

# UC Riverside

## UC Riverside Previously Published Works

### Title

Conservation and Losses of Non-Coding RNAs in Avian Genomes

### Permalink

<https://escholarship.org/uc/item/0d75w22r>

### Journal

PLOS ONE, 10(3)

### ISSN

1932-6203

### Authors

Gardner, Paul P  
Fasold, Mario  
Burge, Sarah W  
[et al.](#)

### Publication Date

2015

### DOI

10.1371/journal.pone.0121797

Peer reviewed

RESEARCH ARTICLE

# Conservation and Losses of Non-Coding RNAs in Avian Genomes

Paul P. Gardner<sup>1,2\*</sup>, Mario Fasold<sup>3,4</sup>, Sarah W. Burge<sup>5</sup>, Maria Ninova<sup>6</sup>, Jana Hertel<sup>3</sup>, Stephanie Kehr<sup>3</sup>, Tammy E. Steeves<sup>1</sup>, Sam Griffiths-Jones<sup>6</sup>, Peter F. Stadler<sup>3,7,8,9,10,11,12</sup>

**1** School of Biological Sciences, University of Canterbury, Christchurch, New Zealand, **2** Biomolecular Interaction Centre, University of Canterbury, Christchurch, New Zealand, **3** Bioinformatics Group, Department of Computer Science; and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstrasse 16-18, D-04107 Leipzig, Germany, **4** ecSeq Bioinformatics, Brandvorwerkstr.43, D-04275 Leipzig, Germany, **5** European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD, UK, **6** Faculty of Life Sciences, University of Manchester, Manchester, United Kingdom, **7** Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany, **8** Fraunhofer Institute for Cell Therapy and Immunology, Perlickstrasse 1, D-04103 Leipzig, Germany, **9** Department of Theoretical Chemistry of the University of Vienna, Währingerstrasse 17, A-1090 Vienna, Austria, **10** Center for RNA in Technology and Health, Univ. Copenhagen, Grønnegårdsvej 3, Frederiksberg C, Denmark, **11** Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501, USA, **12** German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany

\* [paul.gardner@canterbury.ac.nz](mailto:paul.gardner@canterbury.ac.nz)



OPEN ACCESS

**Citation:** Gardner PP, Fasold M, Burge SW, Ninova M, Hertel J, Kehr S, et al. (2015) Conservation and Losses of Non-Coding RNAs in Avian Genomes. *PLoS ONE* 10(3): e0121797. doi:10.1371/journal.pone.0121797

**Academic Editor:** Leonardo Mariño-Ramírez, National Institutes of Health, UNITED STATES

**Received:** November 3, 2014

**Accepted:** February 3, 2015

**Published:** March 30, 2015

**Copyright:** © 2015 Gardner et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All the sequencing data and assemblies are available from the BGI phylogenomics analysis of birds website (<http://phybirds.genomics.org.cn/index.jsp>). The genome annotations and scripts produced by the work presented in this article are available from github (<https://github.com/ppgardne/bird-genomes>).

**Funding:** PPG is supported by a Rutherford Discovery Fellowship, administered by the Royal Society of New Zealand. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Here we present the results of a large-scale bioinformatics annotation of non-coding RNA loci in 48 avian genomes. Our approach uses probabilistic models of hand-curated families from the Rfam database to infer conserved RNA families within each avian genome. We supplement these annotations with predictions from the tRNA annotation tool, tRNAscan-SE and microRNAs from miRBase. We identify 34 lncRNA-associated loci that are conserved between birds and mammals and validate 12 of these in chicken. We report several intriguing cases where a reported mammalian lncRNA, but not its function, is conserved. We also demonstrate extensive conservation of classical ncRNAs (e.g., tRNAs) and more recently discovered ncRNAs (e.g., snoRNAs and miRNAs) in birds. Furthermore, we describe numerous “losses” of several RNA families, and attribute these to either genuine loss, divergence or missing data. In particular, we show that many of these losses are due to the challenges associated with assembling avian microchromosomes. These combined results illustrate the utility of applying homology-based methods for annotating novel vertebrate genomes.

## Introduction

Non-coding RNAs (ncRNAs) are an important class of genes, responsible for the regulation of many key cellular functions. The major RNA families include the classical, highly conserved RNAs, sometimes called “molecular fossils”, such as the transfer RNAs, ribosomal RNAs, RNA components of RNase P and the signal recognition particle [1]. Other classes appear to have

**Competing Interests:** The authors have declared that no competing interests exist.

have evolved more recently, e.g. the small nucleolar RNAs (snoRNAs), microRNAs (miRNAs) and the long non-coding RNAs (lncRNAs) [2].

The ncRNAs pose serious research challenges, particularly for the field of genomics. For example, they lack the strong statistical signals associated with protein coding genes, e.g. open reading frames, G+C content and codon-usage biases [3].

New sequencing technologies have dramatically expanded the rate at which ncRNAs are discovered and their functions are determined [4]. However, in order to determine the full range of ncRNAs across multiple species we require multiple RNA fractions (e.g. long and short), in multiple species, in multiple developmental stages and tissues types. The costs of this approach are still prohibitive in terms of researcher-time and finances. Consequently, in this study we concentrate on bioinformatic approaches, primarily we use homology-based methods (i.e. covariance models (CMs)). We validate the majority of these predictions using RNA-seq. The CM-based approach that we favour, remain state of the art for ncRNA bioinformatic analyses, as they capture both sequence as well as secondary structure constraints on RNAs [5–7]. This has been shown to improve both the sensitivity and specificity rates for homology assignment [8]. For example, the CM based approach for annotating ncRNAs in genomes requires reliable alignments and consensus secondary structures of representative sequences of RNA families, many of which can be found at Rfam [9–14]. These are used to train probabilistic models that score the likelihood that a database sequence is generated by the same evolutionary processes as the training sequences based upon both sequence and structural information [5–7]. The tRNAscan-SE software package uses CMs to accurately predict transfer RNAs [15, 16].

Independent benchmarks of bioinformatic annotation tools have shown that the CM approaches out-perform alternative methods [8], although their sensitivity can be limited for rapidly evolving families such as vault RNAs or telomerase RNA [17].

The publication of 48 avian genomes, including the previously published chicken [18], zebra finch [19] and turkey [20] with the recently published 45 avian genomes [21–27], provides an exciting opportunity to explore conservation of genomic loci that have been associated with ncRNAs in unprecedented detail.

In the following we explore the conservation patterns of the major classes of avian ncRNA loci in further detail. Using homology search tools and evolutionary constraints, we have produced a set of genome annotations for 48 predominantly non-model bird species for ncRNAs that are conserved across the avian species. This conservative set of annotations is expected to contain the core avian ncRNA loci. We focus our report on the unusual results within the avian lineages. These are either unexpectedly well-conserved ncRNAs or unexpectedly poorly-conserved ncRNAs. The former are ncRNA loci that were not expected to be conserved between the birds and the other vertebrates, particularly those ncRNAs whose function is not conserved in birds. The latter are apparent losses of ncRNA loci expected to be conserved; Here, we consider three categories of such “loss”: First, genuine gene losses in the avian lineage where ncRNAs well conserved in other vertebrates are completely absent in birds. Second, “divergence” where ncRNAs have undergone such significant sequence and structural alternations that homology search tools can no longer detect a relationship between other vertebrate exemplars and avian varieties.

Third, “missing” ncRNAs that failed to be captured in the available, largely fragmented, avian genomes. The avian karyotype is characterized by a large number of chromosomes (average  $2n \approx 80$ ) generally consisting of approximately 5 larger “macrochromosomes” and many smaller “microchromosomes” [28–30]. The presence of microchromosomes presents significant assembly challenges [18, 20, 31]. Indeed, of the 48 published avian genomes, 20 of which are high-coverage ( $> 50X$ ), only two were relatively complete chromosomal assemblies when this study was initiated (chicken, zebra finch; [19, 21]) (Chromosomal assemblies of turkey

(NCBI GCF\_000146605.1) and flycatcher (NCBI GCA\_000247815.2) were recently made available). We therefore expect that many ncRNAs in comparative avian genome studies will be missing from the genome assemblies due to microchromosome assembly difficulties.

## Materials and Methods

The 48 bird genome sequences used for the following analyses are available from the phylogenomics analysis of birds website [32, 33].

Bird genomes were searched using the cmsearch program from INFERNAL 1.1 [34, 35] and the covariance models (CMs) from the Rfam database v11.0 [12, 13]. All matches above the curated GA threshold were included. Subsequently, all hits with an E-value greater than  $5 \times 10^{-4}$  were discarded, so only matches which passed the Rfam-curated, model-specific GA threshold, and had an E-value smaller than  $5 \times 10^{-4}$  were retained. The Rfam database classifies non-coding RNAs into hierarchical groupings. The basic units are “families” which are groups of homologous, alignable sequences; “clans” which are groups of un-alignable (or functionally distinct), homologous families; and “classes” which are groups of clans and families with related biological functions e.g. spliceosomal RNAs, miRNAs and snoRNAs [12]; these categories have been used to classify our results.

In order to obtain good annotations of tRNA genes we ran the specialist tRNA-scan version 1.3.1 annotation tool. This method also uses covariance models to identify tRNAs. However it also uses some heuristics to increase the search-speed, annotates the Isoacceptor Type of each prediction and uses sequence analysis to infer if predictions are likely to be functional or tRNA-derived pseudogenes [15, 16].

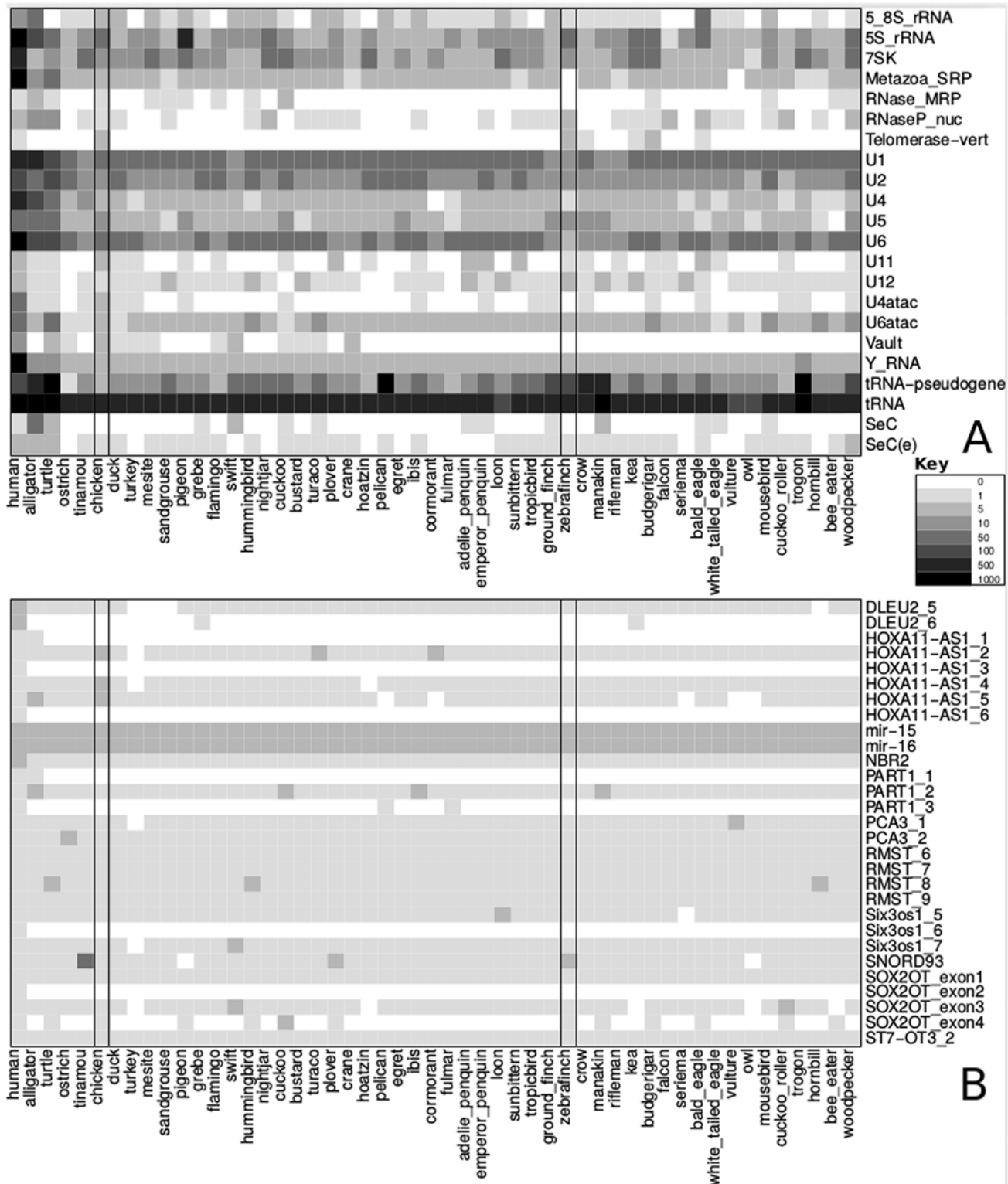
Rfam matches and the tRNA-scan results for families belonging to the same clan were then “competed” so that only the best match was retained for any genomic region [12]. To further increase the specificity of our annotations we filtered out families that were identified in < 10% of the avian genomes that we have analyzed in this work. These filtered families largely corresponded to bacterial contamination or species/clade-specific lncRNAs, miRNAs and snoRNAs that have a high evolutionary turn-over (Fig. O in [S1 Results](#)) [2, 36, 37].

999 microRNA sequence families, previously annotated in at least one vertebrate, were retrieved from miRBase (v19) [38]. Individual sequences or multiple sequence alignments were used to build covariance models with INFERNAL (v1.1rc3) [34, 35], and these models were searched against the 48 bird genomes, and the genomes of the American alligator and the green turtle as out-groups. Hits with E-value < 10 realigned with the query sequences and the resultant multiple sequence alignments manually inspected and edited using RALEE [39]. Those sequences that did not match the characteristics of a microRNA (conserved seed sequence and hairpin secondary structure) were removed from further analysis.

An additional snoRNA homology search was performed with snoStrip [40]. As initial queries we used deuterostomian snoRNA families from human [41], platypus [42], and chicken [43].

The diverse sets of genome annotations were combined and filtered, ensuring conservation in 10% or more of the avian genomes. We collapsed the remaining overlapping annotations into a single annotation. We also generated heatmaps for different groups of ncRNA genes (see [Fig. 1](#) and Figs. A-C in [S1 Results](#)). All the scripts and annotations presented here are available from Github [44].

Chicken ncRNA predictions were validated using two separate RNA-seq data sets (IDs are available in Table C in [S1 Results](#)). The first data set (Bioproject PRJNA204941) contains 971 million reads and comprises 27 samples from 14 different chicken tissues sequenced on Illumina HiSeq2000 using a small RNA-seq protocol [45]. The second data set (SRA accession



**Fig 1. Heatmaps showing the presence/absence and approximate genomic copy-number of “lost, divergent or missing RNAs” and the “unusually, well conserved RNAs”.** On the top we show the families that have been identified as surprising RNA losses, divergence or missing data. In several cases functionally related families have also been included, e.g. the RNA components of the major and minor spliceosomes: U1, U2, U4, U5 and U6; and U11, U12, U4atac, U5 and U6atac, respectively. Below we show the unusually, well conserved RNAs, these are predominantly lincRNAs.

doi:10.1371/journal.pone.0121797.g001

SRP041863) contains 1,46 billion Illumina HiSeq reads sequenced from whole chicken embryo RNA from 7 stages using a strand-specific dUTP protocol [45]. The raw reads were checked for quality and adapters clipped if required by the protocol. Preprocessed reads were mapped to the galGal4 reference genome using SEGEMEHL (version 0.1.9) short read aligner [46] and then overlapped with the ncRNA annotations under consideration of strand information.

## Results

There is substantial gain and loss of lncRNAs and other ncRNA associated loci over evolutionary time [2, 36, 37]. It is difficult to assess how many of these “gains” and “losses” are due to limited bioinformatic sequence alignment tools (these generally fail align correctly below 60–50% sequence identity [47]) or due to genuine gains and losses or data missing from the current genome assemblies. Nevertheless, sequence conservation, generally speaking, provides useful evidence for gene and function conservation.

We have identified 66,879 loci in 48 avian genomes that share sequence similarity with previously characterized ncRNAs and are conserved in > 10% of these avian genomes. These loci have been classified into 626 different families, the majority of which correspond to miRNAs and snoRNAs (summarized in Table 1). Out of necessity we have selected a modest number of families for further discussion. These include the lncRNAs that appear to be conserved between Mammals and Aves and the cases of apparent loss of genes that conserved in most other Vertebrates. The supplementary material (S1 Results) contains further discussions of RNA elements.

**Table 1. A summary of ncRNA genes in human, chicken and all bird genomes.** This table contains the total number of annotated ncRNAs from different RNA types in human, the median number for each of the 48 birds and chicken. The number of chicken ncRNA that show evidence for expression is also indicated (the percentage is given in parentheses). The threshold for determining expression was selected based upon a false positive rate of less than 10%.

ncRNA genes in human, chicken and all bird genomes				
Number in human	median(48 birds)	Number in chicken	Chicken ncRNAs confirmed with RNA-seq	RNA type
62	25.0	34	12 (35.3%)	Long non-coding RNA
356	499.5	427	280 (65.6%)	microRNA
281	120.0	106	90 (84.9%)	C/D box snoRNA
336	85.5	68	48 (70.6%)	H/ACA box snoRNA
34	13.0	12	12 (100.0%)	Small cajal body RNA
1754	48.5	71	32 (45.1%)	Major spliceosomal RNA
58	3.0	6	3 (50.0%)	Minor spliceosomal RNA
525	82.0	122	88 (72.1%)	Cis-regulatory element
316	6.5	9	3 (33.3%)	7SK RNA
1	0.0	2	0 (0.0%)	Telomerase RNA
9	0.0	2	1 (50.0%)	Vault RNA
892	3.0	3	2 (66.7%)	Y RNA
1084	173.5	300	278 (92.7%)	Transfer RNA
80	9.5	4	2 (50.0%)	Transfer RNA pseudogene
941	3.0	4	2 (50.0%)	SRP RNA
607	7.0	22	10 (45.5%)	Ribosomal RNA
4	1.0	2	2 (100.0%)	RNase P/MRP RNA
7340	1080.0	1194	865 (72.4%)	Total

doi:10.1371/journal.pone.0121797.t001

## Unusually well conserved RNAs

The bulk of the “unusually well conserved RNAs” belong to the long non-coding RNA (lncRNA) group. The lncRNAs are a diverse group of RNAs that have been implicated in a multitude of functional processes [48–51]. These RNAs have largely been characterized in mammalian species, particularly human and mouse and have been shown to be rapidly turned-over by evolutionary processes [37]. Consequently, we generally do not expect these to be conserved outside of Mammals. Notable examples include Xist [52] and H19 [53]. There is emerging evidence for the conservation of “mammalian” lncRNAs in Vertebrates [54, 55]), however, like most lncRNAs, the function of these lncRNAs remains largely unknown. Here, we show the conservation of several lncRNAs that have been well-characterized in humans.

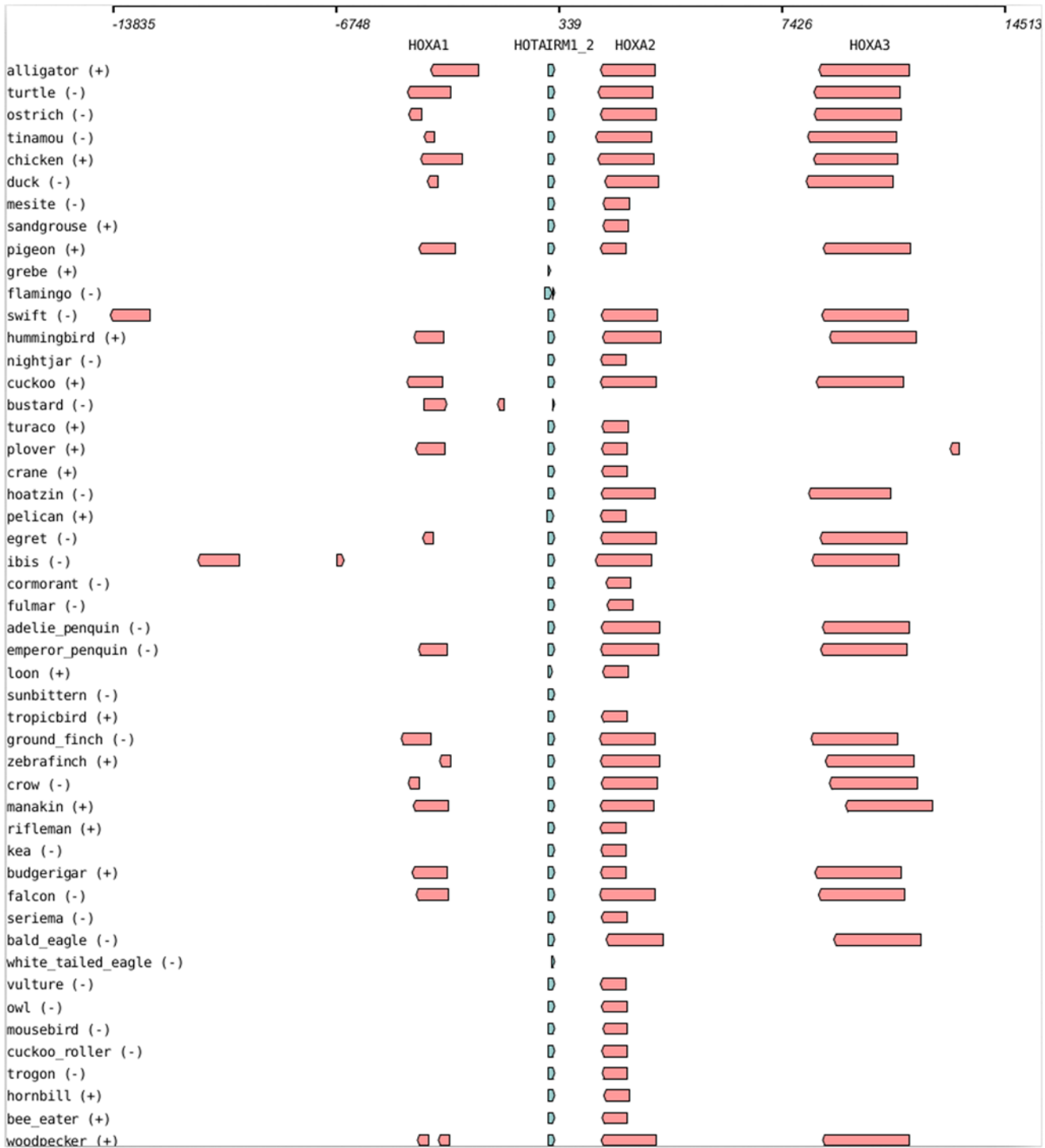
The CM based approach is appropriate for most classes of ncRNA, but the lncRNAs are a particular challenge [50]. CMs cannot model the exon-intron structures of spliced lncRNAs, nor do they deal elegantly with the repeats that many lncRNAs host. Consequently in the latest release of Rfam the lncRNA families that were added were composed of local conserved (and possibly structured elements) within lncRNAs, analogous to the “domains” housed within protein sequences [13]. Whilst some these regions may not reflect functional RNA elements but instead regulatory regions, enhancers or insulators, their syntenic conservation still provides an indication of lncRNA conservation [56].

When analyzing the RNA-domain annotations it is striking that the order (synteny) of many of the lncRNAs with multiple RNA-domains are consistently preserved in the birds. The annotations of these domains lie in the same genomic region, in the same order as in the mammalian homologs. Thus they support a high degree of evolutionary conservation for the entire lncRNA. In particular the HOXA11-AS1, PART1, PCA3, RMST, Six3os1, SOX2OT and ST7-OT3 lncRNAs have multiple, well conserved RNA-domains (See Fig. 1). The syntenic ordering of these seven lncRNAs and the flanking genes are also preserved between the human and chicken genomes (data not shown). We illustrate this in detail for the HOTAIRM1 lncRNA (see Fig. 2 and Fig. M in S1 Results).

The conservation of these “human” lncRNAs among birds suggests they may also be functional in birds. But what these functions may be is not immediately obvious. For example, PART1 and PCA3 are both described as prostate-specific lncRNAs that play a role in the human androgen-receptor pathway [57–59]. Birds lack a prostate but both males and females express the androgen receptor (AR or NR3C4) in gonadal and non-gonadal tissue [60–63]. Thus, we postulate that PART1 and PCA3 also play a role in the androgen-receptor pathway in birds but whether the expression of these lncRNAs are tissue specific is unknown at present.

The HOX cluster lncRNAs HOTAIRM1 (5 RNA-domains), HOXA11-AS1 (6 RNA-domains), and HOTTIP (4 RNA domains) are conserved across the Mammalian and Avian lineages. In the human genome they are located in the HOXA cluster (hg coordinates chr7:27135743–27245922), one of the most highly conserved regions in vertebrate genomes [64], in antisense orientation between HoxA1 and HoxA2, between HoxA11 and HoxA13, and upstream of HoxA13, respectively. Conservation and expression of HOTAIRM1 and HOXA11-AS1 within the HOXA cluster has been studied in some detail in marsupials [65]. Of the 15 RNA-domains five and six representing all three lncRNAs were recovered in the alligator and turtle genomes. All of them appear in the correct order at the expected, syntenically conserved positions within the HOXA cluster. In the birds, where two or more of the HOX cluster lncRNA RNA-domains were predicted on the same scaffold, this gene order and location within HOX was also preserved.

The RMST (Rhabdomyosarcoma 2 associated transcript) RNA-domains 6, 7, 8, and 9 are conserved across the birds. In each bird the gene order was also consistent with the human



**Fig 2. The preservation of gene order (synteny) surrounding the HOTAIRM1 (RF01976) locus across the Avian and other Vertebrate lineages.**

doi:10.1371/journal.pone.0121797.g002



ordering. In the alligator and turtle an additional RNA-domain was predicted in each, these were RNA-domains 2 and 4 respectively, again the ordering of the domains was consistent with human. This suggests that the RMST lncRNA is highly conserved. However, little is known about the function of this RNA. It was originally identified in a screen for differentially expressed genes in two Rhabdomyosarcoma tumor types [66].

In addition, the lncRNA DLEU2 is well conserved across the vertebrates, it is a host gene for two miRNA genes, miR-15 and miR-16, both of which are also well conserved across the vertebrates (see Fig. B in [S1 Results](#)). DLEU2 is thought to be a tumor-suppressor gene as it is frequently deleted in malignant tumours [67, 68].

The NBR2 lncRNA and BRCA1 gene share a bidirectional promoter [69]. Both are expressed in a broad range of tissues. Extensive research on BRCA1 has shown that it is involved in DNA repair [70]. The function of NBR2 remains unknown, yet its conservation across the vertebrates certainly implies a function (See [Fig. 1](#)). We note that the function for this locus may be at the DNA level, however, function at the RNA level cannot be ruled out at this stage.

Of the other classes of RNAs, none showed an unexpected degree of conservation or expansion within the avian lineage. The only exception being the snoRNA, SNORD93. SNORD93 has 92 copies in the tinamou genome, whereas it only has 1–2 copies in all the other vertebrate genomes.

## Unexpectedly poorly conserved ncRNAs: genuine loss, divergence or missing data?

**Genuine loss.** The overall reduction in avian genomic size has been extensively discussed elsewhere [71]. Unsurprisingly, this reduction is reflected in the copy-number of ncRNA genes. Some of the most dramatic examples are the transfer RNAs and pseudogenes which average  $\sim 900$  and  $\sim 580$  copies in the human, turtle and alligator genomes, the average copy-numbers of these drop to  $\sim 280$  and  $\sim 100$  copies in the avian genomes. In addition to reduction in copy-number, the absence of several, otherwise ubiquitous vertebrate ncRNAs, in the avian lineage are suggestive of genuine gene loss.

Namely, mammalian and amphibian genomes contain three loci of clustered microRNAs from the mir-17 and mir-92 families [72]. One of these clusters (cluster II, with families mir-106b, mir-93 and mir-25) was not found in turtles, crocodiles and birds (see Fig. F in [S1 Results](#)). In addition, the microRNA family let-7 is the most diverse microRNA family with 14 paralogs in human. These genes also localize in 7 genomic clusters, together with mir-100 and mir-125 miRNA families (see previous study on the evolution of the let-7 miRNA cluster in [73]). In Sauropsids we observed that cluster A—which is strongly conserved in vertebrates has been completely lost in the avian lineage. Another obvious loss in birds is cluster F, containing two let-7 microRNA paralogs. Cluster H, on the other hand has been retained in all oviparous animals and completely lost later, after the split of Theria (see Fig. G in [S1 Results](#)).

**Divergence.** In order to determine to what extent the absence of some ncRNAs from the infernal-based annotation is caused by sequence divergence beyond the thresholds of the Rfam CMs, we complemented our analysis by dedicated searches for a few of these RNA groups. Our ability to find additional homologs for several RNA families that fill gaps in the abundance matrices ([Fig. 1](#)) strongly suggests that conspicuous absences, in particular of LUCA and LECA RNAs, are caused by incomplete data in the current assemblies and sequence divergence rather than genuine losses.

Vertebrate Y RNAs typically form a cluster comprising four well-defined paralog groups Y1, Y3, Y4, and Y5. In line with [74] we find that the Y5 paralog family is absent from all bird genomes, while it is still present in both alligator and turtle (see Fig. D in [S1 Results](#)). Within

the avian lineage, we find a conserved Y4-Y3-Y1 cluster. Apparently, broken-up clusters are in most cases consistent with breaks (e.g. ends of contigs) in the available sequence assemblies. In several genomes we observe one or a few additional Y RNA homologs unlinked to the canonical Y RNA cluster. These sequences can be identified unambiguously as derived members of one of the three ancestral paralog groups, they almost always fit less well to the consensus (as measured by the CM bit score of paralog group specific covariance models) than the paralog linked to cluster, and there is no indication that any of these additional copies is evolutionarily conserved over longer time scales. We therefore suggest that most or all of these interspersed copies are in fact pseudogenes (see below).

**Missing data.** Seven families of “core” ncRNAs were found in some avian genomes but not others (Fig. 1). These families range in conservation level from being ubiquitous to cellular-life (RNase P and tRNA-sec), present in most Bilateria (vault), present in the majority of eukaryotes (RNase MRP, U4atac and U11) and present in all vertebrates (telomerase) [2]. Therefore, the genuine loss or even diversification of these ncRNA families in the avian lineage is unlikely. Rather, this lack of phylogenetic signal, combined with the fragmented nature of the vast majority of these genomes described above (i.e., of the 48 avian genomes, only the chicken and zebra finch were chromosomally assembled [19, 21] when this project was initiated), suggests the most likely explanation is that these ncRNA families are indicative of missing data. Indeed, of the seven missing ncRNA families, six were found in the chicken genome and three were found in the zebra finch genome. Furthermore, only one of these (RNase MRP) is found on a chicken macrochromosome, and all remaining missing ncRNAs are found on chicken microchromosomes (see Table A in S1 Results). A Fisher's exact test showed that there are significantly more missing ncRNAs on microchromosomes than macrochromosomes,  $P < 10^{16}$  (we use the micro/macro-chromosome assignments from the chicken genome as this is the most complete avian genome). Thus, we suggest that many of these ncRNAs families are missing because: (1) they are predominantly found on microchromosomes [this study] and (2) the vast majority of avian microchromosomes remain unsequenced [21, 31]. Furthermore, there has been minimal chromosomal rearrangement across the avian genome [21]. Therefore, it is likely that the chicken microchromosomal genes are also on microchromosomes in the other avians.

To wit, we performed dedicated searches for a selection of these missing ncRNA families. Here, tRNAscan is tuned for specificity and thus misses several occurrences of tRNA-sec that are easily found in the majority of genomes by `blastn` with  $E \leq 10^{-30}$ . In some cases the sequences appear degraded at the ends, which is likely due to low sequence quality at the very ends of contigs or scaffolds. A `blastn` search also readily retrieves additional RNase P and RNase MRP RNAs in the majority of genomes, albeit only the best conserved regions are captured. In many cases these additional candidates are incomplete or contain undetermined sequence, which explains why they are missed by the CMs [75, 76].

## Classic RNAs: LUCA and LECA

Many RNA families constitute the most evolutionarily conserved genes across all life on this planet [1]. Examples of RNAs derived from the Last Universal Common Ancestor (LUCA) include the transfer RNAs (tRNA), ribosomal RNAs (rRNA), RNA components of RNase P (RNase P RNA), RNase MRP (RNase MRP RNA) and the signal recognition particle (SRP RNA). Other classes of RNA are likely to have been components of the Last Eukaryotic Common Ancestor (LECA). These include the telomerase RNA, major spliceosomal RNAs (U1, U2, U4, U5, and U6) and the minor spliceosomal RNAs (U11, U12, U4atac, and U6atac) [2].

Unsurprisingly, the bulk of these classes of RNAs are well represented across the bird genomes (See Fig. 1). However, there appear to have been “losses” of a few of these RNAs in

certain bird species. Some of these may be due to sequence divergence, of which there are several notable examples e.g. [77–81]. Other apparent loss may be explained by incomplete genome coverage.

A number of the classic RNAs are incorporated into RNA-protein complexes (RNPs) involved in core cellular processes. An example of this are the spliceosomal RNAs. Based upon the presence/absence patterns of the major spliceosomal RNAs they are all well represented in these genome sequences. The exceptions to this observation are the U4 RNA in cormorant and the U5 RNA in the bee eater which are both missing. These two genomes are low coverage, suggesting these genes weren't captured in the current assembly. The minor spliceosomal RNAs are more interesting, the U4atac and U11 snRNAs show widespread patterns of loss, even in some of the high coverage genomes. These RNAs are frequently missed in bioinformatic screens. Indicating either frequent loss [82] or sequences that have diverged beyond the ability of detection by covariance models [83].

The telomerase RNA is also largely missing from the avian annotations. This RNA acts as a template for the telomerase enzyme that extends the telomeres found on chromosome ends. It is only found in the chicken, bald eagle, kea, budgerigar, crow and zebrafinch. Homology searches with the telomerase reverse transcriptase (TERT) protein show that the protein component of the telomerase RNP is conserved across all the bird genomes (data not shown). This pattern of presumably divergent telomerase RNA and conserved telomerase protein has been noted previously, most notably in the fungi [77, 78].

The RNA components of RNase P and RNase MRP also appear to have undergone dramatic losses within the bird lineage. RNase P is required for the maturation of tRNA, the paralogous enzyme, RNase MRP is required for the maturation of rRNA. Each RNP cleaves smaller RNAs from larger transcripts [84]. It is unlikely that these genes have been lost in any of the birds. Homology searches with the RNase associated protein coding genes (POP1, POP4, POP5, POP7, RPP1, RPP14, RPP25, RPP38, RPP40 and RPR2), identified viable homologs of each in all of the bird genomes [85] (data not shown). This suggests that the bird RNase P and MRP RNAs may have diverged slightly from the canonical models.

The 5.8S component of the ribosome in the turtle, turkey bustard, hoatzin, flamingo, tropicbird, seriema, owl, cuckoo roller, trogon, bee eater and falcon appears to have been lost (See Fig. 1). The rRNA repeats are frequently not assembled, consequently it is not surprising to see "losses" in these [86]. Furthermore, the genomes for these species are also low-coverage.

## Small nucleolar RNAs

Small nucleolar RNAs (snoRNAs) are important ncRNAs that participate in the maturation of other functional RNAs [87]. The bulk of the characterised snoRNAs guide either methylation or pseudouridylation modifications, primarily of rRNAs but also spliceosomal RNAs. The two types of modifications are guided by two different types of RNA, the box C/D and the H/ACA snoRNAs respectively, each with a characteristic cohort of motifs and secondary structures [88].

There are 66 ribosomal modification sites, guided by 59 snoRNA families, that are preserved between *H. sapiens* and *S. cerevisiae* [41]. Of these, 45 snoRNA families are conserved in the bird data set. Over a third of the apparent losses of the yeast-human conserved snoRNA families appear to cluster on 2 loci of the ancestral vertebrate genome. We investigated these losses further.

The first cluster is found at chr11:62620797–62622484 on the human genome (hg19) and contains SNORD27, SNORD29 and SNORD31 of the human-yeast conserved snoRNAs. These snoRNAs are located in the inside-out gene SNHG1 which hosts a total of eight C/D box snoRNAs: SNORD25, SNORD26, SNORD27, SNORD28, SNORD29, SNORD22,

SNORD30 and SNORD31 [89]. Each of which are also found in the alligator and turtle genomes within a 3–4 KB locus, yet these have largely been lost in the birds. However, five of the eight snoRNAs are located in the tinamou genome. These are located on the same scaffold and are within 2 KB of each other. This implies that SNHG1 is conserved in the tinamou. Loci with four of the eight snoRNAs can be found in zebrafinch, ground-finch, and bald eagle. Still, three of the eight are located in the ostrich, crow, and cuckoo genomes, again within 2 KB of each other on the same scaffolds. This complex pattern of loss could be attributed to many different models, e.g. multiple losses in birds, poor homology modelling or incomplete genome sequences.

The second cluster is located at chr19:49993222–49994231 on the human genome (hg19) and contains two copies of SNORD33 and one SNORD34 all within a 1 KB genomic region. The turtle and alligator genomes retain the two copies of SNORD33 yet don't have an obvious SNORD34 gene on the same scaffold. Within the bird genomes, the crow and rifleman each retain a single SNORD33 and SNORD34 gene on the same scaffold. While the ground-finch and bald eagle retain a single SNORD33 and the zebrafinch and seriema retain a single SNORD34 (see Fig. C in [S1 Results](#)). In human these snoRNAs are intronic to the host gene, ribosomal protein L13a (RPL13A). Based on BLASTP (version 2.2.18) homology searches for the RPL13A gene, the protein is conserved in the human and turtle genomes and in the bald eagle, crow, rifleman and zebrafinch avian genomes (data not shown). Therefore the RPL13A gene and corresponding intronic snoRNAs show the same conservation pattern. This supports a pattern of loss of the RPL13A gene and the intronic snoRNAs that it hosts in the bird genomes.

## MicroRNAs

MicroRNAs are an important class of non-coding RNA. They have been found in the genomes of Chromalveolata [90, 91], Metazoa [92–94], Mycetozoa [95, 96], Viridiplantae [97–100] and Viruses [101–104]. The miRNAs have been shown to regulate the expression of large numbers of messenger RNAs [105]. The mature miRNA product is generally 22 nucleotides long which is usually processed from a larger RNA that is characterised by a stable hairpin-shaped secondary structure.

Chicken and zebrafinch are the only birds with previously annotated microRNAs. We searched for homologs of these and other vertebrate microRNAs in the genomes of the 48 birds, American alligator and green turtle. Overall, we annotate a total of 16617 putative microRNA loci, homologous to 543 known microRNA genes, of which 487 are annotated in chicken and/or zebra finch, while 56 have been so far known only in non-avian vertebrates. The numbers of annotated loci in the individual species are approximately equal—300–400 per species, except for the turkey (*Meleagris gallopavo*) where we identified 543 sequences homologous to known microRNAs.

In addition, we can confidently identify a further 3 microRNA families that are present in mammals, and turtle and/or crocodile, but not in any avian genome (mir-150, mir-208, mir-590). This suggests that these sequences were lost in the last common ancestor of archosaurs or birds. There are also a number of microRNAs that are predicted to be present in turtles and/or crocodiles, and only a small number of bird genomes. Indeed, there are many missing annotations, species-specific and otherwise, that are not consistent with the consensus phylogeny, and could be due to either incomplete genomes or widespread microRNA loss.

The turkey genome contains a high number (190) of microRNAs so far found only in chicken, which account for the higher number of annotated sequences in this genome compared with other birds. This is consistent with its phylogenetic position as the closest chicken relative

among the examined birds. However, 101 chicken microRNAs have no homolog in the turkey or other bird genomes, suggesting that these genes are chicken-specific. This is consistent with previous reports of large number of species specific microRNAs in all animals, and supports the view of fast microRNA turnover during animal evolution [2].

## Cis-regulatory elements

The cis-regulatory RNAs are a group of RNA structures encoded on mRNAs. Generally they are involved in regulating the expression of the mRNA they are encoded within. Others may recode the translated protein product into an alternate sequence.

This group includes the iron response element (IRE) [106] and the histone 3' UTR (histone3) [107]. These are structured motifs bound by regulatory proteins. The selenocysteine insertion sequence (SECIS) is a structured motif that recodes UGA stop codons to selenocysteines [108] and the GABRA3 stem-loop is a structure recognised by the ADAR enzyme family. This enzyme edits adenine nucleotides to inosine, in this case recoding an isoleucine codon to methionine in exon 9 of the GABRA3 gene [109].

These regulatory elements and others, including an internal ribosome entry site (IRES), potassium channel RNA editing signal (K chan RES), Antizyme RNA frameshifting stimulation element (Antizyme FSE), vimentin 3' UTR protein-binding region (Vimentin3) and a connective tissue growth factor (CTGF) 3' UTR element (CAESAR) are conserved across a diverse group of vertebrates, including the bird lineages explored here (See Fig. 1).

## Pseudogenes

Non-coding RNA derived pseudogenes are a major problem for many ncRNA annotation projects. The human genome, for example, contains > 1 million Alu repeats, which are derived from the SRP RNA [110]. The existing Rfam annotation of the human genome, in particular, contains a number of problematic families that appear to have been excessively pseudogenised. The U6 snRNA, SRP RNA and Y RNA families have 1,371, 941 and 892 annotations in the human genome. These are a heterogeneous mix of pseudogenised, paralogous, diverged or functional copies of these families. Unfortunately, a generalized model of RNA pseudogenes has not been incorporated into the main covariance model package, Infernal. An approach used by tRNAscan [15], is, in theory, generalizable to other RNA families but this remains a work in progress.

It is possible that the avian annotations also contains excessive pseudogenes. However, it has previously been noted that avian genomes are significantly smaller than other vertebrate species [18]. We have also noted a corresponding reduction in the number of paralogs and presumed ncRNA-derived pseudogenes in the avian genomes (see Fig. L in S1 Results). The problematic human families, U6 snRNA, SRP RNA and Y RNA have, for example, just 26, 4 and 3 annotations respectively in the chicken genome and 13, 3 and 3 annotations respectively, on average, in the 48 avian genomes used here. Therefore, we conclude that the majority of our annotations are in fact functional orthologs.

## Experimentally confirmed ncRNAs

The ncRNAs presented here have been identified using homology models and are evolutionarily conserved in multiple avian species. In order to further validate these predictions we have used strand-specific total RNA-seq and small RNA-seq of multiple chicken tissues. After mapping the RNA-seq data to the chicken genome (see Methods for details), we identified a

threshold for calling a gene as expressed by limiting our estimated false-positive rate to approximately 10%. This FDR was estimated using a negative control of randomly selected, un-annotated regions of the genome. Since some regions may be genuinely expressed, the true FDR is potentially lower than 10%. Overall, the number of ncRNAs we have identified in this work that are expressed above background levels is 865 (72.4%) (see [Table 1](#)). This shows that 7.0 times more of our ncRNA predictions are expressed than is expected by chance (Fisher's exact test:  $P < 10^{16}$ ). This number is an underestimate of the fraction of our annotations that are genuinely expressed, as only a fraction of the developmental stages and tissues of chicken have been characterized with RNA-seq. Furthermore, some ncRNAs are expressed in highly specific conditions [[111](#), [112](#)].

The classes of RNAs where the majority of our annotations were experimentally confirmed includes microRNAs, snoRNAs, cis-regulatory elements, tRNAs, SRP RNA and RNase P/MRP RNA. The RNA-seq data could not provide evidence for a telomerase RNA transcript, which are only generally only expressed in embryonic, stem or cancerous tissues. Only a small fraction of the 7SK RNA, the minor spliceosomal RNAs and the lncRNAs could be confirmed with the 10% FDR threshold. There are a number of possible explanations for this: the multiple copies of the 7SK RNA may be functionally redundant and can therefore compensate for one another; The minor spliceosome is, as the name suggests, a rarely used alternative spliceosome; and the lncRNAs are generally expressed at low levels under specific conditions [[111](#), [113](#)]. Nevertheless, 12 of the 34 lncRNA-associated Rfam models were found to be expressed, these included HOTAIRM1, HOXA11-AS1, NBR2, SOX2OT and ST7-OT3 (see Fig. M in [S1 Results](#) for an illustration of RNA expression at the HOTAIRM1 locus).

## Discussion

In this work we have provided a comprehensive annotation of non-coding RNAs in genome sequences using homology-based methods. The homology-based tools have distinct advantages over experimental-based approaches as not all RNAs are expressed in any particular tissue-type or developmental-stage, in fact some RNAs have extremely specific expression profiles, e.g. the *lcy-6* microRNA [[112](#)]. We have identified previously unrecognized conservation of ncRNAs in avian genomes and some surprising "losses" of otherwise well conserved ncRNAs. We have shown that most of these losses are due to difficulties assembling avian microchromosomes rather than *bona fide* gene loss. A large fraction of our annotations have been confirmed using RNA-seq data, which also showed a 7-fold enrichment of expression within our annotations relative to unannotated regions.

The collection of ncRNA sequences is generally biased towards model organisms [[2](#), [87](#)]. However, we have shown that using data from well studied lineages such as mammals can also result in quality annotations of sister taxa such as Aves.

In summary, these results indicate we are in the very early phases of determining the functions of many RNA families. This is illustrated by the fact that the reported functions of some ncRNAs are mammal-specific, yet these are also found in bird genomes.

## Supporting Information

**S1 Results. Supplementary results and discussion.** Further results and discussion of poorly conserved ncRNAs, genomic contamination, additional analyses and the sources of avian sequencing data.

(PDF)

## Acknowledgments

Erich Jarvis (Duke University), Guojie Zhang (BGI-Shenzhen & University of Copenhagen) and Tom Gilbert (University of Copenhagen) for access to data and for invaluable feedback on the manuscript.

Magnus Alm Rosenblad (Univ. of Gothenburg) and Eric Nawrocki (HHMI Janelia Farm) for useful discussions. Matthew Walters for assistance with figures.

We thank Fiona McCarthy (University of Arizona), Carl Schmidt (University of Delaware), Matt Schwartz (Harvard), Igor Ulitsky (Weizmann Institute of Science), Jacqueline Smith and David Burt (Roslin Institute) for providing the RNA-seq data as part of the Avian RNA-seq consortium.

Thanks to @ewanbirney for the following timely tweet: “So. . . missing orthologs to chicken often mean ‘gene might be on the microchromosome’”.

We thank the anonymous referees for providing invaluable suggestions that improved this work.

## Author Contributions

Conceived and designed the experiments: PPG PFS SGJ. Performed the experiments: PPG MF SWB MN JH SK TES SGJ PFS. Analyzed the data: PPG MF SWB MN JH SK TES SGJ PFS. Contributed reagents/materials/analysis tools: PPG MF SWB MN JH SK TES SGJ PFS. Wrote the paper: PPG MF SWB MN JH SK TES SGJ PFS.

## References

1. Jeffares DC, Poole AM, Penny D. Relics from the RNA world. *J Mol Evol*. 1998 Jan; 46(1):18–36. doi: [10.1007/PL00006280](https://doi.org/10.1007/PL00006280) PMID: [9419222](https://pubmed.ncbi.nlm.nih.gov/9419222/)
2. Hoepfner MP, Gardner PP, Poole AM. Comparative analysis of RNA families reveals distinct repertoires for each domain of life. *PLoS Comput Biol*. 2012 Nov; 8(11):e1002752. doi: [10.1371/journal.pcbi.1002752](https://doi.org/10.1371/journal.pcbi.1002752) PMID: [23133357](https://pubmed.ncbi.nlm.nih.gov/23133357/)
3. Rivas E, Eddy SR. Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics*. 2000 Jul; 16(7):583–605. doi: [10.1093/bioinformatics/16.7.583](https://doi.org/10.1093/bioinformatics/16.7.583) PMID: [11038329](https://pubmed.ncbi.nlm.nih.gov/11038329/)
4. Cech TR, Steitz JA. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell*. 2014; 157(1):77–94. doi: [10.1016/j.cell.2014.03.008](https://doi.org/10.1016/j.cell.2014.03.008) PMID: [24679528](https://pubmed.ncbi.nlm.nih.gov/24679528/)
5. Sakakibara Y, Brown M, Hughey R, Mian IS, Sjölander K, Underwood RC, et al. Stochastic context-free grammars for tRNA modeling. *Nucleic Acids Res*. 1994 Nov; 22(23):5112–20. doi: [10.1093/nar/22.23.5112](https://doi.org/10.1093/nar/22.23.5112) PMID: [7800507](https://pubmed.ncbi.nlm.nih.gov/7800507/)
6. Eddy SR, Durbin R. RNA sequence analysis using covariance models. *Nucleic Acids Res*. 1994 Jun; 22(11):2079–88. doi: [10.1093/nar/22.11.2079](https://doi.org/10.1093/nar/22.11.2079) PMID: [8029015](https://pubmed.ncbi.nlm.nih.gov/8029015/)
7. Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: inference of RNA alignments. *Bioinformatics*. 2009 May; 25(10):1335–7. doi: [10.1093/bioinformatics/btp157](https://doi.org/10.1093/bioinformatics/btp157) PMID: [19307242](https://pubmed.ncbi.nlm.nih.gov/19307242/)
8. Freyhult EK, Bollback JP, Gardner PP. Exploring genomic dark matter: a critical assessment of the performance of homology search methods on noncoding RNA. *Genome Res*. 2007 Jan; 17(1):117–125. doi: [10.1101/gr.5890907](https://doi.org/10.1101/gr.5890907) PMID: [17151342](https://pubmed.ncbi.nlm.nih.gov/17151342/)
9. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res*. 2003 Jan; 31(1):439–41. doi: [10.1093/nar/gkg006](https://doi.org/10.1093/nar/gkg006) PMID: [12520045](https://pubmed.ncbi.nlm.nih.gov/12520045/)
10. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res*. 2005 Jan; 33(Database issue):D121–4. doi: [10.1093/nar/gki081](https://doi.org/10.1093/nar/gki081) PMID: [15608160](https://pubmed.ncbi.nlm.nih.gov/15608160/)
11. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, et al. Rfam: updates to the RNA families database. *Nucleic Acids Res*. 2009 Jan; 37(Database issue):D136–40. doi: [10.1093/nar/gkn766](https://doi.org/10.1093/nar/gkn766) PMID: [18953034](https://pubmed.ncbi.nlm.nih.gov/18953034/)
12. Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, et al. Rfam: Wikipedia, clans and the ‘decimal’ release. *Nucleic Acids Res*. 2011 Jan; 39(Database issue):D141–5. doi: [10.1093/nar/gkq1129](https://doi.org/10.1093/nar/gkq1129) PMID: [21062808](https://pubmed.ncbi.nlm.nih.gov/21062808/)

13. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, et al. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* 2013 Jan; 41(Database issue):D226–32. doi: [10.1093/nar/gks1005](https://doi.org/10.1093/nar/gks1005) PMID: [23125362](https://pubmed.ncbi.nlm.nih.gov/23125362/)
14. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, et al. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* 2014 Nov;. PMID: [25392425](https://pubmed.ncbi.nlm.nih.gov/25392425/)
15. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997 Mar; 25(5):955–64. doi: [10.1093/nar/25.5.0955](https://doi.org/10.1093/nar/25.5.0955) PMID: [9023104](https://pubmed.ncbi.nlm.nih.gov/9023104/)
16. Chan PP, Lowe TM. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* 2009 Jan; 37(Database issue):D93–7. doi: [10.1093/nar/gkn787](https://doi.org/10.1093/nar/gkn787) PMID: [18984615](https://pubmed.ncbi.nlm.nih.gov/18984615/)
17. Menzel P, Gorodkin J, Stadler PF. The Tedious Task of Finding Homologous Non-coding RNA Genes. *RNA.* 2009; 15:2075–2082. doi: [10.1261/ma.1556009](https://doi.org/10.1261/ma.1556009) PMID: [19861422](https://pubmed.ncbi.nlm.nih.gov/19861422/)
18. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature.* 2004; 432(7018):695–716. doi: [10.1038/nature03154](https://doi.org/10.1038/nature03154)
19. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, et al. The genome of a songbird. *Nature.* 2010 Apr; 464(7289):757–62. doi: [10.1038/nature08819](https://doi.org/10.1038/nature08819) PMID: [20360741](https://pubmed.ncbi.nlm.nih.gov/20360741/)
20. Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Blomberg LA, et al. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* 2010; 8(9). doi: [10.1371/journal.pbio.1000475](https://doi.org/10.1371/journal.pbio.1000475) PMID: [20838655](https://pubmed.ncbi.nlm.nih.gov/20838655/)
21. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science.* 2014; 346(6215):1311–1320. doi: [10.1126/science.1251385](https://doi.org/10.1126/science.1251385) PMID: [25504712](https://pubmed.ncbi.nlm.nih.gov/25504712/)
22. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science.* 2014; 346(6215):1320–1331. doi: [10.1126/science.1253451](https://doi.org/10.1126/science.1253451) PMID: [25504713](https://pubmed.ncbi.nlm.nih.gov/25504713/)
23. Huang Y, Li Y, Burt DW, Chen H, Zhang Y, Qian W, et al. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat Genet.* 2013 Jul; 45(7):776–83. doi: [10.1038/ng.2657](https://doi.org/10.1038/ng.2657) PMID: [23749191](https://pubmed.ncbi.nlm.nih.gov/23749191/)
24. Zhan X, Pan S, Wang J, Dixon A, He J, Muller MG, et al. Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nat Genet.* 2013 May; 45(5):563–6. doi: [10.1534/genetics.113.154161](https://doi.org/10.1534/genetics.113.154161) PMID: [23525076](https://pubmed.ncbi.nlm.nih.gov/23525076/)
25. Shapiro MD, Kronenberg Z, Li C, Domyan ET, Pan H, Campbell M, et al. Genomic diversity and evolution of the head crest in the rock pigeon. *Science.* 2013 Mar; 339(6123):1063–7. doi: [10.1126/science.1230422](https://doi.org/10.1126/science.1230422) PMID: [23371554](https://pubmed.ncbi.nlm.nih.gov/23371554/)
26. Howard J, Koren S, Phillippy A, Zhou S, Schwartz D, Schatz M, et al. De novo high-coverage sequencing and annotated assemblies of the budgerigar genome. *GigaScience Database.* 2013;.
27. Li J, et al. The genomes of two Antarctic penguins reveal adaptations to the cold aquatic environment; 2014. Submitted.
28. Griffin DK, Robertson LB, Tempest HG, Skinner BM. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet Genome Res.* 2007; 117(1–4):64–77. doi: [10.1159/000103166](https://doi.org/10.1159/000103166) PMID: [17675846](https://pubmed.ncbi.nlm.nih.gov/17675846/)
29. Solinhac R, Leroux S, Galkina S, Chazara O, Feve K, Vignoles F, et al. Integrative mapping analysis of chicken microchromosome 16 organization. *BMC Genomics.* 2010; 11:616. doi: [10.1186/1471-2164-11-616](https://doi.org/10.1186/1471-2164-11-616) PMID: [21050458](https://pubmed.ncbi.nlm.nih.gov/21050458/)
30. Douaud M, Fève K, Gerus M, Fillon V, Bardes S, Gourichon D, et al. Addition of the microchromosome GGA25 to the chicken genome sequence assembly through radiation hybrid and genetic mapping. *BMC Genomics.* 2008; 9:129. doi: [10.1186/1471-2164-9-129](https://doi.org/10.1186/1471-2164-9-129) PMID: [18366813](https://pubmed.ncbi.nlm.nih.gov/18366813/)
31. Ellegren H. The avian genome uncovered. *Trends Ecol Evol.* 2005 Apr; 20(4):180–6. doi: [10.1016/j.tree.2005.01.015](https://doi.org/10.1016/j.tree.2005.01.015) PMID: [16701366](https://pubmed.ncbi.nlm.nih.gov/16701366/)
32. Zhang G, Li B, Li C, Gilbert MTP, Jarvis ED, Wang J, et al. Comparative genomic data of the Avian Phylogenomics Project. *GigaScience.* 2014; 3(26).
33. The Avian Genome Consortium. The phylogenomics analysis of birds website;. [Http://phybirds.genomics.org.cn/index.jsp](http://phybirds.genomics.org.cn/index.jsp).
34. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics.* 2013 Nov; 29(22):2933–5. doi: [10.1093/bioinformatics/btt509](https://doi.org/10.1093/bioinformatics/btt509) PMID: [24008419](https://pubmed.ncbi.nlm.nih.gov/24008419/)
35. Nawrocki EP. Annotating functional RNAs in genomes using Infernal. *Methods Mol Biol.* 2014; 1097:163–97. doi: [10.1007/978-1-62703-709-9\\_9](https://doi.org/10.1007/978-1-62703-709-9_9) PMID: [24639160](https://pubmed.ncbi.nlm.nih.gov/24639160/)



36. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 2011 Sep; 25(18):1915–27. doi: [10.1101/gad.174466](https://doi.org/10.1101/gad.174466) PMID: [21890647](https://pubmed.ncbi.nlm.nih.gov/21890647/)
37. Kutter C, Watt S, Stefflova K, Wilson MD, Goncalves A, Ponting CP, et al. Rapid turnover of long non-coding RNAs and the evolution of gene expression. *PLoS Genet.* 2012; 8(7):e1002841. doi: [10.1371/journal.pgen.1002841](https://doi.org/10.1371/journal.pgen.1002841) PMID: [22844254](https://pubmed.ncbi.nlm.nih.gov/22844254/)
38. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 2014 Jan; 42(Database issue):D68–73. doi: [10.1093/nar/gkt1181](https://doi.org/10.1093/nar/gkt1181) PMID: [24275495](https://pubmed.ncbi.nlm.nih.gov/24275495/)
39. Griffiths-Jones S. RALEE—RNA ALignment editor in Emacs. *Bioinformatics.* 2005 Jan; 21(2):257–9. doi: [10.1093/bioinformatics/bth489](https://doi.org/10.1093/bioinformatics/bth489) PMID: [15377506](https://pubmed.ncbi.nlm.nih.gov/15377506/)
40. Bartschat S, Kehr S, Tafer H, Stadler PF, Hertel J. snoStrip: A snoRNA annotation pipeline; 2014. Preprint.
41. Lestrade L, Weber MJ. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.* 2006 Jan; 34(Database issue):D158–62. doi: [10.1093/nar/gkj002](https://doi.org/10.1093/nar/gkj002) PMID: [16381836](https://pubmed.ncbi.nlm.nih.gov/16381836/)
42. Schmitz J, Zemann A, Churakov G, Kuhl H, Grützner F, Reinhardt R, et al. Retroposed SNOfall—a mammalian-wide comparison of platypus snoRNAs. *Genome Res.* 2008 Jun; 18(6):1005–10. doi: [10.1101/gr.7177908](https://doi.org/10.1101/gr.7177908) PMID: [18463303](https://pubmed.ncbi.nlm.nih.gov/18463303/)
43. Shao P, Yang JH, Zhou H, Guan DG, Qu LH. Genome-wide analysis of chicken snoRNAs provides unique implications for the evolution of vertebrate snoRNAs. *BMC Genomics.* 2009; 10:86. doi: [10.1186/1471-2164-10-86](https://doi.org/10.1186/1471-2164-10-86) PMID: [19232134](https://pubmed.ncbi.nlm.nih.gov/19232134/)
44. Non-coding RNA annotations of bird genomes; 2015. Available from: <https://github.com/ppgardne/bird-genomes>. Accessed 2015 Feb 24.
45. Smith J, Burt DW. The Avian RNAseq Consortium: a community effort to annotate the chicken genome. *bioRxiv.* 2014;.
46. Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, et al. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS computational biology.* 2009; 5(9):e1000502. doi: [10.1371/journal.pcbi.1000502](https://doi.org/10.1371/journal.pcbi.1000502) PMID: [19750212](https://pubmed.ncbi.nlm.nih.gov/19750212/)
47. Gardner PP, Wilm A, Washietl S. A benchmark of multiple sequence alignment programs upon structural RNAs. *Nucleic Acids Res.* 2005; 33(8):2433–2439. doi: [10.1093/nar/gki541](https://doi.org/10.1093/nar/gki541) PMID: [15860779](https://pubmed.ncbi.nlm.nih.gov/15860779/)
48. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.* 2007 Jun; 129(7):1311–23. doi: [10.1016/j.cell.2007.05.022](https://doi.org/10.1016/j.cell.2007.05.022) PMID: [17604720](https://pubmed.ncbi.nlm.nih.gov/17604720/)
49. Chow JC, Yen Z, Ziesche SM, Brown CJ. Silencing of the mammalian X chromosome. *Annu Rev Genomics Hum Genet.* 2005; 6:69–92. doi: [10.1146/annurev.genom.6.080604.162350](https://doi.org/10.1146/annurev.genom.6.080604.162350) PMID: [16124854](https://pubmed.ncbi.nlm.nih.gov/16124854/)
50. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature.* 2009 Mar; 458(7235):223–7. doi: [10.1038/nature07672](https://doi.org/10.1038/nature07672) PMID: [19182780](https://pubmed.ncbi.nlm.nih.gov/19182780/)
51. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell.* 2013 Jul; 154(1):26–46. doi: [10.1016/j.cell.2013.06.020](https://doi.org/10.1016/j.cell.2013.06.020) PMID: [23827673](https://pubmed.ncbi.nlm.nih.gov/23827673/)
52. 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 Jun; 312(5780):1653–5. doi: [10.1126/science.1126316](https://doi.org/10.1126/science.1126316) PMID: [16778056](https://pubmed.ncbi.nlm.nih.gov/16778056/)
53. Smits G, Mungall AJ, Griffiths-Jones S, Smith P, Beury D, Matthews L, et al. Conservation of the H19 noncoding RNA and H19-IGF2 imprinting mechanism in therians. *Nat Genet.* 2008 Aug; 40(8):971–6. doi: [10.1038/ng.168](https://doi.org/10.1038/ng.168) PMID: [18587395](https://pubmed.ncbi.nlm.nih.gov/18587395/)
54. Chodroff RA, Goodstadt L, Sirey TM, Oliver PL, Davies KE, Green ED, et al. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol.* 2010; 11(7):R72. doi: [10.1186/gb-2010-11-7-r72](https://doi.org/10.1186/gb-2010-11-7-r72) PMID: [20624288](https://pubmed.ncbi.nlm.nih.gov/20624288/)
55. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell.* 2011 Dec; 147(7):1537–50. doi: [10.1016/j.cell.2011.11.055](https://doi.org/10.1016/j.cell.2011.11.055) PMID: [22196729](https://pubmed.ncbi.nlm.nih.gov/22196729/)
56. Diederichs S. The four dimensions of noncoding RNA conservation. *Trends in Genetics.* 2014;.
57. Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA, et al. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res.* 1999 Dec; 59(23):5975–9. PMID: [10606244](https://pubmed.ncbi.nlm.nih.gov/10606244/)

58. Lin B, White JT, Ferguson C, Bumgarner R, Friedman C, Trask B, et al. PART-1: a novel human prostate-specific, androgen-regulated gene that maps to chromosome 5q12. *Cancer Res.* 2000 Feb; 60(4):858–63. PMID: [10706094](#)
59. Ferreira LB, Palumbo A, de Mello KD, Sternberg C, Caetano MS, de Oliveira FL, et al. PCA3 noncoding RNA is involved in the control of prostate-cancer cell survival and modulates androgen receptor signaling. *BMC Cancer.* 2012; 12:507. doi: [10.1186/1471-2407-12-507](#) PMID: [23130941](#)
60. Yoshimura Y, Chang C, Okamoto T, Tamura T. Immunolocalization of androgen receptor in the small, preovulatory, and postovulatory follicles of laying hens. *Gen Comp Endocrinol.* 1993 Jul; 91(1):81–9. doi: [10.1006/gcen.1993.1107](#) PMID: [8405895](#)
61. Veney SL, Wade J. Steroid receptors in the adult zebra finch syrinx: a sex difference in androgen receptor mRNA, minimal expression of estrogen receptor alpha and aromatase. *Gen Comp Endocrinol.* 2004 Apr; 136(2):192–9. doi: [10.1016/j.ygcen.2003.12.017](#) PMID: [15028522](#)
62. Fuxjager MJ, Schultz JD, Barske J, Feng NY, Fusani L, Mirzamani A, et al. Spinal motor and sensory neurons are androgen targets in an acrobatic bird. *Endocrinology.* 2012 Aug; 153(8):3780–91. doi: [10.1210/en.2012-1313](#) PMID: [22635677](#)
63. Leska A, Kiezun J, Kaminska B, Dusza L. Seasonal changes in the expression of the androgen receptor in the testes of the domestic goose (*Anser anser f. domestica*). *Gen Comp Endocrinol.* 2012 Oct; 179(1):63–70. doi: [10.1016/j.ygcen.2012.07.026](#) PMID: [22885558](#)
64. Pascual-Anaya J, D’Aniello S, Kuratani S, Garcia-Fernández J. Evolution of Hox gene clusters in deuterostomes. *BMC Developmental Biology.* 2013; 13:26. doi: [10.1186/1471-213X-13-26](#) PMID: [23819519](#)
65. Yu H, Lindsay J, Feng ZP, Frankenberg S, Hu Y, Carone D, et al. Evolution of coding and non-coding genes in HOX clusters of a marsupial. *BMC Genomics.* 2012; 13:251. doi: [10.1186/1471-2164-13-251](#) PMID: [22708672](#)
66. Chan AS, Thorner PS, Squire JA, Zielenska M. Identification of a novel gene NCRMS on chromosome 12q21 with differential expression between rhabdomyosarcoma subtypes. *Oncogene.* 2002 May; 21(19):3029–37. doi: [10.1038/sj.onc.1205460](#) PMID: [12082533](#)
67. Lerner M, Harada M, Lovén J, Castro J, Davis Z, Oscier D, et al. DLEU2, frequently deleted in malignancy, functions as a critical host gene of the cell cycle inhibitory microRNAs miR-15a and miR-16-1. *Exp Cell Res.* 2009 Oct; 315(17):2941–52. doi: [10.1016/j.yexcr.2009.07.001](#) PMID: [19591824](#)
68. Klein U, Lia M, Crespo M, Siegel R, Shen Q, Mo T, et al. The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell.* 2010 Jan; 17(1):28–40. doi: [10.1016/j.ccr.2009.11.019](#) PMID: [20060366](#)
69. Xu CF, Brown MA, Nicolai H, Chambers JA, Griffiths BL, Solomon E. Isolation and characterisation of the NBR2 gene which lies head to head with the human BRCA1 gene. *Hum Mol Genet.* 1997 Jul; 6(7):1057–62. doi: [10.1093/hmg/6.7.1057](#) PMID: [9215675](#)
70. Moynahan ME, Chiu JW, Koller BH, Jasin M. Brca1 controls homology-directed DNA repair. *Mol Cell.* 1999 Oct; 4(4):511–8. doi: [10.1016/S1097-2765\(00\)80202-6](#) PMID: [10549283](#)
71. Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV. Origin of avian genome size and structure in non-avian dinosaurs. *Nature.* 2007 Mar; 446(7132):180–4. doi: [10.1038/nature05621](#) PMID: [17344851](#)
72. Tanzer A, Stadler P. Molecular evolution of a microRNA cluster. *J Mol Biol.* 2004; 339(2):327–35. doi: [10.1016/j.jmb.2004.03.065](#) PMID: [15136036](#)
73. Hertel J, Bartschat S, Wintsche A, C O, The Students of the Bioinformatics Computer Lab 2011, Stadler PF. Evolution of the let-7 microRNA Family. “RNA Biol”. 2012; In press.
74. Mosig A, Guofeng M, Stadler B, Stadler P. Evolution of the vertebrate Y RNA cluster. *Theory in Biosciences.* 2007; 126(1):9–14. doi: [10.1007/s12064-007-0003-y](#) PMID: [18087752](#)
75. Stadler PF, Chen JLL, Hackermüller J, Hoffmann S, Horn F, Khaitovich P, et al. Evolution of Vault RNAs. *Mol Biol Evol.* 2009; 26:1975–1991. doi: [10.1093/molbev/msp112](#) PMID: [19491402](#)
76. Kolbe DL, Eddy SR. Local RNA structure alignment with incomplete sequence. *Bioinformatics.* 2009 May; 25(10):1236–43. doi: [10.1093/bioinformatics/btp154](#) PMID: [19304875](#)
77. Leonardi J, Box JA, Bunch JT, Baumann P. TER1, the RNA subunit of fission yeast telomerase. *Nat Struct Mol Biol.* 2008 Jan; 15(1):26–33. doi: [10.1038/nsmb1343](#) PMID: [18157152](#)
78. Webb CJ, Zakian VA. Identification and characterization of the *Schizosaccharomyces pombe* TER1 telomerase RNA. *Nat Struct Mol Biol.* 2008 Jan; 15(1):34–42. doi: [10.1038/nsmb1354](#) PMID: [18157149](#)
79. Mao C, Bhardwaj K, Sharkady SM, Fish RI, Driscoll T, Wower J, et al. Variations on the tmRNA gene. *RNA Biol.* 2009; 6(4):355–61. doi: [10.4161/ma.6.4.9172](#) PMID: [19617710](#)

80. Lai LB, Chan PP, Cozen AE, Bernick DL, Brown JW, Gopalan V, et al. Discovery of a minimal form of RNase P in *Pyrobaculum*. *Proc Natl Acad Sci U S A*. 2010 Dec; 107(52):22493–8. doi: [10.1073/pnas.1013969107](https://doi.org/10.1073/pnas.1013969107) PMID: [21135215](https://pubmed.ncbi.nlm.nih.gov/21135215/)
81. Chan PP, Cozen AE, Lowe TM. Discovery of permuted and recently split transfer RNAs in Archaea. *Genome Biol*. 2011; 12(4):R38. doi: [10.1186/gb-2011-12-4-r38](https://doi.org/10.1186/gb-2011-12-4-r38) PMID: [21489296](https://pubmed.ncbi.nlm.nih.gov/21489296/)
82. Dávila López M, Rosenblad MA, Samuelsson T. Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. *Nucleic Acids Res*. 2008 May; 36(9):3001–10. doi: [10.1093/nar/gkn142](https://doi.org/10.1093/nar/gkn142) PMID: [18390578](https://pubmed.ncbi.nlm.nih.gov/18390578/)
83. Marz M, Kirsten T, Stadler PF. Evolution of spliceosomal snRNA genes in metazoan animals. *J Mol Evol*. 2008 Dec; 67(6):594–607. doi: [10.1007/s00239-008-9149-6](https://doi.org/10.1007/s00239-008-9149-6) PMID: [19030770](https://pubmed.ncbi.nlm.nih.gov/19030770/)
84. López MD, Rosenblad MA, Samuelsson T. Conserved and variable domains of RNase MRP RNA. *RNA Biol*. 2009 Jul; 6(3).
85. Rosenblad MA, López MD, Piccinelli P, Samuelsson T. Inventory and analysis of the protein subunits of the ribonucleases P and MRP provides further evidence of homology between the yeast and human enzymes. *Nucleic Acids Res*. 2006; 34(18):5145–56. doi: [10.1093/nar/gkl626](https://doi.org/10.1093/nar/gkl626) PMID: [16998185](https://pubmed.ncbi.nlm.nih.gov/16998185/)
86. Floutsakou I, Agrawal S, Nguyen TT, Seoighe C, Ganley AR, McStay B. The shared genomic architecture of human nucleolar organizer regions. *Genome Res*. 2013 Dec; 23(12):2003–12. doi: [10.1101/gr.157941.113](https://doi.org/10.1101/gr.157941.113) PMID: [23990606](https://pubmed.ncbi.nlm.nih.gov/23990606/)
87. Gardner PP, Bateman A, Poole AM. SnoPatrol: how many snoRNA genes are there? *J Biol*. 2010; 9(1):4. doi: [10.1186/jbiol211](https://doi.org/10.1186/jbiol211) PMID: [20122292](https://pubmed.ncbi.nlm.nih.gov/20122292/)
88. Marz M, Gruber AR, Höner Zu Siederdisen C, Amman F, Badelt S, Bartschat S, et al. Animal snoRNAs and scaRNAs with exceptional structures. *RNA Biol*. 2011 Nov; 8(6). doi: [10.4161/rna.8.6.16603](https://doi.org/10.4161/rna.8.6.16603) PMID: [21955586](https://pubmed.ncbi.nlm.nih.gov/21955586/)
89. Tycowski KT, Shu MD, Steitz JA. A mammalian gene with introns instead of exons generating stable RNA products. *Nature*. 1996 Feb; 379(6564):464–6. doi: [10.1038/379464a0](https://doi.org/10.1038/379464a0) PMID: [8559254](https://pubmed.ncbi.nlm.nih.gov/8559254/)
90. Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, et al. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature*. 2010 Jun; 465(7298):617–21. doi: [10.1038/nature09016](https://doi.org/10.1038/nature09016) PMID: [20520714](https://pubmed.ncbi.nlm.nih.gov/20520714/)
91. Huang A, He L, Wang G. Identification and characterization of microRNAs from *Phaeodactylum tricornutum* by high-throughput sequencing and bioinformatics analysis. *BMC Genomics*. 2011; 12:337. doi: [10.1186/1471-2164-12-337](https://doi.org/10.1186/1471-2164-12-337) PMID: [21718527](https://pubmed.ncbi.nlm.nih.gov/21718527/)
92. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*. 1993 Dec; 75(5):843–54. doi: [10.1016/0092-8674\(93\)90529-Y](https://doi.org/10.1016/0092-8674(93)90529-Y) PMID: [8252621](https://pubmed.ncbi.nlm.nih.gov/8252621/)
93. Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*. 2001 Oct; 294(5543):858–62. doi: [10.1126/science.1065062](https://doi.org/10.1126/science.1065062) PMID: [11679671](https://pubmed.ncbi.nlm.nih.gov/11679671/)
94. Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, et al. The expansion of the metazoan microRNA repertoire. *BMC Genomics*. 2006; 7:25. doi: [10.1186/1471-2164-7-25](https://doi.org/10.1186/1471-2164-7-25) PMID: [16480513](https://pubmed.ncbi.nlm.nih.gov/16480513/)
95. Hinas A, Reimegård J, Wagner EG, Nellen W, Ambros VR, Söderbom F. The small RNA repertoire of *Dictyostelium discoideum* and its regulation by components of the RNAi pathway. *Nucleic Acids Res*. 2007; 35(20):6714–26. doi: [10.1093/nar/gkm707](https://doi.org/10.1093/nar/gkm707) PMID: [17916577](https://pubmed.ncbi.nlm.nih.gov/17916577/)
96. Avesson L, Reimegård J, Wagner EG, Söderbom F. MicroRNAs in Amoebozoa: deep sequencing of the small RNA population in the social amoeba *Dictyostelium discoideum* reveals developmentally regulated microRNAs. *RNA*. 2012 Oct; 18(10):1771–82. doi: [10.1261/rna.033175.112](https://doi.org/10.1261/rna.033175.112) PMID: [22875808](https://pubmed.ncbi.nlm.nih.gov/22875808/)
97. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. MicroRNAs in plants. *Genes Dev*. 2002 Jul; 16(13):1616–26. doi: [10.1101/gad.1004402](https://doi.org/10.1101/gad.1004402) PMID: [12101121](https://pubmed.ncbi.nlm.nih.gov/12101121/)
98. Fattash I, Voss B, Reski R, Hess WR, Frank W. Evidence for the rapid expansion of microRNA-mediated regulation in early land plant evolution. *BMC Plant Biol*. 2007; 7:13. doi: [10.1186/1471-2229-7-13](https://doi.org/10.1186/1471-2229-7-13) PMID: [17359535](https://pubmed.ncbi.nlm.nih.gov/17359535/)
99. Axtell MJ, Snyder JA, Bartel DP. Common functions for diverse small RNAs of land plants. *Plant Cell*. 2007 Jun; 19(6):1750–69. doi: [10.1105/tpc.107.051706](https://doi.org/10.1105/tpc.107.051706) PMID: [17601824](https://pubmed.ncbi.nlm.nih.gov/17601824/)
100. Molnár A, Schwach F, Studholme DJ, Thuenemann EC, Baulcombe DC. miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature*. 2007 Jun; 447(7148):1126–9. doi: [10.1038/nature05903](https://doi.org/10.1038/nature05903) PMID: [17538623](https://pubmed.ncbi.nlm.nih.gov/17538623/)
101. Pfeffer S, Zavolan M, Grässer FA, Chien M, Russo JJ, Ju J, et al. Identification of virus-encoded microRNAs. *Science*. 2004 Apr; 304(5671):734–6. doi: [10.1126/science.1096781](https://doi.org/10.1126/science.1096781) PMID: [15118162](https://pubmed.ncbi.nlm.nih.gov/15118162/)

102. Ouellet DL, Plante I, Landry P, Barat C, Janelle ME, Flamand L, et al. Identification of functional microRNAs released through asymmetrical processing of HIV-1 TAR element. *Nucleic Acids Res.* 2008 Apr; 36(7):2353–65. doi: [10.1093/nar/gkn076](https://doi.org/10.1093/nar/gkn076) PMID: [18299284](https://pubmed.ncbi.nlm.nih.gov/18299284/)
103. Pfeffer S, Sewer A, Lagos-Quintana M, Sheridan R, Sander C, Grässer FA, et al. Identification of microRNAs of the herpesvirus family. *Nat Methods.* 2005 Apr; 2(4):269–76. doi: [10.1038/nmeth746](https://doi.org/10.1038/nmeth746) PMID: [15782219](https://pubmed.ncbi.nlm.nih.gov/15782219/)
104. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell.* 2007 Jun; 129(7):1401–14. doi: [10.1016/j.cell.2007.04.040](https://doi.org/10.1016/j.cell.2007.04.040) PMID: [17604727](https://pubmed.ncbi.nlm.nih.gov/17604727/)
105. Lim LP, Lau NC, Garrett-Engle P, Grimson A, Schelter JM, Castle J, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature.* 2005 Feb; 433(7027):769–73. doi: [10.1038/nature03315](https://doi.org/10.1038/nature03315) PMID: [15685193](https://pubmed.ncbi.nlm.nih.gov/15685193/)
106. Stevens SG, Gardner PP, Brown C. Two covariance models for iron-responsive elements. *RNA Biol.* 8(5):792–801. doi: [10.4161/ma.8.5.16037](https://doi.org/10.4161/ma.8.5.16037) PMID: [21881407](https://pubmed.ncbi.nlm.nih.gov/21881407/)
107. López D, Samuelsson T. Early evolution of histone mRNA 3' end processing. *RNA.* 2008 Jan; 14(1):1–10.
108. Lambert A, Lescure A, Gautheret D. A survey of metazoan selenocysteine insertion sequences. *Biochimie.* 2002 Sep; 84(9):953–9. doi: [10.1016/S0300-9084\(02\)01441-4](https://doi.org/10.1016/S0300-9084(02)01441-4) PMID: [12458087](https://pubmed.ncbi.nlm.nih.gov/12458087/)
109. Ohlson J, Pedersen JS, Haussler D, Ohman M. Editing modifies the GABA(A) receptor subunit alpha3. *RNA.* 2007 May; 13(5):698–703. doi: [10.1261/ma.349107](https://doi.org/10.1261/ma.349107) PMID: [17369310](https://pubmed.ncbi.nlm.nih.gov/17369310/)
110. Wheeler TJ, Clements J, Eddy SR, Hubley R, Jones TA, Jurka J, et al. Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Res.* 2013 Jan; 41(Database issue): D70–82. doi: [10.1093/nar/gks1265](https://doi.org/10.1093/nar/gks1265) PMID: [23203985](https://pubmed.ncbi.nlm.nih.gov/23203985/)
111. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences.* 2008; 105(2):716–721.
112. Johnston RJ, Hobert O. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature.* 2003; 426(6968):845–849. doi: [10.1038/nature02255](https://doi.org/10.1038/nature02255) PMID: [14685240](https://pubmed.ncbi.nlm.nih.gov/14685240/)
113. Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddloh JA, et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nature biotechnology.* 2012; 30(1):99–104. doi: [10.1038/nbt.2024](https://doi.org/10.1038/nbt.2024)