Transcriptome and chromatin structure annotation of liver, CD4+ and CD8+ T cells from four livestock species

## Supplementary file 1: Supplementary figures and tables

## List of Figures

S1 RNA-seq processing pipeline overview. RNA-seq reads were processed with STAR for mapping and Cufflinks for transcript modeling using the reference gene annotation as input and for each sample individually. Modelled transcripts were then merged with Cuffmerge to produce a single gene annotation per species, and the transcripts from this novel gene set were further processed with FEELnc to produce a reliable set of IncRNAs. Reference and novel transcripts and gene expressions were quantified in each sample using the STAR/RSEM pipeline. Differential expression analysis was performed on reference and novel genes with $T P M \geq 0.1$ in at least 2 samples using the R/Bioconductor package edgeR.
S2 RNA-seq mapped read classification into genomic domains. For each sample, RNA-seq mapped reads were classified into exonic (1_exon), intronic (2_intr), intergenic (3ig) and other (4 other) reads.
S3 RNA-seq sample PCA based on the expression of reference genes. Only reference genes with a TPM $\geq 0.1$ in at least 2 samples were used.
S4 Gene expression ratio in males versus females on chromosomes $X$ or $Z$. Gene expression was compared in males vs. females using the base 10 logarithm of the TPM values. $y$-axis: the $\log _{2}$ of the male/female expression ratio for the genes located on the chromosome of interest. For each plot, the median of these expression ratios is indicated in the title. Left: genes located on chromosome 1 used as a control. Right: genes located on chromosome X for mammals and $Z$ for chicken. $x$-axis: the position of the gene on the chromosome of interest. 8
S5 Biological Process GO term enrichment analysis for reference genes differentially expressed between liver and T cells. This analysis was performed for each species individually (in columns: cattle, goat, chicken, pig in that order) and for genes over-expressed in liver (top) and over-expressed in T cells (bottom).
S6 Molecular Function GO term enrichment analysis for reference genes differentially expressed between liver and T cells. This analysis was performed for each species individually (in columns: cattle, goat, chicken, pig in that order) and for genes over-expressed in liver (top) and over-expressed in T cells (bottom).
S7 Cellular Compartment GO term enrichment analysis for reference genes differentially expressed between liver and T cells. This analysis was performed for each species individually (in columns: cattle, goat, chicken, pig in that order) and for genes over-expressed in liver (top) and over-expressed in T cells (bottom).
S8 GO term enrichment analysis for reference genes differentially expressed between liver and T cells in all species. This analysis was performed for each GO (in column: Biological Process, Molecular Function, Cellular Compartment in that order) and for genes overexpressed in liver (top) and over-expressed in T cells (bottom).
S9 Distribution of FR-AgENCODE transcripts into four positional (known, extension, alternative, novel) and three coding (mRNA, IncRNA, otherRNA) classes. (A) Distribution of all FR-AgENCODE transcripts into positional and coding classes; (B) and (C) Distribution of FR-AgENCODE transcripts of each coding class into positional classes, and of each positional class into coding classes respectively.
S10 Novel coding FR-AgENCODE genes enrich the set of blood and T cell annotated genes. (A) Venn diagrams of novel coding genes for each triplet of livestock species (two livestock genes are defined as orthologous if they project to the same human gene). (B) Gene set enrichment analysis on the human orthologs of the 93 cattle, 52 goat, 74 chicken and 26 pig genes, using EnrichR (Kuleshov et al. 2016; http://amp.pharm.mssm.edu/Enrichr/) on the ARCS4 Tissue dataset (https://amp.pharm.mssm.edu/archs4/). (C) Gene names of the 12 genes that are common to two livestock species. Out of these 12, 8 are coding for T cell Receptor Alpha or Beta Variable genes.

S12 LncRNA gene features. Genomic structure features (left) and expression (right) are indicated for protein coding genes (coding) and IncRNA genes (lnc). The expression (TPM) is given for the three tissues and the four species; IncRNA genes are between 8- and 43-fold less expressed than protein coding genes (expression median: 0.4 vs .10 TPM respectively).

S13 Syntenic IncRNAs conserved between the 4 livestock species and human. The 6 syntenic IncRNAs conserved in the 4 livestock species and human are represented in green and their surrounding protein coding genes in orange. Distances between the genes are indicated either in base pair or in kilo base pair (k). A distance of 0 means the genes are overlapping. A distance in red means a lower confidence in the orthology relationship for this species, after inspection of the distances found in the other species. When the IncRNA was not known before, a new gene name is proposed (asterisk).
S14 ATAC-seq read pair summary statistics. For each species and sample (labelled by its tissue and animal number), the number of initial read pairs (1_init), of read pairs obtained after trimming (2_trim), mapping (3_map), proper pairing (4_pair), q10 filtering (5_q10), mitochondrial read removal (6_nonmt), and PCR duplicate read removal (7_nondup) are shown.
S15 ATAC-seq peak size distribution. For each species, the size distribution of tissue and merged (i.e. across all tissues) ATAC-seq peaks are provided.
S16 Density of ATAC-seq peaks around starts of novel (i.e. not known) transcripts for cattle (A), goat (B), chicken (C) and pig (D).

S17 ATAC-seq peak distribution into genomic domains. For each species, are provided: the percentage of ATAC-seq peaks that are exonic, intronic, overlapping the TSS of a reference gene extended by 0,1 or 5 Kb on each side, overlapping the TTS of a reference gene extended by 0,1 or 5 Kb on each side and intergenic.
S18 Per sample promoter accessibility for four reference gene expression quartiles in pig.
S19 Schematic illustration of the per-gene and cross-sample RNA-seq versus ATAC-seq correlation analysis.
S20 Correlation between reference gene expression and chromatin accessibility at the TSS for goat. For each pair of reference gene and ATAC-seq peak overlapping the reference gene TSS extended by 1 Kb on each side, the Pearson correlation was computed between the base 10 logarithm of the gene TMM and the base 10 logarithm of the normalized ATAC-seq signal at the peak. The distribution was then plotted for non DE genes (A, top) and for DE genes (B, bottom).
S21 ATAC-seq sample heatmap and hierarchical clustering based on the 1083 level 4 ATACseq peaks. Pairwise similarity between samples is computed as the Pearson correlation between the base 10 logarithm of the normalized reads of the 1083 ATAC-seq peaks common to 4 species. These similarities are plotted as a heatmap, where samples appear both as rows and columns and are labelled by their species, tissue and the sex of the animal. The color of each heatmap cell also reflects the similarity (Pearson correlation) between each sample pair (the lighter, the higher). Hierarchical clustering is performed using one minus the squared Pearson correlation as a distance and the complete linkage aggregation method.
S22 Relation between non promoter proximal chromatin accessibility conservation and differential accessibility. This figure is the same as main Figure 6 but is done restricting to ATAC-seq peaks not overlapping the TSS+-1kb in any of the species where it is present. Phastcons scores of these ATAC-seq peaks were plotted after dividing the human hits according to both their similarity level (between 1 and 3, x-axis, level 4 removed because there were less than 50 peaks in a category) and their differential accessibility (DA) status (DA in at least one species or DA in none of the 4 species, boxplot color). Although the phastcons score obviously increases with the similarity level, it is clear that, for a given similarity level, the phastcons score is higher for DA human hits than for non DA human hits (all similarity levels, p-values $<1.2 \times 10^{-06}$ overall, Wilcoxon tests) (number of elements the boxplots from left to right: 151,783, 19,732, 11,922, 3,317, 2,100, 1,082).
S23 Hi-C read summary statistics. For each species and animal, are plotted the number of: initial read pairs after sequencing (1_initial), pairs with both reads mapped on the genome (2_reported), pairs in a valid configuration, that is when the sum of the distances from the reads to their next Hindlll restriction sites downstream is comprised between 20bp and 1 Kb (3_valid), pairs after removing PCR duplicates (4_valid_rmdup) and the number of read pairs supporting proximity between two different chromosomes (5_trans).

S24 Method for predicting Hi-C A and B compartments. Illustration of the A/B compartment calling workflow using the interaction matrices of the first chromosome in pig. Upper panels: interaction matrices at different steps of the workflow. Lower panel: the first three eigenvectors are shown along the chromosome to illustrate the relevance of PC\#1 as the discriminative value to segregate bins between A and B compartments.
S25 CTCF motif density and local interaction score (left) and Directionality Index (DI, right) within and around Hi-C Armatus and Armatus/Juicer TADs.
S26 Distribution of Hi-C A and B compartments along each chromosome and for each animal. Genome-wide overview of compartment labels per 500 Kb bin in pig for each animal. A general coherence can be observed across replicates. White regions are devoid of any called compartment.

## List of Tables

S1 Completed experiments. Available data per animal, sample and experimental assay. Green: completed experiment. Red (NA): not available. Hi-C was attempted on liver samples only, and succeeded on all species but cattle. Consequently, ATAC-seq was not attempted on the cattle liver samples either.
S2 Genome and reference gene annotation used for each species. . . . . . . . . . . . . . . 33
S3 Software used in the FR-AgENCODE project. . . . . . . . . . . . . . . . . . . . . . . . . . 34
S4 RNA-seq read mapping statistics. Number and proportion of mapped and uniquely mapped RNA-seq read pairs for each species, replicate and tissue.
S5 Differentially Expressed (DE) reference genes. Number of differentially expressed reference genes obtained by the two statistical models (see main text and Methods)
S6 Genes consistently over-expressed in CD4+ compared to CD8+ or reciprocally, in four livestock species. For the 39 genes consistently seen as over-expressed in CD4+ with respect to CD8+ (10 genes) or reciprocally (29 genes), are indicated: the human gene ID, the gene name in a column that indicates the cell type in which the gene is over-expressed, the TPM in human CD4+ and CD8+ cells from the Blueprint project (CD4-positive, alpha-beta T cell and CD8-positive, alpha-beta T cell respectively), whether or not these expression levels are consistent with the differential behavior observed in livestock, the expression in livestock CD4+ and CD8+ (average across samples) and a reference describing the role of the gene in blood cells when available.
S7 FR-AgENCODE transcript positional classification.
S8 FR-AgENCODE novel coding genes and their orthology with human ..... 40
S9 Differentially Expressed (DE) FR-AgENCODE genes. Number of differentially expressedFR-AgENCODE genes obtained by the two statistical models (see main text and Methods).41

S10 IncRNA classification. Number of classified expressed IncRNAs per species. This table includes monoexonic IncRNAs that represent 57-68\% of these transcripts. Loci are bracketed. Fields with an asterisk (*) indicate the existence of IncRNAs unclassified by FEELnc because they are on unassembled contigs; they represent 217, 717, 2,718 and 83 IncRNAs in cattle, goat, chicken and pig respectively and are not listed here.
S11 FR-AgENCODE transcript coding classification.
S12 Number of ATAC-seq peaks per species.44
S13 Number of differentially accessible (DA) ATAC-seq peaks per species. ..... 45

S14 Hi-C read pair mapping statistics. Number of read pairs of different categories. Initial: total number of sequenced read pairs. Reported: pairs with both reads mapped on the genome. Valid: uniquely mapped pairs with an estimated insert size (sum of the distances from the reads to their next downstream Hindlll genomic sites) between 20bp and 1Kb. Valid.rmdup: valid read pairs after duplication removal that were used to build the interaction matrices. Trans: pairs with reads on different chromosomes.
S15 Statistics of Hi-C TADs and A/B compartments. . . . . . . . . . . . . . . . . . . . . . . . . 47

Figure S1: RNA-seq processing pipeline overview. RNA-seq reads were processed with STAR for mapping and Cufflinks for transcript modeling using the reference gene annotation as input and for each sample individually. Modelled transcripts were then merged with Cuffmerge to produce a single gene annotation per species, and the transcripts from this novel gene set were further processed with FEELnc to produce a reliable set of IncRNAs. Reference and novel transcripts and gene expressions were quantified in each sample using the STAR/RSEM pipeline. Differential expression analysis was performed on reference and novel genes with $T P M \geq 0.1$ in at least 2 samples using the R/Bioconductor package edgeR.


Figure S2: RNA-seq mapped read classification into genomic domains. For each sample, RNA-seq mapped reads were classified into exonic ( 1 exon), intronic (2 intr), intergenic (3 ig) and other (4_other) reads.

Projection on the first two PCs


A
Projection on the first two PCs


C

Projection on the first two PCs


B
Projection on the first two PCs


D

Figure S3: RNA-seq sample PCA based on the expression of reference genes. Only reference genes with a TPM $\geq 0.1$ in at least 2 samples were used.


Figure S4: Gene expression ratio in males versus females on chromosomes $\mathbf{X}$ or $\mathbf{Z}$. Gene expression was compared in males vs. females using the base 10 logarithm of the TPM values. $y$-axis: the $\log _{2}$ of the male/female expression ratio for the genes located on the chromosome of interest. For each plot, the median of these expression ratios is indicated in the title. Left: genes located on chromosome 1 used as a control. Right: genes located on chromosome X for mammals and Z for chicken. $x$-axis: the position of the gene on the chromosome of interest.
 Figure S5: Biological Process GO term enrichment analysis for reference genes differentially expressed between liver and T cells. This analysis was performed for each species individually (in columns: cattle, goat, chicken, pig in that order) and for genes over-expressed in liver (top) and over-expressed in T cells (bottom).

Figure S6: Molecular Function GO term enrichment analysis for reference genes differentially expressed between liver and T cells. This analysis was performed for each species individually (in columns:
cattle, goat, chicken, pig in that order) and for genes over-expressed in liver (top) and over-expressed in T cells (bottom).

Figure S7: Cellular Compartment GO term enrichment analysis for reference genes differentially expressed between liver and T cells. This analysis was performed for each species individually (in columns: cattle, goat, chicken, pig in that order) and for genes over-expressed in liver (top) and over-expressed in T cells (bottom).


[^0]


B


C

Figure S9: Distribution of FR-AgENCODE transcripts into four positional (known, extension, alternative, novel) and three coding (mRNA, IncRNA, otherRNA) classes. (A) Distribution of all FR-AgENCODE transcripts into positional and coding classes; (B) and (C) Distribution of FRAgENCODE transcripts of each coding class into positional classes, and of each positional class into coding classes respectively.
 (A) Venn diagrams of novel coding genes for each triplet of livestock species (two livestock genes are defined as orthologous if they project to the same human gene). (B) Gene set enrichment analysis on the human orthologs of the 93 cattle, 52 goat, 74 chicken and 26 pig genes, using EnrichR (Kuleshov et al. 2016; http:
//amp.pharm.mssm.edu/Enrichr/) on the ARCS4 Tissue dataset (https://amp.pharm.mssm.edu/archs4/). (C) Gene names of the 12 genes that are common to two livestock species. Out of these 12, 8 are coding for T cell Receptor Alpha or Beta Variable genes.


A
Projection on the first two PCs


C

Projection on the first two PCs


B

Projection on the first two PCs


D

Figure S11: RNA-seq sample PCA based on FR-AgENCODE gene expression.

genes (expression median: 0.4 vs . 10 TPM respectively).

Figure S13: Syntenic IncRNAs conserved between the 4 livestock species and human. The 6 syntenic IncRNAs conserved in the 4 livestock species and human are represented in green and their surrounding protein coding genes in orange. Distances between the genes are indicated either in base pair or in kilo base pair (k). A distance of 0 means the genes are overlapping. A distance in red means a lower confidence in
the orthology relationship for this species, after inspection of the distances found in the other species. When the IncRNA was not known before, a new gene name is proposed (asterisk).



Figure S14: ATAC-seq read pair summary statistics. For each species and sample (labelled by its tissue and animal number), the number of initial read pairs (1 init), of read pairs obtained after trimming (2 trim), mapping (3_map), proper pairing (4_pair), q10 filtering (5_q10), mitochondrial read removal (6_nonmt), and PCR duplicate read removal (7_nondup) are shown.


Figure S15: ATAC-seq peak size distribution. For each species, the size distribution of tissue and merged (i.e. across all tissues) ATAC-seq peaks are provided.


Figure S16: Density of ATAC-seq peaks around starts of novel (i.e. not known) transcripts for cattle (A), goat (B), chicken (C) and pig (D)


Reference genomic domain
Figure S17: ATAC-seq peak distribution into genomic domains. For each species, are provided: the percentage of ATAC-seq peaks that are exonic, intronic, overlapping the TSS of a reference gene extended by 0,1 or 5 Kb on each side, overlapping the TTS of a reference gene extended by 0 , 1 or 5 Kb on each side and intergenic.

Figure S18: Per sample promoter accessibility for four reference gene expression quartiles in pig.


B
Figure S20: Correlation between reference gene expression and chromatin accessibility at the TSS for goat. For each pair of reference gene and ATAC-seq peak overlapping the reference gene TSS extended by 1 Kb on each side, the Pearson correlation was computed between the base 10 logarithm of the gene TMM and the base 10 logarithm of the normalized ATAC-seq signal at the peak. The distribution was then plotted for non DE genes (A, top) and for DE genes (B, bottom).


Figure S21: ATAC-seq sample heatmap and hierarchical clustering based on the 1083 level 4 ATAC-seq peaks. Pairwise similarity between samples is computed as the Pearson correlation between the base 10 logarithm of the normalized reads of the 1083 ATAC-seq peaks common to 4 species. These similarities are plotted as a heatmap, where samples appear both as rows and columns and are labelled by their species, tissue and the sex of the animal. The color of each heatmap cell also reflects the similarity (Pearson correlation) between each sample pair (the lighter, the higher). Hierarchical clustering is performed using one minus the squared Pearson correlation as a distance and the complete linkage aggregation method.

status
白 non differential
differential

ATAC-seq human hit orthology level
Figure S22: Relation between non promoter proximal chromatin accessibility conservation and differential accessibility. This figure is the same as main Figure 6 but is done restricting to ATAC-seq peaks not overlapping the TSS+-1kb in any of the species where it is present. Phastcons scores of these ATAC-seq peaks were plotted after dividing the human hits according to both their similarity level (between 1 and 3 , $x$-axis, level 4 removed because there were less than 50 peaks in a category) and their differential accessibility (DA) status (DA in at least one species or DA in none of the 4 species, boxplot color). Although the phastcons score obviously increases with the similarity level, it is clear that, for a given similarity level, the phastcons score is higher for DA human hits than for non DA human hits (all similarity levels, p-values $<1.2 \times 10^{-06}$ overall, Wilcoxon tests) (number of elements the boxplots from left to right: $151,783,19,732,11,922,3,317$, 2,100, 1,082).


Figure S23: Hi-C read summary statistics. For each species and animal, are plotted the number of: initial read pairs after sequencing (1_initial), pairs with both reads mapped on the genome (2reported), pairs in a valid configuration, that is when the sum of the distances from the reads to their next HindIII restriction sites downstream is comprised between 20bp and 1 Kb (3_valid), pairs after removing PCR duplicates (4_valid_rmdup) and the number of read pairs supporting proximity between two different chromosomes (5_trans).

1. Merge the biological replicates by adding their read counts for each pair of bins


The Hi-C contact matrices from the 4 replicates (Sus scrofa, chromosome 1)
2. Normalization: matrix balancing and observed/expected (based on distance) (per chromosome)
3. Pearson correlation matrix from each pair of bins


Raw matrix, normalized matrix and correlation matrix (Sus scrofa, chromosome 1)
3. Principal Component Analysis on the bins

=> The sign of the $1^{\text {st }}$ eigenvector (PC\#1) defines the transitions between compartments.

Figure S24: Method for predicting Hi-C A and B compartments. Illustration of the $\mathrm{A} / \mathrm{B}$ compartment calling workflow using the interaction matrices of the first chromosome in pig. Upper panels: interaction matrices at different steps of the workflow. Lower panel: the first three eigenvectors are shown along the chromosome to illustrate the relevance of PC\#1 as the discriminative value to segregate bins between A and B compartments.


## Chicken





C



Figure S25: CTCF motif density and local interaction score (left) and Directionality Index (DI, right) within and around Hi-C Armatus and Armatus/Juicer TADs.

Figure S26: Distribution of Hi-C A and B compartments along each chromosome and for each animal. Genome-wide overview of compartment labels per 500 Kb bin in pig for each animal. A general coherence can be observed across replicates. White regions are devoid of any called compartment.

Table S1: Completed experiments. Available data per animal, sample and experimental assay. Green: completed experiment. Red (NA): not available. Hi-C was attempted on liver samples only, and succeeded on all species but cattle. Consequently, ATAC-seq was not attempted on the cattle liver samples either.

RNA-seq

| Cattle | cattle1 | cattle2 | cattle3 | cattle4 | Goat | goat1 | goat2 | goat3 | goat4 |
| :--- | :---: | :---: | :---: | :---: | :--- | :--- | :--- | :--- | :--- |
| cd4 |  |  |  |  |  |  |  |  |  |
| cd8 | NA |  |  |  |  |  |  |  |  |
| liver |  |  |  |  |  |  |  |  |  |
| Chicken | chicken1 | chicken2 | chicken3 | chicken4 | lig |  |  |  |  |
| cd4 | NA | NA |  |  |  | pig1 | pig2 | pig3 | pig4 |
| cd8 | NA | NA | NA |  | NA |  |  |  |  |
| liver |  |  |  |  |  |  |  |  |  |

ATAC-seq

| Cattle | cattlel | cattle2 | cattle3 | cattle4 | Goat | goat 1 | goat2 | goat3 | goat4 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| cd 4 |  |  |  |  | cd4 |  | NA |  |  |
| cd8 |  |  |  |  | cd8 |  |  |  |  |
| iver | NA | NA | NA | NA | liver |  |  |  |  |
| Chicken | chickenl | chicken2 | chicken3 | chicken4 | Pig | pig1 | pig2 | pig3 | pig4 |
| cd4 |  |  |  | NA | cd4 |  |  |  |  |
| cd8 | NA |  | NA | NA | cd8 |  |  |  |  |
| liver |  |  |  |  | iver |  |  |  | NA |

Hi-C

| Cattle | cattle1 | cattle2 | cattle3 | cattle4 | Goat | goat1 | goat2 | goat3 | goat4 |
| :--- | :---: | :---: | :---: | :---: | :--- | :--- | :--- | :--- | :--- |
| iver | NA | NA | NA | NA | liver |  |  |  |  |
| Chicken | chicken1 | chicken2 | chicken3 | chicken4 | Pig | pig1 | pig2 | pig3 | pig4 |
| iver |  |  |  |  | iver |  |  |  |  |

Table S2: Genome and reference gene annotation used for each species.

| Species | Reference genome assembly |  |  |  | Reference gene annotation |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Version | \# fasta <br> sequences | Size <br> (in Gb) | ContigN50 <br> (in Mb) | Source | \# genes | \# transcripts |
|  | UMD3.1 | 3,317 | 2.67 | 0.1 | Ensembl v90 | 24,616 | 26,740 |
| Goat | CHIR_ARS1 | 29,907 | 2.92 | 26.2 | NCBI vCHIR_ARS1 | 28,931 | 53,266 |
| Chicken | GalGal5 | 23,475 | 1.23 | 2.9 | Ensembl v90 | 24,881 | 38,118 |
| Pig | Sscrofa11.1 | 613 | 2.5 | 48.2 | Ensembl v90 | 25,880 | 49,448 |

Table S3: Software used in the FR-AgENCODE project.

| Software name | Software version | Software URL |
| :--- | :--- | :--- |
| R | 3.3 .3 | https://www.r-project.org/about.html |
| python | 2.7 .2 | https://www.python.org/ |
| samtools | 1.3 .1 | http://www.htslib.org/ |
| bedtools | 2.26 .0 | http://bedtools.readthedocs.io/en/latest/ |
| kentUtils | 302.1 .0 | https://github.com/ENC0DE-DCC/kentUtils |
| bwtool | Nov 2015 version | https://github.com/CRG-Barcelona/bwtool/wiki |
| emboss | 6.4 .0 .0 | http://emboss.sourceforge.net/ |
| fastqc | 0.11 .2 | https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ |
| cutadapt | 1.8 .3 | http://cutadapt.readthedocs.io/en/stable/guide.html |
| trim galore | 0.4 .0 | https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ |
| picardtools | 2.1 .1 | https://broadinstitute.github.io/picard/ |
| STAR | $2.5 .1 b$ | https://github.com/alexdobin/STAR |
| cufflinks | 2.2 .1 | http://cole-trapnell-lab.github.io/cufflinks/ |
| RSEM | 1.3 .0 | http://deweylab.biostat.wisc.edu/rsem/README.html |
| FEELnc | 0.1 .0 | https://github.com/tderrien/FEELnc |
| Bowtie2 | 2.3 .3 .1 | http://bowtie-bio.sourceforge.net/bowtie2/index.shtml |
| Macs2 | 2.1 .1 .20160309 | https://chipster.csc.fi/manual/macs2.html |
| fimo | 4.11 .1 | http://meme-suite.org/doc/fimo.html |
| HiC-Pro | 2.9 .0 | http://nservant.github.io/HiC-Pro/ |
| Armatus | 2,1 | https://www.cs.cmu.edu/~ckingsf/software/armatus/ |
| HiTC | 1.18 .1 | https://bioconductor.org/packages/release/bioc/html/HiTC.html |
| Juicer tools | 1.5 .3 | https://github.com/theaidenlab/juicer/wiki/Juicer-Tools-Quick-Start |
| Last | 956 | http://last.cbrc.jp/ |

Table S4: RNA-seq read mapping statistics. Number and proportion of mapped and uniquely mapped RNA-seq read pairs for each species, replicate and tissue.

| Species | Tissue | Animal | \# read pairs | \# read pairs mapped | \% read pairs mapped | \# uniquely mapped read pairs | \% uniquely mapped read pairs (out of mapped) | \% uniquely mapped read pairs (out of total) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Cattle | cd4 | cattle1 | 115398649 | 111206516 | 96,4 | 108830546 | 97,9 | 94,3 |
|  | cd4 | cattle2 | 118918186 | 114474132 | 96,3 | 112127763 | 98,0 | 94,3 |
|  | cd4 | cattle3 | 121389996 | 113765363 | 93,7 | 108717357 | 95,6 | 89,6 |
|  | cd4 | cattle4 | 116083510 | 112298028 | 96,7 | 110019926 | 98,0 | 94,8 |
|  | cd8 | cattle2 | 117869348 | 113617513 | 96,4 | 111375385 | 98,0 | 94,5 |
|  | cd8 | cattle3 | 118910270 | 114458706 | 96,3 | 111966825 | 97,8 | 94,2 |
|  | cd8 | cattle4 | 121883485 | 117905219 | 96,7 | 115423917 | 97,9 | 94,7 |
|  | liver | cattle1 | 115371212 | 111956767 | 97,0 | 110499020 | 98,7 | 95,8 |
|  | liver | cattle2 | 119870689 | 116728118 | 97,4 | 115092798 | 98,6 | 96,0 |
|  | liver | cattle3 | 136710805 | 132763513 | 97,1 | 130973598 | 98,7 | 95,8 |
|  | liver | cattle4 | 118889544 | 115841946 | 97,4 | 114125584 | 98,5 | 96,0 |
| Goat | cd4 | goat1 | 133365348 | 130269121 | 97,7 | 123559681 | 94,8 | 92,6 |
|  | cd4 | goat2 | 132983117 | 128521637 | 96,6 | 121198446 | 94,3 | 91,1 |
|  | cd4 | goat3 | 133831464 | 129566423 | 96,8 | 122990885 | 94,9 | 91,9 |
|  | cd4 | goat4 | 130150700 | 126874881 | 97,5 | 121461169 | 95,7 | 93,3 |
|  | cd8 | goat1 | 130970787 | 127546631 | 97,4 | 120904297 | 94,8 | 92,3 |
|  | cd8 | goat2 | 131654878 | 128355710 | 97,5 | 121468794 | 94,6 | 92,3 |
|  | cd8 | goat3 | 132923610 | 129276688 | 97,3 | 123235599 | 95,3 | 92,7 |
|  | cd8 | goat4 | 131413248 | 127239980 | 96,8 | 121762207 | 95,7 | 92,7 |
|  | liver | goat1 | 112623152 | 110432315 | 98,1 | 105943818 | 95,9 | 94,1 |
|  | liver | goat2 | 120492856 | 117992580 | 97,9 | 112453956 | 95,3 | 93,3 |
|  | liver | goat3 | 133931176 | 131150414 | 97,9 | 126458023 | 96,4 | 94,4 |
|  | liver | goat4 | 125330385 | 122259872 | 97,6 | 118111570 | 96,6 | 94,2 |
| Chicken | cd4 | chicken3 | 91907306 | 86645024 | 94,3 | 83374510 | 96,2 | 90,7 |
|  | cd4 | chicken4 | 90843387 | 79772579 | 87,8 | 76778803 | 96,2 | 84,5 |
|  | cd8 | chicken4 | 90476514 | 82431332 | 91,1 | 79685312 | 96,7 | 88,1 |
|  | liver | chicken1 | 119681527 | 114944560 | 96,0 | 111122525 | 96,7 | 92,8 |
|  | liver | chicken2 | 107777068 | 103770521 | 96,3 | 100787680 | 97,1 | 93,5 |
|  | liver | chicken3 | 121616481 | 116940968 | 96,2 | 113315177 | 96,9 | 93,2 |
|  | liver | chicken4 | 124089440 | 119195187 | 96,1 | 115442821 | 96,9 | 93,0 |
| Pig | cd4 | pig2 | 116273646 | 113997546 | 98,0 | 109051735 | 95,7 | 93,8 |
|  | cd4 | pig3 | 118260498 | 115042891 | 97,3 | 110755493 | 96,3 | 93,7 |
|  | cd4 | pig4 | 114159672 | 111934574 | 98,1 | 108379228 | 96,8 | 94,9 |
|  | cd8 | pig1 | 231339734 | 225402022 | 97,4 | 219199568 | 97,2 | 94,8 |
|  | cd8 | pig2 | 116383341 | 114513826 | 98,4 | 110146147 | 96,2 | 94,6 |
|  | cd8 | pig3 | 116696412 | 113825826 | 97,5 | 110613640 | 97,2 | 94,8 |
|  | cd8 | pig4 | 117449531 | 114751230 | 97,7 | 108598741 | 94,6 | 92,5 |
|  | liver | pig1 | 116242745 | 113660238 | 97,8 | 110849470 | 97,5 | 95,4 |
|  | liver | pig2 | 117818284 | 115364464 | 97,9 | 111627222 | 96,8 | 94,7 |
|  | liver | pig3 | 110567402 | 108214816 | 97,9 | 105590955 | 97,6 | 95,5 |
|  | liver | pig4 | 114769769 | 112137997 | 97,7 | 109296648 | 97,5 | 95,2 |

Table S5: Differentially Expressed (DE) reference genes. Number of differentially expressed reference genes obtained by the two statistical models (see main text and Methods)

| Differential analysis model | Tissue 1 | Tissue 2 | $\begin{gathered} \hline \log (\text { Tissue } 2 / \\ \text { Tissuel) } \end{gathered}$ | Number of DE reference genes |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | Cattle | Goat | Chicken | Pig |
| 1 (tissue pairs) | cd4 | cd8 | $>0$ | 751 | 1098 | 1504 | 867 |
|  | cd4 | cd8 | $<0$ | 463 | 748 | 592 | 604 |
|  | cd4 | liver | $>0$ | 5048 | 6303 | 5112 | 5725 |
|  | cd4 | liver | $<0$ | 3887 | 4304 | 3351 | 3653 |
|  | cd8 | liver | $\times 0$ | 4890 | 6134 | 4146 | 5665 |
|  | cd8 | liver | $<0$ | 3911 | 4397 | 3087 | 3792 |
| 2 (tcell vs liver) | cd | liver | $>0$ | 4992 | 6188 | 4307 | 5666 |
|  | cd | liver | $<0$ | 3943 | 4384 | 2640 | 3772 |

Table S6: Genes consistently over-expressed in CD4+ compared to CD8+ or reciprocally, in four livestock species. For the 39 genes consistently seen as over-expressed in CD4+ with respect to CD8+ (10 genes) or reciprocally (29 genes), are indicated: the human gene ID, the gene name in a column that indicates the cell type in which the gene is overexpressed, the TPM in human CD4+ and CD8+ cells from the Blueprint project (CD4-positive, alpha-beta T cell and CD8-positive, alpha-beta T cell respectively), whether or not these expression levels are consistent with the differential behavior observed in livestock, the expression in livestock CD4+ and CD8+ (average across samples) and a reference describing the role of the gene in blood cells when available.

| Hemmentid | Gense marne (in colamen correxponding to averexprexalon) |  | TPM in buma |  | Comatitumey betwern likrextock and buman (FC>1) | TPM in catto |  | TPM in toxt |  | TMM in chackem |  | TPM in pob |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | CD4. | cns- | cD. | cns- |  | CD4. | chs. | CD4. | cns. | cD. | cns. | CD. | cns. |
| Ensc000000102245 [1] | CD40LG | . | 270 | 0.5 | ym | 97.9 | 76 | 51.6 | 105 | 164.8 | 75 | 153.4 | 28.7 |
| EN3000000106537 | TSPMN 13 | . | 0.4 | 0.5 | m | 118.5 | 62.6 | 93.9 | 32.6 | 4.0 | 0.5 | 3.4 | 1.0 |
| Ensc0000001074:5s [2] | gatas | . | 60 | 4.0 | yox | 37.7 | 11.5 | 433 | 15.5 | 1135 | 76 | 51.1 | 300 |
| EN3600000113602 | IL.1ML1 | . | NA | Q. 1 | NA | 4.7 | 0.3 | 52 | 2.2 | 53 | 0.2 | 4.2 | 1.0 |
| Ensc000000126333 [3] | CCKT | . | 3120 | 1020 | mox | 33.7 | 11.3 | 164.5 | 44.6 | 96.2 | 21.7 | \$7.1 | 332 |
| EN3600000130396 | ArDs | . | 0.9 | 0.3 | ym | 6.6 | 2.9 | 178 | 2.5 | 3.3 | 0.5 | 2.4 | 0.1 |
| 2nsc000000163599 [4] | CtLa | . | 390 | 20 | ma | 6.1 | 06 | 19.5 | 14 | 6336 | 16.9 | 150 | 2.2 |
| 2N300000017s199 | zc3H12d | . | 170 | 6.0 | max | 30.5 | 13.6 | 457 | 9.4 | 1,384.7 | 134.7 | 263 | 4.5 |
| 2Nsccoocoulis 362 [s] | CD2s | . | 161.0 | 740 | yox | 344 | 9.3 | 91.5 | 11.0 | 2083 | 14.8 | 1074 | 27.6 |
| 2nscroob00178573 [6] | mar | - | 30 | 1.0 | yox | 13.2 | 1.5 | 11.5 | 5.4 | 6.5 | 03 | 150 | 2.5 |
| 2N3c000000139367 | $\bigcirc$ | Ncwil 1 | NA | NA | NA | 5.9 | 450 | so | 159 | 32 | 10.1 | 00 | 02 |
| 2Nscroob001699352 [7] | . | ADMEI | 06 | 20 | yox | 3.9 | 17.9 | 0.6 | 73 | 57.5 | 2193 | 2.7 | 15.2 |
| ENSc00000136966 | . | mycaty | a. 1 | 0.2 | yox | 1.7 | 72 | 0.2 | 0.7 | 00 | 28 | 00 | 14 |
| N3xc00000136573 | . | bLE | 1.0 | 4.0 | ma | 15.9 | 270 | 3.5 | 5312 | 1.9 | 15.4 | 1.5 | 32.8 |
| 2N3c00000182985 | . | Canme | a, | NA | NA | 4.1 | 21.4 | 1.1 | 3.5 | 1.6 | 3.9 | 00 | 0.2 |
| 2nscoocou0271503 [8] | . | CCLS | so | 380 | yox | 33.0 | 1,205. s | 162 | 2,009.6 | 1.2 | 102.1 | 1902 | 2,463.9 |
| 2Nscr000000172116 [9] | . | chsm | 20 | 2100 | yox | 9.7 | 2607 | 17.6 | 514.5 | 0, | 1730 | 1.1 | 115.4 |
|  | . | CKrma | 30 | 340 | max | 23 | 16.3 | 4.0 | 16.2 | 0, 5 | 520 | 3.1 | 37.6 |
| 2N3c00000100592 | . | namal | 4.0 | 3.0 | mo | 53 | so | 2.1 | 2.5 | 03 | a3 | 5.5 | 5.7 |
| ENscrococou39664 [11] | . | napke | a 3 | 0.9 | ma | 1.5 | 8.2 | 0.9 | 10.1 | 30 | 11.7 | 14.9 | 60.1 |
| 2N3c000002113833 | . | mapz | 06 | 0.4 | $\pm$ | a. 1 | 0.7 | 1.9 | 6.9 | a. 1 | as | 0.1 | 09 |
| 2Nsc00000016350s [12] | . | EOMES | 0.9 | 1.0 | yox | 1.0 | 11.3 | 1.0 | 175 | 2.9 | 133.4 | 1.0 | 12.5 |
| 2N3000000139132 | . | FGD4 | 0.2 | 20 | yox | 3.3 | 5.5 | as | 23 | 0.9 | 6.5 | 0.2 | 53 |
| 2Nsc000000160219 | . | GAB3 | 4.0 | 4.0 | mo | 3.3 | 56 | 9.2 | 23.1 | 4.4 | 38.4 | 59 | 27.2 |
| EN36000001123245 | . | Gpip1s | 170 | 260 | ym | 78 | 22.8 | 9.0 | 27.8 | 13.5 | 104.8 | 390 | 2976 |
| ENsc000001003s5 [13] | . | IL.2K8 | 70 | 9.0 | ma | 42.7 | 1316 | 134.9 | 573.8 | 302 | 1907 | 62.6 | 2082 |
| ENSc000000157404 | . | krr | 09 | 20 | ma | 2.4 | 7.9 | 24 | 6.4 | a. 1 | 03 | 00 | 1.2 |
| 2Nscou000036730 [14] | . | LAT2 | 0.5 | 4.0 | ma | 46.1 | 1363 | 200 | 462 | 20 | 11.9 | 22.1 | 123.7 |
| EN3600000159067 | . | Itray | 650 | 550 | mo | 24.0 | 98.1 | 260 | 52.3 | 12.3 | 41.5 | 4.6 | 21.4 |
| 2N3c00000013s186 | . | mass | 0.1 | NA | NA | 0.9 | 5.9 | 2.1 | 6.4 | 16.4 | 32.9 | 00 | 00 |
| EN3600000100311 | . | pmara | 0.9 | NA | NA | 6.1 | 23.5 | 2.2 | 56 | 0.2 | 09 | 0.7 | 5.6 |
| ENsc000000113936 | - | PLEE | 20 | 22.0 | ym | 10.4 | 32.1 | 9.1 | 32.4 | 36.5 | 430 | 53 | 38.9 |
| EN3600000141936 | . | pknmels | so | 20 | mo | 1.5 | 1.5 | 3.4 | 30 | 76 | 3.3 | 3.4 | 3.4 |
| 2xccoocoul9 9915 | - | Muscer 1 A | a. 1 | 0.1 | mo | 3.3 | 5.3 | 2.5 | 10.5 | 276 | 61.5 | 0.3 | 30 |
| N35000000119729 | . | кHOO | 60 | 4.0 | $\pm$ | 18.1 | 74.3 | 14.5 | 45.9 | 0.6 | 8.9 | 2.9 | 16.5 |
| 2N3c0000013615s [15] | . | SPMr | NA | 1.0 | ma | 1.0 | 23.5 | 06 | 11.3 | 20 | 66.0 | 0.1 | 40.5 |
| 2N3000000204634 | . | TBCIDS | 0.6 | 20 | ma | 12.6 | 62.9 | 5.5 | 170 | 03 | 0.6 | 0.2 | as |
| 2N3000000070739 | . | TEGK2 | 6.0 | 60 | mo | 4.5 | 152 | 6.9 | 29.1 | 3.3 | 9.5 | 3.6 | 39.0 |
| ENsc00000169914 | . | ттет | 0.4 | as | yox | 0.4 | 20 | 0.2 | 0.7 | 1.0 | 2.5 | 0.1 | 0.6 |

[^1]







Table S7: FR-AgENCODE transcript positional classification.

| Species | Total \# | known |  | extersion |  | afternative |  | novel |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | \# | \% of total | \# | \% ot total | \# | 90 at total | \# | \%o ot total |
| Catte | 84,971 | 11,736 | 13.8 | 2,500 | 2.9 | 40,813 | 48.0 | 29,922 | 35.2 |
| Gout | 78,091 | 29,520 | 37.8 | 2,583 | 3.3 | 28,891 | 37.0 | 17,097 | 21.9 |
| Cricken | 57,817 | 15,890 | 27.5 | 3,018 | 5.2 | 28,005 | 48.5 | 10,838 | 18.7 |
| Pig | 77,540 | 23,921 | 30.8 | 2,702 | 3.5 | 35,002 | 45.2 | 15,855 | 20.4 |
| Species | knownmRNA |  | knownincRNA |  | known.otherRNA |  |  |  |  |
|  | \# | \% of known | \# | \% of known | \# | \% of known |  |  |  |
| Catte | 11,576 | 98.6 | 13 | 0.1 | 147 | 1.3 |  |  |  |
| Goat | 26,973 | 91.4 | 1,103 | 3.9 | 1,384 | 4.7 |  |  |  |
| Cricken | 14,705 | 92.9 | 882 | 5.5 | 249 | 1.6 |  |  |  |
| Pig | 23,701 | 99.1 | 96 | 0.4 | 124 | 0.5 |  |  |  |
| Species | extersionmRNA |  | extersionincRNA |  | edersion.other RNA |  |  |  |  |
|  | \# | $9 \%$ of extersion | \# | \% of extersion | \# | $9 \%$ of extersion |  |  |  |
| Catte | 2,497 | 99.9 | 0 | 0.0 | 3 | 0.1 |  |  |  |
| Goat | 2,351 | 91.0 | 100 | 3.9 | 132 | 5.1 |  |  |  |
| Cricken | 2,982 | 98.8 | 33 | 1.1 | 3 | 0.1 |  |  |  |
| Pig | 2,094 | 99.7 | 5 | 0.2 | 3 | 0.1 |  |  |  |
| Species | alernativernina |  | aternutivelincRNA |  | aliernative.otherRNA |  |  |  |  |
|  | \# | \% of alternative | \# | 90 of alfernative | \# | 90 of alternative |  |  |  |
| Catte | 40,770 | 99.9 | 0 | 0.0 | 43 | 0.1 |  |  |  |
| Goat | 26,554 | 91.9 | 984 | 3.4 | 1,353 | 4.7 |  |  |  |
| Cricken | 27,470 | 97.9 | 399 | 1.4 | 196 | 0.7 |  |  |  |
| Pig | 34,822 | 99.3 | 226 | 0.6 | 14 | 0.0 |  |  |  |
| Species | novel.mRNA |  | novelincRNA |  | novel.atherRNA |  |  |  |  |
|  | \# | \% of novel | \# | \% of novel | \# | \% of novel |  |  |  |
| Catte | 4,958 | 16.6 | 22,711 | 75.9 | 2,253 | 7.5 |  |  |  |
| Goat | 2,949 | 17.2 | 11,017 | 67.9 | 2,531 | 14.8 |  |  |  |
| Cricken | 2,350 | 21.7 | 6,797 | 62.7 | 1,091 | 15.6 |  |  |  |
| Pig | 2,504 | 15.8 | 12,284 | 77.5 | 1,067 | 6.7 |  |  |  |

Table S8: FR-AgENCODE novel coding genes and their orthology with human

| Species | \# FR- <br> AgENCODE <br> genes | \# novel FRAgENCODE genes | \# novel coding FR- <br> AgENCODE genes | that can be projected to a human gene |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | \# | \% |
| Cattle | 34,296 | 19,324 | 15,611 | 166 | 12.3 |
| Goat | 28,537 | 11,545 | 16,437 | 85 | 6.4 |
| Chicken | 20,408 | 6,597 | 13,149 | 78 | 15.5 |
| Pig | 24,570 | 10,238 | 14,054 | 38 | 9.5 |

Table S9: Differentially Expressed (DE) FR-AgENCODE genes. Number of differentially expressed FR-AgENCODE genes obtained by the two statistical models (see main text and Methods).

| Differential analy sis model | Tissue 1 | Tissue 2 | $\begin{gathered} \hline \log (\text { Tissue } 2 / \\ \text { Tissuel }) \end{gathered}$ | Number of DE FR-AgENCODE genes |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | Cattle | Goat | Chicken | Pig |
| 1 (tissue pairs) | cd4 | cd8 | $>0$ | 1450 | 1614 | 1940 | 1211 |
|  | cd4 | cd8 | $<0$ | 923 | 1056 | 974 | 792 |
|  | cd4 | liver | $>0$ | 12231 | 11782 | 7426 | 9709 |
|  | cd4 | liver | $<0$ | 10311 | 7152 | 5134 | 6386 |
|  | cd8 | liver | >0 | 11810 | 11542 | 5899 | 9880 |
|  | cd8 | liver | $<0$ | 10573 | 7408 | 4331 | 6782 |
| 2 (tcell vs liver) | cd | liver | $>0$ | 12218 | 11773 | 6246 | 10040 |
|  | cd | liver | $<0$ | 10818 | 7448 | 3929 | 6910 |


| cludes monoex asterisk (*) i contigs; they are not listed | exonic Inc dicate th represen here. | NAs th existen 217, 717 | represe <br> of IncP $2,718$ | 57-68\% As unc 83 In | these <br> ified <br> NAs in | anscripts FEELnc <br> ttle, goa | Loci are cause chicke | rackete <br> y are nd pig | ields <br> unass pectiv |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Intergen | RNAs | 88-94\% |  | Geni | NAs (loci) | 12\% |
|  | Total (transcript) | Sam <br> upstr. | rand downstr. | s1kb <br> D | $>1 \mathrm{~kb}$ | Convergent | Exonic antisense | Intronic antisense | Intronic sense |
| Cattle | 22724 | $\begin{gathered} 5522 \\ (4671) \end{gathered}$ | $\begin{gathered} 6438 \\ (5460) \end{gathered}$ | $\begin{gathered} 1450 \\ (1070) \end{gathered}$ | $\begin{gathered} 4494 \\ (3588) \end{gathered}$ | $\begin{gathered} 3328 \\ (2617) \end{gathered}$ | $\begin{aligned} & 1235 \\ & (851) \end{aligned}$ | $\begin{gathered} 241 \\ (173) \end{gathered}$ | $\begin{gathered} \hline 16 \\ (16) \\ \hline \end{gathered}$ |
| Goat | 13864 | $\begin{gathered} 2431 \\ (1940) \\ \hline \end{gathered}$ | $\begin{gathered} 4050 \\ (3312) \end{gathered}$ | $\begin{gathered} 1744 \\ (1241) \end{gathered}$ | $\begin{gathered} 2908 \\ (2233) \\ \hline \end{gathered}$ | $\begin{gathered} 1966 \\ (1496) \\ \hline \end{gathered}$ | $\begin{gathered} 437 \\ (220) \end{gathered}$ | $\begin{gathered} 219 \\ (124) \\ \hline \end{gathered}$ | $\begin{aligned} & 109 \\ & (89) \\ & \hline \end{aligned}$ |
| Chicken | 7502* | $\begin{gathered} 1618 \\ (1390) \end{gathered}$ | $\begin{gathered} 2045 \\ (1728) \end{gathered}$ | $\begin{gathered} \hline 687 \\ (539) \\ \hline \end{gathered}$ | $\begin{aligned} & 1194 \\ & (983) \\ & \hline \end{aligned}$ | $\begin{aligned} & 1093 \\ & (903) \end{aligned}$ | $\begin{gathered} 646 \\ (510) \end{gathered}$ | $\begin{gathered} 157 \\ (124) \end{gathered}$ | $\begin{gathered} 62 \\ (56) \\ \hline \end{gathered}$ |
| Pig | 12587* | $\begin{gathered} 2508 \\ (2100) \\ \hline \end{gathered}$ | $\begin{gathered} 3539 \\ (3075) \end{gathered}$ | $\begin{aligned} & 1117 \\ & (810) \end{aligned}$ | $\begin{gathered} 2466 \\ (2060) \end{gathered}$ | $\begin{gathered} 1746 \\ (1436) \end{gathered}$ | $\begin{gathered} 935 \\ (669) \end{gathered}$ | $\begin{gathered} 216 \\ (178) \end{gathered}$ | $\begin{gathered} 60 \\ (58) \end{gathered}$ |

## Table S11: FR-AgENCODE transcript coding classification.

| Species | Total \#\# | mRNA |  | IncRNA |  | otherRNA |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | \# | \% of total | \# | \% ot total | \# | \% ot total |  |
| Cattle | 84,971 | 59,801 | 70.4 | 22,724 | 26.7 | 2,446 | 2.9 |  |
| Goat | 78,091 | 58,827 | 75.3 | 13,864 | 17.8 | 5,400 | 6.9 |  |
| Chicken | 57,817 | 47,567 | 82.3 | 8,111 | 14.0 | 2,139 | 3.7 |  |
| Pig | 77,540 | 63,721 | 82.2 | 12,611 | 16.3 | 1,208 | 1.6 |  |
| Species | mRNA. known |  | mRNA.extension |  | mRNA.alternative |  | mRNA.novel |  |
|  | \# | \% of mRNA | \# | \% of mRNA | \# | \% of mRNA | \# | \% of mRNA |
| Cattle | 11,576 | 19.4 | 2,497 | 4.2 | 40,770 | 68.2 | 4,958 | 8.3 |
| Goat | 26,973 | 45.9 | 2,351 | 4.0 | 26,554 | 45.1 | 2,949 | 5.0 |
| Chicken | 14,765 | 31.0 | 2,982 | 6.3 | 27,470 | 57.8 | 2,350 | 4.9 |
| Pig | 23,701 | 37.2 | 2,694 | 4.2 | 34,822 | 54.6 | 2,504 | 3.9 |
| Species | IncRNA.known |  | IncRNA.extension |  | IncRNA.alternative |  | IncRNA.novel |  |
|  | \# | \% of IncRNA | \# | \% of IncRNA | \# | \% of IncRNA | \# | \% of IncRNA |
| Cattle | 13 | 0.1 | 0 | 0.0 | 0 | 0.0 | 22,711 | 99.9 |
| Goat | 1,163 | 8.4 | 100 | 0.7 | 984 | 7.1 | 11,617 | 83.8 |
| Chicken | 882 | 10.9 | 33 | 0.4 | 399 | 4.9 | 6,797 | 83.8 |
| Pig | 96 | 0.8 | 5 | 0.0 | 226 | 1.8 | 12,284 | 97.4 |
| Species | otherRNA.known |  | otherRNA.extension |  | otherRNA. alternative |  | otherRNA.novel |  |
|  | \# | \% of otherRNA | \# | \% of otherRNA | \# | \% of otherRNA | \# | \% of otherRNA |
| Cattle | 147 | 6.0 | 3 | 0.1 | 43 | 1.8 | 2,253 | 92.1 |
| Goat | 1,384 | 25.6 | 132 | 2.4 | 1,353 | 25.1 | 2,531 | 46.9 |
| Chicken | 249 | 11.6 | 3 | 0.1 | 196 | 9.2 | 1,691 | 79.1 |
| Pig | 124 | 10.3 | 3 | 0.2 | 14 | 1.2 | 1,067 | 88.3 |

## Table S12: Number of ATAC-seq peaks per species.

| Species | Tissue | \# ATAC-seq peaks |
| :---: | :---: | :---: |
| Cattle | cd4 | 69,661 |
|  | cd8 | 75,295 |
|  | merged | 104,985 |
| Goat | cd4 | 39,526 |
|  | cd8 | 57,084 |
|  | Chicken | liver |
|  | merged | 14,137 |
|  | cd4 | 74,805 |
|  | cd8 | 38,594 |
|  | Pig | liver |
|  | merged | 49,962 |
|  | cd4 | 75,305 |
|  | cd8 | 119,894 |
|  | liver | 80,745 |
|  | merged | 111,457 |
|  |  | 25,885 |
|  |  | 149,333 |

Table S13: Number of differentially accessible (DA) ATAC-seq peaks per species.

| Tissue 1 | Tissue 2 | $\log$ (Tissue2/Tissue1) | Species | Number of DA ATAC-seq peaks |
| :---: | :---: | :---: | :---: | :---: |
| T cell | Liver | $>0$ | Goat | 2,780 |
|  |  | <0 |  | 2,042 |
|  |  | $>0$ | Chicken | 6,663 |
|  |  | <0 |  | 6,991 |
|  |  | $>0$ | Pig | 5,467 |
|  |  | <0 |  | 3,678 |

Table S14: Hi-C read pair mapping statistics. Number of read pairs of different categories. Initial: total number of sequenced read pairs. Reported: pairs with both reads mapped on the genome. Valid: uniquely mapped pairs with an estimated insert size (sum of the distances from the reads to their next downstream HindllI genomic sites) between 20bp and 1 Kb . Valid.rmdup: valid read pairs after duplication removal that were used to build the interaction matrices. Trans: pairs with reads on different chromosomes.

| Species | Animal | initial | reported | valid | valid.rmdup | trans |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Goat | goat1 | $192,807,889$ | $164,130,417$ | $94,760,148$ | $77,324,040$ | $31,728,784$ |
|  | goat2 | $184,098,994$ | $142,479,273$ | $46,280,100$ | $36,346,806$ | $12,858,516$ |
|  | goat3 | $192,081,174$ | $149,494,649$ | $44,712,829$ | $37,115,258$ | $13,227,700$ |
|  | goat4 | $178,929,203$ | $135,758,984$ | $38,435,525$ | $31,232,835$ | $12,691,582$ |
| Chicken | chicken1 | $172,356,821$ | $136,854,885$ | $100,067,668$ | $74,551,182$ | $35,227,545$ |
|  | chicken2 | $182,654,001$ | $152,646,931$ | $122,151,951$ | $86,044,085$ | $30,151,392$ |
|  | chicken3 | $193,520,830$ | $149,648,213$ | $89,183,496$ | $64,824,211$ | $19,603,078$ |
|  | chicken4 | $187,696,586$ | $131,460,531$ | $51,115,059$ | $37,069,637$ | $11,768,987$ |
| Pig | pig1 | $168,050,522$ | $139,712,050$ | $111,642,484$ | $82,782,023$ | $25,956,367$ |
|  | pig2 | $157,480,346$ | $129,447,277$ | $95,004,214$ | $70,483,985$ | $19,076,359$ |
|  | pig3 | $165,285,596$ | $132,403,825$ | $93,115,687$ | $74,144,797$ | $23,560,745$ |
|  | pig4 | $165,529,922$ | $131,590,998$ | $82,504,226$ | $62,212,375$ | $18,705,743$ |


| Table S15: Statistics of Hi-C TADs and A/B compartments. |  |
| :--- | :---: |
| Feature Species Number Min size Mean size Max size Genomic coverage (Mb) Genomic coverage (\%) <br> TADs Pig 10,982 80,000 184,600 $4,520,000$ $2,027.28$  <br>  Chicken 5,362 80,000 148,100 $3,520,000$ 83.25  <br>  Goat 8,990 80,000 219,800 $6,680,000$ 794.28 79.4 <br> A/B comp. Pig 698 343,300 $3,175,000$ $44,000,000$ $2,215.96$ 85.92 <br>  Chicken 578 95,050 $1,596,000$ $10,000,000$ 94 90.51 <br>  Goat 616 426,000 $3,412,000$ $21,000,000$ 222.348 92.101 .87 |  |


[^0]:    Figure S8: GO term enrichment analysis for reference genes differentially expressed between liver
    and T cells in all species. This analysis was performed for each GO (in column: Biological Process,
    Molecular Function, Cellular Compartment in that order) and for genes over-expressed in liver (top) and
    over-expressed in T cells (bottom).

[^1]:    
    

