2017

Domestic chickens activate a piRNA defense against avian leukosis virus

Yu Huining Sun
University of Rochester Medical Center

Li Huitong Xie
University of Rochester Medical Center

Xiaoyu Zhuo
Washington University School of Medicine in St. Louis

Qiang Chen
University of Rochester Medical Center

Dalia Ghoneim
University of Rochester Medical Center

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Authors

Yu Huining Sun, Li Huitong Xie, Xiaoyu Zhuo, Qiang Chen, Dalia Ghoneim, Bin Zhang, Jarra Jagne, Chengbo Yang, and Xin Zhiguo Li

Domestic chickens activate a piRNA defense against avian leukosis virus

Yu Huining Sun[1], Li Huitong Xie[1], Xiaoyu Zhuo[2], Qiang Chen[1], Dalia Ghoneim[1], Bin Zhang[3], Jarra Jagne[4], Chengbo Yang[5], Xin Zhiguo Li[1]*

[1]Center for RNA Biology: From Genome to Therapeutics, Department of Biochemistry and Biophysics, Department of Urology, University of Rochester Medical Center, Rochester, United States; [2]Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, United States; [3]Department of Pathology and Laboratory Medicine, Department of Pediatrics, University of Rochester Medical Center, Rochester, United States; [4]Animal Health Diagnostic Center, Cornell University College of Veterinary Medicine, Ithaca, United States; [5]Department of Animal Science, University of Manitoba, Winnipeg, Canada

Abstract PIWI-interacting RNAs (piRNAs) protect the germ line by targeting transposable elements (TEs) through the base-pair complementarity. We do not know how piRNAs co-evolve with TEs in chickens. Here we reported that all active TEs in the chicken germ line are targeted by piRNAs, and as TEs lose their activity, the corresponding piRNAs erode away. We observed de novo piRNA birth as host responds to a recent retroviral invasion. Avian leukosis virus (ALV) has endogenized prior to chicken domestication, remains infectious, and threatens poultry industry. Domestic fowl produce piRNAs targeting ALV from one ALV provirus that was known to render its host ALV resistant. This proviral locus does not produce piRNAs in undomesticated wild chickens. Our findings uncover rapid piRNA evolution reflecting contemporary TE activity, identify a new piRNA acquisition modality by activating a pre-existing genomic locus, and extend piRNA defense roles to include the period when endogenous retroviruses are still infectious.

DOI: 10.7554/eLife.24695.001

Introduction

A vertebrate germ-line genome faces repeated activation of transposable elements (TEs) as well as integration of new retroviruses that become endogenized. The genome has a marvelous ability to adapt to these challenges (*McClintock, 1984*). Among the adaptive responses, PIWI-interacting RNAs (piRNAs) are essential to protect the integrity of the germ-line genome by targeting 'non-self' sequences through base-pair complementarity. Disruption of piRNA pathways activates TEs in male and female fruit flies (*Wilson et al., 1996*; *Lin and Spradling, 1997*), male and female zebrafish (*Houwing et al., 2008*), and male mice (*Kuramochi-Miyagawa et al., 2004*; *Carmell et al., 2007*). piRNAs bind a specialized sub-family of Argonaute proteins, the PIWI proteins, which are expressed mainly in germ cells (*Kumar and Carmichael, 1998*; *Aravin and Hannon, 2008*; *Farazi et al., 2008*; *Kim et al., 2009*; *Thomson and Lin, 2009*; *Cenik and Zamore, 2011*). piRNAs guide PIWI proteins to their complementary RNA targets. PIWI proteins catalyze an endonucleolytic cleavage between the 10th and 11th positions of the RNA target relative to the piRNA 5′end. The cleaved product can then be loaded into another PIWI protein becoming a secondary piRNA. This results in a 'Ping-Pong' loop that amplifies antisense TE piRNAs (*Brennecke et al., 2007*; *Gunawardane et al., 2007*). The initial triggers of Ping-Pong amplification are produced from discrete genomic loci. The host needs to incorporate the foreign sequences into these piRNA-producing loci to recognize novel 'nonself'

eLife digest Viruses called retroviruses can infect animal cells and merge their genetic information with those of the animal causing damage to the animal's genetic blueprints. Once retroviruses are integrated into a cell they can sometimes get passed down through the generations over the centuries. Almost half of the human genetic code, for example, is made from ancient retroviruses and other foreign sequences. Over time many of these ancient viruses lost the ability to infect other cells and became trapped within cells but they can still jump out and damage the animal's genetic code under certain circumstances. These trapped foreign sequences are called transposable elements.

Animal cells produce molecules called piRNAs to shut down transposable elements. Most piRNAs are produced from genetic information that originally came from integrated retroviruses and that has been hijacked to defend the cell, a similar strategy as Crisper system in bacteria. Domestic chickens produce piRNAs against a virus called avian leukosis virus (or ALV for short) – which commonly infects domestic fowl. The virus also infected the wild ancestors of chickens, known as red jungle fowl, but these birds do not produce piRNAs. This provides an ideal setting to study the evolution of piRNAs in an animal that is not too distantly related to humans (chickens and humans both have backbones, and are therefore both warm-blooded vertebrates).

Sun et al. examined cells from the testicles of domestic chickens and red jungle fowl as an example of the role of piRNAs in protecting genetic information in vertebrates. The investigation revealed that piRNAs against all previously trapped viruses in the chicken's genetic code are produced in chickens to stop them from causing more damage. Sun et al. also observed the creation of piRNAs in chickens in response to ALV that had not yet become trapped in the chicken's genetic code. Importantly, the piRNAs could control these retroviruses while they were still infectious.

The experiments also revealed that piRNAs against ALV are produced from a single copy of ALV that is found in both domestic and wild chickens. The results showed that cells can produce new piRNAs using these pre-existing viral copies within their own genetics. This illustrates that production of piRNA from existing genetic material can be activated in response to certain cues.

Further work will seek to discover how existing genetic information becomes a source of piRNAs. In the United States, 8 billion domestic chickens are consumed each year, and a better understanding of how these birds defend themselves against viral infections could increase the productivity of the poultry industry around the world. Moreover, because other viruses trapped in the chicken's genetic code are related to similar viruses in humans, future discoveries made in this area could help to guide research that will benefit human health as well.

DOI: 10.7554/eLife.24695.002

sequences, an RNA-based immune system similar to CRISPRs in prokaryotes (*Kumar and Chen, 2012*). New piRNA-producing loci can originate by duplication (*Assis and Kondrashov, 2009*), but duplication per se does not directly generate new piRNA sequences. The only known mechanism of new piRNA acquisition comes from studies of fruit flies, in which a TE inserts into an actively expressed piRNA cluster (*Khurana et al., 2011*). However, considering that active piRNA-producing loci represent only a tiny fraction of the genome, and that no preference for insertions of TEs into piRNA-producing loci has been reported (*Kumar and Chen, 2012*), other piRNA acquisition mechanisms remain to be discovered.

Endogenous retroviruses (ERVs) are distributed relative strictly in vertebrate genomes (*Gifford and Tristem, 2003*; *Eickbush and Jamburuthugoda, 2008*). Numerous distinct ERV families have invaded the chicken germ line (*Jurka et al., 2005*), making chicken (*Gallus gallus*) an excellent model to study virus-host interplay. Compared to other TEs in the chicken genome, including the ancient CR1 superfamily (*Vandergon and Reitman, 1994*) and DNA transposons, and largely absent short interspersed nuclear elements (SINEs) (*International Chicken Genome Sequencing Consortium, 2004*), chicken ERVs are more active and have led to phenotypic changes like blue egg-shells (*Wang et al., 2013*) and late feathering (*Boyce-Jacino et al., 1989*). Chicken ERVs can also remain infectious, and may evolve into new viruses through recombination with host genes or exogenous viruses. Avian leukosis virus (ALV) was the first ERV to be discovered (*Temin, 1964*;

Weiss, 1969; *Baluda, 1972*). Uninfected chickens sometimes spontaneously shed infectious ALV subtype E (ALVE) viruses (*Varmus et al., 1972*; *Weiss, 2006*), and the infection can lead to cancer (*Weiss, 2006*). ALV acquisition of a host oncogene, *SRC*, generated a more acute transforming virus, Rous sarcoma virus (RSV) (*Stehelin et al., 1976*). Recombination between ALVE and EAV-HP—a member of the endogenous avian retrovirus (EAV) family, has created ALV subtype J (ALVJ), a new subgroup of ALVs that were associated with myeloid leukosis in meat-type chickens (*Smith et al., 1999*). After spreading to China in 2002, ALVJ mutated to infect egg-laying breeds with a wide spectrum of tumors (*Gao et al., 2010*). Despite the threat of ERVs to the poultry industry, we lack a systematic investigation of ERV activity and piRNA-mediated suppression in the chicken germ line.

Here, we identified the ERV activity and piRNA-producing loci in White Leghorn, the most popular egg-laying breed, dissected the interplay between ERVs and piRNAs, and traced the origination of recently acquired piRNAs in undomesticated wild chickens, Red Jungle Fowl (*Eriksson et al., 2008*). We chose to focus on the White Leghorn because this domestic breed has suffered from ERV activation and thus its ERVs have been extensively studied (*Crittenden, 1991*). White Leghorn lays an average of 280 eggs per year (*Bao et al., 2008*). Breeders have used the late-feathering trait as a convenient marker to select female layers at hatch (*Boyce-Jacino et al., 1989*); however, this trait is linked to a fully infectious ALVE provirus (known as *ev21*) on the Z chromosome, which causes decreased performance (*Smith and Fadly, 1988*; *Fadly and Smith, 1997*). We found that 73 TE families, including ALVE, were active in White Leghorn testis, and all 503 TE insertions absent in Red Jungle Fowl were derived from these TE families. More than 60% of the active TEs belonged to ERV families, indicating that ERVs contribute to most TE activity in chickens. All active TEs are targeted by robust piRNA-mediated suppression. As TEs become inactivated, their targeting piRNAs erode away. We found that the ability of chickens to produce piRNAs targeting ALV is an evolutionarily recent acquisition—White Leghorn produced abundant ALV piRNAs while Red Jungle Fowl did not. The ALV piRNAs in White Leghorn were produced from a truncated ALV provirus that was known to render its host ALV resistant (*Robinson et al., 1981*). The presence of this provirus predated domestication, indicating that the responsible genomic region exists in either an 'on' or 'off' state as a piRNA-producing locus.

Results

Identifying active TEs in domestic fowl

Active ERVs are transcribed and translated, and are able to transpose within the germ line. Comparison of RNA-seq data from 12 tissues (*Brawand et al., 2011*) indicated that chicken ERV families were ubiquitously expressed (*Figure 1—figure supplement 1A*). Because many insertions are truncated, detection of ERV RNAs does not necessarily indicate that they are translated or competent for transposition. It has been shown that chicken ERVs are transcribed and translated in somatic cells (*Bolisetty et al., 2012*). Because the highest expression was in testis and ovary—the only tissues where their expansion can become heritable (*Figure 1—figure supplement 1A*), we decided to perform polysome profile analysis to determine whether ERVs were also being translated in testis. Adult White Leghorn testis lysates were separated in 10–50% sucrose-density gradients by ultracentrifugation (*Figure 1A*). This fractionation separates non-translating ribonucleoproteins, small and large subunits of ribosomes, monosomes and polysomes, as shown by the distribution of rRNA. Actively translated *β-ACTIN, CILI,* and *CIWI* mRNAs co-sedimented with both monosome and polysome fractions, but the *MALAT1* non-coding RNA did not co-sedimented with polysomes. CILI and CIWI are the two PIWI proteins in chickens (*Figure 1—figure supplement 2*). We tested the distribution of *CR1-B* and *CR1-F* families that belong to the CR1 superfamily, as well as the *EAV-HP* and *ALVE* that belong to ERV families. Although CR1 arose prior to the divergence of birds and reptiles and peaked ~45 million years ago (*Vandergon and Reitman, 1994*), CR1-F and CR1-B elements remain able to drive their own transcription (*Wicker et al., 2005*; *Lee et al., 2009*). *CR1-B*, *CR1-F*, *EAV-HP*, and *ALVE* transcripts co-sedimented with polysomes. These profiling results suggest that *CR1-B*, *CR1-F*, *EAV* and *ALVE* insertions are transcribed and translated in testis.

To determine whether the observed co-sedimentation with ribosomes reflects the active translation, we performed ribosome profiling using testis lysates from adult White Leghorn. Ribosome profiling is based on the facts that the ribosome-bound fraction of mRNA is protected from RNase
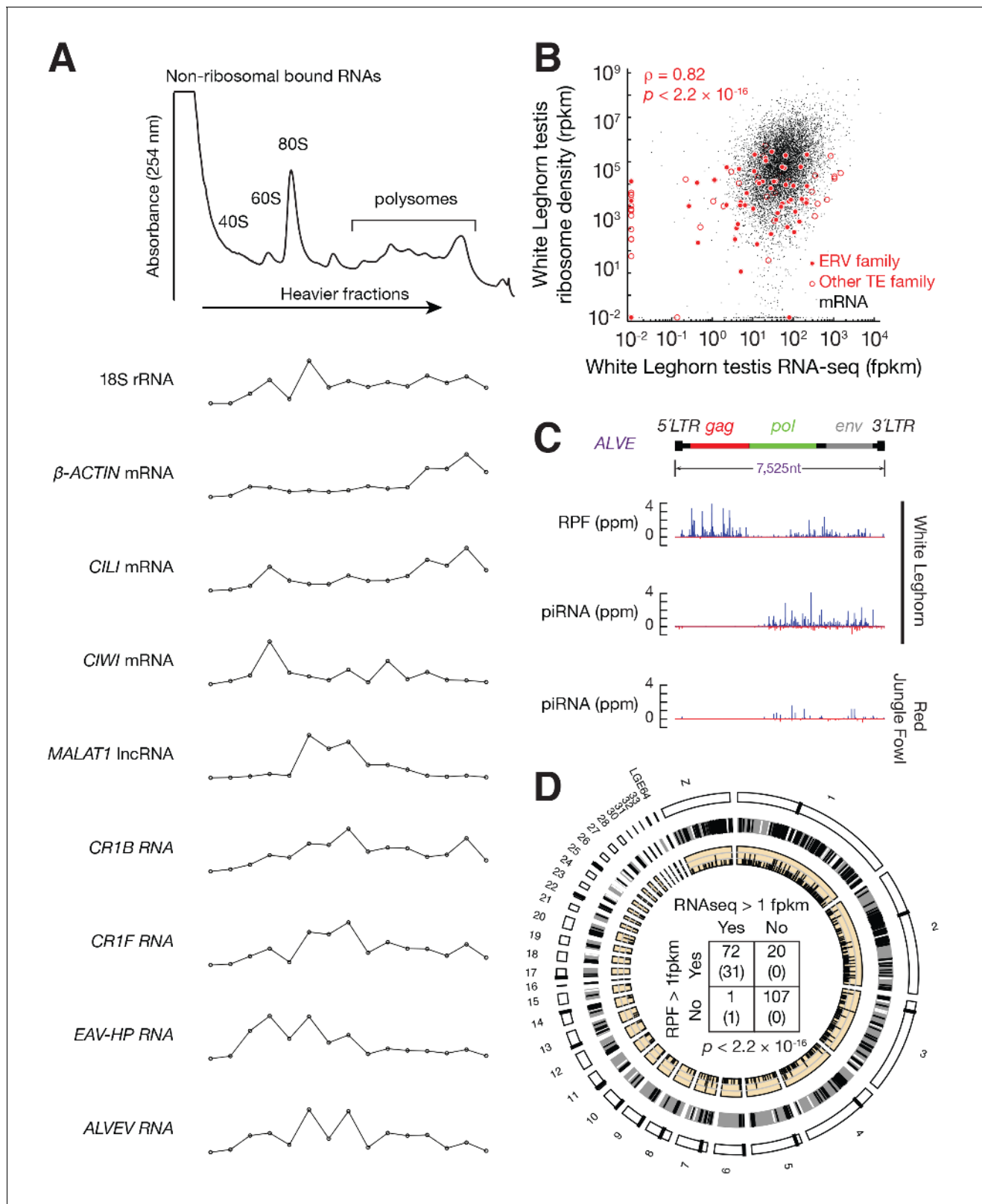
Figure 1. Active ERVs in White Leghorn testis. (**A**) A254 absorbance profile of 10% to 50% sucrose density gradients of testis lysates from adult rooster. From top to bottom, plots show the relative abundance of 18S rRNA, *β-ACTIN* mRNA, *CILI* mRNA, *CIWI* mRNA, chicken *Malat1* lncRNA, *CR1B*, *CR1F*, *EAV-HP*, and *ALVE* quantified by RT-qPCR. Data were normalized to a spike-in control RNA. (**B**) Scatter plots of transcript abundance versus ribosome density. Each black dot represents an mRNA expressed in testis. Each filled red circle represents an ERV family, and each open red circle represents

Figure 1 continued on next page

Figure 1 continued

any other TE family, including DNA transposons and CR1 superfamily; rpkm, reads per kilobase of transcript per million mapped reads; fpkm, fragments per kilobase of transcript per million mapped reads. (**C**) Normalized reads of White Leghorn RPFs (Top), White Leghorn piRNAs (Middle), and Red Jungle Fowl small RNA reads (>23 nt) (Bottom). Blue represents sense mapping reads; Red represents anti-sense mapping reads. The gene organization of ALVE is also shown. *Gag*, group-specific antigen; *Pol*, polymerase; *Env*, envelope protein; ppm, parts per million. (**D**) Circos plot representing the locations, from periphery to center, of cytological position (black lines represent centromeres), piRNA clusters in White Leghorn (Black lines represent conserved piRNA clusters; White lines represent divergent piRNA clusters), putative new insertions discovered by TEMP (tiles) using genomic resequencing of White Leghorn, and 2 × 2 contingency table for Fisher's exact test to assess the significance of the coincidence of transcription and translation of each TE family. The table data correspond to the number of TE families in each category and, in parentheses, the number of TE families in each category with recent transpositions.

The following figure supplements are available for figure 1:

Figure supplement 1. Tissue distribution of ERVs and piRNA pathway genes.

Figure supplement 2. PIWI proteins are conserved between mammals and birds.

Figure supplement 3. Ribosome profiling in adult rooster testes.

Figure supplement 4. A recent ALVE insertion in the *SOX5* gene detected by genome-resequencing of White Leghorn.

digestion in vitro (*Steitz, 1969*), and that the subsequent genome-wide sequencing of ribosome-protected fragments (RPFs) provides a snapshot of in vivo translation (*Ingolia et al., 2009*). RNA fragments protected from RNase A and T1 digestion were isolated from 80S fractions and sequenced (*Figure 1—figure supplement 3A*) (*Ricci et al., 2014*; *Cenik et al., 2015*). Similar to reported RPF sizes in mammals, the RPF sizes in chicken from coding DNA sequences (CDS) ranged from 26–32 nt (*Figure 1—figure supplement 3B*). While RNA-seq reads were distributed throughout the entire set of mRNA transcripts, RPF reads were enriched in CDS regions (*Figure 1—figure supplement 3C*), and RPFs that mapped to CDS regions accounted for 96% of the RPFs that mapped to entire mRNA transcripts. The RPF reads mapping to open read frames displayed an obvious three-nt periodicity (*Figure 1—figure supplement 3D*), reflecting the triplet nature of the genetic code during translation elongation. Based on the enrichment of RPFs at CDS regions and the observed codon periodicity of RPFs, we conclude that the ribosome profiling identified RNAs undergoing translation.

We integrated the ribosome profiling and RNA-seq data in our analysis of transcription and translation of ERVs and other TE families. The ribosome density of each TE family correlated with their steady-state RNA levels ($\rho = 0.82$, $p < 2.2 \times 10^{-16}$) (*Figure 1B*). The median translational efficiency (ratio of ribosome density to transcript abundance) of ERVs was around 1/10 of the median translation efficiency of mRNAs that were expressed in testis. Consistent with our expectation that ALVE was active in White Leghorn, we detected RPFs mapping to ALVE (*Figure 1C*). These ALVE RPF reads displayed a length distribution that was similar to that of RPFs from CDS regions (*Figure 1—figure supplement 3B*), and they displayed codon periodicity (*Figure 1—figure supplement 3D*). These RPFs were distributed throughout the entire ALVE transcripts but with higher abundance at *gag* and *env* than at *pol* (*Figure 1C*). Most transcribed TEs were also translated (the two events, transcription and translation, were significantly associated: Fisher's exact test, $p < 2.2 \times 10^{-16}$; *Figure 1D*). We detected RPFs in 71 of 73 TE families that were transcribed (97.3%). Sometimes RPFs could not be unambiguously assigned to TEs due to their small sizes, resulting in false positive signals on transcriptionally silenced TEs. Based on RNA-seq data, polysome profiles, and ribosome profiling, we conclude that most transcribed TE families in the testis were also being translated.

To detect new transposition events, we aligned the published resequencing data of the White Leghorn genome with ~100X coverage (*Oh et al., 2016*) to the Red Jungle Fowl reference genome. 503 putative TE insertions, absent in Red Jungle Fowl, were distributed throughout the chicken genome (*Figure 1D*, *Supplementary file 1*). Although Red Jungle Fowl are commonly called as the 'ancestor' of domestic chicken, they evolved thousands of years in parallel with domestic chicken after chicken domestication (*West and Zhou, 1988*), therefore our analysis cannot distinct lineage

specific insertions in White Leghorn from lineage specific deletions in Red Jungle Fowl using structure variant alone. De novo ALVE insertions have been reported in domestic fowl (*Crittenden, 1991*), and we detected new ALVE insertions in *SOX5*, as recently reported (*Rutherford et al., 2016*). The identification of a 6 bp target site duplication typical for ALVE (*Figure 1—figure supplement 4*) indicates a transposition event rather than a genome duplication event. No putative new insertion came from the 127 transcriptionally inactive TE families (*Figure 1D*), confirming their inactive state. All putative new insertions came from the 73 actively transcribed TE families. Thus, combing multiple methods to detect active ERVs, we found that active TEs and their insertions in the White Leghorn genome (*Supplementary file 1*). Our data indicated that transposing activity of some TEs has been recent or may have been ongoing.

Chicken piRNAs reflect contemporary TE activity

Active TEs must be tightly controlled in germ cells. Using small RNA-seq data from the adult testis of White Leghorn (*Li et al., 2013*), we detected abundant TE piRNAs, which accounted for 7.8% of total piRNAs, and exhibited a size range peaking at 24–25 nt. These small RNAs were resistant to oxidation (*Figure 2—figure supplement 1A,B*). Oxidation by sodium periodate makes most small RNA species non-accessible for cloning into libraries, but 2′-*O*-methyl-modified 3′ termini protect piRNAs from oxidation (*Ghildiyal et al., 2008*). Like piRNAs in other species, these TE piRNAs typically began with uracil (61.6% of species and 66.7% of reads, *Figure 2—figure supplement 1C*). Almost equal numbers of piRNAs mapped to sense versus antisense strands (median ratio of sense to antisense piRNAs was 1.2) (*Figure 2—figure supplement 1D*), and there was an adenine bias at the 10th position (*Figure 2—figure supplement 1C*), indicating that secondary piRNAs are generated (*Brennecke et al., 2007*; *Gunawardane et al., 2007*). To test whether the anti-sense TE piRNAs were able to guide the PIWI proteins to cleave TE transcripts, we plotted the distance between the 5′ends of anti-sense piRNAs and the 5′ends of sense piRNAs from TE loci. We detected a significant Z score at a distance of 10 nt, a signature of robust Ping-Pong amplification (*Figure 2—figure supplement 1E*) (*Brennecke et al., 2007*; *Gunawardane et al., 2007*). These findings indicate that a piRNA mediated silencing pathway against TEs is active in the chicken germ line.

The expression of piRNAs that target each TE family correlated with overall TE expression ($\rho = 0.81$, $p < 2.2 \times 10^{-16}$, *Figure 2A*; the two events were significantly associated: Fisher's exact test, $p < 2.2 \times 10^{-16}$, *Figure 2B*), although there were exceptions. All the 73 actively expressed TE families were targeted by piRNAs. The presence of this piRNA activity explains why expression of the TEs can be tolerated in White Leghorn. Most inactive TEs (108 families) are not targeted by piRNAs; interestingly, 19 inactive TEs are targeted by piRNAs. Those piRNAs that target inactive TEs exhibit the authentic piRNA length distribution, resistance to oxidation, and first position (1st) U bias (*Figure 2—figure supplement 2A,B*), but they are less abundant than the piRNAs that target active TEs ($p < 2.2 \times 10^{-16}$) (*Figure 2C*). piRNAs that target active TEs display a robust Ping-Pong amplification with a median Z score of 12.2; whereas, the piRNAs that target the inactive TEs do not show significant Ping-Pong amplification (median Z score of 0.41) (*Figure 2C*; Z-score >3.3 corresponds to p<0.01), although both sense and antisense TE piRNAs are produced in equal abundance (*Figure 2—figure supplement 1B*). The lack of a Ping-Pong signature for piRNAs targeting inactive TEs supports the function of Ping-Pong amplification as an adaptive response to TE activation rather than merely a consequence of piRNA production.

Since inactive TEs are no longer a threat to the host genome, piRNAs that target them could either represent a remnant from suppression of past threats, or have acquired new functions beyond TE suppression. To distinguish these two possibilities, we grouped TEs based on their expression and based on TE piRNA expression (*Figure 2A,B*), and compared TE age (*Figure 2D*). If inactive TEs that were targeted by piRNAs have an intermediate age between active TEs and inactive TEs that were not targeted by piRNAs, the targeting more likely reflects a remnant of prior suppression function; if inactive TEs that were targeted by piRNAs are as old as other inactive TEs, it is more likely that these TEs and piRNAs represent possible new functions. We inferred TE age using organism information available in Repbase (*Jurka et al., 2005*). We found that all active TEs that were targeted by piRNAs are recent invaders of the chicken genome. These young TEs are specific to *Gallus gallus* or other *Gallus* genera. ERVs comprised most of these young TEs (47 out of 73 families), which is consistent with the observations that non-ERV TEs lack recent activity in chickens (*International Chicken Genome Sequencing Consortium, 2004*). More than 90% of inactive TEs
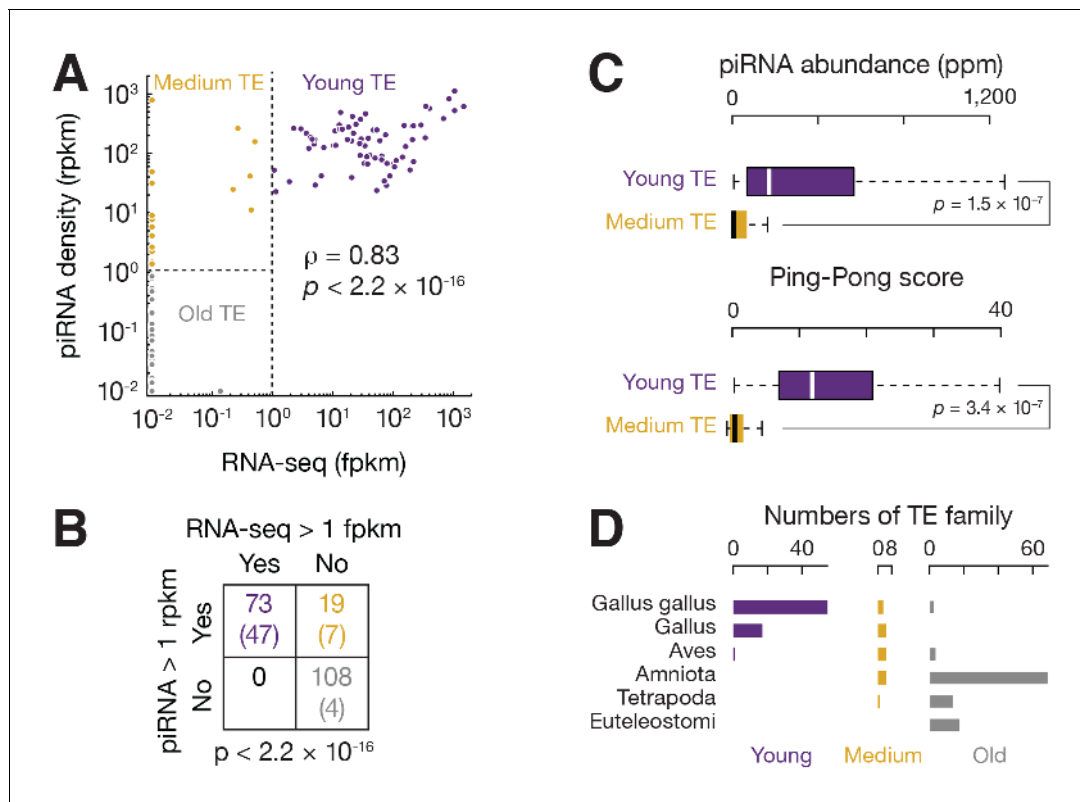
Figure 2. Three groups of TEs based on TE expression and piRNA expression. (A) Scatter plots of TE transcript abundance versus TE piRNA abundance. Each filled circle represents a TE family. Here and in *Figure 2—figure supplements 1* and *2*, young TE in purple, medium TE in yellow, and old TE in grey. (B) 2 × 2 contingency table for Fisher's exact test to assess the significance of the coincidence of the TE transcript abundance and TE piRNA abundance. The table data correspond to the number of TE families in each category and, in parentheses, the number of ERV families in each category. (C) Top, box plots present piRNA abundance per TE family. Bottom, box plots present Ping-Pong amplification score per TE family. (D) Histograms of TE ages.

DOI: 10.7554/eLife.24695.008

The following figure supplements are available for figure 2:

Figure supplement 1. piRNA-mediated TE suppression in rooster testes.
DOI: 10.7554/eLife.24695.009

Figure supplement 2. Medium TE piRNAs are authentic piRNAs.
DOI: 10.7554/eLife.24695.010

that were not targeted by piRNAs had invaded the chicken genome before birds and other amniotes diverged. These old TEs included 80 DNA transposons, 15 CR1s, and 4 ERVs. We found that the 19 inactive TEs that were targeted by piRNAs were of medium age, exhibiting invasion times that were distinct from both old and young TEs (*Figure 2D*, χ^2, $p \leq 2.5 \times 10^{-9}$). From these data, we conclude that piRNA expression reflects TE age—young TEs are targeted by abundant piRNAs, while TE inactivation leads to the erosion of piRNA production. Thus, we designate three TE groups: young, medium, and old based on the expression pattern of TEs and TE piRNAs (*Figure 2A,B*). Our data imply that a rapid turnover of chicken piRNA sequences reflects contemporary TE activity.

ALVE-targeting piRNAs are found in domestic but not wild chickens

piRNAs have not previously been shown to suppress any infectious virus—exogenous or endogenous—in vertebrates; yet, we unexpectedly detected abundant piRNAs mapping to ALVE in adult rooster testis (*Figure 3A*). These ALVE-mapping reads, which peaked at 25 nt were resistant to oxidation, indicating that they were authentic piRNAs. These piRNAs mapped to both strands of ALVE (*Figure 1C*), spanning *env* and the 3′ half of *pol*. The sense piRNAs exhibited a typical 1st U bias, and the antisense piRNAs exhibited a 10th A bias (*Figure 2—figure supplement 2C*), indicating
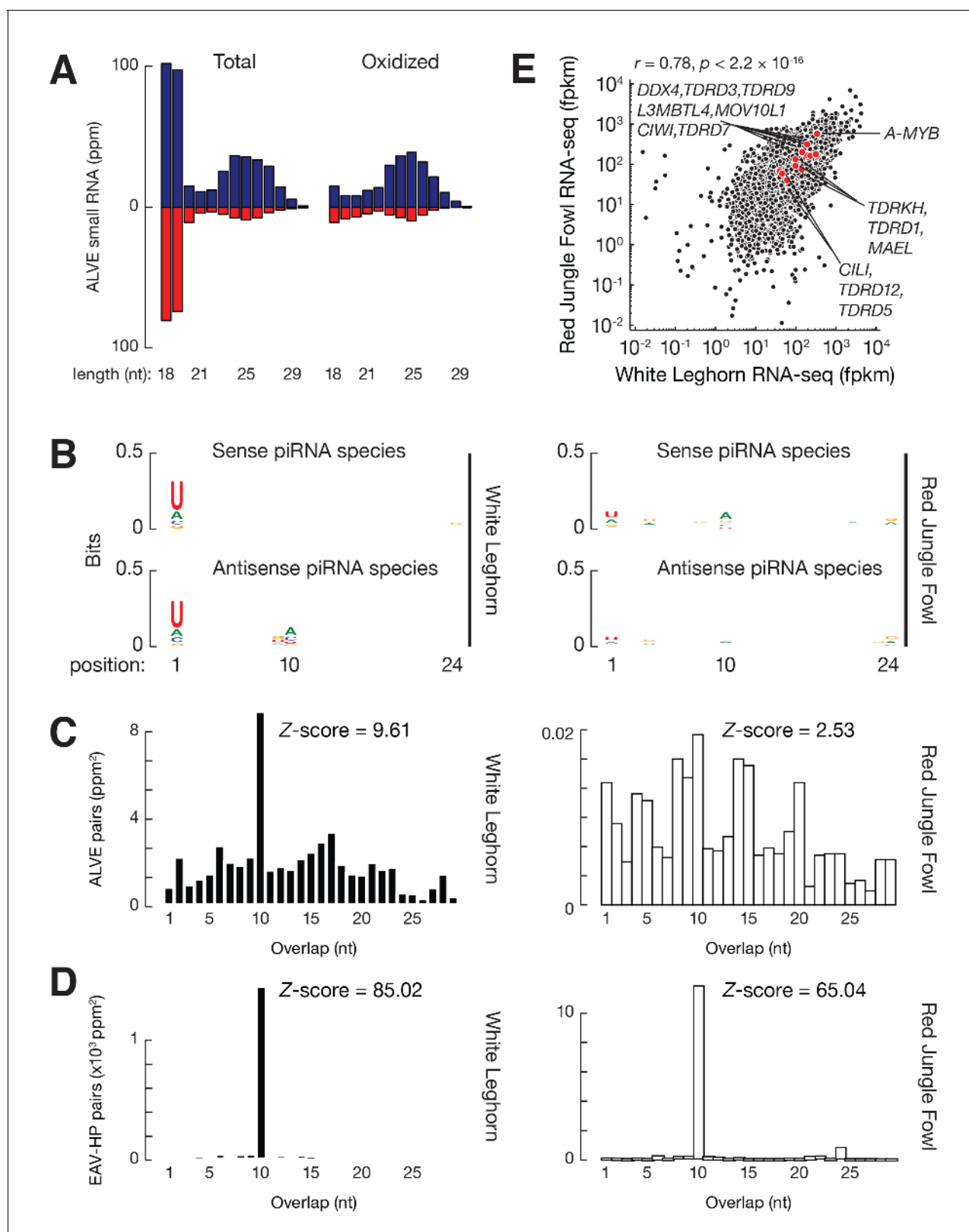
Figure 3. ALVE piRNA acquisition. (**A**) Length distributions of testis small RNAs mapping to ALVE. Blue represents sense mapping piRNAs; Red represents anti-sense mapping piRNAs. (**B**) Sequence logo showing the nucleotide composition of ALVE piRNA species from White Leghorn (*left*) and ALVE species from Red Jungle Fowl (*right*), top, sense ALVE mapping reads, bottom, anti-sense ALVE mapping reads. (**C**) Analysis of the 5′−5′ overlap between ALVE piRNAs from opposite strands was analyzed. Significance of ten-nucleotide overlap ('Ping-Pong') was determined from Z-score.
Figure 3 continued on next page

production of secondary piRNAs. Indeed, robust Ping-Pong amplification signals were detected in ALVE piRNAs (*Figure 3B*). ALVE piRNAs were produced to an abundance of 188 parts per million reads mapped to the genome (ppm), which was roughly half the abundance of EAV-HP piRNAs (359 ppm). The EAV family underwent endogenization prior to *Gallus* speciation, and although its members no longer produce viral particles, they actively transpose and can cause new insertions (*Boyce-Jacino et al., 1992*) as reported in *Supplementary file 1*. Because small RNAs recognize their targets without complete sequence complementarity (*Bartel, 2009*), the roughly 3000 ALVE piRNA species detected can recognize mutated ALVE, which thus might explain why other ALV family members failed to endogenize. The presence of these piRNAs likely improves fitness both by suppressing the mutagenic effects of germ-line activation, and by reducing the numbers of viral copies thereby reducing horizontal transmissions.

ALVE was introduced into the chicken genome following speciation but prior to domestication (*Frisby et al., 1979*). The Red Jungle Fowl genome carries one full length ALVE, ALVE-JFvB (*Weiss and Biggs, 1972*), and one truncated copy, ALVE6 (known as ALVE-JFvA in Red Jungle Fowl and *ev6* in White Leghorn) (*Levin et al., 1994*; *Benkel and Rutherford, 2014*). Using published RNA-seq data from the testis of Red Jungle Fowl (*Necsulea et al., 2014*), we determined the expression of ALVE to be 17.1 fragments per kilobase of transcript per million mapped reads (fpkm), which is roughly one-third of the abundance of EAV-HP (49.3 fpkm). Based on 41 single nucleotide polymorphisms (SNPs) that distinguish ALVE-JFevB and ALVE6, we determined that the two copies were expressed at a 1:5 ratio. Given the level of testicular expression of ALVE, it was surprising that we did not detect robust ALVE piRNAs in Red Jungle Fowl (*Figure 1C*). These ALVE-mapping reads in Red Jungle Fowl had neither a strong U bias (*Figure 3B*), nor significant Ping-Pong amplification (*Figure 3C*). The Ping-Pong analysis method based on the Z-score of piRNA pairs is not affected by sequencing depth (*Zhang et al., 2011*). Therefore, the absence of robust expression of ALVE piRNAs in Red Jungle Fowl indicates that the germ-line endogenization of a new retrovirus is not sufficient to establish piRNA-mediated repression.

To determine whether piRNAs are increased generally for all TEs or specifically for ALVE in White Leghorn, we tested the expression of EAV-HP piRNAs in Red Jungle Fowl. EAV-HP piRNAs exhibited robust Ping-Pong amplification in both White Leghorn and Red Jungle Fowl (*Figure 3D*). Moreover, compared to the non-TE piRNAs, the overall percentage of TE piRNAs did not increase in White Leghorn (χ^2, p=1). We then compared expression of piRNA pathway genes in White Leghorn and Red Jungle Fowl (*Figure 3E*). RNA silencing pathway genes, including *CIWI, CILI, A-MYB* (*Li et al., 2013*), *DDX4* (*Kuramochi-Miyagawa et al., 2010*), *MAEL* (*Soper et al., 2008*), *L3MBTL4* (*Fagegaltier et al., 2016*; *Sumiyoshi et al., 2016*), *MOV10l1* (*Frost et al., 2010*; *Zheng et al., 2010*), *TDRD1* (*Chen et al., 2009*; *Kojima et al., 2009*; *Reuter et al., 2009*; *Wang et al., 2009*), *TDRKH (TDRD2)* (*Saxe et al., 2013*), *TDRD3, TDRD5* (*Yabuta et al., 2011*), *TDRD7* (*Tanaka et al., 2011*), *TDRD9* (*Aravin et al., 2009*; *Shoji et al., 2009*), and *TDRD12* (*Aravin et al., 2009*; *Shoji et al., 2009*), exhibited a median abundance of 164 fpkm in White Leghorn testis, which was not significantly different from expression in Red Jungle Fowl (median abundance of 167 fpkm, p=0.63). This expression of TE piRNAs and piRNA processing genes at approximately the same levels in Red Jungle Fowl and White Leghorn suggests that the activation of piRNAs in White Leghorn is specific to ALVE. The presence of an active piRNA pathway in Red Jungle Fowl indicates that ALVE piRNA expression emerged or was selected subsequent to domestication.

Defining piRNA-producing loci in chickens

The ability of chickens to produce piRNAs against a new ERV provides a rare opportunity to study where new piRNAs are acquired. To identify the genomic source of the ALVE piRNAs, we defined all piRNA-producing loci, so-called piRNA clusters, in White Leghorn. Using our previously developed dynamic programming algorithm (*Li et al., 2013*), in total, we identified 1633 piRNA clusters that

accounted for 0.88% of the chicken genome, and explained 87.3% of total piRNA reads and 81.1% of uniquely mapping piRNAs (*Figure 4A*). These piRNA clusters were distributed on most autosomes and the Z chromosome (*Figure 1D*). Unlike divergently and uni-directionally transcribed mouse piRNA-producing loci, we observed that chicken piRNAs were produced from both strands of piRNA clusters (*Figure 4B*, *Figure 4—figure supplement 1A*) as reported previously (*Li et al., 2013*; *Chirn et al., 2015*), and were derived from convergently transcribed precursors detected by our RNA-seq data (*Figure 4B*, *Figure 4—figure supplement 1A*). Over 70% of clusters (1173 out of 1633) included uniquely mapping piRNAs transcribed from either strand at a level of greater than 10% of total uniquely mapping piRNAs from that cluster. Based on these findings, we conclude that most chicken piRNA-producing loci are transcribed from both strands, and both transcripts are processed into piRNAs.

We tested tissue-specific expression of piRNA precursors using the RNA-seq data to measure the abundance of piRNA cluster transcripts in testis as well as in 11 other tissues (both sexes were included) (*Figure 4C*). The median abundance of piRNA precursors in the testis was 2.6 fpkm, but we were unable to detect these transcripts in other tissues (median abundance = 0). The expression of two PIWI genes, *CILI* and *CIWI*, was also only detected in testis (*Figure 1—figure supplement 1B*). The detection of piRNA precursors and PIWI mRNAs exclusively in testis is consistent with their role in protecting the germ-line genome. The lack of detection of PIWI mRNAs and piRNA precursors in ovary suggests that chicken piRNA pathways display sexual dimorphism. In mouse and chicken, the synthesis of most piRNA precursors and mRNAs of key piRNA pathway genes is driven by the transcription factor A-MYB (*Li et al., 2013*). We found that *A-MYB* is also expressed exclusively in the testis, which may explain the transcriptional activation of piRNA pathways in chicken (*Figure 4C* and *Figure 1—figure supplement 1B*). Thus, the chicken piRNA pathway is transcriptionally activated predominantly in the testis.

A previously defined ALV-resistance locus produces ALVE piRNAs

Based on our finding that the expression of ALVE piRNAs was acquired recently, we reasoned that the ALVE piRNAs were either derived from new ALVE insertions or activated from pre-existing genomic elements. None of the new ALVE insertions in White Leghorn were found within or near the 1633 identified piRNA-producing loci (*Supplementary file 1*). To assess the second possibility, we systematically compared piRNA cluster locations in White Leghorn and Red Jungle Fowl. Using the same parameters in the dynamic programming algorithm, we defined the piRNA clusters in Red Jungle Fowl. Overall, the genomic location of 72% piRNA-producing loci (1168 of 1633) overlapped between the two breeds, but 468 piRNA clusters are specific to White Leghorn (the right circle in *Figure 4D*). The piRNAs from divergent piRNA clusters exhibited authentic piRNA length distribution, resistance to oxidation, and 1st U bias (*Figure 4—figure supplement 1B,C*). In White Leghorn, the conserved piRNA clusters accounted for 77.4% of uniquely mapping piRNAs, and the divergent piRNA clusters only accounted for only 3.7% of uniquely mapping piRNAs. In Red Jungle Fowl, 82.8% of total piRNAs and 74.9% of uniquely mapping piRNAs (*Meunier et al., 2013*) could be explained by the conserved piRNA clusters (*Figure 4A*). Thus, piRNAs are predominantly produced from identical genomic locations but with notable divergence between the breeds.

One divergent piRNA-producing locus (cluster 719) contains the truncated ALVE provirus (*Figure 4E*), ALVE6. Cluster 719 produces abundant piRNAs (147 ppm) in White Leghorn, but in Red Jungle Fowl produces few piRNAs (*Figure 4E*). ALVE6 has lost its 5′LTR, *gag*, and half of *pol*, eliminating its transcriptional promoter (*Tereba, 1981*). The gene structure matches the distributions of piRNAs on ALVE, which starts in the middle of the *pol* gene (*Figure 1C*). This defective provirus has been associated with ALVE resistance (*Robinson et al., 1981*). The wide distribution of ALVE6 in commercial egg-laying breeds has been believed to reflect selection of nonshedders (*Hayward et al., 1980*; *Kuhnlein et al., 1989*; *Smith et al., 1990b*). ALVE6 is the only known ALVE provirus that is present in both White Leghorn and Red Jungle Fowl (*Levin et al., 1994*). In addition to the resequencing analysis of White Leghorn, we used the longer sequencing reads of Sanger sequencing to confirm that although sequence polymorphisms exist, the genomic structure of the ALVE6 locus surrounding regions was remains the same as the reference locus in Red Jungle Fowl (*Figure 4—figure supplement 2A*). These results indicate that ALVE6 existed in the chicken genome prior to domestication.
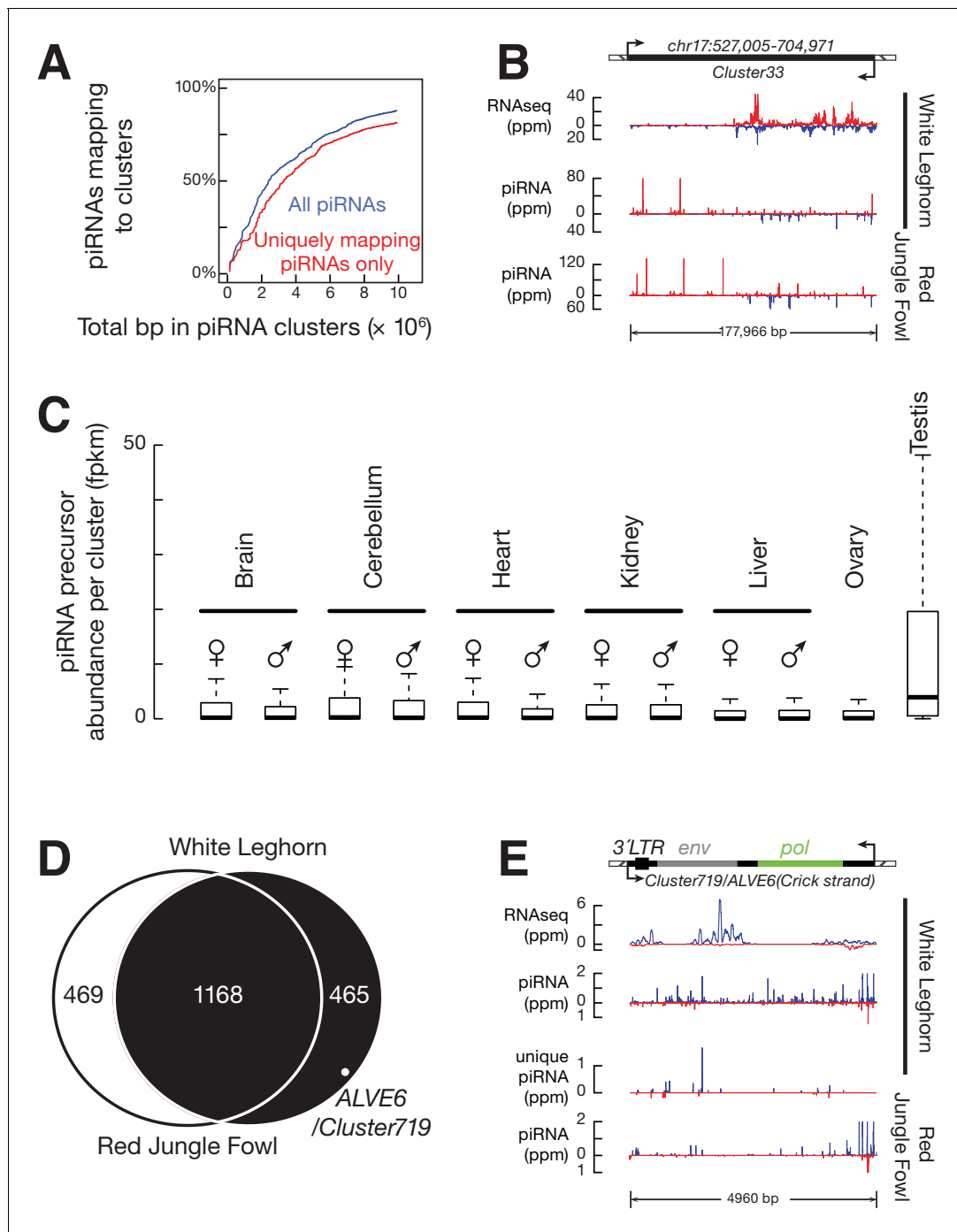
Figure 4. ALVE6 is the primary piRNA-producing locus for viral piRNAs. (**A**) Cumulative distributions for all piRNAs (Blue) and for uniquely mapping piRNAs (Red) on the 1633 piRNA loci in White Leghorn. (**B**) An example of conserved piRNA-producing loci, *Cluster33*, in chicken. Normalized RNA-seq reads of piRNA precursors, piRNAs in White Leghorn, and piRNAs in Red Jungle Fowl. Blue represents Watson-strand piRNAs; Red represents Crick-strand piRNAs. (**C**) Box plots showing abundance of piRNA precursors in 12 chicken tissues. (**D**) Venn diagram showing piRNA clusters defined in Red Jungle Fowl (White) and White Leghorn (Black). (**E**) Normalized RNA-seq reads of piRNA precursors in White Leghorn, White Leghorn piRNAs, unique mapping piRNAs, and Red Jungle Fowl small RNA reads (>23 nt).

The following figure supplements are available for figure 4:

Figure supplement 1. Divergent transcription of piRNA clusters.

Figure supplement 2. ALVE6 existed in chicken genome prior domestication.

Although the ALVE6 locus is defined as a piRNA cluster, it remains possible that ALVE piRNAs in White Leghorn are primarily derived from more recent insertions that occurred during domestication. Each ERV insertion event typically deposits a full-length provirus, as the ERVs are reverse transcribed to double-strand DNAs before their integration into the genome (*Lewinski and Bushman, 2005*). Additionally, an intact ALVE site cannot explain why uniquely mapping piRNAs come from the flanking genomic regions of ALVE6 (*Figure 4E*). Moreover, because ALVE6 has accumulated SNPs that are observed as uniquely mapping reads differentiating ALVE6 from other ALVEs (*Figure 4E*), among the piRNAs overlapped with the 33 SNPs that discriminate new ALVE insertions from ALVE6, 73.3% were expressed from ALVE6 locus and exhibited a pronounced 1st U bias, indicating that ALVE6 was the primary source for ALVE piRNAs. The piRNAs expressed from the new ALVE insertions exhibited a pronounced 10th A bias (*Figure 4—figure supplement 2B*), indicating that they represented secondary piRNAs generated during Ping-Pong amplification of the ALVE piRNAs. Finally, the presence of ALVE6 in chicken suppresses the spontaneous activation of intact ALVE copies, and enhanced fitness has been associated only with truncated ALVE and not with full length ALVE provirus (*Smith et al., 1990a*). Therefore, we conclude that the ALVE piRNAs are primarily produced from the pre-existing ALVE6 locus.

Discussion

In this study, we found that a truncated ALVE provirus gave rise to the piRNAs that target ALVE in White Leghorn. ALVE6 had been identified as a dominant gene that confers resistance to the horizontal spread of spontaneously expressed ALVE (*Robinson et al., 1981*) and to congenital transmission of ALVE (*Smith et al., 1990a*). Although ALVE6 is a defective provirus, and is not infectious, it is highly expressed in domestic fowl (*Hayward et al., 1980*). The truncated provirus produces envelope glycoproteins, and it was proposed that products of ALVE6 compete for cellular receptors (*Robinson et al., 1981*), thus, preventing ALVE replication. However, the expression of envelope proteins from ALVE6 only leads to a 3–4 fold reduction in virus penetration, and does not account for robust resistance to ALVE infection in chickens (*Robinson et al., 1981*). More than 30 years ago, before the discovery of piRNAs, Robinson et al speculated that 'the presence of endogenous virus...would protect the germ line from accumulation of provirus and provirus-associated mutations' (*Robinson et al., 1981*). This type of 'immune response' is also known as viral interference—most hosts are resistant to infection by viruses expressed by their germ-line provirus. Although chicken piRNA mutant is currently not available to estimate the magnitude of restriction rendered by the piRNA pathways, in mouse mutant with disrupted PIWI gene, the transposon expression increased up to 10 fold (*Aravin et al., 2007*). Therefore, our discovery represents a new viral interference mechanism, and provides a critical missing piece to the puzzle: we attribute the ability to protect the chicken germ line, at least partially, to the function of piRNAs produced by truncated ALVEs.

The example presented here of piRNAs targeting an infectious virus in vertebrates represents a previously unappreciated function and history of piRNAs. Generally, there is a clear boundary between TEs and viruses. Most TEs are in a long-term co-evolutionary relationship that minimizes deleterious impacts on the host, but most viruses adopt a more destructive lifestyle that often leads to intense conflict with their hosts (*Feschotte and Gilbert, 2012*). In fruit flies, as a response to different parasites, anti-viral responses are mediated by endogenous siRNAs, and TE silencing is mediated by piRNAs; the functional division of the two small RNA pathways is clear—piRNAs do not appear to play a role in anti-viral defense (*Goic et al., 2013*). Vertebrate genomes, however, contain ERVs (*Malik et al., 2000*), which blurs the boundary between the TEs and viruses. Most ERVs, due to loss-of-function mutations, have lost the ability to make infectious viral particles. This loss occurs evolutionarily, and their infectious capacity is likely maintained for some period of time after endogenization. In addition to ALV in chickens, infectious ERVs have been reported in mammals, including mouse mammary tumor virus (MMTV) (*Moore et al., 1979*), Moloney Murine Leukemia Virus (M-MuLV) (*Stoye and Coffin, 1987*), Koala retrovirus (KoRV) (*Tarlinton et al., 2008*), and Feline ERV-DC (*Anai et al., 2012*). Acting as an essential immune response in germ-line defense, piRNAs would be expected to have evolved before ERVs lost their infectious capacity, in which case piRNAs would have contributed to host defense against the infectious viruses. Our results here expand the function of piRNAs to include resistance to infectious pathogens in vertebrates, and imply that the vertebrate

piRNA pathway has evolved under selection pressures both from the mutagenic effects of TE propagation and from the deleterious effects of activation of infectious ERVs.

The new piRNAs that we report were produced from an existing genomic element, rather than via 'trapping' a new ALVE insertion into a highly expressed piRNA cluster. The birth of new piRNA loci shown here is reminiscent of the origin of new genes from previously noncoding DNA (*Schlötterer, 2015*) that acquire additional regulatory signals for transcription and acquired a functional ORF. The transcription of ALVE6 in White Leghorn could be activated by an adjacent transcriptional promoter as proposed previously (*Tereba, 1981*), or by de novo acquisition of a transcriptional binding motif derived via point mutations. Our detection of ALVE6 expression in Red Jungle Fowl by RNA-seq indicates that transcription alone is not sufficient to become a new piRNA-producing locus. Considering that the genomic structure of ALVE6 is similar between White Leghorn and Red Jungle Fowl, either point mutations or epigenetic changes mark the ALVE6 transcripts for piRNA production in White Leghorn. Although we do not understand the mechanisms that lead to conversion of quiescent genomic regions to emerge as active piRNA producing loci, our work identifies a new mechanism for piRNA acquisition for ERV defense through 'twisting' existing elements.

piRNAs are the most recently discovered family of small silencing RNAs, and questions regarding the biogenesis and function of piRNAs remain. For example, a large proportion of non-TE piRNAs mysteriously enable mammalian sperm production (*Reuter et al., 2011*; *Lim et al., 2015*). Each of the available model organisms, including fruit fly, zebrafish and mouse, exhibits distinct piRNA features, and provides unique insights into piRNA pathway. Chicken piRNAs exhibit unique hybrid features of piRNAs found in other organisms. For example, the convergent transcription of piRNA-producing loci resembles the dual strands in fruit flies (*Brennecke et al., 2007*; *Malone et al., 2009*), but is distinct from that in frog (*Chirn et al., 2015*), zebrafish (*Houwing et al., 2007*), and mice (*Li et al., 2013*). However, unlike piRNAs in fruit flies and zebrafish that derive mainly from TEs, fewer than 10% of chicken piRNAs come from TE regions. The majority of chicken piRNAs come from non-TE regions, similar to piRNAs in adult mouse testis. Thus, the 1633 piRNA-producing loci should provide a valuable resource for the study of chicken piRNAs that will enable us to unify distinct features in model organisms.

Chicken breeding is based on quantitative traits. Putative new insertions, ERV activity, and the capacity to produce piRNAs can be potential contributors to the genetic changes that underlie phenotypic selection. We observed that 58 putative new ERVs inserted into protein coding genes in White Leghorn. Some of these were known to be associated with commercial traits; the others were mapped here for the first time. Each insertion is a mutational event, and has the potential for altering the phenotype. These insertions may contribute to the individual variations of chickens with respect to growth rate, egg production, woody meat, response to heat stress, and resistance to pathogens including newcastle disease virus, avian influenza virus, clostridium, campylobacter, and salmonell. All new insertions are derived from young TEs that are controlled by piRNAs encoded within piRNA-producing loci. Considering the intra-species diversity of piRNA producing loci, both repressive and non-repressive alleles may exist. During selective breeding, it is possible that a genomic region responsible for TE piRNA production is segregated from the active TEs, resulting in TE activation in germ line of F1 generation and increased TE insertions in the offspring of F1. High ALVE levels and increased integrations have been associated with low body weight in chickens (*Ka et al., 2009a*, *2009b*). Therefore, the identified active TEs and piRNA clusters may provide another angle for the discovery of functional polymorphisms underlying quantitative traits, and may also be used to guide breeding to modulate TE activity.

Two observations in our studies indicated the presence of germ-line TE control mechanisms beyond piRNAs. To wit, intact ALVE is present and expressed in Red Jungle Fowl, and upon induction, it can produce infectious viral particles (*Weiss and Biggs, 1972*). Despite this, and despite the absence of ALVE piRNAs, variations of genomic copy numbers of ALVE in Red Jungle Fowl have not been reported. It could be that the somatic TE suppression mechanisms, such as histone modification and DNA methylation, are sufficient to protect Red Jungle Fowl from re-integration. Alternatively, activation, when it occurs, may be extremely deleterious, preventing spread to the general population (*Lu and Clark, 2010*). A second mechanism is suggested by the observation that no piRNA-producing loci or PIWI genes are expressed in the ovary, a site where ERVs are highly expressed. In mice, another small RNA silencing mechanism mediated by endogenous siRNAs is known to protect the murine oocytes from TE attacks (*Tam et al., 2008*; *Watanabe et al., 2008*). As

a consequence, piRNA pathways are not essential for female mouse fertility (*Kuramochi-Miyagawa et al., 2004*; *Carmell et al., 2007*). Although the activation of siRNAs in oocytes is rodent-specific, suggesting that this piRNA-independent defense may not exist in other mammals (*Flemr et al., 2013*; *Rosenkranz et al., 2015*), our studies suggest that sexual dimorphism in piRNA pathway may be a conserved feature.

In conclusion, chicken piRNAs have rapidly evolved to protect the germ-line genome from the contemporary threats. The robust Ping-Pong amplification in piRNAs targeting young TEs reflects an ongoing arms race. When TEs become inactive, the piRNAs gradually erode away, as shown by the low abundance of medium TE piRNAs and the death of piRNAs targeting old TEs. A mystery surrounds the means by which new piRNAs are acquired when a retrovirus is endogenized to a new host. The compact chicken genome, which includes a small fraction of TEs (10%) (*International Chicken Genome Sequencing Consortium, 2004*), permitted pinpointing the ALVE piRNA-producing locus and tracing its evolutionary history. In chickens, ALVE6, as a piRNA-producing locus, exhibits 'on' and 'off' states. Comparative studies among chicken breeds will delineate the molecular events that turn on piRNA production at the ALVE6 locus. The acquisition of piRNA that target a recently invaded ERV, as reported here, represents an opportunity to elucidate the mechanisms by why some transcripts produce piRNAs while some do not.

Materials and methods

Animals
Rooster testes from a 15 months-old White Leghorn of the Cornell Special C strain were used for polysome gradients, ribosomal profiling, RNA-seq, and genomic PCR. The same strain was used to construct the small RNA libraries in our previous studies (*Li et al., 2013*).

Polysome profiling
Testes were flash frozen in liquid nitrogen, and lysed in 1 ml lysis buffer (10 mM Tris-HCl, pH 7.5, 5 mM MgCl$_2$, 100 mM KCl, 1% Triton X-100, 2 mM DTT, 100 µg/ml cycloheximide, and 1× Protease-Inhibitor Cocktail). Lysates were homogenized with a pellet pestle for a total of ten strokes, and incubated at 4°C with inverted rotation for 10 min. The lysates were centrifuged at 1300 *g* for 10 min at 4°C, the supernatant was recovered, and the absorbance at 260 nm was measured. Five A$_{260}$ absorbance units were used for polysome gradients and ribosome profiling.

Samples were loaded on a 10–50% (w/v) linear sucrose gradient (20 mM HEPES-KOH, pH 7.4, 5 mM MgCl$_2$, 100 mM KCl, 2 mM DTT, 100 µg/ml of cycloheximide) and centrifuged in a SW-40ti rotor at 35,000 rpm for 2 hr 40 min at 4°C. Samples were then collected from the top of the gradient using the gradient Fractionation system (BR-188, Brandel, Boca Raton, FL, USA) while monitoring absorbance at 254 nm was measured.

Synthetic spike-in RNAs were added to each collected fraction before RNA purification. The collected fractions were incubated at 42°C in 1% SDS and proteinase K (200 µg/ml) for 45 min. After proteinase K treatment, RNAs were extracted with one volume of Acid phenol (pH 4.5)/chloroform/isoamyl alcohol (25:24:1). The recovered aqueous phase was supplemented with 20 µg glycogen and precipitated with three volumes of 100% ethanol at 4°C for 1 hr. Pellets were washed with 70% ethanol, and RNAs were resuspended in water.

Ribosome profiling
Ribosome profiling was performed as described (*Guo et al., 2010*; *Ingolia et al., 2012*; *Ricci et al., 2014*; *Cenik et al., 2015*) with the following modifications: Cleared testis lysates were incubated with 60 units of RNase T1 (Fermentas, Waltham, MA, USA) and 100 ng of RNase A (Ambion, Waltham, MA, USA) per A$_{260}$ unit for 30 min at room temperature. Samples were loaded on a 10–50% (w/v) linear sucrose gradient, and after centrifuged, the fractions corresponding to 80S monosomes were recovered.

Ribosome profiling Illumina-compatible sequencing libraries were prepared as follows (*Figure 1—figure supplement 3A*): (i) the RPFs were resolved on a 15% acrylamide (19:1) 8 M urea denaturing gel for 1 hr 30 min at 35 W, and fragments ranging from 26 nt to 35 nt were size-selected from the gel; (ii) size-selected RNAs were extracted from the gel slice by electro elution using GeBAflex tubes

(Gerad Biotech, Oxford, OH, USA), and the rRNAs were removed by Ribo-Zero Gold (Epicentre Bio-technologies, Madison, WI, USA); (iii) the 3′ ends of recovered RNAs were dephosphorylated by T4 PNK (New England BioLabs, Ipswich, MA, USA) in MES buffer (100 mM MES-NaOH pH 5.5, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol and 300 mM NaCl) at 37°C for 3 hr, followed by Alkaline Phos-phatase (New England BioLabs) treatment at 37°C for 1 hr; (iv) dephosphorylated RNAs were used in our small RNA library construction protocol with an additional step of 5′ end phosphorylation by T4 PNK (New England BioLabs) using the PNK buffer with 1 mM ATP at 37°C for 1 hr before 5′ ligation.

RNA-seq

Strand-specific RNA-seq libraries were constructed following the TruSeq RNA sample preparation protocol as previously described (*Li et al., 2013*). rRNAs were depleted from total RNAs by Ribo-Zero Gold (Epicentre Biotechnologies, Madison, WI, USA). The library was sequenced using the paired-end 2 × 50 nt platform on a HiSeq 2000.

Quantitative real-time PCR (qRT–PCR)

Extracted RNAs were treated with Turbo DNase (Thermo Fisher, Waltham, MA, USA) for 20 mins at 37°C and then size-selected to isolate RNA $\geq$200 nt (DNA Clean and Concentrator−5, ZYMO RESEARCH, USA) before reverse transcription by SuperScript III (Life Technologies, Carlsbad, CA, USA) at 50°C. Quantitative PCR (qPCR) was performed using the ABI Real-Time PCR Detection System with SYBR Green qPCR Master Mix (Bimake, Houston, TX, USA). Data were analyzed using DART-PCR (*Peirson et al., 2003*). Spike-in RNA was used to normalize RNAs in different fractions. *Supplementary file 1* lists the qPCR primers.

Phylogenetic tree

PIWI protein sequences were obtained from the Ensembl genome browser (SCR_013367). Multiple sequence alignment and neighbour-joining clustering were performed with *clustalw* 2.0.12 (*Thompson et al., 1994*). The R package *ape* (*Paradis et al., 2004*) was used to create the phyloge-netic tree.

TE families

We used 200 chicken TE families that are defined in both Repbase (*Jurka et al., 2005*) and Repeat-Masker (*Smit et al., 2016*, SCR_012954). We downloaded the 233 *Gallus gallus* and ancestral (shared) repeats from Repbase, and first removed the 46 families containing tRNAs, rRNAs, and snRNAs. Because Repbase and RepeatMasker sometimes name TEs differently, we submitted the Repbase repeat sequences to CENSOR (*Kohany et al., 2006*) or to blast to identify the correspond-ing RepeatMasker name. In comparing the TE annotation between RepeatMasker and Repbase, we found that 9 Repbase repeats appeared to be truncations of existing repeats. For example, CAM1_GG appeared to be an incomplete sequence of CR1-C4. Based on the latest chicken genome assembly (Gallus_gallus-5.0), we further removed 12 Repbase repeats did not have corresponding genomic copies. We also noticed that some TEs annotated in the genome by RepeatMasker were not included in the *Gallus gallus* repeats in Repbase. One example is EAV-HP, which is deposited in the archive Repbase21.08, but is classified as being of virus origin rather than chicken origin. We extracted the 34 repeats that are annotated in the current chicken genome by RepeatMasker from the vertebrate archive Repbase. The final total set of 200 TE families and their corresponding names in Repbase and Repeatmasker are listed in *Supplementary file 1*.

General bioinformatics analyses

Analyses were performed using piPipes v1.4 (*Han et al., 2014*). All data from the small RNA-seq, ribosome profiling, RNA-seq, and genome sequencing were analyzed using the latest chicken genome release galGal5 (GCA_000002315.3). Generally, one mismatch is allowed for genome map-ping and three mismatches are allowed for transcriptome mapping. For small RNA analysis, the tran-scriptome included the 200 TE families and 1633 piRNA clusters. For RNA-seq and ribosome profiling analysis, the transcriptome included mRNAs, lncRNAs, piRNA clusters, and TE families.

Supplementary file 1 reports the statistics for the high-throughput sequencing libraries constructed in this study.

In small RNA-seq analysis, reads were mapped to ALVE and EAV-HP sequences before being mapped to the genome, and three mismatches were allowed for alignment. The sequences of EAV-HP and ALVE came from NCBI with id: NC_005947.1 (*Sacco et al., 2000*) and AY013303 (*Johnson and Heneine, 2001*). We analyzed previously published testis small RNA libraries from White Leghorn (GSM1096613) (*Li et al., 2013*), and from Red Jungle fowl (GSM995329) (*Meunier et al., 2013*). Small RNA species with characteristic piRNA length (>23 nt) were defined as piRNAs (*Ghildiyal et al., 2008*). The small RNA libraries from White Leghorn and Red Jungle Fowl were normalized to the sum of all piRNA reads. Oxidized samples were calibrated to the corresponding total small RNA library via the abundance of shared piRNA species. The piRNA abundance per TE or per piRNA cluster is reported either as parts per million reads mapped to the genome (ppm) or reads per kilobase pair per million reads mapped to the genome (rpkm) using a pseudo count of 0.01.

The pair-end total RNA-seq reads were aligned to the genome using TopHat 2.0.12 (*Trapnell et al., 2009*, SCR_013035). Reads were mapped using the '-g 100' flag. The direct transcriptome mapping results were quantified using eXpress (*Roberts and Pachter, 2013*, SCR_006873). The advantage of eXpress lies in the Expectation–Maximization algorithm to apportion multimapping reads, reporting the estimated numbers of fragments in each transcript (*Dempster et al., 1977*). The eXpress results are normalized by the gene compatible reads calculated by Cufflinks per library; and the fpkm (fragments per kilobase of transcript per million mapped reads) value with a pseudo count of 0.01 was used for all analyses. We analyzed our RNA-seq library from testis of White Leghorn and the published RNA-seq libraries from different tissues of Red Jungle Fowl (GSM752557, GSM752558, GSM752559, GSM752560, GSM752561, GSM752562, GSM752563, GSM752564, GSM752565, GSM752566, GSM752567, GSM752568, GSM1064853, GSM1064854, GSM1064855, and GSM1196055) (*Brawand et al., 2011*; *Necsulea et al., 2014*).

Ribosome profiling analysis was done according to the modified small RNA pipeline procedure, but including the junction mapping reads. Ribosome protected fragments (RPFs) 26–32 nt long were selected for further analysis. The RPF abundance per TE or per piRNA cluster was quantified by eXpress, and reported as reads per kilobase pair per million reads mapped to the genome (rpkm) using a pseudo count of 0.01.

The pair-end genome sequencing reads were aligned to the reference genome using BWA-aln (-R 1000) (*Li and Durbin, 2009*, SCR_010910). We analyzed the previously published genome resequencing libraries from White Leghorn (SRX1121834, SRX1121835, SRX1121836) (*Oh et al., 2016*) and we combined the three replicates to increase detection sensitivity. The new transposition events were analyzed by TEMP (*Khurana et al., 2011*; *Zhuang et al., 2014*, SCR_001788). The insertions that are supported by reads at both sides are listed in *Supplementary file 1*.

Statistical analyses were performed in R 3.0.2 (*Team, 2014*, SCR_001905). The significance of the differences was calculated by Wilcoxon rank sum test except as indicated in the text.

Ping-Pong analysis

Ping-Pong amplification was analyzed by the 5′–5′ overlap between piRNA pairs from opposite genomic strands (*Li et al., 2009*). Overlap scores for each overlapping pair were the product of the number of reads of each of the piRNAs from opposite strands. The overall score for each overlap extension (1–30) was the sum of all such products for all chromosomes. Heterogeneity at the 3′ ends of small RNAs was neglected. The Z-score for a 10 bp overlap was calculated using the scores of overlaps from 1–9 and 11–30 as background.

Nucleotide periodicity

Nucleotide periodicity was computed as described (*Pelechano et al., 2015*) with modifications. We first aligned the RPFs to each other using 5′–5′ overlap analysis from the same transcript, and reported the distance spectrum. An annotated ORF is not a prerequisite for this analysis. The distance spectrum of RPFs from mRNAs already showed a 3-nt periodicity pattern. We then transformed the distance spectrum using the 'periodogram' function of the *GeneCycle* package

(*Wichert et al., 2004*) with the 'clone' method. The relative spectral density was calculated by normalizing to the value at the third position.

Rooster piRNA-producing loci detection

We used the same dynamic programming algorithm that we developed previously (*Li et al., 2013*) to identify genomic regions with the highest piRNA density. The oxidized small RNA reads (>23 nt) (SRR772069) were used to define the clusters in White Leghorn, and the small RNA reads (>23 nt) (SRR553601) were used to define the clusters in Red Jungle Fowl. We assumed that piRNA clusters comprise at most 5% of the chicken genome. We first split the genome into one kbp non-overlapping windows, and computed piRNA abundance for each window. The mean of the top 5% of windows was used as the penalty score for the dynamic programming algorithm. The algorithm computes the cumulative piRNA abundance score as a function of the window index along each chromosome. The score at a window is the sum of: the score in the previous window, plus the piRNA abundance in the current window, minus the penalty score; negative scores were reset to 0. The maximum score points to the largest piRNA cluster. We extracted the largest piRNA cluster, recomputed the scores at the corresponding windows, and searched for the next cluster. This process was continued iteratively until the scores for all windows were zero. The boundaries of each cluster were further refined by including those base pairs for which piRNA abundance exceeded the mean piRNA abundance of the top 5% windows. We required a piRNA cluster to have at least one unique mapping read. The coordinates of all 1633 piRNA-producing loci of White Leghorn and whether they are conserved in Red Jungle Fowl are reported in *Supplementary file 1*.

Rooster testis transcriptome annotation

We used Cufflinks v2.2.1 (*Trapnell et al., 2012*, SCR_014597) with parameters of '-u -j 0.2 –min-frags-per-transfrag 40 –overlap-radius 100' to assemble transcripts using the strand specific pair-end RNA-seq data from adult testis of White Leghorn (*Supplementary file 1*). We assembled 59,614 transcripts. Using the TransDecoder/3.3.0 (*Haas et al., 2013*) with the BlastP (retain ORFs with homology to known proteins) and Pfam search (identify common protein domains), we further identified the candidate coding regions of 49,962 mRNA transcripts. We performed our transcriptome analysis on 9505 mRNAs which had an abundance of at least 10 fpkm in testis: among these mRNAs, 9287 were reported in the latest release of RefSeq (GCF_000002315.4, SCR_003496), and 218 were putative novel mRNAs. For lncRNAs, we first selected the 13,103 assembled transcripts that were reported to be lncRNA by RefSeq. Among the 724 lncRNAs with an abundance above 10 fpkm in testis, 218 of these transcripts had ORFs detected by TransDecoder. Following removal these putative false lncRNA, we performed transcriptome analysis on the remaining 502 lncRNAs.

Defining chicken Malat1 sequence using INFERNAL and covariance model

We created a covariance model (CfM) using Infernal v1.1rc1 as previously described (*Nawrocki and Eddy, 2013*, SCR_011809). Briefly, we built a CM based on the human *mascRNA* and *menRNA* alignment using the cmbuild program, which was calibrated for E-value reporting with the cmcalibrate program. We then searched the chicken genome for high scoring hits with the cmsearch program. Default Infernal v1.1rc1 parameters were used for all steps (cmbuild, cmcalibrate, and cmsearch programs). Only one significant hit with an E value below 0.01 was identified by the CM model. This tRNA-like element is from the minus strand of chrUn_AADN03016580:1547–1491 with the E value of 3.8×10^{-9}. Manual inspection of its upstream sequence revealed a MALAT 3′ end like module with two T-rich motifs: TTTTCTTTT and TTTTGCTTTT, and one polyA-like moiety: AAAAAAAGCAAAA. This contig contains 6473 bp and does not harbor TEs. While ESTs mapped of this tRNA-line element, hundreds ESTs mapped to sites spanning the entire upstream region (chrUn_AADN03016580:1492–6473), suggesting that the promoter of this gene lies outside of this contig. Despite the lack of syntenic information, it has been shown to be a human MALAT1 homolog in chicken. The evolution of MALAT1 lncRNA and its 3′end module is reported in a manuscript under review (*Zhang et al., 2017*).

Data access

Acknowledgements

Additional information

Funding

Author contributions

Author ORCIDs

Yu Huining Sun, http://orcid.org/0000-0003-0333-2898
Xiaoyu Zhuo, http://orcid.org/0000-0002-1400-5609
Xin Zhiguo Li, http://orcid.org/0000-0001-6803-4600

Additional files

Supplementary files

• Supplementary file 1. Detailed information and statistics for the sequencing data used in this study. (A) Ribosome profiling sequencing statistics: reads and species. (B) Small RNA sequencing statistics: reads and species. (C) RNA-Seq statistics: reads and species. (D) 200 TE families. (E) TE insertions defined by TEMP. (F) Genome coordinates for the 1633 rooster piRNA-producing loci defined in this study are provided in UCSC BED format (i.e., 0-based) for galGal5. (G) Primers used in this study for qRT-PCR and genomic PCR.
DOI: 10.7554/eLife.24695.015

Major datasets

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Yu Huining Sun, Xin Zhiguo Li	2017	Domestic chickens activate a piRNA defense againstavian leukosis virus	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93559	Publicly available at NCBI Gene Expression Omnibus (accession no: GSE93559)

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Li XZ, Roy CK, Zamore PD	2013	An Ancient Transcription Factor Initiates the Burst of piRNA Production During Early Meiosis in Mouse Testes	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45049	Publicly available at NCBI Gene Expression Omnibus (accession no: GSE45049)
Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H	2013	Evolution of mammalian miRNA genes	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40499	Publicly available at NCBI Gene Expression Omnibus (accession no: GSE40499)
Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW, Zeller U, Khaitovich P, Grützner F, Bergmann S, Nielsen R, Pääbo S, Kaessmann H	2011	The evolution of gene expression levels in mammalian organs	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30352	Publicly available at NCBI Gene Expression Omnibus (accession no: GSE30352)
Necsulea A, Soumillon M, Liechti A, Daish T, Zeller U, Baker J, Grutzner F, Kaessmann H, Warnefors M	2014	The evolution of lncRNA repertoires and expression patterns in tetrapods	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE43520	Publicly available at NCBI Gene Expression Omnibus (accession no: GSE43520)
Oh D, Son B, Mun S, Oh MH, Oh S, Ha J, Yi J, Lee S, Han K	2016	Whole genome resequencing for 3 different domesticated chicken breeds (White Leghorn, Korea domestic and Araucana)	https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP061672	Publicly available at NCBI Sequence Read Archive (accession no: SRP061672)

References

Anai Y, Ochi H, Watanabe S, Nakagawa S, Kawamura M, Gojobori T, Nishigaki K. 2012. Infectious endogenous retroviruses in cats and emergence of recombinant viruses. *Journal of Virology* **86**:8634–8644. doi: 10.1128/JVI.00280-12, PMID: 22674983

Aravin AA, Hannon GJ. 2008. Small RNA silencing pathways in germ and stem cells. *Cold Spring Harbor Symposia on Quantitative Biology* **73**:283–290. doi: 10.1101/sqb.2008.73.058, PMID: 19270082

Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**:744–747. doi: 10.1126/science.1142612, PMID: 17446352

Aravin AA, van der Heijden GW, Castañeda J, Vagin VV, Hannon GJ, Bortvin A. 2009. Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS Genetics* **5**:e1000764. doi: 10.1371/journal.pgen.1000764, PMID: 20011505

Assis R, Kondrashov AS. 2009. Rapid repetitive element-mediated expansion of piRNA clusters in mammalian evolution. *PNAS* **106**:7079–7082. doi: 10.1073/pnas.0900523106, PMID: 19357307

Baluda MA. 1972. Widespread presence, in chickens, of DNA complementary to the RNA genome of avian leukosis viruses. *PNAS* **69**:576–580. doi: 10.1073/pnas.69.3.576, PMID: 4335067

Bao W, Chen G, Li B, Wu X, Shu J, Wu S, Xu Q, Weigend S. 2008. Analysis of genetic diversity and phylogenetic relationships among red jungle fowls and chinese domestic fowls. *Science in China Series C: Life Sciences* **51**:560–568. doi: 10.1007/s11427-008-0076-y, PMID: 18488177

Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**:215–233. doi: 10.1016/j.cell.2009.01.002, PMID: 19167326

Benkel B, Rutherford K. 2014. Endogenous avian leukosis viral loci in the red jungle fowl genome assembly. *Poultry Science* **93**:2988–2990. doi: 10.3382/ps.2014-04309, PMID: 25306461

Bolisetty M, Blomberg J, Benachenhou F, Sperber G, Beemon K. 2012. Unexpected diversity and expression of avian endogenous retroviruses. *mBio* **3**:e00344. doi: 10.1128/mBio.00344-12, PMID: 23073767

Boyce-Jacino MT, O'Donoghue K, Faras AJ. 1992. Multiple complex families of endogenous retroviruses are highly conserved in the genus Gallus. *Journal of Virology* **66**:4919–4929. PMID: 1321278

Boyce-Jacino MT, Resnick R, Faras AJ. 1989. Structural and functional characterization of the unusually short long terminal repeats and their adjacent regions of a novel endogenous avian retrovirus. *Virology* **173**:157–166. doi: 10.1016/0042-6822(89)90231-6, PMID: 2815581

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW, Zeller U, Khaitovich P, Grützner F, Bergmann S, Nielsen R, Pääbo S, Kaessmann H. 2011. The evolution of gene expression levels in mammalian organs. *Nature* **478**:343–348. doi: 10.1038/nature10532, PMID: 22012392

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila. Cell* **128**:1089–1103. doi: 10.1016/j.cell.2007.01.043, PMID: 17346786

Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental Cell* **12**:503–514. doi: 10.1016/j.devcel.2007.03.001, PMID: 17395546

Cenik C, Cenik ES, Byeon GW, Grubert F, Candille SI, Spacek D, Alsallakh B, Tilgner H, Araya CL, Tang H, Ricci E, Snyder MP. 2015. Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Research* **25**:1610–1621. doi: 10.1101/gr.193342.115, PMID: 26297486

Cenik ES, Zamore PD. 2011. Argonaute proteins. *Current Biology* **21**:R446–R449. doi: 10.1016/j.cub.2011.05.020, PMID: 21683893

Chen C, Jin J, James DA, Adams-Cioaba MA, Park JG, Guo Y, Tenaglia E, Xu C, Gish G, Min J, Pawson T. 2009. Mouse piwi interactome identifies binding mechanism of tdrkh tudor domain to arginine methylated miwi. *PNAS* **106**:20336–20341. doi: 10.1073/pnas.0911640106, PMID: 19918066

Chirn GW, Rahman R, Sytnikova YA, Matts JA, Zeng M, Gerlach D, Yu M, Berger B, Naramura M, Kile BT, Lau NC. 2015. Conserved piRNA expression from a distinct set of piRNA cluster loci in eutherian mammals. *PLOS Genetics* **11**:e1005652. doi: 10.1371/journal.pgen.1005652, PMID: 26588211

Crittenden LB. 1991. Retroviral elements in the genome of the chicken: implications for poultry genetics and breeding. *Critical Reviews in Poultry Biology* **3**:73–109.

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*:1–38.

Eickbush TH, Jamburuthugoda VK. 2008. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Research* **134**:221–234. doi: 10.1016/j.virusres.2007.12.010, PMID: 18261821

Eriksson J, Larson G, Gunnarsson U, Bed'hom B, Tixier-Boichard M, Strömstedt L, Wright D, Jungerius A, Vereijken A, Randi E, Jensen P, Andersson L. 2008. Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics* **4**:e1000010. doi: 10.1371/journal.pgen.1000010, PMID: 18454198

Fadly AM, Smith EJ. 1997. Role of contact and genetic transmission of endogenous virus-21 in the susceptibility of chickens to avian leukosis virus infection and tumors. *Poultry Science* **76**:968–973. doi: 10.1093/ps/76.7.968, PMID: 9200232

Fagegaltier D, Falciatori I, Czech B, Castel S, Perrimon N, Simcox A, Hannon GJ. 2016. Oncogenic transformation of Drosophila somatic cells induces a functional piRNA pathway. *Genes & Development* **30**:1623–1635. doi: 10.1101/gad.284927.116, PMID: 27474441

Farazi TA, Juranek SA, Tuschl T. 2008. The growing catalog of small RNAs and their association with distinct argonaute/Piwi family members. *Development* **135**:1201–1214. doi: 10.1242/dev.005629, PMID: 18287206

Feschotte C, Gilbert C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nature Reviews Genetics* **13**:283–296. doi: 10.1038/nrg3199, PMID: 22421730

Flemr M, Malik R, Franke V, Nejepinska J, Sedlacek R, Vlahovicek K, Svoboda P. 2013. A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. *Cell* **155**:807–816. doi: 10.1016/j.cell.2013.10.001, PMID: 24209619

Frisby DP, Weiss RA, Roussel M, Stehelin D. 1979. The distribution of endogenous chicken retrovirus sequences in the DNA of galliform birds does not coincide with avian phylogenetic relationships. *Cell* **17**:623–634. doi: 10.1016/0092-8674(79)90270-8, PMID: 225036

Frost RJ, Hamra FK, Richardson JA, Qi X, Bassel-Duby R, Olson EN. 2010. MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. *PNAS* **107**:11847–11852. doi: 10.1073/pnas.1007158107, PMID: 20547853

Gao YL, Qin LT, Pan W, Wang YQ, Le Qi X, Gao HL, Wang XM. 2010. Avian leukosis virus subgroup J in layer chickens, China. *Emerging Infectious Diseases* **16**:1637–1638. doi: 10.3201/eid1610.100780, PMID: 20875300

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, Zamore PD. 2008. Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* **320**:1077–1081. doi: 10.1126/science.1157396, PMID: 18403677

Gifford R, Tristem M. 2003. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* **26**:291–315. doi: 10.1023/A:1024455415443, PMID: 12876457

Goic B, Vodovar N, Mondotte JA, Monot C, Frangeul L, Blanc H, Gausson V, Vera-Otarola J, Cristofari G, Saleh MC. 2013. RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model Drosophila. *Nature Immunology* **14**:396–403. doi: 10.1038/ni.2542, PMID: 23435119

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation *in* Drosophila. *Science* **315**:1587–1590. doi: 10.1126/science.1140494, PMID: 17322028

Guo H, Ingolia NT, Weissman JS, Bartel DP. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**:835–840. doi: 10.1038/nature09267, PMID: 20703300

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nature Protocols* **8**:1494–1512. doi: 10.1038/nprot.2013.084, PMID: 23845962

Han BW, Wang W, Zamore PD, Weng Z. 2015. piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics* **31**:593–595. doi: 10.1093/bioinformatics/btu647, PMID: 25342065

Hayward WS, Braverman SB, Astrin SM. 1980. Transcriptional products and DNA structure of endogenous avian proviruses. *Cold Spring Harbor Symposia on Quantitative Biology* **44 Pt 2**:1111–1121. doi: 10.1101/SQB.1980. 044.01.120, PMID: 6253186

Houwing S, Berezikov E, Ketting RF. 2008. Zili is required for germ cell differentiation and meiosis in zebrafish. *The EMBO Journal* **27**:2702–2711. doi: 10.1038/emboj.2008.204, PMID: 18833190

Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov DV, Blaser H, Raz E, Moens CB, Plasterk RH, Hannon GJ, Draper BW, Ketting RF. 2007. A role for piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. *Cell* **129**:69–82. doi: 10.1016/j.cell.2007.03.026, PMID: 1741 8787

Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols* **7**:1534–1550. doi: 10.1038/nprot.2012.086, PMID: 22836135

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**:218–223. doi: 10.1126/science.1168978, PMID: 1 9213877

International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**:695–716. doi: 10.1038/ nature03154, PMID: 15592404

Johnson JA, Heneine W. 2001. Characterization of endogenous avian leukosis viruses in chicken embryonic fibroblast substrates used in production of measles and mumps vaccines. *Journal of Virology* **75**:3605–3612. doi: 10.1128/JVI.75.8.3605-3612.2001, PMID: 11264350

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110**:462–467. doi: 10.1159/000084979, PMID: 16093699

Ka S, Kerje S, Bornold L, Liljegren U, Siegel PB, Andersson L, Hallböök F. 2009a. Proviral integrations and expression of endogenous avian leucosis virus during long term selection for high and low body weight in two chicken lines. *Retrovirology* **6**:68. doi: 10.1186/1742-4690-6-68, PMID: 19604406

Ka S, Lindberg J, Strömstedt L, Fitzsimmons C, Lindqvist N, Lundeberg J, Siegel PB, Andersson L, Hallböök F. 2009b. Extremely different behaviours in high and low body weight lines of chicken are associated with differential expression of genes involved in neuronal plasticity. *Journal of Neuroendocrinology* **21**:208–216. doi: 10.1111/j.1365-2826.2009.01819.x, PMID: 19207828

Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, Nowosielska A, Li C, Zamore PD, Weng Z, Theurkauf WE. 2011. Adaptation to P element transposon invasion in Drosophila Melanogaster. *Cell* **147**:1551–1563. doi: 10.1016/j.cell.2011.11.042, PMID: 22196730

Kim VN, Han J, Siomi MC. 2009. Biogenesis of small RNAs in animals. *Nature Reviews Molecular Cell Biology* **10**: 126–139. doi: 10.1038/nrm2632, PMID: 19165215

Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: repbasesubmitter and censor. *BMC Bioinformatics* **7**:474. doi: 10.1186/1471-2105-7-474, PMID: 17064419

Kojima K, Kuramochi-Miyagawa S, Chuma S, Tanaka T, Nakatsuji N, Kimura T, Nakano T. 2009. Associations between PIWI proteins and TDRD1/MTR-1 are critical for integrated subcellular localization in murine male germ cells. *Genes to Cells* **14**:1155–1165. doi: 10.1111/j.1365-2443.2009.01342.x, PMID: 19735482

Kuhnlein U, Sabour M, Gavora JS, Fairfull RW, Bernon DE. 1989. Influence of selection for egg production and Marek's disease resistance on the incidence of endogenous viral genes in White Leghorns. *Poultry Science* **68**: 1161–1167. doi: 10.3382/ps.0681161, PMID: 2554267

Kumar M, Carmichael GG. 1998. Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes. *Microbiology and Molecular Biology Reviews : MMBR* **62**:1415–1434. PMID: 9841677

Kumar MS, Chen KC. 2012. Evolution of animal Piwi-interacting RNAs and prokaryotic CRISPRs. *Briefings in Functional Genomics* **11**:277–288. doi: 10.1093/bfgp/els016, PMID: 22539610

Kuramochi-Miyagawa S, Kimura T, Ijiri TW, Isobe T, Asada N, Fujita Y, Ikawa M, Iwai N, Okabe M, Deng W, Lin H, Matsuda Y, Nakano T. 2004. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131**:839–849. doi: 10.1242/dev.00973, PMID: 14736746

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Takamatsu K, Chuma S, Kojima-Kita K, Shiromoto Y, Asada N, Toyoda A, Fujiyama A, Totoki Y, Shibata T, Kimura T, Nakatsuji N, Noce T, Sasaki H, Nakano T. 2010. MVH in piRNA processing and gene silencing of retrotransposons. *Genes & Development* **24**:887–892. doi: 10.1101/ gad.1902110, PMID: 20439430

Lee SH, Eldi P, Cho SY, Rangasamy D. 2009. Control of chicken CR1 retrotransposons is independent of Dicer-mediated RNA interference pathway. *BMC Biology* **7**:53. doi: 10.1186/1741-7007-7-53, PMID: 19691826

Levin I, Santangelo L, Cheng H, Crittenden LB, Dodgson JB. 1994. An autosomal genetic linkage map of the chicken. *Journal of Heredity* **85**:79–85. doi: 10.1093/oxfordjournals.jhered.a111427, PMID: 7910177

Lewinski MK, Bushman FD. 2005. Retroviral DNA integration–mechanism and consequences. *Advances in Genetics* **55**:147–181. doi: 10.1016/S0065-2660(05)55005-3, PMID: 16291214

Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, Kittler EL, Zapp ML, Klattenhoff C, Schulz N, Theurkauf WE, Weng Z, Zamore PD. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**:509–521. doi: 10.1016/j.cell.2009.04.027, PMID: 19395009

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. doi: 10.1093/bioinformatics/btp324, PMID: 19451168

Li XZ, Roy CK, Dong X, Bolcun-Filas E, Wang J, Han BW, Xu J, Moore MJ, Schimenti JC, Weng Z, Zamore PD. 2013. An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Molecular Cell* **50**:67–81. doi: 10.1016/j.molcel.2013.02.016, PMID: 23523368

Lim SL, Qu ZP, Kortschak RD, Lawrence DM, Geoghegan J, Hempfling AL, Bergmann M, Goodnow CC, Ormandy CJ, Wong L, Mann J, Scott HS, Jamsai D, Adelson DL, O'Bryan MK, Zp Q, O'Bryan MK. 2015. HENMT1 and piRNA stability are required for adult male germ cell transposon repression and to define the spermatogenic program in the mouse. *PLOS Genetics* **11**:e1005620. doi: 10.1371/journal.pgen.1005620, PMID: 26496356

Lin H, Spradling AC. 1997. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. *Development* **124**:2463–2476. PMID: 9199372

Lu J, Clark AG. 2010. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in Drosophila. *Genome Research* **20**:212–227. doi: 10.1101/gr.095406.109, PMID: 19948818

Malik HS, Henikoff S, Eickbush TH. 2000. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Research* **10**:1307–1318. doi: 10.1101/gr.145000, PMID: 10984449

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* **137**:522–535. doi: 10.1016/j.cell.2009.03.040, PMID: 19395010

McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226**:792–801. doi: 10.1126/science.15739260, PMID: 15739260

Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H. 2013. Birth and expression evolution of mammalian microRNA genes. *Genome Research* **23**:34–45. doi: 10.1101/gr.140269.112, PMID: 23034410

Moore DH, Long CA, Vaidya AB, Sheffield JB, Dion AS, Lasfargues EY. 1979. Mammary tumor viruses. *Advances in Cancer Research* **29**:347–418. doi: 10.1016/S0065-230X(08)60850-7, PMID: 89801

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**:2933–2935. doi: 10.1093/bioinformatics/btt509, PMID: 24008419

Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grützner F, Kaessmann H. 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**:635–640. doi: 10.1038/nature12943, PMID: 24463510

Oh D, Son B, Mun S, Oh MH, Oh S, Ha J, Yi J, Lee S, Han K. 2016. Whole genome Re-Sequencing of three domesticated chicken breeds. *Zoological Science* **33**:73–77. doi: 10.2108/zs150071, PMID: 26853871

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**:289–290. doi: 10.1093/bioinformatics/btg412, PMID: 14734327

Peirson SN, Butler JN, Foster RG. 2003. Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Research* **31**:e73–73. doi: 10.1093/nar/gng073, PMID: 12853650

Pelechano V, Wei W, Steinmetz LM. 2015. Widespread Co-translational RNA decay reveals ribosome dynamics. *Cell* **161**:1400–1412. doi: 10.1016/j.cell.2015.05.008, PMID: 26046441

Reuter M, Berninger P, Chuma S, Shah H, Hosokawa M, Funaya C, Antony C, Sachidanandam R, Pillai RS. 2011. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* **480**:264–267. doi: 10.1038/nature10672, PMID: 22121019

Reuter M, Chuma S, Tanaka T, Franz T, Stark A, Pillai RS. 2009. Loss of the Mili-interacting tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nature Structural & Molecular Biology* **16**:639–646. doi: 10.1038/nsmb.1615, PMID: 19465913

Ricci EP, Kucukural A, Cenik C, Mercier BC, Singh G, Heyer EE, Ashar-Patel A, Peng L, Moore MJ. 2014. Staufen1 senses overall transcript secondary structure to regulate translation. *Nature Structural & Molecular Biology* **21**:26–35. doi: 10.1038/nsmb.2739, PMID: 24336223

Roberts A, Pachter L. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods* **10**:71–73. doi: 10.1038/nmeth.2251, PMID: 23160280

Robinson HL, Astrin SM, Senior AM, Salazar FH. 1981. Host susceptibility to endogenous viruses: defective, glycoprotein-expressing proviruses interfere with infections. *Journal of Virology* **40**:745–751. PMID: 6275116

Rosenkranz D, Han CT, Roovers EF, Zischler H, Ketting RF. 2015. Piwi proteins and piRNAs in mammalian oocytes and early embryos: from sample to sequence. *Genomics Data* **5**:309–313. doi: 10.1016/j.gdata.2015.06.026, PMID: 26484274

Rutherford K, Meehan CJ, Langille MG, Tyack SG, McKay JC, McLean NL, Benkel K, Beiko RG, Benkel B. 2016. Discovery of an expanded set of avian leukosis subroup E proviruses in chickens using Vermillion, a novel sequence capture and analysis pipeline. *Poultry Science* **95**:2250–2258. doi: 10.3382/ps/pew194, PMID: 27354549

Sacco MA, Flannery DM, Howes K, Venugopal K. 2000. Avian endogenous retrovirus EAV-HP shares regions of identity with avian leukosis virus subgroup J and the avian retrotransposon ART-CH. *Journal of Virology* **74**: 1296–1306. doi: 10.1128/JVI.74.3.1296-1306.2000, PMID: 10627540

Saxe JP, Chen M, Zhao H, Lin H. 2013. Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. *The EMBO Journal* **32**:1869–1885. doi: 10.1038/emboj.2013.121, PMID: 23714778

Schlötterer C. 2015. Genes from scratch–the evolutionary fate of de novo genes. *Trends in Genetics* **31**:215–219. doi: 10.1016/j.tig.2015.02.007, PMID: 25773713

Shoji M, Tanaka T, Hosokawa M, Reuter M, Stark A, Kato Y, Kondoh G, Okawa K, Chujo T, Suzuki T, Hata K, Martin SL, Noce T, Kuramochi-Miyagawa S, Nakano T, Sasaki H, Pillai RS, Nakatsuji N, Chuma S. 2009. The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Developmental Cell* **17**:775–787. doi: 10.1016/j.devcel.2009.10.012, PMID: 20059948

Smit AFA, Hubley R, Green P. 2016. 2015 RepeatMasker Open-4.0.

Smith EJ, Fadly AM, Crittenden LB. 1990a. Interactions between endogenous virus loci ev6 and ev21. 1. immune response to exogenous avian leukosis virus infection. *Poultry Science* **69**:1244–1250. doi: 10.3382/ps.0691244, PMID: 2172951

Smith EJ, Fadly AM, Crittenden LB. 1990b. Interactions between endogenous virus loci ev6 and ev21. 2. congenital transmission of EV21 viral product to female progency from slow-feathering dams. *Poultry Science* **69**:1251–1256. doi: 10.3382/ps.0691251, PMID: 2172952

Smith EJ, Fadly AM. 1988. Influence of congenital transmission of endogenous virus-21 on the immune response to avian leukosis virus infection and the incidence of tumors in chickens. *Poultry Science* **67**:1674–1679. doi: 10.3382/ps.0671674, PMID: 2853868

Smith LM, Toye AA, Howes K, Bumstead N, Payne LN, Venugopal K. 1999. Novel endogenous retroviral sequences in the chicken genome closely related to HPRS-103 (subgroup J) avian leukosis virus. *Journal of General Virology* **80 (Pt 1**:261–268. doi: 10.1099/0022-1317-80-1-261, PMID: 9934710

Soper SF, van der Heijden GW, Hardiman TC, Goodheart M, Martin SL, de Boer P, Bortvin A. 2008. Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Developmental Cell* **15**:285–297. doi: 10.1016/j.devcel.2008.05.015, PMID: 18694567

Stehelin D, Varmus HE, Bishop JM, Vogt PK. 1976. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **260**:170–173. doi: 10.1038/260170a0, PMID: 176594

Steitz JA. 1969. Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature* **224**:957–964. doi: 10.1038/224957a0, PMID: 5360547

Stoye JP, Coffin JM. 1987. The four classes of endogenous murine leukemia virus: structural relationships and potential for recombination. *Journal of Virology* **61**:2659–2669. PMID: 3039159

Sumiyoshi T, Sato K, Yamamoto H, Iwasaki YW, Siomi H, Siomi MC. 2016. Loss of l(3)mbt leads to acquisition of the ping-pong cycle in Drosophila ovarian somatic cells. *Genes & Development* **30**:1617–1622. doi: 10.1101/gad.283929.116, PMID: 27474440

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, Hannon GJ. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**:534–538. doi: 10.1038/nature06904, PMID: 18404147

Tanaka T, Hosokawa M, Vagin VV, Reuter M, Hayashi E, Mochizuki AL, Kitamura K, Yamanaka H, Kondoh G, Okawa K, Kuramochi-Miyagawa S, Nakano T, Sachidanandam R, Hannon GJ, Pillai RS, Nakatsuji N, Chuma S. 2011. Tudor domain containing 7 (Tdrd7) is essential for dynamic ribonucleoprotein (RNP) remodeling of chromatoid bodies during spermatogenesis. *PNAS* **108**:10579–10584. doi: 10.1073/pnas.1015447108, PMID: 21670278

Tarlinton R, Meers J, Young P. 2008. Biology and evolution of the endogenous koala retrovirus. *Cellular and Molecular Life Sciences* **65**:3413–3421. doi: 10.1007/s00018-008-8499-y, PMID: 18818870

Team RC. 2014. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Temin HM. 1964. Nature of the provirus of rous sarcoma. *National Cancer Institute Monograph* **17**:557–570.

Tereba A. 1981. 5'-terminal deletions are a common feature of endogenous retrovirus loci located on chromosome 1 of white leghorn chickens. *Journal of Virology* **40**:920–926. PMID: 7321106

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**:4673–4680. doi: 10.1093/nar/22.22.4673, PMID: 7984417

Thomson T, Lin H. 2009. The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. *Annual Review of Cell and Developmental Biology* **25**:355–376. doi: 10.1146/annurev.cellbio.24.110707.175327, PMID: 19575643

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**:1105–1111. doi: 10.1093/bioinformatics/btp120, PMID: 19289445

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nature Protocols* **7**:562–578. doi: 10.1038/nprot.2012.016, PMID: 22383036

Vandergon TL, Reitman M. 1994. Evolution of chicken repeat 1 (CR1) elements: evidence for ancient subfamilies and multiple progenitors. *Molecular Biology and Evolution* **11**:886–898. PMID: 7815928

Varmus HE, Weiss RA, Friis RR, Levinson W, Bishop JM. 1972. Detection of avian tumor virus-specific nucleotide sequences in avian cell DNAs (reassociation kinetics-RNA tumor viruses-gas antigen-Rous sarcoma virus, chick cells). *PNAS* **69**:20–24. doi: 10.1073/pnas.69.1.20, PMID: 4333039

Wang J, Saxe JP, Tanaka T, Chuma S, Lin H. 2009. Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis. *Current Biology* **19**:640–644. doi: 10.1016/j.cub.2009.02.061, PMID: 19345100

Wang Z, Qu L, Yao J, Yang X, Li G, Zhang Y, Li J, Wang X, Bai J, Xu G, Deng X, Yang N, Wu C. 2013. An EAV-HP insertion in 5' Flanking region of SLCO1B3 causes blue eggshell in the chicken. *PLoS Genetics* **9**:e1003183. doi: 10.1371/journal.pgen.1003183, PMID: 23359636

Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**:539–543. doi: 10.1038/nature06908, PMID: 18404146

Weiss RA, Biggs PM. 1972. Leukosis and Marek's disease viruses of feral red jungle flow and domestic fowl in Malaya. *JNCI: Journal of the National Cancer Institute* **49**:1713–1725. doi: 10.1093/jnci/49.6.1713, PMID: 411 9166

Weiss RA. 1969. The host range of BRYAN strain rous sarcoma virus synthesized in the absence of helper virus. *Journal of General Virology* **5**:511–528. doi: 10.1099/0022-1317-5-4-511

Weiss RA. 2006. The discovery of endogenous retroviruses. *Retrovirology* **3**:67. doi: 10.1186/1742-4690-3-67, PMID: 17018135

West B, Zhou B-X. 1988. Did chickens go north? new evidence for domestication. *Journal of Archaeological Science* **15**:515–533. doi: 10.1016/0305-4403(88)90080-5

Wichert S, Fokianos K, Strimmer K. 2004. Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics* **20**:5–20. doi: 10.1093/bioinformatics/btg364, PMID: 14693803

Wicker T, Robertson JS, Schulze SR, Feltus FA, Magrini V, Morrison JA, Mardis ER, Wilson RK, Peterson DG, Paterson AH, Ivarie R. 2005. The repetitive landscape of the chicken genome. *Genome Research* **15**:126–136. doi: 10.1101/gr.2438005, PMID: 15256510

Wilson JE, Connell JE, Macdonald PM. 1996. *Aubergine* enhances Oskar translation in the *Drosophila* ovary. *Development* **122**:1631–1639. PMID: 8625849

Yabuta Y, Ohta H, Abe T, Kurimoto K, Chuma S, Saitou M. 2011. TDRD5 is required for retrotransposon silencing, chromatoid body assembly, and spermiogenesis in mice. *The Journal of Cell Biology* **192**:781–795. doi: 10.1083/jcb.201009043, PMID: 21383078

Zhang B, Mao YS, Diermeier S, Novikova IV, Nawrocki EP, Jones TA, Lazar Z, Tung C, Luo W, Eddy S, Sanbonmatsu KY, Spector DL. 2017. Identification and characterization of a class of MALAT1-like genomic loci. *Cell Reports*. In press.

Zhang Z, Xu J, Koppetsch BS, Wang J, Tipping C, Ma S, Weng Z, Theurkauf WE, Zamore PD. 2011. Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and tudor domains. *Molecular Cell* **44**:572–584. doi: 10.1016/j.molcel.2011.10.011, PMID: 22099305

Zheng K, Xiol J, Reuter M, Eckardt S, Leu NA, McLaughlin KJ, Stark A, Sachidanandam R, Pillai RS, Wang PJ. 2010. Mouse MOV10L1 associates with piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *PNAS* **107**:11841–11846. doi: 10.1073/pnas.1003953107, PMID: 20534472

Zhuang J, Wang J, Theurkauf W, Weng Z. 2014. TEMP: a computational method for analyzing transposable element polymorphism in populations. *Nucleic Acids Research* **42**:6826–6838. doi: 10.1093/nar/gku323, PMID: 24753423