Active Enhancer	1.26 ± 0.83
5) Enhancer	6.43 ± 2.73
6) Bivalent enhancer	1.72 ± 1.71
7) Polycomb repressed	8.31 ± 5.82
8) Polycomb repressed & CTCF	0.65 ± 0.81
9) CTCF island	0.63 ± 0.37
10) Enhancer & CTCF	0.59 ± 0.39
11) Poised enhancer	0.70 ± 0.75
12) Active Element	0.60 ± 0.65
13) Active TSS/promoter	1.42 ± 0.38
14) Quiescent/low signal	74.30 ± 6.90

BovReg active enhancers

vs

Line 1 Herefords

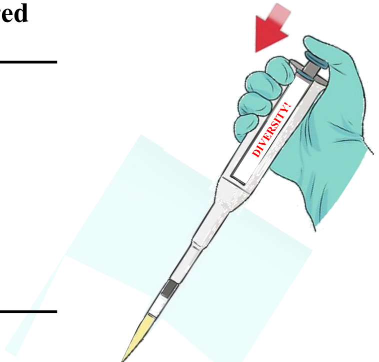
nature COMMUNICATIONS

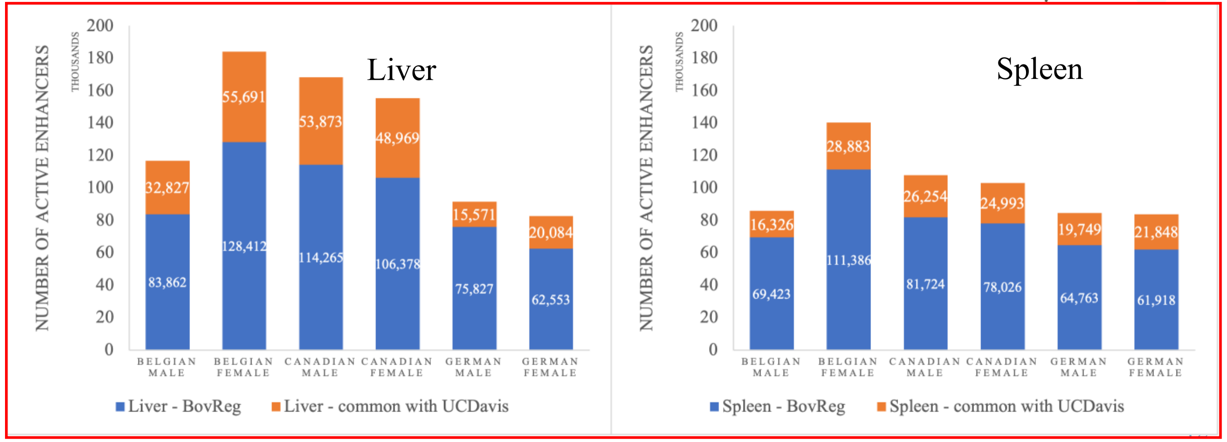
Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research

Colin Kern[1], Ying Wang[1], Xiaoqin Xu[1], Zhangyuan Pan[1], Michelle Halstead[1], Ganrea Chanthavixay[1], Perot Saelao[1], Susan Waters[1], Ruidong Xiang[2,3], Amanda Chamberlain[3], Ian Korf[4], Mary E. Delany[1], Hans H. Cheng[5], Juan F. Medrano[1], Alison L. Van Eenennaam[1], Chris K. Tuggle[6], Catherine Ernst[7], Paul Flicek[8], Gerald Quon[9], Pablo Ross[1✉] & Huaijun Zhou[1✉]

- 8 tissues from 2 Herefords male cattle (14 months old)

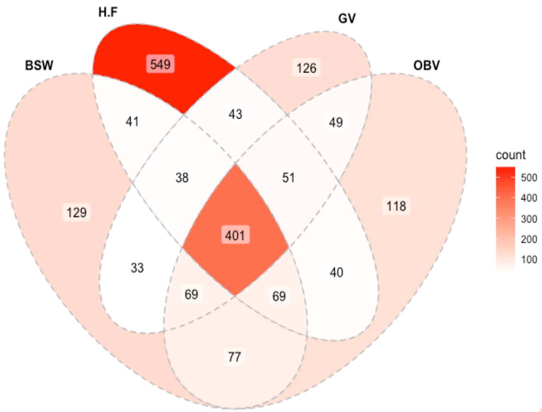
Tissues	Active enhancers (Kern et al., 2021)	Enhancers recaptured BovReg
Cerebellum	16,612	12,993 (**78.21%**)
Cortex	28,235	24,408 (**86.45%**)
Hypothalamus	27,796	17,277 (**62.16%**)
Liver	57,457	46,312 (**80.60%**)
Lung	49,861	38,537 (**77.29%**)
Muscle	28,023	19,442 (**69.38%**)
Spleen	29,707	24,520 (**82.54%**)





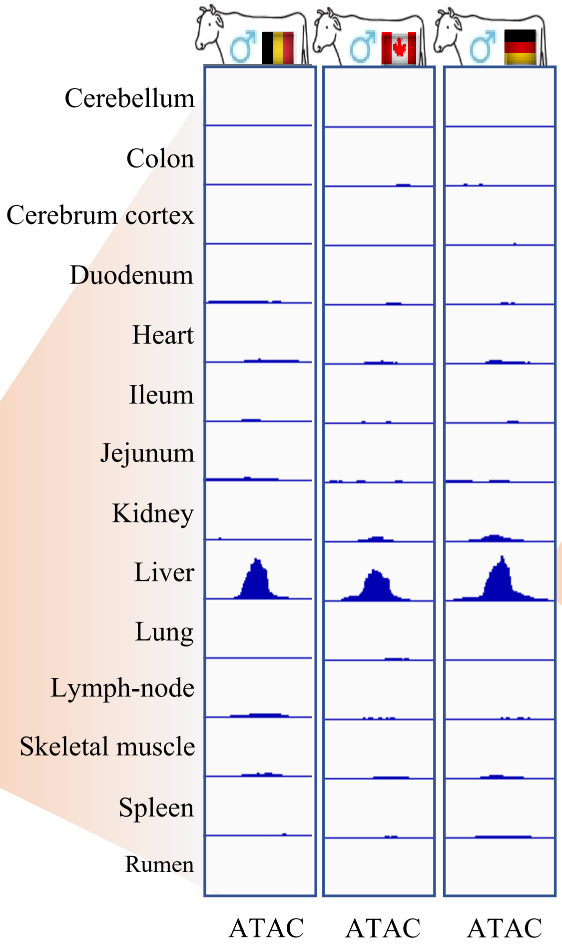
Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

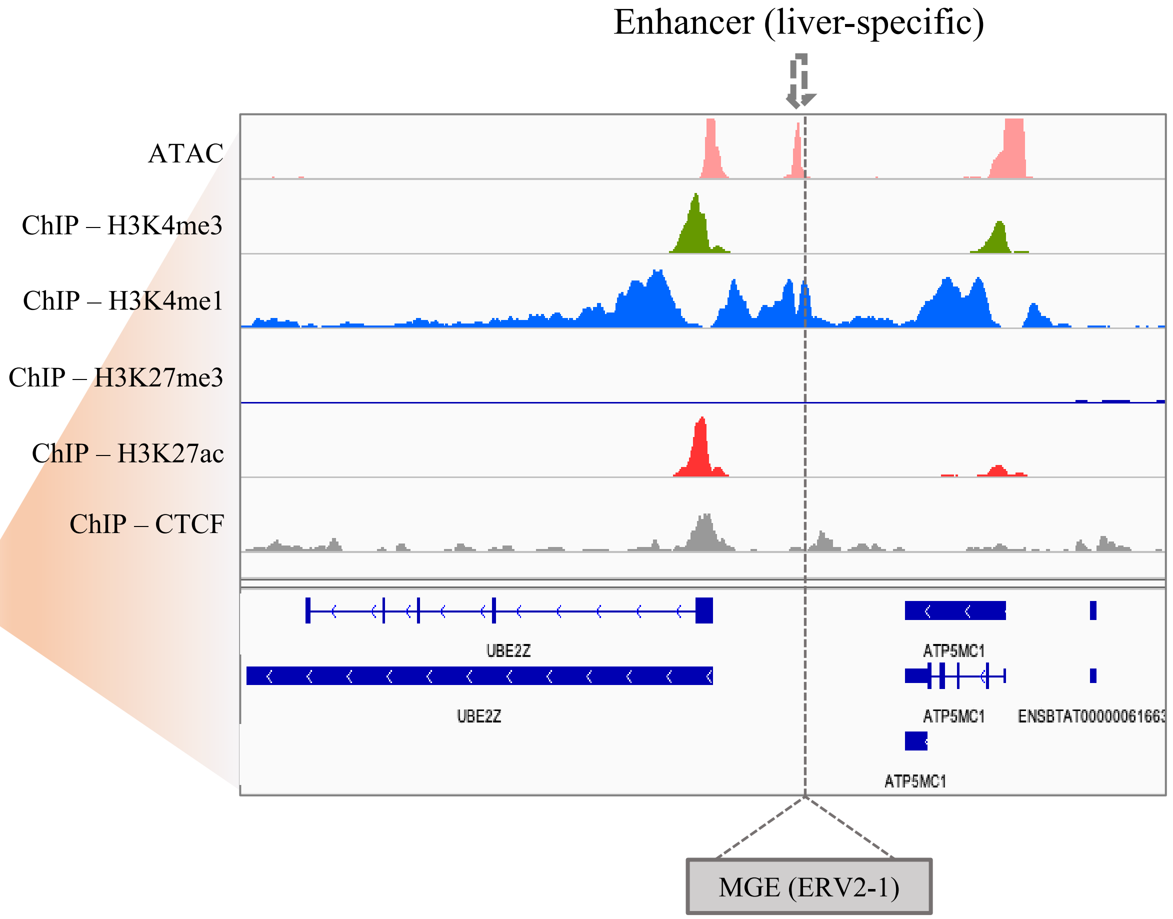
➢ Mobile genetic element co-localized with liver-specific regulatory elements!



Holstein-Friesian (H.F), Brown Swiss (BSW), Grauvieh (GV) and Original Braunvieh (OBV).

> 700 probes in genotyping array (public)





"Given the SV size distribution spike around the size of LTR elements, it is likely such transposable elements will increasingly be identified as a driving force behind bovine phenotypic diversity" - *Pangenome genotyped structural variation improves molecular phenotype mapping in cattle*; *Leonard, Mapel & Pausch, bioRxiv 2023.*

Perspectives & take-home message

Newly annotated genomic features for biology-driven genome selection: the **BovReg** contribution

Genomic features

eQTL in different bovine tissues
(liver, blood, jejunum, rumen, milk cells, muscle, adipose tissue and mammary gland)

Transcriptome map

- 48k genes models including ≥15k potentially novel transcripts
- 1,265 (638 known and 627 novel) miRNAs
- 51,295 transcription start sites (TSS)
- 2,328 TSS-Enhancer regions shared across the three populations

Regulatory regions

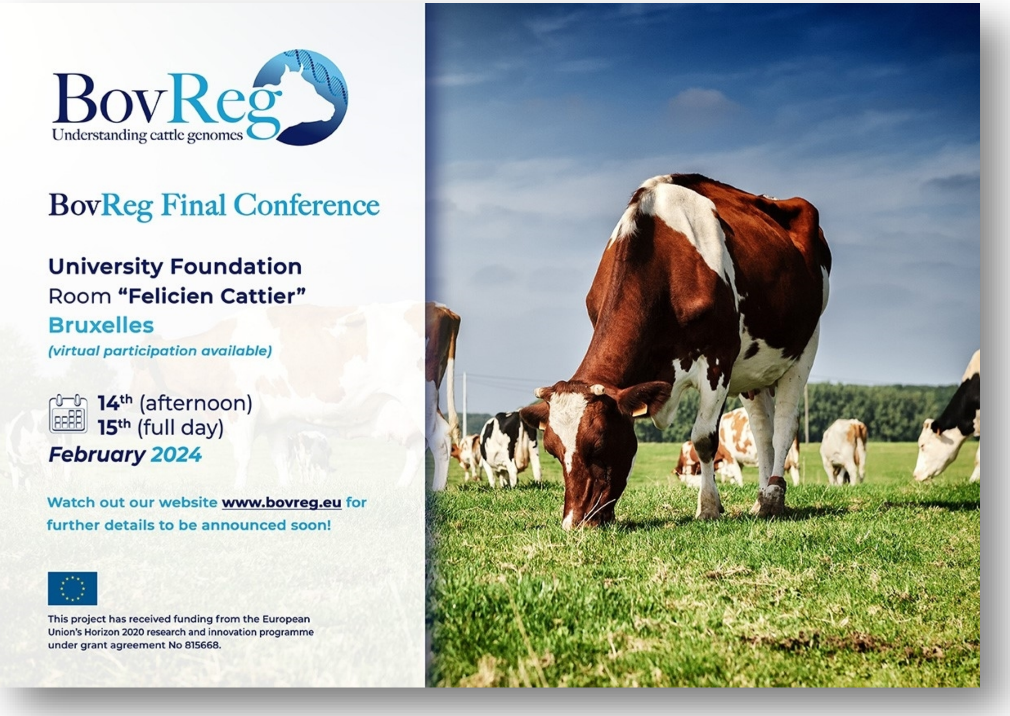
- 925,498 ATAC-Seq consensus peaks
- 327,236 H3K4me3 consensus peaks
- 396,828 H3K4me1 consensus peaks
- 316,262 H3K27me3 consensus peaks
- 454,530 H3K27ac consensus peaks
- 643,198 CTCF consensus peaks

Variations

- Mobile genetic elements (*727 probes, EuroGenomics array*)

LOADING...

WP7/Task7.2: Validating biology-driven genomic selection within and across small breeds
(French, Nordic, Swiss, Canadian beef and dairy breeds)



BovReg
Understanding cattle genomes

BovReg Final Conference

University Foundation
Room **"Felicien Cattier"**
Bruxelles
(virtual participation available)

14th (afternoon)
15th (full day)
February 2024

Watch out our website **www.bovreg.eu** for further details to be announced soon!

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668.

Thank you for your attention!

BovReg
Understanding cattle genomes

www.bovreg.eu

@BovReg BovReg project

@BovReg.H2020 BovReg

WUR, the Netherlands. BovReg General Assembly, October 2022.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668

Disclaimer: the sole responsibility of this presentation lies with the authors. The Research Executive Agency is not responsible for any use that may be made of the information contained therein.