ABSTRACT

Title of Document: IDENTIFICATION AND CHARACTERIZATION OF LONG INTERGENIC NONCODING RNAS ASSOCIATED WITH MAREK'S DISEASE RESISTANCE IN CHICKEN

Fei Zhan

Master of Science, 2013

Directed By: Associate Professor Dr. Jiuzhou Song
Department of Animal and Avian Sciences

Marek's disease (MD) is a highly contagious lymphomatous disease of chicken caused by Marek's disease virus (MDV). MDV has steadily evolved toward increased resistance and virulence over the past decades. A promising strategy for MD prevention and control would be the enhancement of genetic resistance. This study aimed to investigate the roles of long intergenic noncoding RNAs (lincRNAs) in MD resistance and susceptibility in chickens. We reported more than 1000 lincRNA loci in chicken. Computational functional annotation suggested that lincRNAs were involved in a wide range of biological processes. Moreover, distinct lincRNA expression signatures were observed between MD resistance and susceptible chickens. Additionally, a candidate lincRNA termed *linc-stab1* was identified and it may play an important role in MD immune response by regulating a nearby protein-coding

gene *STAB1*. In summary, our results demonstrated that lincRNAs also play an important role in MD resistance and provide good candidates for hypothesis-driven functional studies.

IDENTIFICATION AND CHARACTERIZATION OF LONG INTERGENIC NONCODING RNAS ASSOCIATED WITH MAREK'S DISEASE RESISTANCE IN CHICKEN

By

Fei Zhan

Thesis submitted to the Faculty of the Graduate School of the

University of Maryland, College Park, in partial fulfillment

of the requirements for the degree of

Master of Science

2013

Advisory Committee:

Dr. Jiuzhou Song, Chair
Dr. Héctor Corrada Bravo
Dr. George Liu

Dedication

To my beloved family

Acknowledgements

Firstly, I would like to appreciate my advisor Dr. Jiuzhou Song. Without his guidance, support, and encouragement, my research and thesis work would not have been possible.

Furthermore, I would like to thank my committee members, Dr. Héctor Corrada Bravo and Dr. George Liu, for their insightful comments and suggestions.

In addition, I would also acknowledge all current and former lab members: Dr. Fei Tian, Dr. Apratim Mitra, Dr. Juan Luo, Dr. Yanghua He, Yi Ding, Jose Carrillo, and Chunping Zhao, for their encouragement, suggestions, and friendship.

Finally, I would like to appreciate my family, for their understanding and continuous support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3P-seq	poly(A)-position profiling by sequencing
CAGE	cap analysis of gene expression
CPC	Coding Potential Calculator
dpi	days post infeciton
dscDNA	double stranded cDNA
FANTOM	Funcitonal Annotation of Mouse
FDR	false discovery rate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
G9a	Histone H3 lysine 9 methyltransferase
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
H3K27me3	histone H3 lysine 27 trimethylaiton
H3K36me3	hisotne H3 lysine 36 trimethylation
H3K4me2	hisotne H3 lysine 4 demethylaiton
H3K4me3	histone H3 lysine 4 trimethylaiton
HCC	hepatocellular carcinoma
KS-test	Kolmogorov-Smirnov test
lincRNA	long intergenic noncoding RNA
lncRNA	long noncoding RNA
LSD1	Lysine (K)-specific demethylase 1A
MD	Marek's disease
MDV	Marek's disease virus
MHC	major histocompatibility complex
miRNA	microRNA
MLL	Histone-lysine N-methyltransferase HRX
ncRNA	noncoding RNA
ORF	open reading frame
PCR	polymerase chain reaction
piRNA	Piwi-interacting RNA
Pol II	polymerase II
PRC2	Polycomb Repressive Complex 2
qPCR	real-time polymerase chain reaction
REST	RE1-Silencing Transcription Factor
RIP	RNA immunoprecipitation
SATB1	Special AT-rich sequence-binding protein-1
siRNA	short interfering RNA
snoRNA	small nucleolar RNA
SVM	support vector machine
TDP-43	TAR DNA-binding protein
YY1	YY1 Transcription Factor

Chapter 1: Literature review

Introduction

When the first draft of the human genome was completed in 2000, many researchers were surprised at the low number of genes identified [1]. The latest estimate indicates that our genome have approximately 20,000 protein-coding genes, which represent only ~1.5% of the genome [2, 3]. Nevertheless, large-scale cDNA cloning [4-6], tiling arrays [7-10] and cDNA sequencing (RNA-Seq) [11-13] studies have demonstrated unexpected pervasive transcription in different species ranging from yeast to human. It is reported that up to 70-90% of the human genome will be transcribed at some time during development [8, 13], yielding an unanticipated complex transcriptome with a large population of noncoding transcripts [14]. The functional repertories of metazoan genomes have been greatly expanded by the tens of thousands of noncoding RNAs [15], indicating that the complexity of different organisms may correlate with the noncoding contents [16].

Those RNA molecules that are not translated into proteins are termed noncoding RNAs (ncRNAs). Based on the size of the transcripts, the ncRNAs are classified into small ncRNAs and long ncRNAs (lncRNAs) using an arbitrary cutoff of approximate 200 nt. Small ncRNAs can be further divided into smaller categories such as microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), short interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs). So far, much research has been done on small regulatory ncRNAs and their roles in numerous levels of gene regulation including chromatin architecture, transcription, RNA splicing, RNA editing, and translation have been revealed [17, 18]. However, lncRNAs, which appear to be the largest component of the mammalian noncoding transcriptome, are not well characterized. The lack of sequence conservation, low expression level, and no apparent coding capacity of lncRNAs have caused controversy to their validity and biological significance to the organism [19, 20]. It is argued that some of the lncRNAs may actually originate

from experimental artifacts such as genomic contamination and incomplete reverse transcription [21]. Besides, RNA polymerase II (Pol II) can initiate nonspecific transcription and yield spurious RNA transcripts [20]. Additionally, intensive transcription at one locus frequently “ripples” away into its physical neighboring loci, leading to potential transcription of neighboring regions [22]. Nevertheless, many of the lncRNAs exhibit tissue-specific expression, localization to subcellular compartments, response to stimulus, shared synteny across species, and association with disease, suggesting that they are regulated and unlikely to represent transcriptional “noise” [23, 24]. Although the majority of lncRNAs remain functionally elusive, there exist dozens of well-characterized lncRNAs that play important roles in a wide range of biological processes such as X chromosome inactivation (*Xist*) [25, 26], genomic imprinting (*Kcnq1ot1*, *Air*) [27, 28], and transcriptional regulation (*HOTAIR*, *HOTTIP*, *Mistral*, *lincRNA-p21*, *PANDA*) [29-33].

Due to our limited understanding of the function, mechanism, and evolution, lncRNAs are usually classified based on their relative genomic locations to nearby protein-coding genes (Figure 1.1) [34-36].

(1) Intergenic lncRNAs (also known as long intergenic noncoding RNAs or lincRNAs): lncRNAs located between annotated protein-coding genes.
(2) Intronic lncRNAs: lncRNAs transcribed in either sense or antisense orientation from introns of protein-coding genes but do not overlap with any exons.
(3) Bidirectional lncRNAs: lncRNAs that transcribed from the vicinity (usually < 1kb) of transcription start sites but in divergent directions.
(4) Sense overlapping lncRNAs: lncRNAs that overlap with protein-coding genes and have the same transcriptional directions.
(5) Antisense overlapping lncRNAs: lncRNAs that overlap with protein-coding genes but have the opposite transcriptional directions.

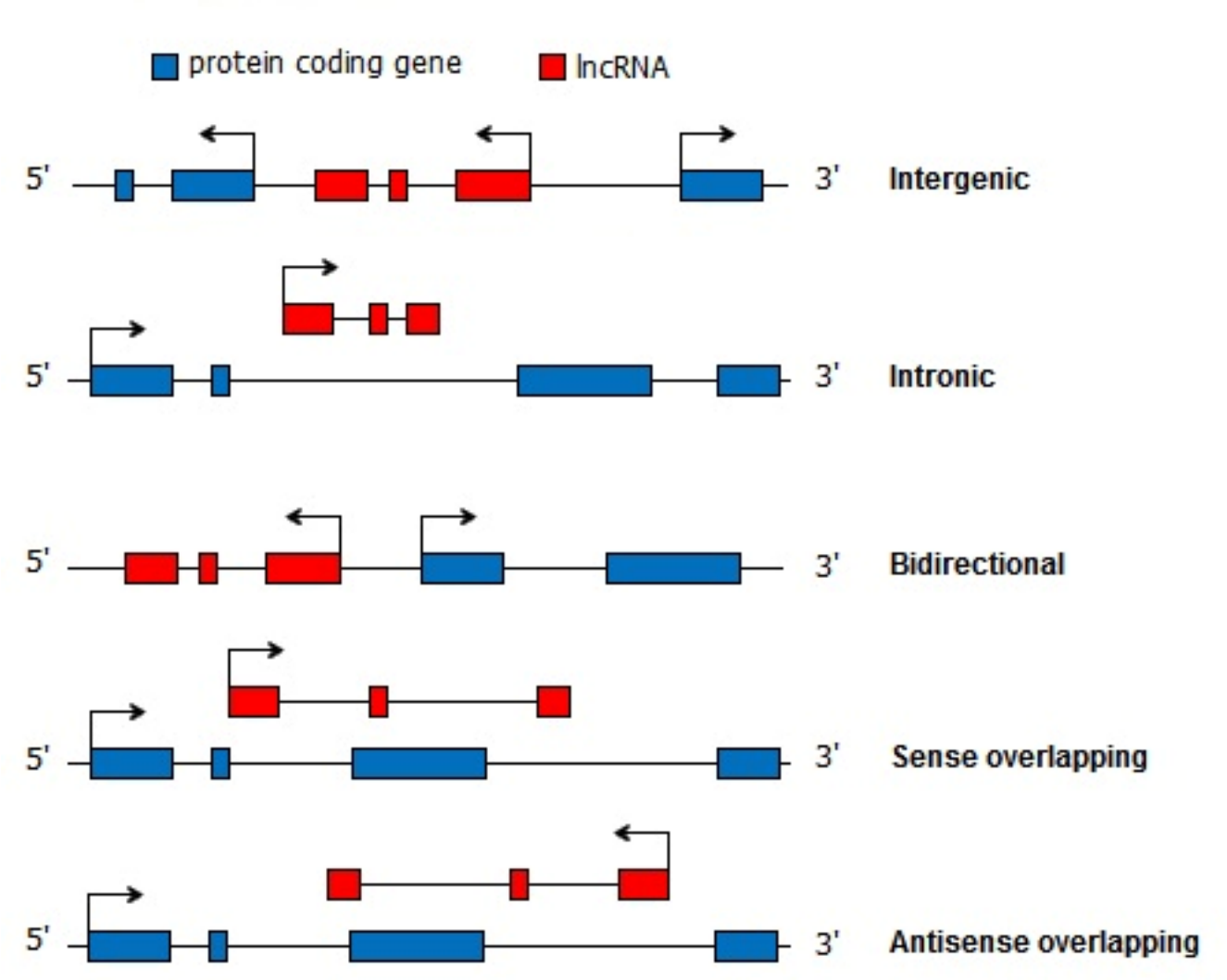


Figure 1.1 LncRNA classifications. LncRNA classification is mainly based on their relative genomic location to nearby protein-coding genes. Coding and noncoding exons are shown in blue and red, respectively. The orientations of transcription are indicated by arrows.

Since long intergenic noncoding RNAs (lincRNAs) do not overlap with any other genes and as a result are relatively easier to define, identify, quantify and interpret compared to other lncRNA categories, recent studies have focused on lincRNAs [37-43]. To date, thousands of lincRNAs have been identified in a variety of species [37-43]. Research shows that lincRNAs share many features in common with mRNAs. For example, most of the lincRNAs are capped, polyadenylated, and usually spliced [44]. Importantly, lincRNAs were involved in a wide range of biological processes including signaling, development, cell cycle, and immune response [37, 39].

In this introduction, we will mainly review the recent progress of genome-wide lincRNAs identification as well as functions and mechanisms that have emerged from some well-characterized examples.

Genome-wide identification of lincRNA

The first genome-wide catalog of lncRNAs was reported by the FANTOM2 (Functional Annotation of Mouse) project in the mouse genome based on automated Sanger sequencing of >60,000 full-length cDNA clones [6, 45], which yielded more than 4000 transcripts as the candidate ncRNAs. With the advent of microarray technology, genomic tiling arrays were widely used for high throughput transcriptome profiling in many species [7, 9, 37, 38] and the number of detected RNA transcripts with no apparent protein-coding capacity has increased exponentially over the past decade. Although tiling arrays were more sensitive than cDNA cloning, there are many technical limitations such as low dynamic range, high false positive rate, low concordance between studies, and difficulties in defining connectivity of transcribed regions [46, 47].

In recent years, next-generation sequencing (NGS) has enabled unprecedented ability to detect novel transcripts by sequencing millions of short cDNA fragments [11, 48, 49]. Together with the advances in computational transcriptome reconstruction [50-52], we are now able to comprehensively identify and characterize noncoding RNAs on a genome-wide scale. After mapping short RNA-Seq reads back to the genome with a splice junction mapper such as TopHat [53], *ab-initio* transcript assemblers such as Cufflinks[50] and Scripture[51] can be used to reconstruct the transcriptome based on mapped reads. However, the repetitive nature of genome sequence and current limitations in sequencing read length make accurate delineation of full-length transcriptional structure not a trivial task, especially for those low abundance transcripts [52]. Although protein-coding transcripts reconstructed by Cufflinks and Scripture have high correspondence, larger discrepancy is observed for noncoding RNAs [41], which

are expressed on average at about 10-fold lower levels than protein-coding genes [38, 39, 41, 51]. In order to achieve better confidence, in some studies both tools were applied for transcript model reconstruction and selected consensus transcripts [39, 41]. Additionally, high-throughput sequencing based methods can pinpoint transcribed regions in the genome using specific features of noncoding RNA molecules. For example, since most lincRNAs are capped and polyadenylated, CAGE (cap analysis of gene expression) [54] and 3P-seq (polyA-position profiling by sequencing) [42, 55] can be used to map transcription start and end sites respectively.

Another major approach is based on chromatin signatures of actively transcribed regions. Using chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-Seq), it is shown that actively transcribed genes are marked by histone H3 lysine 4 trimethylation (H3K4me3) at the promoter regions and histone H3 lysine 36 trimethylation (H3K36me3) along the transcribed regions [56]. By surveying the so called “K4-K36 domains” that located outside annotated protein-coding genes across the genome, about 1600 and 3300 lincRNAs were identified in the mouse and human genome, respectively [37, 38]. This kind of histone pattern based approach is usually coupled with tiling arrays or RNA-Seq to confirm transcription and determine exonic structure of the lincRNAs. A combination of multiple independent evidence would also give better confidence. For instance, by integrating RNA-Seq, H3K4me3 and H3K36me3 chromatin maps, and poly(A)-site mapping data, more than 550 high-confident lincRNAs were identified in zebrafish [42].

Discriminate between lncRNAs and mRNAs

No matter which strategy is used, a fundamental but challenging task is differentiating between protein-coding mRNAs and long noncoding transcripts [21, 57]. Most lncRNAs resemble mRNAs at molecular level and are usually capped, polyadenylated, spliced, and bear mRNA-like chromatin signatures [44].

Besides, emerging evidence has shown that some lncRNAs, such as *SRA* [58], *Oskar* [59] and *VegT* [60], can function through RNA as well as their protein products, which further complicate the classification task.

To efficiently and precisely identify lincRNAs, many attempts have been carried out [21, 57]. One of the most frequently used criteria is the length of open reading frame (ORF) [6]. In eukaryotes a minimum of 100 aa (or 300 nt) is a commonly used cutoff to classify a transcript as protein-coding mRNA. However, many long noncoding transcripts, including some well-characterized ones such as *H19*, *Xist*, *Mirg*, and *HOTAIR*, do have ORFs of at least 100 aa by chance and could be misclassified as mRNA. On the other hand, mRNA can be wrongly classified as ncRNA even if a lower ORF cutoff is used. Although most protein-coding mRNAs are greater than 300 nt in length, a large number of functional peptides as small as 11 aa were identified [61, 62].

Another strategy is searching three-frame translated peptide sequences against known protein or protein domain database (such as Pfam [63]) using BLASTX [64] or HMMER [65]. A transcript is assumed to have coding capacity if any statistically significant hits are found. However, this method is limited by the comprehensiveness and accuracy of current protein database. Moreover, short segments of sequence similarity to mRNAs doesn't necessarily imply that the transcripts will be translated and function as proteins. Research shows that some ncRNAs have evolved from protein-coding genes [66, 67]. For instance, *Xist*, which involve in X chromatin inactivation, has evolved from protein-coding gene *Lnx3* in eutherians by pseudogenization [67].

Similarly, methods that based on evolutionary constrains have also been used [57]. Protein-coding mRNAs bear higher evolutionary pressure to preserve synonymous amino acid contents. Thus, by assessing codon substitution frequency using multiple sequence alignment, ncRNA transcripts should

have higher non-synonymous amino acid change rates under lower pressure to preserve protein-coding capacity.

The most direct evidence for a transcript to be classified as noncoding is that no protein product is produced from the putative ORF. Experimental approaches such as *in vitro* translation assays were used for individual cases [62]. To interrogate translated transcripts in genome-wide scale, ribosome profiling that maps ribosome-associated regions in the genome was used [68]. High throughput mass spectrometry, which can be used to profile the proteome, seems promising in the near future.

All in all, every method discussed above has its own limitations, and there is no single definitive criterion to determine the protein-coding potential of transcripts. A good practice is to combine different strategies to achieve better effects. Tools that can utilize multiple features are also available. For example, features such as ORF length and protein homologs were integrated using SVM (Support Vector Machine) in Coding Potential Calculator (CPC) [69] and CONC (for “coding or non-coding”) [70].

Functions of lncRNA

Recently, so called “guilt by association” studies [37, 39], which infer functions by the correlation of expression between lncRNAs and protein-coding genes, have revealed broad functional roles for lncRNAs in a variety of biological processes. However, currently only a few lncRNAs have experimentally defined functions. Here, emerging roles of lncRNAs in epigenetic regulation and transcription regulation are discussed and exemplified with some well-characterized ones.

LncRNA in X chromosome inactivation: In order to balance the X-linked gene expression dosage

between females and males, one of the X chromosomes is transcriptionally silenced in female. It is shown that this process is mediated by X-inactive-specific transcript (*Xist*), which is a lincRNA transcribed from the X-inactivation center (Xic) [71, 72]. *Xist* is only transcribed from the inactive X chromosome and tethered to Xic by YY1, which serves as a bridge [73]. *Xist* specifically coats the inactive X chromosome and recruits Polycomb repressive complex (PRC2) and LSD1/CoREST/REST complex through 5' and 3' domains, respectively [29]. The whole X chromosome is subsequently inactivated by coupled histone H3 lysine 27 trimethylation (H3K27me3) and histone H3 lysine 4 dimethylation (H3K4me2).

LncRNA in genomic imprinting: LncRNAs can also play important roles in genomic imprinting, by which certain genes are monoallecically expressed in a parent-of-origin fashion. By recruiting chromatin modifying complexes such as PRC2 and G9a, lncRNA can distinguish paternal and maternal alleles and induce allele-specific histone modifications on target chromatin [26, 27, 74]. For example, *Kcnq1ot1* [27], a lncRNA found in the *KCNQ1* locus, is specifically expressed from the paternal allele. It can recruit both PRC2 and G9a to the promoter of *KCNQ1* in *cis* and specifically silence *KCNQ1* expression by inducing H3K27me3 and H3K9me3, resulting in allelic expression of *KCNQ1* only from the maternal allele [27]. Similar to *Kcnq1ot1*, *Air* can target G9a to the chromatin region of *Igf2* and specifically silence the paternal allele expression of *Igf2* through repressive histone marks [74].

Transcriptional regulation through histone modification: *HOTTIP* is a lncRNA transcribed from the distal 5' end of the *HOXA* gene cluster, which coordinates the activation of several 5' *HOXA* genes in *cis* [30]. By directly binding to adapter protein WDR5, *HOTTIP* recruits the histone lysine 4 trimethylases MLL complex. HOTTIP/MLL complex can then be brought into close physical proximity with multiple *HOXA* genes through chromosomal looping, coordinately activating *HOXA* genes by inducing H3K4me3 mark [30]. Another lncRNA called *Mistral*, located between *Hoxa6* and *Hoxa7*, is

also implicated to recruit the MLL complex in *cis* to active neighboring *Hoxa6* and *Hoxa7* [33].

In addition to acting in *cis*, other chromatin-associated lncRNAs can act in *trans*, regulating genes located throughout the genome. *HOTAIR*, which is a lncRNA transcribed from the *HOXC* locus, recruits both PRC2 and LSD1 complexes that induce H3K27me3 and H3K4me2 respectively. It represses *HOXD* gene expression by inducing repressive histone marks across the *HOXD* locus [29, 75].

LncRNAs act as transcriptional coregulators: A p53-induced lncRNA, *lincRNA-p21*, has been reported to regulate p53 mediated apoptosis by physically interacting with transcriptional factor hnRNP-K [31]. By modulating hnRNP-K localization to target genes, lincRNA-p21 can repress expression of genes that are down-regulated for p53 dependent apoptotic response to DNA damage. *PANDA*, which is also a p53 induced lncRNA, is involved in the repression of pro-apoptotic genes by directly interacting with transcription factor NF-YA [32]. Another example of lncRNA-mediated transcriptional repression is related to the regulation of cyclin D1 (*CCND1*). From *CCND1* promoter region, a series of low abundance lncRNAs of variable length are transcribed in response to DNA damage [76]. These lncRNAs remain tethered to the *CCND1* promoter region and allosterically activate RNA-binding protein TLS, which inhibits the activity of histone acetyltransferases CBP and p300, resulting in repressed *CCND1* expression [76].

Other than function as corepressors, lncRNAs can also serve as transcriptional coactivators. *Evf-2* ncRNA, transcribed from the region between *Dlx5* and *Dlx6*, can enhance the activity of Dlx-2 by forming a complex with it and subsequently promote *Dlx5/Dlx6* transcription [77]. Similarly, HSF1 (heat-shock transcription factor 1) activation in heat-shock response is mediated by an ncRNA called *HSR1* (heat-shock RNA-1) [78].

LncRNAs and disease

Aberrant expression of lncRNAs has been shown to associate with different types of cancer and various neurological disorders [79-83]. For example, recent studies report that *HOTAIR* may be involved in modulating chromatin state in a number of cancers and could be a potential target for cancer diagnosis and therapy. *HOTAIR*, located in the *HOXC* locus, is capable of coordinately inducing repressive histone marks H3K27me3 and H3K4me2 by recruiting PRC2 and LSD1 [29, 75]. Gupta et al. [84] shows that *HOTAIR* has increased expression in primary breast tumor and metastases, and thus can be used to predict subsequent metastasis and death. High levels of *HOTAIR* expression is also observed in colorectal cancers [85]. Importantly, high *HOTAIR* expression is reported to tightly correlate with presence of liver metastasis. In another study of hepatocellular carcinoma (HCC), *HOTAIR* expression is elevated compared to normal liver, which suggests the potential of using *HOTAIR* as a prognostic marker for HCC recurrence [86].

Another lincRNA associated with disease is *ANRIL*, which is located in the p15/INK4B-p16/INK4A-p14/ARF tumor suppressor gene cluster, recruits PRC1 and PRC2 complexes [87, 88]. *ANRIL* is involved in the silencing of tumor suppressor genes INK4b-ARF-INK4a and p15/CDKN2B by inducing repressive histone marks [87, 89]. It has been shown that *ANRIL* has altered expression in approximately 30-40% of human tumors [90], suggesting its role in tumorigenesis.

LncRNAs are stable in body fluids, making them ideal candidate biomarkers for noninvasive cancer diagnosis or prognosis [79, 91]. LncRNA *HULC* (highly upregulated in liver cancer), which has high expression levels in HCC patients, can be detected in the blood by conventional PCR methods [92]. Another example is *PCA3* (prostate cancer antigen 3), which is only highly expressed in prostate cancer and levels in urine can serve as an indicator for prostate cancer diagnosis [93]. Similarly, levels of lncRNA *PCAT-1* in urine can be used to identify poor-prognosis prostate patients [79].

Mechanisms of action

Long noncoding RNAs are functionally heterogeneous molecules and might act through a variety of mechanisms. Although only a few dozen of lncRNAs have been well characterized to date, a few potential mechanisms have been proposed [35, 94]. Additional themes will surely emerge as more and more lncRNAs are functionally characterized.

Decoy: LncRNAs can function as decoys for regulatory proteins such as transcription factors and interferes their binding to corresponding targets. *PANDA*, a p53-induced lncRNA, has been reported to interact with the transcription factor NF-YA as a decoy and suppress the expression of pro-apoptotic genes such as *FAS* and *BIK* [32]. Another DNA damage-induced lincRNA, gadd7, can lead to *Cdk6* mRNA degradation by binding to TAR DNA-binding protein 43 (TDP-43) and prevent its interaction with *Cdk6* mRNAs [95]. An additional example is *Gas5*, which is induced upon starvation and releases glucocorticoid receptor from DNA by binding as a decoy [96].

LncRNAs can also function as microRNA sponges and titrate them away from their mRNA targets [97]. It was reported that *CDR1as*, which has 63 conserved binding sites for miR-7, is a miRNA antagonist with high miRNA-binding capacity [97] and expression of the human *CDR1as* gene in zebrafish has a similar phenotype to miR-7 knockdown.

Scaffolds: LncRNAs can serve as scaffolds to bring together multiple proteins into complexes or spatial proximity by binding with distinct domains. For example, *HOTAIR* from the *HOXC* locus is able to recruit both PRC2 and LSD1 complexes that coordinate H3K27me3 and H3K4me2 to silence the *HOXD* cluster in *trans* [29, 75]. Similarly, *Kcnq1ot1* is capable of binding to PRC2 and G9a,

repressing target gene expression by repressive histone marks H3K27me3 and H3K9me3 [27]. *ANRIL*, which is located in a tumor suppressor gene cluster, interacts with both PRC1 and PRC2 to repress the expression of several tumor suppressor genes INK4b-ARF-INK4a and p15/CDKN2B [87-89].

Guides: Several well-characterized lncRNAs may function as guides by recruiting and directing proteins such as chromatin-modifying complexes to specific target regions. *LincRNA-p21*, induced by p53, has been shown to bind with transcription factor hnRNP-K, leading to transcriptional repression at specific genomic loci [31]. Other examples include *HOTTIP* [30] and *Mistral* [33], both of which recruit and direct MLL complex to active *HOXA* genes in *cis*, and *Xist* [25, 26] and *Kcnq1ot1*[27], which induce allele-specific repressive histone marks to silence gene expression.

Rationale and significance

Marek's disease (MD) is a highly contagious lymphomatous and neuropathic disease of chicken caused by Marek's disease virus (MDV), which transforms T lymphocytes and causes mononuclear infiltration of tissues including peripheral nerves, gonad, iris, muscle, visceral organs, and skin [98]. Chicks in most, if not all, commercial settings are commonly exposed to MDV in the first few days of life [99] and infection persists indefinitely. The disease caused severe economic losses until vaccines were developed to control it in 1970s. However, the virus continues to evolve with emergence of new and more virulent strains. As a result, current vaccines are losing efficacy and MD is still one of the major threats to the poultry industry. Better understanding of genetic mechanisms that modulate native and acquired immunity and lead to genetic resistance to MD in chicken will empower us to develop more efficient strategies for control of the disease.

Recent genome-wide transcriptome studies have demonstrated pervasive transcription of noncoding

RNAs in a variety of species including chicken [100]. There is emerging evidence shows that lncRNAs, the largest component of the mammalian noncoding transcriptome, also involved in wide range of biological processes. In recent years, there are some endeavors in revealing functional roles of lncRNAs in immune response [101-103]. In this study, we hypothesize that long intergenic noncoding RNAs (lincRNAs), as an important class of lncRNAs, are associated with MD incidence and resistance in chickens. To test this hypothesis, we used high-throughput transcriptome sequencing (RNA-Seq) data generated from MD resistant and susceptible chickens and cataloged a comprehensive list of lincRNAs that involve in host immune response. By studying the signatures of lincRNA expression, we aim to identify some lincRNAs that involve in MD resistance. We believe that this study will advance our knowledge of immune response to MD and provide clues to developing new strategies to control MD.

Chapter 2: Identification and characterization of lincRNAs associated with Marek's disease resistance in chicken

Introduction

Genome-wide transcriptome studies have demonstrated pervasive transcripts with no protein-coding capacity using large-scale cDNA cloning [4, 5], genomic tiling arrays [7-9] or whole transcriptome sequence (RNA-Seq) [12, 13]. Based on the size of the transcripts, noncoding RNAs are classified into long noncoding RNAs (lncRNAs) and small ncRNAs such as miRNA, snoRNA and siRNA using a somewhat arbitrary cutoff of 200 nt. Much research has been done on small regulatory ncRNAs and their functional roles in different levels of gene regulation including chromatin architecture, transcription, RNA splicing, RNA editing, and translation have been revealed [17, 18]. Although thousands of lncRNAs have been identified in a number of species, only dozens of them have been well characterized so far.

To avoid the complications arising from overlap with other types of genes, recent studies have focused on long intergenic noncoding RNAs (lincRNAs) [37-43]. Although the functions for most lincRNAs remain elusive, dozens of well-characterized ones have suggested a wide spectrum of functions and mechanisms [35, 44, 94]. For example, lincRNAs can recruit repressive chromatin modifying complexes such as PRC2 and G9a to inactivate X chromosome (*Xist*) [26] or to imprint target genes in an allele-specific manner (*Kcnq1ot1*, *Air*) [27, 74]. LincRNAs can also play important roles in transcription regulation through inducing histone modifications (*HOTTIP*, *HOTAIR*, *Mistral*) [29, 30, 33, 75] or act as transcriptional coregulators (*lincRNA-p21*, *PANDA*) [31, 32]. Importantly, lincRNAs are dysregulated in different types of cancer and may associate with cancer and other diseases such as neurological disorders [79-83]. For example, *HOTAIR* has increased expression in primary breast tumor and can be used to predict subsequent metastasis and death [84].

Marek's disease (MD) is a lymphomatous and neuropathic disease of chickens that is caused by Marek's disease virus (MDV). Although the vaccine for MDV infection is widely used since 1970s, MD is still one of the major threats for the poultry industry as MDV has steadily evolved toward increased resistance and virulence over the past decades [104]. A promising strategy for MD prevention and control would be the enhancement of genetic resistance to augment vaccine efficacy. Previous research on host immune response to MDV infection has mainly focused on protein-coding genes. Both major histocompatibility complex (MHC) genes and non-MHC genes have been shown to involve in MD resistance [105-107]. Since lincRNAs are suggested to play important roles in immune response and disease, in this study we identified and characterized lincRNAs that are involved in MD resistance in chicken.

Two inbred chicken lines that have the same MHC haplotype but show dramatic difference in MD incidence were used in the study. RNA-Seq was used to profile the whole transcriptomes of MDV infected and non-infected chickens at three different time points. By using a stringent lincRNA identification pipeline, we reported more than 1000 lincRNA loci in chicken and characterized their genomic and sequence features. Functional analysis based on expression correlation with protein-coding genes associated lincRNAs with various biological processes such as cell cycle, signaling and immune response. Moreover, distinct lincRNA expression profiles were observed between MD resistant and susceptible chickens, suggesting that lincRNAs are involved in host response to MD infection. Notably, a candidate lincRNA termed *linc-satb1* was identified only highly expressed in resistant chickens and may associate with MD resistance by regulation its upstream protein-coding gene *SATB1*. Collectively, results of the study advanced our understandings of immune response to MDV infection and provide good candidates for hypothesis-driven functional studies.

Materials and Methods

Chicken lines & experimental design

Line 6_3 and line 7_2 (USDA-ARS Avian Disease and Oncology Laboratory, East Lansing, Michigan, USA) are two inbred chicken lines with the same MHC haplotype but show dramatic difference in Marek's disease incidence. While line 6_3 is resistant to MD, line 7_2 is highly susceptible. For each chicken line, 12 birds were selected and divided into two groups. One group of the birds was infected with a very virulent (vv+) strain of MDV (648A passage 40) at 5 days after hatching while the other group as control didn't receive any treatment. At 5, 10 and 21 days post infection (dpi) that correspond to early cytolytic, latent, and late cytolytic phase of MDV life cycle, two birds were sacrificed for infected and non-infected groups respectively. A total of 24 samples from bursa of Fabricius were collected and stored in RNAlater solution (QIAGEN) at -80°C for RNA extraction. All procedures were performed following the standard animal ethics and guidelines of ADOL.

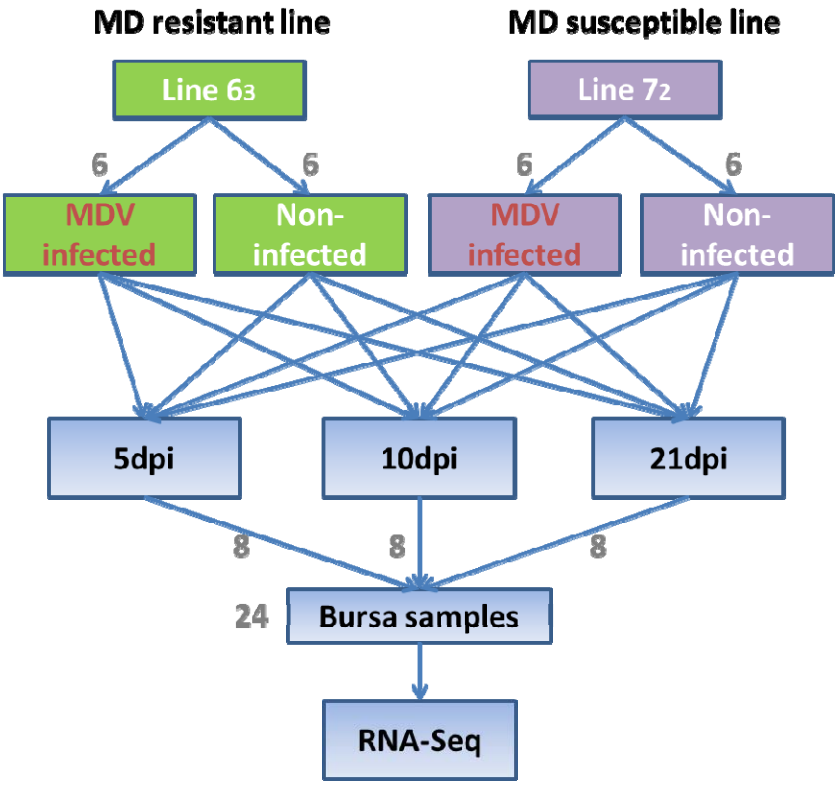


Figure 2.1 Experimental designs. Experimental design for identifying lincRNAs involved in Marek's disease resistance using resistant chicken line 63 and susceptible line 72.

Library preparation & RNA sequencing

Total RNA was isolated using the standard TRIzol (Invitrogen) protocol and mRNA isolation was done by Oligotex mRNA Mini Kit (QIAGEN). And then mRNA was used to synthesize the first and the second strand cDNA by using SuperScriptTM III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) and oligo (dT) 12–18 primers (Invitrogen, Carlsbad, CA, USA). After purification, the double strand cDNA (dscDNA) was fragmented. Then the library for sequencing on the Solexa 1G Genome Analyzer was performed as follow. End repair of the fragmented dscDNA was performed. A 3' A was added the end repaired dscDNA by DNA Polymerase I, Large (Klenow) Fragment. A pair of Solexa adaptors was ligated to the repaired ends by T4 ligase. PCR was used to enrich the purified dscDNA templates. After purification, cluster generation and sequencing analysis were performed on the Illumina Hiseq 2000 following manufacturer protocol.

Public data sources

Chicken genome assembly galGal3 (WUGSC 2.1, May 2006), refGene annotation (galGal3, Feb 2013) and phastCons7way table [108] were downloaded from UCSC Genome Browser (http://genome.ucsc.edu/index.html). Chicken ensGene (WASHUC2.70) annotation was downloaded from Ensembl Genome Browser (http://useast.ensembl.org/index.html). TransMap RefGene for human and mouse were obtained from UCSC Genome Browser. GenBank annotation for human and mouse from NCBI (http://www.ncbi.nlm.nih.gov/) was used for syntenic ortholog identification.

RNA-Seq read mapping

Before mapping, those libraries sequenced to 50bp were trimmed to 30bp by remove the first and last 10bp. RNA-Seq reads for each sample were then mapped to chicken genome individually using the spliced read aligner TopHat (version 2.0.6) [53]. Apart from mapping reads to the reference genome, TopHat is capable of mapping reads that spanning splicing junctions. To maximize the use of exon

junction information derived across all samples, a two iteration mapping strategy was used. In the first iteration, reads were mapped to the chicken genome (galGal3) in order to find potential exon junctions. Since our reads are relatively short (30bp), '--coverage-search' was turned on and '--max-intron-length' was set to 100k, leaving all other parameters as default. Exon junctions identified in each sample were pooled together and a total of 514,704 junctions were identified. With the belief that longer reads can provide additional potential junctions, those eight samples whose reads were 50bp before trimming were mapped to genome using TopHat. All other parameters were to default except '--max-intron-length', which was set to 200k. An additional 47,811 potential exon junctions were identified.

For the second iteration of mapping, all reads were remapped to the genome with exon junctions identified in the first run. Options '--raw-juncs' and '--no-novel-juncs' were used and '--max-intron-length' was set to 100k. The output from this run was used for all subsequent analyses.

RNA-Seq transcriptome assembly

Transcriptomes were assembled individually for each sample with Cufflinks (version 2.0.2) [50]. Cufflinks uses spliced reads to determine the exonic structure of transcripts and a parsimonious set of transcripts is constructed based on mapped RNA-Seq reads. Cufflinks was run with '--GTF-guide' and Ensembl gene annotation was supplied to guide RABT (Reference Annotation Based Transcript) assembly. Besides, options '-frag-bias-correct' and '-multi-read-correct' were turned on to improve transcript abundance estimation. Transcriptomes from all samples were then merged together with cuffmerge to build a consensus set of transcripts across samples.

Identification of chicken lincRNAs

To identify long intergenic noncoding RNAs (lincRNAs), transcripts were compared to genome annotations (ensGene and refGene) with Cuffcompare to exclude those that overlap with protein-

coding genes, pseudogenes, and ncRNAs other than lincRNAs. Remaining transcripts located in the intergenic regions were then subject to the following five filters to identify candidate lincRNAs.

Step 1: To filter out single exon transcripts and those smaller than 200 bases. This step removes suspicious single exon transcripts that may originate from nonspecific transcription initiated by Pol II or DNA contaminations in the cDNA library. Besides by the definition of lincRNA, short transcripts less than 200bp were excluded.

Step 2: By using Coding Potential Calculator (CPC) [69], coding potential score was calculated for each transcript (including reverse complement strand). CPC distinguishes protein-coding transcripts from noncoding RNAs based on the homology and open reading frame features of the input transcripts and a score between -1 and 1 is calculated for each transcript. A transcript was classified as noncoding if potential scores for both forward and reverse complement strands were less than zero. If any of the transcripts within a candidate lincRNA locus was classified as protein-coding, this locus was excluded.

Step 3: To filter out transcripts originated from unannotated protein-coding genes, all candidate lincRNAs were searched against a non-redundant protein database using BLASTX [64]. Transcripts with hits longer than 30bp and e-value less than 0.001 were considered to have significant protein-coding hits. A lincRNA locus was kept if none of the transcripts in the loci has any significant protein-coding hits.

Step 4: In order to exclude transcripts that contain known protein domains, all transcripts were translated into amino acid sequences in all three reading frames. HMMER [65] was used to identify any known protein domain by searching against the Pfam database (Pfam 27.0) [63] and transcripts with significant Pfam hits were excluded.

Step 5: For transcriptome assembly with Cufflinks [50], spliced reads spanning exon junctions were used as important evidence to determine the transcript exonic structures. Thus, the validity of the identified lincRNA transcripts relies on the quality of those spliced reads. However we found that for some candidate lincRNAs, reads spanning their exon junctions have poor sequencing quality or mapping quality. Therefore, to ensure the quality of the lincRNA transcripts, for each putative lincRNA transcript we check the average mapping quality score of mapped spliced reads and the number of spliced reads with highest quality (50). Those candidate lincRNAs without any mapped spliced read of quality 50 or with average quality score less than 10 were discarded.

Gene expression estimation and normalization

The expression level of all protein-coding genes and lincRNAs were estimated using Cufflinks [50] in its expression abundance estimation mode with upper quantile normalization. To get more comprehensive coverage of annotated genes, refGene and ensGene annotations were combined together. The expression levels were represented in FPKM (Fragments Per Kilobase of transcript per Million mapped reads) [50]. We then normalized the expression by taking log2 transformation. For biological replicates, mean expression levels were used.

Clustering of gene expression profiles

To get the expression patterns of lincRNAs, log2 fold change between infected and non-infected chicken were calculated at different time points for each chicken line. Those lincRNAs with maximum expression values across all three different time points greater than 1 FPKM and have fold changes greater or equal than 2 in at least one time point were selected. Expression profiles were then clustered using hierarchal clustering with complete linkage and visualized using heatmaps.

Expression correlation of lincRNAs and their neighboring genes

The nearest neighboring protein-coding genes were identified for each lincRNA locus. We used both refGene and ensGene annotations to identify nearest 3' and 5' neighbors of lincRNAs. When the nearest protein-coding gene was annotated in both of refGene and ensGene annotations, refGene annotation was used. Ensembl gene annotation was used when the distance from the gene to the lincRNA locus was smaller than nearest refGene annotation.

Pearson correlation coefficients were calculated for each lincRNA and its neighboring protein-coding gene pair. Briefly, we took the average FPKM values for biological replicates and created log2 transformed expression vectors that represent gene expression levels across 12 different conditions. Pearson correlation coefficients were then calculated using the expression vectors.

As controls, we also generated neighboring protein-protein gene pairs and random protein-coding gene pairs through random sampling. As described above, we calculated the Pearson correlation coefficients for the two control sets. Finally, Kolmogorov-Smirnov test was applied to compare the strength of expression correlation.

Functional enrichment analysis of lincRNA's neighboring protein-coding genes

To test whether lincRNAs were preferentially located near protein-coding genes with specific functions, DAVID (http://david.abcc.ncifcrf.gov/) [109, 110] was applied to test the enrichment of Gene Ontology (GO) terms for neighboring protein-coding genes of lincRNAs. Representative significant GO terms were selected.

Control regions for lincRNAs

A control set consisted of 1225 transcripts was generated for lincRNAs. To this end, a transcript

structure matched region (with same length and exonic structure) was randomly sampled from unannotated region of the genome for each candidate lincRNA transcript. Regions that were within 10kb of any annotated protein-coding genes were excluded. This randomly generated list of transcripts was used as control for lincRNAs in the study.

Correlation matrix between lincRNAs and protein-coding genes

For each pair of lincRNA and protein-coding gene, we calculated a Pearson correlation coefficient for expressions across 12 conditions. The expression vectors were normalized with FPKM and log2 transformed. A correlation coefficient matrix was generated with pairwise combinations of lincRNAs and protein-coding genes. The matrix was then hierarchically clustered with Euclidian distance metric and complete linkage. A heatmap was generated based on the clustered correlation matrix.

Gene set enrichment analysis

For each lincRNA locus, Pearson correlation coefficients were calculated with all annotated protein-coding genes. We then ranked the protein-coding gene list by their correlation coefficients. Gene Set Enrichment Analysis (GSEA) [111] was used to identify significantly enriched gene sets in the ranked gene list. GSEA uses weighted Kolmogorov-Smirnov test to see whether any predefined sets of related genes exhibit unusual behavior in the current expression profile. Specifically, we were testing whether any gene sets were enriched in the top or bottom of the correlation coefficients ranked gene list. In our study, 825 GO gene sets were downloaded from Molecular Signatures Database v4.0 [111]. Each gene set consists of genes annotated by the same biological process GO term. Those gene sets less than 15 or larger than 500 were excluded from the analysis. To calculate FDR values, gene sets were permuted 1000 times. A gene set was considered significant enriched if its FDR was less than 0.05. Gene sets and lincRNA loci without any significant association with others were removed. An association matrix was constructed between lincRNAs and significant GO gene sets (or functional GO terms). The matrix was

first clustered by GO gene sets based on the hierarchical tree of GO terms. GO gene sets with similar GO functions were grouped in the same cluster. Then the matrix was hierarchically clustered by lincRNA loci. A heatmap was drawn based on the clustered association matrix.

Sequence conservation of lincRNAs

To assess the sequence conservation of lincRNAs, phastCons7way table from UCSC was used. PhastCons7way table contains sequence evolutionary conservation information generated from Multiz alignments of 7 vertebrates. PhastCons scores [108] were calculated using a phylogenetic hidden Markov model. For each lincRNA, we calculated the mean phastCons score along the transcript region. Similarly, we calculated mean phastCons scores for refGene exon regions, refGene intron regions and lincRNA control set. A cumulative frequency was plotted based on the distribution of phastCons scores.

Syntenic orthologs of chicken lincRNAs in human and mouse

TransMap maps genes and related annotations in one species to another using synteny-filtered BLASTZ alignment to determine the most likely orthologs. Based on TransMap information, chicken orthologous genes in human and mouse can be identified. To find syntenic orthologs for chicken lincRNAs, we firstly determined orthologous genes in human and mouse for the lincRNA adjacent protein-coding genes. If both neighboring protein-coding genes of a lincRNA have corresponding orthologs in human or mouse, we checked whether the orthologous gene pair was also adjacent with conserved relative orientation in the human or mouse genome. If they were positionally equivalent, noncoding transcripts annotated between them (if any) were considered as candidate syntenic orthologs for chicken lincRNAs in human and mouse. Only lincRNAs on autosomes were included in this analysis.

Experimental validation of lincRNAs

The protocols of mRNA extraction and dscDNA synthesis are identical as mentioned above. Real-time PCR using SYBR Green PCR Kit were utilized to validate multiple unannotated transcripts coupled with the conventional PCR based on iCycler iQ PCR System (Bio-Rad). The primers for conventional PCR amplification were designed using Primer3 (http://fokker.wi.mit.edu/primer3/input.htm) and confirmed by Oligo 6.0. The melting temperatures were 55-65°C and the length of the amplicons was between 80-500 bp. These primer pairs were designed to span over exons based on chicken genome. Chicken genomic DNA was used as control in PCR amplification.

The primers for real-time PCR assay were designed using Oligo 6.0. The melting temperature was set at 60°C and the length of the amplicons was between 50-200bp. The primer pairs were designed within exons. Replicates were performed for RT-qPCR reactions. qPCR reaction was run using the program as follow: pre-incubation (95°C for 10 min), 40 cycles of amplification (95°C for 10 s, 60°C for 10 s, and 72°C for 10s), melting curves using a heat ramp and cool down. Cycle threshold values (Ct values) were obtained from iCycler iQ PCR software. The expression of lincRNAs and genes was normalized against housekeeping gene of *GAPDH* cDNA in the corresponding samples. The relative fold enrichment of each treatment group was calculated by comparing the enrichment value for the given primer pair to *GAPDH*.

Results

Assembly of chicken transcriptome

In our study, RNA-Seq was used to profile the transcriptome of two inbred chicken lines that show dramatic difference in Marek's Disease incidence. Following the standard Illumina mRNA-Seq protocol, poly-adenylated RNA was purified and converted to cDNA libraries for sequencing. Approximately 600 million short reads were obtained from 24 samples in total. By using a two iteration

mapping strategy, nearly 70% of the reads (more than 400 million) were mapped to chicken genome (galGal3) with TopHat (Table 2.1).

Table 2.1 Read mapping statistics

Line	DPI	Treatment	Raw reads	Mapped reads	Unmapped reads	Percent mapped
L63	5dpi	Infected	26,436,718	22,757,281	3,679,437	86.08%
			25,362,239	21,849,245	3,512,994	86.15%
		Non-infected	20,830,331	13,103,489	7,726,842	62.91%
			20,217,983	14,613,688	5,604,295	72.28%
	10dpi	Infected	32,070,953	27,650,629	4,420,324	86.22%
			27,656,512	23,698,520	3,957,992	85.69%
		Non-infected	19,643,972	11,361,656	8,282,316	57.84%
			17,741,686	11,756,499	5,985,187	66.26%
	21dpi	Infected	24,764,470	19,245,364	5,519,106	77.71%
			16,377,980	14,098,784	2,279,196	86.08%
		Non-infected	23,120,214	17,400,805	5,719,409	75.26%
			12,829,650	9,403,356	3,426,294	73.29%
L72	5dpi	Infected	27,855,862	19,824,345	8,031,517	71.17%
			31,178,311	20,499,964	10,678,347	65.75%
		Non-infected	33,818,194	21,314,161	12,504,033	63.03%
			31,334,208	20,067,622	11,266,586	64.04%
	10dpi	Infected	21,258,758	13,979,594	7,279,164	65.76%
			20,030,690	13,809,847	6,220,843	68.94%
		Non-infected	24,005,256	16,038,502	7,966,754	66.81%
			28,146,097	19,191,739	8,954,358	68.19%
	21dpi	Infected	30,141,835	15,226,575	14,915,260	50.52%
			29,783,547	12,667,613	17,115,934	42.53%
		Non-infected	24,061,054	15,483,473	8,577,581	64.35%
			29,227,316	21,234,014	7,993,302	72.65%

Transcriptomes were assembled individually for each sample using Cufflinks and merged together to build consensus transcript models. In total, 151385 transcripts were identified in 104531 loci across all experimental conditions.

Identification of lincRNAs in chicken

By comparing with RefSeq and Ensembl gene annotations using cuffcompare, transcripts that overlap with protein-coding genes, pseudogenes, and ncRNAs other than lincRNAs were removed. To identify lincRNAs from remaining novel transcripts, aforementioned filtering steps were applied to remove any

unannotated protein-coding genes and suspicious transcripts that may originate from transcript assembly errors or library contaminations. A total of 1056 candidate lincRNA loci with 1225 lincRNA transcripts were identified (Figure 2.2).

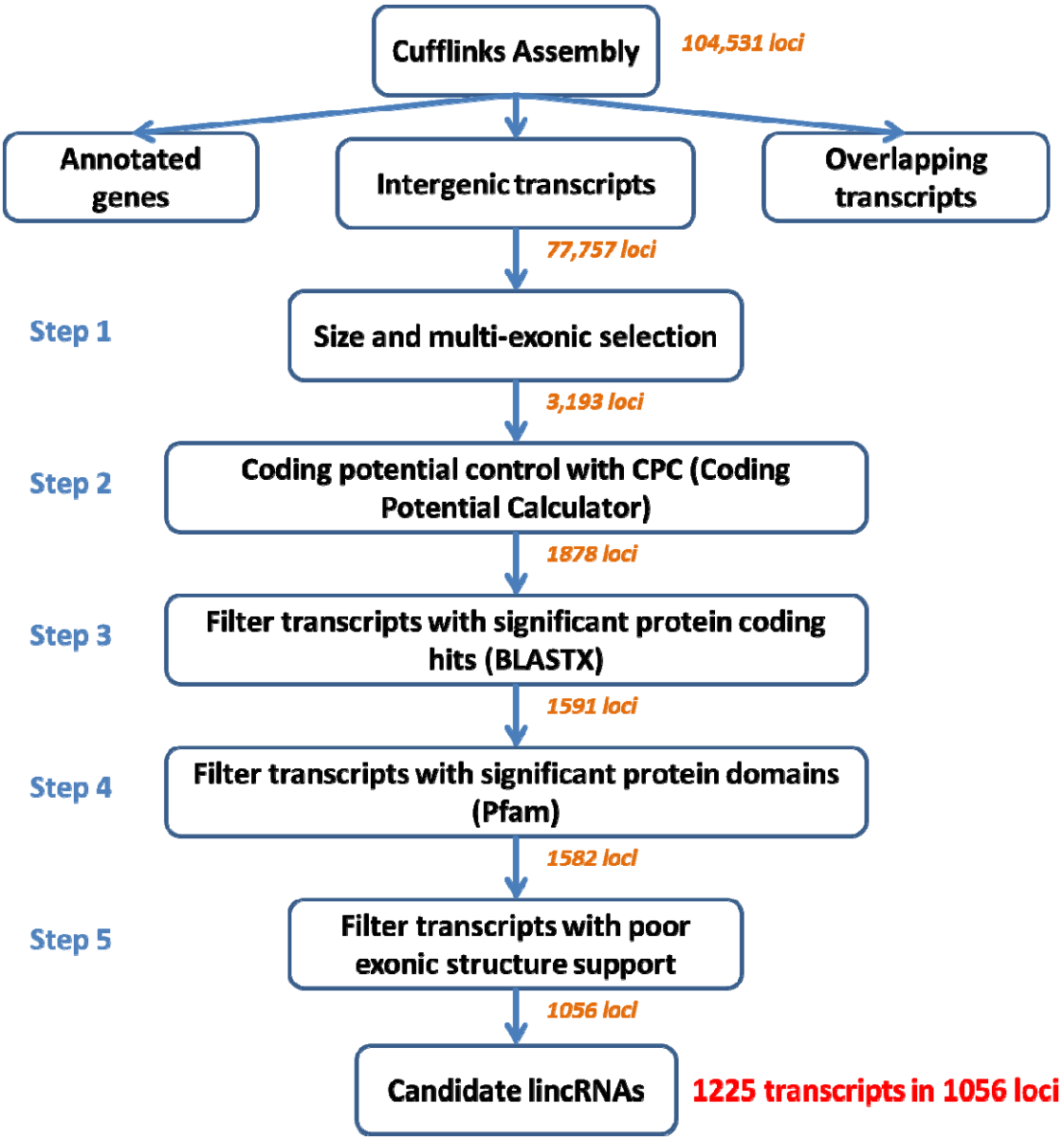


Figure 2.2 LincRNA identification pipeline. After transcriptome assembly with Cufflinks, intergenic transcripts were selected and subjected to five filtering steps that remove unannotated protein-coding genes and spurious transcripts. A total of 1225 candidate lincRNA transcripts were identified in 1056 loci.

Basic lincRNA properties

LincRNAs identified in our study are shorter and have fewer transcripts within each locus compared to

protein-coding genes. The length of lincRNA transcripts ranges from a few hundred bases up to 9kb (with average length about 1kb compared to > 3kb for protein-coding genes) as shown in Figure 2.3A. Whereas protein-coding genes (Ensembl annotation) have about 21exons on average (data not shown here), most of the lincRNAs identified only have two exons (on average 2.2 exons per transcript) (Figure 2.3B). However, this could be a conservative estimate since the lincRNA transcript assembly maybe incomplete due to low expression levels.

By comparing the expression levels of lincRNA and protein-coding genes (Figure 2.3C), we found that the overall expression level of lincRNAs is approximately 8 fold lower than protein-coding genes, which is consistent with previous observations in other species [39, 41]. Furthermore, the phastCons score [108], which is a measure of sequence conservation based on multiple sequence alignments, was used to evaluate the sequence conservation of lincRNA. The mean phastCons scores along the transcript regions were calculated for lincRNAs, exons and introns of RefSeq genes, and control regions for lincRNAs, respectively (Figure 2.3D). As expected, the randomly selected control regions show the lowest conservation; RefSeq exonic and intronic regions have stronger and moderate sequence conservation, respectively, as compared to lincRNAs. The lower phastCons score indicates that lincRNAs have a relative high sequence variation possibly due to a lower selection pressure.

Collectively, lincRNAs identified in our study have fewer exons, shorter transcript length, significantly lower expression levels, and less sequence conservation compared to protein-coding genes. These observations are consistent with studies on other species [39-42].

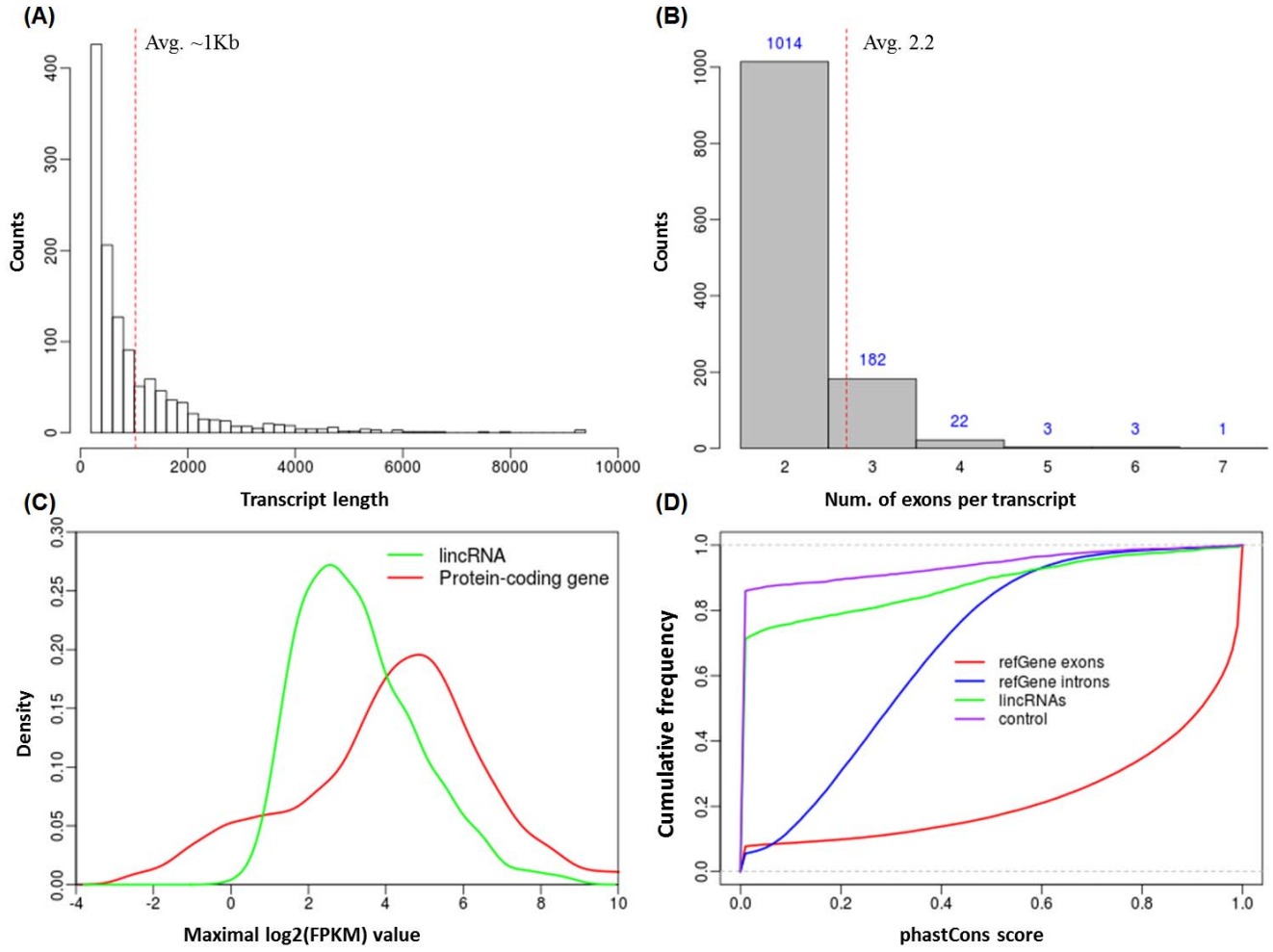


Figure 2.3 Basic properties of lincRNAs. **(A)** LincRNA transcript length distribution. The average length is about 1kb and is marked by red dot line on the figure. **(B)** Number of exons for lincRNA transcripts. As marked by red dot line on the figure, on average there are 2.2 exons per transcripts. **(C)** LincRNA expression levels compared to protein-coding genes. The overall expression level for lincRNAs is much lower than that of protein coding genes. **(D)** LincRNA sequence conservation as measured by phastCons scores. LincRNAs show intermediate sequence conservation as compared to intergenic control regions and protein coding genes.

Precursors for small regulatory RNAs

It was reported that some lncRNAs are actually precursors for small regulatory RNA. For example, H19 was shown to host miR-675 in its first exon [112, 113] and 10 highly conserved snoRNAs were derived from lncRNA Gas5 [114]. In our lincRNA catalog, significant miRNA and snoRNA domains were identified for 7 and 5 lincRNAs, respectively, by searching against Rfam database [115], suggesting that they may be precursors for small regulatory RNAs (Table 2.2). However instead of

producing small RNAs, these lincRNAs may also function independently. For instance, these conserved RNA domains may direct specific binding with DNA or RNA sequence.

Table 2.2 Precursors of small regulatory RNAs

Target name	Accession	Query name	Description
microRNAs			
mir-684	RF00876	TCONS_00000218	microRNA mir-684
mir-207	RF00802	TCONS_00000218	microRNA mir-207
mir-279	RF00754	TCONS_00012819	microRNA mir-279
MIR828	RF01026	TCONS_00078633	microRNA MIR828
mir-573	RF01040	TCONS_00025556	microRNA mir-573
miR-430	RF01413	TCONS_00038360	microRNA miR-430
mir-279	RF00754	TCONS_00064993	microRNA mir-279
mir-207	RF00802	TCONS_00074891	microRNA mir-207
snoRNAs			
TB11Cs4H1	RF01539	TCONS_00001014	Trypanosomatid snoRNA TB11Cs4H1
TB11Cs4H1	RF01539	TCONS_00000421	Trypanosomatid snoRNA TB11Cs4H1
sR11	RF01150	TCONS_00014387	Small nucleolar RNA sR11
snoZ152	RF00350	TCONS_00014387	Small nucleolar RNA Z152/R70/R12/
ceN47	RF01637	TCONS_00057778	C. elegans snoRNA ceN47
SNORA55	RF00431	TCONS_00074562	Small nucleolar RNA SNORA55
telomerase RNAs			
Telomerase-vert	RF00024	TCONS_00074562	Vertebrate telomerase RNA
Telomerase-vert	RF00024	TCONS_00074562	Vertebrate telomerase RNA

LincRNA and their neighboring protein-coding genes

For each lincRNA locus, we found its directly adjacent upstream and downstream protein-coding genes. The distances from lincRNAs to their neighboring protein-coding genes range from a few bases up to 5Mb. The median distance was about 4.5Kb and approximately 65% of the protein-coding neighbors were within 10Kb (Figure 4A).

To test whether lincRNAs were co-expressed with protein-coding neighbors, Pearson correlations of expression levels between lincRNAs and neighboring protein-coding genes were calculated (Figure

4B). We observed stronger expression correlations for lincRNAs and their neighboring protein-coding genes than randomly selected protein-coding gene pairs (Kolmogorov-Smirnov test). However, there is no significant difference in correlation compared to adjacent protein-coding gene pairs. This result is similar to previous studies in human, mouse and zebrafish [39, 41, 116].

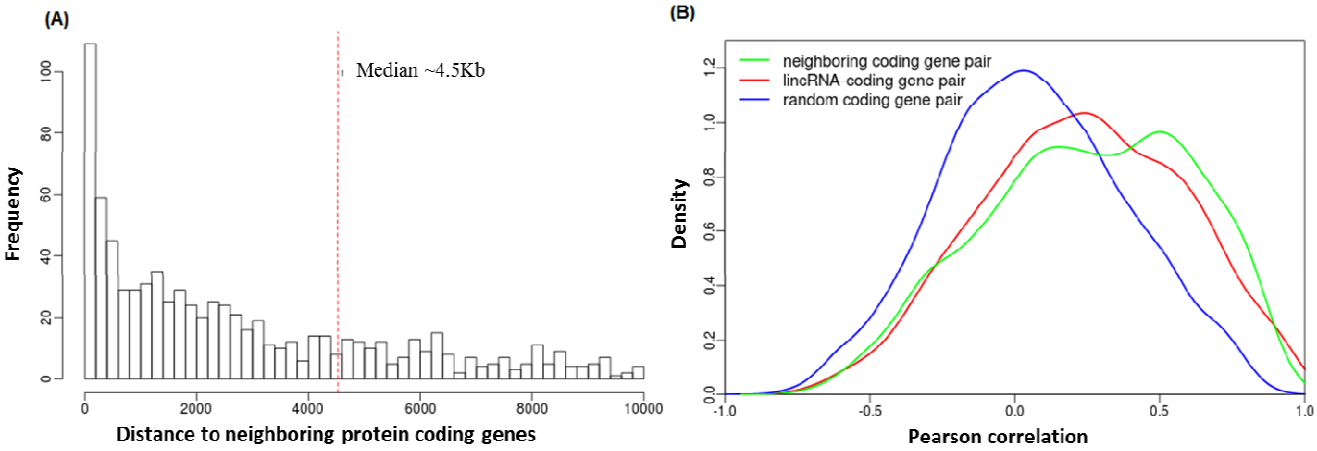


Figure 2.4 LincRNAs and neighboring protein-coding genes. **(A)** The distribution for the distance from lincRNAs to their neighboring protein-coding genes. The median distance is about 4.5kb as marked by red dot line. **(B)** Comparison of expression correlation. Pearson correlation coefficients were calculated for (a) lincRNA loci and their nearest protein-coding genes, (b) adjacent protein-coding genes, and (c) randomly selected protein-coding genes. A density plot was then generated based on correlation coefficients for all three groups.

To test whether lincRNAs were preferentially located in the vicinity of protein-coding genes with specific functional annotations, Gene Ontology (GO) enrichment was computed for lincRNA neighboring genes that were within 10kb. Our results (Table 2.3) indicated that those neighboring genes were significantly enriched ($p < 0.005$) in regulation of cell death, vasculature development, regulation of transcription, gland development, regulation of RNA metabolic process, and lymphocyte activation.

Table 2.3 Selected enriched GOs for lincRNA neighboring protein-coding genes

Term	Description	Count	P-value
GO:0042981	regulation of apoptosis	19	0.000991
GO:0001568	blood vessel development	11	0.002544
GO:0045449	regulation of transcription	44	0.031890

GO:0032268	regulation of cellular protein metabolic process	12	0.004934
GO:0006796	phosphate metabolic process	30	0.027536

Assigning function through expression correlation

Through lincRNA expression correlation with protein-coding genes, previous studies have shown that lincRNAs may involve in biological processes including regulation of transcription, signaling, regulation of development [37, 39, 41, 117], etc. To computationally assign functions to the identified lincRNAs, we constructed a correlation matrix by calculating Pearson correlations between lincRNA loci and protein-coding genes expressions across 12 groups (2 chicken lines, 2 infection states, and 3 different dpi). Strong correlations (either positive or negative) were found between lincRNA loci and certain groups of protein-coding genes (Supplementary Figure 2). We then applied Gene Set Enrichment Analysis (GSEA) to Pearson correlation coefficient-ranked protein-coding genes for each lincRNA. An association matrix was built between 618 lincRNA loci and 159 significantly enriched functional gene sets defined by GO terms. The matrix was first clustered by GO gene set and then hierarchically clustered by lincRNAs.

Based on GO functional terms, GO gene sets were clustered into 8 distinct functional groups with gene sets in each group share similar annotations. As shown in Figure 2.5, we found several sets of lincRNAs associated with functional groups such as signaling, cell cycle, development, transcription, DNA repair and immune response. Of the lincRNA annotated (618 in total), more than 30 percent showed association with signaling. For the rest lincRNAs, most of them were associated with cell cycle, DNA repair and transcription. Almost the same set of lincRNAs was involved in all these three functional groups, which may be explained by the fact that these three functions were highly related. Although most the lincRNAs were only associated with specific functions, a small portion of lincRNAs

(~45) shown a wide spectrum of functionality and were associated with multiple distinct functional groups.

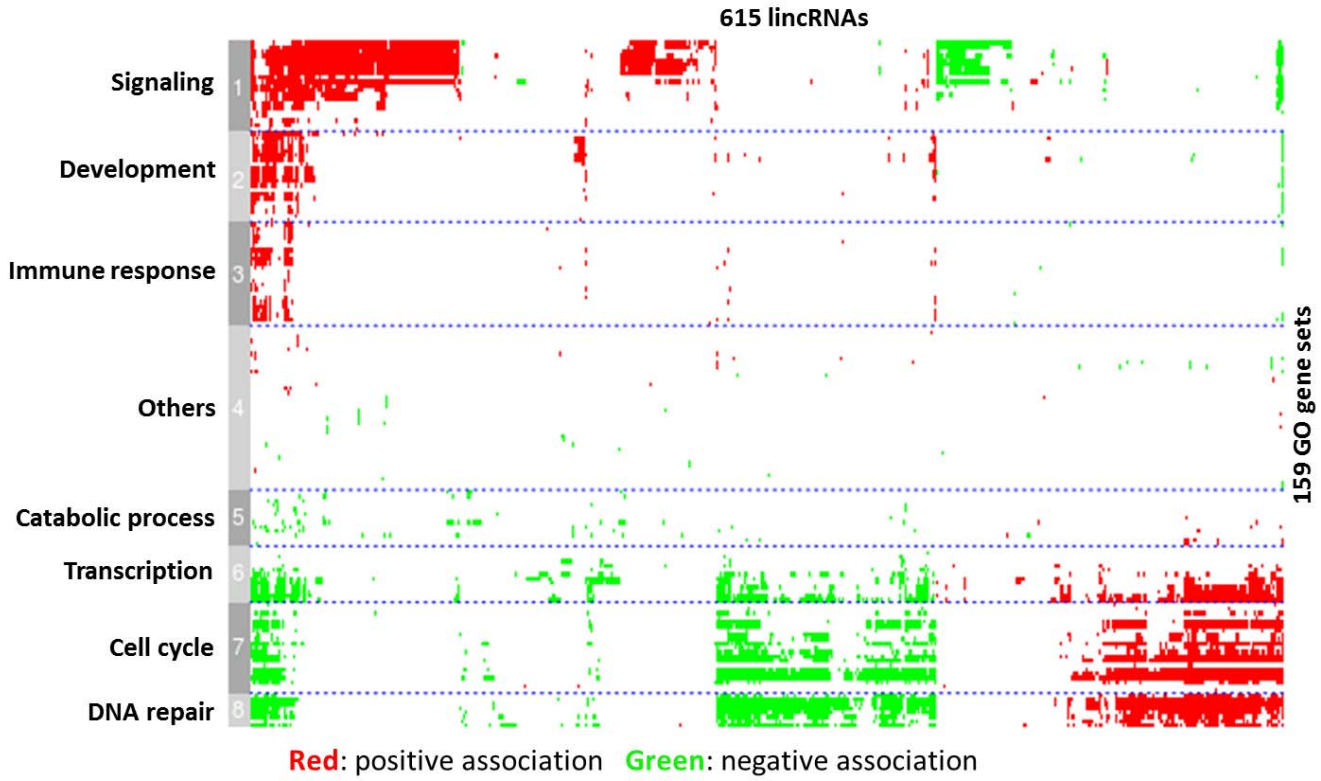


Figure 2.5 Computational lincRNA function annotation. Expression-based association matrix of 615 lincRNA loci (column) and 159 functional gene sets (row). Red represents positive association, green represents negative association, and white represents no significant association. Based on GO gene sets, the matrix was classified into 8 clusters and GO function for each cluster is labeled on the left.

Syntenic lincRNAs of chicken in human and mouse

To investigate the evolutionary conservation of the lincRNAs, we compared the identified chicken lincRNAs with human and mouse. Other than using sequence similarity, potential lincRNA orthologs were identified through searching positional equivalent noncoding RNAs in the human and mouse genome (Figure 2.6). For 759 lincRNA loci located on autosomes, TransMap refGene of chicken was used to determine orthologous genes in human and mouse for protein-coding neighbors of lincRNAs. We found that around 75% (594 and 560 for human and mouse respectively) of neighboring protein-

coding gene pairs that both have human or mouse orthologs were also adjacent in the human or mouse genome. We then identified putative syntenic lincRNA orthologs for chicken by searching noncoding RNAs located in between orthologous gene pairs. About 312 lincRNAs in chicken were observed having positional equivalent noncoding RNAs in the human or mouse genome. Most of these potential syntenic lincRNA orthologs were annotated as miscRNAs and 25 were annotated microRNAs and ncRNA. Importantly, more than 25 were annotated as lincRNAs in the human genome. Detailed inspection of these positionally equivalent noncoding RNAs reveals two lincRNA orthologs located in the *HOXA* cluster. *HOTTIP* (HOXA transcript at the distal tip), which is located at the 5' tip of the *HOXA* locus, is one of the only few well-characterized mammalian lincRNAs [30]. In chicken genome, one positional equivalent lincRNA locus was identified as the *HOTTIP* ortholog. Another lincRNA ortholog *HOXA11-AS*, located between *HOXA11* and *HOXA13*, was identified nearby. Orthologs for these two lincRNAs were also found at positional equivalent locations in the mouse genome, indicating that they were highly conserved during the independent evolution. By investigating the expression correlation of *HOTTIP* homolog in chicken with *HOXA* cluster, we found that *HOTTIP* homolog in chicken was positively correlated with HOXA genes in the cluster, which is consistent with the *HOTTIP*'s function as a 5' *HOXA* gene activator. Interestingly, as the distance to the lincRNA increased (from *HOXA13* to *HOXA1*), the correlation coefficient decreased as well, which may be explained by the *cis* regulatory nature of *HOTTIP* lincRNA.

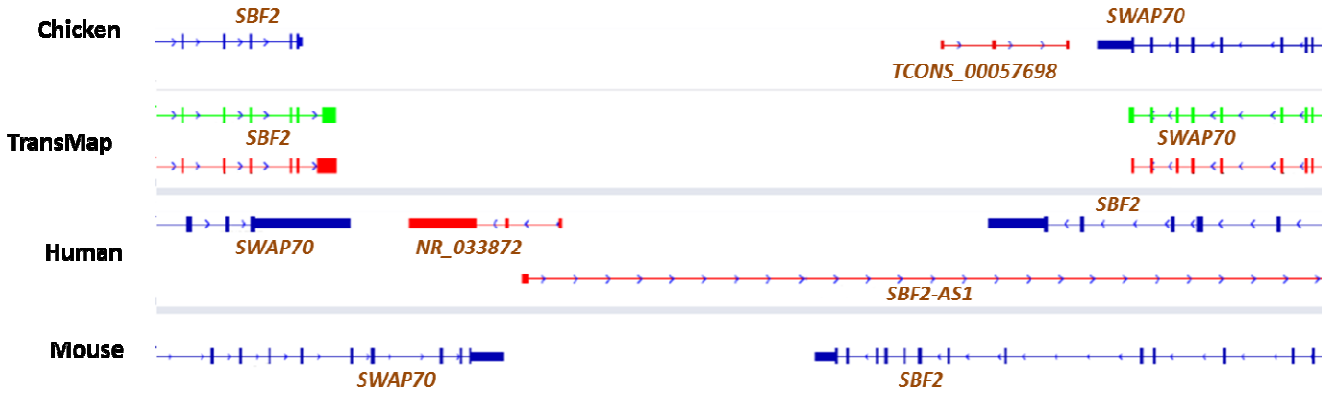


Figure 2.6 An example for syntenic lincRNAs. Transcript TCONS_00057698, is a lincRNA that identified in chicken and located in the upstream of *SWAP70* with opposite transcriptional direction. In human genome, an annotated noncoding RNA, NR_033972, is positionally equivalent with TCONS_00057698. Besides, both noncoding RNAs have three exons. In mouse genome, no positionally equivalent noncoding RNA was identified.

Additionally, we identified 49 (~6%) chicken lincRNAs mapped to human or mouse protein-coding regions. It is likely that these lincRNAs were unannotated protein-coding genes or derived from ancestral protein-coding genes.

LincRNA expression signatures in resistant and susceptible chickens

To systematically investigate the expression changes of lincRNAs induced by MDV infection, we identified significantly differential expressed lincRNAs between infected and non-infected chickens at three different time points for each line. By requiring minimum highest expression level greater than 1 FPKM and a minimum fold change greater or equal than two, a total of 425 and 636 lincRNA loci were identified significantly differential expressed in at least one time points for line 6_3 and line 7_2, respectively (Figure 2.7).

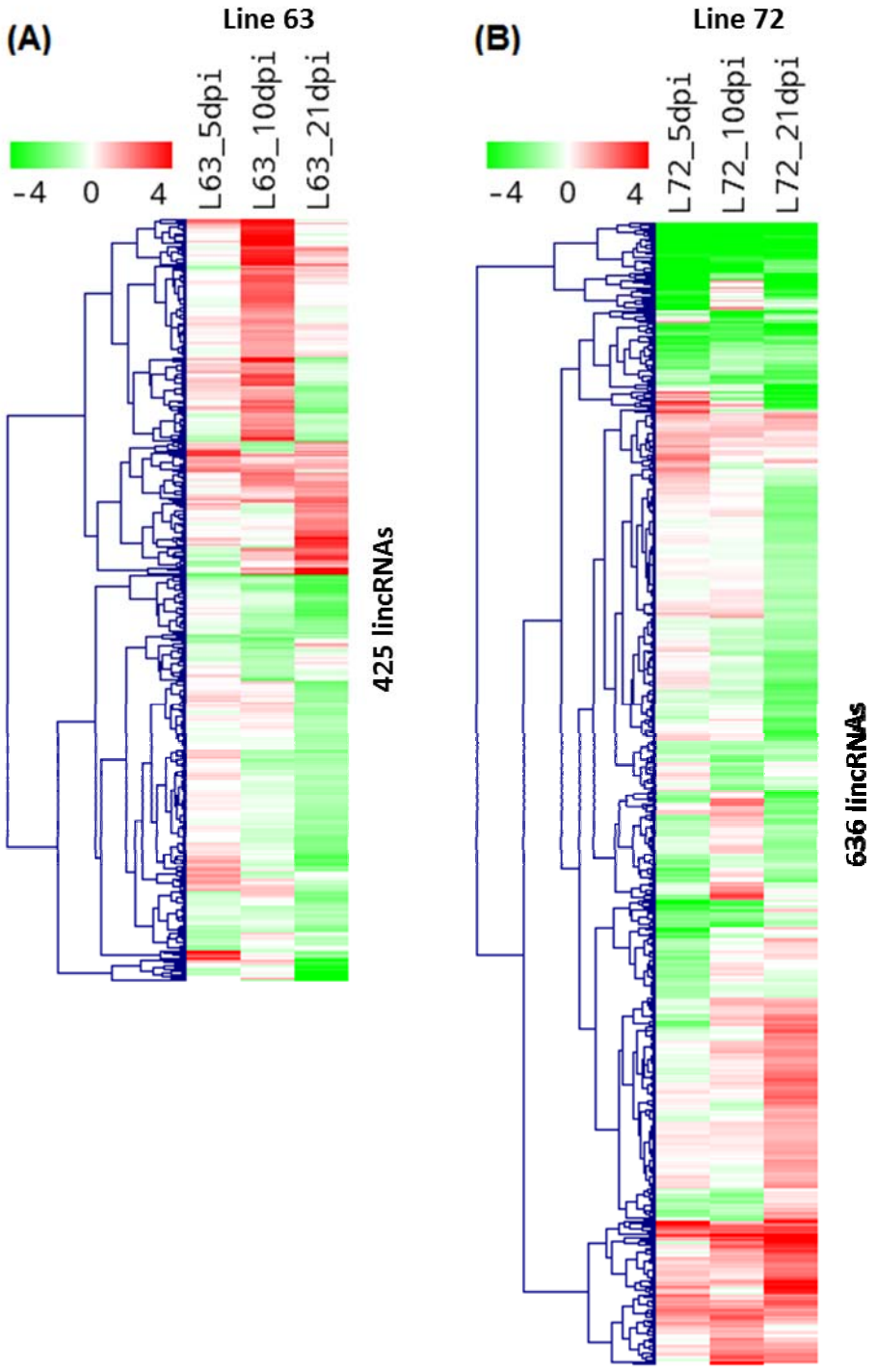


Figure 2.7 LincRNA expression signatures. Expression profiles of differentially expressed lincRNAs between infected and non-infected chickens at three different time points each line. LincRNAs were represented as rows and different experimental conditions were represented as columns. Each value represent log2 ratio of a lincRNA expression level in infected chicken compared to non-infected chicken. The matrix was clustered by lincRNAs using hieratical clustering with complete linkage. Red indicates up-regulation and green represents down-regulation. **(A)** Expression profile for 425 differentially expressed lincRNAs in line 6_3; **(B)** Expression profile for 636 differentially expressed lincRNAs in line 7_2

From the computational lincRNAs annotation, a total of 30 lincRNAs were suggested either positively or negatively associated with immune response related GO terms. Those lincRNAs were selected and their expression profiles were investigated (Figure 2.8) across time points and chicken lines. Interestingly, distinct expression patterns were observed between resistant line 6_3 and susceptible line 7_2. In resistant line 6_3, those lincRNAs were mostly up-regulated in infected chickens at 5dpi and highly expressed at 10dpi. In the contrary, for susceptible line 7_2, most lincRNAs were down-regulated after infection and upregulation were mostly seen at 21dpi. These distinct expression signatures suggest that those lincRNAs were involved in immune response to MD.

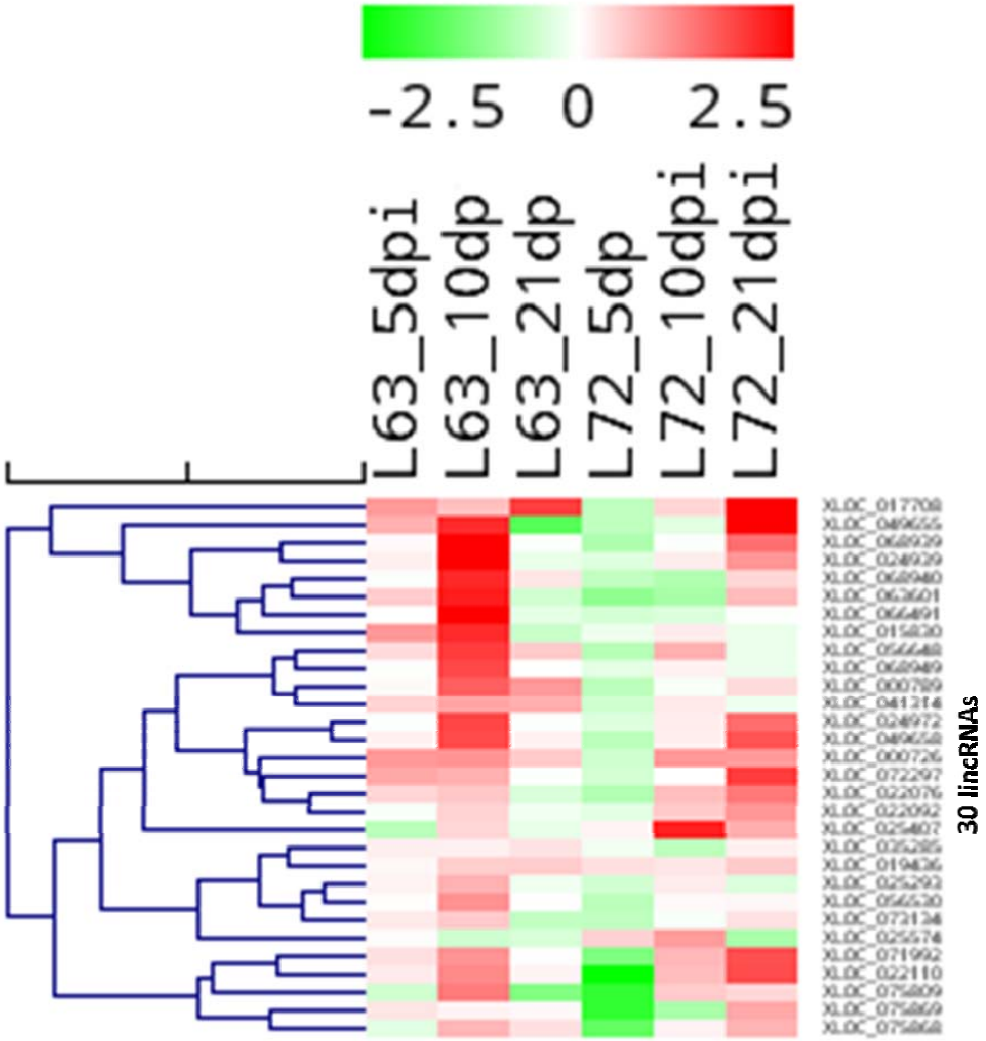


Figure 2.8 Expression signatures for immune response related lincRNAs. Expression profile for 30 lincRNAs that were annotated involving immune response through co-expression based functional annotation. LincRNAs were represented as rows and different experimental conditions were

represented as columns. Each value represent log2 ratio of a lincRNA expression level in infected chicken compared to non-infected chicken for specific chicken line and time points. Red indicates up-regulation and green represents down-regulation. The matrix was clustered by lincRNAs using hieratical clustering with complete linkage.

A lincRNA that may be involved in Marek's disease resistance

Most importantly, a lincRNA (*linc-satb1*) that may associate with Marek's disease resistance was identified in the upstream region of the *SATB1* gene (Figure 2.9A). Based on expression correlation with other protein-coding genes, *linc-satb1* was found to positively associate with defense response, inflammatory response, lymphocyte activation and response to external stimulus; and to negatively associate with cell cycle related functions such as cell cycle process, DNA replication, etc. Moreover, *linc-satb1* is only highly expressed in infected birds of MD-resistant line 6_3 at 10dpi (Figure 2.9B), which corresponds to the latent phase of MDV infection. All these observations suggest that linc-satb1 may associate with immune response to Marek's disease virus infection.

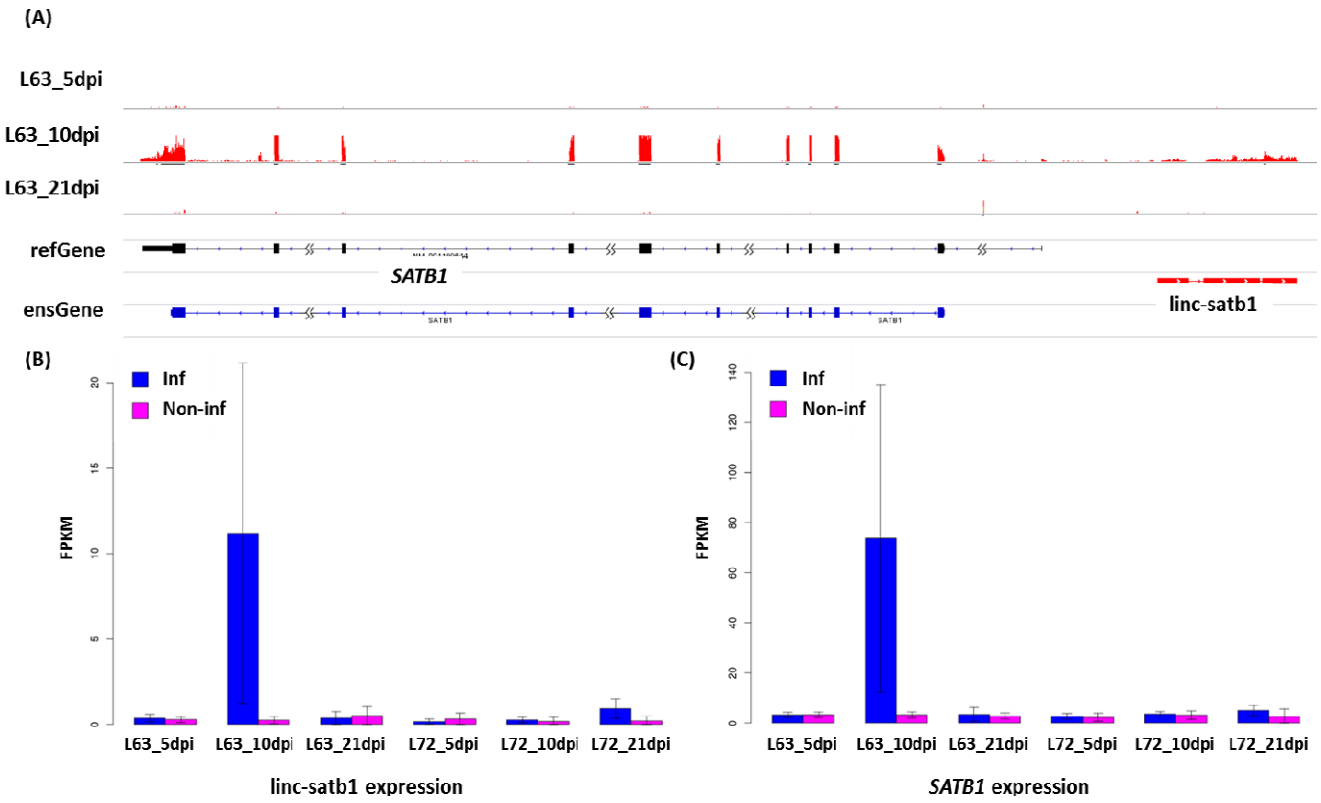


Figure 2.9 LincRNA *linc-satb1* may associate with immune response to Marek's disease. (A)

Relative genomic location is shown for *linc-satb1* (red) and *SATB1* gene (black for refGene annotation, and blue for Ensembl annotation). Expression levels in line 6_3 chickens at 5, 10, and 21dpi are shown in the above three panels. **(B)** Expression levels for *linc-satb1* that was estimated using RNA-Seq data. **(C)** Estimated expression value for protein-coding gene *SATB1*. Infection samples are shown in blue and non-infected control samples are shown in magenta.

A potential homolog of *linc-satb1*, termed GM20098/NR_045095, was found in mouse genome. GM20098 is annotated as a validated miscRNA and has the same synteny with *linc-satb1* in chicken. Both *linc-satb1* and GM20098 have three exons and comparable transcript length. However, like many positional equivalent lincRNAs reported in other species [42], no significant sequence similarity was detected between these two transcripts; and only a few short stretches of sequence are highly similar.

A strong positive correlation (Pearson correlation coefficient ~0.95) was observed between in expression levels of *linc-satb1* and its nearby protein-coding gene *SATB1* (Figure 2.9C), suggesting that *linc-satb1* may play an important role in the immune response by activating *SATB1*, which is a genome organizer that regulates chromatin structure and a transcription factor that controls a large number of genes that involve in T cell development and activation [118, 119]. Notably, we found that the expression of T cell lymphocyte associated activators such as *LEF1*, *TCF7*, and *CCR7* were highly induced for line 6_3 infected birds at 10dpi. More importantly, in agreement with the expression pattern of *SATB1* and *linc-satb1*, cytotoxic T cell co-receptor CD8A and CD8B were only highly expressed in line 6_3 birds at 10dpi. These two glycoproteins are of great importance in cell-mediated immunity, through which cytotoxic T cells recognize infected cells by binding to MHC class I molecules that display virus antigens. CD8+ T cells then induced apoptosis of infected cells and prevented tumorigenesis.

Experimental validation of lincRNAs

To gain confidence in our transcript structure, 17 lincRNA transcripts in 12 loci were verified by

ordinary PCR reaction with dscDNA as template from 24 individual samples. The results showed that four lincRNA structures were confirmed against genomic DNA of chicken, which was capable to amplify one or more fragments in the chicken genome and served as control (Figure 2.10).

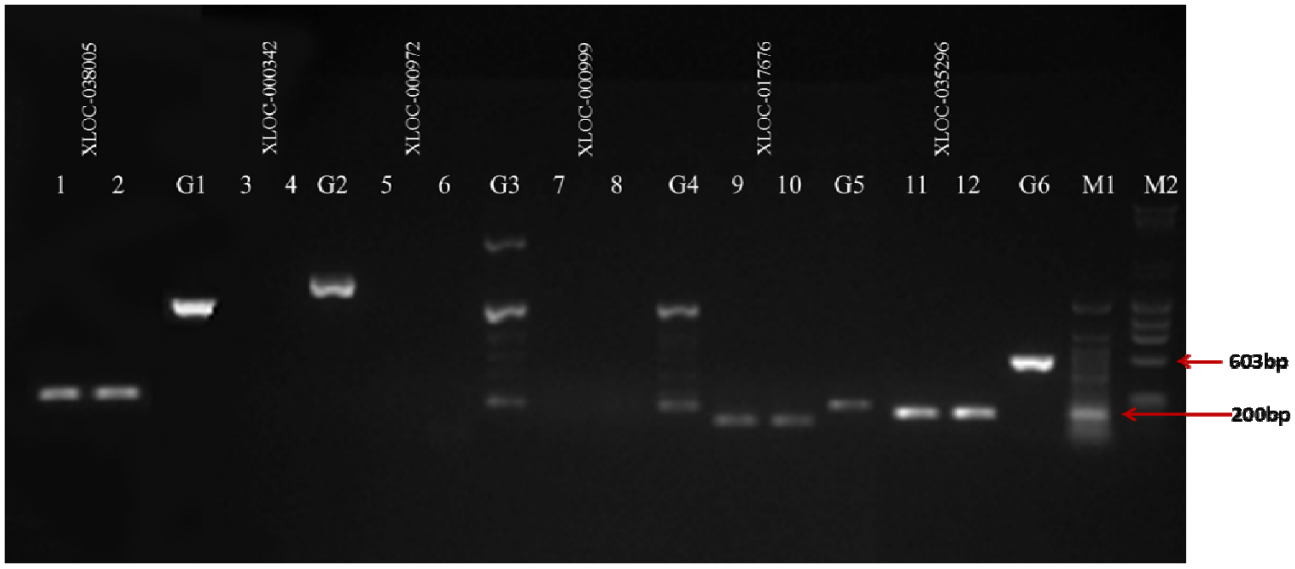


Figure 2.10 Exonic structure validations of candidate lincRNAs based on ordinary PCR. Lane 1-12: amplified fragments for target lincRNAs as the template of dscDNA and corresponding lincRNAs were indicated in the figure; Lane G1-G6: amplified fragments for specific lincRNAs as the template of chicken genomic DNA; Lane M1 and M2: 50bp DNA ladder and broad range DNA ladder (NEB).

In order to confirm several candidate differentially expressed lincRNAs among different treatment groups, quantitative PCR (qPCR) of six lincRNAs in various treatment conditions were conducted and four lincRNAs showed significant differential expression between compared samples (Figure 2.11).

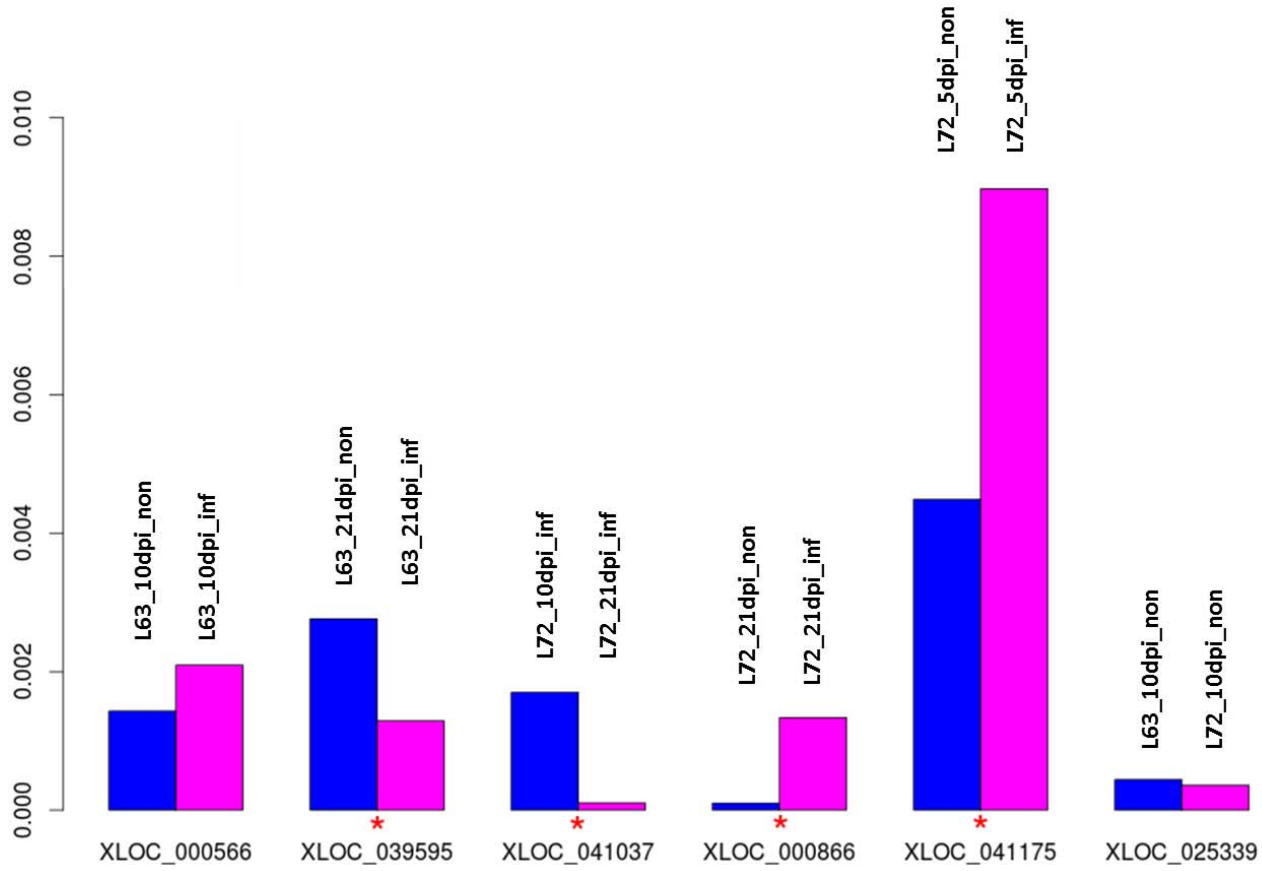


Figure 2.11 Differential lincRNA expression validations based on qPCR assay. Different comparisons were labeled on the graph. LincRNA loci with more than two fold changes are considered differentially expressed and are marked by red star. Four out of six comparisons were validated with qPCR.

Based on previous information, qPCR on lincRNA *linc-satb1* (XLOC_024939) and *SATB1* was conducted to confirm co-expression of linc-satb1 and *SATB1* gene (Figure 2.12). The experimental results confirmed our RNA-Seq based expression estimation (Pearson correlation coefficients are 0.998 and 0.999 for *linc-satb1* and *SATB1*) and linc-satb1 and *SATB1* displayed significant high expression in line 63 infected chickens at 10dpi, in contrast to other treatment groups, which showed no change.

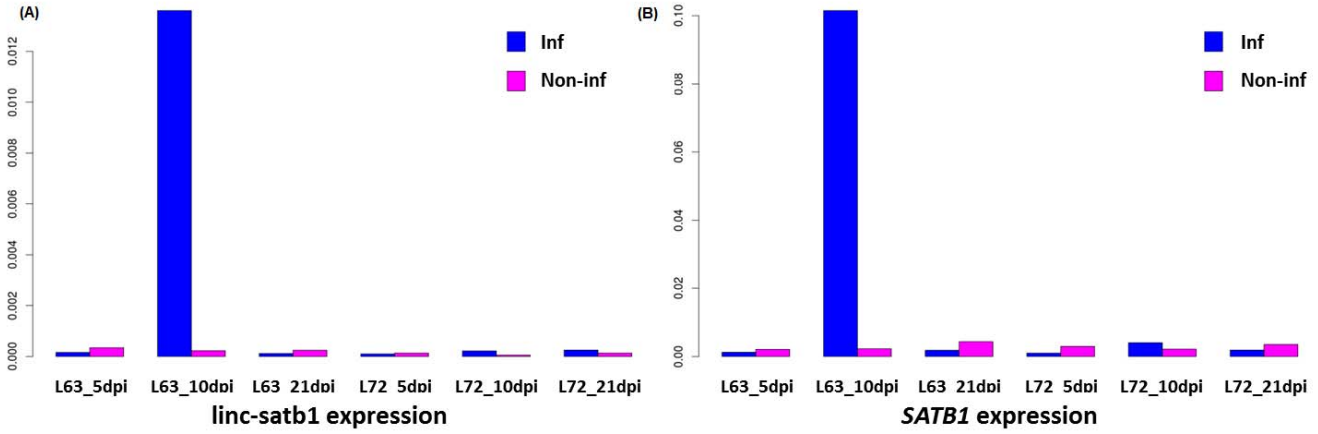


Figure 2.12 Expression validations for linc-satb1 and satb1. Expression validation for lincRNA *linc-satb1* **(A)** and protein-coding gene *SATB1* **(B)** in different treatment groups based on qPCR assay. Infected samples are shown in blue and non-infected control samples are shown in magenta.

Discussion

In this study, RNA-seq was used to interrogate the whole transcriptome of chicken bursa at three different time points that corresponding to important phases of MDV infection. By using *ab-initio* transcriptome assembly following by stringent lincRNA identification pipeline, we ascertained more than 1000 candidate lincRNA loci in chicken. We found that the identified lincRNAs shared similar properties as those reported in mammalian genomes, including significant lower expression, shorter transcript length, fewer exons, and less sequence conservation as compared to known protein-coding genes. Significant small regulatory RNA domains were identified in 12 lincRNAs, suggesting that they may precursors for miRNAs or snoRNAs. In accordance with previous reports [39, 41, 42], we did not detect higher correlation for expression profiles between lincRNAs and their nearby protein-coding gene than adjacent protein-coding genes themselves. However, the correlation is much stronger than randomly selected protein-coding gene pairs. In addition, GO term enrichment analysis of lincRNA neighboring genes indicates that lincRNAs were preferentially located in the vicinity of protein-coding genes that are related to apoptosis, cell death, and transcription regulation.

We found that the sequence conservation for lincRNAs, which was measured based on phastCons score, was significant lower than protein-coding regions but under higher selection pressure than random selected control regions. Despite the relative low sequence conservation, shared synteny was detected for a number of chicken lincRNAs in human and mouse, including some well-characterized lincRNAs such as *HOTTIP*. It is reported that < 6% of zebrafish lincRNAs have detectable sequence conservation with human or mouse lincRNAs [42] and only ~12% of human and mouse lincRNAs appear to be conserved in the other species [41, 120]. These results suggest that unlike protein-coding mRNA that bears high selection pressure to preserve synonymous amino acid sequences, lncRNAs may be subject to constrains by structure or sequence-specific interactions that only preserve short regions of sequence [121]. The properties of a number of well-characterized lncRNAs, such as *Xist*, *Kcnq1ot1*, *HOTTIP* and *HOTAIR*, that scaffold or guide chromatin-modifying complexes by direct binding, are in favor of this hypothesis. Another explanation for at least some of lncRNAs without strict sequence conservation is that active transcription rather than the ncRNA transcript is important [122]. It is shown that some lncRNAs may exert their regulatory functions by helping establishing open chromatin domains and promote the accessibility or protein-coding genes to RNA polymerases [123, 124].

By inferring lincRNA function based on coexpression with protein-coding genes, we found that lincRNAs are mainly involved in signaling, cell cycle, development, transcription regulation, DNA repair and immune response. Previous research has shown that Marek's disease increases DNA damage and oxidative stress in chickens [125]. As suggested by computational function annotation, it is likely that lincRNAs are also involved in the process of DNA repair. As expected, a large number of lincRNAs were differentially expressed between infected and non-infected chickens or between different chicken lines. Detailed examination of immune response associated lincRNAs revealed distinct expression patterns between resistant line 6_3 and susceptible line 7_2 across three different time

points, suggesting their roles in MD resistance. Importantly, a lincRNA termed *linc-satb1* that may associate with Marek's disease resistance was identified specifically highly expressed in infected line 6_3 birds at 10 dpi. Computational lincRNA functional annotation based on expression correlation indicates that *linc-satb1* is positively associated with GO terms such as defense response, inflammatory response, lymphocyte activation and response to external stimulus, suggesting its role in immune response to Marek's disease infection. *Linc-satb1* may exert its function by activating the expression of *SATB1* as strong expression correlation was observed between *linc-satb1* and *SATB1*. By inducing *SATB1* expression, a large number of immune response related genes were upregulated and cell-mediated immunity may be activated to destroy MDV infected cells. Collectively, our analysis indicates that *linc-stab1* mediated *SATB1* gene activation may partially explain the different immune response to MD between these two chicken lines.

Like many well-characterized lncRNAs, *linc-satb1* may activate *SATB1* by recruiting chromatin-modifying complexes in *cis*. Nevertheless, other mechanisms of action for *linc-satb1* are also possible. Instead of acting in cis, *linc-satb1* may act in *trans*, regulating genes located in distal regions. To further investigate the function of *linc-satb1*, additional experimental assays are necessary. For instance, we can overexpress *linc-satb1* in line 7_2 chickens and see whether they can gain resistance to Marek's disease. In addition, RNA immunoprecipitation (RIP) can be used to test whether *SATB1* is activated by *linc-satb1* through recruitment of histone modifying complexes or transcriptional factors.

In summary, in this study more than 1000 chicken lincRNA loci were identified. Computational functional analysis indicated that the putative lincRNAs are associated with a wide range of functions, including immune response. Detailed examination showed that several lincRNAs have different expression signatures in the bursa of MD resistant and susceptible chickens. One of the lincRNA termed *linc-satb1* exhibit strong association with immune response to MD and may play an important

role in MD resistance by regulating nearby protein-coding gene *SATB1*. Our catalog of lincRNAs together with their functional annotations provides excellent candidates for hypothesis-driven experiments. We believe that subsequently overexpression, knockdown or knockout experiments would give us a clear picture about functions of these lincRNAs and their roles in MD immune response.

Chapter 3: Conclusions and discussion

Although vaccines have been developed, Marek's disease (MD) is still one of the major threatens to poultry industry. To understand the mechanisms involved in the host response to MDV, we use MD resistant line 6_3 and MD susceptible line 7_2 as our model. These two lines have the same MHC haplotype but shows distinct susceptibility to MDV, providing a unique model to non-MHC associated MD resistance. In recent years, emerging evidence has shown that lncRNAs, as the largest component of the noncoding transcriptome, can also play important roles in a variety of biological processes including immune response. However, current researches on lncRNAs mainly focus on mammalian genomes such as human and mouse. Only a few studies about lncRNAs were seen on livestock animals such as chicken [100] and cattle [126]. Besides, although noncoding RNAs have been reported involve in chicken immune responses [101], the relevance of lncRNAs in host response to MDV infection has not been systemically studied yet.

One major goal of the present study is to identify a comprehensive catalog of long intergenic noncoding RNAs (lincRNAs) in chicken that challenged by MDV. By using RNA-Seq data generated from MDV-infected chickens and non-infected chickens, we are able to identify more than 1000 lincRNAs loci. A systemically characterization shows that those lincRNAs share similar properties as those reported in mammalian genomes. Comparing to known protein-coding genes, we found chicken lincRNAs have significant lower expression, short transcript length, fewer exons, and less sequence conservation. GO term enrichment analysis of lincRNA neighboring genes indicates that lincRNAs were preferentially located in the vicinity of protein-coding genes that are related to apoptosis and transcription regulation. Co-expression based lincRNAs function annotation shows that they are mainly involved in signaling, cell cycle, development, transcription regulation, DNA repair and immune response. In addition, for a number of identified chicken lincRNAs, positional equivalent noncoding transcripts were identified in human or mouse genome.

Another goal of the study is to systematically investigate the functional roles of lincRNAs in host response to MDV infection. Whole transcriptome profiling of the identified lincRNAs revealed large number of differentially expressed loci, indicating the lincRNAs are involved in the host immune response. Detailed examination of immune response related lincRNAs revealed a lincRNAs termed *linc-satb1* that may promote cell-mediated immunity to MD through regulating nearby protein-coding gene *SATB1*, which is an important genome organizer and transcription factor involved in T cell development and activation. However, the detailed mechanism of the regulation is elusive and need additional gain- or loss-of-function studies.

In conclusion, the present study reported a comprehensive catalog of chicken lincRNAs and a number of lincRNAs that may associate with MD resistance. Our catalog of lincRNAs together with their functional annotations provides deep insights regarding to the mechanisms of Marek's disease resistance and susceptibility.

Supplementary

Supplementary Table 1 Library sequencing statistics

Line	DPI	Treatment	File	Total raw reads	Length
L63	5dpi	Infected	chicken_Bursa_L63_5dpi_inf_1	26436718	30bp
			chicken_Bursa_L63_5dpi_inf_2	25362239	30bp
		Non-infected	chicken_Bursa_L63_5dpi_non_1	20830331	30bp
			chicken_Bursa_L63_5dpi_non_2	20217983	30bp
	10dpi	Infected	chicken_Bursa_L63_10dpi_inf_1	32070953	30bp
			chicken_Bursa_L63_10dpi_inf_2	27656512	30bp
		Non-infected	chicken_Bursa_L63_10dpi_non_1	19643972	30bp
			chicken_Bursa_L63_10dpi_non_2	17741686	30bp
	21dpi	Infected	chicken_Bursa_L63_21dpi_inf_1	24764470	30bp
			chicken_Bursa_L63_21dpi_inf_2	16377980	30bp
		Non-infected	chicken_Bursa_L63_21dpi_non_1	23120214	50bp
			chicken_Bursa_L63_21dpi_non_2	12829650	50bp
L72	5dpi	Infected	chicken_Bursa_L72_5dpi_inf_1	27855862	30bp
			chicken_Bursa_L72_5dpi_inf_2	31178311	30bp
		Non-infected	chicken_Bursa_L72_5dpi_non_1	33818194	50bp
			chicken_Bursa_L72_5dpi_non_2	31334208	50bp
	10dpi	Infected	chicken_Bursa_L72_10dpi_inf_1	21258758	30bp
			chicken_Bursa_L72_10dpi_inf_2	20030690	30bp
		Non-infected	chicken_Bursa_L72_10dpi_non_1	24005256	50bp
			chicken_Bursa_L72_10dpi_non_2	28146097	50bp
	21dpi	Infected	chicken_Bursa_L72_21dpi_inf_1	30141835	30bp
			chicken_Bursa_L72_21dpi_inf_2	29783547	30bp
		Non-infected	chicken_Bursa_L72_21dpi_non_1	24061054	50bp
			chicken_Bursa_L72_21dpi_non_2	29227316	50bp

Supplementary Table 2 LincRNAs and their nearest neighboring protein coding genes

Chromosome	Start	End	Transcript	Nearest gene	Distance
chr1	3364074	3366587	TCONS_00000027	ENSGALT00000010225	12747
chr1	6824147	6841053	TCONS_00000731	ENSGALT00000016041	1171
chr1	6824174	6841109	TCONS_00000029	ENSGALT00000016041	1198
chr1	7899090	7904015	TCONS_00000735	ENSGALT00000010717	251
chr1	11405637	11414123	TCONS_00000039	ENSGALT00000039793	28856
chr1	11405637	11424659	TCONS_00000040	ENSGALT00000039793	18320
chr1	14736227	14739292	TCONS_00000051	ENSGALT00000013259	2157
chr1	14738450	14739292	TCONS_00000053	ENSGALT00000013259	2157
chr1	25605939	25617123	TCONS_00000080	ENSGALT00000014754	26817
chr1	26841646	26853546	TCONS_00000766	ENSGALT00000038937	122
chr1	28634101	28641056	TCONS_00000091	ENSGALT00000015373	1351
chr1	29076125	29077832	TCONS_00000093	ENSGALT00000038867	739
chr1	32510066	32512277	TCONS_00000777	ENSGALT00000015735	30973

chr1	32510066	32517985	TCONS_00000778	ENSGALT00000015735	25265
chr1	36177822	36180373	TCONS_00000104	NM_205001	457
chr1	36177872	36186494	TCONS_00000796	NM_205001	507
chr1	36780139	36783424	TCONS_00000801	ENSGALT00000016103	99
chr1	36780139	36783599	TCONS_00000802	ENSGALT00000016103	99
chr1	37371130	37375132	TCONS_00000805	ENSGALT00000016200	2543
chr1	38332779	38339294	TCONS_00000109	ENSGALT00000016553	2229
chr1	38332780	38339322	TCONS_00000809	ENSGALT00000016553	2230
chr1	39803318	39813005	TCONS_00000813	ENSGALT00000016642	1227
chr1	41340038	41362614	TCONS_00000121	ENSGALT00000016802	1190
chr1	46227327	46230344	TCONS_00000125	ENSGALT00000018400	52268
chr1	46770295	46774311	TCONS_00000128	ENSGALT00000018422	2697
chr1	47213527	47219780	TCONS_00000826	ENSGALT00000018479	427
chr1	47217388	47220480	TCONS_00000131	ENSGALT00000030638	587
chr1	51565206	51574286	TCONS_00000146	ENSGALT00000019514	640
chr1	52147641	52149451	TCONS_00000148	ENSGALT00000019690	4800
chr1	52249532	52255647	TCONS_00000845	ENSGALT00000019693	20214
chr1	52869924	52877921	TCONS_00000847	ENSGALT00000030597	449
chr1	55565482	55567409	TCONS_00000854	ENSGALT00000037715	64371
chr1	55690633	55693781	TCONS_00000856	ENSGALT00000020629	182
chr1	57799530	57802424	TCONS_00000175	ENSGALT00000020872	76
chr1	58001978	58010774	TCONS_00000176	ENSGALT00000020872	92469
chr1	58197272	58206786	TCONS_00000864	ENSGALT00000020873	73
chr1	58200499	58207485	TCONS_00000866	ENSGALT00000020873	3300
chr1	58203181	58207485	TCONS_00000867	ENSGALT00000020873	5982
chr1	58203181	58207485	TCONS_00000869	ENSGALT00000020873	5982
chr1	58203181	58207485	TCONS_00000870	ENSGALT00000020873	5982
chr1	58203181	58207485	TCONS_00000871	ENSGALT00000020873	5982
chr1	58203181	58224673	TCONS_00000873	ENSGALT00000020873	5982
chr1	58604015	58612396	TCONS_00000182	ENSGALT00000020934	492
chr1	61447279	61456153	TCONS_00000886	ENSGALT00000021101	1970
chr1	61449330	61461281	TCONS_00000192	ENSGALT00000021101	4021
chr1	61454535	61471709	TCONS_00000890	ENSGALT00000021101	9226
chr1	61457882	61471653	TCONS_00000894	ENSGALT00000021101	12573
chr1	61691981	61693420	TCONS_00000195	ENSGALT00000021115	968
chr1	64011632	64013029	TCONS_00000205	ENSGALT00000021284	2638
chr1	66832438	66839546	TCONS_00000209	ENSGALT00000021444	83296
chr1	74363339	74365876	TCONS_00000921	ENSGALT00000030506	7319
chr1	74525413	74531789	TCONS_00000227	ENSGALT00000037288	78271
chr1	74528248	74531789	TCONS_00000228	ENSGALT00000037288	81106
chr1	75346250	75350404	TCONS_00000925	ENSGALT00000037278	1341
chr1	75351021	75370306	TCONS_00000231	ENSGALT00000037278	6112
chr1	75914769	75915674	TCONS_00000927	ENSGALT00000027934	1620
chr1	78463032	78465015	TCONS_00000931	ENSGALT00000023176	1528
chr1	78495608	78499189	TCONS_00000236	ENSGALT00000023174	1196

chr1	79320448	79323668	TCONS_00000238	ENSGALT00000023057	5748
chr1	80238625	80242069	TCONS_00000242	ENSGALT00000023348	68
chr1	80730769	80734545	TCONS_00000940	ENSGALT00000023673	382
chr1	81419139	81425567	TCONS_00000245	ENSGALT00000029277	37777
chr1	81635014	81641760	TCONS_00000247	NM_205118	7235
chr1	81635014	81642126	TCONS_00000248	NM_205118	6869
chr1	81635154	81642552	TCONS_00000249	NM_205118	6443
chr1	81642767	81647938	TCONS_00000944	NM_205118	1057
chr1	82359055	82363524	TCONS_00000251	ENSGALT00000023883	12242
chr1	82655152	82656660	TCONS_00000254	ENSGALT00000023929	2899
chr1	83617291	83637496	TCONS_00000953	ENSGALT00000024199	637
chr1	83633683	83637496	TCONS_00000955	ENSGALT00000024199	637
chr1	83651536	83666847	TCONS_00000958	ENSGALT00000024199	3639
chr1	83654299	83670472	TCONS_00000267	ENSGALT00000024217	1495
chr1	83662800	83666847	TCONS_00000960	ENSGALT00000024217	5120
chr1	86271529	86275954	TCONS_00000973	NM_001012820	2329
chr1	87543510	87544173	TCONS_00000279	ENSGALT00000036916	2421
chr1	89988931	89989891	TCONS_00000288	ENSGALT00000024772	1558
chr1	95251624	95254533	TCONS_00001011	ENSGALT00000036845	8297
chr1	95721538	95723828	TCONS_00000312	ENSGALT00000024971	17050
chr1	95721538	95728104	TCONS_00000313	ENSGALT00000024971	12774
chr1	95722474	95734852	TCONS_00000316	ENSGALT00000024971	6026
chr1	95724021	95728104	TCONS_00000317	ENSGALT00000024971	12774
chr1	95725567	95732901	TCONS_00001015	ENSGALT00000024971	7977
chr1	95729222	95730425	TCONS_00000319	ENSGALT00000024971	10453
chr1	95731249	95734852	TCONS_00000321	ENSGALT00000024971	6026
chr1	95905438	95907203	TCONS_00000323	ENSGALT00000024988	24155
chr1	95912903	95913666	TCONS_00000324	ENSGALT00000024988	17692
chr1	95918657	95921428	TCONS_00001017	ENSGALT00000024988	9930
chr1	98467988	98482758	TCONS_00001019	ENSGALT00000025002	4722
chr1	98475700	98485586	TCONS_00000328	ENSGALT00000025002	1894
chr1	98475942	98482651	TCONS_00000329	ENSGALT00000025002	4829
chr1	98478697	98485544	TCONS_00000331	ENSGALT00000025002	1936
chr1	1.07E+08	1.07E+08	TCONS_00001032	ENSGALT00000025470	107
chr1	1.08E+08	1.08E+08	TCONS_00001035	ENSGALT00000036785	47977
chr1	1.12E+08	1.12E+08	TCONS_00001055	ENSGALT00000025993	59067
chr1	1.13E+08	1.13E+08	TCONS_00001056	ENSGALT00000026021	3608
chr1	1.13E+08	1.13E+08	TCONS_00000371	ENSGALT00000026100	7297
chr1	1.14E+08	1.14E+08	TCONS_00001061	ENSGALT00000026124	3055
chr1	1.15E+08	1.15E+08	TCONS_00000378	ENSGALT00000026163	87043
chr1	1.16E+08	1.16E+08	TCONS_00001071	ENSGALT00000036657	6755
chr1	1.16E+08	1.16E+08	TCONS_00000380	ENSGALT00000036657	26352
chr1	1.17E+08	1.17E+08	TCONS_00000385	ENSGALT00000026238	1298
chr1	1.18E+08	1.18E+08	TCONS_00000387	ENSGALT00000026271	13977
chr1	1.19E+08	1.19E+08	TCONS_00000391	ENSGALT00000026256	2965

chr1	1.26E+08	1.26E+08	TCONS_00001086	ENSGALT00000026765	5406
chr1	1.29E+08	1.29E+08	TCONS_00001101	ENSGALT00000035437	18436
chr1	1.29E+08	1.29E+08	TCONS_00000421	ENSGALT00000035437	13214
chr1	1.3E+08	1.3E+08	TCONS_00001106	ENSGALT00000026826	3044
chr1	1.32E+08	1.32E+08	TCONS_00001118	ENSGALT00000030281	474
chr1	1.34E+08	1.34E+08	TCONS_00000437	ENSGALT00000027004	2567
chr1	1.34E+08	1.34E+08	TCONS_00000438	ENSGALT00000027004	2790
chr1	1.34E+08	1.34E+08	TCONS_00000439	ENSGALT00000027004	8641
chr1	1.34E+08	1.34E+08	TCONS_00000440	ENSGALT00000027004	12284
chr1	1.36E+08	1.36E+08	TCONS_00001133	ENSGALT00000030248	6487
chr1	1.38E+08	1.38E+08	TCONS_00000448	ENSGALT00000027090	2640
chr1	1.39E+08	1.39E+08	TCONS_00000455	ENSGALT00000027116	30
chr1	1.41E+08	1.41E+08	TCONS_00001148	ENSGALT00000027137	101672
chr1	1.41E+08	1.41E+08	TCONS_00000457	ENSGALT00000027141	68809
chr1	1.41E+08	1.41E+08	TCONS_00000463	ENSGALT00000027148	847
chr1	1.43E+08	1.43E+08	TCONS_00001154	ENSGALT00000027219	1622
chr1	1.48E+08	1.48E+08	TCONS_00000487	ENSGALT00000027257	6436
chr1	1.52E+08	1.52E+08	TCONS_00000496	ENSGALT00000027322	17897
chr1	1.59E+08	1.59E+08	TCONS_00000515	ENSGALT00000027351	27892
chr1	1.59E+08	1.59E+08	TCONS_00000516	ENSGALT00000027351	27892
chr1	1.66E+08	1.66E+08	TCONS_00001209	ENSGALT00000023466	39226
chr1	1.66E+08	1.66E+08	TCONS_00001210	ENSGALT00000023466	39226
chr1	1.66E+08	1.66E+08	TCONS_00001211	ENSGALT00000023466	1069
chr1	1.66E+08	1.66E+08	TCONS_00000536	ENSGALT00000023466	2429
chr1	1.66E+08	1.66E+08	TCONS_00000537	ENSGALT00000023466	10536
chr1	1.67E+08	1.67E+08	TCONS_00001218	ENSGALT00000027377	300744
chr1	1.71E+08	1.71E+08	TCONS_00001235	ENSGALT00000027421	4490
chr1	1.74E+08	1.74E+08	TCONS_00000564	ENSGALT00000030136	12339
chr1	1.74E+08	1.74E+08	TCONS_00001242	ENSGALT00000030136	16432
chr1	1.74E+08	1.74E+08	TCONS_00001243	ENSGALT00000030136	21585
chr1	1.74E+08	1.74E+08	TCONS_00000566	ENSGALT00000030134	2662
chr1	1.74E+08	1.74E+08	TCONS_00000567	ENSGALT00000030134	17852
chr1	1.74E+08	1.74E+08	TCONS_00000568	ENSGALT00000030134	17852
chr1	1.74E+08	1.74E+08	TCONS_00000569	ENSGALT00000030134	17852
chr1	1.74E+08	1.74E+08	TCONS_00001245	ENSGALT00000030134	52652
chr1	1.77E+08	1.77E+08	TCONS_00000581	ENSGALT00000027550	1328
chr1	1.79E+08	1.79E+08	TCONS_00000587	ENSGALT00000027590	16602
chr1	1.79E+08	1.79E+08	TCONS_00001264	ENSGALT00000027595	21681
chr1	1.83E+08	1.83E+08	TCONS_00000602	ENSGALT00000027677	566
chr1	1.85E+08	1.85E+08	TCONS_00001284	ENSGALT00000027745	3088
chr1	1.87E+08	1.87E+08	TCONS_00000612	ENSGALT00000027766	4461
chr1	1.87E+08	1.87E+08	TCONS_00001293	ENSGALT00000027770	3125
chr1	1.87E+08	1.87E+08	TCONS_00001295	ENSGALT00000027770	191
chr1	1.87E+08	1.87E+08	TCONS_00001296	ENSGALT00000027770	191
chr1	1.87E+08	1.87E+08	TCONS_00001297	ENSGALT00000036427	713

chr1	1.87E+08	1.87E+08	TCONS_00001298	ENSGALT00000036427	456
chr1	1.87E+08	1.87E+08	TCONS_00001300	ENSGALT00000036426	150
chr1	1.87E+08	1.87E+08	TCONS_00000622	ENSGALT00000036426	6298
chr1	1.9E+08	1.9E+08	TCONS_00001311	ENSGALT00000027810	20500
chr1	1.9E+08	1.9E+08	TCONS_00001317	ENSGALT00000027842	40089
chr1	1.92E+08	1.92E+08	TCONS_00000647	ENSGALT00000030060	4032
chr1	1.93E+08	1.94E+08	TCONS_00001328	ENSGALT00000027884	207
chr1	1.98E+08	1.98E+08	TCONS_00000669	ENSGALT00000017334	4991
chr1	1.98E+08	1.98E+08	TCONS_00000673	ENSGALT00000002458	1277
chr1	1.98E+08	1.98E+08	TCONS_00001346	ENSGALT00000031372	3695
chr1	1.99E+08	1.99E+08	TCONS_00001351	ENSGALT00000036390	8940
chr2	650115	670852	TCONS_00025910	ENSGALT00000008096	496
chr2	1730034	1733888	TCONS_00025914	ENSGALT00000008443	153
chr2	2250390	2253304	TCONS_00025436	ENSGALT00000008560	1269
chr2	3834105	3835425	TCONS_00025928	ENSGALT00000008839	5572
chr2	3837593	3841542	TCONS_00025440	ENSGALT00000008851	8076
chr2	5152421	5159595	TCONS_00025938	ENSGALT00000009704	2747
chr2	5159781	5161780	TCONS_00025444	ENSGALT00000009704	562
chr2	6710529	6712015	TCONS_00025944	ENSGALT00000010141	699
chr2	8597259	8599685	TCONS_00025946	ENSGALT00000010436	53
chr2	14405081	14413862	TCONS_00025966	ENSGALT00000011752	357
chr2	16143155	16149470	TCONS_00025970	ENSGALT00000031243	4164
chr2	17952125	17961234	TCONS_00025976	ENSGALT00000038352	1234
chr2	24762142	24769241	TCONS_00025485	ENSGALT00000017404	139
chr2	26571865	26575272	TCONS_00025995	ENSGALT00000017444	58
chr2	28294272	28312705	TCONS_00026001	ENSGALT00000017557	20375
chr2	32520927	32523266	TCONS_00025501	ENSGALT00000018005	1486
chr2	32613780	32619836	TCONS_00025502	ENSGALT00000037816	1643
chr2	32615275	32619836	TCONS_00025503	ENSGALT00000037816	1643
chr2	32615462	32619836	TCONS_00025504	ENSGALT00000037816	1643
chr2	32618571	32619836	TCONS_00025505	ENSGALT00000037816	1643
chr2	32632797	32637075	TCONS_00025507	ENSGALT00000037816	10896
chr2	32632830	32637075	TCONS_00025508	ENSGALT00000037816	10929
chr2	32635166	32637075	TCONS_00025511	ENSGALT00000037816	13265
chr2	32635445	32636805	TCONS_00026016	ENSGALT00000037816	13544
chr2	35231115	35236359	TCONS_00025519	NM_001199644	4335
chr2	40446683	40455119	TCONS_00025530	ENSGALT00000018727	1237
chr2	40451868	40455119	TCONS_00025531	ENSGALT00000018727	1237
chr2	41163519	41170420	TCONS_00026030	ENSGALT00000018781	4072
chr2	44403267	44405452	TCONS_00026044	ENSGALT00000019499	473
chr2	46092999	46101147	TCONS_00025547	ENSGALT00000019710	332
chr2	48671521	48674095	TCONS_00025552	ENSGALT00000019978	22935
chr2	48671634	48676002	TCONS_00025553	ENSGALT00000019978	21028
chr2	48673471	48676661	TCONS_00025554	ENSGALT00000019978	20369
chr2	50687363	50691235	TCONS_00025556	ENSGALT00000037439	7859

chr2	50829057	50831384	TCONS_00025559	ENSGALT00000037437	656
chr2	51776706	51778129	TCONS_00025562	ENSGALT00000020179	10458
chr2	51776706	51780077	TCONS_00025563	ENSGALT00000020179	8510
chr2	51776706	51782486	TCONS_00025564	ENSGALT00000020179	6101
chr2	51959446	51961300	TCONS_00025565	ENSGALT00000037413	1353
chr2	55292379	55300787	TCONS_00026068	ENSGALT00000020233	1835
chr2	57070408	57071058	TCONS_00025572	ENSGALT00000020598	427
chr2	58058189	58060619	TCONS_00025574	ENSGALT00000031087	8
chr2	59733769	59736398	TCONS_00026084	ENSGALT00000020680	3043
chr2	59736952	59741209	TCONS_00025578	ENSGALT00000020680	6226
chr2	59736952	59759489	TCONS_00026086	ENSGALT00000020680	6226
chr2	59741434	59756406	TCONS_00025584	ENSGALT00000020680	10708
chr2	59752721	59754715	TCONS_00025585	ENSGALT00000020680	21995
chr2	59760691	59766435	TCONS_00025586	ENSGALT00000020680	29965
chr2	59760691	59770789	TCONS_00025587	ENSGALT00000020680	29965
chr2	59760691	59770789	TCONS_00025588	ENSGALT00000020680	29965
chr2	59760691	59771946	TCONS_00025589	ENSGALT00000020680	29965
chr2	59760691	59772665	TCONS_00025590	ENSGALT00000020680	29965
chr2	59822587	59829704	TCONS_00026107	ENSGALT00000020688	68307
chr2	61337400	61339198	TCONS_00026109	ENSGALT00000020693	13791
chr2	62696447	62697307	TCONS_00026119	ENSGALT00000020748	3312
chr2	62763554	62780100	TCONS_00026120	ENSGALT00000020748	53393
chr2	62764834	62780100	TCONS_00026121	ENSGALT00000020748	54673
chr2	62777008	62794743	TCONS_00026122	ENSGALT00000020748	66847
chr2	62839306	62843431	TCONS_00026123	ENSGALT00000020748	129145
chr2	62897214	62899267	TCONS_00026124	ENSGALT00000020752	154499
chr2	62897214	62902687	TCONS_00026125	ENSGALT00000020752	151079
chr2	63963850	63966556	TCONS_00026130	ENSGALT00000020793	64086
chr2	64170206	64177290	TCONS_00025609	ENSGALT00000031049	5839
chr2	64170206	64177290	TCONS_00025610	ENSGALT00000031049	5839
chr2	66884971	66886280	TCONS_00026141	ENSGALT00000020896	47544
chr2	66884971	66887475	TCONS_00026142	ENSGALT00000020896	47544
chr2	67164181	67182733	TCONS_00025618	ENSGALT00000037220	72
chr2	67658522	67661665	TCONS_00025619	ENSGALT00000020927	5341
chr2	67658522	67666198	TCONS_00026146	ENSGALT00000020927	808
chr2	67684498	67687468	TCONS_00026147	ENSGALT00000020927	4557
chr2	67691140	67692679	TCONS_00026148	ENSGALT00000020927	11199
chr2	68062237	68062840	TCONS_00026152	NM_205006	135729
chr2	68066590	68069957	TCONS_00026153	NM_205006	128612
chr2	68074472	68077233	TCONS_00026155	NM_205006	121336
chr2	68076160	68077901	TCONS_00025626	NM_205006	120668
chr2	68076160	68077901	TCONS_00025627	NM_205006	120668
chr2	71968471	71988282	TCONS_00026162	ENSGALT00000021079	569085
chr2	81121552	81124645	TCONS_00025650	ENSGALT00000021302	4350
chr2	86000312	86003532	TCONS_00026182	ENSGALT00000021443	78282

chr2	86045503	86053186	TCONS_00026188	ENSGALT00000021443	28628
chr2	86483711	86486004	TCONS_00025660	ENSGALT00000021448	2667
chr2	91708453	91712199	TCONS_00026200	ENSGALT00000022087	117
chr2	92861507	92870884	TCONS_00025683	ENSGALT00000022185	30863
chr2	92881155	92881969	TCONS_00025685	ENSGALT00000022185	19778
chr2	93121176	93135408	TCONS_00025687	ENSGALT00000022218	25991
chr2	93742746	93743440	TCONS_00026214	ENSGALT00000036936	1334
chr2	1.01E+08	1.01E+08	TCONS_00026226	ENSGALT00000022620	12543
chr2	1.04E+08	1.04E+08	TCONS_00025712	ENSGALT00000030933	92
chr2	1.06E+08	1.06E+08	TCONS_00025725	ENSGALT00000024173	126
chr2	1.06E+08	1.06E+08	TCONS_00025726	ENSGALT00000024173	661
chr2	1.1E+08	1.1E+08	TCONS_00025736	ENSGALT00000024473	16139
chr2	1.17E+08	1.17E+08	TCONS_00025763	ENSGALT00000036736	3750
chr2	1.19E+08	1.19E+08	TCONS_00025769	ENSGALT00000025011	441
chr2	1.19E+08	1.19E+08	TCONS_00025770	ENSGALT00000025011	441
chr2	1.21E+08	1.21E+08	TCONS_00026266	ENSGALT00000030885	6246
chr2	1.23E+08	1.23E+08	TCONS_00025775	ENSGALT00000025212	288
chr2	1.24E+08	1.24E+08	TCONS_00026270	ENSGALT00000025271	2570
chr2	1.25E+08	1.25E+08	TCONS_00026278	ENSGALT00000036659	18024
chr2	1.25E+08	1.25E+08	TCONS_00025790	ENSGALT00000029878	41530
chr2	1.26E+08	1.26E+08	TCONS_00025791	ENSGALT00000025396	3135
chr2	1.26E+08	1.26E+08	TCONS_00026279	ENSGALT00000025409	3490
chr2	1.29E+08	1.29E+08	TCONS_00025798	ENSGALT00000025642	2499
chr2	1.3E+08	1.3E+08	TCONS_00025802	ENSGALT00000030841	83
chr2	1.32E+08	1.32E+08	TCONS_00026302	ENSGALT00000036620	1103
chr2	1.34E+08	1.34E+08	TCONS_00025813	ENSGALT00000030825	34982
chr2	1.34E+08	1.34E+08	TCONS_00026306	ENSGALT00000030825	45093
chr2	1.4E+08	1.4E+08	TCONS_00026318	ENSGALT00000025974	1362
chr2	1.4E+08	1.4E+08	TCONS_00026319	ENSGALT00000025974	2007
chr2	1.42E+08	1.42E+08	TCONS_00025828	ENSGALT00000025965	17994
chr2	1.42E+08	1.42E+08	TCONS_00026325	NM_001012555	3003
chr2	1.44E+08	1.44E+08	TCONS_00026329	ENSGALT00000026354	1565
chr2	1.45E+08	1.45E+08	TCONS_00025852	ENSGALT00000026309	20016
chr2	1.46E+08	1.46E+08	TCONS_00026339	ENSGALT00000026300	2292
chr2	1.46E+08	1.46E+08	TCONS_00026340	ENSGALT00000026300	11296
chr2	1.52E+08	1.52E+08	TCONS_00025886	ENSGALT00000026043	3462
chr2	1.52E+08	1.52E+08	TCONS_00026361	ENSGALT00000026043	29624
chr2	1.53E+08	1.53E+08	TCONS_00026366	ENSGALT00000026040	167361
chr3	1145772	1151986	TCONS_00042114	ENSGALT00000012786	34259
chr3	1509781	1512185	TCONS_00042117	ENSGALT00000012766	311766
chr3	2415715	2428553	TCONS_00042123	ENSGALT00000040236	2574
chr3	3059806	3080923	TCONS_00042127	ENSGALT00000013435	2053
chr3	3066226	3075987	TCONS_00041724	ENSGALT00000013435	6989
chr3	3070225	3080926	TCONS_00042129	ENSGALT00000013435	2050
chr3	3071206	3080910	TCONS_00042130	ENSGALT00000013435	2066

chr3	5342942	5346740	TCONS_00042144	ENSGALT00000031981	40839
chr3	5582070	5583481	TCONS_00042146	ENSGALT00000031980	514
chr3	5597839	5618047	TCONS_00041747	ENSGALT00000014211	8156
chr3	8765318	8777314	TCONS_00041753	ENSGALT00000014542	31462
chr3	8765574	8777314	TCONS_00041754	ENSGALT00000014542	31718
chr3	8765574	8777314	TCONS_00042154	ENSGALT00000014542	31718
chr3	8860124	8862563	TCONS_00042156	ENSGALT00000014538	102
chr3	9803452	9806581	TCONS_00042160	ENSGALT00000014322	33641
chr3	10016675	10020651	TCONS_00041757	ENSGALT00000014275	26935
chr3	10561438	10566873	TCONS_00042168	ENSGALT00000039449	19518
chr3	10622154	10624228	TCONS_00042174	ENSGALT00000039449	80234
chr3	13390431	13392596	TCONS_00041766	ENSGALT00000033195	19
chr3	13437721	13440268	TCONS_00041768	ENSGALT00000039344	8258
chr3	17542693	17544765	TCONS_00041773	ENSGALT00000023376	355
chr3	20474227	20477082	TCONS_00041782	ENSGALT00000031945	12
chr3	23461383	23464244	TCONS_00041789	ENSGALT00000016046	1164
chr3	23465315	23472346	TCONS_00042197	ENSGALT00000016046	5096
chr3	23650800	23656859	TCONS_00042200	ENSGALT00000016057	591
chr3	24016216	24017377	TCONS_00041792	ENSGALT00000016090	1190
chr3	24262105	24264976	TCONS_00042204	ENSGALT00000016119	249
chr3	24262105	24264976	TCONS_00042205	ENSGALT00000016119	249
chr3	26666608	26669143	TCONS_00042214	NM_001031052	1484
chr3	28583614	28584802	TCONS_00041799	ENSGALT00000016303	3911
chr3	28854924	28856452	TCONS_00042217	ENSGALT00000031906	1652
chr3	28872292	28884654	TCONS_00042218	ENSGALT00000031906	19020
chr3	31390427	31393612	TCONS_00042226	ENSGALT00000016501	150
chr3	33528730	33530748	TCONS_00041821	ENSGALT00000037966	243
chr3	34838070	34840788	TCONS_00041826	ENSGALT00000017274	638
chr3	34838074	34840788	TCONS_00041827	ENSGALT00000017274	638
chr3	35125694	35127392	TCONS_00042235	ENSGALT00000017313	11088
chr3	35181842	35186096	TCONS_00042237	ENSGALT00000017328	126
chr3	42471892	42484978	TCONS_00042250	ENSGALT00000031842	485
chr3	46113142	46115898	TCONS_00042260	ENSGALT00000018844	8077
chr3	49854076	49865662	TCONS_00042270	ENSGALT00000037597	54
chr3	51041416	51045587	TCONS_00042273	ENSGALT00000021168	134
chr3	51970900	51973501	TCONS_00042275	ENSGALT00000022168	3652
chr3	52183387	52184801	TCONS_00042278	ENSGALT00000037555	6380
chr3	52187920	52189122	TCONS_00041882	ENSGALT00000037555	10913
chr3	52258096	52267573	TCONS_00042281	ENSGALT00000022182	709
chr3	52258096	52272891	TCONS_00042282	ENSGALT00000022182	709
chr3	53214997	53217394	TCONS_00042287	ENSGALT00000022248	55647
chr3	54576872	54597066	TCONS_00041893	ENSGALT00000022379	33124
chr3	54585055	54597066	TCONS_00041894	ENSGALT00000022379	41307
chr3	54611657	54621059	TCONS_00041896	ENSGALT00000022399	47706
chr3	54611657	54622968	TCONS_00041897	ENSGALT00000022399	45797

chr3	54611657	54624867	TCONS_00041898	ENSGALT00000022399	43898
chr3	54611825	54628911	TCONS_00041899	ENSGALT00000022399	39854
chr3	54621331	54624646	TCONS_00042293	ENSGALT00000022399	44119
chr3	54654210	54666903	TCONS_00041901	ENSGALT00000022399	1862
chr3	56768011	56771749	TCONS_00041913	ENSGALT00000022492	572
chr3	57182189	57190815	TCONS_00041914	ENSGALT00000022595	7546
chr3	62288428	62289334	TCONS_00042319	ENSGALT00000023945	20253
chr3	63578488	63586392	TCONS_00041934	ENSGALT00000037386	24031
chr3	68706366	68709835	TCONS_00042333	ENSGALT00000024233	24913
chr3	68713994	68716094	TCONS_00042335	ENSGALT00000024233	32541
chr3	69880551	69888512	TCONS_00042337	ENSGALT00000024677	32497
chr3	69887456	69896604	TCONS_00042338	ENSGALT00000024677	39402
chr3	70041847	70045912	TCONS_00041956	ENSGALT00000024698	30
chr3	70042171	70045912	TCONS_00041957	ENSGALT00000024698	30
chr3	70603375	70612217	TCONS_00042340	ENSGALT00000024716	2507
chr3	70605961	70612217	TCONS_00041958	ENSGALT00000024716	5093
chr3	70605961	70626543	TCONS_00041959	ENSGALT00000024718	2329
chr3	70967510	70969199	TCONS_00041964	ENSGALT00000024823	9870
chr3	74222914	74226196	TCONS_00041970	ENSGALT00000024978	4077
chr3	74410153	74413204	TCONS_00041973	ENSGALT00000024980	2619
chr3	75574562	75576210	TCONS_00042359	ENSGALT00000025128	21083
chr3	77846939	77849790	TCONS_00042366	ENSGALT00000025147	12283
chr3	78298497	78310173	TCONS_00042369	ENSGALT00000025431	7429
chr3	78299542	78302204	TCONS_00041978	ENSGALT00000025431	8474
chr3	79507375	79516561	TCONS_00041985	ENSGALT00000025521	85136
chr3	79513342	79531229	TCONS_00042375	ENSGALT00000025521	70468
chr3	79518123	79532113	TCONS_00042376	ENSGALT00000025521	69584
chr3	79525805	79529764	TCONS_00042379	ENSGALT00000025521	71933
chr3	79528899	79531229	TCONS_00042380	ENSGALT00000025521	70468
chr3	80447138	80447863	TCONS_00041992	ENSGALT00000025551	2094
chr3	81194067	81196159	TCONS_00042387	ENSGALT00000025577	213
chr3	84238389	84247061	TCONS_00041999	ENSGALT00000025651	3879
chr3	84303547	84306904	TCONS_00042001	ENSGALT00000025673	800
chr3	84557387	84557876	TCONS_00042002	ENSGALT00000025682	88660
chr3	85166135	85170093	TCONS_00042005	ENSGALT00000025707	441
chr3	90724999	90727332	TCONS_00042401	ENSGALT00000026295	6973
chr3	91209928	91216979	TCONS_00042015	ENSGALT00000031634	254
chr3	93661778	93674156	TCONS_00042404	ENSGALT00000036946	1916
chr3	99361676	99364901	TCONS_00042420	ENSGALT00000026520	84
chr3	1.01E+08	1.01E+08	TCONS_00042040	ENSGALT00000036892	114891
chr3	1.01E+08	1.01E+08	TCONS_00042041	ENSGALT00000036892	114891
chr3	1.01E+08	1.01E+08	TCONS_00042042	ENSGALT00000036892	115837
chr3	1.04E+08	1.04E+08	TCONS_00042435	ENSGALT00000026577	68156
chr3	1.05E+08	1.05E+08	TCONS_00042049	ENSGALT00000026595	13407
chr3	1.07E+08	1.07E+08	TCONS_00042441	ENSGALT00000006362	2223

chr3	1.08E+08	1.08E+08	TCONS_00042445	ENSGALT00000026819	26124
chr3	1.08E+08	1.08E+08	TCONS_00042446	ENSGALT00000026819	32033
chr3	1.1E+08	1.1E+08	TCONS_00042089	ENSGALT00000026867	12177
chr3	1.1E+08	1.1E+08	TCONS_00042455	NM_001008462	82
chr3	1.13E+08	1.13E+08	TCONS_00042467	ENSGALT00000026967	913
chr3	1.13E+08	1.13E+08	TCONS_00042468	ENSGALT00000026980	3694
chr4	1356867	1359719	TCONS_00050824	ENSGALT00000006567	5031
chr4	1360249	1363380	TCONS_00050825	ENSGALT00000006567	1370
chr4	1402650	1413442	TCONS_00050830	ENSGALT00000032424	4979
chr4	1965260	1967210	TCONS_00050837	ENSGALT00000007868	2570
chr4	2150340	2150909	TCONS_00050554	ENSGALT00000008645	924
chr4	2503476	2510658	TCONS_00050843	ENSGALT00000009343	1580
chr4	5134951	5138938	TCONS_00050563	ENSGALT00000010691	5631
chr4	5134951	5143237	TCONS_00050564	ENSGALT00000010691	5631
chr4	5134951	5150290	TCONS_00050565	ENSGALT00000010721	936
chr4	5158474	5159749	TCONS_00050854	ENSGALT00000032371	962
chr4	9035147	9039516	TCONS_00050859	ENSGALT00000011407	95
chr4	9131634	9133395	TCONS_00050860	ENSGALT00000032365	25005
chr4	9538973	9541705	TCONS_00050574	ENSGALT00000011586	3254
chr4	12134975	12139398	TCONS_00050869	ENSGALT00000012370	62
chr4	12281340	12293247	TCONS_00050870	ENSGALT00000012483	193
chr4	12651311	12652108	TCONS_00050872	ENSGALT00000012638	1219
chr4	12651311	12652834	TCONS_00050873	ENSGALT00000012638	493
chr4	14003300	14004974	TCONS_00050582	ENSGALT00000013217	2572
chr4	18178821	18186531	TCONS_00050884	ENSGALT00000014861	5366
chr4	18183199	18187752	TCONS_00050885	ENSGALT00000014861	4145
chr4	18333959	18342893	TCONS_00050886	ENSGALT00000014934	235
chr4	21121372	21126748	TCONS_00050889	ENSGALT00000015036	5435
chr4	21182146	21186232	TCONS_00050603	ENSGALT00000015038	34871
chr4	21187150	21188773	TCONS_00050604	ENSGALT00000015038	39875
chr4	21356679	21375618	TCONS_00050605	ENSGALT00000015067	160
chr4	21360477	21374802	TCONS_00050892	ENSGALT00000015067	976
chr4	26363485	26377955	TCONS_00050620	ENSGALT00000015737	136
chr4	30329370	30333413	TCONS_00050911	ENSGALT00000015895	904
chr4	30833169	30836522	TCONS_00050916	ENSGALT00000015989	95
chr4	31724684	31730197	TCONS_00050921	ENSGALT00000016121	917
chr4	34476276	34484331	TCONS_00050645	ENSGALT00000038482	26161
chr4	35789457	35796220	TCONS_00050651	ENSGALT00000016708	2575
chr4	35799257	35800969	TCONS_00050652	ENSGALT00000038409	12270
chr4	36100550	36105715	TCONS_00050655	ENSGALT00000016805	4602
chr4	39010686	39012473	TCONS_00050935	ENSGALT00000017138	1317
chr4	39126077	39139196	TCONS_00050664	ENSGALT00000017142	60083
chr4	40243649	40253316	TCONS_00050669	ENSGALT00000017236	27044
chr4	40245242	40263240	TCONS_00050939	ENSGALT00000017236	28637
chr4	40247853	40251098	TCONS_00050671	ENSGALT00000017236	31248

chr4	40721298	40739186	TCONS_00050942	ENSGALT00000017266	3820
chr4	40764926	40767097	TCONS_00050673	ENSGALT00000017270	4456
chr4	47670849	47702414	TCONS_00050690	ENSGALT00000018134	22299
chr4	47885710	47886207	TCONS_00050959	ENSGALT00000018245	842
chr4	48446564	48450030	TCONS_00050960	ENSGALT00000018386	1511
chr4	51015355	51027029	TCONS_00050700	ENSGALT00000018673	137
chr4	51019578	51025937	TCONS_00050965	ENSGALT00000018673	1229
chr4	51148860	51153325	TCONS_00050701	ENSGALT00000018712	2031
chr4	51528567	51532467	TCONS_00050968	ENSGALT00000018842	7636
chr4	51535727	51540072	TCONS_00050969	ENSGALT00000018860	12660
chr4	54446172	54456717	TCONS_00050981	ENSGALT00000019307	1281
chr4	55160725	55164128	TCONS_00050709	ENSGALT00000019318	2829
chr4	56610191	56611622	TCONS_00050989	ENSGALT00000019591	2965
chr4	57521361	57534668	TCONS_00050990	ENSGALT00000019606	138664
chr4	58778642	58780410	TCONS_00050993	ENSGALT00000037680	4687
chr4	58789040	58792730	TCONS_00050994	ENSGALT00000037680	2471
chr4	58790384	58795954	TCONS_00050717	ENSGALT00000037680	3815
chr4	61504902	61511584	TCONS_00051000	ENSGALT00000019988	681
chr4	61678578	61694063	TCONS_00050732	ENSGALT00000037625	4806
chr4	61678578	61694319	TCONS_00051001	ENSGALT00000037625	4806
chr4	61678650	61684517	TCONS_00050733	ENSGALT00000037625	4878
chr4	62674111	62676405	TCONS_00051002	ENSGALT00000020111	175
chr4	62766751	62776575	TCONS_00051006	ENSGALT00000020122	176
chr4	62771149	62776524	TCONS_00051007	ENSGALT00000020122	227
chr4	62772213	62776561	TCONS_00051008	ENSGALT00000020122	190
chr4	64247721	64252121	TCONS_00050735	ENSGALT00000032114	10735
chr4	66979639	66987181	TCONS_00051016	ENSGALT00000022394	117
chr4	70771886	70772799	TCONS_00050744	ENSGALT00000023074	39668
chr4	70913187	70926858	TCONS_00051024	ENSGALT00000023091	778
chr4	70921708	70926109	TCONS_00050745	ENSGALT00000023091	9299
chr4	71465136	71468791	TCONS_00051026	ENSGALT00000041403	1932
chr4	76039905	76045687	TCONS_00050757	ENSGALT00000023246	3286
chr4	76050560	76054030	TCONS_00050759	ENSGALT00000023251	4733
chr4	76050750	76053855	TCONS_00050760	ENSGALT00000023251	4908
chr4	78755689	78762205	TCONS_00050761	ENSGALT00000037312	5352
chr4	81416003	81416951	TCONS_00051039	ENSGALT00000032056	1194
chr4	82385390	82406822	TCONS_00051042	ENSGALT00000025035	12178
chr4	86083653	86102716	TCONS_00051050	ENSGALT00000041399	28638
chr4	86228156	86229719	TCONS_00051053	ENSGALT00000025327	1927
chr4	86304814	86318333	TCONS_00050783	ENSGALT00000025330	12604
chr4	87067563	87078330	TCONS_00050785	ENSGALT00000037161	4656
chr4	87079714	87085673	TCONS_00050787	ENSGALT00000037161	16807
chr4	87212045	87221041	TCONS_00051056	ENSGALT00000037161	149138
chr4	87703763	87705416	TCONS_00050792	ENSGALT00000025345	2567
chr4	87703763	87707019	TCONS_00050793	ENSGALT00000025345	964

chr4	93001495	93036042	TCONS_00050802	ENSGALT00000025833	60541
chr4	93009777	93023076	TCONS_00050803	ENSGALT00000025833	68823
chr4	93018334	93032662	TCONS_00050805	ENSGALT00000025833	77380
chr4	93019164	93025886	TCONS_00050806	ENSGALT00000025833	78210
chr4	93026180	93045402	TCONS_00050807	ENSGALT00000025839	76076
chr4	93026706	93046808	TCONS_00050808	ENSGALT00000025839	74670
chr4	93107323	93120607	TCONS_00050809	ENSGALT00000025839	871
chr4	93117049	93120607	TCONS_00050810	ENSGALT00000025839	871
chr5	70364	71629	TCONS_00057680	ENSGALT00000005394	1433
chr5	5711125	5717269	TCONS_00057841	ENSGALT00000019632	252
chr5	8240232	8241733	TCONS_00057845	ENSGALT00000008463	58
chr5	8683987	8688936	TCONS_00057693	ENSGALT00000008698	63
chr5	8684064	8688936	TCONS_00057694	ENSGALT00000008698	63
chr5	8841710	8864003	TCONS_00057846	ENSGALT00000008731	1086
chr5	10229531	10236861	TCONS_00057698	ENSGALT00000009233	1616
chr5	10288516	10294033	TCONS_00057699	ENSGALT00000009297	339
chr5	11403868	11408115	TCONS_00057850	ENSGALT00000009719	1767
chr5	14305557	14322361	TCONS_00057855	ENSGALT00000010455	7576
chr5	14305557	14322363	TCONS_00057856	ENSGALT00000010455	7574
chr5	14708850	14710796	TCONS_00057857	ENSGALT00000010536	728
chr5	15303482	15314888	TCONS_00057705	ENSGALT00000032700	4167
chr5	18153641	18155994	TCONS_00057877	ENSGALT00000038796	965
chr5	18298363	18298989	TCONS_00057714	ENSGALT00000012208	360
chr5	20161628	20172087	TCONS_00057716	ENSGALT00000012671	2292
chr5	20187700	20189408	TCONS_00057717	ENSGALT00000012671	1811
chr5	21228002	21231142	TCONS_00057882	ENSGALT00000012898	304
chr5	24237251	24237824	TCONS_00057885	ENSGALT00000022785	641
chr5	24394992	24398432	TCONS_00057721	ENSGALT00000038538	3070
chr5	25069034	25072232	TCONS_00057724	ENSGALT00000013183	6986
chr5	25074581	25076146	TCONS_00057725	ENSGALT00000013183	12533
chr5	26335303	26348323	TCONS_00057890	ENSGALT00000032651	16203
chr5	26705299	26720214	TCONS_00057728	ENSGALT00000032642	1788
chr5	26705308	26720214	TCONS_00057729	ENSGALT00000032642	1797
chr5	26705308	26720214	TCONS_00057730	ENSGALT00000032642	1797
chr5	28011273	28013858	TCONS_00057734	ENSGALT00000014915	210
chr5	29251758	29252422	TCONS_00057900	ENSGALT00000015283	41594
chr5	31066523	31073169	TCONS_00057907	ENSGALT00000015455	17976
chr5	31069293	31074154	TCONS_00057745	ENSGALT00000015455	20746
chr5	31070321	31071409	TCONS_00057910	ENSGALT00000015455	21774
chr5	32859708	32860928	TCONS_00057751	ENSGALT00000037920	111
chr5	32955802	32981537	TCONS_00057918	ENSGALT00000015850	366
chr5	32965115	32981537	TCONS_00057920	ENSGALT00000015850	366
chr5	34753510	34758242	TCONS_00057759	ENSGALT00000016008	903
chr5	37083002	37086421	TCONS_00057929	ENSGALT00000016236	145
chr5	38339236	38340410	TCONS_00057768	ENSGALT00000016285	1483

chr5	39058098	39070904	TCONS_00057771	ENSGALT00000016444	30910
chr5	40730733	40737776	TCONS_00057778	ENSGALT00000016811	1019
chr5	41133910	41134594	TCONS_00057939	ENSGALT00000016888	1968
chr5	41133910	41135480	TCONS_00057940	ENSGALT00000016888	1082
chr5	41272453	41280074	TCONS_00057941	ENSGALT00000016892	1388
chr5	43093465	43101340	TCONS_00057943	ENSGALT00000017186	135
chr5	43094792	43098912	TCONS_00057944	ENSGALT00000017186	2563
chr5	44633117	44634230	TCONS_00057783	ENSGALT00000017230	272
chr5	45840579	45843752	TCONS_00057951	ENSGALT00000017352	3689
chr5	49806623	49812310	TCONS_00057962	ENSGALT00000018138	460068
chr5	55416396	55420875	TCONS_00057801	ENSGALT00000037417	4345
chr5	56776769	56782278	TCONS_00057807	ENSGALT00000019409	159
chr6	9159674	9162617	TCONS_00062039	ENSGALT00000041074	8781
chr6	9829724	9835786	TCONS_00061923	ENSGALT00000004876	232
chr6	10316878	10331780	TCONS_00061927	ENSGALT00000041024	17088
chr6	11399507	11413546	TCONS_00062053	ENSGALT00000006047	38989
chr6	11418116	11418792	TCONS_00062057	ENSGALT00000006047	57598
chr6	11829424	11830379	TCONS_00062060	ENSGALT00000006533	884
chr6	13282804	13287272	TCONS_00062066	ENSGALT00000007666	1512
chr6	13397231	13408524	TCONS_00061947	ENSGALT00000007743	25577
chr6	13407756	13408524	TCONS_00062067	ENSGALT00000007743	36102
chr6	17666179	17669799	TCONS_00062075	ENSGALT00000008855	3467
chr6	18025815	18028253	TCONS_00061951	ENSGALT00000009109	195
chr6	19636401	19643808	TCONS_00062083	ENSGALT00000009968	2122
chr6	20325941	20332704	TCONS_00062090	ENSGALT00000039927	2774
chr6	22222015	22223478	TCONS_00061974	ENSGALT00000011251	11991
chr6	22675813	22676877	TCONS_00061977	ENSGALT00000011320	81070
chr6	25238379	25247458	TCONS_00062103	ENSGALT00000013503	1260
chr6	25721614	25723209	TCONS_00062107	ENSGALT00000013711	5068
chr6	30961491	30965701	TCONS_00061997	ENSGALT00000015143	7
chr6	33502264	33504335	TCONS_00062120	ENSGALT00000015798	634
chr6	33534726	33548795	TCONS_00062124	ENSGALT00000015798	10131
chr6	33546633	33548963	TCONS_00062126	ENSGALT00000015798	22038
chr6	34124624	34144329	TCONS_00062003	ENSGALT00000041355	129689
chr7	72795	75996	TCONS_00064920	ENSGALT00000003560	10043
chr7	125994	128966	TCONS_00064921	ENSGALT00000003581	6062
chr7	1135664	1138323	TCONS_00064821	ENSGALT00000004147	55
chr7	2718091	2719524	TCONS_00064927	ENSGALT00000004522	2842
chr7	3619005	3627017	TCONS_00064928	ENSGALT00000039949	652
chr7	3619005	3627447	TCONS_00064929	ENSGALT00000039949	222
chr7	5706512	5708018	TCONS_00064938	ENSGALT00000033786	3217
chr7	10803047	10804514	TCONS_00064834	ENSGALT00000012904	405
chr7	12882148	12883981	TCONS_00064946	ENSGALT00000013776	18742
chr7	12882148	12883981	TCONS_00064947	ENSGALT00000013776	18742
chr7	12986702	12993912	TCONS_00064948	ENSGALT00000013797	5242

chr7	13016240	13021107	TCONS_00064949	ENSGALT00000013797	112
chr7	13018901	13021107	TCONS_00064838	ENSGALT00000013797	2773
chr7	13336213	13349747	TCONS_00064841	ENSGALT00000039047	13593
chr7	13666588	13668201	TCONS_00064842	ENSGALT00000014015	7193
chr7	13700320	13709185	TCONS_00064952	ENSGALT00000014022	2639
chr7	15173235	15188086	TCONS_00064963	ENSGALT00000014439	1614
chr7	17991452	17996192	TCONS_00064968	ENSGALT00000022743	1351
chr7	17991452	17997677	TCONS_00064969	ENSGALT00000038663	180
chr7	18319402	18322426	TCONS_00064976	ENSGALT00000015190	12214
chr7	19543078	19553357	TCONS_00064869	ENSGALT00000015585	20709
chr7	21188652	21200547	TCONS_00064878	ENSGALT00000017839	263
chr7	21201462	21202142	TCONS_00064879	ENSGALT00000017873	1974
chr7	23466344	23467990	TCONS_00064880	ENSGALT00000038329	2589
chr7	28532987	28535203	TCONS_00064891	ENSGALT00000019123	16
chr7	28533500	28535203	TCONS_00064892	ENSGALT00000019123	16
chr7	29536330	29538332	TCONS_00064992	ENSGALT00000019286	3501
chr7	32429133	32429390	TCONS_00064996	ENSGALT00000020184	49790
chr7	34734326	34737293	TCONS_00064999	ENSGALT00000020306	255
chr7	35033280	35036206	TCONS_00064901	ENSGALT00000020339	72241
chr7	35034708	35036206	TCONS_00064902	ENSGALT00000020339	73669
chr7	37919748	37920625	TCONS_00065009	ENSGALT00000020497	11805
chr8	134403	142365	TCONS_00067882	ENSGALT00000034240	7974
chr8	251664	255987	TCONS_00067883	ENSGALT00000034240	125235
chr8	1252582	1253413	TCONS_00067884	ENSGALT00000003064	350
chr8	1515969	1517056	TCONS_00067795	ENSGALT00000028061	4836
chr8	1754594	1757436	TCONS_00067886	ENSGALT00000034139	71917
chr8	1758942	1760617	TCONS_00067887	ENSGALT00000034139	76265
chr8	1760829	1786792	TCONS_00067888	ENSGALT00000034139	78152
chr8	1798332	1807576	TCONS_00067890	ENSGALT00000034139	115655
chr8	2480428	2487154	TCONS_00067801	ENSGALT00000003516	171
chr8	2480428	2487190	TCONS_00067893	ENSGALT00000003516	171
chr8	4184470	4185107	TCONS_00067809	ENSGALT00000039986	1556
chr8	7298716	7311044	TCONS_00067824	ENSGALT00000007207	8573
chr8	15721112	15743440	TCONS_00067924	ENSGALT00000033345	21503
chr8	21257152	21265613	TCONS_00067935	ENSGALT00000016435	4850
chr8	21842390	21844716	TCONS_00067936	ENSGALT00000016635	4045
chr8	21999445	22002037	TCONS_00067937	ENSGALT00000016747	27632
chr8	24857604	24887800	TCONS_00067872	ENSGALT00000017177	4221
chr9	1589749	1591030	TCONS_00070332	ENSGALT00000001805	1058
chr9	1644669	1647285	TCONS_00070263	ENSGALT00000001834	215
chr9	2399730	2405163	TCONS_00070265	ENSGALT00000003200	4332
chr9	2508295	2509083	TCONS_00070334	ENSGALT00000003376	4976
chr9	3376022	3404146	TCONS_00070340	ENSGALT00000003448	109488
chr9	3613730	3618529	TCONS_00070343	ENSGALT00000003640	3537
chr9	3975017	3987883	TCONS_00070270	ENSGALT00000003692	196026

chr9	3976371	3993858	TCONS_00070271	ENSGALT00000003692	197380
chr9	4613624	4629045	TCONS_00070273	ENSGALT00000010800	17497
chr9	5484803	5486113	TCONS_00070275	ENSGALT00000041158	293
chr9	5708086	5708888	TCONS_00070355	ENSGALT00000010346	7636
chr9	6137812	6143445	TCONS_00070277	ENSGALT00000009279	282
chr9	6141669	6143445	TCONS_00070278	ENSGALT00000009271	1516
chr9	6159798	6167162	TCONS_00070358	ENSGALT00000009271	1852
chr9	6293650	6296522	TCONS_00070279	ENSGALT00000008903	1599
chr9	6294058	6297679	TCONS_00070362	ENSGALT00000008903	442
chr9	6332885	6339162	TCONS_00070363	ENSGALT00000033720	10
chr9	6377611	6382823	TCONS_00070364	ENSGALT00000008893	8730
chr9	6402901	6406329	TCONS_00070365	ENSGALT00000008892	8002
chr9	6595618	6596406	TCONS_00070366	ENSGALT00000008844	7375
chr9	9415797	9419335	TCONS_00070374	NM_001083920	6388
chr9	9436141	9455459	TCONS_00070375	ENSGALT00000040805	9309
chr9	9516730	9526234	TCONS_00070376	ENSGALT00000008191	269
chr9	10302435	10311491	TCONS_00070378	ENSGALT00000007951	7897
chr9	11278604	11282478	TCONS_00070296	ENSGALT00000004533	1769
chr9	13551598	13556652	TCONS_00070386	ENSGALT00000011075	105
chr9	14619913	14627579	TCONS_00070300	ENSGALT00000011686	9118
chr9	16483447	16490275	TCONS_00070402	ENSGALT00000012571	1979
chr9	16500005	16502332	TCONS_00070307	ENSGALT00000012599	16971
chr9	18769795	18774289	TCONS_00070313	ENSGALT00000014520	650
chr9	20819496	20821108	TCONS_00070406	ENSGALT00000014937	4394
chr9	21437690	21438684	TCONS_00070320	ENSGALT00000015276	929
chr9	21475368	21477155	TCONS_00070321	ENSGALT00000039303	167
chr9	23747466	23750276	TCONS_00070325	ENSGALT00000015551	134
chr9	24529219	24533999	TCONS_00070412	ENSGALT00000016761	204
chr9	24991720	25010054	TCONS_00070328	ENSGALT00000016875	1685
chr10	1711842	1715490	TCONS_00012713	ENSGALT00000034778	2696
chr10	5991546	5996797	TCONS_00012807	ENSGALT00000005967	1461
chr10	5994835	5998131	TCONS_00012808	ENSGALT00000005967	4750
chr10	6508762	6533515	TCONS_00012720	ENSGALT00000034544	142
chr10	6508762	6533515	TCONS_00012721	ENSGALT00000034544	142
chr10	6510651	6533515	TCONS_00012723	ENSGALT00000034544	142
chr10	6513683	6533515	TCONS_00012729	ENSGALT00000034544	142
chr10	6528294	6533515	TCONS_00012730	ENSGALT00000034544	142
chr10	7559496	7563789	TCONS_00012811	ENSGALT00000006501	472
chr10	11380172	11387544	TCONS_00012817	ENSGALT00000007734	383
chr10	11667107	11672666	TCONS_00012736	ENSGALT00000038883	892
chr10	11928011	11933765	TCONS_00012738	ENSGALT00000034280	5164
chr10	12221998	12226833	TCONS_00012819	ENSGALT00000009095	1661
chr10	12952189	12965579	TCONS_00012821	ENSGALT00000009722	9314
chr10	13863287	13884061	TCONS_00012824	ENSGALT00000010326	1277
chr10	13864889	13879729	TCONS_00012748	ENSGALT00000010326	5609

chr10	13872217	13883930	TCONS_00012825	ENSGALT00000010326	1408
chr10	13906833	13910039	TCONS_00012826	ENSGALT00000010330	8977
chr10	14875288	14890803	TCONS_00012755	ENSGALT00000010970	7855
chr10	18679204	18685506	TCONS_00012771	ENSGALT00000011398	1450
chr10	18679204	18685917	TCONS_00012772	ENSGALT00000011398	1039
chr10	19614067	19615547	TCONS_00012836	ENSGALT00000038458	419
chr10	21431523	21443696	TCONS_00012839	ENSGALT00000013068	1502
chr10	21431918	21446988	TCONS_00012840	ENSGALT00000013068	1897
chr10	21991189	21994789	TCONS_00012845	ENSGALT00000013365	313
chr11	1927796	1928335	TCONS_00014321	ENSGALT00000004428	68
chr11	2675673	2676604	TCONS_00014384	ENSGALT00000034387	27211
chr11	3341847	3343341	TCONS_00014325	ENSGALT00000005367	9105
chr11	3348838	3351078	TCONS_00014327	ENSGALT00000005367	16096
chr11	4652451	4671782	TCONS_00014387	NM_001185147	2358
chr11	6948162	6952832	TCONS_00014336	ENSGALT00000006090	208
chr11	6961627	6963669	TCONS_00014338	ENSGALT00000006125	548
chr11	10538428	10539531	TCONS_00014397	ENSGALT00000007444	633
chr11	10937033	10940528	TCONS_00014344	ENSGALT00000007766	1505
chr11	10959266	10971502	TCONS_00014398	ENSGALT00000007766	5375
chr11	10968378	10971502	TCONS_00014399	ENSGALT00000007766	14487
chr11	19652112	19656912	TCONS_00014362	ENSGALT00000023100	356
chr11	21890968	21892265	TCONS_00014420	ENSGALT00000005098	85
chr12	177493	191428	TCONS_00016101	ENSGALT00000002200	8253
chr12	846219	849830	TCONS_00016172	ENSGALT00000002691	101
chr12	1482909	1502319	TCONS_00016178	ENSGALT00000036292	2521
chr12	1497785	1502319	TCONS_00016180	ENSGALT00000034435	11499
chr12	1497785	1502319	TCONS_00016181	ENSGALT00000034435	11499
chr12	1573580	1576467	TCONS_00016111	ENSGALT00000034435	198
chr12	2258622	2259090	TCONS_00016182	ENSGALT00000003959	1012
chr12	4768742	4775022	TCONS_00016195	ENSGALT00000007907	13507
chr12	5889628	5905902	TCONS_00016128	ENSGALT00000008142	506
chr12	7168112	7169728	TCONS_00016145	ENSGALT00000008512	1411
chr12	11302164	11307079	TCONS_00016154	ENSGALT00000010371	1108
chr12	14013614	14033524	TCONS_00016157	ENSGALT00000011955	456
chr12	14017145	14033513	TCONS_00016221	ENSGALT00000011955	467
chr12	16997612	17000244	TCONS_00016160	ENSGALT00000012632	1794
chr12	18780160	18789520	TCONS_00016228	ENSGALT00000033297	12608
chr12	18794297	18795878	TCONS_00016230	ENSGALT00000033297	26745
chr12	19702856	19722944	TCONS_00016162	ENSGALT00000033255	6455
chr12	20321284	20332885	TCONS_00016232	ENSGALT00000013831	2832
chr12	20386809	20394719	TCONS_00016233	ENSGALT00000013856	449
chr12	20421296	20427468	TCONS_00016164	ENSGALT00000013876	8648
chr12	20427990	20435642	TCONS_00016168	ENSGALT00000013922	3050
chr13	1310513	1321887	TCONS_00018103	ENSGALT00000041268	903
chr13	1367056	1376147	TCONS_00018108	ENSGALT00000001396	2116

chr13	2230688	2232135	TCONS_00018118	ENSGALT00000003907	1429
chr13	2245138	2248239	TCONS_00018054	ENSGALT00000003907	15879
chr13	7557919	7572251	TCONS_00018061	ENSGALT00000002291	5731
chr13	8231777	8235956	TCONS_00018129	ENSGALT00000023522	19571
chr13	8543235	8544747	TCONS_00018133	ENSGALT00000001881	6925
chr13	10448034	10457529	TCONS_00018144	ENSGALT00000005272	1115
chr13	14003987	14016388	TCONS_00018081	ENSGALT00000009699	3401
chr13	18625809	18631978	TCONS_00018152	ENSGALT00000039739	1910
chr13	18636050	18645593	TCONS_00018153	ENSGALT00000012278	3939
chr13	18638493	18640630	TCONS_00018089	ENSGALT00000012278	8902
chr13	18646698	18648265	TCONS_00018156	ENSGALT00000012278	1267
chr13	18704459	18720203	TCONS_00018091	ENSGALT00000012325	409
chr13	18715167	18718255	TCONS_00018157	ENSGALT00000012325	2357
chr13	18716156	18719128	TCONS_00018158	ENSGALT00000012325	1484
chr14	920041	921833	TCONS_00019902	ENSGALT00000005095	1260
chr14	1836226	1847194	TCONS_00019903	ENSGALT00000041045	26850
chr14	1851311	1859137	TCONS_00019904	ENSGALT00000041045	41935
chr14	2328273	2333286	TCONS_00019844	ENSGALT00000006456	4305
chr14	2328385	2329049	TCONS_00019845	ENSGALT00000006456	8542
chr14	2377333	2378837	TCONS_00019905	ENSGALT00000004936	3045
chr14	2377333	2378872	TCONS_00019906	ENSGALT00000004936	3010
chr14	2809593	2812461	TCONS_00019909	ENSGALT00000006722	41303
chr14	4012047	4018401	TCONS_00019915	ENSGALT00000007098	140
chr14	4157780	4158339	TCONS_00019917	ENSGALT00000007337	6159
chr14	4575466	4588254	TCONS_00019919	ENSGALT00000033954	2227
chr14	5950077	5956646	TCONS_00019858	ENSGALT00000008468	6286
chr14	5950077	5959219	TCONS_00019924	ENSGALT00000008468	6286
chr14	6453130	6454700	TCONS_00019925	ENSGALT00000009263	383
chr14	6683283	6690867	TCONS_00019862	ENSGALT00000009629	2669
chr14	6685727	6701824	TCONS_00019863	ENSGALT00000009629	5113
chr14	6686115	6690867	TCONS_00019864	ENSGALT00000009629	5501
chr14	6687351	6704449	TCONS_00019927	ENSGALT00000009629	6737
chr14	7447306	7451112	TCONS_00019868	ENSGALT00000010366	1989
chr14	7570506	7573419	TCONS_00019869	ENSGALT00000010426	107
chr14	13446425	13455617	TCONS_00019877	ENSGALT00000012703	2507
chr14	13446929	13455802	TCONS_00019940	ENSGALT00000012703	3011
chr14	15453425	15470665	TCONS_00019955	ENSGALT00000003152	649
chr14	15816162	15819426	TCONS_00019888	ENSGALT00000003923	737
chr14	15818323	15819426	TCONS_00019889	ENSGALT00000003923	2898
chr15	570621	572469	TCONS_00021283	ENSGALT00000034860	4012
chr15	1806844	1816138	TCONS_00021332	ENSGALT00000034777	118182
chr15	4474082	4476078	TCONS_00021337	ENSGALT00000004577	124
chr15	4734977	4739610	TCONS_00021296	ENSGALT00000004933	9227
chr15	5293126	5295951	TCONS_00021339	ENSGALT00000039382	641
chr15	7155082	7174074	TCONS_00021301	ENSGALT00000034183	2437

chr15	8213883	8214484	TCONS_00021347	ENSGALT00000034014	98
chr15	8527787	8529387	TCONS_00021304	ENSGALT00000010319	729
chr15	8721638	8722275	TCONS_00021349	ENSGALT00000010703	12172
chr15	9087441	9092348	TCONS_00021305	ENSGALT00000038413	54
chr15	9191277	9196288	TCONS_00021353	ENSGALT00000011126	3068
chr15	9383075	9385111	TCONS_00021355	ENSGALT00000011307	350
chr15	10086167	10089013	TCONS_00021358	ENSGALT00000011968	255
chr15	10364718	10365716	TCONS_00021309	ENSGALT00000012531	170
chr15	10364718	10368510	TCONS_00021310	ENSGALT00000012531	170
chr15	10364976	10368510	TCONS_00021312	ENSGALT00000012531	428
chr15	10364976	10368510	TCONS_00021313	ENSGALT00000012531	428
chr15	11713252	11714888	TCONS_00021315	ENSGALT00000013375	4972
chr17	1445000	1450771	TCONS_00022620	ENSGALT00000014440	973
chr17	6122262	6129522	TCONS_00022588	ENSGALT00000034679	8050
chr17	6122293	6125170	TCONS_00022589	ENSGALT00000034679	12402
chr17	6122293	6129522	TCONS_00022591	ENSGALT00000034679	8050
chr17	6184032	6195959	TCONS_00022593	ENSGALT00000006983	5232
chr17	6184034	6197572	TCONS_00022594	ENSGALT00000006983	3619
chr17	6186536	6195959	TCONS_00022595	ENSGALT00000006983	5232
chr17	6186843	6195959	TCONS_00022596	ENSGALT00000006983	5232
chr17	6187228	6187866	TCONS_00022637	ENSGALT00000006983	13325
chr17	6191900	6197572	TCONS_00022597	ENSGALT00000006983	3619
chr17	6192178	6195882	TCONS_00022638	ENSGALT00000006983	5309
chr17	6774231	6775650	TCONS_00022639	ENSGALT00000006004	348
chr17	6786437	6801268	TCONS_00022599	ENSGALT00000006004	1742
chr17	7078678	7095872	TCONS_00022640	ENSGALT00000005827	17286
chr17	7613535	7616528	TCONS_00022602	ENSGALT00000021564	559
chr17	7615184	7619121	TCONS_00022603	ENSGALT00000021564	2208
chr17	7832890	7836007	TCONS_00022604	ENSGALT00000039109	2359
chr17	9602582	9609783	TCONS_00022614	ENSGALT00000034915	2723
chr18	7761	20910	TCONS_00023520	ENSGALT00000023010	1228
chr18	7812	8770	TCONS_00023521	ENSGALT00000023010	13368
chr18	644713	646232	TCONS_00023477	ENSGALT00000001464	907
chr18	3378044	3381699	TCONS_00023533	ENSGALT00000002374	76
chr18	3715654	3734338	TCONS_00023491	ENSGALT00000002493	27249
chr18	3956622	3970707	TCONS_00023535	ENSGALT00000002646	36634
chr18	3963178	3983216	TCONS_00023493	ENSGALT00000002646	24125
chr18	3968559	3969644	TCONS_00023494	ENSGALT00000002646	37697
chr18	4041721	4046041	TCONS_00023538	ENSGALT00000039189	213
chr18	4623576	4626564	TCONS_00023543	ENSGALT00000003475	3594
chr18	4829326	4830987	TCONS_00023547	ENSGALT00000038959	2652
chr18	6849506	6855960	TCONS_00023553	ENSGALT00000038687	6190
chr18	6888498	6889065	TCONS_00023502	ENSGALT00000034655	527
chr18	8119331	8120860	TCONS_00023556	ENSGALT00000006974	68224
chr18	8212013	8215184	TCONS_00023506	ENSGALT00000006978	119

chr18	9537130	9537546	TCONS_00023564	ENSGALT00000011353	307
chr18	10649805	10651510	TCONS_00023569	ENSGALT00000012947	1179
chr19	5781010	5781909	TCONS_00024632	ENSGALT00000006350	1713
chr19	6821638	6822378	TCONS_00024639	ENSGALT00000039391	8085
chr19	7578466	7579454	TCONS_00024673	ENSGALT00000039210	1119
chr19	7947171	7949928	TCONS_00024641	ENSGALT00000008517	197
chr19	8157726	8172303	TCONS_00024643	ENSGALT00000008680	5369
chr19	8173536	8180083	TCONS_00024644	ENSGALT00000034733	5923
chr19	8954217	8962517	TCONS_00024646	ENSGALT00000008953	1898
chr19	9335461	9336244	TCONS_00024676	ENSGALT00000009214	3329
chr19	9922959	9925991	TCONS_00024649	ENSGALT00000041385	8877
chr20	241104	247066	TCONS_00036025	ENSGALT00000040660	21304
chr20	2102460	2106088	TCONS_00035954	ENSGALT00000003268	3415
chr20	2104475	2105946	TCONS_00036034	ENSGALT00000003268	3557
chr20	3545957	3563907	TCONS_00035963	ENSGALT00000034569	33770
chr20	3580141	3582986	TCONS_00036038	ENSGALT00000034569	14691
chr20	3589315	3592965	TCONS_00036040	ENSGALT00000034569	4712
chr20	3736351	3748907	TCONS_00036043	ENSGALT00000005634	51
chr20	4850480	4853814	TCONS_00035973	ENSGALT00000005832	1976
chr20	4859531	4862237	TCONS_00035974	ENSGALT00000023262	1771
chr20	4882129	4887544	TCONS_00035975	ENSGALT00000023262	362
chr20	5049616	5055116	TCONS_00036049	ENSGALT00000040069	951
chr20	5343149	5348710	TCONS_00035981	ENSGALT00000034377	7064
chr20	5345263	5346938	TCONS_00035982	ENSGALT00000034377	9178
chr20	5438057	5440447	TCONS_00036055	ENSGALT00000006853	211
chr20	7711772	7713653	TCONS_00035988	ENSGALT00000008238	1367
chr20	8197924	8207518	TCONS_00036063	NM_001245982	4475
chr20	8446542	8450937	TCONS_00036064	ENSGALT00000009160	2257
chr20	9595882	9596362	TCONS_00036002	ENSGALT00000009828	1894
chr20	9597606	9597897	TCONS_00036069	ENSGALT00000009828	359
chr20	9704689	9712416	TCONS_00036071	ENSGALT00000009906	1147
chr20	9707388	9709495	TCONS_00036072	ENSGALT00000009906	4068
chr20	10167611	10169008	TCONS_00036012	ENSGALT00000010709	20954
chr20	10697665	10699259	TCONS_00036079	ENSGALT00000011732	1708
chr20	13301831	13311779	TCONS_00036024	ENSGALT00000012932	18860
chr20	13486403	13501069	TCONS_00036094	ENSGALT00000012969	5020
chr20_random	55449	70576	TCONS_00037665	ENSGALT00000004894	280
chr21	222734	224492	TCONS_00037740	ENSGALT00000000739	2614
chr21	2944292	2963446	TCONS_00037680	ENSGALT00000003488	3738
chr21	2944591	2957322	TCONS_00037681	ENSGALT00000003488	4037
chr21	2945282	2949879	TCONS_00037682	ENSGALT00000003488	4728
chr21	2949948	2957325	TCONS_00037747	ENSGALT00000003488	9394
chr21	2957495	2969514	TCONS_00037748	ENSGALT00000003488	16941
chr21	2958831	2961645	TCONS_00037686	ENSGALT00000003488	18277
chr21	5521198	5528204	TCONS_00037698	ENSGALT00000006827	2002

chr21	5526730	5528204	TCONS_00037699	ENSGALT00000006827	2002
chr21	6518418	6519451	TCONS_00037708	ENSGALT00000007645	464
chr21	6530494	6532365	TCONS_00037757	ENSGALT00000034897	9617
chr21	6764668	6772056	TCONS_00037734	NM_204717	3088
chr21	6770103	6772761	TCONS_00037760	NM_204717	8523
chr21	6785981	6787425	TCONS_00037737	ENSGALT00000016394	399
chr22	434038	436396	TCONS_00038360	ENSGALT00000000273	685
chr22	2837164	2848105	TCONS_00038382	ENSGALT00000006746	1774
chr23	125	8746	TCONS_00038859	ENSGALT00000000425	2451
chr23	85617	87455	TCONS_00038862	ENSGALT00000000498	2664
chr23	712706	713368	TCONS_00038874	ENSGALT00000035354	105119
chr23	767625	769844	TCONS_00038832	ENSGALT00000035354	48643
chr23	1607834	1623735	TCONS_00038839	ENSGALT00000001194	51
chr23	2511593	2518017	TCONS_00038899	ENSGALT00000001414	727
chr23	4370832	4371643	TCONS_00038852	ENSGALT00000003361	356
chr23	4774934	4775512	TCONS_00038908	ENSGALT00000003990	3052
chr23	5354190	5358126	TCONS_00038854	ENSGALT00000005202	699
chr23	5502775	5513888	TCONS_00038855	ENSGALT00000035005	2359
chr23	5511102	5518137	TCONS_00038856	ENSGALT00000005601	4339
chr23	5576920	5578255	TCONS_00038858	ENSGALT00000005662	5354
chr24	203835	211451	TCONS_00039597	ENSGALT00000000479	4684
chr24	2905855	2913641	TCONS_00039606	ENSGALT00000010445	2190
chr24	3032974	3044306	TCONS_00039637	ENSGALT00000010518	24131
chr24	3811870	3818454	TCONS_00039608	ENSGALT00000040848	2287
chr24	3811870	3818576	TCONS_00039609	ENSGALT00000040848	2165
chr24	3815030	3818454	TCONS_00039611	ENSGALT00000040848	2287
chr24	5710537	5715708	TCONS_00039652	ENSGALT00000040594	372
chr24	5712712	5713546	TCONS_00039654	ENSGALT00000040594	2547
chr25	138079	139247	TCONS_00040330	ENSGALT00000022715	10637
chr25	1004785	1006511	TCONS_00040320	ENSGALT00000000945	1993
chr25	1350631	1357731	TCONS_00040322	NM_204121	1703
chr25	1353370	1357731	TCONS_00040323	ENSGALT00000040737	4288
chr25	1353508	1357731	TCONS_00040324	ENSGALT00000040737	4288
chr25	1696939	1700216	TCONS_00040337	ENSGALT00000029552	2870
chr25_random	71807	73419	TCONS_00040439	ENSGALT00000000457	15576
chr26	290244	291526	TCONS_00040455	ENSGALT00000000894	1161
chr26	1254736	1259887	TCONS_00040460	ENSGALT00000000599	6515
chr26	1603326	1606739	TCONS_00040462	ENSGALT00000040571	2997
chr26	2171253	2172558	TCONS_00040463	ENSGALT00000001140	3482
chr26	2332813	2336026	TCONS_00040485	ENSGALT00000001265	104
chr26	2376874	2382949	TCONS_00040486	ENSGALT00000001312	1393
chr26	2976903	2978149	TCONS_00040470	ENSGALT00000002093	299
chr26	3601256	3604510	TCONS_00040487	ENSGALT00000002953	455
chr26	3898378	3901453	TCONS_00040488	ENSGALT00000004046	10069
chr26	4017360	4023524	TCONS_00040472	ENSGALT00000004295	34

chr26	4190804	4192428	TCONS_00040489	ENSGALT00000004418	1094
chr26	4859077	4860020	TCONS_00040493	ENSGALT00000005513	8106
chr26	4875931	4876487	TCONS_00040495	ENSGALT00000005533	71
chr27	89791	105325	TCONS_00040981	ENSGALT00000000263	5905
chr27	2783251	2783971	TCONS_00040991	ENSGALT00000000832	3190
chr27	3123330	3124764	TCONS_00040994	ENSGALT00000020429	216
chr27	3398578	3401917	TCONS_00041020	ENSGALT00000040289	73
chr27	3800031	3801880	TCONS_00040999	ENSGALT00000000616	2046
chr27	4032403	4033079	TCONS_00041024	ENSGALT00000002573	2201
chr28	582965	587316	TCONS_00041355	ENSGALT00000000587	237
chr28	782535	783258	TCONS_00041357	ENSGALT00000035203	174
chr28	867021	868517	TCONS_00041398	ENSGALT00000041279	848
chr28	1002778	1004367	TCONS_00041401	ENSGALT00000001137	225
chr28	1008346	1010349	TCONS_00041402	ENSGALT00000041266	1071
chr28	1730375	1744619	TCONS_00041408	ENSGALT00000001502	2074
chr28	1730628	1747866	TCONS_00041363	ENSGALT00000001502	2327
chr28	1745573	1747866	TCONS_00041366	ENSGALT00000001527	5695
chr28	3832254	3838337	TCONS_00041382	ENSGALT00000005994	452
chr28	3835490	3838337	TCONS_00041383	ENSGALT00000005994	452
chr28	3856801	3863462	TCONS_00041419	ENSGALT00000040857	1227
chr28	4292325	4293664	TCONS_00041420	ENSGALT00000016336	2309
chrE22C19W28_E50C23	14184	16947	TCONS_00072591	ENSGALT00000023582	217
chrE22C19W28_E50C23	14230	16426	TCONS_00072608	ENSGALT00000023582	263
chrE22C19W28_E50C23	41139	41729	TCONS_00072609	ENSGALT00000009635	177
chrE22C19W28_E50C23	427410	430786	TCONS_00072601	ENSGALT00000029664	3758
chrE64	7316	8299	TCONS_00072728	ENSGALT00000021976	2236
chrE64_random	506131	515758	TCONS_00072737	ENSGALT00000041301	63139
chrUn_random	4989	11895	TCONS_00074273	ENSGALT00000036210	2824
chrUn_random	141797	146663	TCONS_00074276	ENSGALT00000036204	9502
chrUn_random	616670	620910	TCONS_00072768	ENSGALT00000007538	579
chrUn_random	673721	678944	TCONS_00074288	ENSGALT00000036198	36180
chrUn_random	1036716	1043314	TCONS_00072777	ENSGALT00000022645	34920
chrUn_random	1041627	1045309	TCONS_00074295	ENSGALT00000022645	39831
chrUn_random	1263494	1271310	TCONS_00072785	ENSGALT00000036187	31686
chrUn_random	1576124	1597241	TCONS_00072790	ENSGALT00000036183	1262
chrUn_random	1588740	1597241	TCONS_00072791	ENSGALT00000036183	1262
chrUn_random	2282878	2293902	TCONS_00072803	ENSGALT00000012864	114435
chrUn_random	2342438	2342792	TCONS_00074342	ENSGALT00000036178	160257
chrUn_random	2581977	2602151	TCONS_00072812	ENSGALT00000036177	13497
chrUn_random	2693750	2697078	TCONS_00072815	ENSGALT00000036176	24598
chrUn_random	2865368	2870049	TCONS_00072822	ENSGALT00000001886	87893
chrUn_random	3002471	3005931	TCONS_00072829	NM_204253	2465
chrUn_random	3152091	3153115	TCONS_00074374	ENSGALT00000029722	7796
chrUn_random	3166054	3166479	TCONS_00074375	ENSGALT00000029722	21759
chrUn_random	3570971	3571661	TCONS_00072839	ENSGALT00000036169	5821

chrUn_random	4204477	4205971	TCONS_00074396	ENSGALT00000017767	36299
chrUn_random	4622053	4625257	TCONS_00074408	ENSGALT00000036160	2334
chrUn_random	4847858	4848440	TCONS_00072867	ENSGALT00000036159	39723
chrUn_random	4847858	4851584	TCONS_00072868	ENSGALT00000036159	36579
chrUn_random	5373559	5377667	TCONS_00072882	ENSGALT00000022937	4134
chrUn_random	5373569	5377667	TCONS_00072883	ENSGALT00000022937	4144
chrUn_random	5548601	5549333	TCONS_00074430	ENSGALT00000013764	1577
chrUn_random	5791293	5793515	TCONS_00072891	ENSGALT00000012270	2049
chrUn_random	6240179	6249095	TCONS_00072897	ENSGALT00000036149	29863
chrUn_random	6240201	6241121	TCONS_00072898	ENSGALT00000036149	37837
chrUn_random	6477818	6479108	TCONS_00074447	NM_204271	22412
chrUn_random	6753686	6773249	TCONS_00074453	NM_001030561	12848
chrUn_random	7313564	7325651	TCONS_00074468	ENSGALT00000036138	56156
chrUn_random	7433932	7441087	TCONS_00072927	ENSGALT00000031417	48389
chrUn_random	7624712	7628859	TCONS_00074475	ENSGALT00000017450	2502
chrUn_random	7799993	7812522	TCONS_00072945	ENSGALT00000036133	58580
chrUn_random	8842383	8843694	TCONS_00072977	ENSGALT00000036119	41684
chrUn_random	9607007	9622928	TCONS_00074520	ENSGALT00000025590	52806
chrUn_random	9615571	9618976	TCONS_00074522	ENSGALT00000025590	61370
chrUn_random	9664915	9676700	TCONS_00074525	ENSGALT00000025590	110714
chrUn_random	9964090	9967227	TCONS_00074539	ENSGALT00000014280	22210
chrUn_random	10170363	10187788	TCONS_00074544	ENSGALT00000036107	6363
chrUn_random	10170368	10180038	TCONS_00073009	ENSGALT00000036107	6368
chrUn_random	10173082	10183801	TCONS_00073010	ENSGALT00000036107	9082
chrUn_random	10381414	10388019	TCONS_00073023	ENSGALT00000036105	52270
chrUn_random	10385884	10390154	TCONS_00073026	ENSGALT00000036105	50135
chrUn_random	10666610	10674777	TCONS_00074558	ENSGALT00000036103	14191
chrUn_random	10881650	10894099	TCONS_00073040	ENSGALT00000029418	23483
chrUn_random	11290057	11302477	TCONS_00073066	ENSGALT00000036095	67313
chrUn_random	11474382	11481922	TCONS_00073069	ENSGALT00000029317	4957
chrUn_random	11481313	11483661	TCONS_00073070	ENSGALT00000029317	3218
chrUn_random	11860302	11881592	TCONS_00073090	ENSGALT00000000543	39562
chrUn_random	11883658	11901137	TCONS_00073092	ENSGALT00000000543	62918
chrUn_random	11925791	11938616	TCONS_00073097	ENSGALT00000000543	105051
chrUn_random	11999739	12001606	TCONS_00074613	ENSGALT00000016183	84229
chrUn_random	12502710	12503493	TCONS_00074620	ENSGALT00000018733	23176
chrUn_random	12988439	12991616	TCONS_00073122	ENSGALT00000010130	8402
chrUn_random	13082665	13083333	TCONS_00073128	NM_001031572	16087
chrUn_random	13705722	13708454	TCONS_00074650	ENSGALT00000029278	34530
chrUn_random	13802325	13805032	TCONS_00074655	ENSGALT00000029569	94264
chrUn_random	13920903	13930737	TCONS_00074660	ENSGALT00000013991	2304
chrUn_random	14346222	14347799	TCONS_00074676	ENSGALT00000012206	2624
chrUn_random	14707133	14718445	TCONS_00074681	ENSGALT00000022296	14504
chrUn_random	14707140	14714334	TCONS_00073161	ENSGALT00000022296	18615
chrUn_random	14978507	14979235	TCONS_00074694	ENSGALT00000014078	2008

chrUn_random	15053383	15059705	TCONS_00073174	ENSGALT00000036081	22793
chrUn_random	15279608	15282387	TCONS_00073180	ENSGALT00000000937	91599
chrUn_random	16084885	16088869	TCONS_00073210	ENSGALT00000029416	2995
chrUn_random	16105811	16117749	TCONS_00074720	ENSGALT00000029416	23921
chrUn_random	16106079	16117749	TCONS_00074722	ENSGALT00000029416	24189
chrUn_random	16159382	16163237	TCONS_00073213	ENSGALT00000029416	77492
chrUn_random	16874496	16875093	TCONS_00073231	ENSGALT00000002927	2068
chrUn_random	16991459	16997734	TCONS_00073233	NM_001044680	8112
chrUn_random	17397755	17398350	TCONS_00073241	ENSGALT00000029130	1676
chrUn_random	17468475	17474762	TCONS_00074755	NM_001031591	33913
chrUn_random	18654010	18656337	TCONS_00073262	ENSGALT00000023482	33228
chrUn_random	18809438	18817114	TCONS_00073267	ENSGALT00000036036	29992
chrUn_random	18820414	18840281	TCONS_00073268	ENSGALT00000036036	6825
chrUn_random	18885310	18888219	TCONS_00074790	ENSGALT00000018763	3327
chrUn_random	20310808	20315306	TCONS_00073303	ENSGALT00000020495	91492
chrUn_random	20383053	20401810	TCONS_00074822	NM_204396	13849
chrUn_random	20394080	20401810	TCONS_00074825	NM_204396	13849
chrUn_random	20400987	20413688	TCONS_00073311	NM_204396	1971
chrUn_random	20406088	20413688	TCONS_00073312	NM_204396	1971
chrUn_random	20459923	20460587	TCONS_00073315	ENSGALT00000016176	20110
chrUn_random	20941754	20944908	TCONS_00073335	ENSGALT00000020936	153
chrUn_random	21229077	21231727	TCONS_00073341	ENSGALT00000036012	27397
chrUn_random	21268736	21283449	TCONS_00074851	ENSGALT00000036012	9355
chrUn_random	22634039	22642160	TCONS_00073378	ENSGALT00000029877	25478
chrUn_random	22818660	22823000	TCONS_00074884	ENSGALT00000018918	28306
chrUn_random	23293693	23296736	TCONS_00074891	ENSGALT00000029521	6263
chrUn_random	23453606	23458231	TCONS_00073407	ENSGALT00000035958	12496
chrUn_random	23497510	23501587	TCONS_00074897	ENSGALT00000035955	173
chrUn_random	23829114	23832968	TCONS_00073414	NM_205480	47774
chrUn_random	23937788	23949830	TCONS_00073418	NM_205480	56407
chrUn_random	24018001	24028969	TCONS_00073424	ENSGALT00000019240	2419
chrUn_random	24253219	24253736	TCONS_00074914	ENSGALT00000012064	22633
chrUn_random	24288772	24307955	TCONS_00073430	ENSGALT00000007001	1161
chrUn_random	24722230	24722624	TCONS_00074920	ENSGALT00000035947	825
chrUn_random	24924782	24925865	TCONS_00074930	ENSGALT00000033142	79978
chrUn_random	25543117	25543546	TCONS_00074941	ENSGALT00000031846	7159
chrUn_random	25604753	25620776	TCONS_00073441	ENSGALT00000021010	15559
chrUn_random	25780309	25781722	TCONS_00074943	ENSGALT00000035934	6339
chrUn_random	26851780	26864551	TCONS_00074955	ENSGALT00000028263	56709
chrUn_random	27168233	27173523	TCONS_00073462	ENSGALT00000035910	2995
chrUn_random	27180041	27185169	TCONS_00074958	ENSGALT00000035910	1729
chrUn_random	27186843	27205205	TCONS_00073463	ENSGALT00000035910	8531
chrUn_random	27493232	27497324	TCONS_00073475	ENSGALT00000022363	1142
chrUn_random	27508111	27510479	TCONS_00073476	ENSGALT00000022363	7489
chrUn_random	27947989	27958162	TCONS_00074980	ENSGALT00000021767	3261

chrUn_random	27947989	27958164	TCONS_00074981	ENSGALT00000021767	3259
chrUn_random	28208596	28214332	TCONS_00074985	ENSGALT00000022320	537
chrUn_random	29273610	29292143	TCONS_00074997	ENSGALT00000029669	8486
chrUn_random	29406051	29408593	TCONS_00074999	ENSGALT00000035869	7452
chrUn_random	29407164	29408595	TCONS_00073518	ENSGALT00000035869	8565
chrUn_random	29477954	29484821	TCONS_00073520	ENSGALT00000021977	1228
chrUn_random	29593307	29607135	TCONS_00073523	ENSGALT00000023556	47569
chrUn_random	29706691	29707290	TCONS_00073526	ENSGALT00000035865	1206
chrUn_random	30802520	30806205	TCONS_00073553	ENSGALT00000021968	1757
chrUn_random	30815952	30818678	TCONS_00075035	ENSGALT00000000134	4593
chrUn_random	31303620	31304598	TCONS_00075044	ENSGALT00000035839	19081
chrUn_random	31407610	31409145	TCONS_00073575	ENSGALT00000021700	833
chrUn_random	31470490	31471393	TCONS_00075046	ENSGALT00000022112	196
chrUn_random	32462595	32463354	TCONS_00073586	NM_204802	112634
chrUn_random	32462595	32463859	TCONS_00073587	NM_204802	112634
chrUn_random	32462595	32467446	TCONS_00073588	NM_204802	112634
chrUn_random	32760180	32781125	TCONS_00075075	ENSGALT00000035828	41825
chrUn_random	33067368	33078269	TCONS_00075077	ENSGALT00000022230	24949
chrUn_random	33904219	33917691	TCONS_00073614	ENSGALT00000035812	113406
chrUn_random	34168522	34180166	TCONS_00075092	ENSGALT00000035811	22192
chrUn_random	35168223	35169813	TCONS_00073630	ENSGALT00000022441	12683
chrUn_random	35689207	35694294	TCONS_00073635	ENSGALT00000022698	1289
chrUn_random	35690704	35694294	TCONS_00073636	ENSGALT00000022698	2786
chrUn_random	35694991	35717257	TCONS_00073638	ENSGALT00000022698	7073
chrUn_random	35718988	35722928	TCONS_00073644	ENSGALT00000022698	31070
chrUn_random	35723590	35724844	TCONS_00075118	ENSGALT00000022698	35672
chrUn_random	35814191	35819856	TCONS_00075122	ENSGALT00000022591	75845
chrUn_random	35824969	35828794	TCONS_00073648	ENSGALT00000022591	66907
chrUn_random	35861831	35862339	TCONS_00073651	ENSGALT00000022591	33362
chrUn_random	36079848	36081711	TCONS_00075136	ENSGALT00000001345	23438
chrUn_random	36346913	36365206	TCONS_00075139	ENSGALT00000008327	2300
chrUn_random	37028124	37036059	TCONS_00073668	ENSGALT00000023185	15594
chrUn_random	37789386	37801509	TCONS_00073682	NM_001001337	8022
chrUn_random	37798172	37814900	TCONS_00075170	NM_001001337	16808
chrUn_random	37961474	37993214	TCONS_00075174	ENSGALT00000004805	39737
chrUn_random	37966742	37976732	TCONS_00075175	ENSGALT00000004805	56219
chrUn_random	37967518	37976732	TCONS_00075176	ENSGALT00000004805	56219
chrUn_random	38078901	38093391	TCONS_00075179	ENSGALT00000004805	36206
chrUn_random	38243520	38246282	TCONS_00075183	ENSGALT00000032890	7365
chrUn_random	38247327	38251874	TCONS_00075184	ENSGALT00000029992	9018
chrUn_random	38247327	38256363	TCONS_00075185	ENSGALT00000029992	4529
chrUn_random	38247327	38260313	TCONS_00075186	ENSGALT00000029992	579
chrUn_random	38329275	38337626	TCONS_00073698	ENSGALT00000004807	11273
chrUn_random	38331741	38334969	TCONS_00073699	ENSGALT00000004807	13930
chrUn_random	38332256	38334969	TCONS_00073700	ENSGALT00000004807	13930

chrUn_random	38356023	38356840	TCONS_00073701	ENSGALT00000004807	278
chrUn_random	38394165	38394959	TCONS_00075191	ENSGALT00000004807	38420
chrUn_random	38519104	38519521	TCONS_00075194	ENSGALT00000023126	1111
chrUn_random	38592955	38598426	TCONS_00073703	ENSGALT00000029406	4776
chrUn_random	38606628	38626865	TCONS_00073704	ENSGALT00000035750	14618
chrUn_random	38781303	38784944	TCONS_00073706	ENSGALT00000015228	62770
chrUn_random	39814013	39817885	TCONS_00073715	ENSGALT00000012242	5887
chrUn_random	41249402	41250528	TCONS_00073732	ENSGALT00000035729	2162
chrUn_random	41340649	41345484	TCONS_00073735	ENSGALT00000035726	85895
chrUn_random	41489918	41509309	TCONS_00073741	ENSGALT00000022869	5385
chrUn_random	41606183	41608284	TCONS_00075246	NM_001012403	48560
chrUn_random	42188930	42196687	TCONS_00075258	NM_204604	18604
chrUn_random	42231868	42235006	TCONS_00073753	NM_204604	61542
chrUn_random	42613649	42617183	TCONS_00073758	ENSGALT00000029291	52088
chrUn_random	42725891	42727459	TCONS_00075272	ENSGALT00000035712	65416
chrUn_random	42837493	42857350	TCONS_00073761	ENSGALT00000029333	2869
chrUn_random	43136991	43140899	TCONS_00075284	ENSGALT00000017918	1891
chrUn_random	43138191	43140899	TCONS_00075285	ENSGALT00000017918	3091
chrUn_random	43291208	43299501	TCONS_00073767	ENSGALT00000009996	35166
chrUn_random	43317101	43324657	TCONS_00075288	ENSGALT00000009996	10010
chrUn_random	43497487	43513208	TCONS_00075291	ENSGALT00000035707	6530
chrUn_random	43794707	43797624	TCONS_00073773	ENSGALT00000035703	8035
chrUn_random	44084821	44093817	TCONS_00073778	ENSGALT00000035701	6151
chrUn_random	45597084	45600071	TCONS_00073813	ENSGALT00000023283	48327
chrUn_random	46235247	46243383	TCONS_00075333	NM_204309	11083
chrUn_random	46985301	47000853	TCONS_00073835	ENSGALT00000021236	39964
chrUn_random	47971358	47971661	TCONS_00073867	ENSGALT00000004942	23430
chrUn_random	48314551	48315460	TCONS_00075385	ENSGALT00000035654	9994
chrUn_random	48392135	48406514	TCONS_00073878	ENSGALT00000000216	10706
chrUn_random	48588529	48590088	TCONS_00073883	ENSGALT00000035652	14797
chrUn_random	48588530	48599510	TCONS_00073884	ENSGALT00000035652	14798
chrUn_random	48589726	48610585	TCONS_00075390	ENSGALT00000035652	15994
chrUn_random	49306569	49316736	TCONS_00073904	ENSGALT00000035643	3968
chrUn_random	49709326	49725025	TCONS_00075421	ENSGALT00000012539	4697
chrUn_random	49731075	49732486	TCONS_00073909	ENSGALT00000012539	26446
chrUn_random	49759981	49778143	TCONS_00073911	ENSGALT00000013374	796
chrUn_random	50383274	50393626	TCONS_00075433	ENSGALT00000035634	5043
chrUn_random	50433043	50437661	TCONS_00073921	ENSGALT00000034271	12027
chrUn_random	50520865	50525532	TCONS_00075438	ENSGALT00000035632	46613
chrUn_random	50563313	50580713	TCONS_00075440	ENSGALT00000035632	89061
chrUn_random	51457769	51459524	TCONS_00075458	ENSGALT00000035622	37945
chrUn_random	51458968	51478573	TCONS_00075459	ENSGALT00000035622	18896
chrUn_random	51492057	51493120	TCONS_00073950	ENSGALT00000035622	4349
chrUn_random	51492057	51493120	TCONS_00073951	ENSGALT00000035622	4349
chrUn_random	51574737	51584090	TCONS_00073953	ENSGALT00000035619	6328

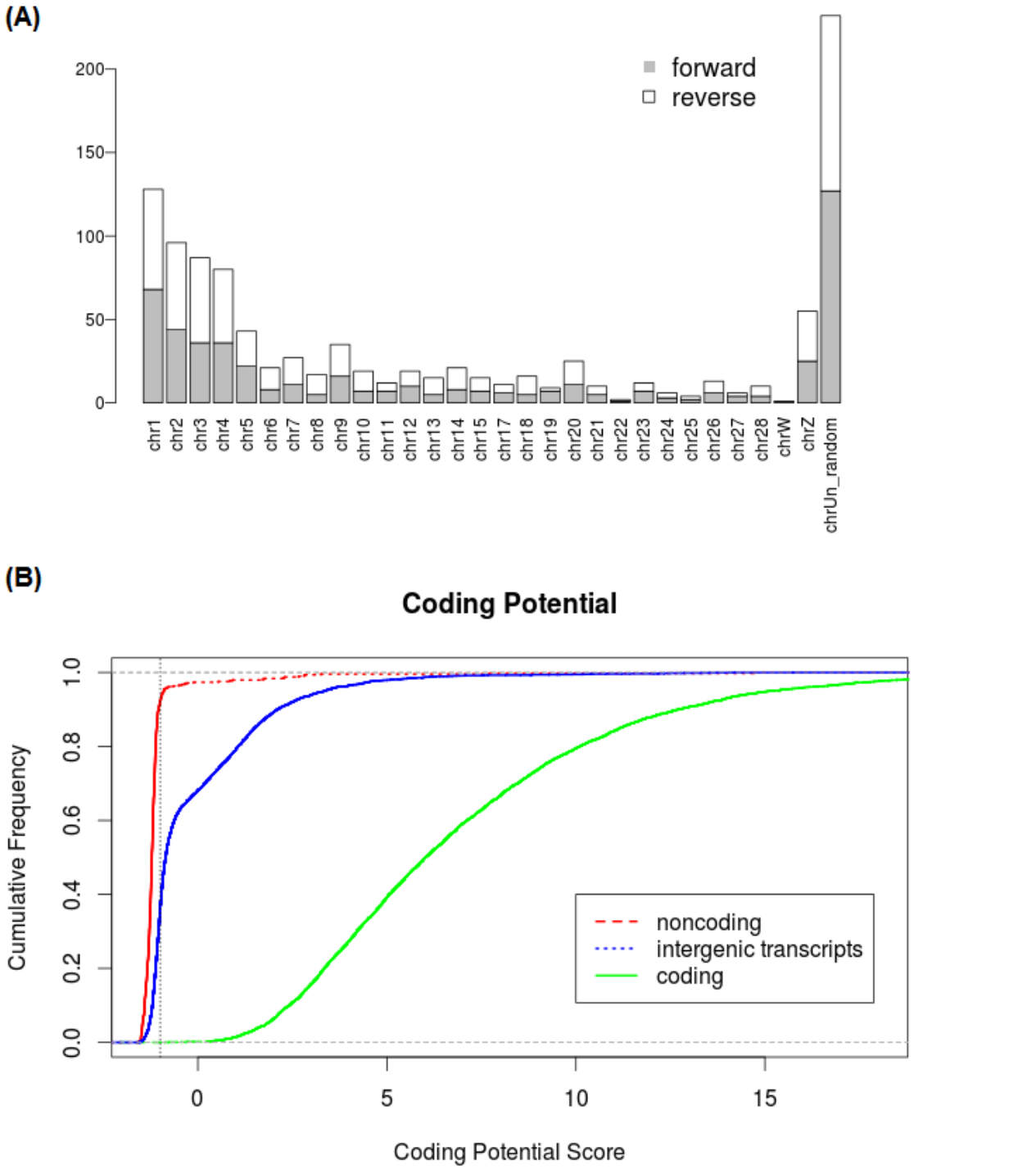
chrUn_random	51697425	51702756	TCONS_00073954	ENSGALT00000022546	2864
chrUn_random	51798325	51804736	TCONS_00073956	ENSGALT00000035617	20410
chrUn_random	52294707	52307059	TCONS_00075476	ENSGALT00000029816	39527
chrUn_random	52296848	52307061	TCONS_00073963	ENSGALT00000029816	41668
chrUn_random	52360315	52365559	TCONS_00073965	ENSGALT00000035607	73938
chrUn_random	52546696	52553484	TCONS_00075487	ENSGALT00000035607	105907
chrUn_random	52710997	52727291	TCONS_00073978	ENSGALT00000035606	43484
chrUn_random	52738832	52757189	TCONS_00073979	ENSGALT00000035606	71319
chrUn_random	53164012	53178827	TCONS_00075516	ENSGALT00000022175	63973
chrUn_random	53755512	53789648	TCONS_00075540	ENSGALT00000000021	36899
chrUn_random	53757465	53775262	TCONS_00075543	ENSGALT00000000021	38852
chrUn_random	53767770	53775671	TCONS_00075544	ENSGALT00000000021	49157
chrUn_random	53841344	53864281	TCONS_00075545	ENSGALT00000000021	122731
chrUn_random	53922422	53924061	TCONS_00074007	ENSGALT00000000051	139801
chrUn_random	54027431	54031310	TCONS_00075550	ENSGALT00000000051	32552
chrUn_random	54027431	54045548	TCONS_00075551	ENSGALT00000000051	18314
chrUn_random	54100645	54106266	TCONS_00074021	ENSGALT00000035602	6376
chrUn_random	54275339	54276834	TCONS_00074032	ENSGALT00000035598	19905
chrUn_random	54275339	54282384	TCONS_00074033	ENSGALT00000035598	19905
chrUn_random	54298250	54310333	TCONS_00075561	ENSGALT00000035598	42816
chrUn_random	54338898	54348321	TCONS_00074034	ENSGALT00000035597	29879
chrUn_random	54551238	54568246	TCONS_00075571	ENSGALT00000035592	502
chrUn_random	54554803	54568246	TCONS_00074041	ENSGALT00000035592	4067
chrUn_random	54567460	54580922	TCONS_00074043	ENSGALT00000035592	16724
chrUn_random	55084436	55098096	TCONS_00074056	ENSGALT00000021819	2717
chrUn_random	55353207	55367956	TCONS_00075606	ENSGALT00000001817	4791
chrUn_random	55358882	55364532	TCONS_00074064	ENSGALT00000001817	8215
chrUn_random	55975302	55975799	TCONS_00075619	NM_204893	13893
chrUn_random	56716054	56721032	TCONS_00075629	ENSGALT00000028318	224047
chrUn_random	57354700	57371321	TCONS_00075638	ENSGALT00000035571	64069
chrUn_random	57400722	57420643	TCONS_00074110	ENSGALT00000035571	110091
chrUn_random	57605360	57619174	TCONS_00075649	ENSGALT00000000023	7478
chrUn_random	57808710	57820159	TCONS_00075653	ENSGALT00000023449	42155
chrUn_random	58165638	58178260	TCONS_00074130	ENSGALT00000035565	9733
chrUn_random	59384438	59407780	TCONS_00075704	ENSGALT00000035550	30864
chrUn_random	59384438	59407980	TCONS_00075705	ENSGALT00000035550	30664
chrUn_random	59387187	59391150	TCONS_00074165	ENSGALT00000035550	47494
chrUn_random	59398457	59402245	TCONS_00074169	ENSGALT00000035550	36399
chrUn_random	59539955	59551439	TCONS_00075709	ENSGALT00000035545	8336
chrUn_random	59663336	59663825	TCONS_00075712	ENSGALT00000022647	17624
chrUn_random	59663336	59666458	TCONS_00075713	ENSGALT00000022647	14991
chrUn_random	60708252	60722837	TCONS_00075731	ENSGALT00000035527	46386
chrUn_random	60991504	61001503	TCONS_00075735	ENSGALT00000035523	5129
chrUn_random	61169322	61189545	TCONS_00074200	ENSGALT00000018516	5758
chrUn_random	61355576	61367876	TCONS_00075738	ENSGALT00000035520	51991

chrUn_random	61472959	61483604	TCONS_00074208	ENSGALT00000005124	1151
chrUn_random	61514718	61523087	TCONS_00074211	ENSGALT00000005124	42910
chrUn_random	61669604	61686863	TCONS_00075751	ENSGALT00000001115	35290
chrUn_random	61679684	61684608	TCONS_00075752	ENSGALT00000001115	45370
chrUn_random	61680394	61684608	TCONS_00074215	ENSGALT00000001115	46080
chrUn_random	61683045	61686305	TCONS_00074217	ENSGALT00000001115	48731
chrUn_random	61697710	61709702	TCONS_00075756	ENSGALT00000001115	63396
chrUn_random	61947671	61957268	TCONS_00074222	ENSGALT00000031740	126052
chrUn_random	61952463	61959700	TCONS_00074223	ENSGALT00000031740	123620
chrUn_random	62029142	62032357	TCONS_00074225	ENSGALT00000031740	50963
chrUn_random	62140576	62155292	TCONS_00074226	ENSGALT00000035514	27878
chrUn_random	62237210	62254861	TCONS_00074231	ENSGALT00000023064	36817
chrUn_random	62337355	62339570	TCONS_00074236	ENSGALT00000035512	1217
chrUn_random	62455655	62456938	TCONS_00075770	ENSGALT00000016529	38214
chrUn_random	62953433	62962303	TCONS_00074250	NM_204996	6828
chrUn_random	63805216	63811823	TCONS_00075795	ENSGALT00000035497	1135
chrW	224735	245860	TCONS_00078132	ENSGALT00000016349	10538
chrW_random	19053	20802	TCONS_00078149	ENSGALT00000022067	9906
chrZ	2415324	2417530	TCONS_00078458	ENSGALT00000002566	1546
chrZ	7132654	7135946	TCONS_00078209	ENSGALT00000009334	5483
chrZ	8241361	8243477	TCONS_00078471	ENSGALT00000003394	123
chrZ	8241361	8243477	TCONS_00078472	ENSGALT00000003394	123
chrZ	8554068	8556571	TCONS_00078475	ENSGALT00000039281	879
chrZ	9055650	9058643	TCONS_00078480	ENSGALT00000005018	1640
chrZ	10338689	10341335	TCONS_00078234	ENSGALT00000005609	263
chrZ	11389323	11395118	TCONS_00078240	ENSGALT00000005924	833
chrZ	13019641	13025001	TCONS_00078244	ENSGALT00000023979	397
chrZ	15881321	15897120	TCONS_00078522	ENSGALT00000032852	1380
chrZ	16319073	16327744	TCONS_00078256	ENSGALT00000023725	1700
chrZ	21534202	21542312	TCONS_00078264	ENSGALT00000023909	5980
chrZ	21539913	21542312	TCONS_00078265	ENSGALT00000023909	5980
chrZ	21541231	21541995	TCONS_00078534	ENSGALT00000023909	6297
chrZ	22388307	22388909	TCONS_00078536	ENSGALT00000006998	397
chrZ	23554814	23560213	TCONS_00078271	ENSGALT00000024108	2589
chrZ	23558691	23560364	TCONS_00078273	ENSGALT00000024108	6466
chrZ	23749235	23755971	TCONS_00078542	ENSGALT00000024082	3295
chrZ	25972608	25978977	TCONS_00078545	ENSGALT00000038490	17468
chrZ	26083684	26104412	TCONS_00078275	ENSGALT00000024201	453
chrZ	27349823	27351552	TCONS_00078279	ENSGALT00000034683	1157
chrZ	28622084	28628915	TCONS_00078289	ENSGALT00000032836	21509
chrZ	28622094	28633682	TCONS_00078290	ENSGALT00000032836	16742
chrZ	31972262	31976345	TCONS_00078297	ENSGALT00000024368	82961
chrZ	33383978	33387476	TCONS_00078302	ENSGALT00000024342	1085
chrZ	34450136	34474956	TCONS_00078563	ENSGALT00000024385	820
chrZ	36901055	36903111	TCONS_00078313	ENSGALT00000024477	63

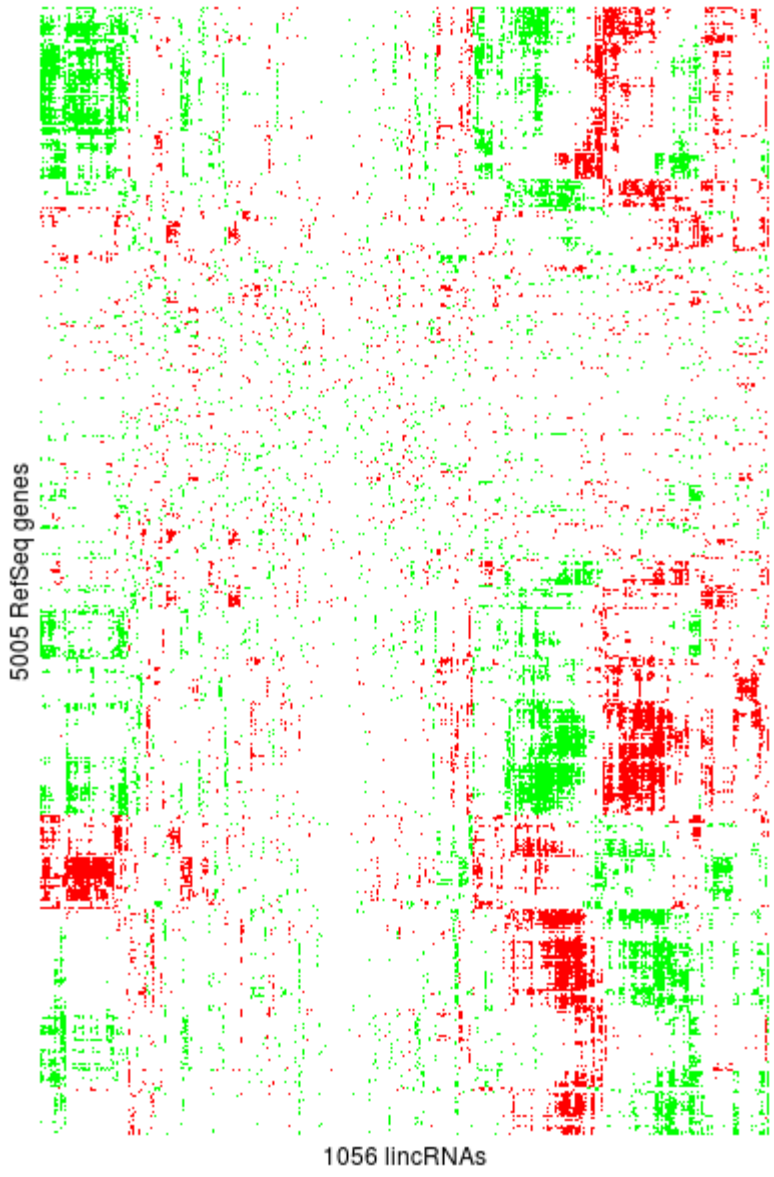
chrZ	36921933	36931747	TCONS_00078574	ENSGALT00000024478	3128
chrZ	37275900	37279006	TCONS_00078576	ENSGALT00000024491	2231
chrZ	41468118	41469035	TCONS_00078583	ENSGALT00000020603	3674
chrZ	42926651	42931298	TCONS_00078585	ENSGALT00000037939	24205
chrZ	42968344	42969716	TCONS_00078321	ENSGALT00000037939	65898
chrZ	43014751	43029608	TCONS_00078588	ENSGALT00000037939	112305
chrZ	43327554	43329120	TCONS_00078589	ENSGALT00000032807	131
chrZ	43984130	43999323	TCONS_00078331	ENSGALT00000021979	10082
chrZ	46556472	46557711	TCONS_00078337	ENSGALT00000000344	4272
chrZ	46618041	46623976	TCONS_00078338	ENSGALT00000037869	1460
chrZ	52246653	52251369	TCONS_00078349	ENSGALT00000037818	55
chrZ	52246653	52251369	TCONS_00078350	ENSGALT00000037818	55
chrZ	52246653	52251369	TCONS_00078352	ENSGALT00000037818	55
chrZ	52246653	52251369	TCONS_00078353	ENSGALT00000037818	55
chrZ	52246653	52251369	TCONS_00078354	ENSGALT00000037818	55
chrZ	52246653	52251369	TCONS_00078355	ENSGALT00000037818	55
chrZ	52521329	52523670	TCONS_00078614	ENSGALT00000024821	595
chrZ	55061088	55073788	TCONS_00078368	ENSGALT00000023680	2893
chrZ	59083703	59086944	TCONS_00078629	ENSGALT00000031494	43516
chrZ	59248235	59255032	TCONS_00078630	ENSGALT00000031494	1690
chrZ	59249362	59255032	TCONS_00078631	ENSGALT00000031494	2817
chrZ	59443581	59451502	TCONS_00078633	ENSGALT00000017720	63
chrZ	62362634	62373277	TCONS_00078642	ENSGALT00000037654	2667
chrZ	62363113	62369782	TCONS_00078643	ENSGALT00000037654	3146
chrZ	64395316	64402184	TCONS_00078649	ENSGALT00000025246	89
chrZ	65679517	65679953	TCONS_00078652	ENSGALT00000032768	2358
chrZ	66276506	66288629	TCONS_00078655	ENSGALT00000028205	182337
chrZ	66288089	66306293	TCONS_00078407	ENSGALT00000028205	164673
chrZ	68750376	68756920	TCONS_00078663	ENSGALT00000037562	22
chrZ	68754870	68769266	TCONS_00078417	ENSGALT00000037562	4516
chrZ	69734432	69736225	TCONS_00078424	ENSGALT00000003486	6443
chrZ	73772806	73777173	TCONS_00078436	ENSGALT00000008552	1837
chrZ	74293183	74307350	TCONS_00078675	ENSGALT00000008724	6140
chrZ	74342376	74345984	TCONS_00078440	ENSGALT00000026510	33727
chrZ	74342376	74346445	TCONS_00078441	ENSGALT00000026510	33266
chrZ	74348386	74349740	TCONS_00078678	ENSGALT00000026510	29971
chrZ	74348720	74354460	TCONS_00078445	ENSGALT00000026510	25251
chrZ	74351996	74354460	TCONS_00078447	ENSGALT00000026510	25251
chrZ	74351996	74361326	TCONS_00078448	ENSGALT00000026510	18385
chrZ	74370635	74379545	TCONS_00078687	ENSGALT00000026510	166
chrZ	74370635	74379545	TCONS_00078688	ENSGALT00000026510	166
chrZ	74492646	74503056	TCONS_00078689	ENSGALT00000026508	4386

Supplementary Table 3 Fold changes of immune response associated lincRNAs

lincRNAs	L63_5dpi	L63_10dpi	L63_21dpi	L72_5dpi	L72_10dpi	L72_21dpi
XLOC_068940	-0.038	2.112	0.253	-0.589	-0.758	0.406
XLOC_000789	0.059	1.577	0.000	-0.649	-0.074	0.342
XLOC_063601	0.496	2.229	-0.448	-1.070	-0.852	0.662
XLOC_068939	0.067	2.985	-0.032	-0.777	-0.032	1.413
XLOC_024972	-0.024	1.817	-0.022	-0.328	0.184	1.441
XLOC_024939	0.153	3.969	-0.205	-0.310	0.186	1.026
XLOC_049658	0.156	1.767	0.146	-0.649	0.172	1.678
XLOC_071992	0.314	1.052	0.029	-1.244	0.711	1.720
XLOC_022076	0.403	0.556	-0.361	-0.742	0.636	1.321
XLOC_022110	0.204	1.155	0.111	-2.478	0.650	1.761
XLOC_075869	0.256	0.135	0.073	-2.010	-0.830	0.846
XLOC_075809	-0.448	1.293	-1.206	-2.014	0.514	0.398
XLOC_025293	0.125	0.761	-0.129	-0.439	0.195	-0.338
XLOC_035285	0.171	0.148	0.280	-0.118	-0.631	0.164
XLOC_075868	-0.267	0.712	0.297	-1.529	0.130	0.745
XLOC_056648	0.359	1.903	0.507	-0.690	0.773	-0.203
XLOC_068949	-0.019	1.799	-0.004	-0.263	0.141	-0.180
XLOC_041314	0.422	0.905	0.791	-0.467	0.212	-0.160
XLOC_000726	0.000	1.062	0.518	-0.463	0.000	0.000
XLOC_066491	-0.009	3.031	-0.250	-0.417	-0.345	-0.011
XLOC_056530	0.031	1.072	0.025	-0.661	0.062	0.080
XLOC_015830	0.000	2.092	-0.516	-0.149	0.189	-0.190
XLOC_072297	0.857	0.783	-0.017	-0.428	0.000	1.904
XLOC_022092	-0.065	0.434	-0.131	-0.261	0.527	1.027
XLOC_049655	0.773	2.079	-1.621	-0.634	-0.283	2.763
XLOC_017708	0.989	0.624	1.874	-0.616	0.410	2.580
XLOC_019436	0.097	0.487	0.491	0.311	0.239	0.504
XLOC_025407	-0.659	0.401	-0.210	0.110	2.180	0.803
XLOC_073134	0.210	0.529	-0.581	-0.609	-0.068	0.294
XLOC_025574	0.049	-0.466	-0.379	0.449	0.968	-0.822



Supplementary Figure 1 Basic properties of lincRNAs. **(A)** LincRNA loci distribution on each chromosome and strand. 507 and 549 lincRNA loci were on forward and reverse strand, respectively. On each individual chromosome, lincRNA loci were also almost equally distributed on both strands. Roughly, the number of lincRNA loci identified on each chromosome is proportional to its size. One exception is chrUn_random, on which 232 candidate lincRNA loci were identified. **(B)** Distribution of coding potential scores as calculated by Coding Potential Calculator. For protein-coding genes (green), almost all transcripts have coding potential score greater than 1. The coding potential for our candidate lincRNAs (blue) were much lower and are comparable to annotated noncoding transcripts (red).



Supplementary Figure 2 expression correlation between lincRNAs and protein-coding genes. LincRNAs were shown as columns, and RefSeq genes were shown as rows. Each point represents the Pearson correlation coefficient between certain lincRNAs and protein-coding gene. Red means positive correlation and green means negative correlation. The matrix was clustered using hierarchical clustering with complete linkage.

References

1. Ewing B, Green P: **Analysis of expressed sequence tags indicates 35,000 human genes**. *Nature genetics* 2000, **25**(2):232-234.
2. **Finishing the euchromatic sequence of the human genome**. *Nature* 2004, **431**(7011):931-945.
3. Clamp M, Fry B, Kamal M, Xie X, Cuff J, Lin MF, Kellis M, Lindblad-Toh K, Lander ES: **Distinguishing protein-coding and noncoding genes in the human genome**. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(49):19428-19433.
4. Stolc V, Gauhar Z, Mason C, Halasz G, van Batenburg MF, Rifkin SA, Hua S, Herreman T, Tongprasit W, Barbano PE *et al*: **A gene expression map for the euchromatic genome of Drosophila melanogaster**. *Science* 2004, **306**(5696):655-660.
5. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C *et al*: **The transcriptional landscape of the mammalian genome**. *Science* 2005, **309**(5740):1559-1563.
6. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H *et al*: **Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs**. *Nature* 2002, **420**(6915):563-573.
7. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S *et al*: **Global identification of human transcribed sequences with genome tiling arrays**. *Science* 2004, **306**(5705):2242-2246.
8. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE *et al*: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project**. *Nature* 2007, **447**(7146):799-816.
9. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL *et al*: **RNA maps reveal new RNA classes and a possible function for pervasive transcription**. *Science* 2007, **316**(5830):1484-1488.
10. Rinn JL, Euskirchen G, Bertone P, Martone R, Luscombe NM, Hartman S, Harrison PM, Nelson FK, Miller P, Gerstein M *et al*: **The transcriptional activity of human Chromosome 22**. *Genes & development* 2003, **17**(4):529-540.
11. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq**. *Nat Methods* 2008, **5**(7):621-628.
12. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M: **The transcriptional landscape of the yeast genome defined by RNA sequencing**. *Science* 2008, **320**(5881):1344-1349.
13. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F *et al*: **Landscape of transcription in human cells**. *Nature* 2012, **489**(7414):101-108.
14. Kapranov P, Willingham AT, Gingeras TR: **Genome-wide transcription and the implications for genomic organization**. *Nature reviews Genetics* 2007, **8**(6):413-423.
15. Ponting CP: **The functional repertoires of metazoan genomes**. *Nature reviews Genetics* 2008, **9**(9):689-698.
16. Taft RJ, Pheasant M, Mattick JS: **The relationship between non-protein-coding DNA and eukaryotic complexity**. *BioEssays : news and reviews in molecular, cellular and developmental biology* 2007, **29**(3):288-299.
17. Mattick JS, Makunin IV: **Small regulatory RNAs in mammals**. *Human molecular genetics* 2005, **14 Spec No 1**:R121-132.
18. Storz G, Vogel J, Wassarman KM: **Regulation by small RNAs in bacteria: expanding frontiers**. *Molecular cell* 2011, **43**(6):880-891.

19. Brosius J: **Waste not, want not--transcript excess in multicellular eukaryotes**. *Trends Genet* 2005, **21**(5):287-288.
20. Struhl K: **Transcriptional noise and the fidelity of initiation by RNA polymerase II**. *Nature structural & molecular biology* 2007, **14**(2):103-105.
21. Dinger ME, Pang KC, Mercer TR, Mattick JS: **Differentiating protein-coding and noncoding RNA: challenges and ambiguities**. *PLoS computational biology* 2008, **4**(11):e1000176.
22. Ebisuya M, Yamamoto T, Nakajima M, Nishida E: **Ripples from neighbouring transcription**. *Nature cell biology* 2008, **10**(9):1106-1113.
23. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ *et al*: **Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs**. *Cell* 2004, **116**(4):499-509.
24. Ravasi T, Suzuki H, Pang KC, Katayama S, Furuno M, Okunishi R, Fukuda S, Ru K, Frith MC, Gongora MM *et al*: **Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome**. *Genome research* 2006, **16**(1):11-19.
25. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N: **Requirement for Xist in X chromosome inactivation**. *Nature* 1996, **379**(6561):131-137.
26. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT: **Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome**. *Science* 2008, **322**(5902):750-756.
27. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C: **Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation**. *Molecular cell* 2008, **32**(2):232-246.
28. Sleutels F, Zwart R, Barlow DP: **The non-coding Air RNA is required for silencing autosomal imprinted genes**. *Nature* 2002, **415**(6873):810-813.
29. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY: **Long noncoding RNA as modular scaffold of histone modification complexes**. *Science* 2010, **329**(5992):689-693.
30. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA *et al*: **A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression**. *Nature* 2011, **472**(7341):120-124.
31. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M *et al*: **A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response**. *Cell* 2010, **142**(3):409-419.
32. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P *et al*: **Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters**. *Nature genetics* 2011, **43**(7):621-629.
33. Bertani S, Sauer S, Bolotin E, Sauer F: **The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin**. *Molecular cell* 2011, **43**(6):1040-1046.
34. Ponting CP, Oliver PL, Reik W: **Evolution and functions of long noncoding RNAs**. *Cell* 2009, **136**(4):629-641.
35. Rinn JL, Chang HY: **Genome regulation by long noncoding RNAs**. *Annu Rev Biochem* 2012, **81**:145-166.
36. Kung JT, Colognori D, Lee JT: **Long noncoding RNAs: past, present, and future**. *Genetics* 2013, **193**(3):651-669.
37. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP *et al*: **Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals**. *Nature* 2009, **458**(7235):223-227.

38. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A *et al*: **Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression**. *Proc Natl Acad Sci U S A* 2009, **106**(28):11667-11672.
39. Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, Fan L, Sandelin A, Rinn JL, Regev A *et al*: **Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis**. *Genome Res* 2012, **22**(3):577-591.
40. Young RS, Marques AC, Tibbit C, Haerty W, Bassett AR, Liu JL, Ponting CP: **Identification and properties of 1,119 candidate lincRNA loci in the Drosophila melanogaster genome**. *Genome Biol Evol* 2012, **4**(4):427-442.
41. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL: **Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses**. *Genes Dev* 2011, **25**(18):1915-1927.
42. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP: **Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution**. *Cell* 2011, **147**(7):1537-1550.
43. Ponjavic J, Ponting CP: **The long and the short of RNA maps**. *BioEssays : news and reviews in molecular, cellular and developmental biology* 2007, **29**(11):1077-1080.
44. Ulitsky I, Bartel DP: **lincRNAs: Genomics, Evolution, and Mechanisms**. *Cell* 2013, **154**(1):26-46.
45. Numata K, Kanai A, Saito R, Kondo S, Adachi J, Wilming LG, Hume DA, Hayashizaki Y, Tomita M: **Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection**. *Genome research* 2003, **13**(6B):1301-1306.
46. Agarwal A, Koppstein D, Rozowsky J, Sboner A, Habegger L, Hillier LW, Sasidharan R, Reinke V, Waterston RH, Gerstein M: **Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays**. *BMC Genomics* 2010, **11**:383.
47. van Bakel H, Nislow C, Blencowe BJ, Hughes TR: **Most "dark matter" transcripts are associated with known genes**. *PLoS Biol* 2010, **8**(5):e1000371.
48. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y: **RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays**. *Genome research* 2008, **18**(9):1509-1517.
49. Wilhelm BT, Marguerat S, Goodhead I, Bahler J: **Defining transcribed regions using RNA-seq**. *Nat Protoc* 2010, **5**(2):255-266.
50. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation**. *Nature biotechnology* 2010, **28**(5):511-515.
51. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C *et al*: **Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs**. *Nature biotechnology* 2010, **28**(5):503-510.
52. Garber M, Grabherr MG, Guttman M, Trapnell C: **Computational methods for transcriptome annotation and quantification using RNA-seq**. *Nat Methods* 2011, **8**(6):469-477.
53. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq**. *Bioinformatics* 2009, **25**(9):1105-1111.
54. Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, Sasaki D, Imamura K, Kai C, Harbers M *et al*: **CAGE: cap analysis of gene expression**. *Nat Methods* 2006, **3**(3):211-222.
55. Jan CH, Friedman RC, Ruby JG, Bartel DP: **Formation, regulation and evolution of**

Caenorhabditis elegans 3'UTRs. *Nature* 2011, **469**(7328):97-101.
56. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP *et al*: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells**. *Nature* 2007, **448**(7153):553-560.
57. Frith MC, Bailey TL, Kasukawa T, Mignone F, Kummerfeld SK, Madera M, Sunkara S, Furuno M, Bult CJ, Quackenbush J *et al*: **Discrimination of non-protein-coding transcripts from protein-coding mRNA**. *RNA biology* 2006, **3**(1):40-48.
58. Leygue E: **Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected relevance to breast cancer**. *Nucl Recept Signal* 2007, **5**:e006.
59. Jenny A, Hachet O, Zavorszky P, Cyrklaff A, Weston MD, Johnston DS, Erdelyi M, Ephrussi A: **A translation-independent role of oskar RNA in early Drosophila oogenesis**. *Development* 2006, **133**(15):2827-2833.
60. Kloc M, Wilk K, Vargas D, Shirato Y, Bilinski S, Etkin LD: **Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of Xenopus oocytes**. *Development* 2005, **132**(15):3445-3457.
61. Kondo T, Hashimoto Y, Kato K, Inagaki S, Hayashi S, Kageyama Y: **Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA**. *Nature cell biology* 2007, **9**(6):660-665.
62. Galindo MI, Pueyo JI, Fouix S, Bishop SA, Couso JP: **Peptides encoded by short ORFs control development and define a new eukaryotic gene family**. *PLoS Biol* 2007, **5**(5):e106.
63. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL *et al*: **The Pfam protein families database**. *Nucleic acids research* 2008, **36**(Database issue):D281-288.
64. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic acids research* 1997, **25**(17):3389-3402.
65. Eddy SR: **A new generation of homology search tools based on probabilistic inference**. *Genome Inform* 2009, **23**(1):205-211.
66. Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC: **Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana**. *Nature genetics* 2004, **36**(12):1282-1290.
67. Duret L, Chureau C, Samain S, Weissenbach J, Avner P: **The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene**. *Science* 2006, **312**(5780):1653-1655.
68. Ingolia NT, Lareau LF, Weissman JS: **Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes**. *Cell* 2011, **147**(4):789-802.
69. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G: **CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine**. *Nucleic acids research* 2007, **35**(Web Server issue):W345-349.
70. Liu J, Gough J, Rost B: **Distinguishing protein-coding from non-coding RNAs through support vector machines**. *PLoS genetics* 2006, **2**(4):e29.
71. Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S: **The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus**. *Cell* 1992, **71**(3):515-526.
72. Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF: **The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus**. *Cell* 1992, **71**(3):527-542.
73. Jeon Y, Lee JT: **YY1 tethers Xist RNA to the inactive X nucleation center**. *Cell* 2011, **146**(1):119-133.

74. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P: **The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin**. *Science* 2008, **322**(5908):1717-1720.
75. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E *et al*: **Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs**. *Cell* 2007, **129**(7):1311-1323.
76. Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R: **Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription**. *Nature* 2008, **454**(7200):126-130.
77. Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD: **The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator**. *Genes & development* 2006, **20**(11):1470-1484.
78. Shamovsky I, Ivannikov M, Kandel ES, Gershon D, Nudler E: **RNA-mediated response to heat shock in mammalian cells**. *Nature* 2006, **440**(7083):556-560.
79. Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, Kominsky HD *et al*: **Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression**. *Nature biotechnology* 2011, **29**(8):742-749.
80. Brunner AL, Beck AH, Edris B, Sweeney RT, Zhu SX, Li R, Montgomery K, Varma S, Gilks T, Guo X *et al*: **Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers**. *Genome biology* 2012, **13**(8):R75.
81. Ziats MN, Rennert OM: **Aberrant expression of long noncoding RNAs in autistic brain**. *J Mol Neurosci* 2013, **49**(3):589-593.
82. Esteller M: **Non-coding RNAs in human disease**. *Nat Rev Genet* 2011, **12**(12):861-874.
83. Salta E, De Strooper B: **Non-coding RNAs with essential roles in neurodegenerative disorders**. *Lancet Neurol* 2012, **11**(2):189-200.
84. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL *et al*: **Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis**. *Nature* 2010, **464**(7291):1071-1076.
85. Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, Tanaka F, Shibata K, Suzuki A, Komune S *et al*: **Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers**. *Cancer research* 2011, **71**(20):6320-6326.
86. Yang Z, Zhou L, Wu LM, Lai MC, Xie HY, Zhang F, Zheng SS: **Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation**. *Annals of surgical oncology* 2011, **18**(5):1243-1250.
87. Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM: **Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a**. *Molecular cell* 2010, **38**(5):662-674.
88. Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y: **Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene**. *Oncogene* 2011, **30**(16):1956-1962.
89. Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H: **Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA**. *Nature* 2008, **451**(7175):202-206.
90. Kim WY, Sharpless NE: **The regulation of INK4/ARF in cancer and aging**. *Cell* 2006, **127**(2):265-275.
91. Reis EM, Verjovski-Almeida S: **Perspectives of Long Non-Coding RNAs in Cancer**

Diagnostics. *Frontiers in genetics* 2012, **3**:32.
92. Panzitt K, Tschernatsch MM, Guelly C, Moustafa T, Stradner M, Strohmaier HM, Buck CR, Denk H, Schroeder R, Trauner M *et al*: **Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA**. *Gastroenterology* 2007, **132**(1):330-342.
93. Shappell SB, Fulmer J, Arguello D, Wright BS, Oppenheimer JR, Putzi MJ: **PCA3 urine mRNA testing for prostate carcinoma: patterns of use by community urologists and assay performance in reference laboratory setting**. *Urology* 2009, **73**(2):363-368.
94. Wang KC, Chang HY: **Molecular mechanisms of long noncoding RNAs**. *Molecular cell* 2011, **43**(6):904-914.
95. Liu X, Li D, Zhang W, Guo M, Zhan Q: **Long non-coding RNA gadd7 interacts with TDP-43 and regulates Cdk6 mRNA decay**. *The EMBO journal* 2012, **31**(23):4415-4427.
96. Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP: **Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor**. *Sci Signal* 2010, **3**(107):ra8.
97. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M *et al*: **Circular RNAs are a large class of animal RNAs with regulatory potency**. *Nature* 2013, **495**(7441):333-338.
98. Bacon LD, Hunt HD, Cheng HH: **A review of the development of chicken lines to resolve genes determining resistance to diseases**. *Poult Sci* 2000, **79**(8):1082-1093.
99. Gimeno IM, Witter RL, Hunt HD, Lee LF, Reddy SM, Neumann U: **Marek's disease virus infection in the brain: virus replication, cellular infiltration, and major histocompatibility complex antigen expression**. *Vet Pathol* 2001, **38**(5):491-503.
100. Li T, Wang S, Wu R, Zhou X, Zhu D, Zhang Y: **Identification of long non-protein coding RNAs in chicken skeletal muscle using next generation sequencing**. *Genomics* 2012, **99**(5):292-298.
101. Ahanda ML, Ruby T, Wittzell H, Bed'Hom B, Chausse AM, Morin V, Oudin A, Chevalier C, Young JR, Zoorob R: **Non-coding RNAs revealed during identification of genes involved in chicken immune responses**. *Immunogenetics* 2009, **61**(1):55-70.
102. Pang KC, Dinger ME, Mercer TR, Malquori L, Grimmond SM, Chen W, Mattick JS: **Genome-wide identification of long noncoding RNAs in CD8+ T cells**. *Journal of immunology* 2009, **182**(12):7738-7748.
103. Peng X, Gralinski L, Armour CD, Ferris MT, Thomas MJ, Proll S, Bradel-Tretheway BG, Korth MJ, Castle JC, Biery MC *et al*: **Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling**. *MBio* 2010, **1**(5).
104. Davison TF, Nair V, Institute for Animal Health (Great Britain): **Marek's disease : an evolving problem**. Amsterdam ; Boston: Elsevier; 2004.
105. Parvizi P, Abdul-Careem MF, Haq K, Thanthrige-Don N, Schat KA, Sharif S: **Immune responses against Marek's disease virus**. *Anim Health Res Rev* 2010, **11**(2):123-134.
106. Bacon LD, Hunt HD, Cheng HH: **Genetic resistance to Marek's disease**. *Curr Top Microbiol Immunol* 2001, **255**:121-141.
107. Sarson AJ, Parvizi P, Lepp D, Quinton M, Sharif S: **Transcriptional analysis of host responses to Marek's disease virus infection in genetically resistant and susceptible chickens**. *Animal genetics* 2008, **39**(3):232-240.
108. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S *et al*: **Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes**. *Genome research* 2005, **15**(8):1034-1050.
109. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources**. *Nat Protoc* 2009, **4**(1):44-57.

110. Huang da W, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists**. *Nucleic acids research* 2009, **37**(1):1-13.
111. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(43):15545-15550.
112. Cai X, Cullen BR: **The imprinted H19 noncoding RNA is a primary microRNA precursor**. *RNA* 2007, **13**(3):313-316.
113. Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, Smits G, Reik W: **The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r**. *Nature cell biology* 2012, **14**(7):659-665.
114. Smith CM, Steitz JA: **Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes**. *Molecular and cellular biology* 1998, **18**(12):6897-6909.
115. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A: **Rfam 11.0: 10 years of RNA families**. *Nucleic acids research* 2013, **41**(Database issue):D226-232.
116. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L *et al*: **lincRNAs act in the circuitry controlling pluripotency and differentiation**. *Nature* 2011, **477**(7364):295-300.
117. Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Solda G, Simons C *et al*: **Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation**. *Genome research* 2008, **18**(9):1433-1445.
118. Cai S, Han HJ, Kohwi-Shigematsu T: **Tissue-specific nuclear architecture and gene expression regulated by SATB1**. *Nature genetics* 2003, **34**(1):42-51.
119. Yasui D, Miyano M, Cai S, Varga-Weisz P, Kohwi-Shigematsu T: **SATB1 targets chromatin remodelling to regulate genes over long distances**. *Nature* 2002, **419**(6907):641-645.
120. Church DM, Goodstadt L, Hillier LW, Zody MC, Goldstein S, She X, Bult CJ, Agarwala R, Cherry JL, DiCuccio M *et al*: **Lineage-specific biology revealed by a finished genome assembly of the mouse**. *PLoS Biol* 2009, **7**(5):e1000112.
121. Pang KC, Frith MC, Mattick JS: **Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function**. *Trends Genet* 2006, **22**(1):1-5.
122. Wilusz JE, Sunwoo H, Spector DL: **Long noncoding RNAs: functional surprises from the RNA world**. *Genes & development* 2009, **23**(13):1494-1504.
123. Hirota K, Miyoshi T, Kugou K, Hoffman CS, Shibata T, Ohta K: **Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs**. *Nature* 2008, **456**(7218):130-134.
124. Uhler JP, Hertel C, Svejstrup JQ: **A role for noncoding transcription in activation of the yeast PHO5 gene**. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(19):8011-8016.
125. Keles H, Fidan AF, Cigerci IH, Kucukkurt I, Karadas E, Dundar Y: **Increased DNA damage and oxidative stress in chickens with natural Marek's disease**. *Vet Immunol Immunopathol* 2010, **133**(1):51-58.
126. Huang W, Long N, Khatib H: **Genome-wide identification and initial characterization of bovine long non-coding RNAs from EST data**. *Animal genetics* 2012, **43**(6):674-682.