Animal (2020), **14:10,** pp 2003–2013 © The Author(s), 2020. Published by Cambridge University Press on behalf of The Animal Consortium
doi:10.1017/S1751731120000841

animal

Review: Long non-coding RNA in livestock

B. Kosinska-Selbi[1], M. Mielczarek[1,2] and J. Szyda[1,2†]

[1]*Biostatistic Group, Department of Genetics, Wroclaw University of Environmental and Life Sciences, Kożuchowska 7, Wrocław 51-631, Poland;* [2]*Department of Cattle Breeding, National Research Institute of Animal Production, Krakowska 1, Balice 32-083, Poland*

(Received 25 July 2019; Accepted 27 March 2020; First published online 8 May 2020)

*Less than 2% of mammalian genomes code for proteins, but 'the majority of its bases can be found in primary transcripts' – a phenomenon termed the pervasive transcription, which was first reported in 2007. Even though most of the transcripts do not code for proteins, they play a variety of biological functions, with regulation of gene expression appearing as the most common one. Those transcripts are divided into two groups based on their length: small non-coding RNAs, which are maximally 200 bp long, and long non-coding RNAs (**lncRNA**s), which are longer than 200 nucleotides. The advances in next-generation sequencing methods provided a new possibility of investigating the full set of RNA molecules in the cell. In this review, we summarized the current state of knowledge on lncRNAs in three major livestock species –* Sus scrofa, Bos taurus *and* Gallus gallus, *based on the literature and the content of biological databases. In the NONCODE database, the largest number of identified lncRNA transcripts is available for pigs, but cattle have the largest number of lncRNA genes. Poultry is represented by less than a half of records. Genomic annotation of lncRNAs showed that the majority of them are assigned to introns (pig, poultry) or intergenic (cattle). The comparison with well-annotated human and mouse genomes indicates that such annotation is a result of lack of proper lncRNA annotation data. Since lncRNAs play an important role in genomic studies, their characterization in farm animals' genomes is critical in bridging the gap between genotype and phenotype.*

Keywords: cattle, genomic annotation, non-coding transcripts, pig, poultry

Implications

Long non-coding RNAs are common transcripts existing in genomes, including those of livestock. An individual transcriptome contains more long non-coding RNAs than messenger RNA molecules. Our literature review shows significant impact of long non-coding RNAs on a variety of phenotypes relevant to livestock production and welfare, which is exhibited by the role of long non-coding RNAs as modifiers of the expression of protein-coding genes. In humans, long non-coding RNAs exist which act as disease biomarkers what can also be anticipated for livestock. The better understanding of individual- and tissue-specific variability in long non-coding RNAs expression is important for more precise exploitation of genetic variation of livestock phenotypes.

Introduction

Beginning from the discovery of transfer RNA (**tRNA**) and ribosomal RNA (**rRNA**) in the 1950s, non-coding RNAs (**ncRNA**s) with biological roles have been known for some 60 years (Palazzo and Lee 2015). Recent advances in next-generation sequencing methods have been proving new possibilities of investigating the full set of RNA molecules in genomes and have led to the increase in the number of studies carried out on RNA. Specifically, RNA-seq is a technique used for a whole transcriptome sequencing which, together with computational methods, allows for transcriptome reconstructing and the quantification of gene expression. The method overcomes shortcomings of microarray technology by offering a more comprehensive coverage of whole transcriptomes, and it is not limited to known sequences (Uchida *et al.*, 2017). Therefore, RNA-seq offers a remarkable opportunity to genome-wide annotation and characterization of long non-coding RNAs (**lncRNA**s) (Xiao *et al.*, 2018). According to recent knowledge, less than 2% of the genome codes for proteins, but 'the majority of its bases can be found in primary transcripts' – a phenomenon termed the pervasive transcription, which was first reported by the ENCODE Project Consortium (2007). Recently, Lee *et al.* (2018) stated that only 1~2% of the genome has a protein-coding potential, while the reminder forms ncRNA molecules. Because of typically low expression levels (comparing to protein-coding transcripts), they are described as 'transcription noise' (Ma *et al.*, 2013).

Non-coding RNA is classified into two groups: short and long non-coding RNA (Nie *et al.*, 2012). Transcripts shorter than 200 nucleotides are termed small non-coding RNA

†E-mail: joanna.szyda@upwr.edu.pl

https://doi.org/10.1017/S1751731120000841 Published online by Cambridge University Press

and include Piwi-interacting RNA, small interfering RNA, microRNA (**miRNA**), rRNA, tRNA, small nucleolar RNA and small nuclear RNA (Storz, 2002). Transcripts longer than 200 nucleotides are classified as lncRNAs among which lncRNAs that located in-between genes are termed long intergenic non-coding RNA (**lincRNA**) (Wang and Chang, 2011; Zheng *et al*., 2018).

In livestock, genome-wide association studies based on single-nucleotide polymorphisms (**SNP**s) led to the identification of many mutations causal for the phenotypes of commercial interest, still most of the significant SNPs fall into genomic regions not covered by genic DNA (Goddard *et al*., 2016; www.animalgenome.org/QTLdb/). However, despite this large number of significant SNPs identified, their joint effects do not account for all of the phenotypic variations observed in traits routinely measured in livestock – the phenomenon called 'missing heritability' first introduced into human genetics by Manolio *et al.* (2009). Since lncRNA may be one of the potential causes of missing heritability, it is of interest for livestock genomics. Therefore, the aim of our study was to characterize the current state of knowledge on lncRNA in three major species of farm animals: *Bos taurus, Sus scrofa* and *Gallus gallus.*

Long non-coding RNA detection workflow

Given a raw RNA-seq data, a typical workflow for the identification and annotation of lncRNAs is composed of two major parts – a part which is common to processing all RNA-seq data and a part dedicated to lncRNA.

In the common part, the first step (1a) involves generation of a control report for sequence quality, which is typically done using the FastQC software (Andrews *et al*., 2010). In the second step (2a), raw sequence reads are pre-processed by filtering out contaminations from sequencing adapters and by removing or trimming low-quality reads. The minimum threshold for read quality score is usually set to 20 (Wu *et al*., 2018). The Trimmomatic software (Bolger *et al*., 2014) is a popular tool for pre-processing of raw RNA-seq data. In the next step (3a), cleaned sequences are mapped to the reference genome with the most commonly used software tools being: Tophat (Trapnell *et al*., 2009), Tophat2 (Kim *et al*., 2013), Bowtie2 (Langmead and Salzberg 2012) or HISAT2 (Kim *et al*., 2015). Finally in the last step (4a), assembling of the uniquely mapped sequence reads into transcripts is most often done by Cufflinks (Trapnell *et al*., 2010) or StringTie (Pertea *et al*., 2015).

In order to proceed with the analysis specific to lncRNA, several transcript filtration steps are to be pre-imposed on the annotated RNA-seq data. The first step (1b) is to identify only novel transcripts, which do not correspond to the protein-coding part of the genome and do not represent previously annotated lncRNAs. It is most often performed using the Cuffcompare function of Cufflinks, which classifies all the available assembled transcripts based on their annotation to the pre-specified reference genome in the GFF format. Transcripts classified as 'A transfrag falling entirely within a reference intron', 'Unknown, intergenic transcript' or 'Exonic overlap with reference on the opposite strand' are then selected for downstream analysis. The genomic annotation of lncRNA in livestock is still very scarce as compared to human or mouse genomes; therefore, for livestock species, a much larger number of novel lncRNA transcripts is expected than for humans. Such potential candidate sequences for lncRNA are then subjected to several filtering steps (2b), which can differ depending on the analytic approach. The most important filtering steps comprise (i) filtering by length – removing transcripts shorter than 200 bp and transcripts longer than 10 000 bp containing a single exon, (ii) filtering by sequence content – removing transcripts overlapping with repeat or low complexity regions defined in the reference genome assembly (note, that for livestock genomes, this information is still limited), (iii) filtering by expression level – removing extremely high and extremely low expressed transcripts using the FPKM measure (fragments per kilo base of transcript per million mapped reads) to quantify the expression level; removal thresholds can either be arbitrarily chosen or estimated dynamically from the available data and (iv) filtering by protein-coding potential – various approaches involving removing transcripts containing known protein-coding domains using, for example, Transeq (El-Gebali *et al*., 2019) or HMMER (Eddy *et al*., 2011; Finn *et al*., 2011) software; removing transcripts with a significant hit in the Pfam database using, for example, PfamScan software (Bateman *et al*., 2002; Finn *et al*., 2014); removing transcripts, which products show similarity to known proteins from the RefSeq non-redundant protein database or the UniRef90 database (Suzek *et al*., 2015) using BLASTX (Altschul *et al*., 1990); removing transcripts based on their protein-coding potential level estimated, for example, by Coding Potential Calculator (CPC) (Kong *et al*., 2007), Coding Potential Assessment Tool (CPAT) (Wang *et al*., 2013), Coding Non-Coding Index (CNCI) software (Sun *et al*., 2013) or predictor of Long non-coding RNAs and mEssenger RNAs based on an improved K-mer scheme (**PLEK**) (Li *et al*., 2014). The next step (3b) comprises merging of the RNA-seq data corresponding to novel lncRNAs, defined by the above workflow, and the known lncRNAs with positions defined in databases such as the ALDB (Li *et al*., 2015) or the NONCODE. Note that various authors apply different thresholds for the percent of sequence identity and the percent of length of the aligned sequence to call lncRNAs. The final downstream analysis (4b) of the combined data sets depends on the experimental hypothesis and on the underlying experimental design. Most typically, it involves (i) the comparison of lncRNA expression levels between experimental conditions using, for example, DESeq2 software (Love *et al*., 2014), (ii) identification of target genes of differentially expressed lncRNAs, which can either be done by a dedicated software, for example, LncTar (Li *et al*., 2015), by considering the physical proximity between a lncRNA and a protein-coding gene, or by considering high correlations between the expression level of a lncRNA and an messenger RNA (**mRNA**), and (iii) the functional annotation of target genes to metabolic pathways and/or gene ontologies using, for example, DAVID (Huang *et al*., 2009), KOBAS (Xie *et al*., 2011) programs, or the in-house tools provided by annotation databases,

https://doi.org/10.1017/S1751731120000841 Published online by Cambridge University Press

for example, the Gene Ontology database enrichment analysis tool (geneontology.org; Eilbeck *et al.*, 2005) or the Reactome database analysis tool (reactome.org).

Function of long non-coding RNA

As it can be seen from the number of lncRNA genes or transcripts listed in Tables 1 to 3, there are more lncRNAs than mRNA molecules transcribed from a DNA template. Moreover, lncRNAs can be identified within various cell compartments, such as nucleus, nucleolus, cytoplasm and mitochondria. This reflects the variety of functions which they exhibit on cell metabolism. Technically, lncRNAs functions are related to (i) transcription through either chromatin modifications by interaction with enzymes or through interactions with transcriptional machinery proteins and miRNAs; (ii) post-transcriptional regulations of mRNA molecules, such as capping, alternative splicing, editing, transport, translation, degradation and stability; and (iii) epigenetic

Table 1 *Recent (since 2016) studies on long non-coding RNA (lncRNA), including long intergenic non-coding RNA (lincRNA), in* Sus scrofa *on a genome-wide scale*

Study	DOI	Number of lncRNA genes/lncRNA transcripts detections	Number of analyzed individuals/samples	Analyzed tissues
Chen *et al.* (2019)	10.3389/fgene.2019.00196	1078 (lincRNA)	36	*Longissimus dorsi* muscle and subcutaneous fat
Fang *et al.* (2019)	10.7717/peerj.6577	13 520	One type of cell	IPEC-J2 Cell line cultivated DMEM-F12 medium supplemented with 5% FBS
Kumar *et al.* (2019)	10.1016/j.gene.2019.04.014	6808	16	Back fat
Shi *et al.* (2019)	10.3389/fgene.2019.00160	252 (lincRNA)	20	Adipose tissue
Yang *et al.* (2019)	10.3389/fgene.2019.00409	4456	9	Pineal gland
Wang *et al.* (2019b)	10.1111/age.12849	3827	6	Ovary
Che *et al.* (2018)	10.1371/journal.pone.0193552	15 040	6	Spleen
Jin *et al.* (2018)	10.3390/genes9090443	19 310	6	Lung
Kern *et al.* (2018)	10.1186/s12864-018-5037-7	14 429	2	Adipose, cerebellum, cortex, hypothalamus, liver, lung, muscle and spleen
Li *et al.* (2018)	10.2217/epi-2017-0117	713	3	Liver
Liang *et al.* (2018)	10.1111/age.12720	53 468	Database	Various tissues
Liu *et al.* (2018)	10.3390/ijms19061722	2076	8	Ovary
Miao *et al.* (2018)	10.1016/j.bbrc.2018.06.028	4910	6	Adipose
Shen *et al.* (2018)	10.1159/000494794	3368	16	Liver
Yan *et al.* (2018)	10.7717/peerj.5997	2144	15	Spleen
Zhao *et al.* (2018)	10.2217/epi-2017-0149	18 676	9	33 tissues
Zou *et al.* (2018)	10.3389/fgene.2018.00102	1032	12	Muscle
Gao *et al.* (2017)	10.2527/jas2016.1297	5153 (novel)	21	Muscle
Sun *et al.* (2017)	10.2527/jas.2016.0867	5566 (differentially expressed)	10	Muscle
Weng *et al.* (2017)	10.1186/s12864-017-3907-z	343	6	Adipose
Weng *et al.* (2017)	10.1016/j.ygeno.2017.07.001	15 528		Testis
Yang *et al.* (2017)	10.1155/2017/6152582	2139 (novel)	>16	Heart, kidney, liver, lung, muscle, small intestine, spleen and stomach
Yu *et al.* (2017)	10.18632/oncotarget.18269	4868	6	Fat, liver and muscle
Xing *et al.* (2017)	10.1111/asj.12777	8946	6	Muscle
Zou *et al.* (2017a)	10.1038/s41598-017-07998-9	323 (lincRNA)	6	Muscle
Zou *et al.* (2017b)	10.3390/genes8080203	759 (lincRNA)	6	Muscle
Li *et al.* (2016a)	10.1038/srep38333	7618 (novel lincRNA)	5	Abdominal fat, embryo, endometrium, heart, kidney, liver, lung and muscle
Ran *et al.* (2016)	10.1095/biolreprod.115.136911	752	6	Testis
Shen *et al.* (2016)	10.3390/genes7070034	1018	6	Thyroid gland
Wang *et al.* (2016)	10.1038/srep20238	2805	12	Endometrium

https://doi.org/10.1017/S1751731120000841 Published online by Cambridge University Press

Table 2 *Recent (since 2016) studies on long non-coding RNA (lncRNA), including long intergenic non-coding RNA (lincRNA), in* Bos taurus *on a genome-wide scale*

Study	DOI	Number of lncRNA genes/lncRNA transcripts detections	Number of analyzed individuals/samples	Analyzed tissues
Choi *et al.* (2019)	10.1080/19768354.2018.1512522	304	9	Skeleton muscle and adipose tissues
Gao *et al.* (2019)	10.3389/fgene.2019.00646	23 735	6	Testis
Wang *et al.* (2019a)	10.1038/s41598-018-38462-x	11 561	6	Sperm
Zeng *et al.* (2019)	10.3168/jds.2019-16257	3481	24 Samples from different lactation periods	Milk exosomes
Cai *et al.* (2018)	10.3389/fgene.2018.00281	6450	6	Mammary gland
Ibeagha-Awemu *et al.* (2018a)	10.3390/genes9030142	1568 4243	32	Ileum, rumen
Ibeagha-Awemu *et al.* (2018b)	10.3390/ijms19113610	4995	12	Mammary gland
Kern *et al.* (2018)	10.1186/s12864-018-5037-7	7235	2	Adipose, cerebellum, cortex, hypothalamus, liver, lung, muscle and spleen
Li *et al.* (2018)	10.1016/j.omtn.2018.07.003	13 580	3 fetuses 3 adults	Heart, spleen, kidney, liver, lung, stomach, small intestine and muscle
Weikard *et al.* (2018)	10.18632/oncotarget.24898	1042	12	Jejunum mucosa
Yang *et al.* (2018)	10.1186/s12864-018-4974-5	3746 (differentially expressed)	8	Mammary gland
Zheng *et al.* (2018)	10.3168/jds.2018-14900	1181	4	Mammary gland
Liu *et al.* (2017)	10.1111/age.12539	7188	3	Muscle
Ma *et al.* (2017)	10.1016/j.rvsc.2017.09.020	1236	2	Cell lines
Tong *et al.* (2017)	10.1186/s12864-017-3858-4	184 (lincRNA)	5	Mammary gland

modifications manifested by the regulation of imprinting (Bhat and Jones 2016; Fernandes *et al.*, 2019). On the organism level, lncRNAs are known to be abnormally expressed in many diseases with the most predominant influence on cancer and viral infections thereby playing a role of biomarkers. A practical example of lncRNA being a biomarker in human cancers is HOTAIR, whose overexpression results in development and metastases of several cancer types in humans (Lorenzi *et al.*, 2019). In the healthy physiological state, lncRNAs play a role in organ differentiation during embryogenesis (Grote and Herrmann, 2015) as well as in the process of aging (Xing *et al.*, 2017). Most of the applications related to livestock investigate the functional annotation of lncRNAs, manifested by Gene Ontologies (**GO**) and KEGG pathways assigned to their target genes. These functions however are strongly related to the experimental design applied in each particular study and thus not of a universal nature. For example, in a recent study on pigs, Chen *et al.* (2019) applied the DAVID software to constructed clusters composed of GO terms and KEGG pathways characteristic for genes, which were targets of lncRNAs differentially expressed in relation to growth performance. In addition, You *et al.* (2019) applied the DAVID software for functional clustering of GO term and KEGG pathways related to target genes of lncRNAs differentially expressed in white leghorn chicken infected by the Marek's disease virus and in a control, healthy group. In cattle, an example of functional annotation is the study of Gao *et al.* (2019). Using KOBAS, the authors tested a functional enrichment in GO terms and KEGG pathways of genes targeted by lncRNAs differentially expressed in two developmental stages of testis.

Long non-coding RNA in livestock

Sus scrofa. Among livestock, the largest number of identified lncRNA transcripts is available for pigs' genomes, amounting to 29 585 transcripts in the NONCODE 5.0 database (Zhao *et al.*, 2015). Moreover, Liang *et al.* (2018) developed the Pig LncRNANet database (lnc.rnanet.org), which stores authors' own lncRNA discoveries, as well as results from other published studies. Although not fully mature, the database provides valuable bioinformatic functions, such as sequence BLAST, lncRNA sequence visualization including overlaps with QTL and SNV positions, as well as the visualization of transcript expression levels in various tissues. Currently, the Pig LncRNANet database contains 53 468 lncRNA records and is thereby the most comprehensive pig lncRNA catalog.

Most of the studies carried out on pigs are logically related to differential expression of all lncRNAs or only lincRNAs for

https://doi.org/10.1017/S1751731120000841 Published online by Cambridge University Press

Table 3 *Recent (since 2016) studies on long non-coding RNA (lncRNA), including long intergenic non-coding RNA (lincRNA), in* Gallus gallus *on a genome-wide scale*

Study	DOI	Number of lncRNA genes/lncRNA transcripts detections	Number of analyzed individuals/ samples	Analyzed tissues
Yin *et al.* (2020)	10.1016/j.ygeno.2019.02.003	6832	40	Ovary, magnum, isthmus and uterus
Li *et al.* (2019)	10.1371/journal.one.0215006	4404	6	Breast muscle tissue
Xu *et al.* (2019)	10.3382/ps/pez434	4127	3	Liver
You *et al.* (2019)	10.1186/s12864-019-5625-1	2819	17	Spleen
Adetula *et al.* (2018)	10.1038/s41598-018-31301-z	9977	14	Uterovaginal junction
Hong *et al.* (2018)	10.3390/ijms19082359	6900	1	Breast, bone marrow, cerebellum, cerebrum, comb, eye, fascia, gall bladder, gizzard, heart, immature egg, kidney, liver, lung, mature egg, pancreas, skin, shank, spleen and uterus
Kern *et al.* (2018)	10.1186/s12864-018-5037-7	9393	2	Adipose, cerebellum, cortex, hypothalamus, liver, lung, muscle and spleen
Li *et al.* (2018)	10.3390/genes9010034	2484	3	Muscle
Liu *et al.* (2018)	10.1038/s41598-018-25103-6	959 (Differentially expressed)	12	Ovary
Peng *et al.* (2018)	10.1016/j.ygeno.2018.09.012	8691	10	Ovary
Ren *et al.* (2018b)	10.1139/gen-2017-0114	9343	12	Skeletal muscle
Ren *et al.* (2018a)	10.1186/s12864-018-4754-2	1376	6	Trachea
Wu *et al.* (2018)	10.1186/s12864-018-4891-7	2220	25	Liver
Xu *et al.* (2018)	10.1186/s12864-018-5301-x	39 907	21	Cerebrum
Cao *et al.* (2017)	10.18632/oncotarget.14971	15 412	60	Vein and skin
Liu *et al.* (2017)	10.1038/s41598-017-08738-9	2597	6	Testis
Kuo *et al.* (2017)	10.1186/s12864-017-3691-9	> 20 000	9	Brain and embryo
Muret *et al.* (2017)	10.1186/s12711-016-0275-0	2193	16	Liver and adipose
Qiu *et al.* (2017)	10.1016/j.dci.2017.06.015	4804	2 Cell lines	CEF: chicken embryo fibroblastst, HD11: macrophage cell line
Wang *et al.* (2017b)	10.1186/s12862-017-1036-6	4754	821 Transcriptomes	Multiple
Zhang *et al.* (2017a)	10.1371/journal.pone.0172389	25 435	12	Intramuscular preadipocytes
Zhang *et al.* (2017b)	10.1534/g3.116.037069	27 023	12	Preadipocytes from abdominal adipose tissue

traits included into the selection goal of most breeds. These comprise growth performance expressed by analyzing transcription in muscle tissue either in comparison to other tissues (Chen *et al.*, 2019), in comparison between animal groups (Zou *et al.*, 2017a), between breeds with different growth performance characteristics (Gao *et al.*, 2017; Sun *et al.*, 2017; Yu *et al.*, 2017) or between different developmental stages (Zou *et al.*, 2018). In the context of growth performance, the expression in other tissues has also been considered – intramuscular adipose tissue (Miao *et al.*, 2018) or back fat tissue (Chen *et al.*, 2019; Kumar *et al.*, 2019). Besides growth performance, meat quality is also an important meat production trait, for which the expression of lincRNA has recently been considered, for example, by Zou *et al.* (2017b). Some studies analyzed the impact of testosterone deficiency on the expression of lncRNAs, either by comparing intact and castrated males (Xing *et al.*, 2017; Wang *et al.*, 2017a) or different male developmental stages (Ran *et al.*, 2016; Sun *et al.*, 2017). Apart from production, lncRNA expression was also assessed for traits related to female reproduction by the comparison of expression in different developmental stages of ovaries (Liu *et al.*, 2018; Wang *et al.*, 2019b) or in endometrium of pregnant and non-pregnant females (Wang *et al.*, 2016). Other experimental designs were devoted to the analysis of expression in specific organs, such as spleen (Chen *et al.*, 2019; Yan *et al.*, 2018) in the context of resistance against pathogenic infections, pineal (Yang *et al.*, 2019), liver (Li *et al.*, 2018) or lung (Jin *et al.*, 2018). Profiles of lncRNA expression in disease were considered for porcine circovirus-associated disease by Fang *et al.* (2019) and for intrauterine growth restriction by Shen *et al.* (2018). Moreover, instead of inter-group expression comparisons, some studies focused on identification and genomic annotation of lncRNAs in various tissues (Li *et al.*, 2016a; Liu *et al.*, 2017; Yang *et al.*, 2017; Zhao *et al.*, 2018).

The most recent studies on lncRNA detection in *Sus scrofa* conducted on a genome-wide scale are summarised in Table 1.

Bos taurus. Although less lncRNA transcripts have been identified for cattle than for pigs, the former is the livestock species with the largest number of lncRNA genes

https://doi.org/10.1017/S1751731120000841 Published online by Cambridge University Press

(22 227 in NONCODE 5.0). The first genome-wide catalog of bovine intergenic lncRNAs was provided by the study of Huang *et al.* (2012) who identified 449 lncRNAs located in 405 intergenic regions, using public bovine-specific expressed sequence tag sequences. After that, the majority of studies of lncRNAs were related to their expression in the mammary gland (e.g., Cai *et al.*, 2018; Yang *et al.*, 2018; Ibeagha-Awemu *et al.*, 2018b) or milk exosomes (Zeng *et al.*, 2019) in relation to dairy production. In addition, the role of lncRNAs in beef production was addressed by assessing the expression in *longissimus thoraci* by Billerey *et al.* (2014), as well as in muscles and adipose tissues by Choi *et al.* (2019). Long non-coding transcriptome of male reproduction traits was analyzed by Wang *et al.* (2019a) in sperm samples with divergent motility as well as by Gao *et al.* (2019) in different testis development stages. Among other phenotypes, Weikard *et al.* (2013) focused on lncRNA expression in pigmented and non-pigmented bovine skin samples, Weikard *et al.* (2018) – on lncRNA expression dependent on energy metabolism associated with different diets and Ibeagha-Awemu *et al.* (2018a) – on lncRNA expression in ileum and rumen during different developmental stages. A comprehensive genome-wide annotation of lncRNA expressed in 18 tissues was presented by Koufariotis *et al.* (2015).

The most recent studies on lncRNA detection in *Bos taurus* conducted on a genome-wide scope are summarized in Table 2.

Gallus gallus. With 12 850 lncRNA transcripts, corresponding to 9527 genes, poultry is represented by less than half of the number of records in the NONCODE database than the above-mentioned mammalian livestock species. However, Kou *et al.* (2017) comparing the complexity of human and chicken transcriptomes suggested that chicken transcriptome is similar in complexity to the human transcriptome.

During the last few years, the chicken genome has been intensively investigated in the context of lncRNA. Since meat performance is one of the economically most important poultry phenotypes, a large number of studies relate to the expression of lncRNA in tissues related to growth: muscle (Li *et al.*, 2016b; Cai *et al.*, 2017; Ren *et al.*, 2017; Li *et al.*, 2018; Ren *et al.*, 2018a), adipose tissue (Muret *et al.*, 2017; Zhang *et al.*, 2017a and 2017b) as well as to meat quality (e.g., Li *et al.*, 2019). The influence of lncRNA on egg laying performance was analyzed by Peng *et al.* (2018). Another economically important one in poultry group of traits is immune response. In the context of lncRNA expression, it was analyzed by Qiu *et al.* (2017) and Hu *et al.* (2018) for the resistance to the Avian leukosis virus J, by Ren *et al.* (2018b) for the resistance to *Cryptosporidium baileyi.* Long non-coding RNA expression changes in the presence of selenium deficiency were addressed by Fan *et al.* (2017) and Cao *et al.* (2017). Fertility was studied in the context of female reproduction by Liu *et al.* (2018) and Yin *et al.* (2020) for ovary, by Adetula *et al.* (2018) for uterovaginal tissue as well as by Yin *et al.* (2020) for oviduct. Male reproduction traits were assessed by lncRNA expression study in sperm with differential motility (Liu *et al.*, 2017). In addition, lncRNA expression study exists, which focused on the development of specific organs, such as liver (Muret *et al.*, 2017; Wu *et al.*, 2018; Xu *et al.*, 2019), brain (Xu *et al.* 2018) and ovary (Liu *et al.*, 2018). Other recently analyzed phenotypes comprise lncRNA expression landscape related to chicken domestication addressed by Wang *et al.* (2017b) as well as the expression related to feathers and skin black color studied by Hong *et al.* (2018).

The most recent studies on lncRNA detection in *Gallus gallus* conducted on a genome-wide scope are summarized in Table 3.

Genomic annotation of long non-coding RNAs

One of the most complete and therefore most widely used databases that store lncRNAs is the NONCODE. It provides the collection and annotation of ncRNAs, especially lncRNAs, in 17 species, including livestock. The current version of the NONCODE (5.0) contains 548 640 transcripts, identified either by the RNA-seq, expression microarrays, or based on the literature. Yet, the database is far from complete while most of the studies listed in Tables 1 to 3 report a very large number of novel lncRNAs. For instance, Kern *et al.* (2018) reported that only 18.3% of pig, 1.7% of cattle and 5.7% of poultry lncRNA transcripts from the NONCODE database overlapped with transcripts detected in their analysis.

We annotated lncRNAs of five species from the NONCODE database (accessed on 20 November 2019) using the Variant Effect Predictor software (McLaren *et al.*, 2010). 51 453 lncRNA genes were annotated to the Sscrofa11.1 (Warr *et al.*, 2019; accessed from www.ensembl.org/Sus_scrofa on 20 November 2019 with the corresponding Ensembl ID GCA_000003025.6) reference genome, 25 683 lncRNA genes were annotated to the ARS-UCD1.2 reference genome of *Bos taurus* (Shamimuzzaman *et al.*, 2020; accessed from www.ensembl.org/Bos_taurus on 20 November 2019 with the corresponding Ensembl ID GCA_002263795.2) and 22 843 lncRNA genes were annotated to the GRCg6a reference genome of *Gallus gallus* (accessed from www.ensembl.org/Gallus_gallus on 20 November 2019 with the corresponding Ensembl ID GCA_000002315.5). These livestock annotations were compared to the well-annotated mouse GRCm38 (318 287 lncRNA genes) and human GRCh38 (616 532 lncRNA genes) genomes. Since genomic coordinates of lncRNAs were defined in relation to different versions of reference genomes, we converted them to the most current assembly of each species using the liftOver software (Kent *et al.*, 2002). Species-specific annotations are visualized in Figure 1. The distribution pattern of lncRNA for *Homo sapiens* and *Mus musculus* is very similar. The observed difference between those two species and the livestock species demonstrates the incompleteness of the annotation of lncRNA in livestock. Precisely, a more complete annotation of the lncRNA in humans and mice is manifested by the fact that the categories representing the actual proper annotation for lncRNA (non-coding transcript exon variant, non-coding transcript variant and regulatory region variant)

https://doi.org/10.1017/S1751731120000841 Published online by Cambridge University Press

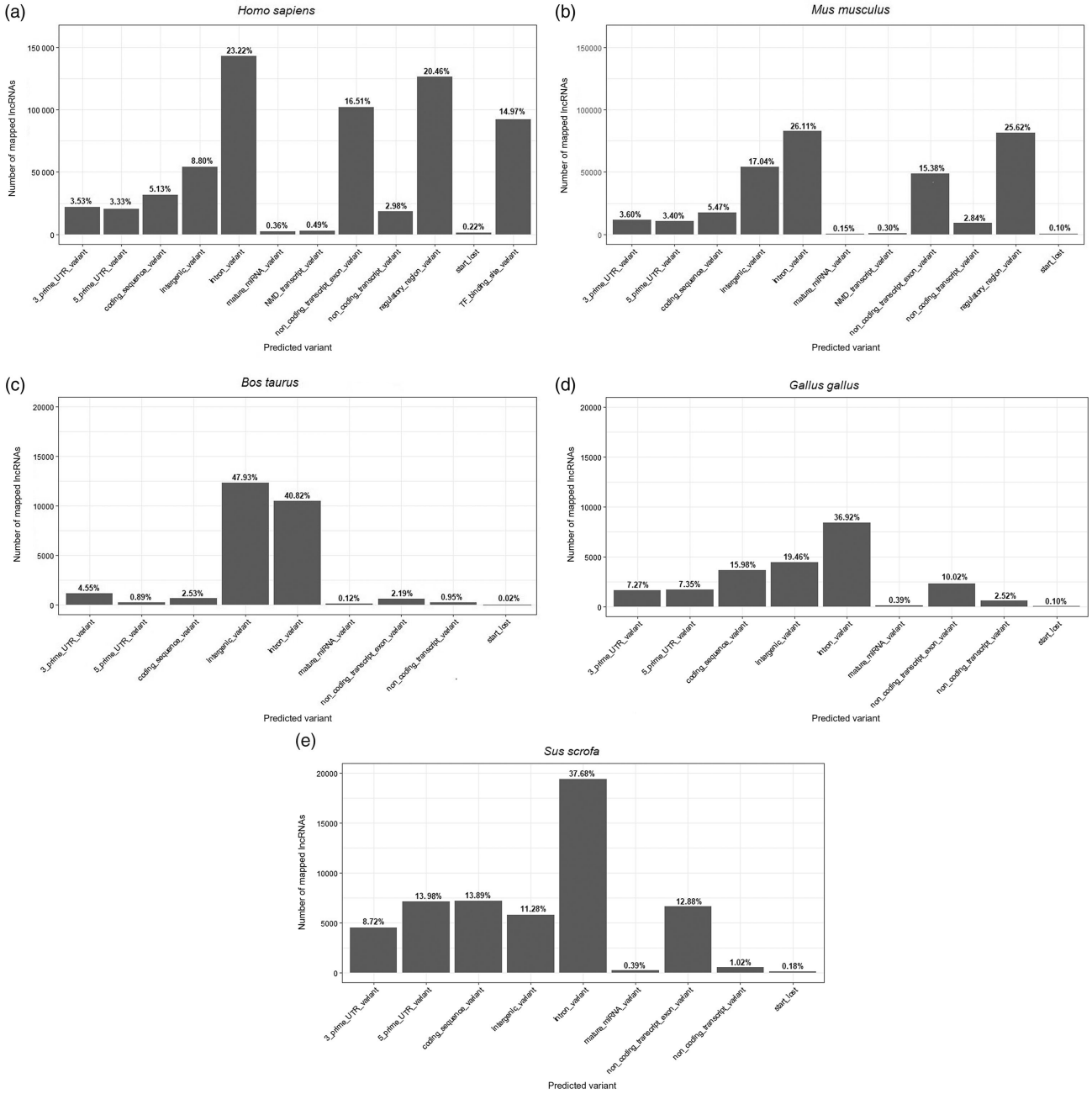

Figure 1 Genomic annotation of long non-coding RNA genes from the NONCODE database for (a) human (*Homo sapiens*), (b) mouse (*Mus musculus*), (c) cattle (*Bos taurus*), (d) poultry (*Gallus gallus*) and (e) pig (*Sus scrofa*). UTR = Untranslated Transcribed Region.

make up 39.35% of all annotations in *Homo sapiens* and 43.84% in *Mus musculus*, but only 13.90% in *Sus scrofa*, 12.54% in *Gallus gallus* and as little as 3.14% in *Bos taurus*. In livestock, the missing ncRNA annotation seems to be predominantly assigned to intergenic and intron sequences.

Conclusions

The above-mentioned studies demonstrate that lncRNAs play important roles not only in the regulation of gene expression (as it was originally emphasized, see e.g., the earlier review of Mercer *et al.*, 2009), but also in numerous other aspects of normal physiology and diseases. Compared to the number of transcripts for model organisms (*Homo sapiens* and *Mus musculus*), livestock has a relatively small number of deposited transcripts in biological databases; therefore, the main course of future research is to further improve the annotation of the non-coding part of livestock genomes. The major challenge associated with lncRNA analysis is a poor accuracy of transcript detection, which involves many data filtration stages. Since the identification workflows of lncRNAs detection change dynamically, there is a great need for defining standardized pipelines. The problem is demonstrated by a typically very large number of 'novel' transcripts reported by each study. As already mentioned above, Kern *et al.* (2018) reported a very low overlap between the NONCODE database and

https://doi.org/10.1017/S1751731120000841 Published online by Cambridge University Press

transcripts detected in their analysis. Koufariotis *et al.* (2015) validated 87.47% of lncRNAs expressed in liver based on samples from the same individual, but validation across animals was much lower – 55.27% of validated lncRNAs in blood. Of course, the problem with low detection accuracy is typical for all high-throughput technologies, but in the case of lncRNA detection, it is enhanced by the fact that the expression levels of lncRNAs are typically low. We need to bear in mind that such low repeatability of lncRNA detection is not necessarily always associated with false-positive or false-negative detections. The expression of lncRNA genes is not only low but also extremely tissue-specific (see e.g., graphical summary offered by the Pig LncRNANet database), related to physiological state of the individual. Moreover, since 2018, studies demonstrating a regulation of lncRNA expression by gut microbiota have emerged (Dempsey *et al.*, 2018; Li and Cui, 2018), adding a new insight into the complexity of the transcriptome landscape.

Today's trend is to receive economic efficiency, by fully exploiting the genetic information. That is why it is important to gain knowledge on all mechanisms controlling gene expression, such as lncRNA molecules.

Acknowledgements

None.

B. Kosinska-Selbi 0000-0003-4709-5556
M. Mielczarek 0000-0002-1086-9119
J. Szyda 0000-0001-9688-0193

Conflict of interest

None.

Ethics statement

Not applicable.

Software and data repository resources

None of the data was deposited in an official repository.

References

ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermüller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaöz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Löytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, NISC Comparative Sequencing Program, Baylor College of Medicine Human Genome Sequencing Center, Washington University Genome Sequencing Center, Broad Institute, Children's Hospital Oakland Research Institute, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN, Yu Y, Ruan Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrímsdóttir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B and Jong PJ 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799–816.

Adetula AA, Gu L, Nwafor CC, Du X, Zhao S and Li S 2018. Transcriptome sequencing reveals key potential long non-coding RNAs related to duration of fertility trait in the uterovaginal junction of egg-laying hens. Scientific Reports 4, 13185.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410.

Andrews S 2010. FastQC: a quality control tool for high throughput sequence data. Retrieved on 15 January 2020 from http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffin-Jones S, Howe KL, Marshall M and Sonnhammer ELL 2002. The Pfam protein families database. Nucleic Acids Research 30, 276–280.

Bhat S and Jones WD 2016. An accelerated miRNA-based screen implicates Atf-3 *Drosophila odorant* receptor expression. Scientific Reports 6, 20109.

Billerey C, Boussaha M, Esquerré D, Rebours E, Djari A, Meersseman C, Kloop C, Gautheret D and Rocha D 2014. Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. BMC Genomics 15, 499.

Bolger AM, Lohse M and Usadel B 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30, 2114–2120.

Cai B, Li Z, Ma M, Wang Z, Han P, Abdalla BA, Nie Q and Zhang X 2017. LncRNA-Six1 encodes a micropeptide to activate six1 in cis and is involved in cell proliferation and muscle growth. Frontiers in Physiology 8, 230.

Cai W, Li C, Liu S, Zhou C, Yin H, Song J, Zhang Q and Zhang S 2018. Genome wide identification of novel long non-coding RNAs and their potential associations with milk proteins in Chinese Holstein cows. Frontiers in Genetics 9, 281.

Cao C, Fan R, Zhao J, Zhao X, Yang J, Zhang Z and Xu S 2017. Impact of exudative diathesis induced by selenium deficiency on LncRNAs and their roles in the oxidative reduction process in broiler chick veins. Oncotarget 8, 20695–20705.

Che T, Li D, Jin L, Fu Y, Liu Y, Liu P, Wang Y, Tang Q, Ma J, Wang X, Jiang A, Li X and Li M 2018. Long non-coding RNAs and mRNAs profiling during spleen development in pig. PLoS ONE 13, e0193552.

Chen L, Shi G, Chen G, Li J, Li M, Zou C, Fang Ch and Li Ch 2019. Transcriptome analysis suggests the roles of long intergenic non-coding RNAs in the growth performance of weaned piglets. Frontiers in Genetics 18, 10–196.

Choi JY, Shin D, Lee HJ and Oh JD 2019. Comparison of long noncoding RNA between muscles and adipose tissues in *Hanwoo* beef cattle. Animal Cells and Systems 23, 50–58.

Dempsey J, Zhang A and Cui JY 2018. Coordinate regulation of long non-coding RNAs and protein-coding genes in germ-free mice. BMC Genomics 19, 834.

Eddy SR 2011. Accelerated profile HMM searches. PLOS Computational Biology 7, e1002195.

https://doi.org/10.1017/S1751731120000841 Published online by Cambridge University Press

Eilbeck K, Lewis SE, Mungall CJ, Yandell M, Stein L, Durbin R and Ashburner M 2005. The sequence ontology: a tool for the unification of genome annotations. Genome Biology 6, R44.

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE and Rober Finn RD 2019. The Pfam protein families database in 2019. Nucleic Acids Research 47, D427–D432.

Fan R, Cao C, Zhao X, Shi Q, Zhao J and Xu S 2017. Downregulated long non-coding RNA ALDBGALG0000005049 induces inflammation in chicken muscle suffered from selenium deficiency by regulating stearoyl-CoA desaturase. Oncotarget 8, 52761–52774.

Fang M, Yang Y, Wang N, Wang A, He Y, Wang J, Jiang Y and Deng Z 2019. Genome-wide analysis of long non-coding RNA expression profile in porcine circovirus 2-infected intestinal porcine epithelial cell line by RNA sequencing. PeerJ 7, e6577.

Fernandes LGV, Guaman LP, Vasconcellos SA, Heinemann MB, Picardeau M, and Nascimento ALTO 2019. Gene silencing on RNA-guided catalycally inactive Cas9 (dCas9): a new tool for genetic engineering in *Leptospira*. Scientific Reports 9, 1839.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J and Sonnhammer ELL 2014. Pfam: the protein families database. Nucleic Acids Research 42, D222–D230.

Finn RD, Clements J and Eddy SR 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Research 39, W29–W37.

Gao PF, Guo XH, Du M, Cao GQ, Yang QC, Pu ZD, Wang ZY, Zhang Q, Li M, Jin YS, Wang XJ, Liu H and Li BG 2017. LncRNA profiling of skeletal muscles in Large White pigs and Mashen pigs during development. Journal of Animal Science 95, 4239–4250.

Gao Y, Li S, Lai Z, Zhou Z, Wu F, Huang Y, Lan X, Lei C, Chen H and Dang R 2019. Analysis of long non-coding RNA and mRNA expression profiling in immature and Mature Bovine (*Bos taurus*) Testes. Frontiers in Genetics 10, 646.

Goddard ME, Kemper KE, MacLeod IM, Chamberlain AJ and Hayes BJ 2016. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. Biological Sciences 283, 20160569.

Grote P and Herrmann BG 2015. Long noncoding RNAs in organogenesis: making the difference. Trends in Genetics 31, 329–35.

Hong H, Chai HH, Nam K, Lim D, Lee KT, Do YJ, Cho CY and Nam JW 2018. Non-coding transcriptome maps across twenty tissues of the Korean Black Chicken, Yeonsan Ogye. International Journal of Molecular Sciences 19, 2359.

Hu X, Chen S, Jia C, Xue S, Dou C, Dai Z, Xu H, Sun Z, Geng T and Cui H 2018. Gene expression profile and long non-coding RNA analysis, using RNA-Seq, in chicken embryonic fibroblast cells infected by avian leukosis virus. Journal of Archives of Virology 163, 639–647.

Huang DW, Sherman BT, Zheng X, Yang J, Imamichi T, Stephens R and Lempicki R 2009. Extracting biological meaning from large gene lists with DAVID. Current Protocols in Bioinformatics 27, 13.11.1–13.11.13.

Huang W, Long N and Khatib H 2012. Genome-wide identification and initial characterization of bovine long non-coding RNAs from EST data. Animal Genetics 43, 674–682.

Ibeagha-Awemu EM, Do DN, Dudemaine PL, Fomenky BE and Bissonnette N 2018a. Integration of lncRNA and mRNA transcriptome analyses reveals genes and pathways potentially involved in calf intestinal growth and development during the early weeks of life. Genes 9, 142.

Ibeagha-Awemu EM, Li R, Dudemaine PL, Do DN and Bissonnette N 2018b. Transcriptome analysis of long non-coding RNA in the bovine mammary gland following dietary supplementation with linseed oil and safflower oil. International. Journal of Molecular Sciences 19, 3610.

Jin L, Hu S, Tu T, Huang Z, Tang Q, Ma J, Wang X, Li X, Zhou X, Shuai S and Li M 2018. Global long noncoding RNA and mRNA expression changes between prenatal and neonatal lung tissue in pigs. Genes 5, 9.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM and Haussler D 2002. The human genome browser at UCSC. Genome Research 12, 996–1006.

Kern C, Wang Y, Chitwood J, Korf I, Delany M, Cheng H, Medrano JF, Van Eenennaam AL, Ernst C, Ross P and Zhou H 2018. Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species. BMC Genomics 19, 684.

Kim D, Langmead B and Salzberg SL 2015. HISAT: a fast-spliced aligner with low memory requirements. Nature Methods 12, 357–360.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R and Salzberg SL 2013. Tophat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14, R36.

Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L and Gao G 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Research 35, W345–W349.

Koufariotis LT, Chen YP, Chamberlain A, Vander Jagt C and Hayes BJ 2015. A catalogue of novel bovine long noncoding RNA across 18 tissues. PLoS ONE 10, e0141225.

Kumar H, Srikanth K, Park W, Lee SH, Choi BH, Kim H, Kim YM, Cho ES, Kim JH, Lee JH, Jung JY, Go GW, Lee KT, Kim JM, Lee J, Lim D and Park JE 2019. Transcriptome analysis to identify long non coding RNA (lncRNA) and characterize their functional role in back fat tissue of pig. Gene 703, 71–82.

Kuo RI, Tseng E, Eory L, Paton IR, Archibald AL and Burt DW 2017. Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human. BMC Genomics 18, 323.

Langmead B and Salzberg SL 2012. Fast gapped-read alignment with bowtie 2. Nature Methods 9, 357–359.

Lee Y, Park C, Lee S, Lee D and Kim J 2018. Isolation and functional examination of the long non-coding RNA *Redrum*. Molecules and Cells 41, 134–139.

Li A, Zhang J, Zhou Z, Wang L, Liu Y and Liu Y 2018. ALDB: a domestic-animal long noncoding RNA database. PLoS One 10, e0124003.

Li AM, Zhang JY and Zhou ZY 2014. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. BMC Bioinformatics 15, 311.

Li CY and Cui JY 2018. Regulation of protein-coding gene and long noncoding RNA pairs in liver of conventional and germ-free mice following oral PBDE exposure. PLoS ONE 13, e0201387.

Li D, Li F, Jiang K, Zhang M, Han R, Jiang R, Li Z, Tian Y, Yan F, Kang X and Sun G 2019. Integrative analysis of long noncoding RNA and mRNA reveals candidate lncRNAs responsible for meat quality at different physiological stages in Gushi chicken. PLoS ONE 14, e0215006.

Li J, Gao Z, Wang X, Liu H, Zhang Y and Liu Z 2016a. Identification and functional analysis of long intergenic noncoding RNA genes in porcine pre-implantation embryonic development. Scientific Reports 6, 38333.

Li J, Ma W, Zeng P, Wang J, Geng B, Yang J and Cui Q 2015. LncTar: a tool for predicting the RNA targets of long noncoding RNAs. Briefings in Bioinformatics 16, 806–812.

Li Z, Ouyang H, Zheng M, Cai B, Han P, Abdalla BA, Nie Q and Zhang X 2016b. Integrated analysis of long non-coding RNAs (lncRNAs) and mRNA expression profiles reveals the potential role of lncRNAs in skeletal muscle development of the chicken. Frontiers in Physiology 7, 687.

Liang G, Yang Y, Li H, Yu H, Li X, Tang Z and Li K 2018. LncRNAnet: a comprehensive *Sus scrofa* lncRNA database. Animal Genetics 49, 632–635.

Liu L, Xiao Q, Gilbert ER, Cui Z, Zhao X, Wang Y, Yin H, Li D, Zhang H and Zhu Q 2018. Whole-transcriptome analysis of atrophic ovaries in broody chickens reveals regulatory pathways associated with proliferation and apoptosis. Scientific Reports 8, 7231.

Liu Y, Sun Y, Li Y, Bai H, Xue F, Xu S, Xu H, Shi L, Yang N and Chen J 2017. Analyses of long non-coding RNA and mRNA profiling using RNA sequencing in chicken testis with extreme sperm motility. Scientific Reports 7, 9055.

Lorenzi L, Avila Cobos F, Decock A, Everaert C, Helsmoortel H, Lefever S, Verboom K, Volders PJ, Speleman F, Vandesompele J and Mestdagh P 2019. Long noncoding RNA expression profiling in cancer: challenges and opportunities. Genes Chromosomes Cancer 58, 191–199.

Love MI, Huber W and Anders S 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550.

Ma L, Bajic VB and Zhang Z 2013. On the classification of long non-coding RNAs. RNA Biology 10, 924–933.

Ma Q, Li L, Tang Y, Fu Q, Liu S, Hu S and Qiao J 2017. Analyses of long non-coding RNAs and mRNA profiling through RNA sequencing of MDBK cells at different stages of bovine viral diarrhea virus infection. Research in Veterinary 115, 508–516.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA and Visscher PM 2009. Finding the missing heritability of complex diseases. Nature 461, 747–753.

https://doi.org/10.1017/S1751731120000841 Published online by Cambridge University Press

McLaren W, Pritchard B, Rios D, Chen Y, Flicek P and Cunningham F 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 26, 2069–2070.

Mercer TR, Dinger ME and Mattick JS 2009. Long non-coding RNAs: insights into functions. Nature Reviews. Genetics 10, 155–159.

Miao Z, Wang S, Zhang J, Wei P, Guo L, Liu D, Wang Y and Shi M 2018. Identification and comparison of long non-conding RNA in Jinhua and Landrace pigs. Biochemical and Biophysical Research Communications 506, 765–771.

Muret K, Klopp C, Wucher V, Esquerré D, Legeai F, Lecerf F, Desert C, Boutin MS, Jehl F, Acloque H, Giuffra E, Djebali S, Foissac S, Derrien T and Lagarrigue S 2017. Long noncoding RNA repertoire in chicken liver and adipose tissue. Genetics Selection Evolution 49, 6.

Nie L, Wu HJ, Hsu JM, Chang SS, Labaff AM, Li CW, Wang Y, Hsu JL and Hung MC 2012. Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer. American Journal of Translational Research 4, 127–150.

Palazzo AF and Lee ES 2015. Non-coding RNA: what is functional and what is junk? Frontiers in Genetics 6, 2.

Peng Y, Chang L, Wang Y, Wang R, Hu L, Zhao Z, Geng L, Liu Z, Gong Y, Li J, Li X and Zhang C 2018. Genome-wide differential expression of long noncoding RNAs and mRNAs in ovarian follicles of two different chicken breeds. Genomics 7543, 30268–30274.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT and Salzberg SL 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology 33, 290–295.

Qiu L, Li Z, Chang G, Bi Y, Liu X, Xu L, Zhang Y, Zhao W, Xu Q and Chen G 2017. Discovery of novel long non-coding RNAs induced by subgroup J avian leukosis virus infection in chicken. Developmental and Comparative Immunology 76, 292–302.

Ran M, Chen B, Li Z, Wu M, Liu X, He C, Zhang S and Li Z 2016. Systematic identification of long noncoding RNAs in immature and mature porcine testes. Biology of Reproduction 94, 77.

Ren GJ, Fan XC, Liu TL, Wang SS and Zhao GH 2018a. Genome-wide analysis of differentially expressed profiles of mRNAs, lncRNAs and circRNAs during *Cryptosporidium baileyi* infection. BMC Genomics 19, 356.

Ren T, Li Z, Zhou Y, Liu X, Han R, Wang Y, Yan F, Sun G, Li H and Kang X 2018b. Sequencing and characterization of lncRNAs in the breast muscle of Gushi and Arbor Acres chickens. Genome 61, 337–347.

Ren T, Zhou Y, Zhou Y, Tian W, Gu Z, Zhao S, Chen Y, Han R, Liu X and Li Z 2017. Identification and association of novel lncRNA pouMU1 gene mutations with chicken performance traits. Journal of Genetics 96, 941–950.

Shamimuzzaman MD, Le JJ, Unni DR, Diesh CM, Triant DA, Walsh AT, Tayal A, Connat GC, Hagen DE and Elsik CG 2020. Bovine Genome Database: new annotation tools for new reference genome. Nucleic Acids Research 48, D676–D681.

Shen L, Zhang S, Li Q, Fu Y, Tang G, Jiang Y, Li M, Wang J, Li X, Che L and Zhu L 2018. The landscape of non-coding RNA in an adult pig model of intrauterine growth restriction. Cellular Physiology and Biochemistry 50, 1764–1778.

Shen Y, Mao H, Huang M, Chen L, Chen J, Cai Z, Wang Y and Xu N 2016. Long noncoding RNA and mRNA expression profiles in the thyroid gland of two phenotypically extreme pig breeds using ribo-zero RNA sequencing. Genes 7, 34.

Shi G, Chen L, Chen G, Zou C, Li J, Li M, Fang Ch and Li Ch 2019. Identification and functional prediction of long intergenic non-coding RNAs related to subcutaneous adipose development in pigs. Frontiers in Genetics 10, 160.

Storz G 2002. An expanding universe of noncoding RNAs. Science 296, 1260–1263.

Sun J, Xie M, Huang Z, Li H, Chen T, Sun R, Wang J, Xi Q, Wu T and Zhang Y 2017. Integrated analysis of non-coding RNA and mRNA expression profiles of 2 pig breeds differing in muscle traits. Journal of Animal Science 95, 1092–1103.

Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R and Zhao Y 2013. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Research 41, e166.

Suzek BE, Wang Y, Huang H, McGarvey PB and Wu CW 2015. The UniProt Consortium, UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics 31, 926–932.

Tong C, Chen Q, Zhao L, Ma J, Ibeagha-Awemu EM and Zhao X 2017. Identification and characterization of long intergenic noncoding RNAs in bovine mammary glands. BMC Genomics 18, 468.

Trapnell C, Pachter L and Salzberg SL 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111.

Trapnell C, Williams B, Pertea G, Mortazavi A, Kwan G and van Baren MJ, Pachter L 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology 28, 511–515.

Uchida S 2017. High-throughput methods to detect long non-coding RNAs. High Throughput 6, 12.

Wang J, Hua L, Chen J, Zhang J, Bai X, Gao B, Li C, Shi Z, Sheng W, Gao Y and Xing B 2017a. Identification and characterization of long non-coding RNAs in subcutaneous adipose tissue from castrated and intact full-sib pair Huainan male pigs. BMC Genomics 18, 542.

Wang KC and Chang HY 2011. Molecular mechanisms of long noncoding RNAs. Molecular Cell 43, 904–914.

Wang L, Park HJ, Dasari S, Wang S and Kocher JP 2013. CPAT: coding-potential assessment tool using an alignment-free logical regression model. Nucleic Acids Research 41, e74.

Wang X, Yang C, Guo F, Zhang Y, Ju Z, Jiang Q, Zhao X, Liu Y, Zhao H, Wang J, Sun Y, Wang C, Zhu H and Huang J 2019a. Integrated analysis of mRNAs and long noncoding RNAs in the semen from Holstein bulls with high and low sperm motility. Scientific Reports 9, 2092.

Wang Y, Xue S, Liu X, Liu H, Hu T, Qiu X, Zhang J and Lei M 2016. Analyses of long non-coding RNA and mRNA profiling using RNA sequencing during the pre-implantation phases in pig endometrium. Scientific Reports 6, 20238.

Wang YM, Xu HB, Wang MS, Otecko NO, Ye LQ, Wu DD and Zhang YP 2017b. Annotating long intergenic non-coding RNAs under artificial selection during chicken domestication. BMC Evolutionary Biology 17, 192.

Wang Z, Yang Y, Li S, Li K and Tang Z 2019b. Analysis and comparison of long non-coding RNAs expressed in the ovaries of Meishan and Yorkshire pigs. Animal Genetics 50, 660–669.

Warr A, Affara N, Aken B, Bickhart DM, Beiki H, Billis K, Chow W, Eory L, Heather HA, Flicek P, Girón CG, Griffin DK, Hall R, Hannum G, Hourlier T, Howe K, Hume DA, Izuogu O, Kim K, Koren S, Liu H, Manchanda N, Martin FJ, Nonneman DJ, O'Connor RE, Phillippy AM, Gary GA, Rosen BD, Rund LA, Sargent CA, Schook LB, Schroeder SG, Schwartz AS, Skinner BM, Talbot R, Tseng E, Tuggle ChK, Watson M, Smith TPL and Archibald AL 2019. An improved pig reference genome sequence to enable pig genetics and genomics research. bioRxiv 668921.

Weikard R, Hadlich F, Hammon HM, Frieten D, Gerbert C, Koch C, Dusel G and Kuehn C 2018. Long noncoding RNAs are associated with metabolic and cellular processes in the jejunum mucosa of pre-weaning calves in response to different diets. Oncotarget 9, 21052–21069.

Weikard R, Hadlich F and Kuehn C 2013. Identification of novel transcripts and noncoding RNAs in bovine skin by deep next generation sequencing. BMC Genomics 14, 789.

Weng B, Ran M, Chen B, He C, Dong L and Peng F 2017. Genome-wide analysis of long non-coding RNAs and their role in postnatal porcine testis development. Genomics 109, 446–456.

Wu S, Liu Y, Guo W, Cheng X, Ren X, Chen S, Li X, Duan Y, Sun Q and Yang X 2018. Identification and characterization of long noncoding RNAs and mRNAs expression profiles related to postnatal liver maturation of breeder roosters using Ribo-zero RNA sequencing. BMC Genomics 19, 498.

Xiao Y, Hu J and Yin W 2018. Systematic identification of non-coding RNAs: a bioinformatics perspective. Advances in Experimental Medicine and Biology 1094, 9–18.

Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY and Wei L 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Research 39, W316–W322.

Xing B, Bai X, Guo H, Chen J, Hua L, Zhang J, Ma Q, Ren Q, Wang H and Wang J 2017. Long non-coding RNA analysis of muscular responses to testosterone deficiency in Huainan male pig. Animal Science Journal 88, 1451–1456.

Xu E, Zhang L, Yang H, Shen L, Feng Y, Ren M and Xiao Y 2019. Transcriptome profiling of the liver among the prenatal and postnatal stages in chickens. Poultry Science 98, 7030–7040.

Xu Z, Che T, Li F, Tian K, Zhu Q, Mishra SK, Dai Y, Li M and Li D 2018. The temporal expression patterns of brain transcriptome during chicken development and ageing. BMC Genomics 19, 917.

Yan Z, Huang X, Sun W, Yang Q, Shi H, Jiang T, Li S, Wang P and Gun S 2018. Analyses of long non-coding RNA and mRNA profiling in the spleen of diarrheic piglets caused by *Clostridium perfringens* type C. PeerJ 6, e5997.

https://doi.org/10.1017/S1751731120000841 Published online by Cambridge University Press

Yang B, Jiao B, Ge W, Zhang X, Wang S, Zhao H and Wang X 2018. Transcriptome sequencing to detect the potential role of long non-coding RNAs in bovine mammary gland during the dry and lactation period. BMC Genomics 19, 605.

Yang Y, Zhou R, Li W, Liu Y, Zhang Y, Ao H, Li H and Li K 2019. Dynamic transcriptome analysis reveals potential long lon-coding RNAs governing postnatal pineal development in pig. Frontiers in Genetics 10, 409.

Yang Y, Zhou R, Zhu S, Li X, Li H, Yu H and Li K 2017. Systematic identification and molecular characteristics of long noncoding RNAs in pig tissues. BioMed Research International 2017, 6152582.

Yin Z, Lian L, Zhu F, Zhang ZH, Hincke M, Yang N and Hou ZhCh 2020. The transcriptome landscapes of ovary and three oviduct segments during chicken (Gallus gallus) egg formation. Genomics 112, 243–251.

You Z, Zhang Q, Liu C, Song J, Yang N and Lian L 2019. Integrated analysis of lncRNA and mRNA repertoires in Marek's disease infected spleens identifies genes relevant to resistance. BMC Genomics 20, 245.

Yu L, Tai L, Zhang L, Chu L, Li Y and Zhou L 2017. Comparative analyses of long non-coding RNA in lean and obese pigs. Oncotarget 8, 41440–41450.

Zeng B, Chen T, Xie MY, Luo JY, He JJ, Xi QY, Sun JJ and Zhang YL 2019. Exploration of long noncoding RNA in bovine milk exosomes and their stability during digestion in vitro. Journal of Dairy Science 102, 6726–6737.

Zhang T, Zhang X, Han K, Zhang G, Wang J, Xie K and Xue Q 2017b. Genome-wide analysis of lncRNA and mRNA expression during differentiation of abdominal preadipocytes in the chicken. G3 7, 953–966.

Zhang T, Zhang X, Han K, Zhang G, Wang J, Xie K, Xue Q and Fan X 2017a. Analysis of long noncoding RNA and mRNA using RNA sequencing during the differentiation of intramuscular preadipocytes in chicken. PLoS ONE 12, e0172389.

Zhao P, Zheng X, Feng W, Wang H, Kang H, Ning C, Du H, Yu Y, Li B, Zhao Y and Liu JF 2018. Profiling long noncoding RNA of multi-tissue transcriptome enhances porcine noncoding genome annotation. Epigenomics 10, 301–320.

Zhao Y, Li H, Fang S, Kang Y, Wu W, Hao Y, Li Z, Bu D, Sun N, Zhang MQ and Chen R 2015. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. Nucleic Acids Research 44, D203–D208.

Zheng X, Ning C, Zhao P, Feng W, Jin, Y, Zhou L, Yu Y and Liu J 2018. Integrated analysis of long noncoding RNA and mRNA expression profiles reveals the potential role of long noncoding RNA in different bovine lactation stages. Journal of Dairy Science 101, 1–13.

Zou C, Li J, Luo W, Li L, Hu A, Fu Y, Hou Y and Li C 2017a. Transcriptome analysis reveals long intergenic non-coding RNAs involved in skeletal muscle growth and development in pig. Scientific Reports 7, 8704.

Zou C, Li L, Cheng X, Li C, Fu Y, Fang C and Li C 2018. Identification and functional analysis of long intergenic non-coding RNAs underlying intramuscular fat content in pigs. Frontiers in Genetics 9, 102.

Zou C, Li S, Deng L, Guan Y, Chen D, Yuan X, Xia T, He X, Shan Y and Li Ch 2017b. Transcriptome analysis reveals long intergenic noncoding RNAs contributed to growth and meat quality differences between Yorkshire and Wannanhua Pig. Genes 8, E203.

https://doi.org/10.1017/S1751731120000841 Published online by Cambridge University Press