Systematic identification and characterization of chicken (*Gallus gallus*) ncRNAs

Yong Zhang[1], Jie Wang[2], Shoujun Huang[1], Xiaopeng Zhu[2], Jun Liu[3], Ning Yang[4], Dandan Song[5], Rimao Wu[1], Wei Deng[2], Geir Skogerbø[2], Xiu-Jie Wang[3], Runsheng Chen[2,*] and Dahai Zhu[1,*]

[1]National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100005, [2]Bioinformatics Laboratory, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, [3]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Datun Road, Beijing 100101, [4]Department of Animal Genetics and Breeding, College of Animal Science and Technology, China Agricultural University, Beijing 100094 and [5]Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

ABSTRACT

Recent studies have demonstrated that non-coding RNAs (ncRNAs) play important roles during development and evolution. Chicken, the first genome-sequenced non-mammalian amniote, possesses unique features for developmental and evolutionary studies. However, apart from microRNAs, information on chicken ncRNAs has mainly been obtained from computational predictions without experimental validation. In the present study, we performed a systematic identification of intermediate size ncRNAs (50–500 nt) by ncRNA library construction and identified 125 chicken ncRNAs. Importantly, through the bioinformatics and expression analysis, we found the chicken ncRNAs has several novel features: (i) comparative genomic analysis against 18 sequenced vertebrate genomes revealed that the majority of the newly identified ncRNA candidates is not conserved and most are potentially bird/ chicken specific, suggesting that ncRNAs play roles in lineage/species specification during evolution. (ii) The expression pattern analysis of intronic snoRNAs and their host genes suggested the coordinated expression between snoRNAs and their host genes. (iii) Several spatio-temporal specific expression patterns suggest involvement of ncRNAs in tissue development. Together, these findings provide new clues for future functional study of ncRNAs during development and evolution.

INTRODUCTION

The completion of an increasing number of genome sequencing projects have revealed that sequences corresponding to protein-coding genes only comprise around 1.5–2% of vertebrate genomes (1–4). Accumulating evidence suggests that considerable proportions of eukaryotic genomes are transcribed as non-coding RNAs (ncRNAs) (5–8). A number of different types of ncRNAs have been identified (9–14), ranging in size from around 20 nt, e.g. miRNAs and piRNAs, to over 10 000 nt. Importantly, a growing body of evidence has shown that ncRNAs can act as key regulators of development and other biological processes, supporting the notion that ncRNAs are indispensable players in the control of eukaryotic life/biology (15–20).

The chicken (*Gallus gallus*) possesses unique features for the study of developmental processes of organogenesis and has for several decades been used as a model organism for vertebrate developmental biology. The domesticated chicken originates from the red jungle fowl (*G. gallus*), and shares a common ancestor with mammals ∼310 million years ago (Mya). The chicken was the first non-mammalian amniote to have its genome sequenced; the genome size is about 40% of the average mammal

genome and contains 20 000–23 000 predicted protein-coding genes. As a representative of the birds, chicken occupies an important evolutionary position which bridges the evolutionary gap between mammals and other vertebrates (4), and some genomic syntenic regions are more conserved between human and chicken than between human and mouse (21). Therefore, comparative genomic analyses of the chicken genome could provide important information for functional and evolutionary studies of vertebrate protein coding and non-protein-coding genes.

Genome-wide identification of chicken ncRNAs has mainly relied on computational predictions by searching for sequences with homology to ncRNAs identified in other species. The first draft of the chicken genome sequencing project only predicted 571 RNA genes, mainly including highly conserved tRNAs, rRNAs, snRNAs, RNase P, Y RNA, SRP, 7SK and a few conserved miRNAs and snoRNAs (4). The updated data from ENSEMBL (Release 52) annotated 1026 chicken ncRNA genes due to an increased number of predicted snoRNAs and miRNAs based on searching for homologs of human and/or mouse RNAs (560 miRNAs and 148 snoRNAs) (22). Very recently, Qu's group reported 155 chicken snoRNAs predicted by snoSeeker (23). Thus far, 5732 human and 3287 mouse RNA genes, respectively, have been annotated by the ENSEMBL Release 52, and based on these numbers it is reasonable to expect that there exist at least 3000–4000 ncRNA genes in the chicken genome that have not yet been identified. The identification of ncRNAs escaping bioinformatical/computational prediction will rely on experimental strategies and systematic screening. Screenings for chicken microRNAs have been performed by several groups (24–26), but genome-wide identification of other types of chicken ncRNAs by experimental RNomics has yet not been reported and this prompted us to perform a profiling of the chicken non-miRNA ncRNAs. In the present study, we have focused on the systematic identification of intermediate size ncRNAs with lengths in the size range 50–500 nt.

A majority of known ncRNAs in the size range 50–500 nt are snoRNAs, snRNAs, tRNAs, 5SrRNAs and 5.8SrRNAs. During the past few years, a number of intermediate size ncRNAs have been identified in a variety of organisms spanning from *Escherichia coli* to *Homo sapiens* by experimental approaches or computational prediction or combination of the two strategies (9,27–31). snoRNAs, which mainly function in modification of ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs) and tRNAs (32,33), represent one of the largest groups of functional ncRNAs currently known in eukaryotic cells. Based on their sequence and structural features, snoRNAs can be classified into two families, the C/D box snoRNAs that generally guide 2′-O-methylation and the H/ACA box snoRNAs that guide rRNA pseudouridylation, both by pairing with targets via sequence complementarity (17,34). Although it has been reported that some human C/D box snoRNAs are independently transcribed as evidenced by the presence of methylated guanosine caps at their 5′-ends (35), most

vertebrate snoRNAs are encoded in the intronic regions of protein-coding genes with ribosomal or other translation-related functions (36). In general, the biogenesis of intron-encoded snoRNAs is dependent on the transcription of their host genes (37). snoRNAs were originally assumed to guide rRNA modification, however, recent findings have demonstrated that transcripts with snoRNA characteristics can also target other types of RNAs such as mRNAs and thereby regulate their post-transcriptional expression or splicing patterns. The C/D box snoRNA HBII-52 regulated alternative splice site usage in a reporter construct carrying exons IV, Va, Vb and VI of the serotonin receptor 5-HT$_{2C}$R when transfected into immortalized neuronal cells (19). In another report it was shown that an artificial C/D box snoRNA could induce pre-mRNA splicing impairment (38). In addition, it has been demonstrated that a subset of H/ACA box snoRNAs located in the Cajal bodies are involved in the regulation of telomerase RNA localization (36,39). Moreover, tissue-specific and developmentally regulated expression indicates that some snoRNAs have regulatory functions during development (40,41). In summary, previous findings suggest the existence of numerous snoRNAs in eukaryotic genomes with functional roles extending beyond those merely related to ribosomal RNA modification. Therefore, identification and characterization of broad sets of snoRNAs and other intermediate size regulatory ncRNAs in various organisms is crucial for a better understanding of ncRNA functions in development and evolution.

In this study, we carried out a genome-wide systematic identification of chicken ncRNAs by constructing chicken ncRNAs libraries (50–500 nt). A total of 125 unique ncRNAs were cloned, including 68 recently predicted ncRNA candidates (23). We examined the sequence conservation of the ncRNAs among 18 vertebrate species and found that most of the newly identified ncRNAs are potentially chicken- or bird-specific. The expression pattern analysis of intronic snoRNAs and their host genes suggested the coordinated expression between snoRNAs and their host genes. Finally, several ncRNAs exhibited specific tissue and temporal expression patterns, indicating functional roles in tissue development.

MATERIALS AND METHODS

Identification and classification of chicken ncRNAs

Total RNA was isolated from chicken tissues at embryonic Day 14, post-hatch 1 day and post-hatch 4 weeks. At each time-point, five randomly selected male individuals were sacrificed for tissue sampling. Full-length ncRNA specific libraries of both capped and uncapped transcripts were generated according to a previously described method (30) with modifications. Total RNA was fractionated on Qiagen-tips with a 0.6–1.0 M NaCl gradient elution of QRW2 buffer (Qiagen RNA/DNA handbook). Highly abundant rRNAs (5.8S rRNAs and 5S rRNAs) and snRNAs (U1 snRNA, U2 snRNA, U4 snRNA and U5 snRNA) were removed from the small RNA fraction (50–500 nt) with

the Ambion MicrobExpress kit. The remaining RNAs were dephosphorylated with calf intestine alkaline phophatase (Fermentas), and ligated to the 3′ adaptor with T4 RNA ligase (Fermentas). After removal of excessive 3′ adaptor, the ligation products were split into two aliquots, of which one was treated with PolyNucleotide Kinase (PNK, Fermentas) to phosphorylate non-capped RNA, and the other was treated with Tobacco Acid Pyrophosphatase (TAP, Epicentre) to remove 5′-end methyl-guanosine caps from capped RNA. Thereafter, both samples were ligated to the 5′ adaptor and reverse transcribed with Thermoscript reverse transcriptase (Invitrogen) using oligo 3RT as the RT primer. The cDNA was amplified by PCR for 13 cycles using Platinum Taq (Invitrogen) with 3RT and 5CD primers, cloned into pGEM-T vector (Promega) and sequenced. All oligonucleotide sequences used in this study are shown in Supplementary Data file 4.

After removing redundant sequences, the obtained unique ncRNA sequences were first used as queries in a BlastN search against the NCBI nucleotide database and the Ensemble annotated chicken genome sequences to remove matches to rRNAs, tRNAs and mRNAs. The remaining sequences were used for ncRNA prediction and classification using the INFERNAL (42) and SnoReport software (43). The genomic loci of snoRNAs were obtained with the BLAT program (44), and comparisons of cloned snoRNAs with snoRNAs recorded in NCBI, Ensemble and Rfam databases was performed according to the genomic loci of snoRNAs. Known snoRNAs from other species were downloaded from the Rfam database. Transcripts that could not be annotated by any former databases or tools were regarded as unclassified ncRNAs.

For the experimentally identified snoRNA candidates, their potential targets were computationally predicted based on their base complementary information by using pairwise alignment. We searched against several known chicken non-coding transcripts including four rRNAs (18S, 28S, 5.8S, 5S) and eight snRNAs (U1, U2, U4a, U4b, U4b1, U4c, U4x and U5a).

Northern blot hybridization

Total RNA extracted from eight chicken tissues, including heart, liver, brain, lung, kidney, intestine, spleen and skeletal muscle at different development stages, were separated by 8% PAGE (7 M urea) and transferred to Nylon membrane (N+, Amersham). Antisense RNAs against the ncRNAs were labeled with Digoxigenin-11-UTP by *in vitro* transcription with T7 and SP6 RNA polymerase. The RNA blots were hybridized in ULTRAhyb (Ambion) at 68°C overnight, washed with $2\times$ SSC/0.1% SDS washing buffer at 68°C for 2×5 min, following stringent washing with $0.1\times$ SSC/0.1% SDS washing buffer at 68°C for 2×30 min. Thereafter, the RNA blots were blocked with blocking buffer for 30–60 min at room temperature and incubated for 30 min with anti-DIG-AP (1 : 10 000 diluted in blocking buffer). Hybridization signals were detected by CDP-star (Roche). Chemiluminescent signals were exposed

to X-ray film. For detection of snoRNA expression in different tissues, RNA samples from chicken at 2 weeks post hatching were used.

Analysis of snoRNA host genes

Intronic ncRNAs were identified according to the ENSEMBL release 52 *Gallus gallus* genome annotation version WASHUC2. The Gene Ontology analysis of host genes of intronic snoRNAs was carried out using the GOEAST software (45). The microarray data used in this study covered expression profiles of chicken skeletal muscle tissues at post-hatch 1 day, 2, 4, 6 and 8 weeks. Three biological replicates were performed for each time point, and the average signal for each gene over the three biological replicates were used for further analysis. The expression changes shown in Figure 4B were calculated as the expression of a gene at a time point minus its expression at the immediate preceding time point.

Conservation analysis of ncRNAs

The genomic sequences of chicken and 18 vertebrate genomes were downloaded from the UCSC genome browser. The conservation of chicken ncRNAs in the 18 vertebrate genomes was determined by using chicken ncRNAs as queries in BLAT searches of other genomes (44). Conservation scores were calculated based on the maximal alignment length and the identity of the alignment of the BLAT hits on each vertebrate genome. The conservation status of all cloned ncRNAs was clustered hierarchically according to their conservation scores in the 18 genomes using the hclust package in R.

Preparation and transfection of chicken fetal myoblasts

Chicken fetal myoblasts (CFMs) were prepared from pectoralis muscle tissue from 11th day White Leghorn chicken embryos according to a modified protocol described by Yang *et al.* (46). Briefly, the pectoralis muscle tissue was collected and washed with phosphate buffered saline (PBS). The tissue was minced and digested with 0.1% collagenase type I in PBS buffer at 37°C for 30 min with constantly shaking. The digested tissue was collected by centrifuging, resuspended in PBS buffer by repeated pipetting and filtrated to remove aggregated cells and myotubes. The collected cells were resuspended in α-MEM medium, seeded into a 100-mm dish, and cultured for 1 h at 37°C. Rapidly adhering cells were mainly composed of fibroblasts and were discarded. Myoblasts remaining in suspension were collected, transferred to a six-well gelatin-coated plate at 1×10^5 cells/cm^2 with α-MEM containing 10% Horse Serum, 100 IU/ml penicillin and 100 μg/ml streptomycin (Invitrogen), and cultured for 24 h at 37°C in humidified 95% air–5% CO_2. The myoblasts were transfected with the Lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's instructions. The transfected cells were lysed after 48 h of transfection, and total RNA was extracted using the TRIzol reagent (Invitrogen). The primers for *pfkm* minigene generation are shown in Supplementary Data file 4.

RESULTS

Identification and Classification of Chicken ncRNAs

Full-length intermediate size ncRNAs (50–500 nt) enriched libraries were constructed according to a previously described method (30) with some modifications. This procedure ensured that the library contained a substantial fraction of full-length ncRNA clones with defined 5′ and 3′ termini. The total RNA used for library construction was extracted from a mix of 10 chicken tissues (heart, liver, brain, lung, kidney, intestine, glandular stomach, gizzard, spleen, skeletal muscle) harvested at three developmental stages (14 days old embryo, 1 day and 4 weeks chicken after hatching). Altogether, 3468 clones from two full-length cDNA libraries were sequenced. After removing matches to tRNAs and rRNAs, the remaining 201 sequences were taken as ncRNA candidates. Among these, 33 clones with a match to EST or RefSeq RNA sequences were excluded in this report because they can not be detected the same size band as obtained from library sequencing, suggesting they could be mRNA degradation products. The 43 other ncRNA candidates were also excluded because they did not yield detectable expression signals by northern blot. The remaining 125 ncRNAs candidates validated by northern blot included 90 ncRNA loci predicted by the ENSEMBL or Rfam (22,47) and 35 novel candidates (Figure 1, Supplementary Data files 1 and 2). The previously predicted 90 ncRNAs included 77 snoRNAs, 10 snRNA, two Y RNAs and one SRP RNA. The 35 novel ncRNA candidates were comprised of 25 transcripts with possible snoRNA characteristics, one transcript predicted to be a snRNA, and nine transcripts that could not be assigned to any known class of ncRNAs (Table 1). Among the 25 novel snoRNA candidates, three C/D box snoRNA candidates could target chicken rRNAs, whereas 10 of the novel snoRNA candidates might be orphan snoRNAs with no apparent target in chicken rRNAs or snRNAs. The other 12 novel snoRNA candidates were also independently predicted by Qu's group (23) (Table 1 and Supplementary Data file 3). The expression of the 35 novel ncRNA candidates was shown in Figure 2. They are ubiquitously expressed in the eight tested chicken tissues, the single exception being Ggn39, which is specifically expressed in skeletal muscle and heart tissue. The expression patterns of the other ncRNAs were shown in Supplementary Figure S1.

Potentially bird-specific ncRNAs

To analyze the evolutionary conservation of the ncRNAs, we searched for homologs of all the 125 chicken ncRNAs in the genomes of the lizard and 6 fish and 11 mammalian species, and hierarchically clustered the ncRNAs according to their conservation scores. The 125 chicken ncRNAs could be grouped into seven clusters according to their sequence conservation scores in each of the 18 queried genomes (Figure 3A), and one ncRNA from each of the seven clusters was selected as a representative and shown in Figure 3B. As shown in Figure 3A, compared to the highly conserved snRNAs, previously
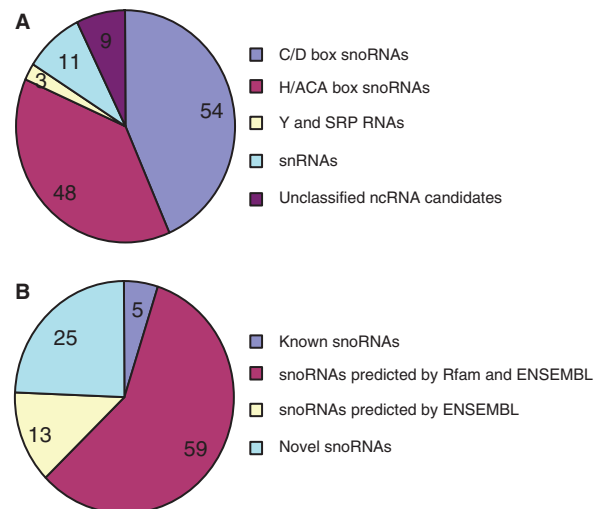
Figure 1. Classification of cloned chicken ncRNAs. (**A**) Classification of unique ncRNA sequences. (**B**) Distribution of cloned snoRNAs.

predicted snoRNAs are less conserved, whereas the newly identified snoRNAs and unclassified ncRNAs are poorly or not conserved. Ten snRNAs, two Y RNA and SRP RNA are highly conserved and grouped into clusters 1 and 2 (e.g. Ggn130_U4 and Ggn85_Y RNA in Figure 3B), and only the predicted snRNA (Ggn103) shows relatively low conservation (cluster 7). Unlike snRNAs, snoRNAs are generally less conserved. Of the 77 previously predicted snoRNAs, none are grouped into cluster 1 and only two are grouped into cluster 2 (Ggn112_SCARNA8 and Ggn6_SNORD18), whereas 36 (46.7%) belong to cluster 6, which is composed of transcripts with very limited sequence conservation (e.g. Ggn75_SNORD36 in Figure 3B). Interestingly, 16 snoRNAs, fall into clusters 3–5 (e.g. Ggn46, Ggn9 and Ggn120 in Figure 3B) and are only conserved in the lizard and/or platypus genomes but not in other mammals, suggesting the existence of an ancient group of snoRNAs that have been preserved in the reptile–bird lineage.

The 35 newly identified snoRNAs and unclassified ncRNA candidates are all not well conserved. In total, 22 of 25 novel snoRNA candidates (88%) are grouped into cluster 7 which show no conservation in any other examined genome/organism (Figure 3A). Six of the nine unclassified ncRNA candidates (67%) are also unconserved and are grouped into cluster 7 (e.g. Ggn39 in Figure 3B). Consistent with their low conservation score in mammalian genomes, only one ncRNA out of 35 novel ncRNAs can be detected in human and mouse tissues by northern blot (Figure 2). Interestingly, three snoRNAs (Ggn5, Ggn51 and Ggn112) can be detected in human but not in mouse tissues (Figure 3C and Supplementary Figure 2). In support of this observation, genomic sequence analysis showed these three snoRNAs have higher sequences similarity between chicken and human than between chicken and mouse, indicating that these chicken snoRNAs are shared by chicken and a limited number of mammalian lineages.

Table 1. General information for 35 novel ncRNA candidates identified in this study

GGN_ID	GenBank_ID	Reads_no	Length	ncRNA_type	Target	Genomic _location in_chick	chr_location_ in_zebra finch	Seq_identity_ with_zebra finch
GGN11	EU240230	2	85	CD	28S rRNA	Intergenic	No hit	No hit
GGN20	EU240238	1	61	CD	28S rRNA	Intergenic	No hit	No hit
GGN37	EU240254	1	70	CD	28S rRNA	Intergenic	chr4 13645396 13645464 +	85.30%
GGN86	EU240302	9	96	CD	Orphan	Intergenic	chr22_random 663757 663840 +	92.50%
GGN120	EU240333	1	71	CD	Orphan	Intronic	No hit	No hit
GGN138	EU240346	1	58	CD	Orphan	Intergenic	No hit	No hit
GGN148	EU240352	2	66	CD	Orphan	No_genomic_loci	No hit	No hit
GGN100[a]	EU240315	3	78	CD	18S rRNA	Intronic	chr7 21747676 21747750 −	85.40%
GGN71[a]	EU240287	1	77	CD	18S rRNA	Intronic	chr5 61606460 61606536 −	85.80%
GGN107[a]	EU240321	6	68	CD	28S rRNA	Intronic	chr8 3321788 3321851 −	87.50%
GGN52[a]	EU240268	1	66	CD	18S rRNA	Intronic	chr8 3323611 3323676 −	87.90%
GGN34[a]	EU240252	1	66	CD	28S rRNA	Intronic	chr2 121527086 121527152 −	88.20%
GGN108[a]	EU240322	2	99	CD	28S rRNA	Intronic	chr8 17728445 17728543 +	91.00%
GGN80[a]	EU240296	4	82	CD	18S rRNA	Intronic	chr17 6882327 6882410 +	95.90%
GGN82[a]	EU240298	2	67	CD	18S rRNA	Intronic	chr19 7185695 7185736 +	97.50%
GGN17[a]	EU240236	1	70	CD	28S rRNA	Intronic	No hit	No hit
GGN79	EU240295	3	133	HACA	Orphan	Intronic	chr15 7584209 7584339 −	86.30%
GGN126	EU240339	1	123	HACA	Orphan	Intergenic	chr5 9310745 9310867 +	87.90%
GGN72	EU240288	5	129	HACA	Orphan	Intronic	chr1 80099319 80099434 +	90.60%
GGN87	EU240303	3	130	HACA	Orphan	Intergenic	chr23 92846 92969 −	92.40%
GGN58	EU240274	11	134	HACA	Orphan	Intergenic	No hit	No hit
GGN56	EU240272	2	135	HACA	Orphan	Intronic	No hit	No hit
GGN32[a]	EU240250	1	136	HACA	28S rRNA	Intergenic	chr1 57671975 57672107 +	90.00%
GGN123[a]	EU240336	5	130	HACA	18S rRNA	Intronic	No hit	No hit
GGN74[a]	EU240290	2	137	HACA	28S rRNA	Intronic	No hit	No hit
GGN103	EU240318	4	122	snRNA		Intergenic	No hit	No hit
GGN141	EU240348	1	123	Unclassified		Intergenic	chr1 80104054 80104169 +	83.70%
GGN67[a]	EU240283	2	69	Unclassified		Intronic	chr1 90202832 90202890 +	88.20%
GGN136	EU240344	1	115	Unclassified		Intergenic	chr20 6905097 6905198 −	90.90%
GGN105	EU240320	3	62	Unclassified		Intronic	chr8 3324079 3324144 −	91.70%
GGN68	EU240284	1	227	Unclassified		Intronic	LGE22 223377 223624 −	92.40%
GGN46[a]	EU240262	1	76	Unclassified		Intronic	chr1A 50331437 50331502 +	100%
GGN46[a]	EU240262	1	76	Unclassified		Intronic	chr1A 50337731 50337796 +	100%
GGN147	EU240351	1	96	Unclassified		Intronic	No hit	No hit
GGN16	EU240235	1	56	Unclassified		Intronic	No hit	No hit
GGN39	EU240256	1	67	Unclassified		Intronic	No hit	No hit

[a]ncRNAs independently predicted by Qu's group (23).

We further searched for homologs of 35 novel ncRNA candidates in the recently released zebra finch (*Taeniopygia guttata*) genome. Twenty one of the novel transcripts have homologs in the zebra finch with sequence identity above 85%, whereas the remaining 14 ncRNAs had no hits in the current version of the zebra finch genome (Table 1). Although this may be due to the incomplete genomic sequence available for the zebra finch it is also possible that some of these ncRNAs are specific to chicken.

Host gene analysis of intron-encoded snoRNAs

Of the 102 snoRNAs, 89 were encoded by intronic regions of protein-coding genes and 11 were located in intergenic regions (Table 2). For two of the snoRNA transcripts no genomic loci could be found in the current version of the chicken genome sequence. The 11 non-intronic snoRNAs appear all to be unique transcriptional units. For the intronic snoRNAs, we found 12 protein-coding genes each hosting two intronic snoRNAs, two genes hosting three snoRNAs and two genes hosting four snoRNAs (Supplementary Data file 2). Only in two cases, a single

intron encodes the two ncRNA loci, and in both cases the two loci within the same intron encoded identical ncRNA transcripts (Supplementary Figure S3).

We performed functional analysis of the snoRNA host genes based on Gene Ontology annotations. Host genes of both C/D box and H/ACA box snoRNAs were enriched with functional annotation relating to ribosome biogenesis, translation and cellular protein metabolic processes. We also found that host genes of both C/D box and H/ACA box snoRNAs were enriched for some other functions, including initiation of DNA duplication, protein transportation, regulation of translation initiation, regulation of cell cycle and other processes (Figure 4A and Supplementary Figure S4).

To further explore the possible coordinated expression pattern between intronic snoRNAs and their host genes, we examined the expression of snoRNA host genes using our recently published Affymetrix microarray data (48). RNA samples from chicken skeletal muscle tissues at 1 day, or 2, 4, 6 and 8 weeks after hatching were used for the array analysis. The results show that in general, the expression of snoRNA host genes did not vary much
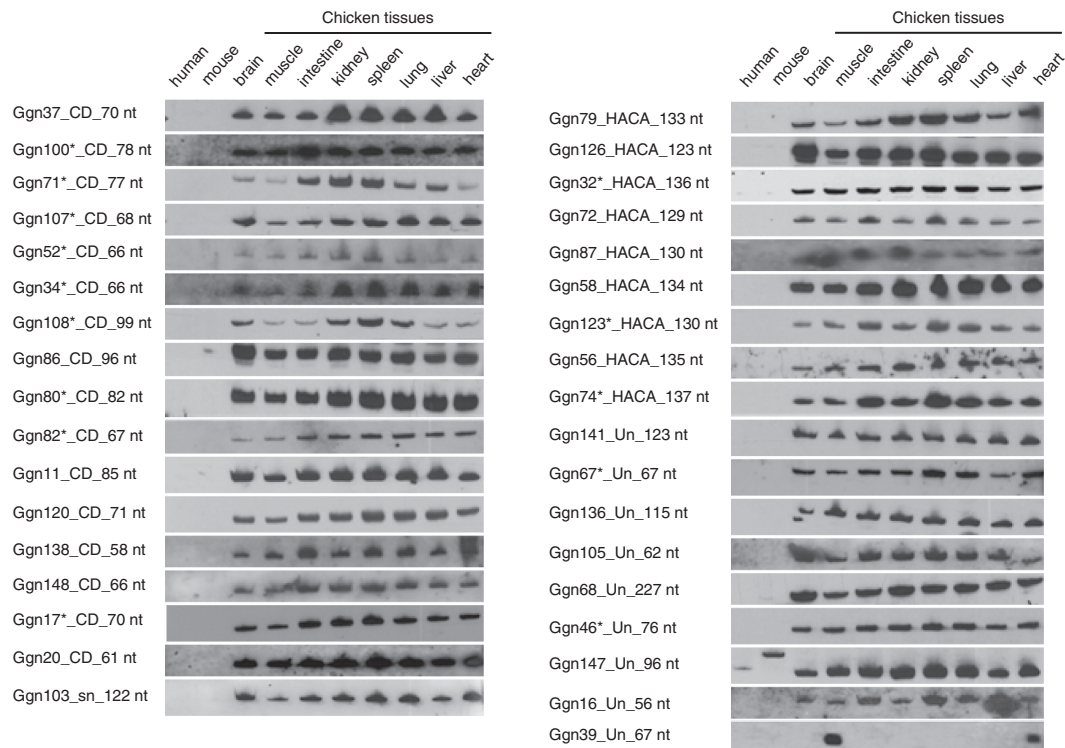
Figure 2. Tissue expression pattern of 35 novel ncRNA candidates. The expression pattern of 35 novel ncRNA candidates was examined by northern blot analysis of total RNA from chicken heart, liver, lung, spleen, kidney, intestine, skeletal muscles and brain. Total RNA from human and mouse skeletal muscle were included in each blot in order to test for possible expression of the ncRNA candidates in different species. The ncRNA candidates are labeled as GgnX_class_size on the left of each northern blot. (For example: Ggn37_CD_70 nt means Ggn37 is a predicted C/D box snoRNA with length 70 nt. 'Un' denotes 'unclassified'.) Asterisks indicate that ncRNAs were independently predicted by Qu's group (23).

across the examined developmental stages, and that genes with ribosome related functions had least variation in expression (Figure 4B). Host genes with functions related to cell cycle and apoptosis showed the most variable expression across developmental stages (Figure 4B). Correspondingly, snoRNAs encoded in intronic regions of ribosome related genes also exhibited constant expression across different developmental stages, whereas snoRNAs hosted by non-ribosome related genes showed variable expression levels during postnatal development (Supplementary Figure S5). It is worth noting that snoRNA host genes with ribosome and translation related functions had much higher expression levels than the average coding gene expression, whereas the expression levels of host genes with other known functions were relatively low (Figure 4C). Similarly, such expression discrepancy also observed in the expression pattern of snoRNAs during chicken muscle development (Supplementary Figure S5).

Tissue and developmental stage-specific expression of chicken ncRNAs

The ncRNA expression patterns may yield important clues to their functions. Thus, we investigated the expression of ncRNAs in various chicken tissues, including heart, liver, lung, spleen, kidney, intestine, brain and skeletal muscle. All examined ncRNAs were

detected by northern blot analysis (Supplementary Figure S1). Unlike snRNAs and Y-RNAs, which showed constant expression levels across all examined tissues, some of the snoRNAs tended to have lower expression in brain, muscle and heart than in other tissues (Figure 5A). Interestingly, we found that one of the unclassified ncRNA (Ggn39) was only expressed in the chicken heart and skeletal muscle (Figure 5A). To look at the possible functional roles of the ncRNAs during skeletal muscle development, we examined the expression pattern of these ncRNAs in chicken skeletal muscle tissues at different developmental stages by northern blot analysis. Our results showed that several of the snoRNAs had developmentally regulated expression patterns (Figure 5B and Supplementary Figure S5). For example, snoRNAs Ggn74, Ggn3 and Ggn7 were highly expressed during embryonic development, whereas snoRNA Ggn5 was more abundantly expressed after hatch (Figure 5B). Ggn39, which is specifically expressed in skeletal muscle and heart, was predominantly expressed at the post-hatching stages (Figure 5B), suggesting that this unclassified ncRNA might have specific functions in myogenesis during postnatal development.

To further investigate the possible involvement of ncRNAs in chicken myogenesis, we analyzed the expression pattern of the 105 ncRNAs in the skeletal muscle tissues of two chicken lines with divergent
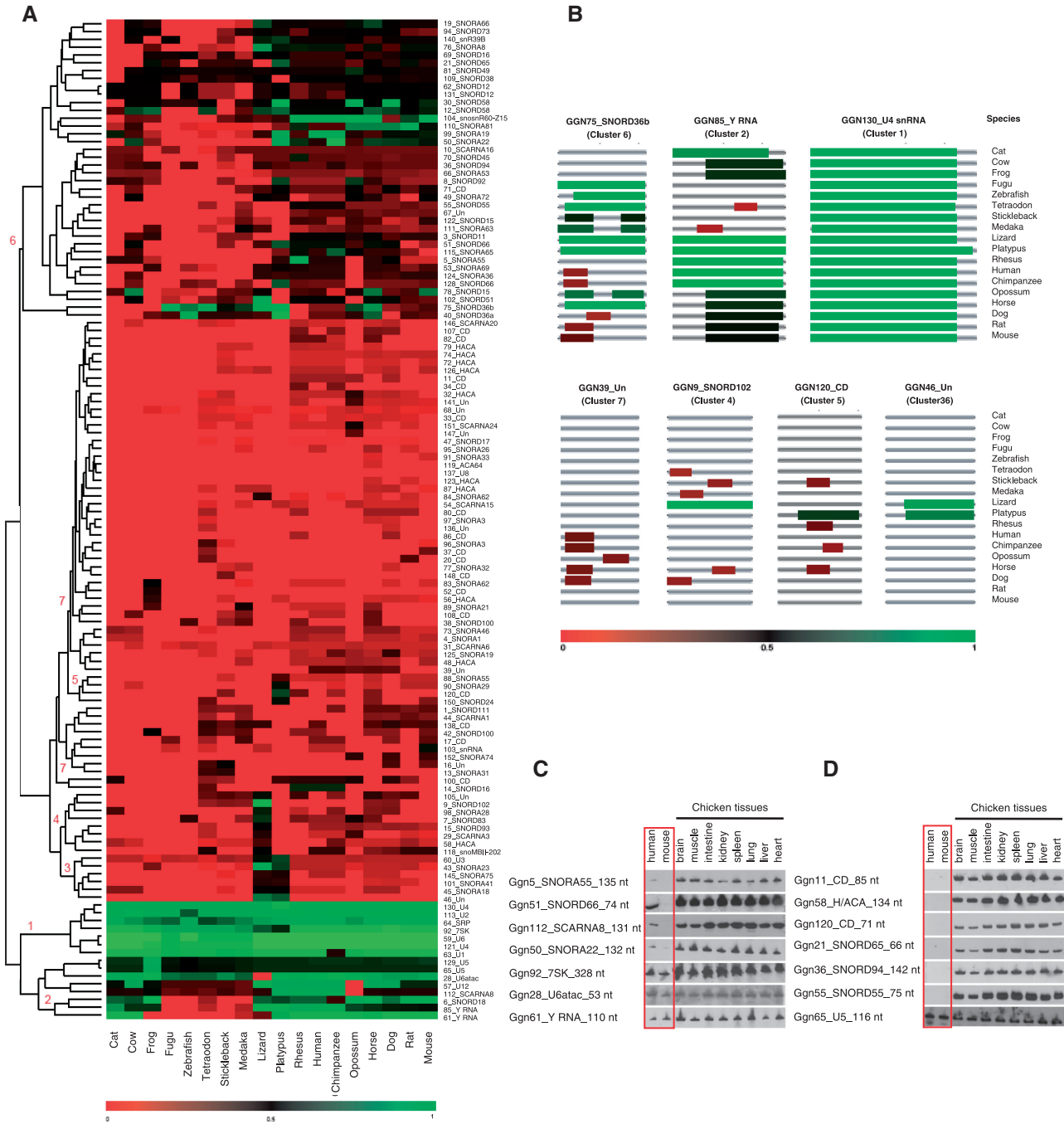
Figure 3. Comparative genomics analysis of 125 chicken ncRNAs. (**A**) Heatmap of the conservations scores of 125 chicken ncRNAs in 18 vertebrate genomes. The conservation degree decreases as color changes from green to red. The ncRNAs were hierarchically clustered based on the conservation scores in the 18 genomes. (**B**) Representatives of ncRNAs from each of the seven clusters. (**C**) Representatives of conserved ncRNAs. The identity of each ncRNA is shown on the left of each blot. The tested chicken tissues are indicated on the upper of the figure. Human and mouse total RNA samples used in the experiments are as described in Figure 2. (**D**) Representatives of less conserved ncRNAs. The U5 snRNA was included as a loading control.

muscle growth rates (fast growing 'broiler' chicken for meat production and slow growing 'layer' chicken for egg production). We found that several ncRNAs were differentially expressed in skeletal muscle tissues between broilers and layers. The expression levels of Ggn68, Ggn91 and Ggn89 were significantly higher in layers than in broilers at post-hatching 2 and 4 weeks (Figure 5C), whereas Ggn44, Ggn4 and Ggn29 showed higher expression in broilers than in layers (Figure 5D). The differential expression patterns of these ncRNAs suggest that they may be important regulators for the divergent muscle growth phenotype of broilers and layers.

Table 2. Eleven snoRNAs with intergenic loci

Ggn_ID	GenBank_ID	Annotation	Reads_no	Length	Sequences
Ggn29	EU240247	ENSGALT00000042330 :SCARNA3	1	143	CTGGAAGCTGTTGGGACCAGTTGGGACCGGTGGTGGGCAA AACAAGTGTTTTGTCCACTCTTGCAGTCTTCCTAGAGT AAAGTGGAGGTCTCTGTCTGCTTTAGGAGAGCCAACT AAACTGGATCGTTCCCCTCCATACATGT
Ggn119[a]	EU240332	RF01225:ACA64	1	134	TTGGCACAGTCTGAAAATCACCTGCGGTCTCCCGTGGCCCC CGGTGAGATAACCGTGCCGACAGCATAGGGAACATCA AAGCTTATTGCCCACGGTGACAGTGTGTGGGGAGTGA AACCCGTGTTCCCATAACC
Ggn38	EU240255	ENSGALT00000042195 :SNORD100	1	72	ATGAAAATTTGGCTCCCTCTACTGAAACAAATGAGGAGAAC TGCCTTAATCGTTACCCAAATCTTCTGAGAT
Ggn11	EU240230	Predicted CD box snoRNA	2	85	TGCTCAATGATGATCATCCTCCTTTGATTACCTGGTGAGGT AATATGAGAGGACATGGAATAATTTCACCGGCGAAC TGAGAGCA
Ggn138	EU240346	Predicted CD box snoRNA	1	58	AGTGCAGTGATCATAAACCAAAGCTGAACATGGAGCTCAGT GTGGTTTATTGTAGAGG
Ggn20	EU240238	Predicted CD box snoRNA	1	61	GCTGGTGATGAGATAGTTATCCCTGTCCGAAACGTTCCTCTG TGGAAGCGTGACTCTGAGG
Ggn37	EU240254	Predicted CD box snoRNA	1	70	CGTCCGATGATGAACCTCAATGCTGTTCACATCCTGACACGC CGTGACGAGCGCTGTCGAGCTGAGGACG
Ggn86	EU240302	Predicted CD box snoRNA	9	96	CGATCCTTCCGGTTCATAGCAAATGATGAATGGGAGTTGCAC GCGGCTGCGTGACGTGTGCGCCCTTTGTTACGACGTGC ACAGCCCCTTCTGAGC
Ggn58	EU240274	Predicted HACA box snoRNA	11	134	CATGCCCCAGTCGTGTTGCAGATATGGCTGTAGTGCCATGTT TGTGTCATTAGGTGGCAGAAAGGAAAAGGCTGTGTCTT TGCTAATGCTCTGAAACCGGTGAGCACTCAGGAATGAC TAGCAACCTGACAAAT
Ggn87	EU240303	Predicted HACA box snoRNA	3	130	CTGCATGTTAATCCAAGAGCTGTGGCTCTGACGTAGCTGC AGGTCTCCAACAACATGCAAGAGCAACGGGAAGGTCT TTGACTGCTCGGCCTCTTCTGCCTGTTGCTGTCACTCAC CCCTCCTATATATT
Ggn126	EU240339	Predicted HACA box snoRNA	1	123	GGCTCGCGCAATTCCAAACCTGACAGTGGTTCTGGTTTGCT GTCAGCCTCATAGAGCAAAAGCGAGGGTTTATTCACT GAAAAGGTGAAGCCGTTTCCTTTTTTCCTTGCCTCCC TCACACTC

[a]Independently predicted by Qu's group (23).

Ggn39 is generated from the *pfkm* gene in a splicing-dependent manner

Mapping the novel transcripts to the chicken genome showed that Ggn39 is located in the second intron of the phosphofructokinase M (*pfkm*) gene (Figure 6A). The size of this intron is 68 nt, whereas the cloned Ggn39 transcript is 67 nt long corresponding to 1–67 bp of intron sequence. Northern blot analysis showed that Ggn39 and its host gene *pfkm* were specifically detected in skeletal and heart muscle tissues, and are most abundantly expressed during postnatal myogenesis (Figure 6B). The similarity in spatio-temporal expression-patterns of the ncRNA and host gene loci suggests that Ggn39 may be generated by intron splicing from the *pfkm* pre-mRNA. We tested this possibility by transfecting the chicken myoblasts with a wild-type *pfkm* minigene (pcDNA3-*pfkm*) comprising exons 1, 2, 3 and introns 1, 2 and the first 37 bp of intron 3, and a mutant *pfkm* minigene containing a mutation in the 5′ splicing site (GT to CT) of the second intron (pcDNA3-*pfkm*-mutant) (Figure 6C). Northern blot analysis demonstrated that the Ggn39 ncRNA was only detected in the myoblasts transfected with the wild-type minigene but not in myoblast transfected with the mutant construct (Figure 6D), suggesting that Ggn39 is generated in a splicing-dependent manner.

DISCUSSION

Increasing numbers of ncRNAs are being identified from various model organisms by experimental RNomics combined with bioinformatics analyses (49,50). Herein, we have performed the first systematic profiling of the chicken intermediate size ncRNA transcriptome by experimental RNomics, and cloned 125 chicken ncRNAs. Comparative genomic analysis with blast against 18 sequenced vertebrate genomes showed that only a few (5 out of 35) of the novel ncRNAs have homologs in non-avian vertebrates, and the remaining novel ncRNAs are likely to be specific to the bird or chicken lineages. Our findings are in line with recent studies on other species. Computer analysis of genomic tiling array data revealed 300 putative candidates for primate specific ncRNAs (51). Unclassified novel ncRNAs often show little sequence conservation and low cellular concentrations (30). Less
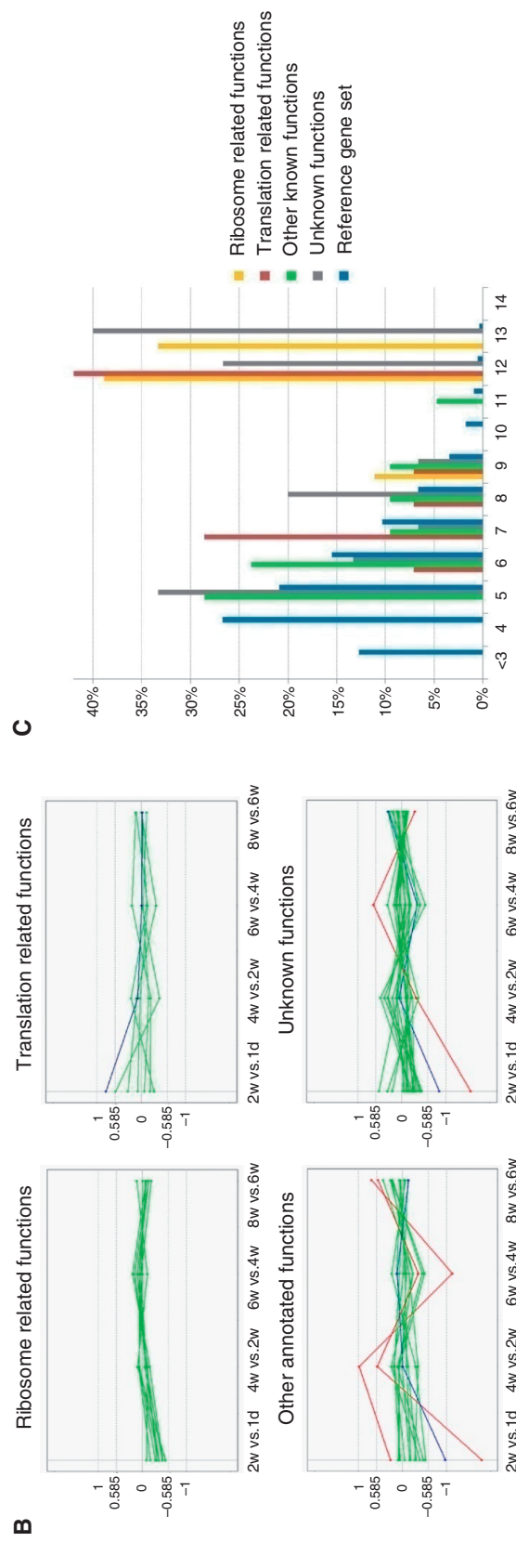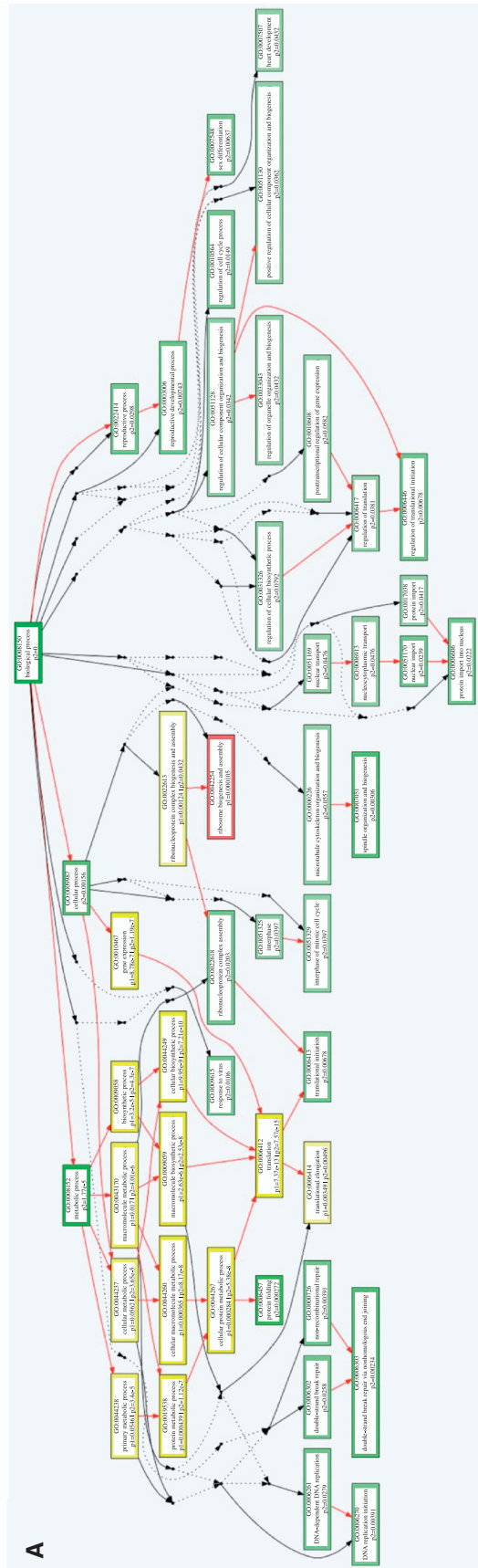
Figure 4. Analysis of snoRNA host genes. (**A**) Functional analysis of snoRNA host genes according to the Biological Process category of Gene Ontology. GO terms enriched among host genes of C/D box snoRNAs and H/ACA box snoRNAs are shown as red and green boxes, respectively, and GO terms enriched among host genes of both C/D box and H/ACA box snoRNAs are shown as yellow boxes. The *P*-values for the enrichment of each GO term are shown within the boxes. Non-enriched GO terms in the hierarchical tree of GO terms are shown as dots to reduce the figure size. (**B**) Expressional changes of snoRNA host gene in chicken skeletal muscle tissue across different developmental stages. The log$_2$ value of the expression difference between post-hatch time points 1 day, 2, 4, 6 and 8 weeks are shown. Genes with expression variance less than 1.5-fold between any two continues time points are shown as green lines, genes with expression variance larger than 1.5-fold but less than 2-fold between any two continues time points are shown as blue lines, and genes with expression variance larger than 2-fold between any two continues time points are shown as red lines. (**C**) Expression of snoRNA host genes in chicken skeletal muscle tissues. The log$_2$ values of the average expression level of each snoRNA host gene across all examined time points are shown on the *X*-axis. The proportions of snoRNA host genes within each expression level range are shown on the *Y*-axis. The expression of all genes detected on the microarray is used as the reference gene set.
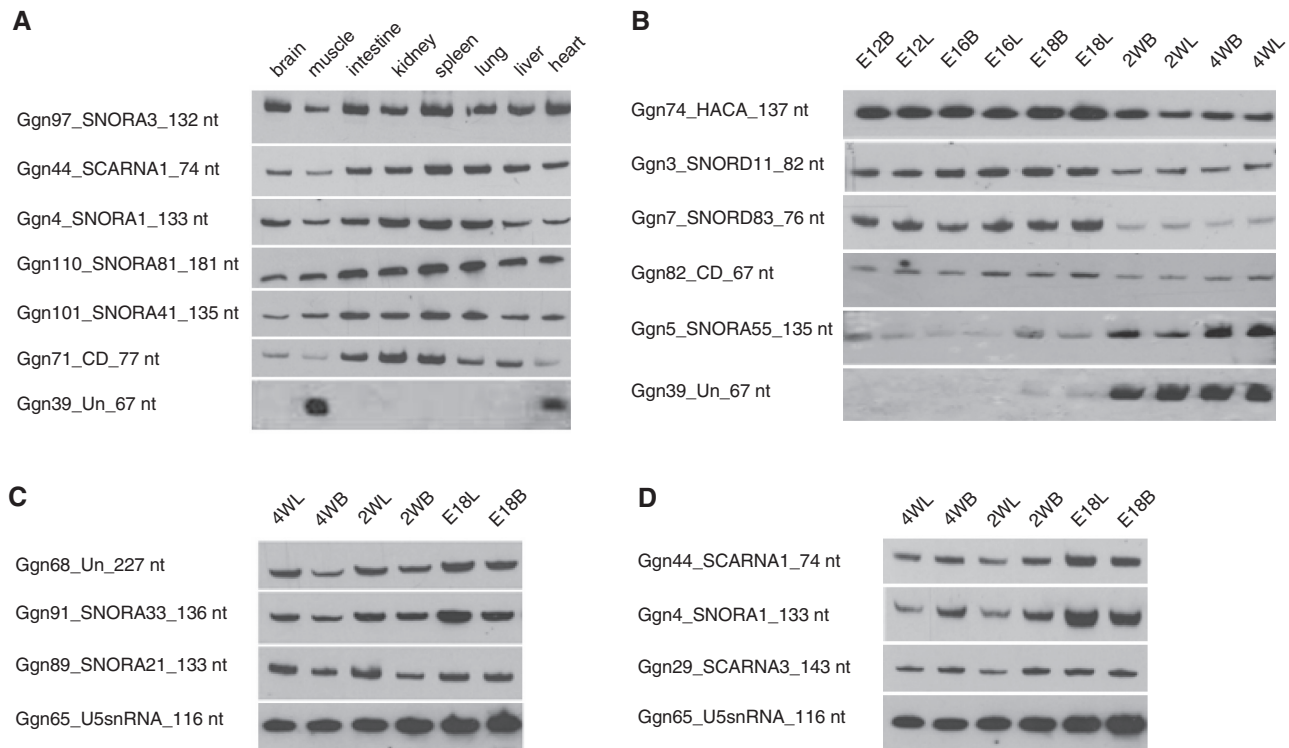
Figure 5. Spatio-temporal ncRNA expression. (**A**) Tissue-specific ncRNA expression. (**B**) ncRNAs with developmentally regulated expression in chicken muscle tissues. The tested samples were from *Pectoralis major* muscles at different developmental stages of broiler (B) and layer (L) chickens, including embryonic Days 12, 16, 18 (E12B, E12L, E16B, E16L, E18B, E18L) and 2 and 4 weeks after hatch (2WB, 2WL, 4WB, 4WL). (**C**) Three ncRNAs with higher expression level in skeletal muscle of layer (L) than of broiler (B). (**D**) Three ncRNAs with higher expression level in skeletal muscle of broiler (B) than of layer (L).

conserved ncRNAs might be important factors in animal development (8), and a remaining challenge for a better understanding of ncRNA function is therefore the detection of less conserved, low abundance ncRNAs in different organisms. Together with previous reports, the identification of chicken specific ncRNAs indicated that there may exist numerous types of ncRNAs that are specific to a lineage or even a single species.

It has been reported that a majority of the identified vertebrate snoRNAs are encoded in introns of protein-coding genes whose functions are mostly related to synthesis, structure or function of the translational apparatus (36). This mode of gene organization and expression provides a molecular basis for functional prediction for recently discovered ncRNA**s.** Similar analyses have been proved invaluable for predicting the functional roles of miRNAs host genes (52). In this study, we have sought to combine GO annotation on host gene function and cellular localizations with experimentally validated expression profiles of snoRNA host genes to predict possible functions of the intronic snoRNAs. We have found that snoRNA host genes encode proteins with a broad spectrum of biological roles ranging from DNA duplication initiation, protein transportation, translation initiation, regulation of cell cycle and other developmental processes. Our results indicate that the host genes of H/ACA have a broader spectrum of cellular locations than do host genes of C/D box

snoRNAs, suggesting that H/ACA and C/D box snoRNAs may function in different cellular processes. This observation was further supported by the microarray data on host gene expression during skeletal muscle development. snoRNA host genes with functions related to ribosome biogenesis and translation were much more abundantly expressed during myogenesis than host genes with other functions (such as cell cycle and apoptosis), which showed lower and more variable expression across the developmental stages. Our findings could provide information for further studies of the functions of snoRNAs from host genes with different biological functions, as well as for the study of the functional and regulatory coordination between snoRNAs and their host genes during development.

The spatio-temporally regulated expression patterns of the novel ncRNAs provide further functional information about the role of these ncRNAs during development. It is well accepted that some ncRNAs, especially some microRNAs and piRNAs, have specific spatio-temporal expression patterns and play functional roles in cell differentiation and organogenesis during development. For instance, miRNAs with muscle-specific expression are involved in development and disease of the skeletal and heart muscle (53,54). In general, the spatio-temporally regulated expression of other classes of ncRNAs, particularly the unclassified ncRNAs, has been less well explored. We investigated this question by northern blot
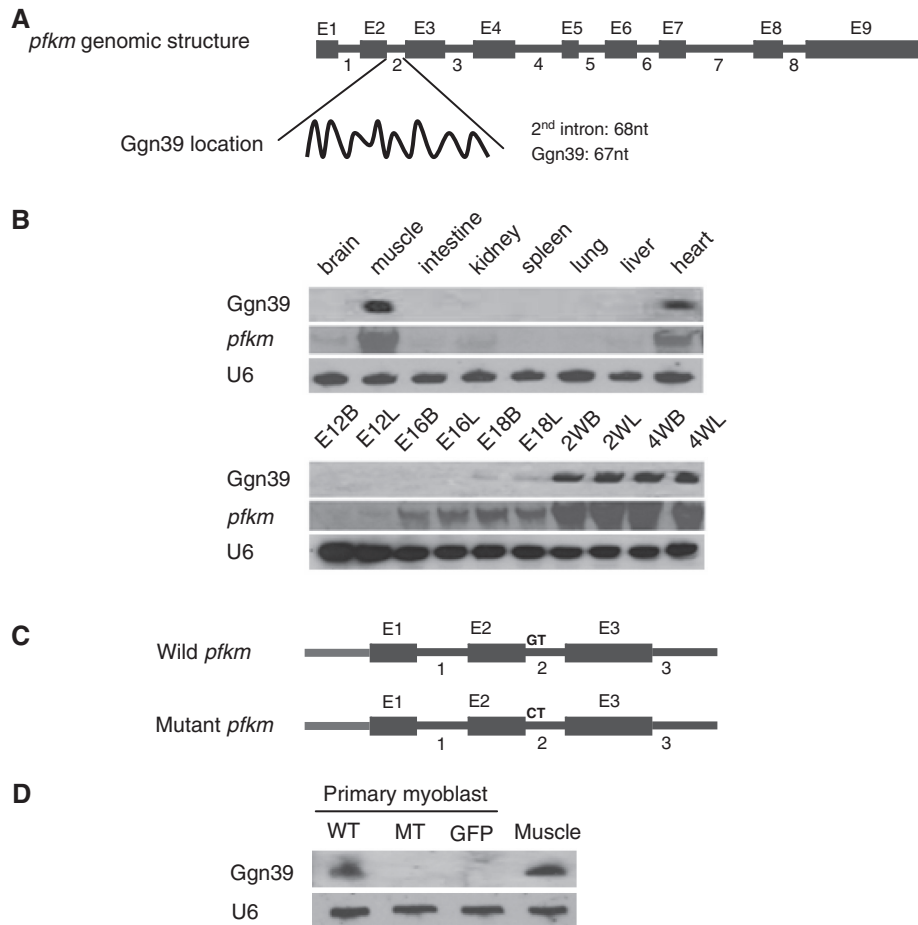
Figure 6. Ggn39 is generated in a splicing-dependent manner. (**A**) The genomic organization of Ggn39. Boxes and lines represent exons (E1–E9) and introns (1–8), respectively, of the *pfkm* gene. (**B**) The spatio-temporal expression pattern of Ggn39 and its host gene *pfkm*. The expression patterns of Ggn39 and *pfkm* were examined by northern blot analysis of total RNA from chicken tissues. The corresponding antisense RNA labeled with DIG-11-UTP were used as probes. U6 snRNA was used as loading control. (**C**) Schematic structure of the *pfkm* minigene and the second intron 5′ splicing site mutant. (**D**) The biogenesis of Ggn39 is dependent on splicing of the second intron of the *pfkm* gene. Chicken primary myoblasts were transfected with wild-type (WT) and mutant (MT) *pfkm* minigenes. GFP denotes pcDNA3-EGFP that served as transfection control. Northern blot analysis was performed on total RNAs prepared from the transfected cells. As the Ggn39 is highly expressed in skeletal muscle, chicken skeletal muscle RNA was used as positive control for the northern blot experiments.

analysis of a substantial number of the ncRNAs identified in the present work. Our data showed that the majority of the novel ncRNAs exhibited ubiquitous tissue distribution; some were abundantly expressed only in a few tissues. Most interestingly, we found that one of the unclassified ncRNAs (Ggn39) was generated from the second intron of *pfkm* gene in a splicing-dependent manner, and that Ggn39 ncRNA, like its host gene *pfkm*, showed specific expression in the chicken heart and skeletal muscle cells, especially during postnatal development. As a key enzyme in the glucose metabolism, PFKM plays important roles in the metabolism and development of skeletal muscle (55). The coordinated expression pattern of Ggn39 and its host gene suggests a functional correlation between Ggn39 and PFKM which may provides useful information for further studies of the metabolic and developmental functions of Ggn39 and its host gene *pfkm*.

Recently, brain-specific snoRNAs have been implicated in brain development and neurological disease (41,56,57).

In chicken we identified an H/ACA-box snoRNA (Ggn146) encoded in the intron of the chicken ortholog of the Williams Beuren syndrome chromosome region 22 gene. Since Williams Beuren syndrome is caused by microdeletion within the syntenic region containing this gene in human (58), it is an intriguing possibility that an ortholog of Ggn146 may also be involved in the pathogenesis of the Williams Beuren syndrome in human. The broiler and layer chickens are an excellent model systems for the study of molecular mechanisms involved in skeletal muscle development because of their significant muscle growth differences during myogenesis. Up to now, the molecular genetics underlying the divergent muscle growth rate between broilers and layers have remained unclear. Similar to muscle-specific miRNAs (e.g. miRNA-1), ncRNAs that are differentially expressed in skeletal muscle tissues between broiler and layer chickens might function as muscle growth regulators. Identification of these ncRNAs could facilitate ncRNA functional studies and provide essential

information for a better understanding of the functional roles of ncRNAs during development.

ACCESSION NUMBERS

EU240220, EU240222–EU240239, EU240246–EU240250, EU240252–EU240263, EU240265–EU240328, EU240331–EU240352, EU240354–EU240355, EU240357.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Gibbs,R.A., Weinstock,G.M., Metzker,M.L., Muzny,D.M., Sodergren,E.J., Scherer,S., Scott,G., Steffen,D., Worley,K.C., Burch,P.E. *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **428**, 493–521.
2. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
3. Waterston,R.H., Lindblad-Toh,K., Birney,E., Rogers,J., Abril,J.F., Agarwal,P., Agarwala,R., Ainscough,R., Alexandersson,M., An,P. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
4. Hillier,L.W., Miller,W., Birney,E., Warren,W., Hardison,R.C., Ponting,C.P., Bork,P., Burt,D.W., Groenen,M.A., Delany,M.E. *et al.* (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, **432**, 695–716.
5. Cheng,J., Kapranov,P., Drenkow,J., Dike,S., Brubaker,S., Patel,S., Long,J., Stern,D., Tammana,H., Helt,G. *et al.* (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, **308**, 1149–1154.
6. Kapranov,P., Cawley,S.E., Drenkow,J., Bekiranov,S., Strausberg,R.L., Fodor,S.P. and Gingeras,T.R. (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science*, **296**, 916–919.
7. Bertone,P., Stolc,V., Royce,T.E., Rozowsky,J.S., Urban,A.E., Zhu,X., Rinn,J.L., Tongprasit,W., Samanta,M., Weissman,S. *et al.* (2004) Global identification of human transcribed sequences with genome tiling arrays. *Science*, **306**, 2242–2246.
8. Birney,E., Stamatoyannopoulos,J.A., Dutta,A., Guigo,R., Gingeras,T.R., Margulies,E.H., Weng,Z., Snyder,M., Dermitzakis,E.T., Thurman,R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
9. Huttenhofer,A., Kiefmann,M., Meier-Ewert,S., O'Brien,J., Lehrach,H., Bachellerie,J.P. and Brosius,J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, **20**, 2943–2953.
10. Numata,K., Kanai,A., Saito,R., Kondo,S., Adachi,J., Wilming,L.G., Hume,D.A., Hayashizaki,Y. and Tomita,M. (2003) Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res.*, **13**, 1301–1306.
11. Mercer,T.R., Dinger,M.E., Sunkin,S.M., Mehler,M.F. and Mattick,J.S. (2008) Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. USA*, **105**, 716–721.
12. Aravin,A., Gaidatzis,D., Pfeffer,S., Lagos-Quintana,M., Landgraf,P., Iovino,N., Morris,P., Brownstein,M.J., Kuramochi-Miyagawa,S., Nakano,T. *et al.* (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, **442**, 203–207.
13. Girard,A., Sachidanandam,R., Hannon,G.J. and Carmell,M.A. (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, **442**, 199–202.
14. Ruby,J.G., Jan,C., Player,C., Axtell,M.J., Lee,W., Nusbaum,C., Ge,H. and Bartel,D.P. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell*, **127**, 1193–1207.
15. Storz,G., Altuvia,S. and Wassarman,K.M. (2005) An abundance of RNA regulators. *Annu. Rev. Biochem.*, **74**, 199–217.
16. Plasterk,R.H. (2006) Micro RNAs in animal development. *Cell*, **124**, 877–881.
17. Kiss,T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.*, **20**, 3617–3622.
18. Prasanth,K.V. and Spector,D.L. (2007) Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. *Genes Dev.*, **21**, 11–42.
19. Kishore,S. and Stamm,S. (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*, **311**, 230–232.
20. Couzin,J. (2008) MicroRNAs make big impression in disease after disease. *Science*, **319**, 1782–1784.
21. Burt,D.W., Bruley,C., Dunn,I.C., Jones,C.T., Ramage,A., Law,A.S., Morrice,D.R., Paton,I.R., Smith,J., Windsor,D. *et al.* (1999) The dynamics of chromosome evolution in birds and mammals. *Nature*, **402**, 411–413.
22. Hubbard,T.J., Aken,B.L., Ayling,S., Ballester,B., Beal,K., Bragin,E., Brent,S., Chen,Y., Clapham,P., Clarke,L. *et al.* (2009) Ensembl 2009. *Nucleic Acids Res.*, **37**, D690–D697.
23. Shao,P., Yang,J.H., Zhou,H., Guan,D.G. and Qu,L.H. (2009) Genome-wide analysis of chicken snoRNAs provides unique implications for the evolution of vertebrate snoRNAs. *BMC Genomics*, **10**, 86.
24. Shao,P., Zhou,H., Xiao,Z.D., He,J.H., Huang,M.B., Chen,Y.Q. and Qu,L.H. (2008) Identification of novel chicken microRNAs and analysis of their genomic organization. *Gene*, **418**, 34–40.
25. Burnside,J., Ouyang,M., Anderson,A., Bernberg,E., Lu,C., Meyers,B.C., Green,P.J., Markis,M., Isaacs,G., Huang,E. *et al.* (2008) Deep sequencing of chicken microRNAs. *BMC Genomics*, **9**, 185.
26. Glazov,E.A., Cottee,P.A., Barris,W.C., Moore,R.J., Dalrymple,B.P. and Tizard,M.L. (2008) A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. *Genome Res.*, **18**, 957–964.
27. Yuan,G., Klambt,C., Bachellerie,J.P., Brosius,J. and Huttenhofer,A. (2003) RNomics in Drosophila melanogaster: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res.*, **31**, 2495–2507.
28. Eddy,S.R. (2002) Computational genomics of noncoding RNA genes. *Cell*, **109**, 137–140.
29. Vogel,J., Bartels,V., Tang,T.H., Churakov,G., Slagter-Jager,J.G., Huttenhofer,A. and Wagner,E.G. (2003) RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.*, **31**, 6435–6443.
30. Deng,W., Zhu,X., Skogerbo,G., Zhao,Y., Fu,Z., Wang,Y., He,H., Cai,L., Sun,H., Liu,C. *et al.* (2006) Organization of the Caenorhabditis elegans small non-coding transcriptome: genomic features, biogenesis, and expression. *Genome Res.*, **16**, 20–29.
31. Vitali,P., Royo,H., Seitz,H., Bachellerie,J.P., Huttenhofer,A. and Cavaille,J. (2003) Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res.*, **31**, 6543–6551.
32. Maxwell,E.S. and Fournier,M.J. (1995) The small nucleolar RNAs. *Annu. Rev. Biochem.*, **64**, 897–934.

33. Zemann,A., op de Bekke,A., Kiefmann,M., Brosius,J. and Schmitz,J. (2006) Evolution of small nucleolar RNAs in nematodes. *Nucleic Acids Res.*, **34**, 2676–2685.
34. Balakin,A.G., Smith,L. and Fournier,M.J. (1996) The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell*, **86**, 823–834.
35. Tycowski,K.T., Aab,A. and Steitz,J.A. (2004) Guide RNAs with 5′ caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. *Curr. Biol.*, **14**, 1985–1995.
36. Kiss,T., Fayet,E., Jady,B.E., Richard,P. and Weber,M. (2006) Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harb. Symp. Quant. Biol.*, **71**, 407–417.
37. Filipowicz,W. and Pogacic,V. (2002) Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol.*, **14**, 319–327.
38. Semenov,D.V., Vratskih,O.V., Kuligina,E.V. and Richter,V.A. (2008) Splicing by exon exclusion impaired by artificial box c/d RNA targeted to branch-point adenosine. *Ann. NY Acad. Sci.*, **1137**, 119–124.
39. Lukowiak,A.A., Narayanan,A., Li,Z.H., Terns,R.M. and Terns,M.P. (2001) The snoRNA domain of vertebrate telomerase RNA functions to localize the RNA within the nucleus. *RNA*, **7**, 1833–1844.
40. Nicholls,R.D. and Knepper,J.L. (2001) Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annu. Rev. Genomics Hum. Genet.*, **2**, 153–175.
41. Cavaille,J., Buiting,K., Kiefmann,M., Lalande,M., Brannan,C.I., Horsthemke,B., Bachellerie,J.P., Brosius,J. and Huttenhofer,A. (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci. USA*, **97**, 14311–14316.
42. Klein,R.J. and Eddy,S.R. (2003) RSEARCH: finding homologs of single structured RNA sequences. *BMC Bioinformatics*, **4**, 44.
43. Hertel,J., Hofacker,I.L. and Stadler,P.F. (2008) SnoReport: computational identification of snoRNAs with unknown targets. *Bioinformatics*, **24**, 158–164.
44. Kent,W.J. (2002) BLAT–the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
45. Zheng,Q. and Wang,X.J. (2008) GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Res.*, **36**, W358–W363.
46. Yang,W., Zhang,Y., Ma,G., Zhao,X., Chen,Y. and Zhu,D. (2005) Identification of gene expression modifications in myostatin-stimulated myoblasts. *Biochem. Biophys. Res. Commun.*, **326**, 660–666.
47. Gardner,P.P., Daub,J., Tate,J.G., Nawrocki,E.P., Kolbe,D.L., Lindgreen,S., Wilkinson,A.C., Finn,R.D., Griffiths-Jones,S., Eddy,S.R. *et al.* (2009) Rfam: updates to the RNA families database. *Nucleic Acids Res.*, **37**, D136–D140.
48. Zheng,Q., Zhang,Y., Chen,Y., Yang,N., Wang,X.J. and Zhu,D. (2009) Systematic identification of genes involved in divergent skeletal muscle growth rates of broiler and layer chickens. *BMC Genomics*, **10**, 87.
49. Pang,K.C., Stephen,S., Dinger,M.E., Engstrom,P.G., Lenhard,B. and Mattick,J.S. (2007) RNAdb 2.0–an expanded database of mammalian non-coding RNAs. *Nucleic Acids Res.*, **35**, D178–D182.
50. Lestrade,L. and Weber,M.J. (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.*, **34**, D158–D162.
51. Zhang,Z., Pang,A.W. and Gerstein,M. (2007) Comparative analysis of genome tiling array data reveals many novel primate-specific functional RNAs in human. *BMC Evol. Biol.*, **7 (Suppl. 1)**, S14.
52. Baskerville,S. and Bartel,D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
53. Zhao,Y., Ransom,J.F., Li,A., Vedantham,V., von Drehle,M., Muth,A.N., Tsuchihashi,T., McManus,M.T., Schwartz,R.J. and Srivastava,D. (2007) Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell*, **129**, 303–317.
54. Chen,J.F., Mandel,E.M., Thomson,J.M., Wu,Q., Callis,T.E., Hammond,S.M., Conlon,F.L. and Wang,D.Z. (2006) The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat. Genet.*, **38**, 228–233.
55. Wegener,G. and Krause,U. (2002) Different modes of activating phosphofructokinase, a key regulatory enzyme of glycolysis, in working vertebrate muscle. *Biochem. Soc. Trans.*, **30**, 264–270.
56. Mehler,M.F. and Mattick,J.S. (2007) Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiol. Rev.*, **87**, 799–823.
57. Peters,J. (2008) Prader-Willi and snoRNAs. *Nat. Genet.*, **40**, 688–689.
58. Pober,B.R., Johnson,M. and Urban,Z. (2008) Mechanisms and treatment of cardiovascular disease in Williams-Beuren syndrome. *J. Clin. Invest.*, **118**, 1606–1615.