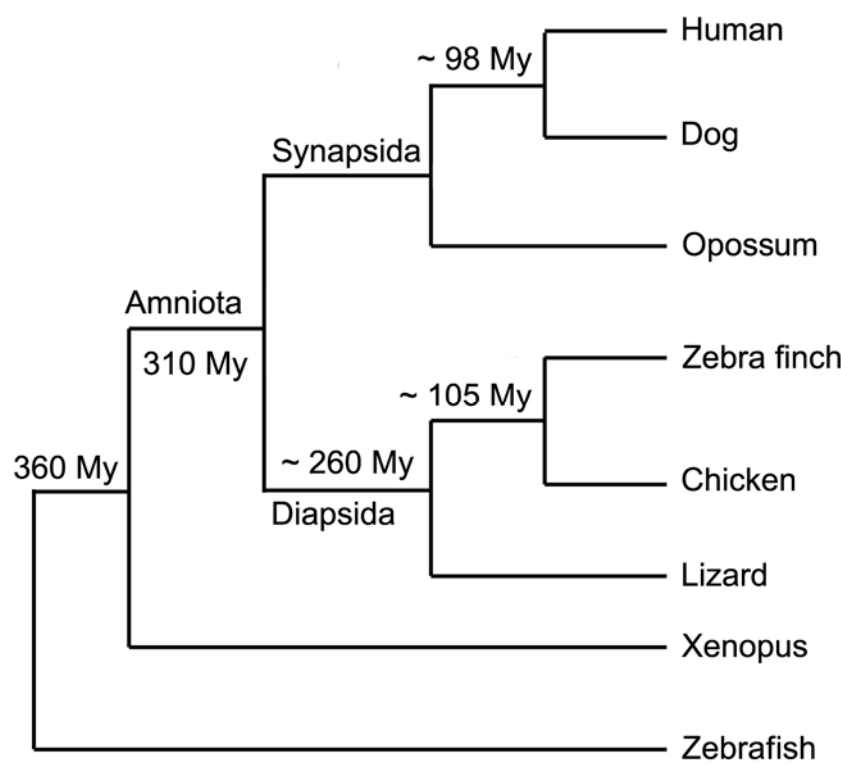


Supplemental Figure S1



Supplemental Figure S1. Summary of phylogenetic relationships of vertebrate species addressed in this study. The tree structure and estimated divergence times are used with modifications from Hedges et al. (Hedges et al. 2006) and Kumar and Hedges (Kumar and Hedges 1998)

Supplemental Figures S2a-S2t

Fig S2a. *Annexin A6* gene (*ANXA6*)

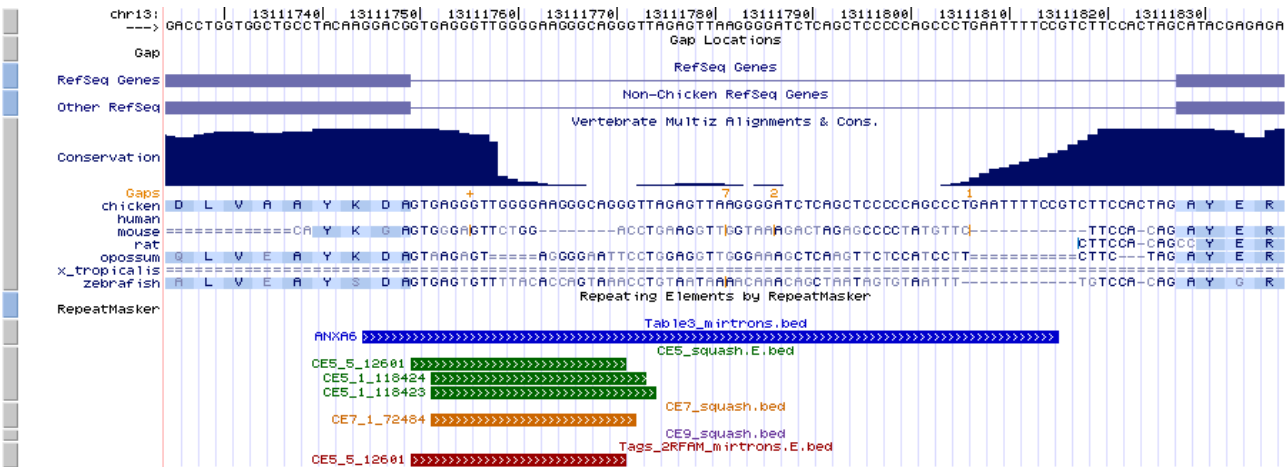


Fig S2b. *cyclin A2* (*CCNA2*)

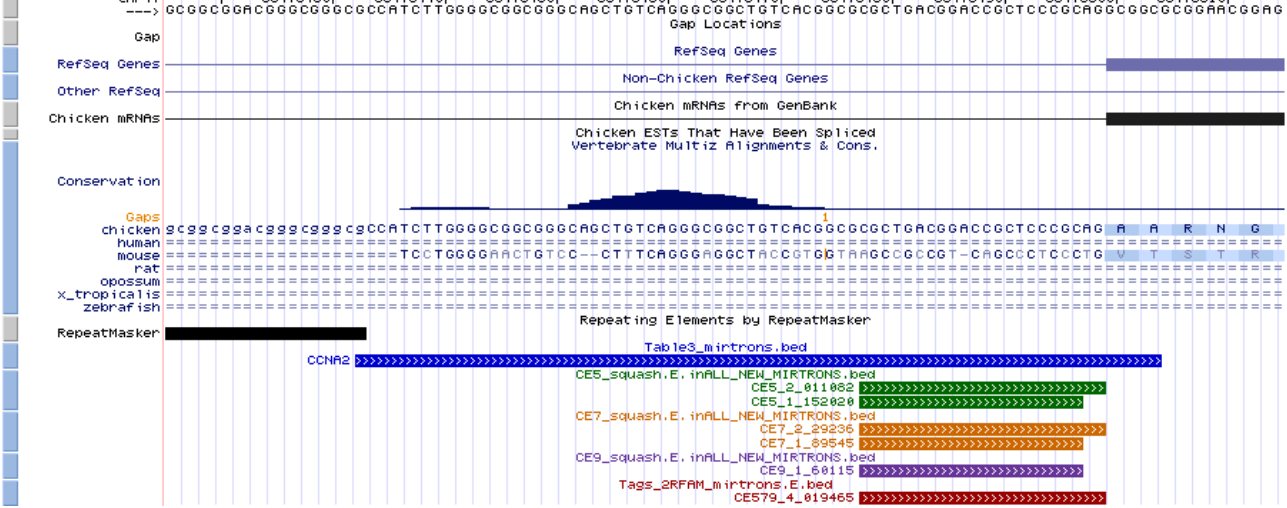


Fig S2c. *Williams Beuren syndrome chromosome region 27* (*WBSCR27*)

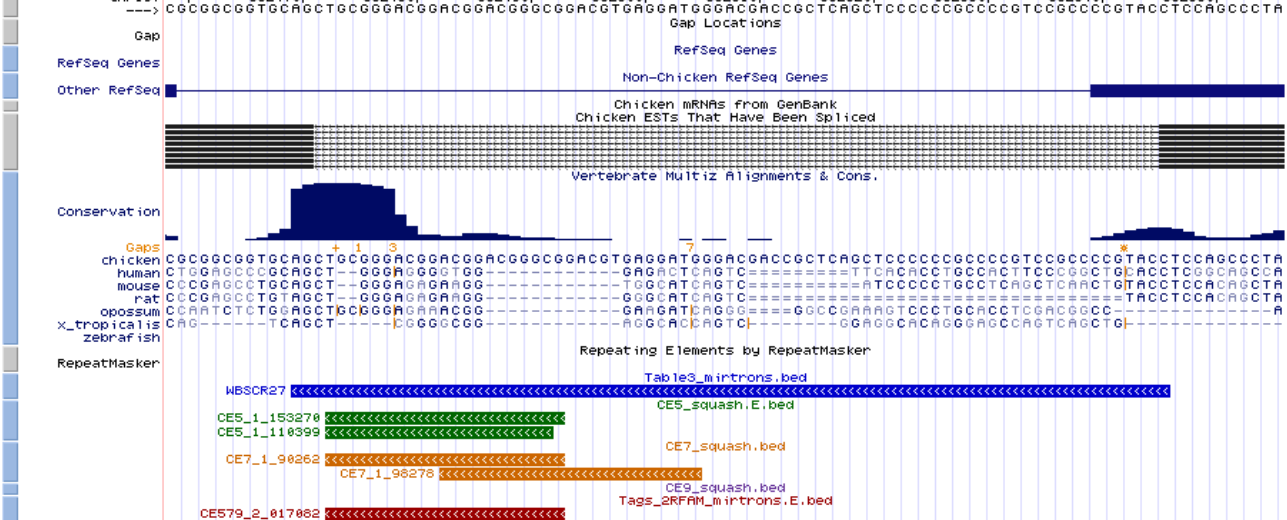


Fig S2g. epidermal growth factor (EGF) receptor family gene (ERBB2)

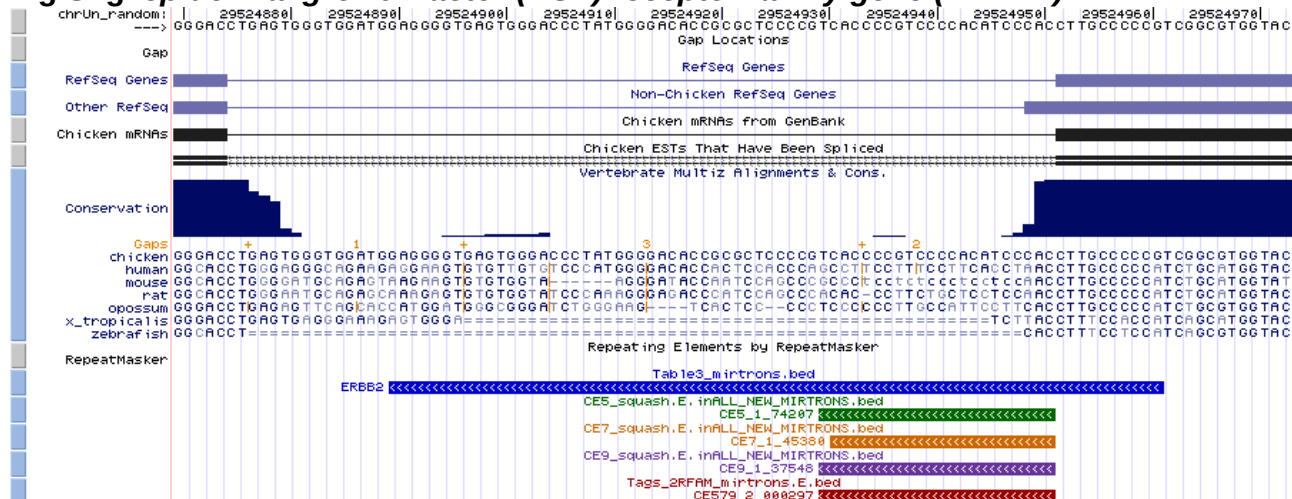


Fig S2h. trinucleotide repeat containing 5 (TNRC5)

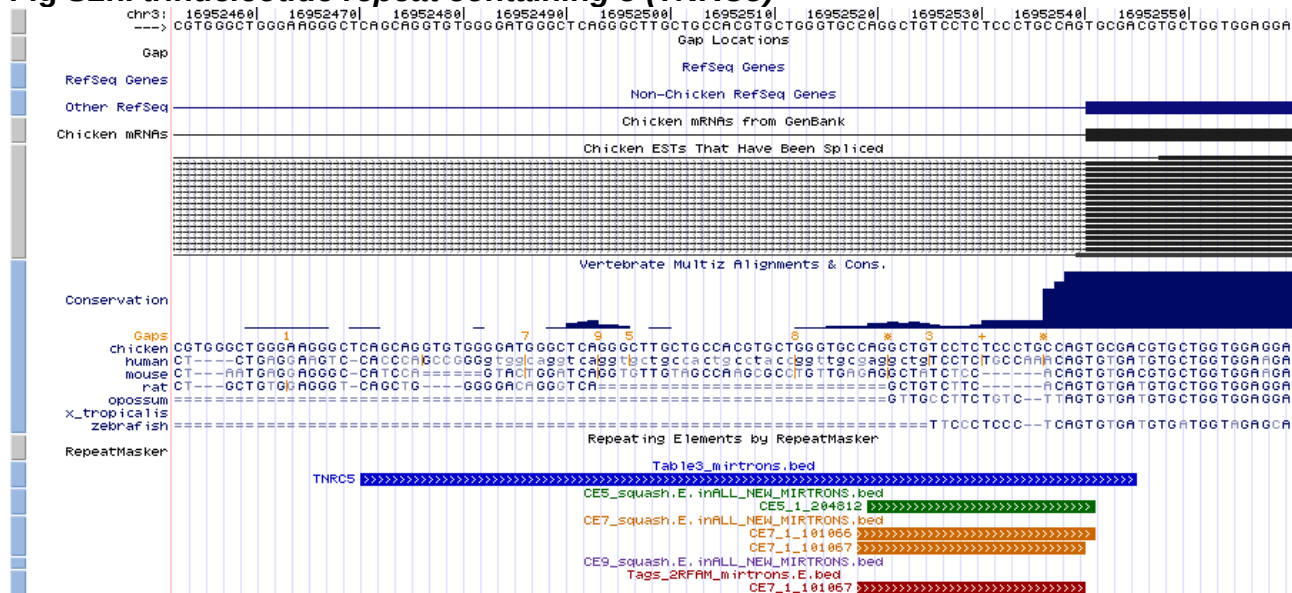


Fig S2i. ribosomal RNA processing 12 homolog gene (RRP12)

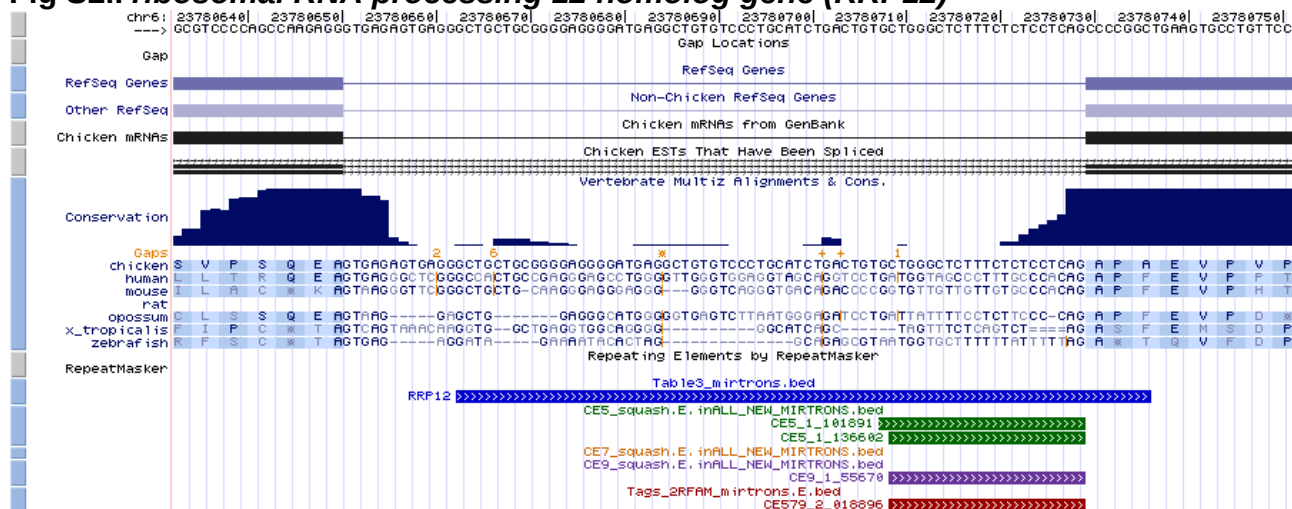


Fig S2j. nucleolar protein 5A gene (NOL5A)

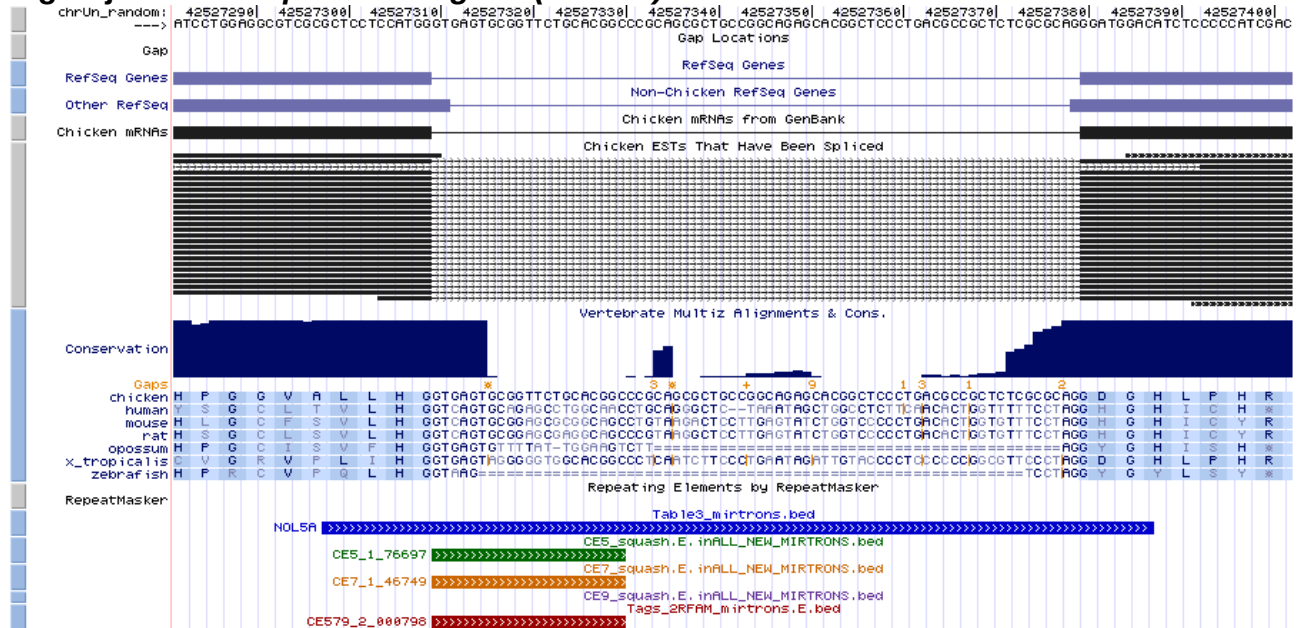


Fig S2k. Furin gene

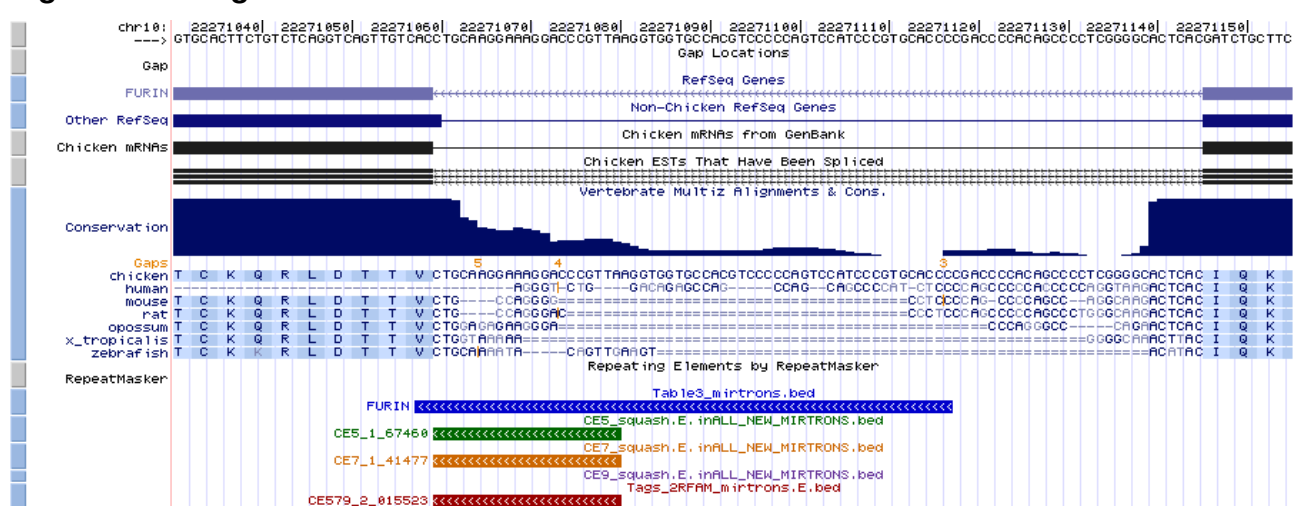


Fig S2l. p53-associated parkin-like cytoplasmic protein gene (PARC)

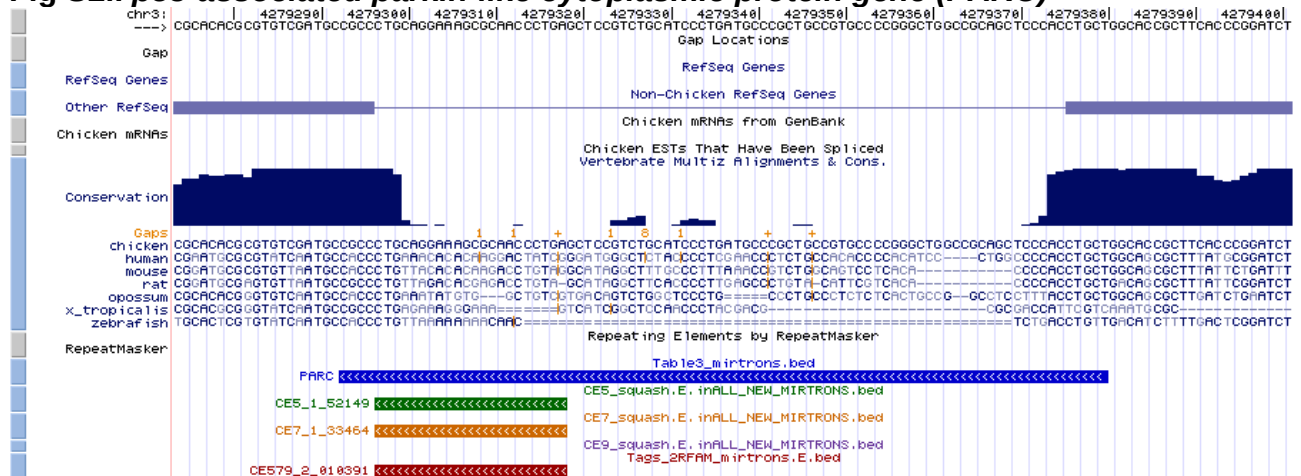


Fig S2m. *wingless-type MMTV integration site family gene (WNT3)*

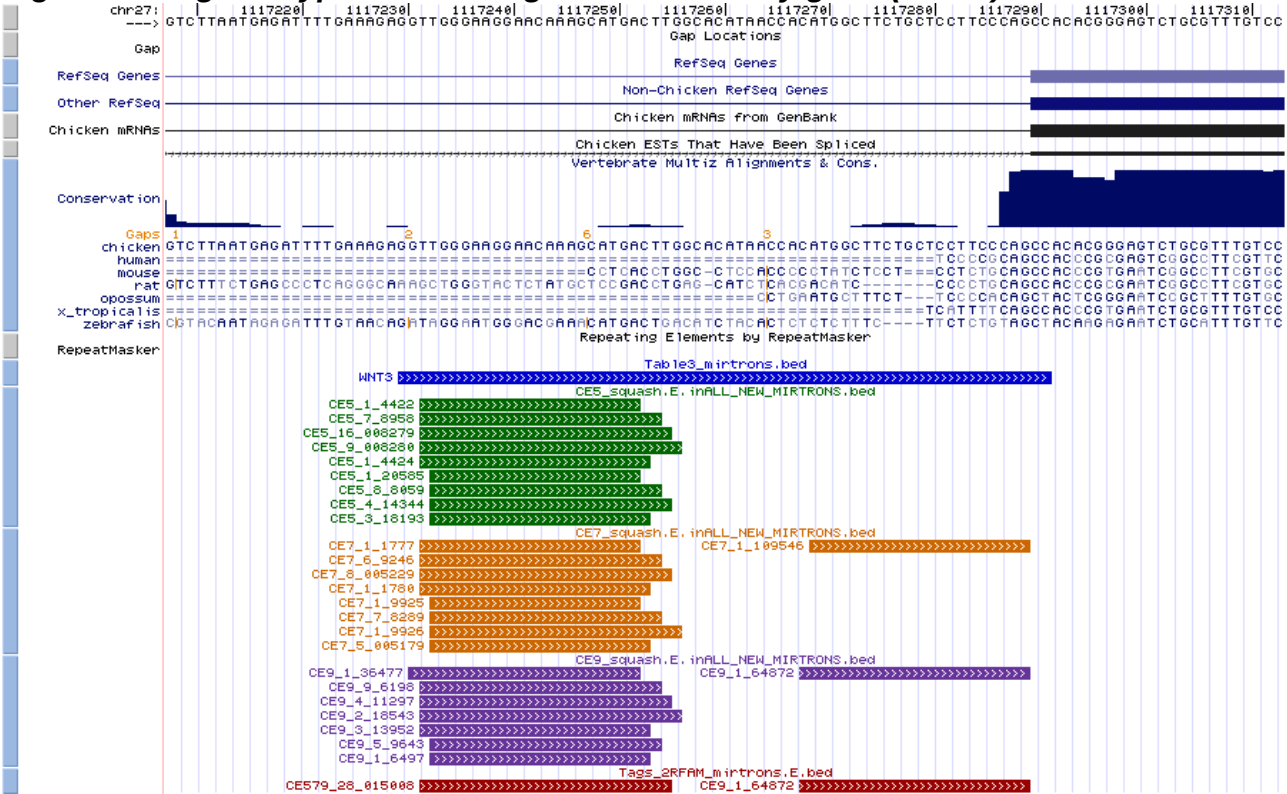


Fig S2n. *ADAM metalloproteinase gene (ADAMTS10)*

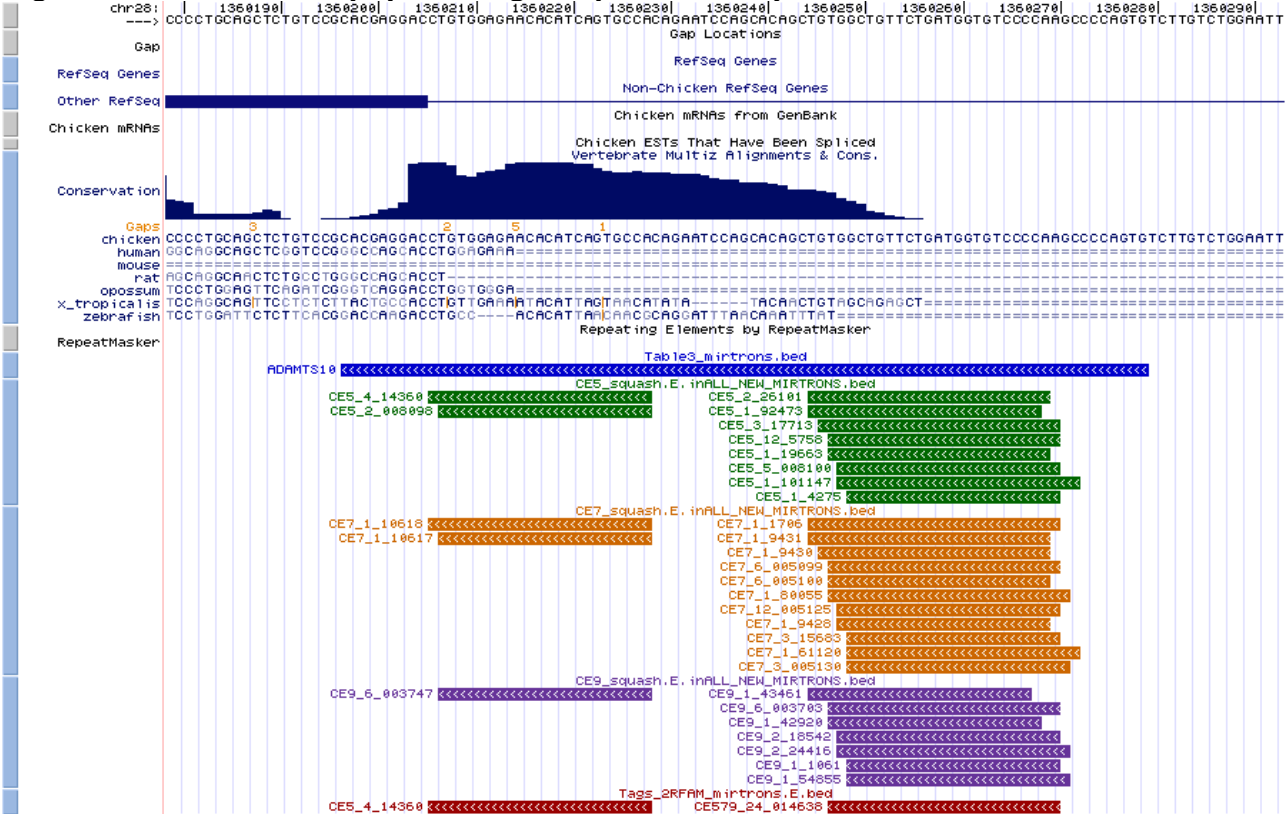


Fig S2o. transmembrane protein 63B (Tmem63B)

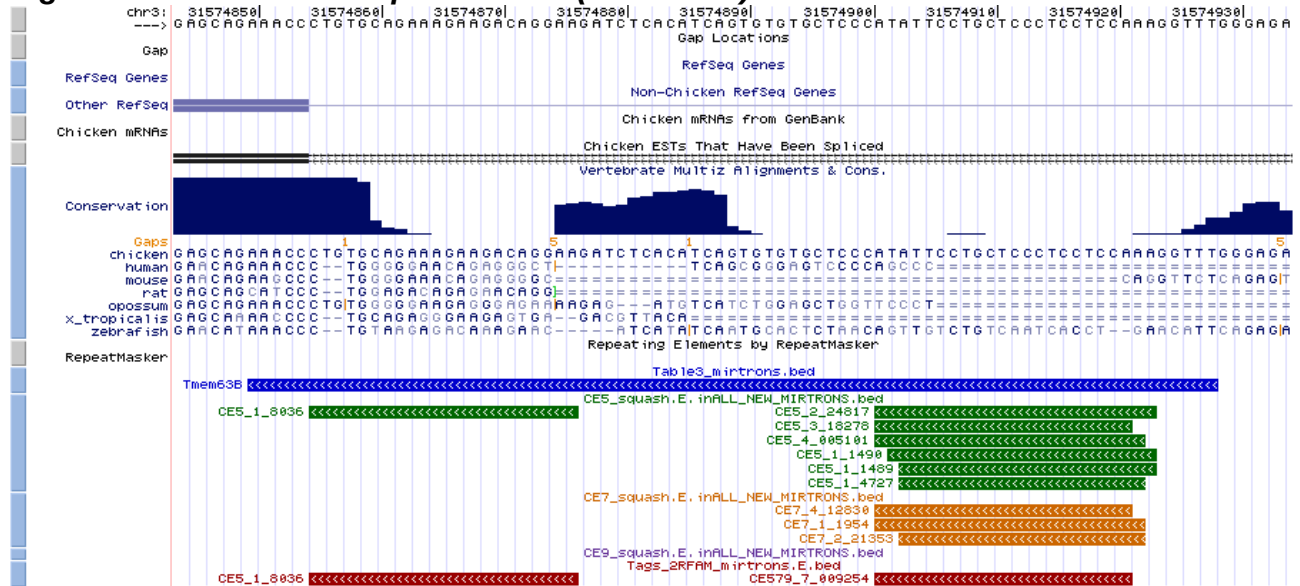


Fig S2p. sex comb on midleg homolog 1 gene (Scmh1)

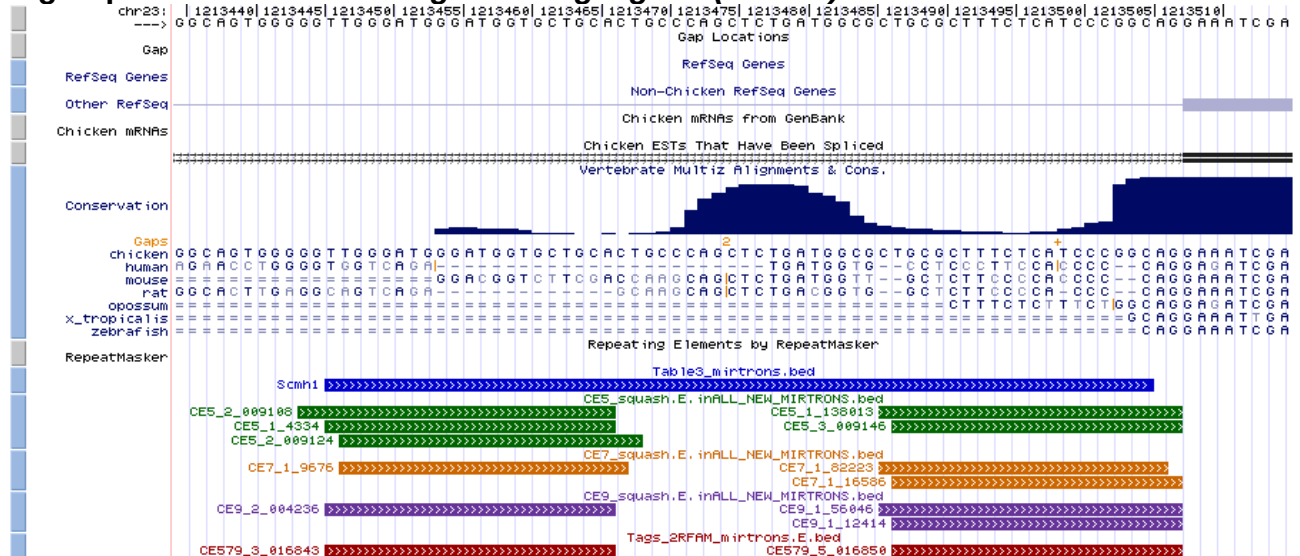


Fig S2q. cofactor required for SP1 activation (CRSP3)

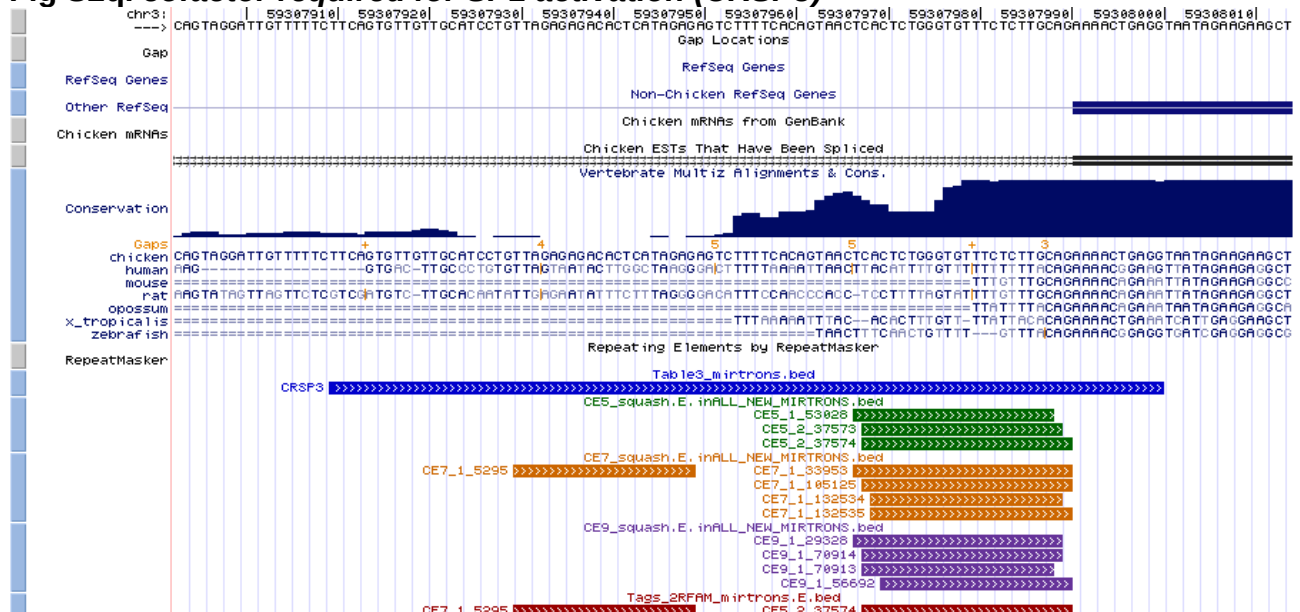
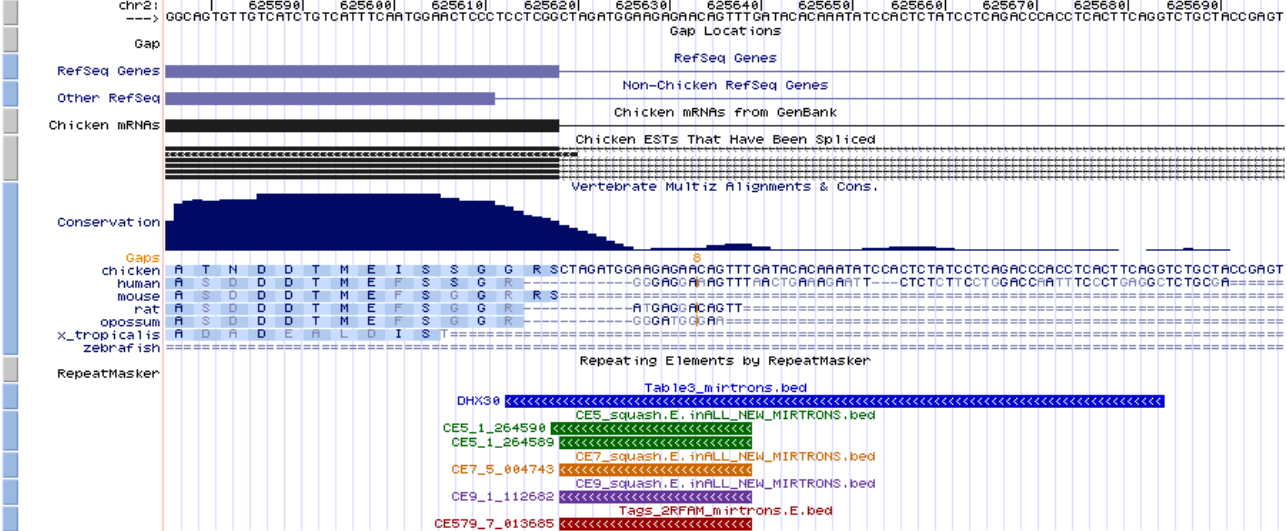
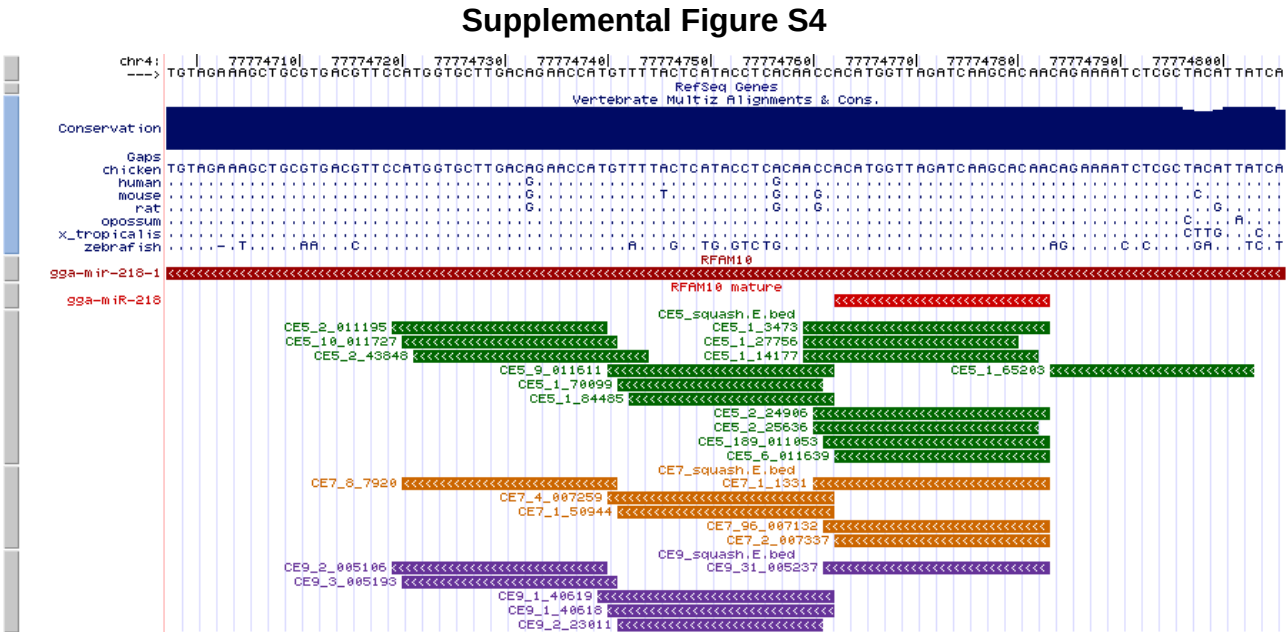
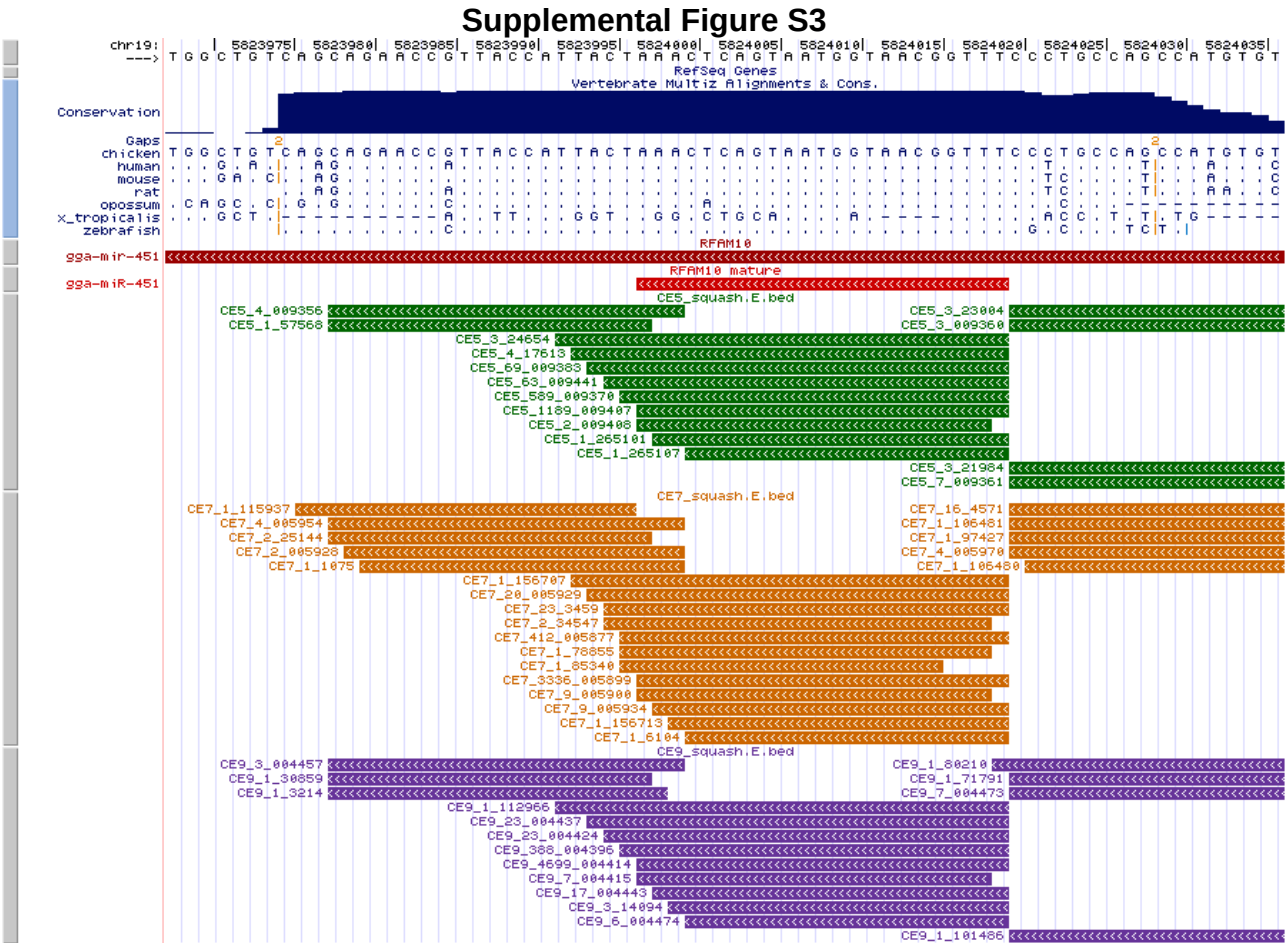


Fig S2r. DEAH-box helicase gene (DHX30)



Supplemental Figures S2a-S2r. Mirtron candidates listed in the Table 3. The figures shows UCSC genome browser screens displaying relative positions of RefSeq genes (top), predicted miRNA hairpin-like precursors (blue), and sequence tags originating from this loci in three small RNA libraries: CE5 - green, CE7 - dark orange, CE9 – magenta, and the consensus sequence for the miRBase submission (deep red).



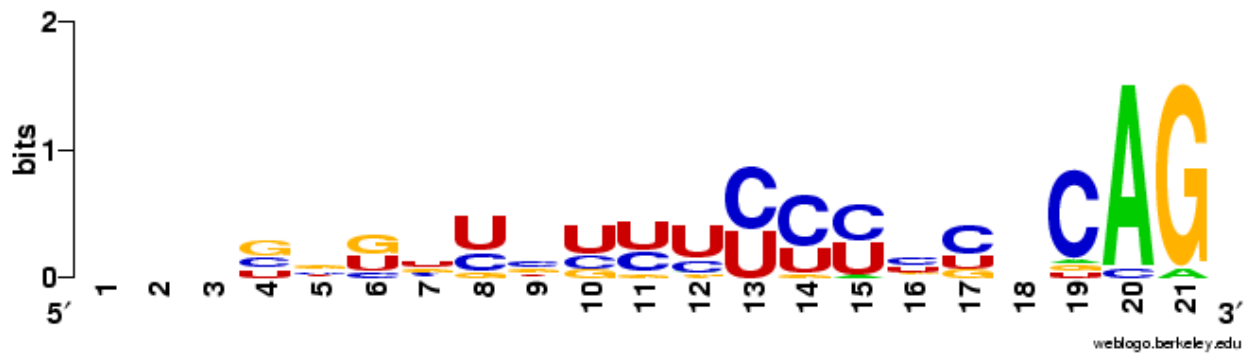
Supplemental Figures S3, S4. Atypical pattern of miRNA processing. The figure shows UCSC genome browser screens displaying gga-miR-451 and gga-mir-218-1 (red), their hairpin precursors (deep red), and sequence tags originating from these loci in three small RNA libraries: CE5 - green, CE7 - dark orange, and CE9 - magenta. Arrow directions indicate positive or negative strand of genomic DNA. Identical bases in the multiple sequence alignment are shown as dots.

Supplemental Figure S5

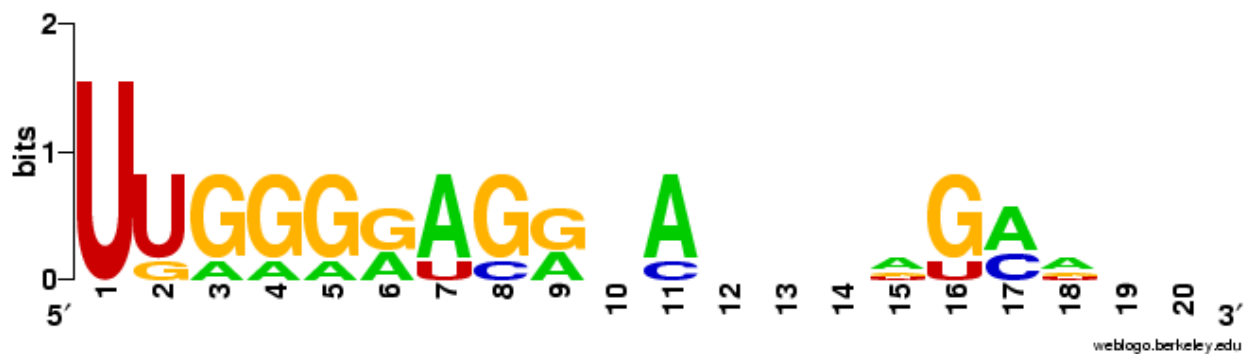
A.



B.



C.



Supplemental Figure S5. Sequence logos of the aligned ends of mirton-derived sequence tags. The logos were generated using WebLogo 2.8 (<http://weblogo.berkeley.edu/>). (A) The logo was generated using 4 representative sequence tags anchored to the donor (5') splice site. (B) The logo was generated using 16 representative sequence tags anchored to the acceptor (3') splice site. (C) The logo was generated using 5 representative sequence tags from 'hanging' 5' arm of the pre-miRNA not anchored to splice site junction.

Supplemental Table S2

RNA class	CE5		CE7		CE9	
	# reads	% of total	# reads	% of total	# reads	% of total
total	3257227		3251820		3187949	
known miRNAs	2069643	63.5	2036387	62.6	2152321	67.5
repeats	13998	0.7	7773	0.2	8945	0.3
rRNA	89204	2.7	35237	1.1	31464	1.0
tRNA	17630	0.5	21660	0.7	23239	0.7
snRNA	914	0.03	1027	0.03	391	0.01
snoRNA	562	0.02	333	0.01	241	0.01
coding exons	2253	0.07	775	0.02	569	0.02
other	1063023	32.6	1148628	35.3	970779	30.5

Supplemental Table S2. Distribution of sequence reads originating from known RNA classes in different libraries. Note “other” includes all unclassified sequences including new miRNAs.

Supplemental Table S3

RNA class	CE5		CE7		CE9	
	# tags	% of total	# tags	% of total	# tags	% of total
total	59049		23067		17187	
known miRNAs	2290	3.9	2151	9.3	2085	12.1
repeats	4137	7.0	1358	5.9	1126	6.6
rRNA	5147	8.7	3205	13.9	2688	15.6
tRNA	1666	2.8	1371	5.9	1359	7.9
snRNA	531	0.9	436	1.9	273	1.6
snoRNA	333	0.6	173	0.7	152	0.9
coding exons	1954	3.3	611	2.6	491	2.9
other	42991	72.8	13762	59.7	9013	52.4

Supplemental Table S3. Distribution of unique sequence tags originating from known RNA classes in different libraries. Note “other” includes all unclassified sequences including new miRNAs.

Supplemental Table S4			
	CE5	CE7	CE9
# total uniq tags	59049	23067	17187
# tags with multiple reads	16715	10332	7344
% of total tags	28.3	44.8	42.7
# tags with single reads	42334	12735	9843
% of total tags	71.7	55.2	57.3

Supplemental Table S4. Singleton vs. multiple sequence tags.

Supplemental Table S5

Name	Zebra Finch	Lizard	Xenopus	Danio	Opossum	Dog	Human
gga-mir-124a-2	Yes	Yes	Yes	Yes	Yes	Yes	Yes
gga-mir-101a-2	Yes	Yes	Yes	Yes	Yes	Yes	Yes
gga-mir-301a-2	Yes	Yes	Yes	Yes	Yes	Yes	Yes
gga-mir-10a	Yes	Yes	Yes	Yes	Yes	Yes	Yes
gga-mir-15c	Yes	nd	Yes	Yes	nd	nd	nd
gga-mir-458	nd	Yes	nd	Yes	nd	nd	nd
gga-mir-130c-1	Yes	nd	Yes	Yes	nd	nd	nd
gga-mir-146b	Yes	nd	Yes	Yes	nd	nd	nd
gga-mir-193a	Yes	Yes	nd	Yes	nd	nd	Yes
gga-mir-730	Yes	Yes	nd	Yes	nd	nd	nd
gga-mir-454a	Yes	Yes	Yes	Yes	nd	nd	Yes
gga-mir-16c	Yes	Yes	Yes	nd	nd	nd	nd
gga-mir-551b	Yes	nd	nd	nd	nd	Yes	Yes
CE5_40_2300	Yes	Yes	Yes	nd	Yes	nd	nd
CE9_1_76734	Yes	Yes	nd	nd	Yes	nd	nd
CE5_1_188053	Yes	nd	nd	nd	Yes	Yes	nd
CE5_1_6191	Yes	nd	Yes	nd	nd	nd	nd
CE5_1_37323	Yes	Yes	nd	nd	nd	nd	nd
CE9_1_43609	Yes	Yes	Yes	nd	nd	nd	nd
CE5_290_511	Yes	nd	nd	nd	nd	nd	nd
CE5_248_576	Yes	nd	nd	nd	nd	nd	nd
CE5_170_752	Yes	nd	nd	nd	nd	nd	nd
CE5_88_1293	Yes	nd	nd	nd	nd	nd	nd
CE5_120_997	Yes	nd	nd	nd	nd	nd	nd
CE5_28_3012	Yes	nd	nd	nd	nd	nd	nd
CE5_18_4177	Yes	nd	nd	nd	nd	nd	nd
CE7_1_15272	Yes	nd	nd	nd	nd	nd	nd
CE9_1_444	Yes	nd	nd	nd	nd	nd	nd
CE5_1_187210	Yes	nd	nd	nd	nd	nd	nd
CE5_1_21731	Yes	nd	nd	nd	nd	nd	nd
CE5_70_1466	Yes	nd	nd	nd	nd	nd	nd
CE7_1_28368	Yes	nd	nd	nd	nd	nd	nd
CE7_1_32577	Yes	nd	nd	nd	nd	nd	nd
CE9_1_1680	Yes	nd	nd	nd	nd	nd	nd
CE9_1_54991	Yes	nd	nd	nd	nd	nd	nd
CE9_1_73248	Yes	nd	nd	nd	nd	nd	nd
CE9_1_79496	Yes	nd	nd	nd	nd	nd	nd
CE7_2_32677	Yes	nd	nd	nd	nd	nd	nd
CE5_1_226537	Yes	nd	nd	nd	nd	nd	nd
CE5_2_34022	Yes	nd	nd	nd	nd	nd	nd
CE5_3_17773	Yes	nd	nd	nd	nd	nd	nd

See page 15 for the Table Legend

Supplemental Table S6

Name / ID	# reads	Conservation	Genomic coordinates	Strand
gga-mir-124a	97979	Yes	chr20:8681786-8681862	+
gga-mir-10a	41052	Yes	chrUn_random:379303-379377	-
gga-mir-101a-2	32820	Yes	chr8:29051917-29051993	-
gga-mir-146b	11321	Yes	chr4:92169270-92169399	-
gga-mir-16c	8510	Yes	chr4:4048688-4048759	-
gga-mir-130c-1	4290	Yes	chr19:7145026-7145120	-
gga-mir-22a	3082	Yes	chr19:5352093-5352197	-
gga-mir-15c	2099	Yes	chr4:4049054-4049130	-
CE5_101_1119	1610	No	chrZ:44167488-44167594	-
CE5_290_511	1241	Yes	chr3:76659762-76659835	+
gga-mir-301a-2	863	Yes	chr19:7144738-7144828	-
gga-mir-454a-2	833	Yes	chr15:399849-399931	-
gga-mir-193a	654	Yes	chr18:6423769-6423846	+
CE5_248_576	618	Yes	chr3:78710220-78710299	+
gga-mir-551b	550	Yes	chr9:21966414-21966514	-
CE5_170_752	538	Yes	chr7:1330063-1330139	-
CE5_4_14525	525	No	chr10:14823389-14823454	-
CE5_1_6191	506	Yes	chr2:1721333-1721406	+
gga-mir-458	463	Yes	chr13:8034157-8034273	-
CE5_2_28639	442	No	chr7:10846-10929	+
CE9_1_79254	393	No	chr20:10846980-10847083	-
CE5_18_4177	336	Yes	chr1:130934899-130934979	+
CE5_88_1293	306	Yes	chr15:769595-769666	+
CE5_120_997	275	Yes	chrUn_random:9521374-9521457	-
CE9_1_33234	238	No	chr21:4121864-4121940	+
CE5_1_6611	179	No	chr1:194622058-194622155	-
CE5_12_5775	174	No	chr7:25293056-25293161	+
CE9_1_113116	138	No	chr2:69711199-69711298	+
CE5_28_3012	105	Yes	chrZ:34596485-34596561	+
CE5_3_18212	104	No	chrZ:16041926-16041998	-
CE9_1_1680	103	Yes	chr3:36895819-36895922	+
CE5_40_2300	102	Yes	chr3:99798397-99798474	+
CE5_70_1466	92	Yes	chr4:11023001-11023097	-
CE5_1_107270	74	No	chr11:20587340-20587444	-
CE5_2_26268	64	No	chr8:17124282-17124357	-
CE9_1_90871	62	No	chr13:12926220-12926297	-
CE5_1_24027	59	No	chr15:7810091-7810171	-
CE5_30_2849	57	No	chr2:3751022-3751096	-
CE5_2_31688	54	No	chr9:23692586-23692675	+
CE5_10_6713	51	No	chr12:13130850-13130948	+

See page 15 for the Table Legend

Supplemental Table S8

Gene name	RefSeq ID	# of reads	Position of reads	Genomic coordinates	Strand
Ccnl2	NM_001039577*	86	5p	chr21:2197500-2197561	+
AP1S2	NM_00100626	11	5p	chr1:125276272-125276345	+
Unknown	BU269957*	8	3p	chr3:43902379-43902460	+
ACTR1A	NM_001007948	7	5p	chr6:24600334-24600422	-
LARP4	NM_001012700	6	5p	chrE22C19W28_E50C23: 194939-195017	-
HSD11B1L	NM_001001201	5	5p	chr28_random:35314-35410	+
PER2	NM_204262	5	5p	chr9:6543163-6543226	-
NRXN1	NM_004801*	4	3p	chr3:6459627-6459716	+
plk3	BU105809*	4	5p	chr8:21500442-21500525	+
NFRKB	NM_001006317	3	5p	chr24:1568451-1568512	+
VAMP4	NM_003762	3	5p	chr8:4814316-4814369	-
HBEGF	NM_204849	3	5p	chr13:1419189-1419251	-
ACCN2	NM_001040467	2	5p	chrE22C19W28_E50C23: 284501-284558	-
GAPDH	NM_204305	2	3p	chr1:80092526-80092605	+
Hspca	CO635775*	2	5p	chr5:51983783-51983866	-
PIK3C2B	NM_002646*	2	5p	chr26:1586290-1586358	-
LHX6	DQ082894*	2	5p	chr17:9433952-9434042	-
SDK2	NM_204538	2	5p	chr18:9051012-9051095	-
WBSCR22	NM_001039332	2	3p	chr19:73930-73992	+
VAMP4	NM_003762	2	5p	chr8:4815052-4815113	-
RPUSD4	NM_032795*	2	5p	chr24:315408-315471	-

Supplemental Table S8. List of chicken mirtron candidates.

Supplemental Table S5. Summary of the evolutionary conservation of new miRNA candidates. **Yes** – refers to either known orthologous miRNA in other species or highly similar sequence detected by BLAT / BLAST. **nd** – ‘not determined’, neither known orthologous miRNA nor similar candidate sequence were detected.

Supplemental Table S6. The most abundant new chicken miRNAs. Number of reads refers to the counts of most frequently sequenced reads present in all three embryonic libraries. Note that the total number of reads, including less common reads is higher.

REFERENCES

- Hedges, S.B., J. Dudley, and S. Kumar. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**: 2971-2972.
- Kumar, S. and S.B. Hedges. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**: 917-920.