Supplemental Information, Harding *et al*.

Supplemental Methods

Small RNA-seq library preparation and sequencing

Total RNA was obtained by Trizol extraction from *X. tropicalis* embryos and treated with DNAse I (Worthington Biochemical Corporation). Small RNA libraries were prepared as described (Hafner et al. 2008). Briefly, small RNAs of 18–30 nt were size-selected by gel purification and RNA adapters ligated on using T4 RNA ligase (Promega). cDNA was synthesized and used as a template to amplify the small RNA library by PCR. The PCR library was adjusted to 10 nM concentration and single-end sequenced using the small RNA sequencing primer (Illumina) on the Illumina platform on a Genome Analyzer IIx at the Cambridge Research Institute. The small RNA libraries were sequenced in a technical repeat at the London Research Institute. In both sequencing runs, 36 bp of sequence was obtained.

mRNA-seq library preparation and sequencing

Total RNA was obtained by Trizol extraction from *X. tropicalis* embryos and the mRNA-seq sample prep kit (Illumina) was used to prepare the mRNA libraries. Briefly, polyadenylated RNA was selected, fragmented and cDNA was synthesized using Superscript II reverse transcriptase (Invitrogen). Double-stranded cDNA was then synthesized using T4 DNA polymerase and used for Illumina sample preparation. 72 bp of single-end mRNA sequence was obtained.

Novel miRNA prediction pipeline

Unannotated small RNA tags (Table S2) with read numbers > 10 and length of 22–23 nt were aligned to the genome with up to two mismatches. Overlapping small RNA reads form tag-contigs. Bases in the tag-contig were assigned a weighted depth and tag-contigs with a maximum weighted depth greater than one and contiguous length

greater than 22 nt were retained for the next step of novel miRNA prediction. The genomic sequence of the tag-contig was extended by 10 nucleotides 5' and 60 nucleotides 3' or vice versa to form a candidate miRNA precursor sequence whose secondary structure was predicted using Mfold (Ambros et al. 2003; Zuker 2003). Putative miRNA hairpins were manually screened against Ambros criteria (Ambros et al. 2003) to select candidate miRNAs for experimental detection.

Cell culture

The mammalian cell lines were all cultured in DMEM supplemented with 10% FCS except the HUVECs which were cultured in EGM-2 Bullet-Kit media with supplements (CC-3162, Lonza) in collagen pre-coated flasks (BD Biocoat) at 5% CO2. NCI-H460 cells were cultured in RPMI/10% FCS. Total RNA was extracted as described for *X. tropicalis* above.

Primers

small RNA qPCR primers	
poly(T) adapter for cDNA synthesis	5'GCGAGCACAGAATTAATACGACTCACT ATAGGTTTTTTTTTTTTVN
miRNA qPCR reverse primer	5' GCGAGCACAGAATTAATACGACTCAC
Forward primers	
miR-F	5' GAGAAAGTGCTTCTCGTTCGGCTGA
xtr-miR-427	5' AAAGTGCTTTCTGTTTTGGGCG
xtr-miR-A	5' AGCAAATCTGTTGGTTTGTACAAAC
xtr-miR-C	5' TGCGCTCGGTGGATAGAAGACGTGA
xtr-miR-148a	5' TCAGTGCACTACAGAACTTTGT
xtr-miR-206	5' TGGAATGTAAGGAAGTGTGTGG
bta-miR-140	5' TACCACAGGGTAGAACCACGGA
gga-miR-30a-3p	5' CTTTCAGTCGGATGTTTGCAGC
Control primer pairs	
X. tropicalis odc	F: GGCCACACTGGCAACTCATGC R: CGGTGTGCGCTCAGTTCTGGT
X. tropicalis U6	F: ATGGCCCCTGCGCAAGGATG R: miRNA qPCR reverse primer
Human *gapdh*	F: CTTCAACAGCGACACCCACT R: GTGGTCCAGGGGTCTTACTC

Northern blot oligos	
miR-F	5'TCAGCCGAACGAGAAGCACTTTCTC

siteRNA cluster qRT-PCR oligos	
arhgef10l	F: AGTAGTAGTGGGGGGCGCAACCAAA R: CCCTTCCCTTGTGTCTCTTCCCTTGT
map7d2	F: GGTGCAACCAAACGATTGCTC R: CACTACATCTTTTCCCTTGTGTCTCT
gabbr2	F: AGGAGTCGGCCAGGAACACAT R: CCTTCTCTTAACCTTCAAAGCTCTC
polr3b	F: AACCAAACTATTGCTCTGTGAGG R: TCTCTTCCCTCGTGTCTCTGTAT
kiaa1468	F: CAAATGATTGCTCTGCGAGG R: GGAATGCCCTCCCTGATTTTAA
Common group 6	F: ACGAAGGAGTCGGCCAGGAA R: TTCCTCTGCTCCTCACTACATCT
cdk15	F: AGTAAGGGCAGGGGCACG R: TCCGTTCGTCTCCGCCTACT
Gabbr1	F: AGTACCAGAAGCCAGAGAGGATTA R: TCCGCCTACTACCTGCTCG
galk2	F: ATTGAATATGTGGGTAAGAGCTGAAG R: AGTCTGAAATTTCTCAGCTTCTCTGT
tdp1 cluster 1	F: TGGCAAGCAAAGGGGCACGA R: ACTCAGCCTGCCTGACGACGA
tdp1 cluster 2	F: TGGATAATCAGAATAGTCAGG R: TGGTGCTTTTCTGAGCCTTGTG

mRNA qRT-PCR oligos	
arhgef10l	F: GCTCCGGGCAGCACCAACTT R: CAGGACGTACCGAGGGCCGA
map7d2	F: GGCTCTGCCGAAGCGCTCAT R: ACCGGTGACACCTTGGGGCT
gabbr2	F: TGATGCCCGCTTTCCGCCTG R: TCAGCACCAGGGAGAGCGCA
polr3b	F: TGCCCGAGCAAGAGGTCCCA R: GGCCGCACTTCCCACAGACG
kiaa1468	F: TTGTGGGGCCCAGCACAAGC R: AAGCAAGCGAGGCAGCAGGG
cdk15	F: AAGGGCTCGGAGTAACAGCG R: TTCCAAAAGGCAGGCCCCTC
gabbr1	F: CTCAGCCCCATCTGAATGAT R: AGACGTGCCTGACACACAAG
galk2	F: TGCCTTGGTCTGCTGTGCTG R: ATGGATTGGTCCATGCCGCC
tdp1	F: CCCTGCTGGTGGCTCATTGCC R: ACCATGCCAGGTGCTGGGAA
xbra	F: CCTGTGGATGAGGTTCAAGG R: CACGCTCACCTTTAGAACTGG

ChIP qPCR oligos	
arhgef10l TSS	F: ACTACTGGCGGGGCACTACT R: CCTAATACATGGTAGCCAGC
arhgef10l small RNA cluster	F: AGTAGTAGTGGGGGGCGCAACCAAA R: CCCTTCCCTTGTGTCTCTTCCCTTGT
jak2 TSS	F: GTAATACCAAGTAATCCCCAT R: ATGACTGAAATCGATGGAACC
jak2 small RNA cluster	F: GTTGTTGCCGTCACCAAGGGG R: CCAATTCCCACTGAGGTGCAAGC
tdp1 TSS	F: GGTAGGACCTCAGCTTCCCAA R: ATGCTTTTCTGGCATCGGAGC
tdp1 small RNA cluster	F: TGGCAAGCAAAGGGGCACGA R: ACTCAGCCTGCCTGACGACGA
kiaa1468 TSS	F: TTGGGTGTTGTGGGGCCCAGCA R: TGAAGGTAGCGAGAAGCAAGC
kiaa1468 small RNA cluster	F: CAAATGATTGCTCTGCGAGG R: GGAATGCCCTCCCTGATTTTAA
derl2 TSS	F: CACCGGTGGAATCTGCATATA R: CCATTGGAGGCAGAGCGTACTG
derl2 small RNA cluster	F: GGGGCACAACCAAGCGTTT R: CAAAGAATATCTTACAAACTCCTTC
Gabbr1 TSS	F: GCTCAGGGACAGCCCAATGT R: ATAACACCATGGCGTGCCCA
Gabbr1 small RNA cluster	F: AGTACCAGAAGCCAGAGAGGATTA R: TCCGCCTACTACCTGCTCG

Small RNA library oligos	
5' RNA Adapter (26 nt)	5' GUUCAGAGUUCUACAGUCCGACGAUC
3' RNA Adapter (23 nt)	5' P-UCGUAUGCCGUCUUCUGCUUGUidT
RT primer	5' CAAGCAGAAGACGGCATACGA
PCR Primer 1	5' CAAGCAGAAGACGGCATACGA
PCR Primer 2	5' AATGATACGGCGACCACCGACAGGTTCAGA GTTCTACAGTCCGA
Small RNA Sequencing Primer	5' CGACAGGTTCAGAGTTCTACAGTCCGACGAC

Table S1. Small RNA-seq mapping statistics for the original sequencing run and a technical repeat.

small RNA library	Number of raw reads	% of raw reads aligning to genome with 0 mismatches	Number of small RNA tags
stage 8	15,824,233	57.7	672,032
stage 8 technical repeat	23,119,446	54.7	696,569
stage 10	15,338,864	59.3	387,173
stage 10 technical repeat	21,565,804	57.1	358,909
stage 18	14,739,870	54.7	566,646
stage 18 technical repeat	27,893,758	48.9	599,334
stage 10 animal	14,297,841	57.2	551,798
stage 10 animal technical repeat	20,704,206	56.0	528,987
stage 10 vegetal	15,171,802	48.8	718,228
stage 10 vegetal technical repeat	26,455,794	44.8	834,281

Table S2. Annotation of the *X. tropicalis* small RNAome in early development.

small RNA library	% reads annotated as miRNAs	% reads annotated as piRNAs*	% Rfam reads	% Fantom3 reads	% Hinv, Evofold, Antisense ncRNA pipeline and literature data set reads	% Unannotated genomic reads
stage 8	0.17	0.37	1.10	0.21	3.8×10^{-4}	98.1
stage 8 technical repeat	0.15	0.35	1.45	2.47	0.35	95.2
stage 10	0.32	0.31	1.19	0.15	5.0×10^{-4}	98.0
stage 10 technical repeat	0.27	0.33	1.79	1.27	0.21	96.1
stage 18	1.56	0.56	1.17	0.26	3.3×10^{-4}	96.5
stage 18 technical repeat	1.21	0.56	1.94	2.25	0.29	93.8
stage 10 animal	0.67	0.88	2.20	0.49	1.1×10^{-3}	95.8
stage 10 animal technical repeat	0.49	1.04	3.13	4.91	0.21	93.2
stage 10 vegetal	0.33	0.68	1.41	0.32	8.2×10^{-4}	97.3
stage 10 vegetal technical repeat	0.25	0.70	2.94	4.91	0.29	90.9

Summary of the percentage of trimmed reads that align with zero mismatches to miRNA sequences (mature miRNAs, miRNA-star and hairpin sequences from miRbase v14 and miRNAs from RNAdb 2.0 (http://research.imb.uq.edu.au/rnadb/rnadb2_archive.htm)), *piRNA sequences (from RNAdb 2.0), Rfam sequences (Rfam 9.1), Fantom3 sequences (mouse non-coding RNAs) and other non-coding RNA databases in the small RNA libraries from the original sequencing run and a technical repeat sequencing run. The percentage of unannotated reads that align to the genome is shown.

*These sequences were annotated as piRNAs, but the majority also match fragments of tRNAs and snoRNAs.

Table S3. The most frequently sequenced microRNAs in *X. tropicalis* early development.

small RNA library	microRNA	% of total miRNA reads
stage 8	hsa-miR-423-5p	9.9
	xtr-let-7c	6.3
	xtr-let-7f	6.0
	xtr-miR-148a	5.5
	xtr-miR-101a	5.3
stage 10	hsa-miR-423-5p	11.6
	dme-miR-184	7.9
	bmo-miR-184	7.8
	xtr-let-7f	6.1
	xtr-let-7c	5.0
stage 18	xtr-miR-206	42.4
	mmu-miR-10b	5.7
	rno-miR-10a-3p	5.7
	xtr-miR-130b	4.7
	mmu-miR-466i	4.3
stage 10 animal	dme-miR-184	13.3
	bmo-miR-184	13.2
	hsa-miR-423-5p	7.6
	xtr-let-7f	4.8
	xtr-let-7c	4.2
stage 10 vegetal	hsa-miR-423-5p	9.9
	xtr-let-7f	6.1
	bmo-miR-184	5.5
	xtr-miR-101a	5.3
	cfa-miR-378	4.6

Supplemental Figure legends

Figure S1. qPCR validation of dynamic and localized miRNA expression.

Total RNA from whole *X. tropicalis* embryos at stages 8, 10 and 18 and stage 10 embryos dissected into animal (A) and vegetal (V) hemispheres was prepared for qPCR (see Methods). miRNAs present in the small RNA libraries were detected by qPCR and RNA levels were calculated relative to *odc* or, in the case of xla-miR-148a, *U6*. Error bars are standard deviations from triplicate repeats. qPCR profiles (blue) were compared to the normalized miRNA read number (orange) obtained by small RNA sequencing.

Figure S2. Unannotated small RNAs map to introns and upstream of genes.

(*A*) Small RNAs map to introns of the *tdp1* gene reproducibly in all five small RNA libraries.

(*B*) Small RNA alignments and mRNA-seq alignments to the *gnai2* gene are shown. This is an example of a cluster of siteRNAs that is antisense to the gene.

(*C*) An example of small RNA alignments upstream of *tmem5* is shown.

The small RNA-seq and mRNA-seq reads are visualized in IGV as described in the legend to Figure 5.

Figure S3. siteRNA clustering.

siteRNAs were grouped according to their alignment to regions of TEs (Table 1). Representative siteRNA cluster sequences were aligned using Clustal Omega. Adenine, cytosine, guanine and thymidine are highlighted respectively in yellow, orange, red and blue. Note that the *arhge10l* siteRNA cluster contains three repeats of the group 7 core sequence.

References

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M et al. 2003. A uniform system for microRNA annotation. *RNA* **9:** 277-279.

Hafner M, Landgraf P, Ludwig J, Rice A, Ojo T, Lin C, Holoch D, Lim C, Tuschl T. 2008. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* **44:** 3-12.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31:** 3406-3415.

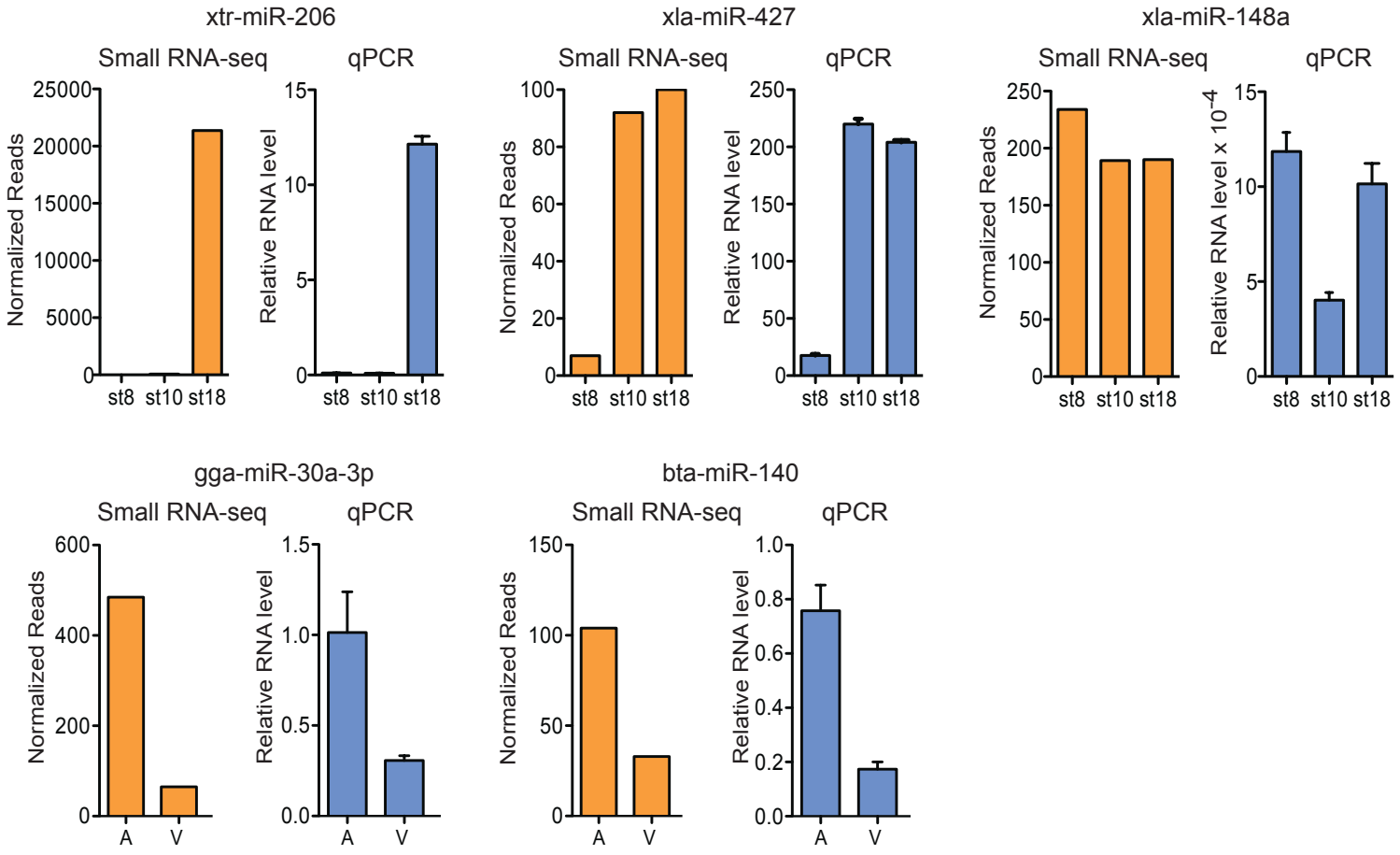
xtr-miR-206
Small RNA-seq
qPCR
Normalized Reads
Relative RNA level
st8 st10 st18
xla-miR-427
Small RNA-seq
qPCR
Normalized Reads
Relative RNA level
st8 st10 st18
xla-miR-148a
Small RNA-seq
qPCR
Normalized Reads
Relative RNA level x 10-4
st8 st10 st18
gga-miR-30a-3p
Small RNA-seq
qPCR
Normalized Reads
Relative RNA level
A V
bta-miR-140
Small RNA-seq
qPCR
Normalized Reads
Relative RNA level
A V

Figure S1

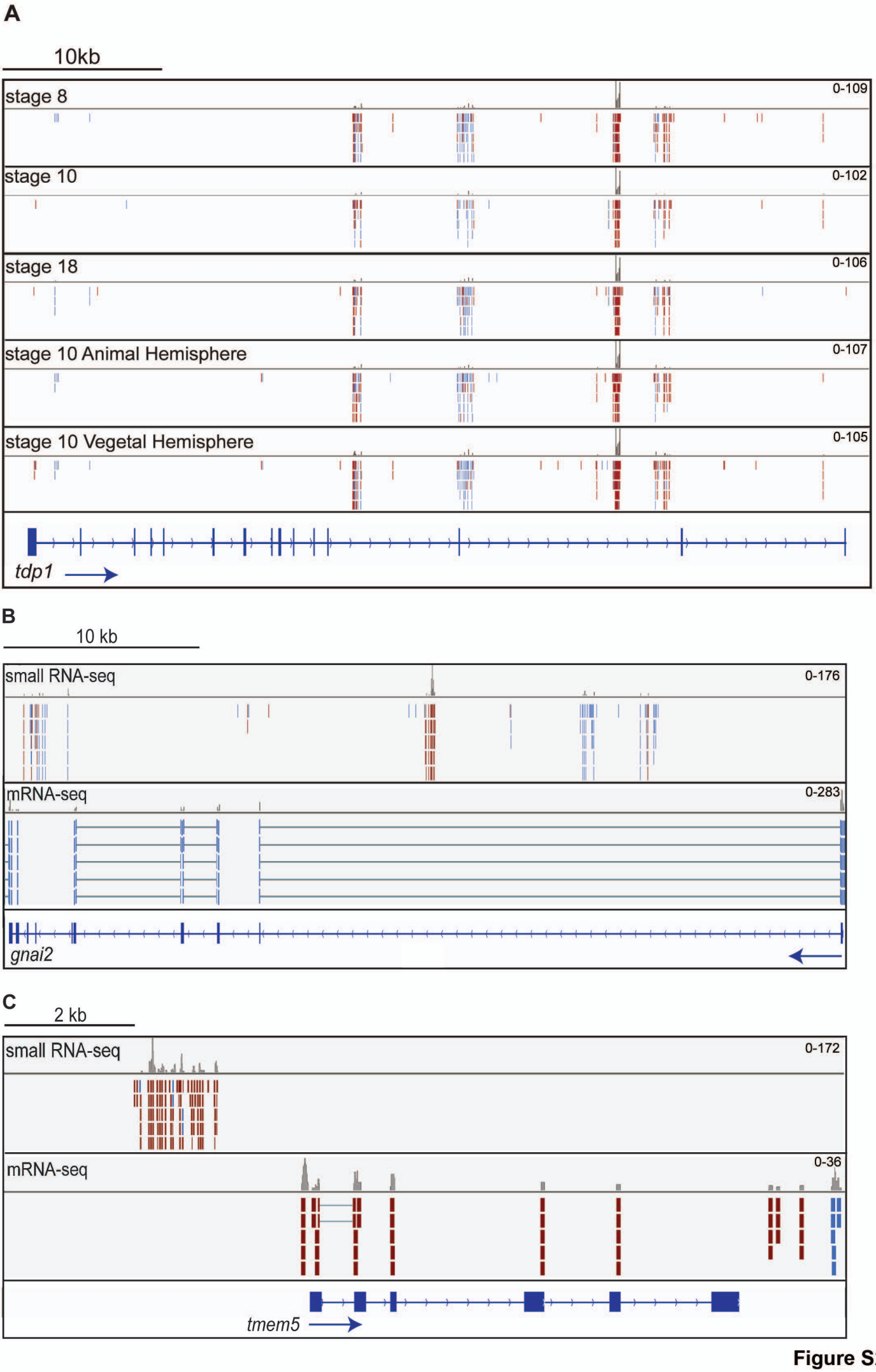
A
10kb
stage 8
0-109
stage 10
0-102
stage 18
0-106
stage 10 Animal Hemisphere
0-107
stage 10 Vegetal Hemisphere
0-105
tdp1
B
10 kb
small RNA-seq
0-176
mRNA-seq
0-283
gnai2
C
2 kb
small RNA-seq
0-172
mRNA-seq
0-36
tmem5

Figure S2

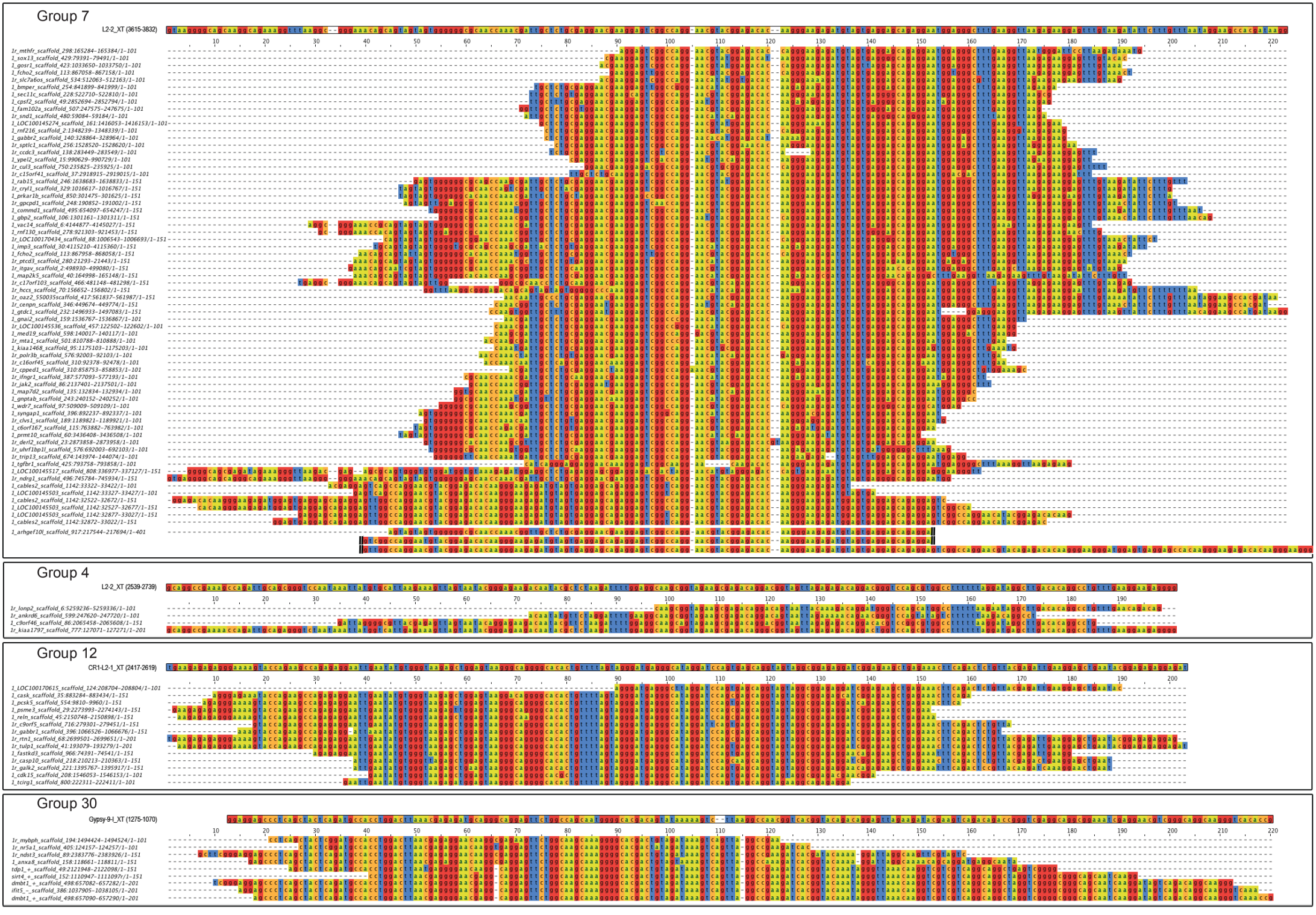
Group 7
L2-2_XT (3615-3832)
1r_mthfr_scaffold_298:165284–165384/1–101
1_sox13_scaffold_429:79393–79491/1–101
1_gosr1_scaffold_423:1033650–1033750/1–101
1_fcho2_scaffold_113:867058–867158/1–101
1r_slc7a6os_scaffold_534:512063–512163/1–101
1_bmper_scaffold_254:841899–841999/1–101
1_sec11c_scaffold_228:522710–522810/1–101
1_cpsf2_scaffold_49:2852694–2852794/1–101
1_fam102a_scaffold_507:247575–247675/1–101
1r_snd1_scaffold_480:59084–59184/1–101
1_LOC100145274_scaffold_161:1416053–1416153/1–101
1_rnf216_scaffold_2:1348239–1348339/1–101
1_gabbr2_scaffold_140:328864–328964/1–101
1r_sptlc1_scaffold_256:1528520–1528620/1–101
1r_ccdc3_scaffold_138:283449–283549/1–101
1_ypel2_scaffold_15:990629–990729/1–101
1r_cul3_scaffold_750:235825–235925/1–101
1r_c15orf41_scaffold_37:2918915–2919015/1–101
1_rab15_scaffold_246:1638683–1638833/1–151
1r_cryl1_scaffold_329:1016617–1016767/1–151
1_prkar1b_scaffold_850:301475–301625/1–151
1r_gpcpd1_scaffold_248:190852–191002/1–151
1_commd1_scaffold_495:654097–654247/1–151
1_gbp2_scaffold_106:1301161–1301311/1–151
1_vac14_scaffold_6:4144877–4145027/1–151
1_rnf130_scaffold_278:921303–921453/1–151
1r_LOC100170434_scaffold_88:1006543–1006693/1–151
1_imp3_scaffold_30:4115210–4115360/1–151
1_fcho2_scaffold_113:867958–868058/1–151
1r_ptcd3_scaffold_280:21293–21443/1–151
1r_itgav_scaffold_2:4989930–4990080/1–151
1_map2k5_scaffold_40:164998–165148/1–151
1r_c17orf103_scaffold_466:481148–481298/1–151
1r_hccs_scaffold_70:156652–156802/1–151
1r_oaz2_550035scaffold_417:561837–561987/1–151
1r_cenpn_scaffold_346:449674–449774/1–151
1_gtdc1_scaffold_232:1496933–1497083/1–151
1_gnai2_scaffold_159:1536767–1536867/1–101
1r_LOC100145536_scaffold_457:122502–122602/1–101
1_med19_scaffold_598:140017–140117/1–101
1r_mta1_scaffold_501:810788–810888/1–101
1_kiaa1468_scaffold_95:1175103–1175203/1–101
1r_polr3b_scaffold_576:92003–92103/1–101
1r_c16orf45_scaffold_310:92378–92478/1–101
1r_cpped1_scaffold_310:858753–858853/1–101
1r_ifngr1_scaffold_387:577093–577193/1–101
1r_jak2_scaffold_86:2137401–2137501/1–101
1_map7d2_scaffold_135:132834–132934/1–101
1_gnptab_scaffold_243:240152–240252/1–101
1_wdr7_scaffold_97:509009–509109/1–101
1_syngap1_scaffold_396:892237–892337/1–101
1r_clvs1_scaffold_189:1189821–1189921/1–101
1_c6orf167_scaffold_115:763882–763982/1–101
1_prmt10_scaffold_60:3436408–3436508/1–101
1r_derl2_scaffold_23:2873858–2873958/1–101
1r_uhrf1bp1l_scaffold_576:692003–692103/1–101
1r_trip13_scaffold_674:143974–144074/1–101
1_tgfbr1_scaffold_425:793758–793858/1–101
1_LOC100145517_scaffold_808:336977–337127/1–151
1r_ndrg1_scaffold_496:745784–745934/1–151
1_cables2_scaffold_1142:33322–33422/1–101
1_LOC100145503_scaffold_1142:33327–33427/1–101
1_cables2_scaffold_1142:32522–32672/1–151
1_LOC100145503_scaffold_1142:32527–32677/1–151
1_LOC100145503_scaffold_1142:32877–33027/1–151
1_cables2_scaffold_1142:32872–33022/1–151
1_arhgef10l_scaffold_917:217544–217694/1–401
Group 4
L2-2_XT (2539-2739)
1r_lonp2_scaffold_6:5259236–5259336/1–101
1r_ankrd6_scaffold_599:247620–247720/1–101
1_c9orf46_scaffold_86:2065458–2065608/1–151
1r_kiaa1797_scaffold_777:127071–127271/1–201
Group 12
CR1-L2-1_XT (2417-2619)
1_LOC100170615_scaffold_124:208704–208804/1–101
1_cask_scaffold_35:883284–883434/1–151
1_pcsk5_scaffold_554:9810–9960/1–151
1r_psme3_scaffold_29:2273993–2274143/1–151
1_reln_scaffold_45:2150748–2150898/1–151
1r_c9orf5_scaffold_716:279301–279451/1–151
1r_gabbr1_scaffold_396:1066526–1066676/1–151
1r_rtn1_scaffold_68:2699501–2699651/1–201
1r_tulp1_scaffold_41:193079–193279/1–201
1_fastkd3_scaffold_966:74391–74541/1–151
1r_casp10_scaffold_218:210213–210363/1–151
1r_galk2_scaffold_221:1395767–1395917/1–151
1_cdk15_scaffold_208:1546053–1546153/1–101
1_tcirg1_scaffold_800:222311–222411/1–101
Group 30
Gypsy-9-I_XT (1275-1070)
1r_mybph_scaffold_194:1494424–1494524/1–101
1r_nr5a1_scaffold_405:124157–124257/1–101
1r_ndst3_scaffold_89:2383776–2383926/1–151
1_anxa8_scaffold_158:118661–118811/1–151
tdp1_+_scaffold_49:2121948–2122098/1–151
sirt4_+_scaffold_152:1110947–1111097/1–151
dmbt1_+_scaffold_498:657082–657282/1–201
ifit5_-_scaffold_386:1037905–1038105/1–201
dmbt1_+_scaffold_498:657090–657290/1–201

Figure S3