Supplemental Methods

Preprocessing of HTS reads

As the two first of the eight libraries sequenced by Fasteris SA (Geneva) had reads limited to 35 nt in length, all HTS reads were truncated to 35 nt prior to analysis. To speed up subsequent computations, HTS reads from each library were first pooled by sequence and absolute counts encoded in sequence identifiers. Notably, quality values were discarded.

To remove sequencing adapters, each unique sequence was globally aligned with the Gex Adapter 2 (5'-TCGTATGCCGTCTTCTGCTTG; Illumina) using NEEDLE (EMBOSS package 6.1.0; Rice et al. 2000) with gap penalties of 10.0 and 0.5 (opening and extension, respectively). NEEDLE output was then automatically processed to trim sequences with a 3'-aligned region ≥ 10 nt having at most 3 edits (mismatch or gap) relatively to the adapter. These somewhat liberal thresholds were selected to maximize the number of usable reads.

After a repooling step carried out for each individual library, a data set combining all trimmed reads from the eight libraries was assembled (~240,000 unique sequences representing ~48,000,000 reads). This merged data set was then used to predict miRNA and snoRNA precursors, thus ensuring that all animals shared the same list of predictions.

Prediction of miRNA precursors

Ovine miRNA precursors were predicted genome-wide using core algorithms of MIRDEEP (Friedlander et al. 2008). For the *DLK1-GTL2* domain, we had our own ovine sequence (Charlier et al. 2001), whereas the cow genome (bosTau4 assembly; Liu et al. 2009) was used as a proxy for the remaining of the genome (due to the incompleteness of the sheep genome).

In practice, unique sequences were first mapped on the ovine domain or on the unmasked cow contigs using SOAP.CONTIG 1.11 (Li et al. 2008). The following parameters were selected as a reasonable compromise between speed and sensitivity considering that genome-wide predictions would be based on heterologous mapping: seed size (8 nt), maximum number of mismatches (2), maximum gap size (2 nt), maximum number of equal best hits (10 and all reported).

SOAP output was then filtered to discard sequences mapping to more than five genomic locations (as suggested in MIRDEEP protocol) and re-formatted to match the requirements of the EXCISE-CANDIDATES component of MIRDEEP pipeline. For the *DLK1-GTL2* domain, this step resulted in the excision of ~1,000 candidate precursors, while ~180,000 candidates were excised from bovine contigs.

Following MIRDEEP protocol, candidate precursors were submitted to RNAFOLD 1.6.5 (McCaskill 1990) using default parameters. In parallel, SOAP.SHORT was used to re-map unique sequences on candidate precursors with the same parameters as above except that the maximum number of equal best hits was raised to 10,000. RNAFOLD structures and SOAP output re-formatted as a 'signature' file were finally input to MIRDEEP core component, along with a list of metazoan miRNAs extracted from miRBase 12 (Griffiths-Jones 2006) for scoring conserved seeds. MIRDEEP was configured to make use of RANDFOLD 2.0 (Bonnet et al. 2004) and the score cut-off was kept to the default value (1).

Curation of miRNA precursors

For the ovine *DLK1-GTL2* domain, 49 precursors were predicted by MIRDEEP, of which 9 were unknown (miR-154b, miR-323b, miR-323c, miR-411b and miR-3955 to 3959) and one absent from miRBase (miR-376e) yet described in (Seitz et al. 2004). Moreover, careful examination revealed that

six precursors had been missed by MIRDEEP in spite of corresponding reads in our HTS libraries (see below for detailed protocol). In two cases (miR-494 and 665), precursors were not even excised because they would have been > 140 nt, which is the internal threshold of the EXCISE-CANDIDATES component. Three other candidates (miR-431, 433, 1197) were rejected because they failed to fold in a canonical stem-loop, while the last one (miR-412) had a score just below the cut-off. Finally, manual analysis of a MLAGAN (Brudno et al. 2003) multiple alignment of the domain showed that six more ovine precursors were conserved at ≥ 70% with their human orthologues (with no deletion in the mature sequence), though no read could be associated with these (miR-300, 337, 453, 654, 770, 889). Hence, the curated catalogue for the *DLK1-GTL2* domain comprises 61 miRNAs (49 predicted by MIRDEEP and 12 unpredicted yet very likely real precursors).

Of the 540 bovine precursors retained by MIRDEEP, 71 were discarded as false positives resulting from widespread mapping of mutated reads related to highly expressed miRNAs (miR-1, 206, 378, and the let-7 family). Specifically, taking advantage of the BLASTN tool of the miRBase 12 web interface, we filtered out precursors for which the first hit to the predicted mature sequence was any of these mature miRNAs with a bitscore < 35 bits. Using standalone NCBI BLASTN 2.2.22 (Altschul et al. 1997) to map the remaining bovine precursors on the ovine *DLK1-GTL2* domain, we further removed 46 precursors that were perfectly syntenic with their ovine orthologues. This let us with 423 bovine precursors identified by heterologous mapping of our ovine HTS reads. Even if miR-1 mature dominates our libraries (see main text), we noticed that only one (on BTA13) of its two known loci is actually predicted by MIRDEEP. Again, this failure is due to the candidate precursor on BTA24 being > 140 nt.

To assemble a genome-wide MIRDEEP catalogue of ovine miRNAs, the reduced bovine catalogue was combined to the raw ovine catalogue for the *DLK1-GTL2* domain (without the 12 unpredicted precursors), which resulted in a list of 472 precursors. The latter was used for all statistical analyses

pertaining to miRNA expression, whereas the curated ovine catalogue limited to the domain (61 precursors) was used in Fig. 1 and to study miRNA affinity for *DLK1*.

Annotation of miRNA precursors

To annotate our catalogue, we compared the 472 ovine precursors to the 615 bovine precursors present in miRBase 14. As the largest part of our catalogue had been predicted by heterologous mapping on the cow genome, orthologues of known bovine miRNAs were straightforward to identify by comparison of genomic locations. For all these precursors, we nevertheless checked that the ovine mature sequence (deduced from HTS reads) had at most two mismatches with its bovine counterpart. The remaining precursors were subdivided into three categories according to the outcome of two successive standalone BLASTN analyses against bovine then non-ruminant miRNA species present in miRBase 14 (using a word length of 12 nt). If any of the ovine "species" (i.e., mature, star, 5p, 3p; see below) aligned to the first hit with a bitscore > 32 bits, the corresponding ovine precursor was annotated either as a paralogue of a known bovine miRNA or as an orthologue of a non-ruminant miRNA, depending on the genomic source of the hit. Otherwise, the ovine precursor was considered unknown. In the *DLK1-GTL2* domain, manual annotation was preferred to automated annotation for miR-376e and miR-411b.

Clustering of the ovine precursors into families based on sequence similarity was carried out with the MCL algorithm (Enright et al. 2002). In a first step, all precursors were compared to each other with standalone BLASTN, again using a word length of 12 nt for maximum sensitivity. Then, the MCLBLASTLINE component of the MCL package was used to build and partition a graph from the pairwise bitscores. To limit false homologies, a bitscore cut-off (> 35 bits) was set after visual inspection and mining of the BLASTN report. The unique inflation parameter *I* was set to 2.0 as

described in the main text. The 484 precursors (including the 12 unpredicted precursors) clustered into 287 families, of which 224 were singletons. Note that family #61 involving one unpredicted precursor from the ovine *DLK1-GTL2* domain is not mentioned in the main text but shown in Fig. 1.

Prediction of C/D snoRNAs in the *DLK1-GTL2* domain

Ovine and bovine snoRNAs in the *DLK1-GTL2* domain were predicted with HMMER 2.3.2 (Durbin et al. 1998) using HMM profiles built from human snoRNAs. Human sequences for 14qI and 14qII families were downloaded from snoRNABase 3 (Lestrade and Weber 2006) and separately aligned with CLUSTALW 1.83 (Thompson et al. 1994). The HMMBUILD component of HMMER was then used to compute the two profiles and HMMSEARCH to scan the two ruminant loci. Domain-specific E-value cut-offs for 14qI and 14qII profiles were set to 1e-10 and 1e-04, respectively. We selected these values empirically as those yielding the closest set of snoRNAs in ruminants (10 x 14qI and 35/36 x 14qII) in comparison to the human domain (9 x 14qI and 31 x 14qII). As some snoRNAs were predicted by both profiles, the boundary between the two families that was eventually drawn might be slightly off. The lone 14q0 was localized in the bovine domain by mapping the known ovine sequence with standalone BLASTN. Finally, using the MLAGAN multiple alignment of the domain, we manually identified orthologues among ovine (including an additional copy of 14qII missed by HMMER), bovine, human and murine snoRNAs based on synteny and sequence similarity.

Gene annotation and conservation analyses in the *DLK1-GTL2* domain

To determine exon/intron boundaries for protein-coding and start/end positions for long non-coding transcripts in the ovine domain, we combined experimental evidence obtained in sheep with

human, murine and bovine transcripts annotated in public databases. Briefly, non-ovine unspliced transcripts were downloaded from Ensembl 57 using BIOMART (Haider et al. 2009) and the CONVERT tool of the UCSC genome browser to target the region orthologous to the ovine domain in hg19 (chr14:101,136,416-101,545,413), mm9 (chr12:110,645,537-110,991,765) and bosTau4 (chr21:65,669,039- 66,065,774). All transcripts were then jointly mapped on our ovine contigs using standalone BLASTN with an E-value threshold of 0.001. Finally, individual hits with an identity percentage > 80% that belonged to the same annotated gene were merged *in silico* to deduce crude ovine coordinates.

Sequence conservation for *MIRG* and *MEG8* was computed with PLOTCON (EMBOSS package 6.1.0) using a window size of 8 nt. Aligned blocks spanning the whole set of human miRNAs (miRBase 14) or snoRNAs (snoRNABase 3), respectively for *MIRG* and *MEG8*, were assembled using the Galaxy portal (Giardine et al. 2005) based on the 46-way MULTIZ alignment (UCSC genome browser). Aside from hg19, mm9 and bosTau4, blocks also contained sequences from panTro2, ponAbe2, rheMac2, rn4, cavPor3, equCab2, and canFam2 (all built with the Syntenic Net method). To the exception of miR-1193 that was manually added (as it is missing from miRBase), these analyses included only short RNA species described in human, yet generally conserved across mammals.

Quantitative analyses of HTS reads

To estimate miRNA expression from HTS data, unique sequences from each individual library were first re-mapped on the 472 precursors of the curated catalogue with SOAP.SHORT (see above for parameter values). Then, SOAP output was parsed to filter out reads with indels and those mapping on the reverse strand (relatively to the precursor sequence). Further, when a unique sequence had more than one best hit (with at most two mismatches), its absolute count (as encoded in the

identifier) was uniformly distributed among equal best hits. Two main outputs were generated from the automated parsing of this controlled mapping step: (i) empirical base counts at each position of each precursor and (ii) a quantitative inventory of all unique molecules (i.e., with a unique combination of start and end positions) spawned by each precursor. In both cases, HTS libraries were also parsed separately so that comparisons between genotypes (or individual animals) remained possible in addition to merged analyses. Empirical base counts were used to deduce genome-wide ovine sequences from HTS reads (see above) as well as to examine imprinting and editing, whereas the inventory was processed as follows.

Unique molecules from each precursor were first sorted by descending count and by location on the precursor (to break ties). Starting with the most abundant, molecules were then assigned to growing "stacks" based on their location. Stack height and width enlarged as required to accommodate new molecules except when this would result in merging two neighbouring "stacks". Instead, those rare molecules overlapping two stacks were discarded. At the end of this process, each precursor had from one to three non-overlapping stacks of molecules of defined height and width. Stacks were classified in "species" according to their relative height: if the 2nd highest was >= 0.15 as abundant as the highest, both were qualified as '5p' and '3p', depending on their location; otherwise, the highest was qualified as 'mature' and the 2nd highest (if present) as 'star'. The few additional stacks were considered minor (e.g., loop) and discarded. Though this standard nomenclature (Landgraf et al. 2007) is used in Suppl. Table S1, it had no practical influence on statistical analyses, with all species from each precursor equally considered in computations (see below). Within each stack, molecules were ranked according to their abundance. Except for isomir statistics reported in the main text, the most abundant molecule was selected to represent its whole species (in terms of both sequence and start/end positions), while all molecules contributed to the height of the stack corresponding to the species.

To assemble input tables for statistical analyses across libraries, stack heights for each precursor (at most two) were combined using species from the merged data set as a guide. Specifically, a stack from a particular library was included only if start/end positions of its most abundant molecule fell within 3 nt of those of the merged data set for the considered precursor. This approach allowed us to ignore species types during combination while discarding on a case-by-case basis libraries in which a given species was not comparable with those from other libraries. Resulting tables were normalised and further analysed with the R statistical software package (R Development Core Team, Vienna, Austria) as described in the main text. For the analyses restricted to the *DLK1-GTL2* domain, sub-tables were generated by filtering complete tables on precursor locations. Similar tables were assembled for snoRNAs by separately mapping HTS reads to the 47 "precursor" sequences identified in the ovine *DLK1-GTL2* domain.

Analyses of additional HTS libraries

Seven new HTS libraries were sequenced and analysed in house using the Illumina Sequencing Analysis Software 1.3. Contrary to the first batch of libraries, quality values were taken into account to filter reads using default parameters of the pipeline (i.e., those with a “chastity” less than 0.6 on two or more bases among the first 25 bases were discarded). After filtering, the average number of reads per library was very close to that found in the original experiment (6,452,524 vs. 6,324,668) yet with a much wider range (2,309,996—8,494,025 vs. 5,222,920—6,685,342). New reads were trimmed to 35 nt and their adapter removed before further processing. The corresponding merged data set had ~430,000 unique sequences representing ~43,500,000 reads (for seven libraries).

To allow comparison with the eight libraries already analysed, new sequences were simply re-mapped on the previously established miRNA/snoRNA catalogues and the output parsed as above.

Since statistical analyses actually required the use of the 15 HTS libraries, input tables from the first and second batch had to be combined in some way. This was carried out through a 'full join' between the two input tables where miRNA species rows from a given precursor were combined whenever start/end positions of the most abundant molecule of their respective merged data sets fell within 3 nt of each other. Even if reasonable, this strategy had the somewhat undesirable effect of generating two complementary 'star' rows for 28 precursors with unstable 'star' species. Given that 51 more miRNA species were only found in the second batch (37 in the other way around), this explains the difference between the numbers of species mentioned in the main text when respectively considering the first batch alone (747) or the two batches at the same time (747+28+51=826).

Comparison of HTS, Exiqon and Taqman data

As Exiqon probes had been designed for human miRNAs, it was crucial to identify those that could hybridize to ovine miRNAs to avoid analysing false negatives. To this end, we only considered Exiqon probes for which the target human mature sequence (as provided by Exiqon) was found unaltered in our ruminant precursors and at a location within 3 nt of the start position of the corresponding miRNA species (averaged across HTS1 and HTS2 merged data sets). In an attempt to limit many-to-many relationships in combined tables, we further discarded probes that were reported by Exiqon as targeting more than a single miRNA. This led to a combined table of 265 miRNA species. The eight Taqman probes were selected to ensure proper hybridization to ovine miRNAs and manually assigned to their HTS and/or Exiqon counterparts.

For the analyses of relative expression of domain-encoded miRNAs broken by genotype, HTS and Exiqon data were combined as above. However, for four Exiqon probes rejected by our stringent

combination strategy (miR-377-3p, miR-412-3p, miR-654-3p, miR-655-3p), we nevertheless observed the expected genotype effect, which indicates that these probes specifically hybridize to their targets. Therefore, they were manually added to the corresponding data sets.

Analyses of non-miRNA HTS reads

In contrast to other studies dealing with high throughput sequencing of short RNAs (e.g., Rathjen et al. 2009), an overwhelming fraction (~90—93%) of our HTS reads could be affiliated to one or more miRNA precursors. The remaining reads consisted either in 35-nt sequences lacking a recognizable sequencing adapter or in shorter (apparently non-miRNA) sequences followed by a regular adapter. To ensure that we did not miss any interesting short RNA species, both classes of "orphan" reads were pooled (within each batch of HTS libraries) and sequentially analyzed as summarized in Suppl. Fig. 18.

First, single base tracts were filtered out using a very simple heuristic: HTS reads (without adapter when recognizable) having >= 60% N's or >= 90% of either A, C, G, T (excluding N's) were tagged as poly-N (or A, C, G, T) and dropped from further analyses. For simplicity, these are reported as 'poly-A' since the latter amounts for 86—98% of the HTS reads identified with this strategy. Next, we assembled a custom database compiling the two sequencing adapters, all miRBase 14 precursors, human snoRNAs from snoRNABase 3, as well as the non-redundant content of Repbase 14.10 (Jurka et al. 2005). Using standalone BLASTN with an E-value threshold of 0.01, the remaining reads were then annotated by sequence similarity to the first database hit (when available). As this BLAST analysis was deliberately less stringent than the original SOAP mapping, it is not surprising to retrieve 41-49% of orphan reads corresponding to known miRNAs. In many cases, these new "miRNA" reads were either too short (< 19 nt) or too mutated (> 2 mismatches; indels) to be

effectively mapped by SOAP, while a third class consisted of reads mapping to real miRNA precursors missed by MIRDEEP. Among reads matching Repbase entries, we identified fragments of tRNAs, of both small- and large-subunit rRNAs, and of various other repeated sequences. Finally, the reads still to be annotated were likewise compared to a database of bovine cDNAs (built from Ensembl 56 using BIOMART). Beside the expected fragments of highly-expressed protein genes, this third analysis yielded a few additional tRNAs and snoRNAs, which were obviously absent from Repbase. At the end of the annotation process, 21—29% of the orphan reads remained unidentified (reported as 'unknown').

As shown in Suppl. Fig. 18, the qualitative breakdown of orphan reads is very similar between the two batches of HTS libraries, albeit relative abundance differs quite markedly. This is especially striking for two pairs of categories that appear to be somehow in competition for cloning or sequencing: adapters vs. poly-A and miR-1 vs. miR-378.

To look for endo-siRNAs (Nilsen 2008) in our orphan reads (in the two batches of HTS libraries), we used FINDPEAKS 4.0.8 (Fejes et al. 2008) followed by a manual analysis of the resulting peaks. In practice, orphan reads identified (in the BLASTN analysis) as poly-A tracts, sequencing adapters, miRNAs or snoRNAs were removed before proceeding. Reads were also filtered on length (19—25 nt) to ensure efficient mapping on the bovine genome. Though this eliminated all the remaining 35-nt sequences, visual inspection of randomly selected read batches confirmed that most dropped sequences were altered beyond recognition. Using SOAP.CONTIG with the same parameter values as for the prediction of miRNA precursors (see above), the qualifying reads were mapped on the unmasked cow contigs. Then, SOAP output was filtered to discard sequences mapping to more than five genomic locations (to avoid repetitive elements) and those mapping to unknown chromosome fragments (too numerous to handle with FINDPEAKS). After conversion to BED format, the filtered SOAP output was further processed by two accessory components of FINDPEAKS (SEPARATEREADS and

SORTFILES). FINDPEAKS itself was finally invoked with the following parameters: distribution type (3 = native or sequence coverage), minimum peak size (1), maximum PET (paired end tag) size (1,000 nt), and BEDGRAPH output. For the manual analysis, BEDGRAPH files for the two HTS experiments were simultaneously uploaded to the UCSC genome browser and all peaks above 400 in windows of 20,000 bp were examined by eye (89 chromosomal segments in total). This analysis did not provide any evidence for endo-siRNAs, and evaluated peaks corresponded to previously described categories of RNA, namely tRNAs, rRNAs, other repeat sequences and small pieces of protein coding genes that are known to be highly expressed in skeletal muscle .

A distinct analysis was carried out for the *DLK1-GTL2* domain. Briefly, the two initial SOAP mappings used to excise candidate precursors (see above) were directly processed by the FINDPEAKS pipeline using the same parameter values as genome-wide, except for the minimum peak size (10). This yielded 111 and 146 peaks, respectively for the first and second batch of HTS libraries. Peak locations were then compared to those of the 49 precursors predicted by MIRDEEP to remove peaks corresponding to already identified miRNAs. Finally, the remaining peaks were examined by eye and a coverage cut-off was set at 50, which resulted into 6 expressed precursors not predicted by MIRDEEP to be added to the domain miRNA catalogue (see above), as well as 8 more "stacks" that might correspond to additional miRNAs but were not further considered.

Evaluation of miRNA affinity for *DLK1*

To assess miRNA affinity for the *DLK1* transcript, we assembled two different catalogues of miRNA species encoded by the *DLK1-GTL2* domain. The first one was limited to ovine miRNAs and contained the most abundant representative (when expressed) of the 110 species identified after the first HTS experiment, as well as 4 species specific to the second HTS experiment, 6 alternative molecules

reaching > 70% of the most expressed molecule, and 7 edited variants, thus amounting to 127 distinct sequences, all considered equally. In contrast, the second catalogue included 70 human and 77 murine "orthologues" in addition to the 114 (110+4) most abundant ovine miRNA species. When ovine species displayed an alternative 5p extremity relative to miRBase annotation, we considered both possibilities by extracting additional shifted orthologues from human and murine precursors (and in the other way around). This concerned miR-127 (5p), 381 (3p), 409 (3p), 411a (5p), 431 (3p), 487a (3p), 539 (5p), and 654 (5p). Altogether, the second catalogue included 140 triplets of sequences (114 "ovine" species, 8 shifted species, 2 human- and 15 murine-specific species, as well as the human/murine version of miR-668) and was used to compute multiorganism (MO) scores summed across the three orthologous sequences of the *DLK1* transcript (see main text for details).

Two different scoring engines were used. For G-scores, we used a customized version of COMPSEQ (EMBOSS package 6.1.0) allowing to look for k-mers longer than 6 nt. M-scores were computed with MIRANDA 3.0-sept2008 (Betel et al. 2008) using default parameter values, except that we set a conservative score cut-off (140) to avoid spurious weak matches and to speed up computations. Note that using older versions of MIRANDA yielded surprisingly contrasted predictions for some miRNA species (data not shown).

Shuffling was carried out separately for each region of the *DLK1* transcripts (5'-UTR, CDS, 3'-UTR). In all cases, the algorithm was as follows. Using a 3-nt overlapping sliding window, the numbers of occurrences of all nucleotide triplets in the region are recorded. From these counts, we deduce (1) probabilities for starting dinucleotides and (2) conditional probabilities for single base extensions. To generate shuffled sequences, a semi-empirical strategy was preferred to a purely probabilistic approach after comparison of the performances of the two methods (data not shown). First, a starting dinucleotide is randomly drawn from the empirical distribution. Then, the sequence is extended until the required length by drawing the next nucleotide conditionally on the last two

nucleotides. For each nucleotide, we first try to draw from empirical pools containing the true number of single base extensions for each prefix. If this fails (because the current prefix pool has been exhausted before reaching the end of the nascent sequence), we draw from the empirical distribution of single base extensions corresponding to the current prefix. To maximize the odds to get shuffled sequences as close as possible to the true sequence in compositional terms, we generate twice the required number of sequences, rank candidates according to the sum of triplet count differences with the original sequence, and keep the upper half of the list. Hence, to get 10,000 shuffled variants of the 3'-UTR of *DLK1*, we actually generated 20,000 sequences, of which we discarded the 10,000 sequences that were the least similar in trinucleotide composition to the real 3'-UTR.

After scoring all sequences, p-values were computed by comparing each of the species or "quadrille" scores obtained with the true sequence of interest to the corresponding distribution of scores obtained with the shuffled sequences. To account for multiple testing, simple Bonferroni adjustments were achieved by dividing the 0.05 significance threshold by the number of tests.

GO Analysis of miRNA targets

Conserved target sites for conserved miRNA families were downloaded from the human TargetScan 5.1 website (Friedman et al. 2009). From this initial multispecies list, we first discarded all rows corresponding to non-bovine sites. Then, we further filtered rows to keep only those sites targeted by one of the 24 conserved miRNA families with representatives in the *DLK1-GTL2* domain: miR-127, miR-134, miR-136, miR-154, miR-299/299-3p, miR-300, miR-329/362-3p, miR-370, miR-376/376ab/376b-3p, miR-376c, miR-377, miR-379, miR-382, miR-410, miR-411, miR-431, miR-433,

miR-487/487b, miR-494, miR-495/1192, miR-496, miR-543, miR-544, miR-758. This resulted into a list of 4,798 sites spread among 2,832 bovine 3'-UTRs.

Control sets of genes of the same sample size were randomly drawn among the 8,458 bovine genes present in the TargetScan list. This is noteworthy as it means that control genes have all at least one conserved target site for a conserved miRNA family, which could lead to an under-representation of some functional classes (e.g., house-keeping genes). This was carried out either 10,000 or 200,000 times, depending on the total size of GO graph used in the analysis (see main text for details). Whenever a given gene was selected (either randomly or through a cognate miRNA), all the associated GO terms got a hit. This was done without weighting for the number of miRNA target sites in the 3'-UTRs. At the end of the random sampling, the enrichment was statistically evaluated as for *DLK1* affinity, i.e., by comparing the number of hits obtained by each individual GO term with the true set of genes to the distribution of the number of hits obtained by the same term with the sets of control genes. Bonferroni adjustments used the number of GO (or GO Slim) terms actually associated to the bovine TargetScan list (6,930 for the whole GO graph and 55 for the GO Slim graph).

GO associations were fetched from the EBI GOA project (Barrell et al. 2009). We used the human GOA annotation (release 81, 21 January 2010) since TargetScan only provides human gene symbols, even for bovine orthologues. The mapping to GO Slim terms was performed using the corresponding GOA Slim map also available at the EBI FTP server. Annotation of GO terms was achieved by using a local mirror of the GO termdb downloaded from http://www.geneontology.org/ on March 1st, 2010.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17): 3389-3402.

Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C, Apweiler R. 2009. The GOA database in 2009--an integrated Gene Ontology Annotation resource. *Nucleic Acids Res* **37**(Database issue): D396-403.

Betel D, Wilson M, Gabow A, Marks DS, Sander C. 2008. The microRNA.org resource: targets and expression. *Nucleic Acids Res* **36**(Database issue): D149-153.

Bonnet E, Wuyts J, Rouze P, Van de Peer Y. 2004. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* **20**(17): 2911-2917.

Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S. 2003. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* **13**(4): 721-731.

Charlier C, Segers K, Karim L, Shay T, Gyapay G, Cockett N, Georges M. 2001. The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting their imprinting status. *Nat Genet* **27**(4): 367-369.

Durbin R, Eddy SR, Krogh A, Mitchison G. 1998. *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge Univ Pr.

Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**(7): 1575-1584.

Fejes AP, Robertson G, Bilenky M, Varhol R, Bainbridge M, Jones SJ. 2008. FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics* **24**(15): 1729-1730.

Friedlander MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, Rajewsky N. 2008. Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol* **26**(4): 407-415.

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**(1): 92-105.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* **15**(10): 1451-1455.

Griffiths-Jones S. 2006. miRBase: the microRNA sequence database. *Methods Mol Biol* **342**: 129-138.

Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A. 2009. BioMart Central Portal--unified access to biological data. *Nucleic Acids Res* **37**(Web Server issue): W23-27.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**(1-4): 462-467.

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**(7): 1401-1414.

Lestrade L, Weber MJ. 2006. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res* **34**(Database issue): D158-162.

Li R, Li Y, Kristiansen K, Wang J. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**(5): 713-714.

Liu Y, Qin X, Song XZ, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y et al. 2009. Bos taurus genome assembly. *BMC Genomics* **10**: 180.

McCaskill JS. 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**(6-7): 1105-1119.

Nilsen TW. 2008. Endo-siRNAs: yet another layer of complexity in RNA silencing. *Nat Struct Mol Biol* **15**(6): 546-548.

Rathjen T, Pais H, Sweetman D, Moulton V, Munsterberg A, Dalmay T. 2009. High throughput sequencing of microRNAs in chicken somites. *FEBS Lett* **583**(9): 1422-1426.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**(6): 276-277.

Seitz H, Royo H, Bortolin ML, Lin SP, Ferguson-Smith AC, Cavaille J. 2004. A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res* **14**(9): 1741-1748.

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 Res* **22**(22): 4673-4680.

Supplemental Figures and Tables

Supplemental Figure 1: Working model for polar overdominance at the ovine *CLPG* locus.

Supplemental Figure 2: Number of microRNA precursors [predicted by MIRDEEP (Friedlander et al. 2008) based on deep sequencing of small RNA libraries from skeletal muscle of sheep] that map to the different bovine chromosomes. The bovine genome sequence was used as reference for the MIRDEEP analysis except for 390 Kb of ovine sequence corresponding to the core of the *DLK1-GTL2* domain encompassing the *CLPG* mutation. U: unassigned sequence contigs.

Supplemental Figure 3: (A) Ratio of reads mapping to the 5p arm for the 472 precursors predicted by MIRDEEP. The graph illustrates the excess of miRNAs spawned from the 5p arm. The shaded areas cover precursors for which either the 5p or 3p arm generates > 87% of the reads, thus defining a star species issued from the other arm. **(B)** Base pair composition of miRNAs derived from the 5p arm (upper panel) or 3p arm (lower panel). miRNA from the populous miR-2284 family were excluded from the analysis. Data are shown for the second HTS experiment, but were in essence identical to that of the first HTS experiment. Position one of the miRNA was defined as the starting position of the most expressed isomir.

Supplemental Figure 4: Plot of average pairwise similarities of sliding 8-nt windows across 10 mammals (human, chimpanzee, orangutan, rhesus macaque, guinea pig, mouse, rat, cow, horse, dog) of the *MIRG* **(A)** and *MEG8* **(B)** genes. The position and identifier of the miRNAs **(A)** and C/D snoRNAs **(B)** are given as horizontal orange bars. The coincidence between the peaks of conservation and miRNAs but not C/D snoRNAs positions is striking.

Supplemental Figure 5: Correlation between the expression levels of 826 miRNA species in skeletal muscle (longissimus dorsi) of eight-week old sheeps representing the four possible *CLPG* genotypes

(+/+; CLPG^{Mat}/+^{Pat} = C/+; +^{Mat}/CLPG^{Pat} = +/C; CLPG/CLPG = C/C) estimated from the number of HTS reads obtained in two independent experiments (HTS1 and HTS2). Two animals were analyzed per genotype, except *+/C* for which only one animal was sequenced twice (HTS1 + HTS2). Raw read numbers were divided by the total number of reads for the corresponding animal and multiplied by the average total number of reads across animals. X- and Y-axis correspond to log-base-2 of the corresponding normalized number of reads for the first (HTS1) and second (HTS2) sequencing experiment, respectively. R_P and R_S are Pearson and Spearman correlation coefficients, respectively.

Supplemental Figure 6: Number of sequence reads mapping to the 5p and 3p arm of the miR-382 precursor in the first (HTS1) and second (HTS2) sequencing experiments for the eight animals representing the four possible *CLPG* genotypes. It can for instance be seen that in HTS1, two individuals *(+/C 1* and *C/C 2*) showed 5p/3p ratios that were in opposite direction as the other individuals. The HTS2 experiment did not confirm this finding for *C/C 2*.

Supplemental Figure 7: (A&B) Correlation between the expression levels of 265 miRNA species in skeletal muscle (longissimus dorsi) of eight-week old sheep representing the four possible *CLPG* genotypes *(+/+; CLPG^{Mat}/+^{Pat} = C/+; +^{Mat}/CLPG^{Pat} = +/C; CLPG/CLPG = C/C*) estimated from (i) the number of HTS reads (A: HTS1 and B: HTS2) and (ii) fluorescent intensities obtained by hybridizing total Hy3-labelled RNA on Exiqon miRCURY™ LNA (Version 9.2 - updated to miRbase 11.0). Raw read numbers were divided by the total number of reads for the corresponding animal and multiplied by the average total number of reads across animals. Raw fluorescent intensities were divided by the total intensity of the corresponding animal and multiplied by the average total intensity across animals. X- and Y-axis correspond to log-base-2 of the corresponding normalized number of HTS reads and normalized intensities, respectively. R_P and R_S are Pearson and Spearman correlation coefficients, respectively. **(C&D)** Number of miRNA pairs ranked congruently (green) or incongruently (red) with regard to expression level as assessed by HTS (C: HTS1 and D: HTS2) versus

Exiqon. miRNA pairs are categorized according to the fold difference in expression level assessed on the Exiqon arrays (X-axis). The numbers under the bars correspond to the % of incongruent pairs in the corresponding category.

Supplemental Figure 8: Comparison of miRNA expression levels (relative to *let-7d*) in skeletal muscle of 8-week old animals representing the four possible *CLPG* genotypes *(+/+; $CLPG^{Mat}/+^{Pat}$ = C/+; $+^{Mat}/CLPG^{Pat}$ = +/C; CLPG/CLPG = C/C*), estimated by: (i) HTS (**A**: HTS1 and **B**:HTS2), (ii) Exiqon array-hybridization and (iii) QRT-PCR. Relative expression levels are measured on a log-base-2 scale. Bars corresponding to miRNAs also studied by QRT-PCR are colored as indicated in the legend.

Supplemental Figure 9: Quantitative RT-PCR analysis of the expression level of *DLK1* and *GTL2* in longissimus dorsi of eight week old sheep representing the four possible genotypes at the *CLPG* locus: *+/+* (gray bars), $CLPG^{Mat}/+^{Pat}$ (red bars), $+^{Mat}/CLPG^{Pat}$ (blue bars), *CLPG/CLPG* (purple bars). Expression levels are relative to the *+/+* individual marked as "Ref". Error bars correspond to standard errors of the estimates over triplicates.

Supplemental Figure 10: Illustration of the need for a correction of $\log_2(i/m)$ to account for the deviations of the average (across miRNAs) of this parameter between individuals. Two individuals are shown, corresponding to a *+/+* (A) and a *CLPG/CLPG* animal (B). Each dot corresponds to a miRNA mapping either to the *DLK1-GTL2* domain (red diamonds) or not (gray circles). The X-axis corresponds to $\log_2$ of the ratio between the number of adjusted reads for the individual over the average number of adjusted reads across the eight samples. The Y-axis corresponds to $\log_2$ of the ratio between the normalized Hy3 fluorescence of the sample over the normalized Hy5 fluorescence of a pool of the eight samples. It can be seen that for most of the miRNAs, both individuals have low $\log_2(i/m)$ despite the fact that miRNA read numbers were adjusted for the total number of mappable reads for a given individual. For unexplained reasons, some animals indeed had zero or near zero

reads for miRNAs that were characterized by low level expression, yet with considerably higher read numbers in most animals. The correction, based on the average $\log_2$ ratio over non *DLK1-GTL2* miRNAs (to avoid erasing the *CLPG* effect if it exists) shifts the cloud to the right.

Supplemental Figure 11: Average expression level, relative to the mean expression level of seven individuals sequenced twice (HTS1 and HTS2), of 25 small RNA species derived from C/D snoRNAs within the *DLK1-GTL2* domain in skeletal muscle of eight sheep sorted by *CLPG* genotype *(*gray: *+/+;* blue: $+^{Mat}/CLPG^{Pat}$; red: $CLPG^{Mat}/+^{Pat}$*;* purple: *CLPG/CLPG*). Error bars correspond to 1.96 x the standard error of the estimate.

Supplemental Figure 12: (A) $\text{Log}_{10}(1/p)$ values of the effect of *CLPG* genotype on the expression level of 265 small RNAs in skeletal muscle of eight 8-week old sheep. Expression levels were estimated from fluorescence intensity on Exiqon arrays**.** The statistical significance of the *CLPG* effect was estimated by ANOVA. Gray vertical bars correspond to miRNAs outside of the *DLK1-GTL2* domain, red vertical bars to miRNAs from the *DLK1-GTL2* domain and orange vertical bars to small RNAs derived from C/D snoRNA precursors. Horizontal black lines correspond to the nominal (plain line) and Bonferroni-adjusted (dotted line) 5% significance thresholds. Horizontal blue bars mark the different chromosomes (right Y-axis). UN: correspond to unassigned sequence contigs. (**B).** Average expression level, relative to the mean expression level of seven individuals sequenced twice (HTS1 and HTS2), of 34 miRNAs from the *DLK1-GTL2* domain in skeletal muscle of eight sheep sorted by *CLPG* genotype *(*gray: *+/+;* blue: $+^{Mat}/CLPG^{Pat}$; red: $CLPG^{Mat}/+^{Pat}$*;* purple: *CLPG/CLPG*). Error bars correspond to 1.96 x the standard error of the estimate.

Supplemental Figure 13: Relative abundance of miRNA from the *DLK1-GTL2* domain in skeletal muscle of eight 8-week old sheep representing the four possible *CLPG* genotypes (**A**: *+/+;* **B:** $CLPG^{Mat}/+^{Pat}$ *= C/+;* **C:** $+^{Mat}/CLPG^{Pat}$ *= +/C;* **D:** *CLPG/CLPG = C/C*). Upper half: Reads derived from the

corresponding miRNA expressed as a percentage of the total number of reads derived from all miRNAs in the *DLK1-GTL2* domain. Lower half: Fluorescence intensity for the corresponding miRNA expressed as a percentage of the total fluorescence for all miRNAs in the *DLK1-GTL2* domain. miRNAs that could not be interrogated on the Exiqon array are marked by asterisks. The colors of the vertical bars distinguish the individuals and sequencing experiments as indicated in the inlet. The limits of the long non-coding RNA precursors are marked by the blue bars underneath the graph.

Supplemental Figure 14: Experiment-wide log(1/p)-values (vertical red lines) of the effect of *CLPG* genotype on the expression level of 265 miRNAs in longissimus dorsi of eight 8-week old animals (2x *+/+;* 2x $CLPG^{Mat}/+^{Pat}$*;* 2x $+^{Mat}/CLPG^{Pat}$*;* 2X *CLPG/CLPG*). The analysis combines HTS and array hybridization data. In this analysis, the actual p-value of the miRNA-specific F-statistics was determined from permutation data rather than from tables. The horizontal blue bars mark the limits between the different chromosomes (right Y-axis). The horizontal black lines correspond to the nominal (full line) and Bonferroni-corrected (265 tests; dotted line) 5% significance thresholds.

Supplemental Figure 15: (A) Statistical significance (log(1/p)) of the affinity of ovine miRNAs in the *DLK1-GTL2* domain for the 5'UTR, coding sequence (ORF) and 3'UTR of the ovine *PEG11*. The affinity was measured using either G- (blue) or M-scores (orange) as defined in the text. The last pair of bars ("quad") at the right of the graph correspond to the quadrille scores, the remaining bars to the species-scores and are labeled accordingly. p-values were determined using the sequence shuffling test described in the main text. Species-scores require a Bonferroni correction for 127 (or 140) independent tests. **(B)** Same as in (A) except that the scores are "multiorganism (MO) scores" combining information from sheep, human and mouse.

Supplemental Figure 16: (A) Frequency distribution of read numbers ($\log_{10}$) mapping to previously known miRNA precursors. **(B)** Same for newly detected miRNA precursors.

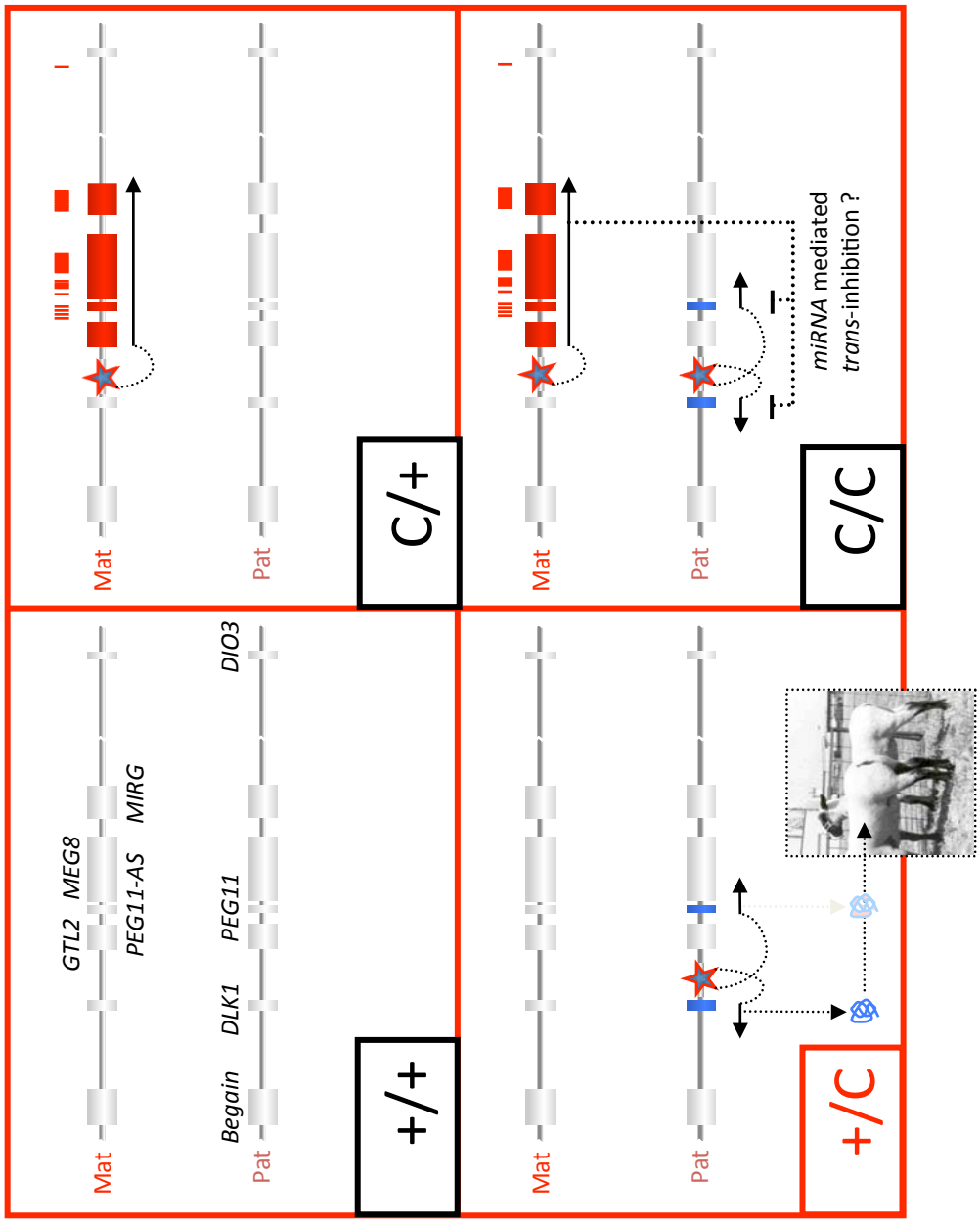
Supplemental Figure 17: (A) Small RNAs derived from C/D snoRNA precursors mapping to the *DLK1-GTL2* domain. Isomirs are aligned with their cognate precursors, and numbers of corresponding reads indicated. The position of the most abundant 5p (red) and 3p (green) derived small RNAs on the RNAFOLD-predicted hairpin is shown. **(B)** CLUSTALW alignment of the 12 C/D snoRNAs. Consensus sequence and localization of C/D and C'/D' boxes are displayed.

Supplemental Table 1: Genome-wide MIRDEEP miRNA catalogue built from HTS data. Precursor names are taken from miRBase 14, based on sequence similarity of the corresponding miRNA species. Chromosomal locations are expressed in bovine coordinates, except for the *DLK1-GTL2* domain where coordinates are relative to the revised GenBank entry AF354168. For clarity, only non-singleton family numbers are shown. Species types (mature, star, 5p, 3p) were determined as in Landgraf et al. (2007). Note that species types, sequences and seeds are based on HTS1 data only, with the most abundant isomir from each arm selected as the representative of its species. Read numbers are computed on the merged data sets and correspond to the height of the stack for each miRNA species. See main text and Supplemental Methods for details.

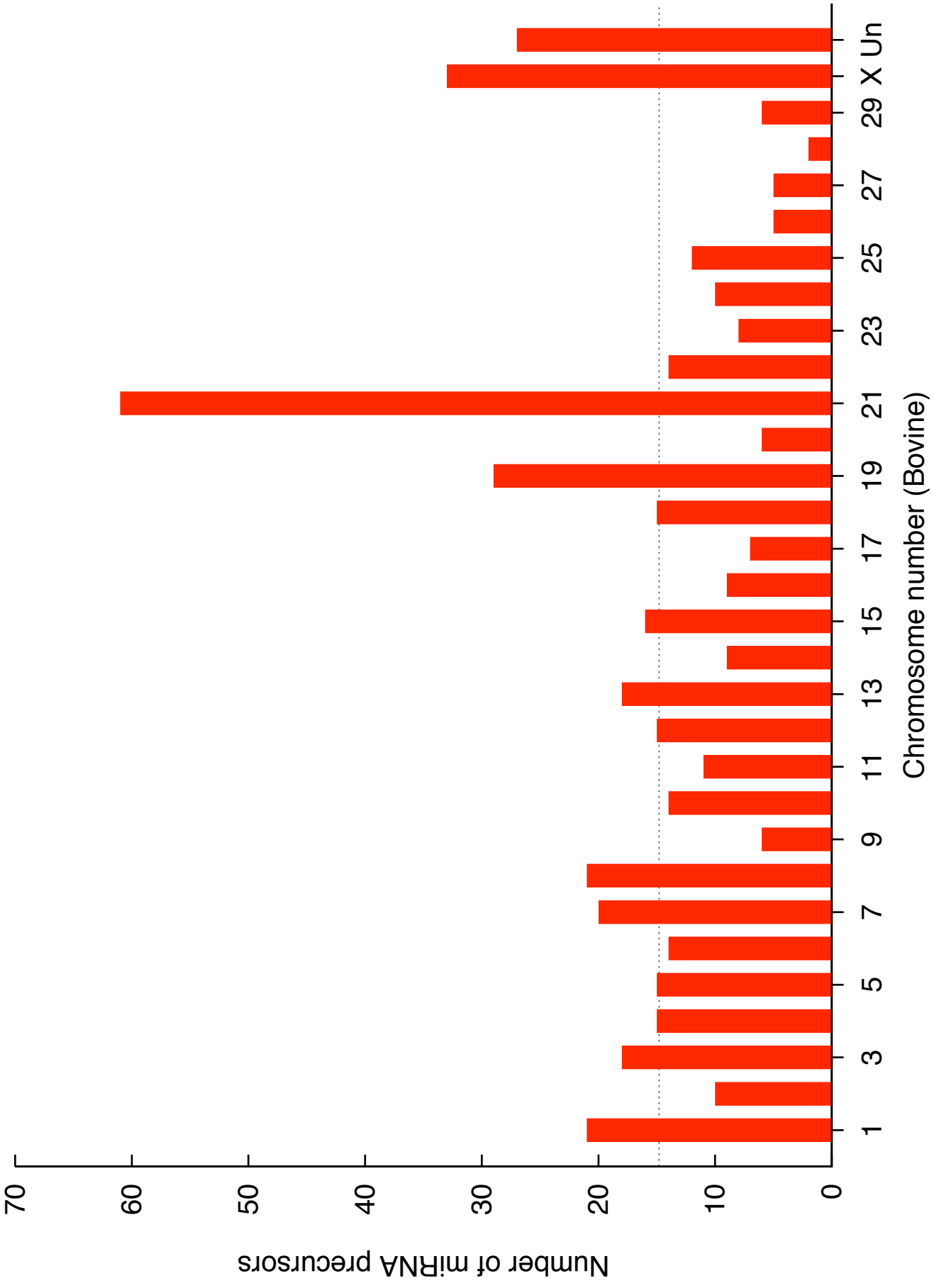
Supplemental Table 2: (A) Gene ontology terms enriched amongst mammalian genes with conserved target sites for miRNAs in the *DLK1-GTL2* domain. **(B&C)** List of bovine genes targeted by miRNAs encoded in the *DLK1-GTL2* domain for which the GO annotation was significantly enriched. Only 3'-UTR conserved target sites for conserved miRNA families identified in the human TargetScan were considered. To be included, a given gene has to be (1) targeted by at least one the miRNA families with representatives in the *DLK1-GTL2* domain and (2) associated to one or more of the eight most enriched GO terms. B & C correspond to two variants of the GO analysis (Slim and whole; see main text and Supplemental Methods for details).

Supplemental Table 3: Mouse and sheep primer sequences used in RNA editing analysis of miRNA precursors from *DLK1-GTL2* domain.

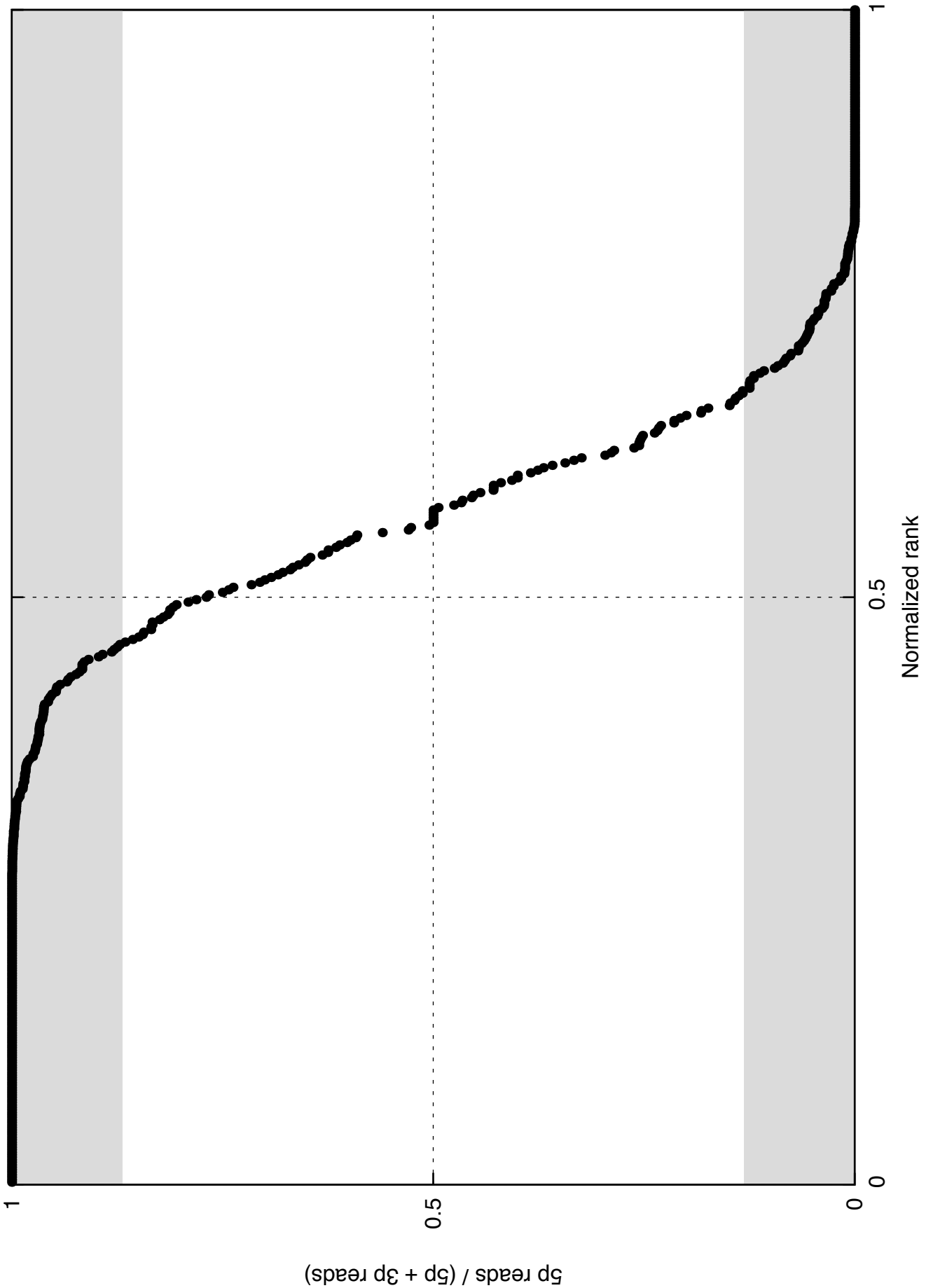
Suppl. Fig. 1


GTL2
MEG8
PEG11-AS
MIRG
DIO3
PEG11
DLK1
Begain
Mat
Pat
+/+
C/+
C/C
+/C
miRNA mediated
trans-inhibition ?

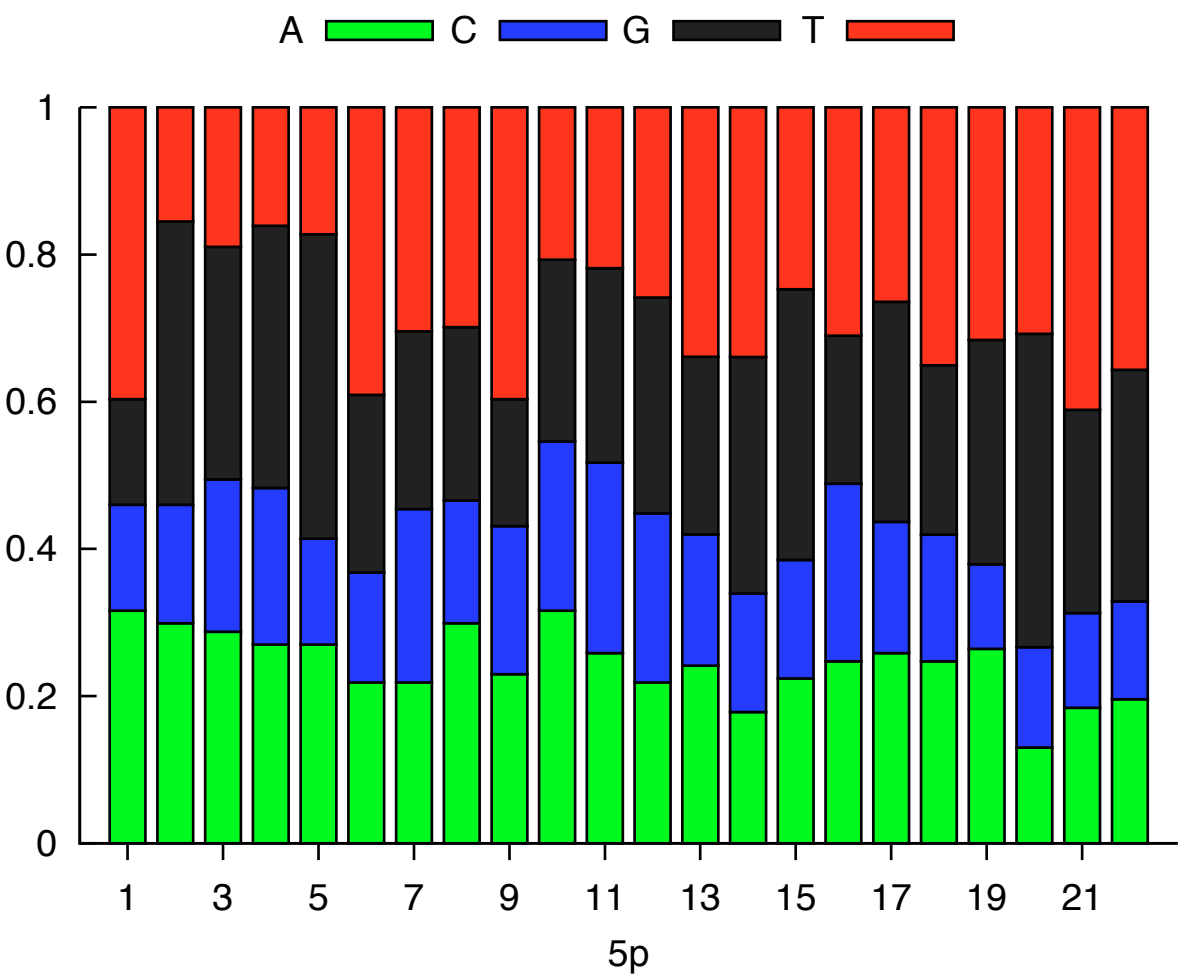
Suppl. Fig. 2

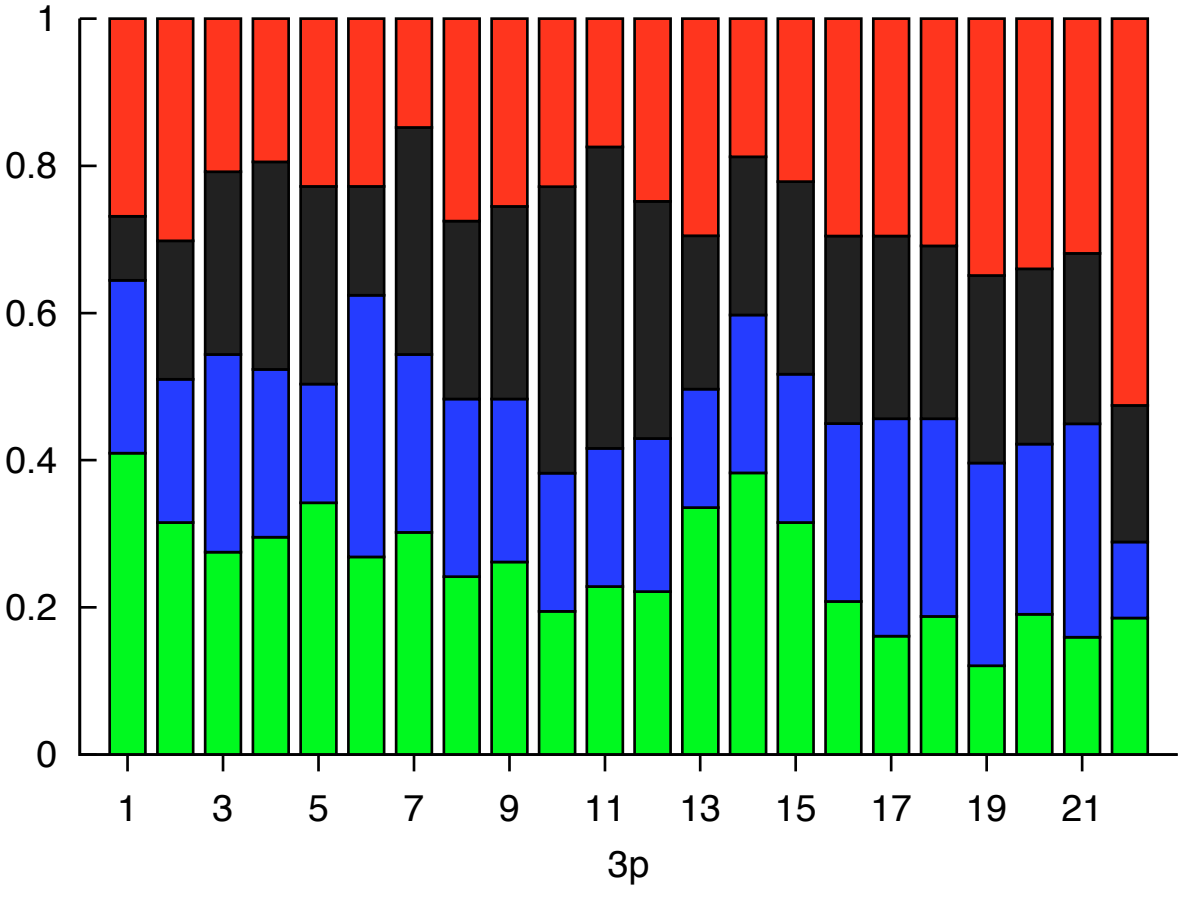

Number of miRNA precursors
0
10
20
30
40
50
60
70
Chromosome number (Bovine)
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
X
Un

Suppl. Fig. 3A

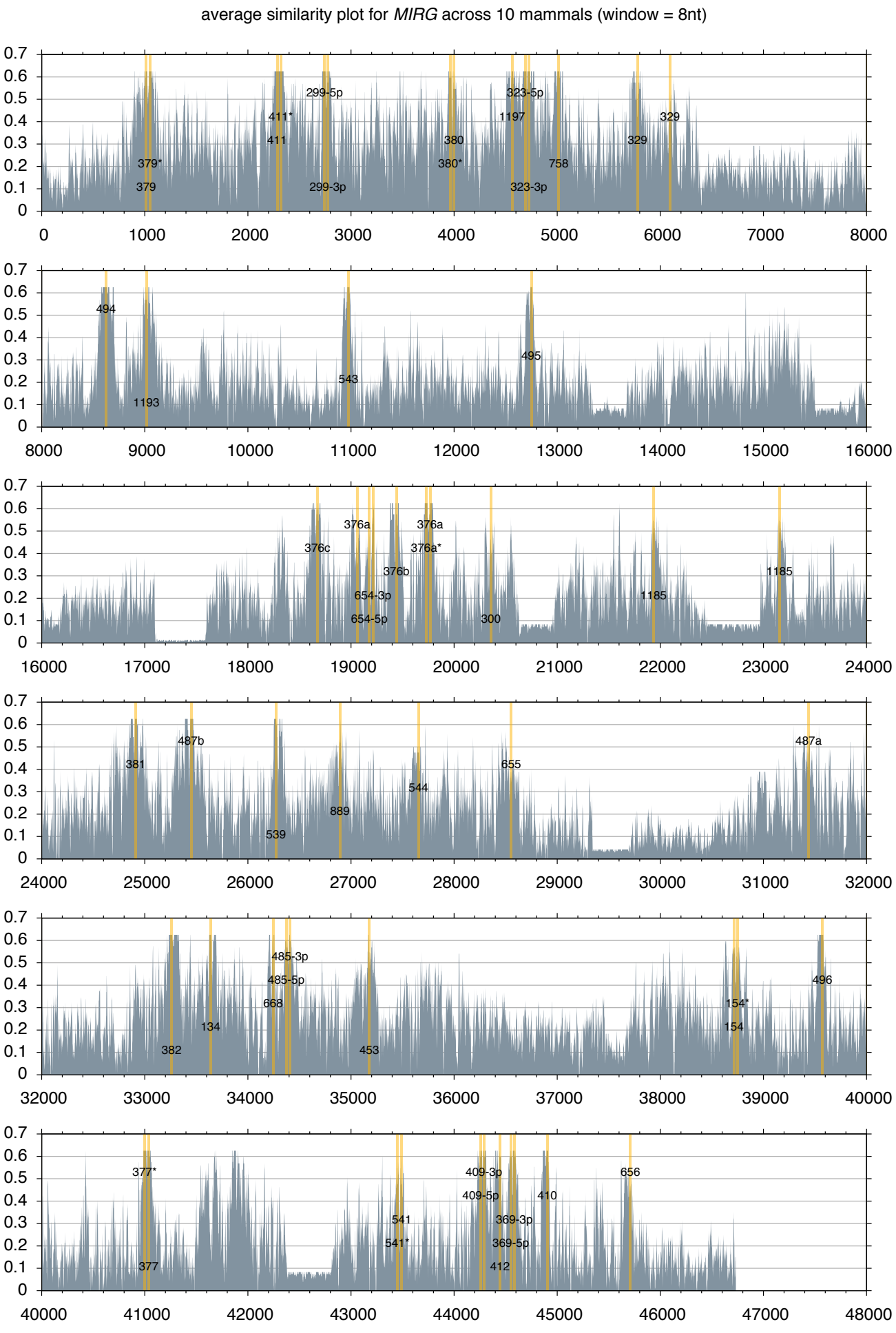


Suppl. Fig. 3B

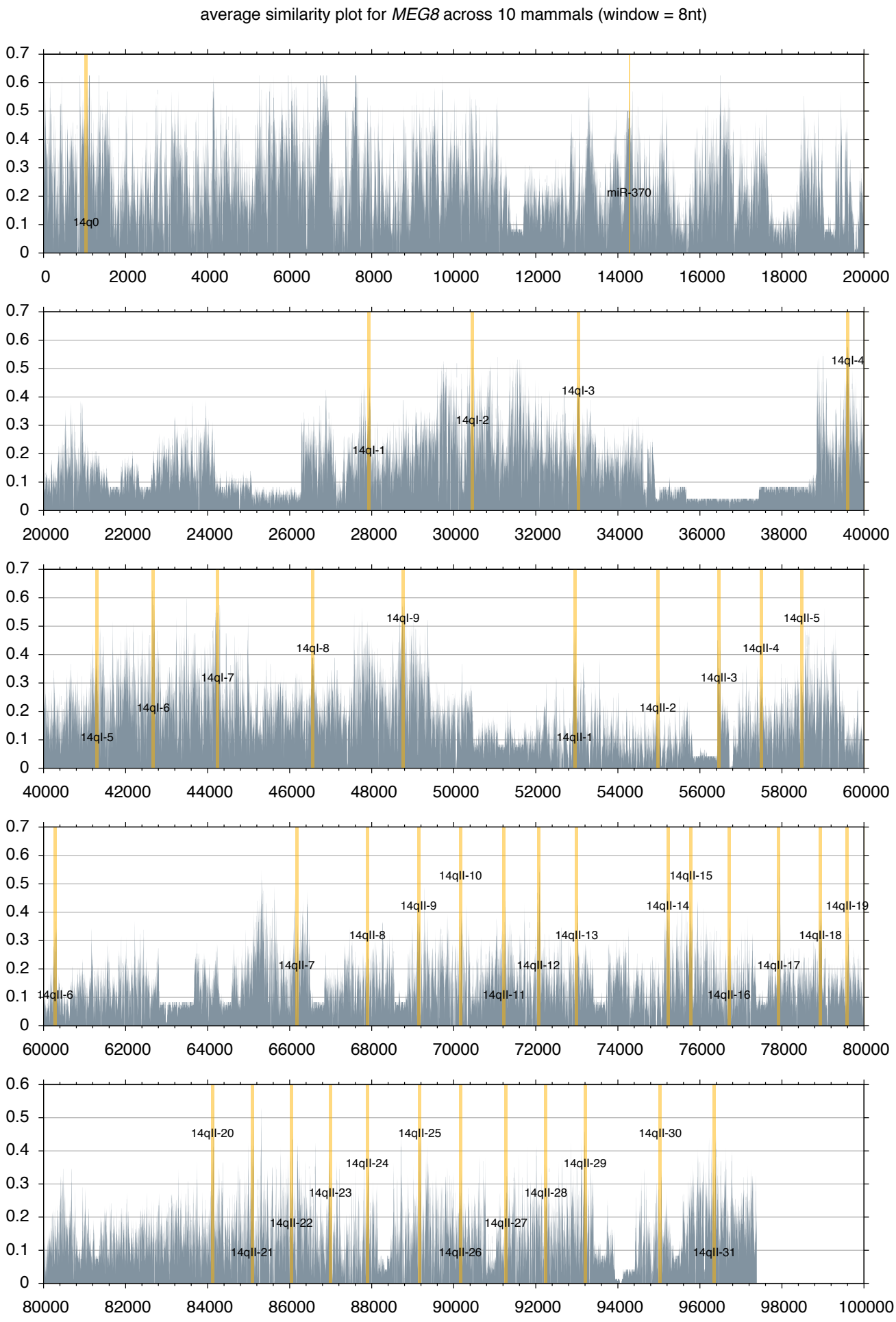

A
C
G
T
1
0.8
0.6
0.4
0.2
0
1
3
5
7
9
11
13
15
17
19
21
5p


1
0.8
0.6
0.4
0.2
0
1
3
5
7
9
11
13
15
17
19
21
3p

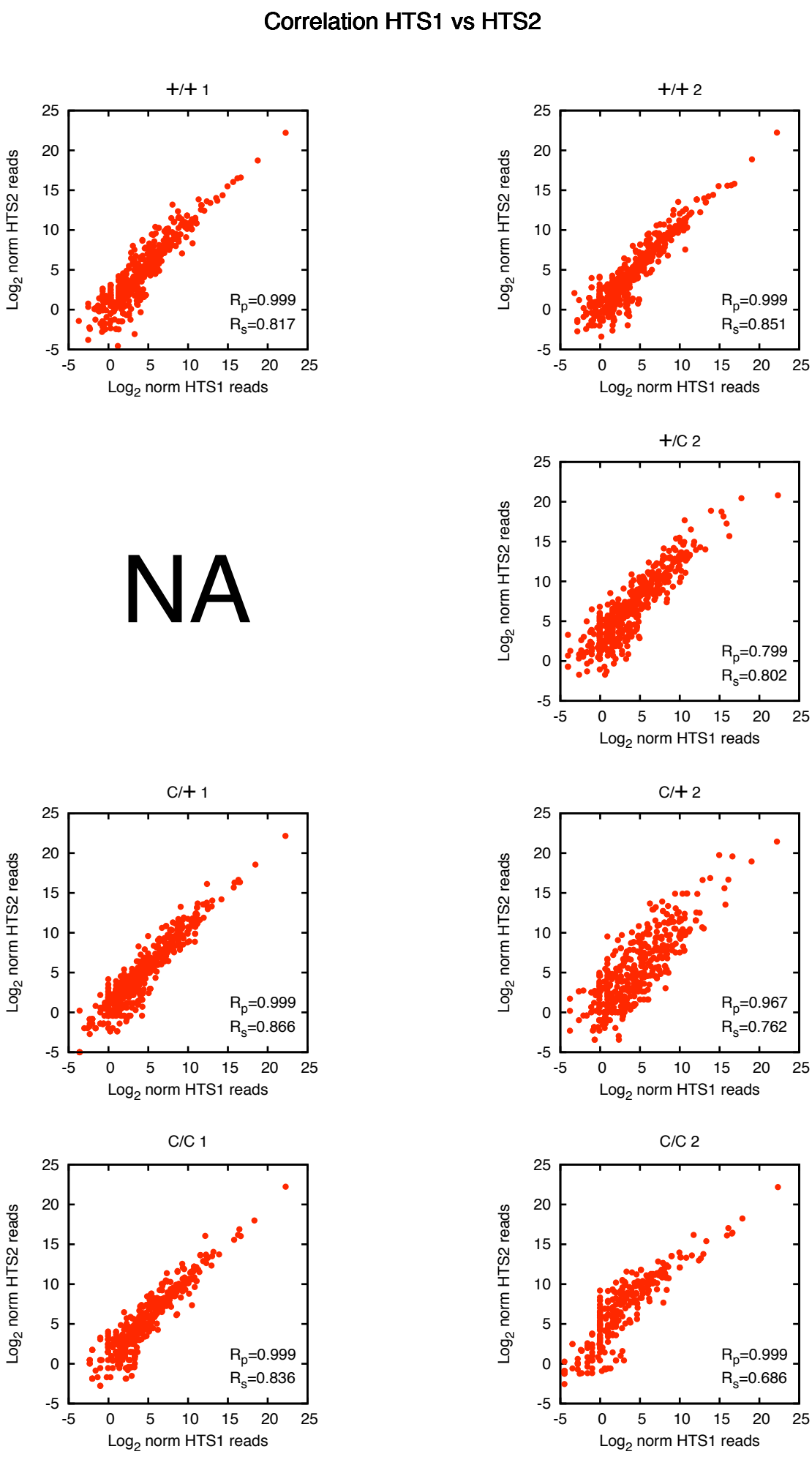
Suppl. Fig. 4A


average similarity plot for *MIRG* across 10 mammals (window = 8nt)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1000 2000 3000 4000 5000 6000 7000 8000
379*
379
411*
411
299-5p
299-3p
380
380*
323-5p
323-3p
1197
758
329
329
494
1193
543
495
8000 9000 10000 11000 12000 13000 14000 15000 16000
376a
376a
376c
376a*
376b
654-3p
654-5p
300
1185
1185
16000 17000 18000 19000 20000 21000 22000 23000 24000
381
487b
539
889
544
655
487a
24000 25000 26000 27000 28000 29000 30000 31000 32000
382
134
668
485-5p
485-3p
453
154*
154
496
32000 33000 34000 35000 36000 37000 38000 39000 40000
377*
377
541
541*
409-3p
409-5p
410
369-3p
369-5p
412
656
40000 41000 42000 43000 44000 45000 46000 47000 48000

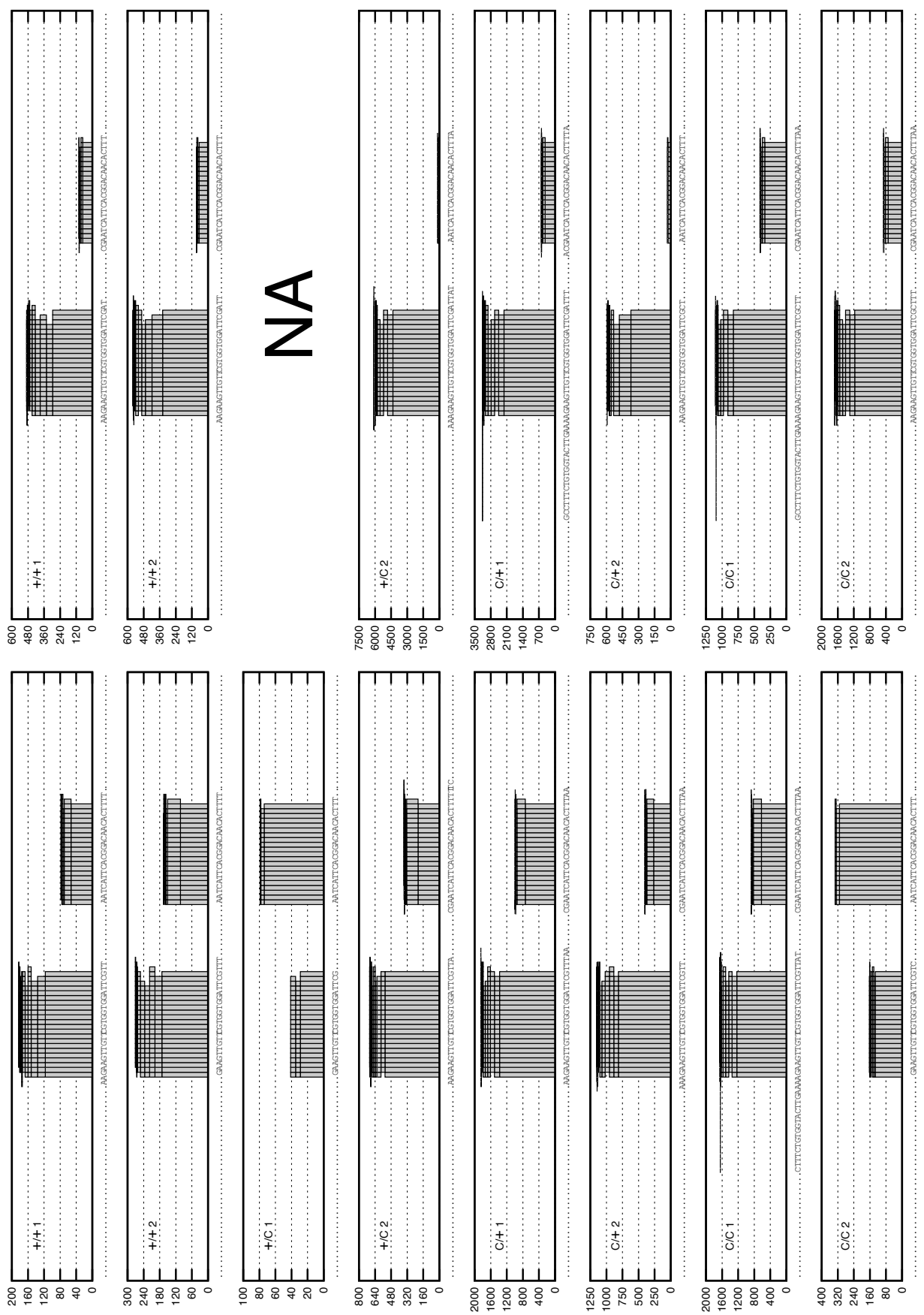
Suppl. Fig. 4B


average similarity plot for *MEG8* across 10 mammals (window = 8nt)
14q0
miR-370
14qI-1
14qI-2
14qI-3
14qI-4
14qI-5
14qI-6
14qI-7
14qI-8
14qI-9
14qII-1
14qII-2
14qII-3
14qII-4
14qII-5
14qII-6
14qII-7
14qII-8
14qII-9
14qII-10
14qII-11
14qII-12
14qII-13
14qII-14
14qII-15
14qII-16
14qII-17
14qII-18
14qII-19
14qII-20
14qII-21
14qII-22
14qII-23
14qII-24
14qII-25
14qII-26
14qII-27
14qII-28
14qII-29
14qII-30
14qII-31

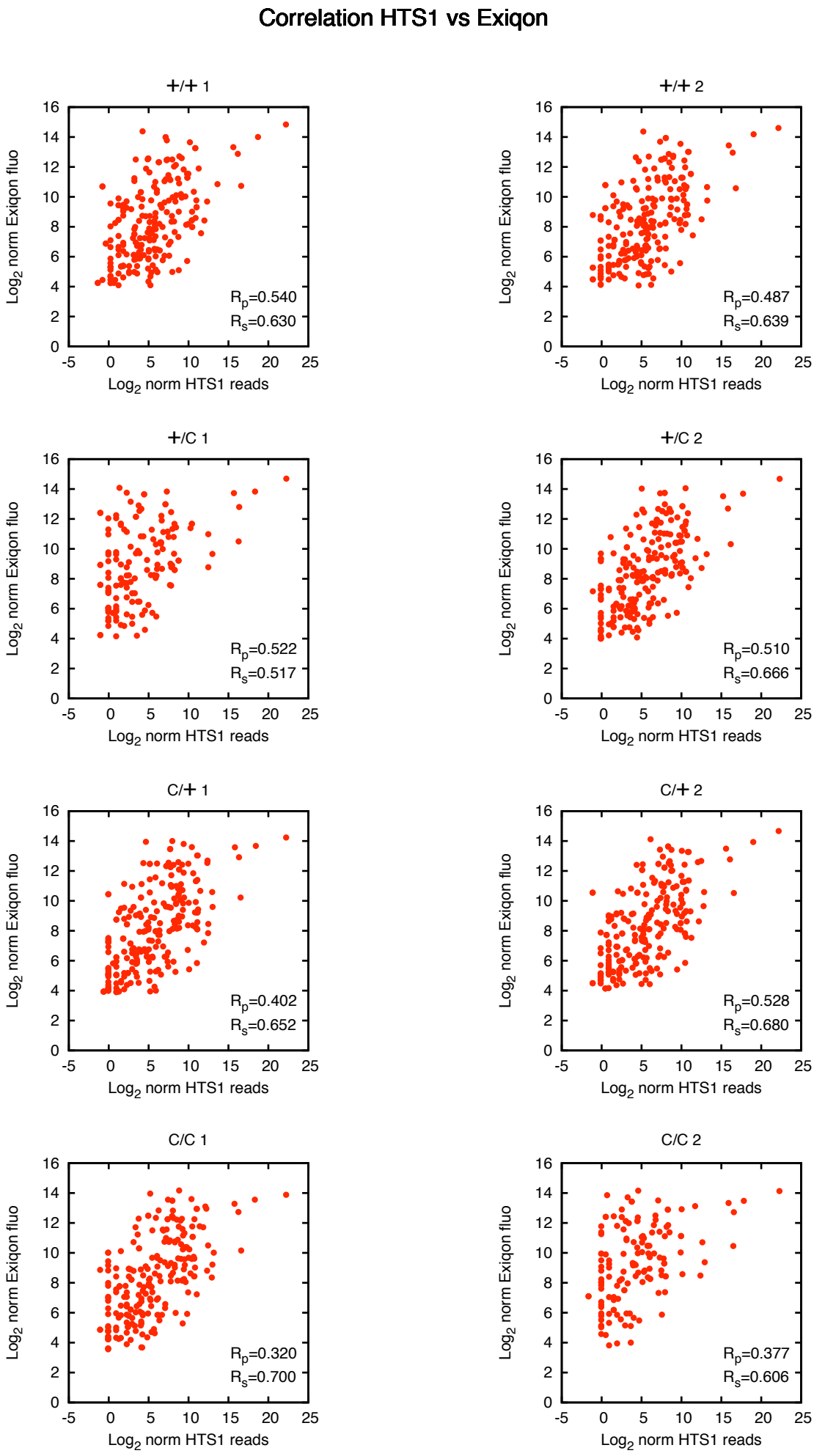
Suppl. Fig. 5


Correlation HTS1 vs HTS2
+/+ 1
Log2 norm HTS2 reads
Log2 norm HTS1 reads
Rp=0.999
Rs=0.817
+/+ 2
Rp=0.999
Rs=0.851
NA
+/C 2
Rp=0.799
Rs=0.802
C/+ 1
Rp=0.999
Rs=0.866
C/+ 2
Rp=0.967
Rs=0.762
C/C 1
Rp=0.999
Rs=0.836
C/C 2
Rp=0.999
Rs=0.686

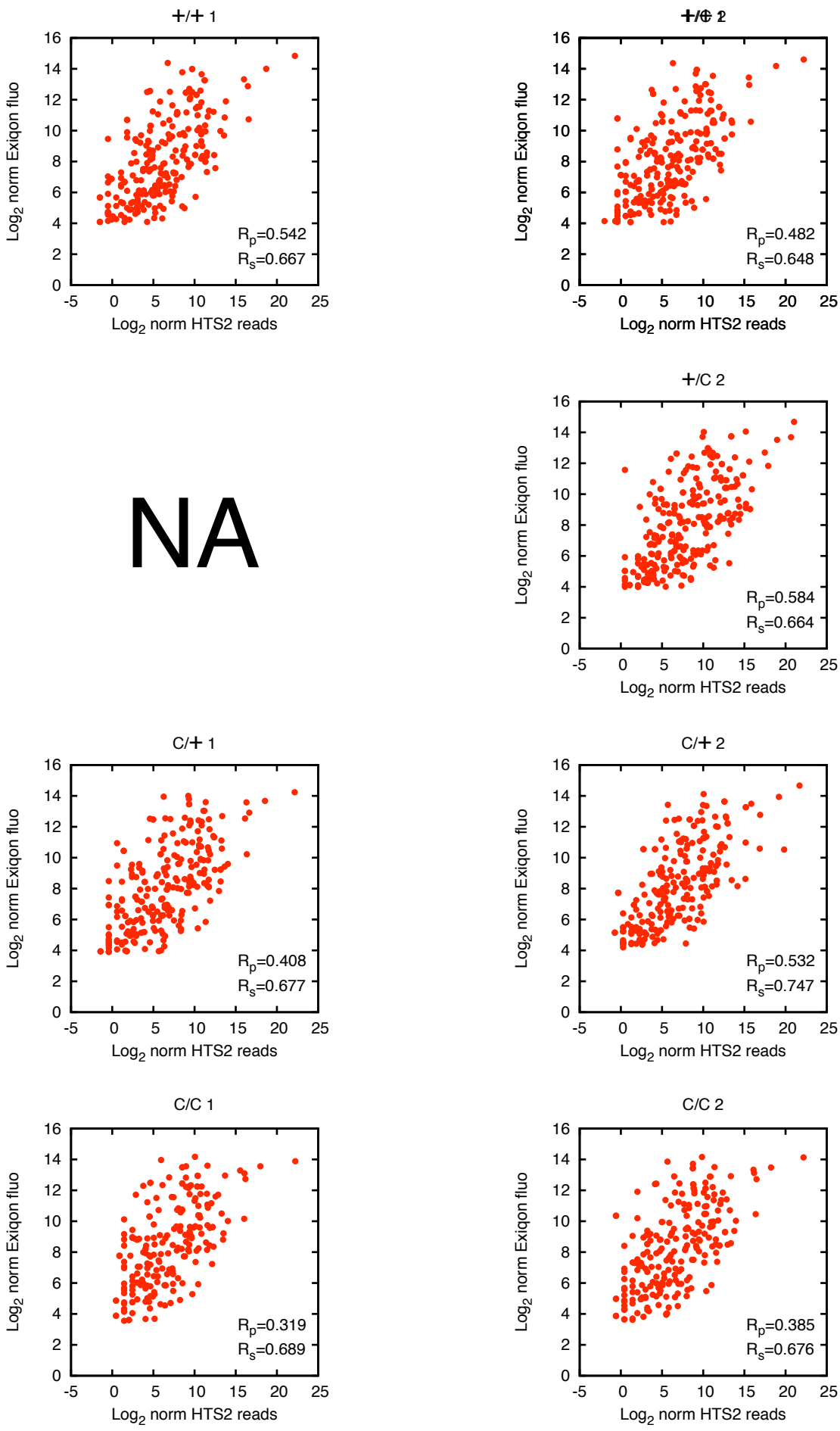
Suppl. Fig. 6



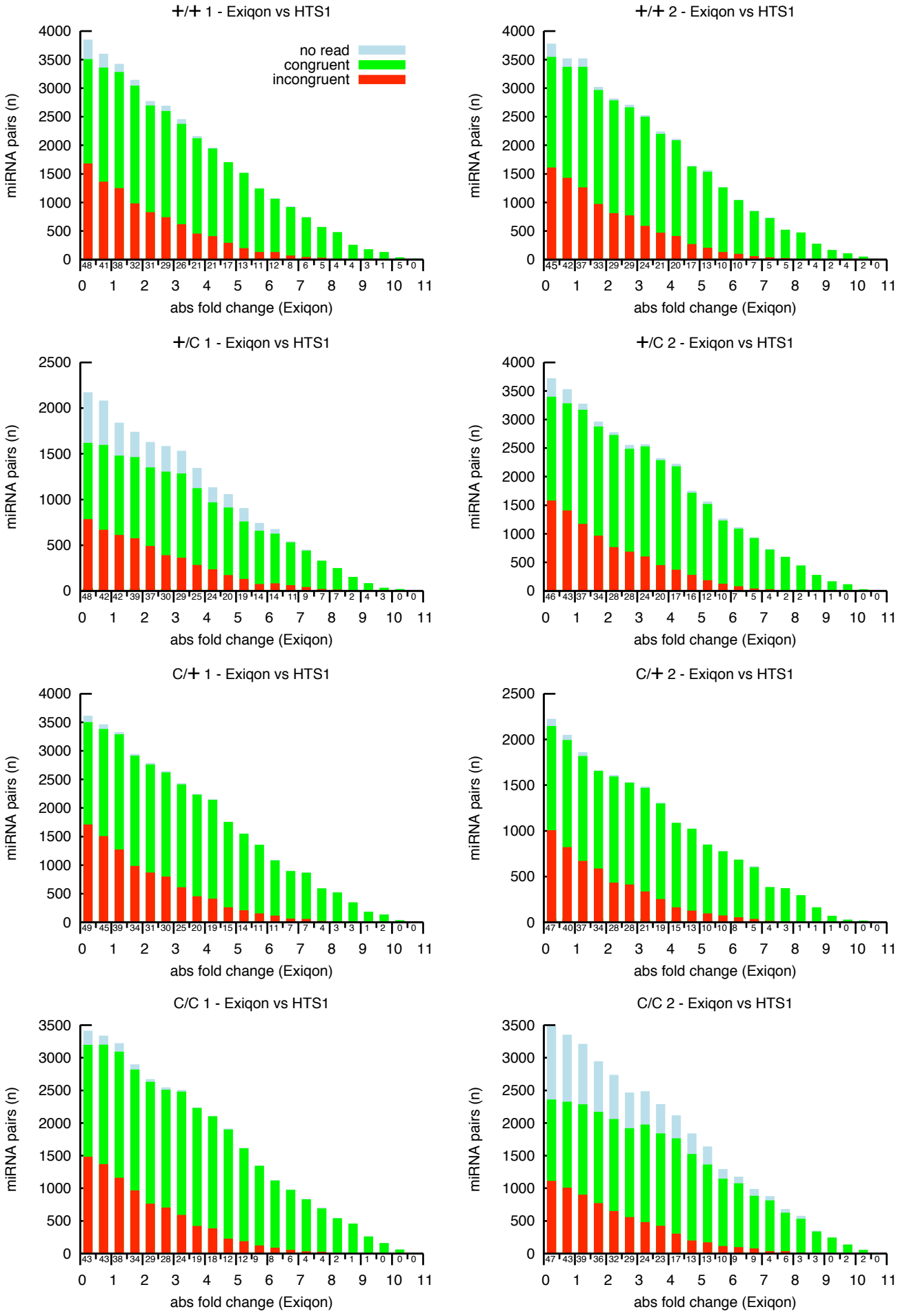
Suppl. Fig. 7A


Correlation HTS1 vs Exiqon
+/+ 1
Rp=0.540
Rs=0.630
+/+ 2
Rp=0.487
Rs=0.639
+/C 1
Rp=0.522
Rs=0.517
+/C 2
Rp=0.510
Rs=0.666
C/+ 1
Rp=0.402
Rs=0.652
C/+ 2
Rp=0.528
Rs=0.680
C/C 1
Rp=0.320
Rs=0.700
C/C 2
Rp=0.377
Rs=0.606
Log2 norm Exiqon fluo
Log2 norm HTS1 reads

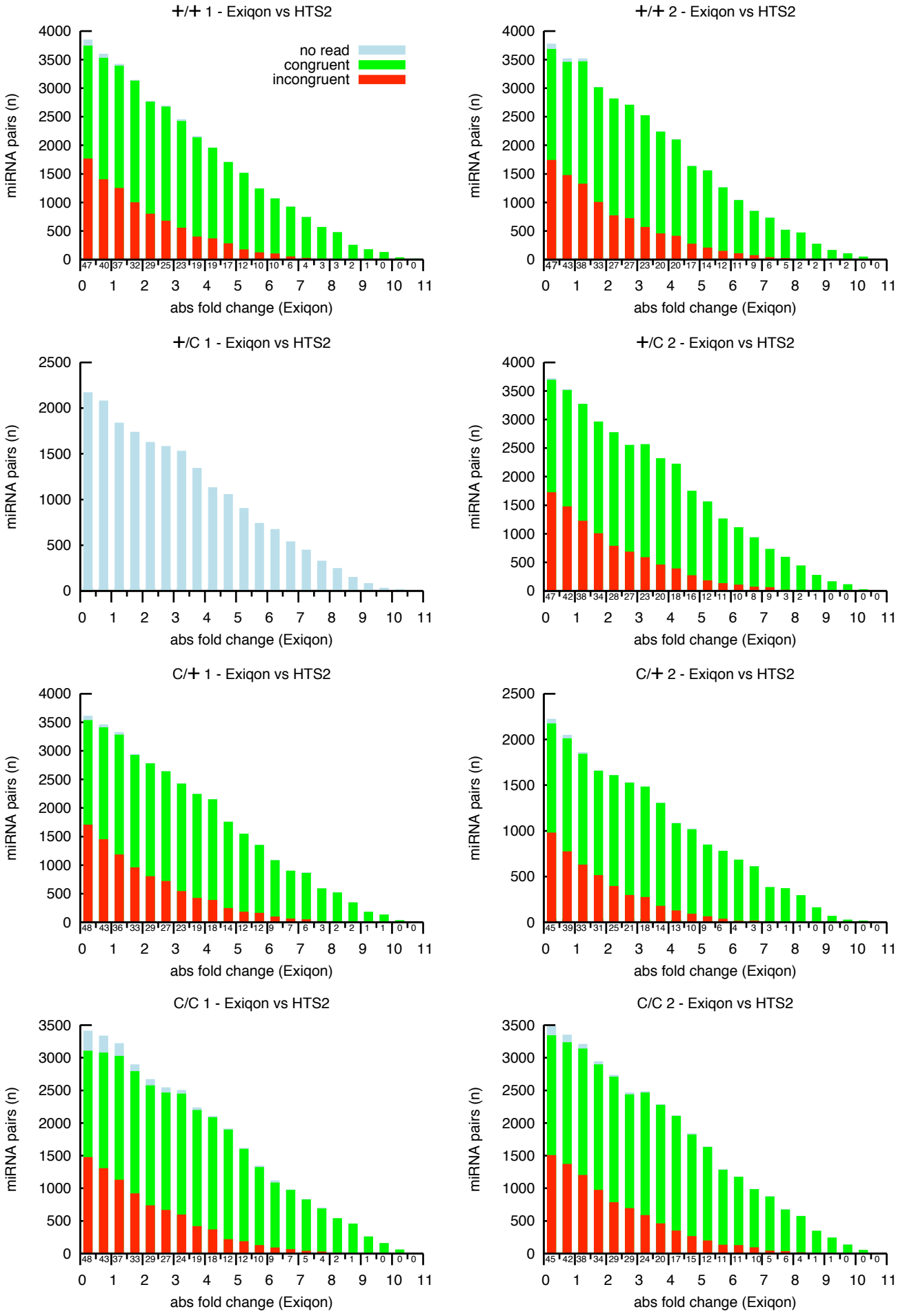
Suppl. Fig. 7B


Correlation HTS2 vs Exiqon
+/+ 1
Rp=0.542
Rs=0.667
+/+ 2
Rp=0.482
Rs=0.648
NA
+/C 2
Rp=0.584
Rs=0.664
C/+ 1
Rp=0.408
Rs=0.677
C/+ 2
Rp=0.532
Rs=0.747
C/C 1
Rp=0.319
Rs=0.689
C/C 2
Rp=0.385
Rs=0.676
Log2 norm Exiqon fluo
Log2 norm HTS2 reads

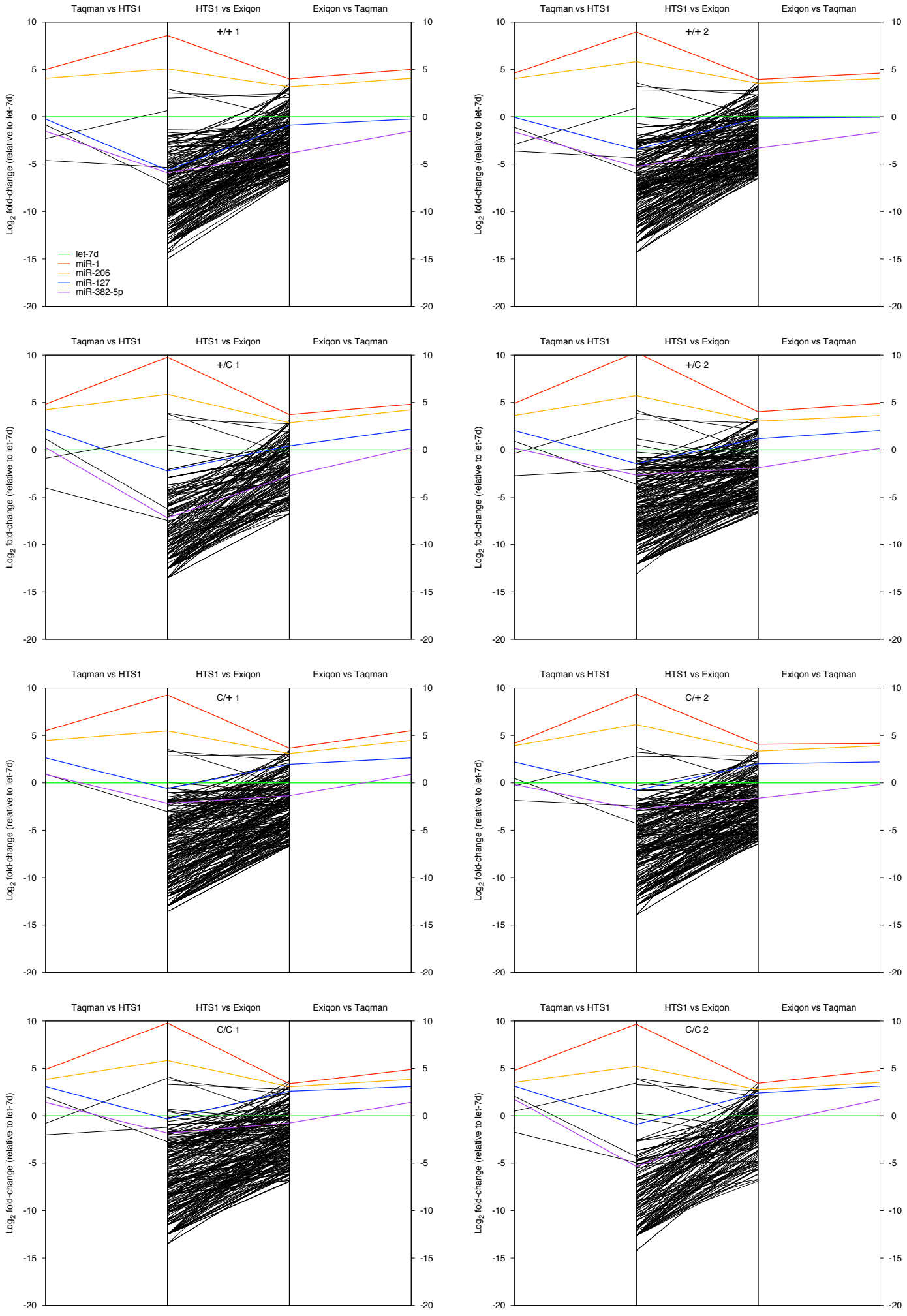
Suppl. Fig. 7C


+/+ 1 - Exiqon vs HTS1
+/+ 2 - Exiqon vs HTS1
+/C 1 - Exiqon vs HTS1
+/C 2 - Exiqon vs HTS1
C/+ 1 - Exiqon vs HTS1
C/+ 2 - Exiqon vs HTS1
C/C 1 - Exiqon vs HTS1
C/C 2 - Exiqon vs HTS1
no read
congruent
incongruent
miRNA pairs (n)
abs fold change (Exiqon)

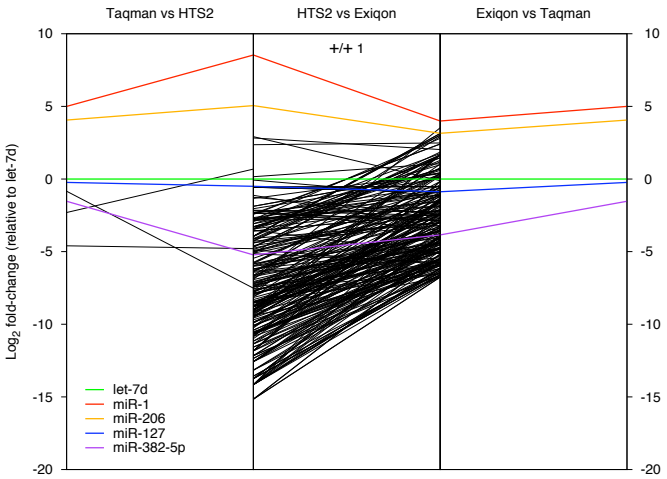
Suppl. Fig. 7D

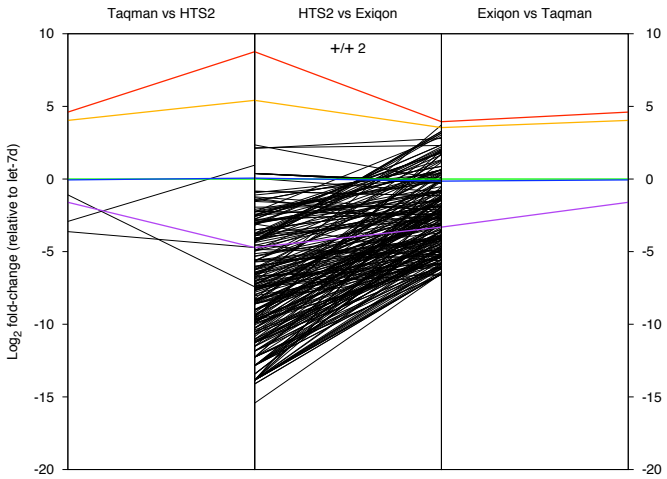

+/+ 1 - Exiqon vs HTS2
+/+ 2 - Exiqon vs HTS2
+/C 1 - Exiqon vs HTS2
+/C 2 - Exiqon vs HTS2
C/+ 1 - Exiqon vs HTS2
C/+ 2 - Exiqon vs HTS2
C/C 1 - Exiqon vs HTS2
C/C 2 - Exiqon vs HTS2
no read
congruent
incongruent
miRNA pairs (n)
abs fold change (Exiqon)

Suppl. Fig. 8A

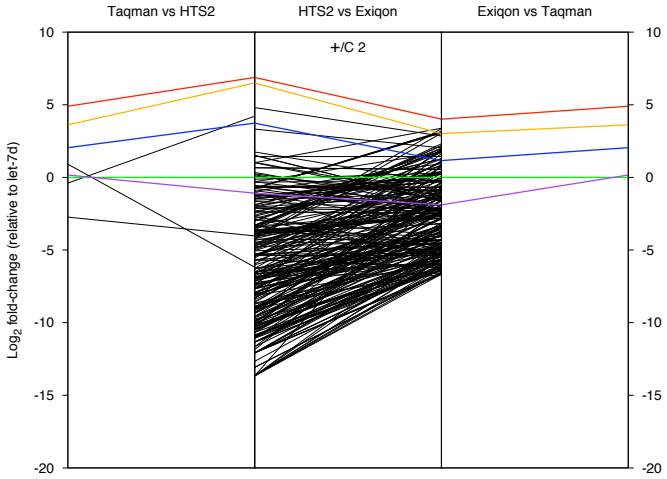

Taqman vs HTS1
HTS1 vs Exiqon
Exiqon vs Taqman
+/+ 1
+/+ 2
+/C 1
+/C 2
C/+ 1
C/+ 2
C/C 1
C/C 2
Log2 fold-change (relative to let-7d)
let-7d
miR-1
miR-206
miR-127
miR-382-5p
10
5
0
-5
-10
-15
-20

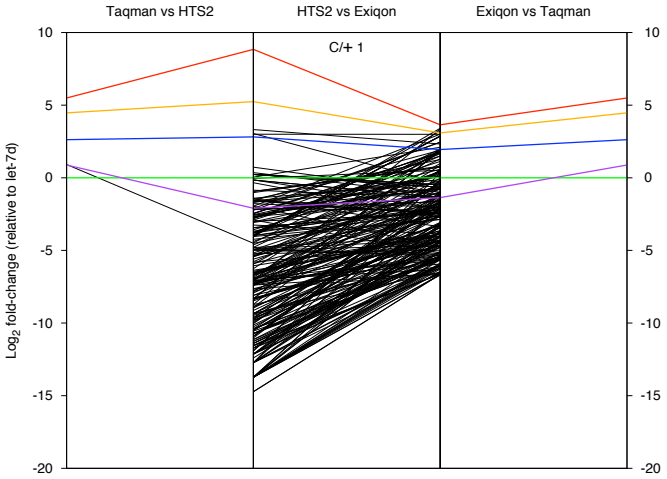
Suppl. Fig. 8B

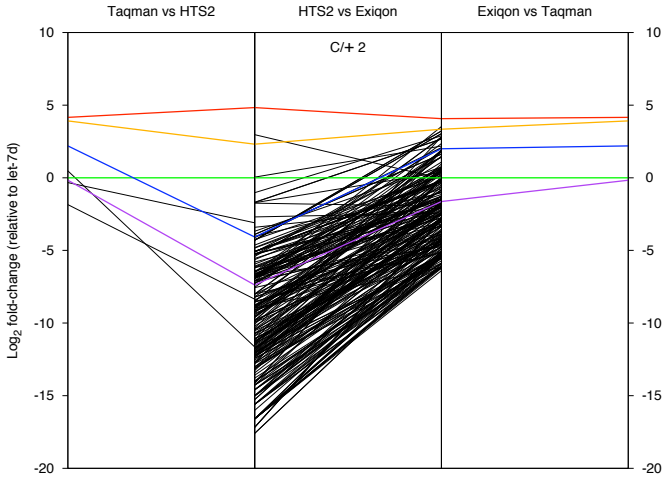


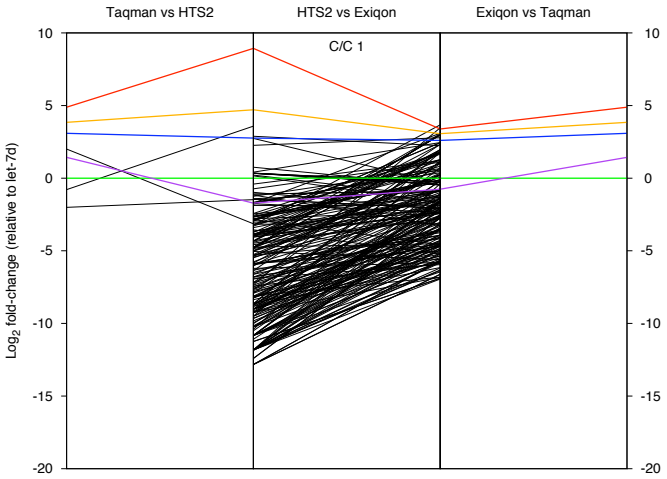


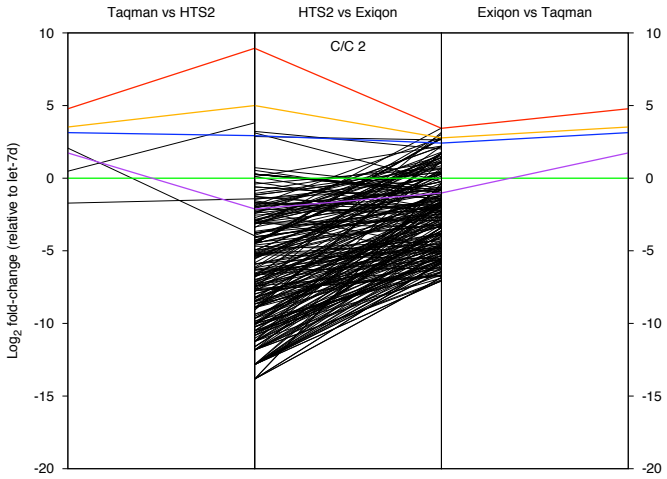
NA



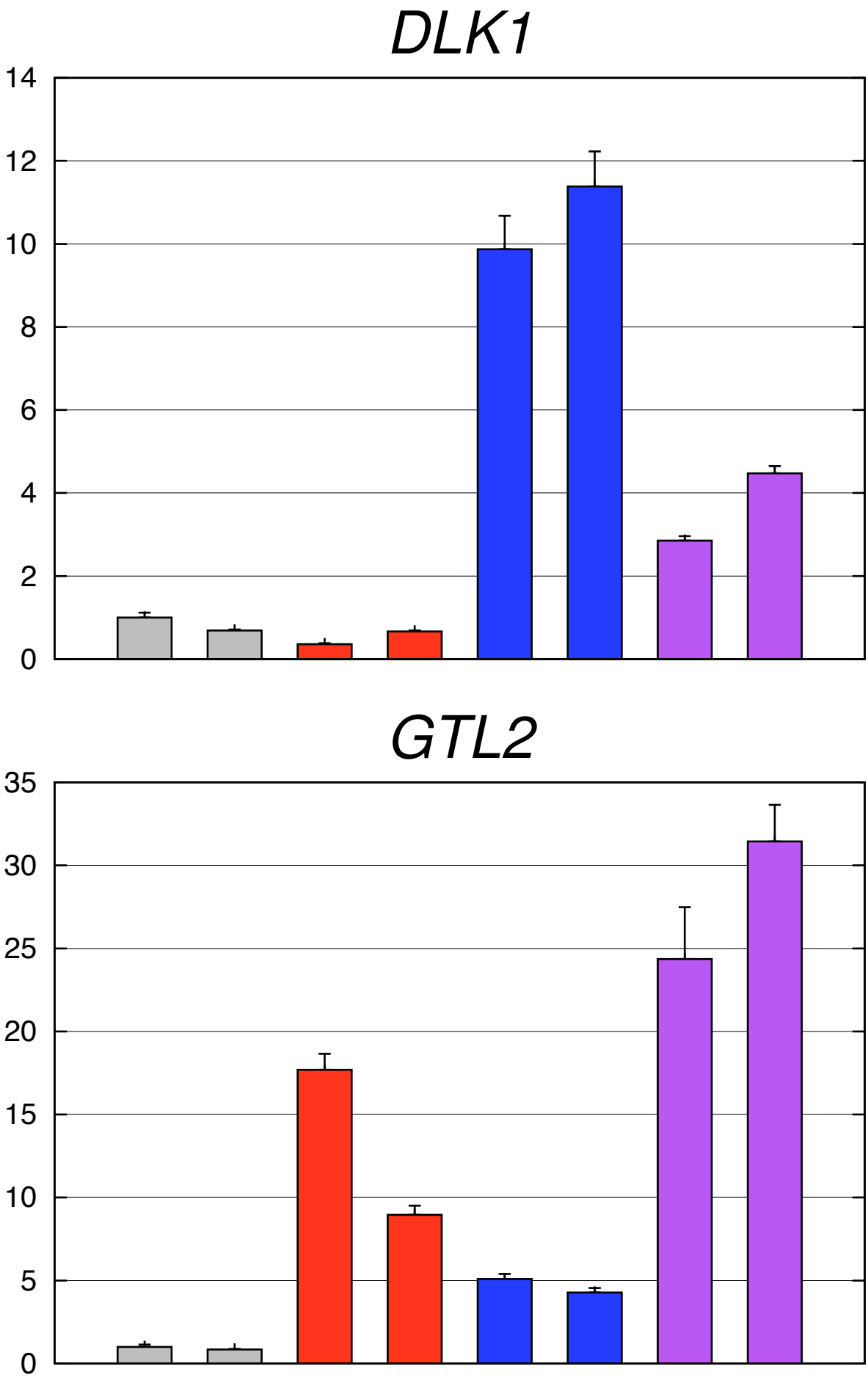








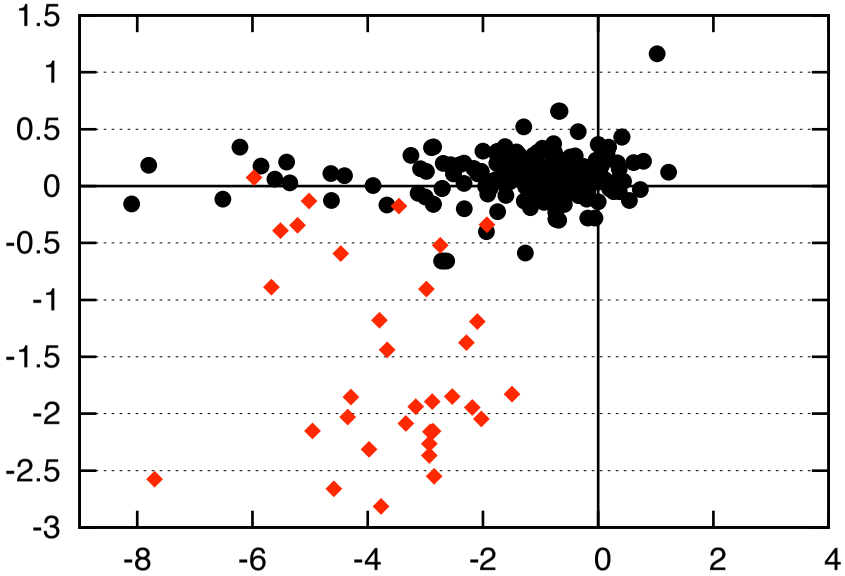
Suppl. Fig. 9

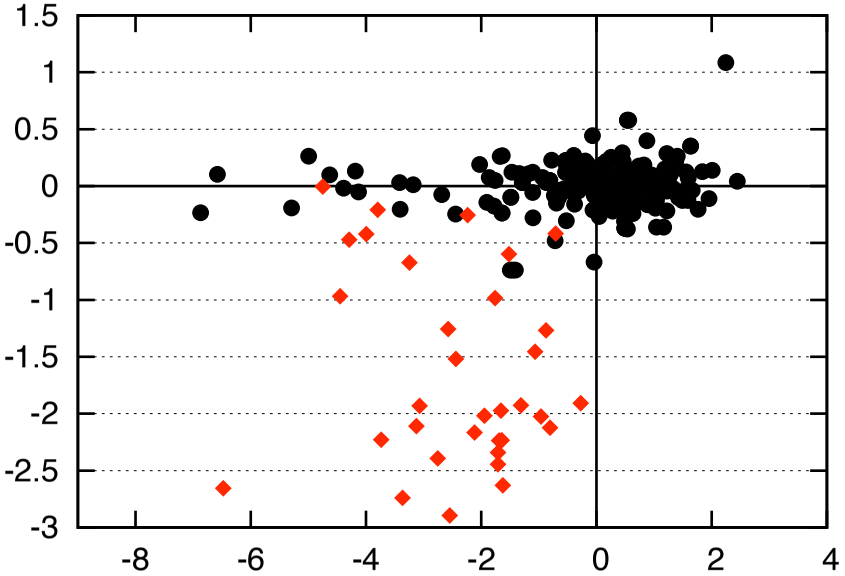


BEFORE CORRECTION

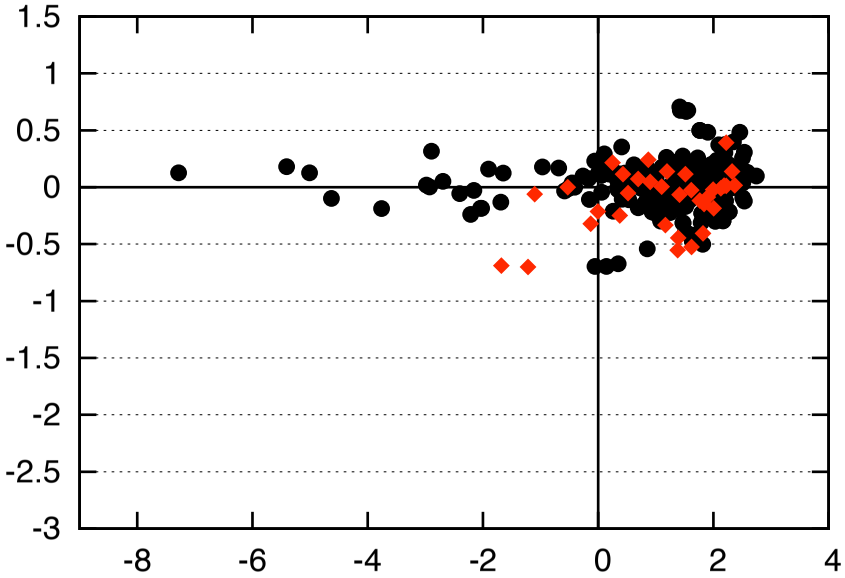
AFTER CORRECTION

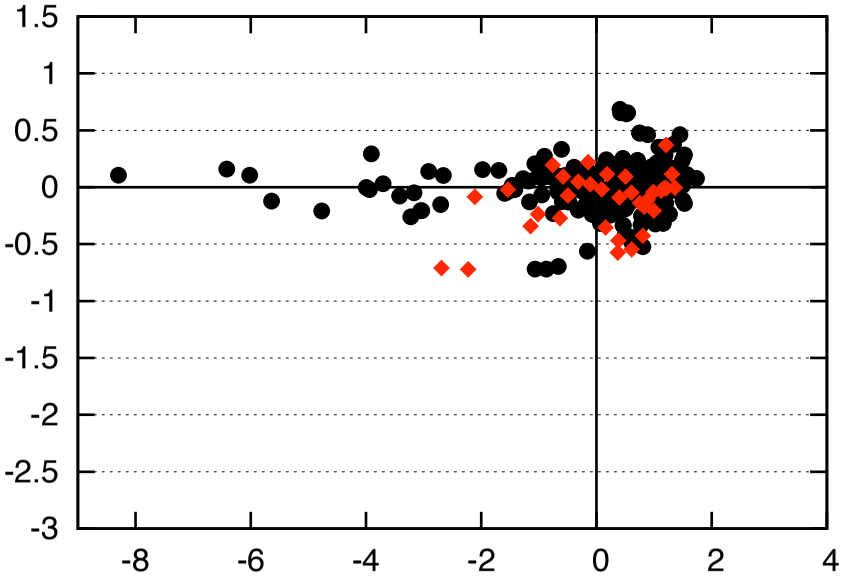
A: +/+



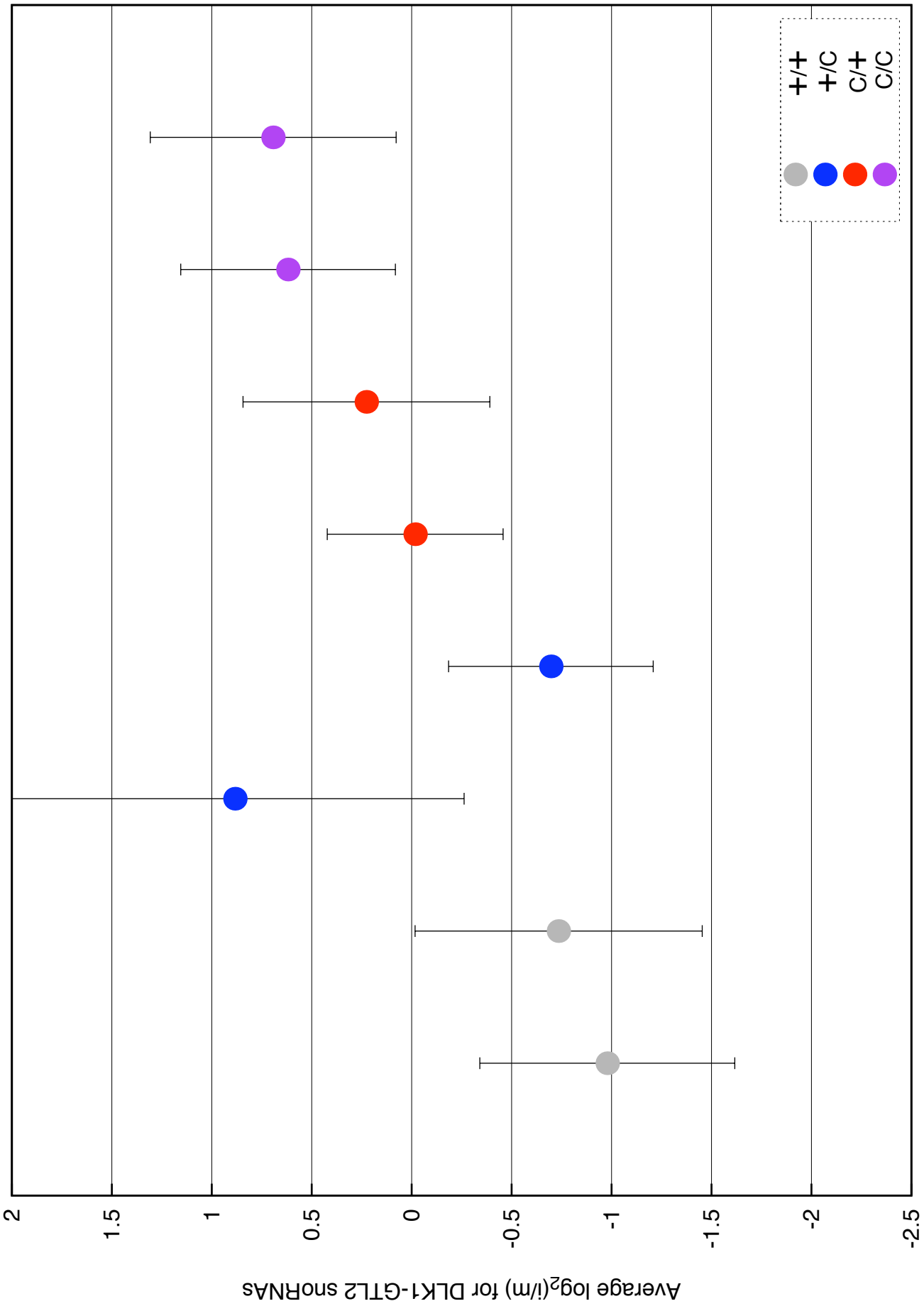


B : CLPG/CLPG

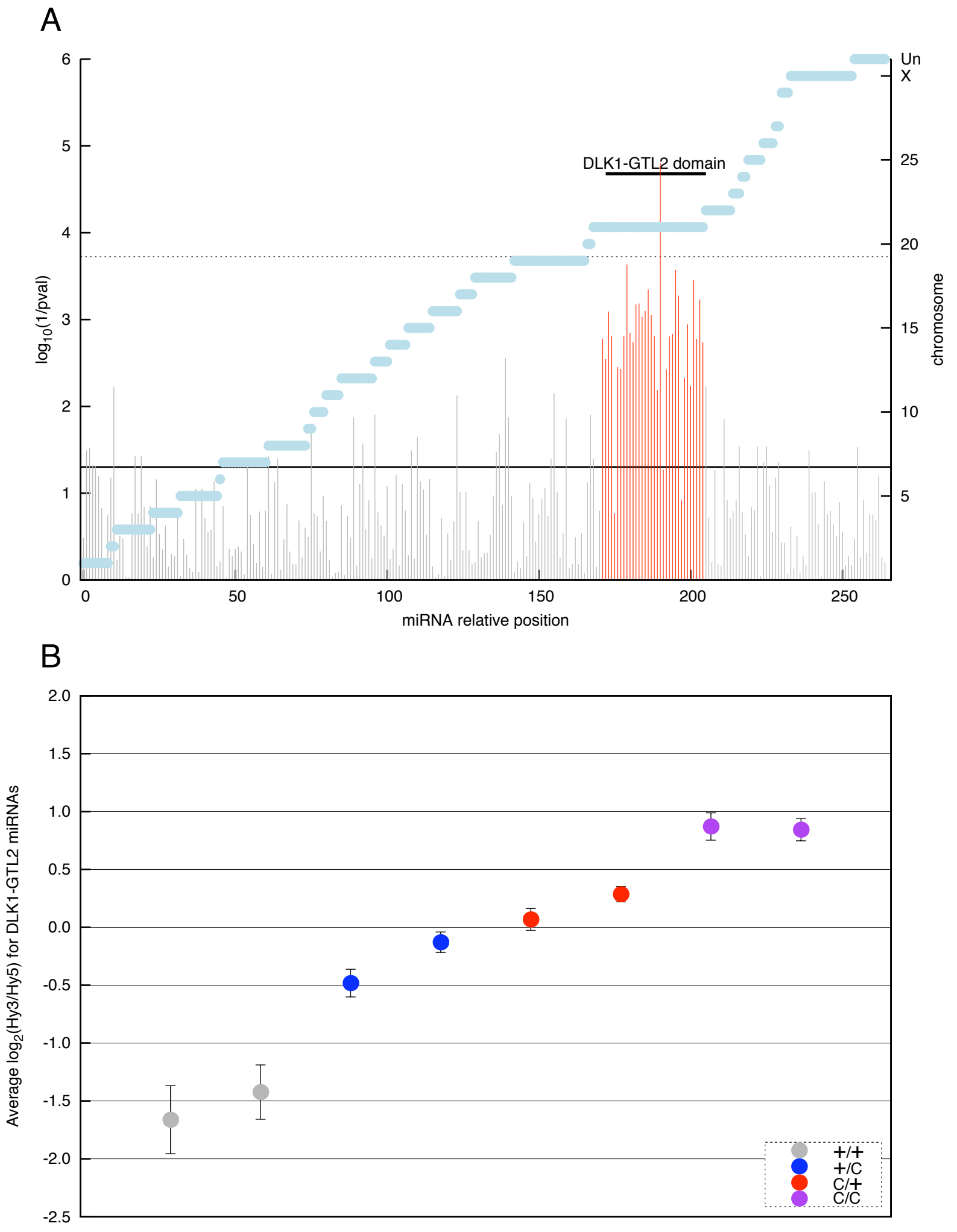




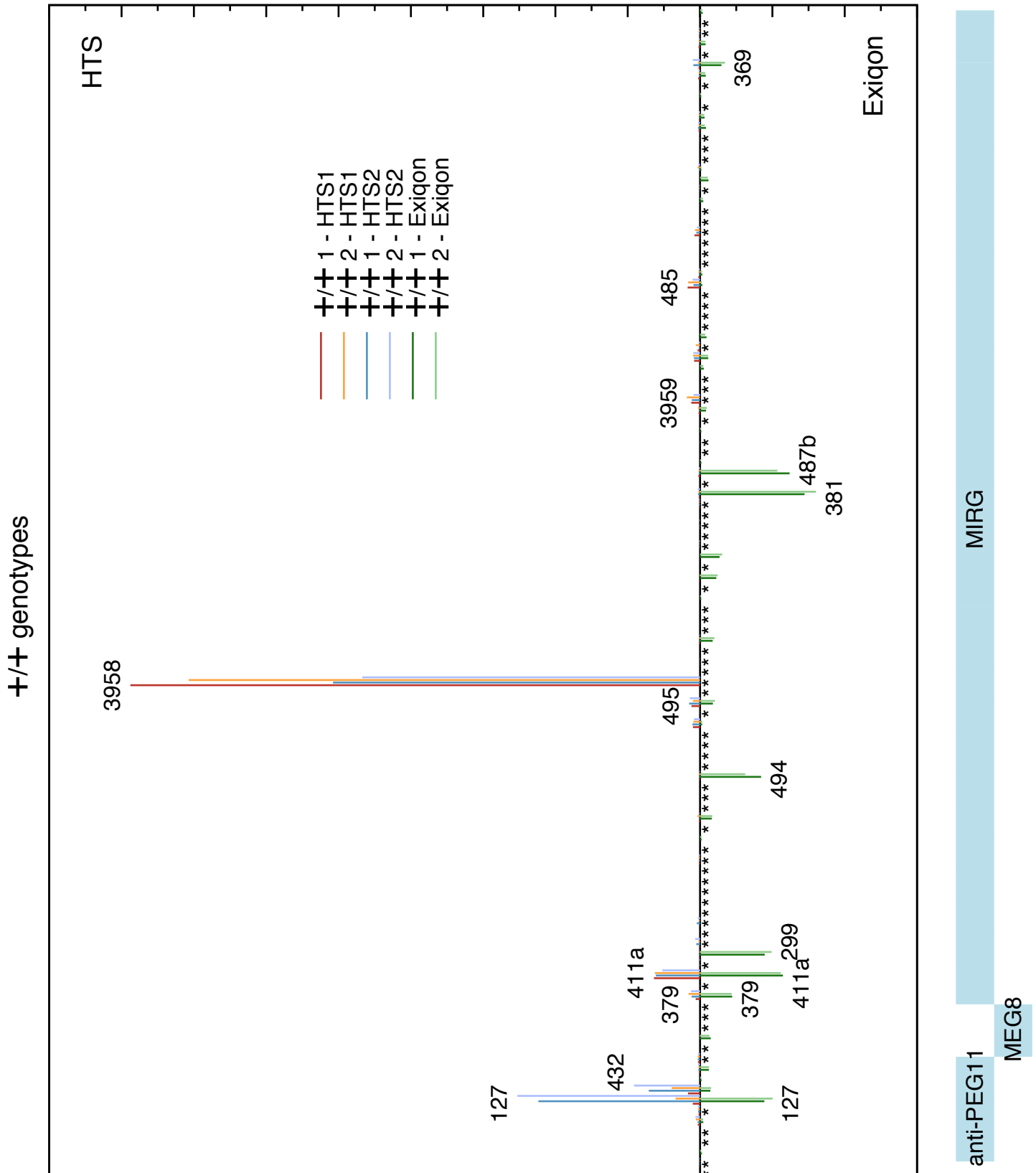
Suppl. Fig. 11


Average log2(i/m) for DLK1-GTL2 snoRNAs
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
+/+
+/C
C/+
C/C

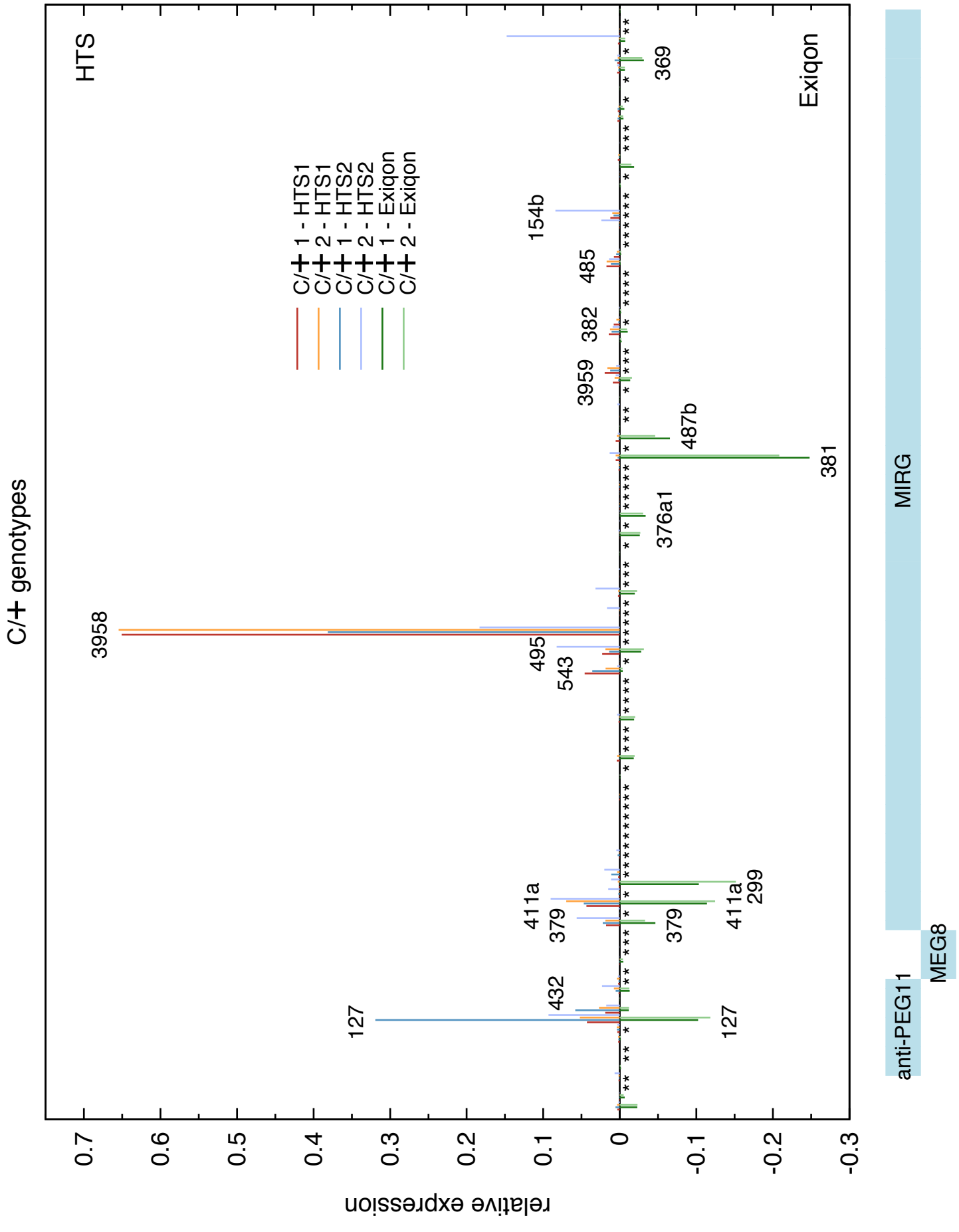
Suppl. Fig. 12


A
DLK1-GTL2 domain
log10(1/pval)
chromosome
miRNA relative position
0
50
100
150
200
250
Un
X
B
Average log2(Hy3/Hy5) for DLK1-GTL2 miRNAs
+/+
+/C
C/+
C/C

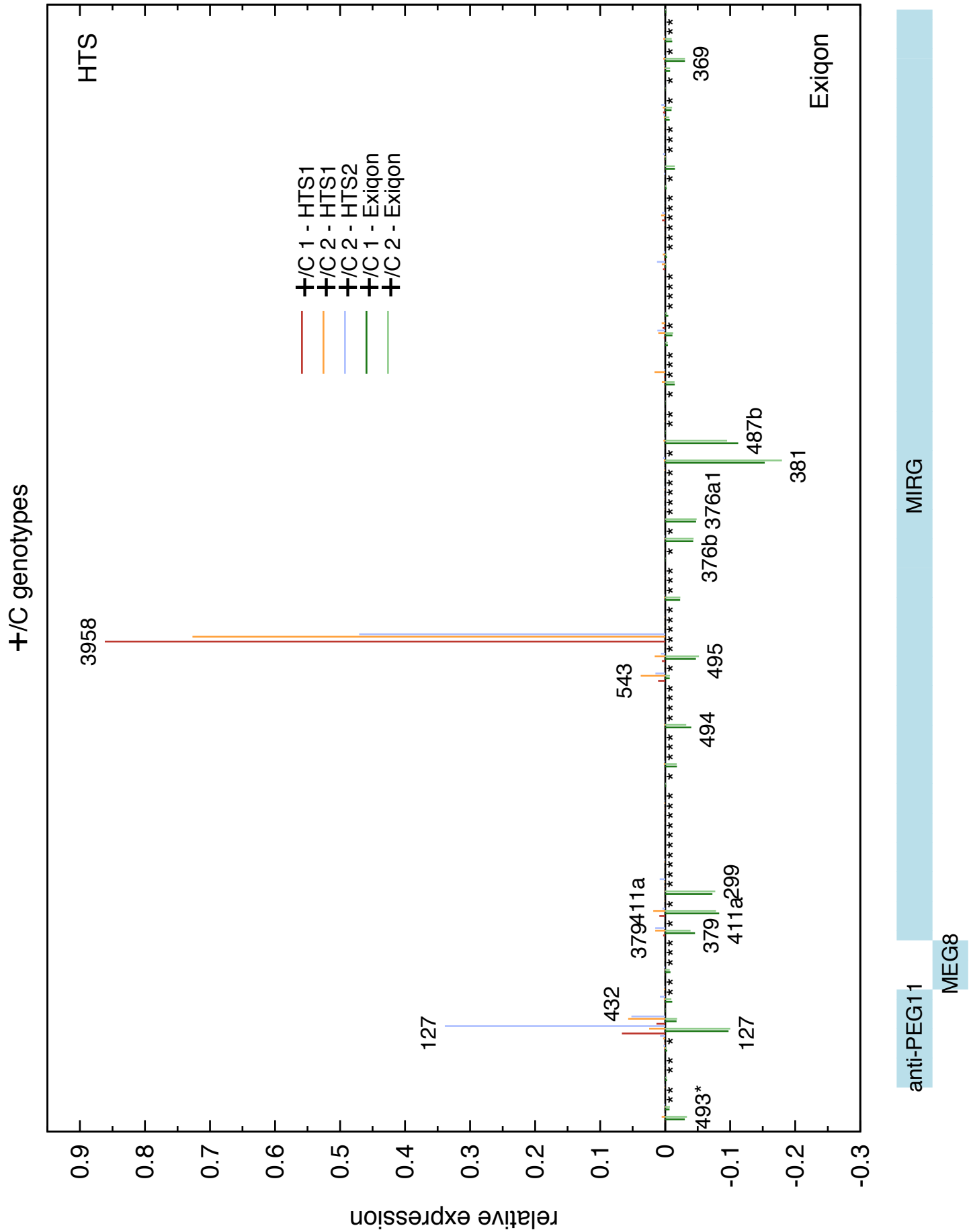
Suppl. Fig. 13A



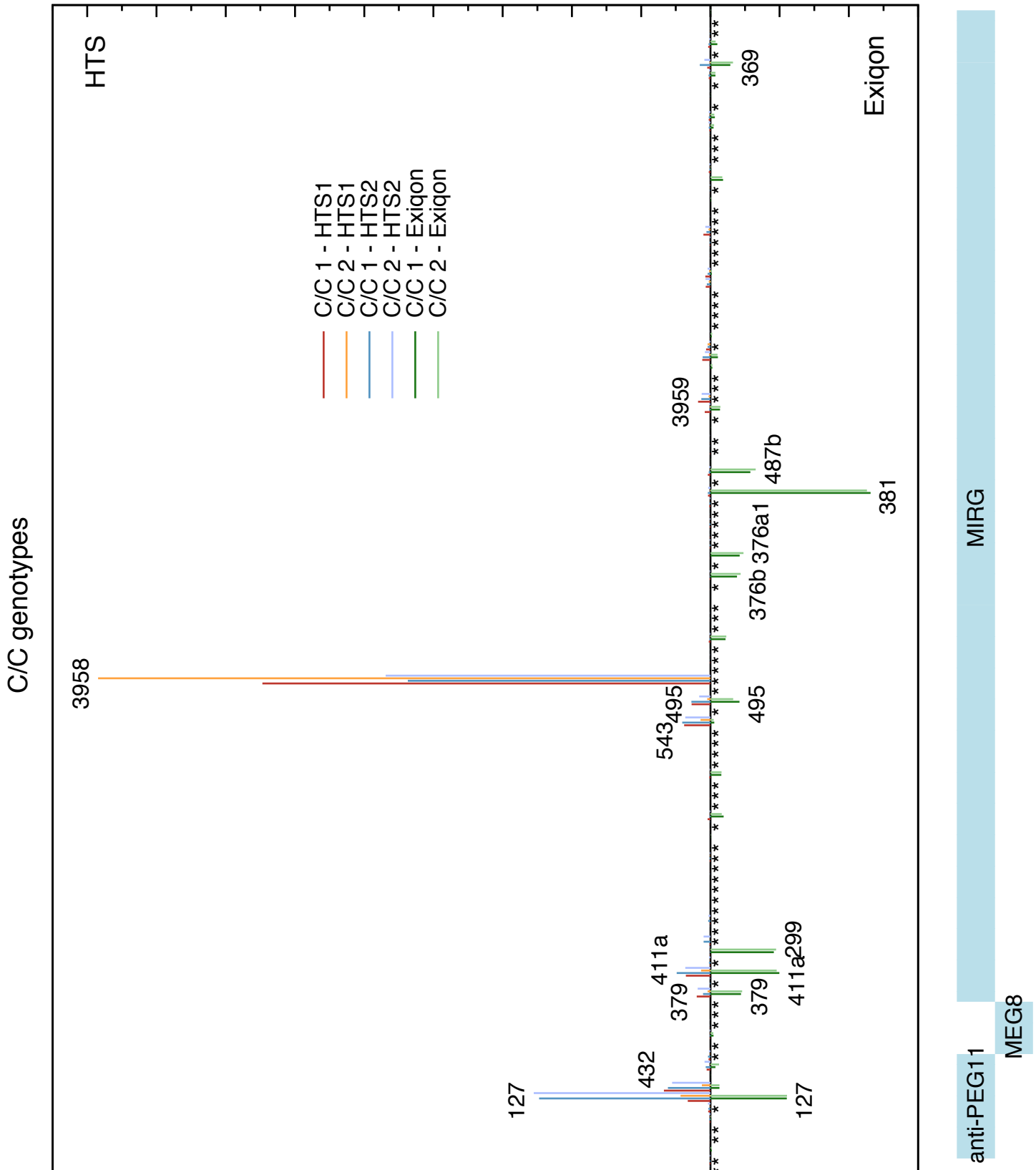
Suppl. Fig. 13B



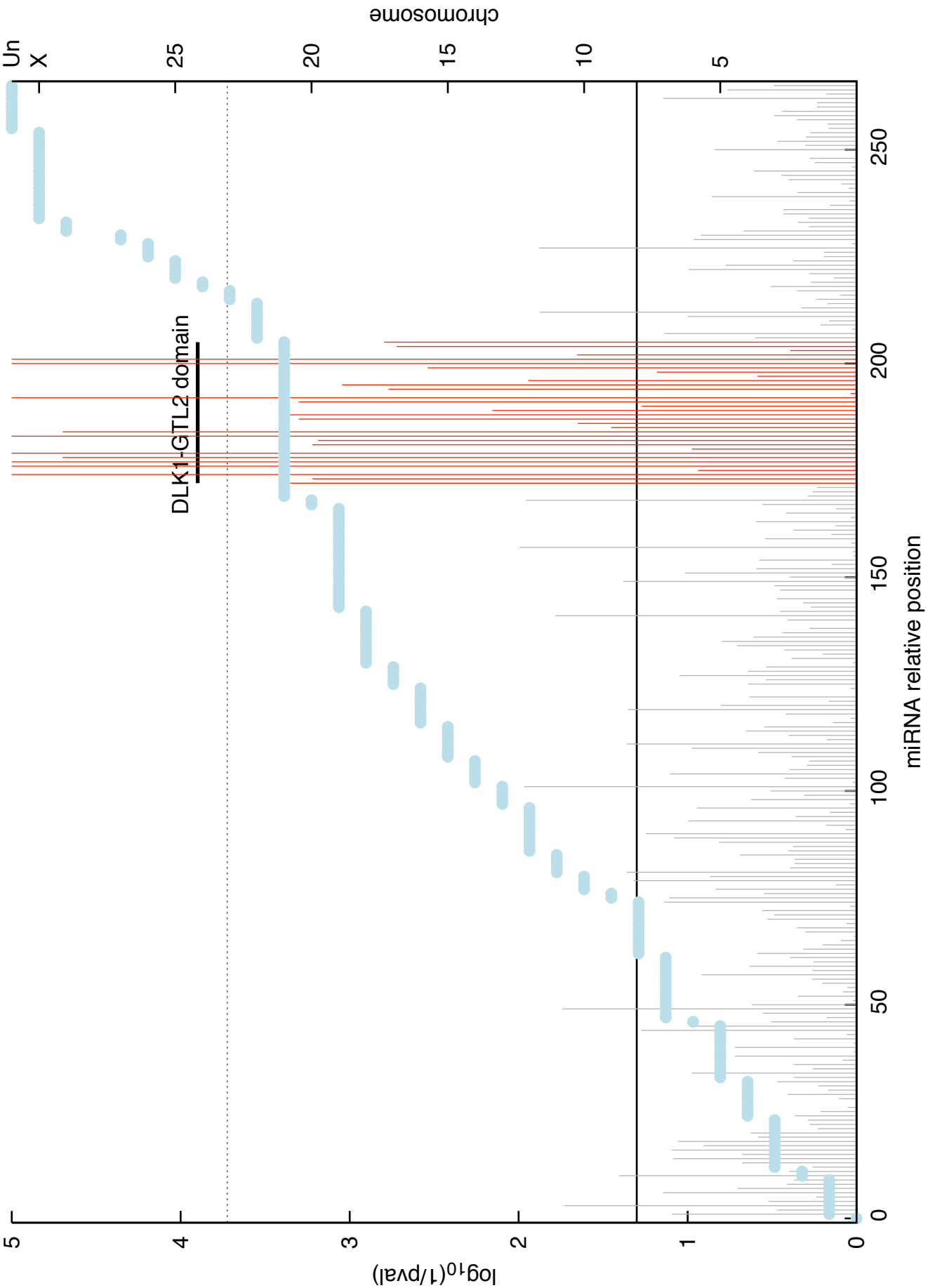
Suppl. Fig. 13C



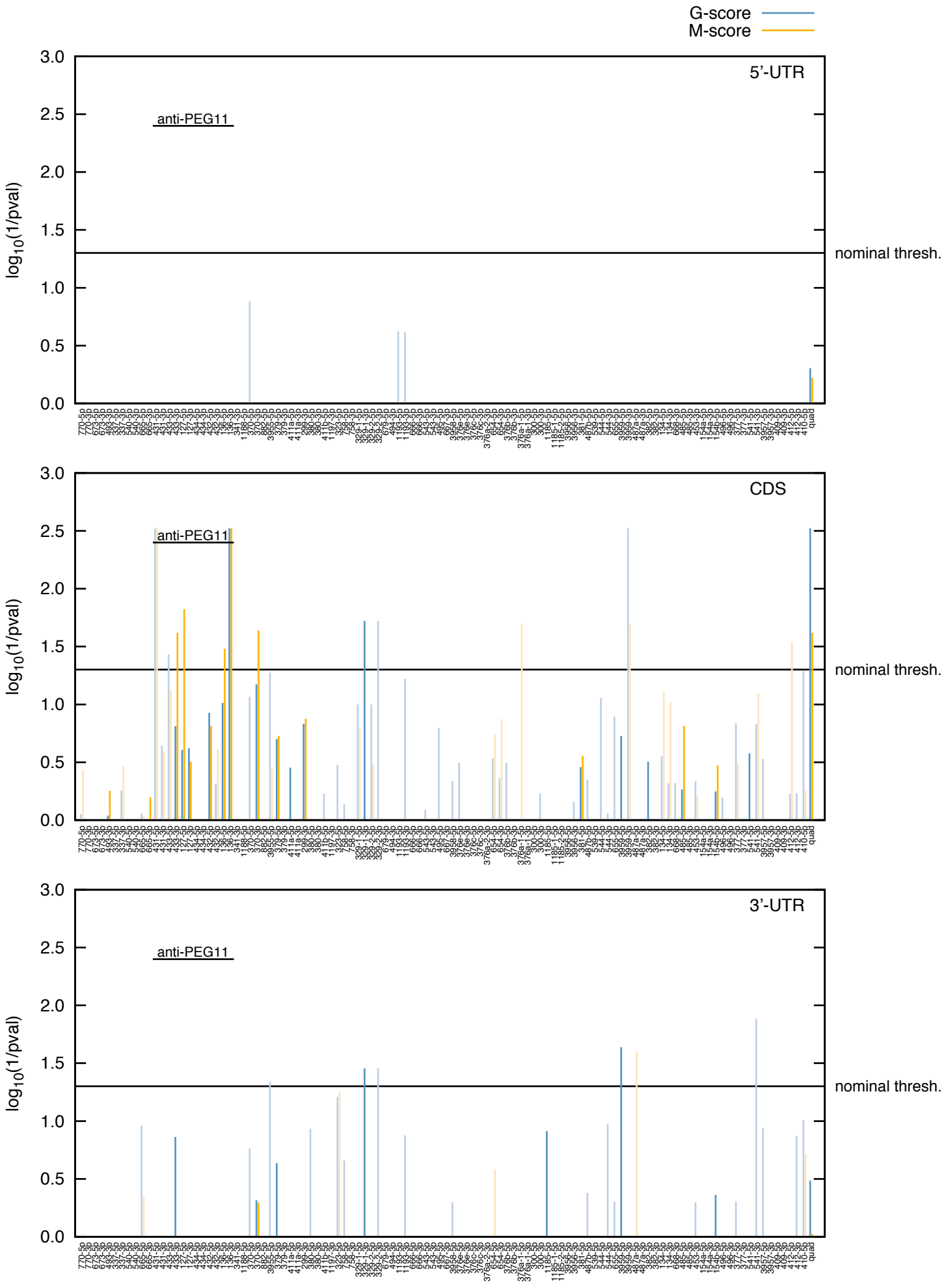
Suppl. Fig. 13D



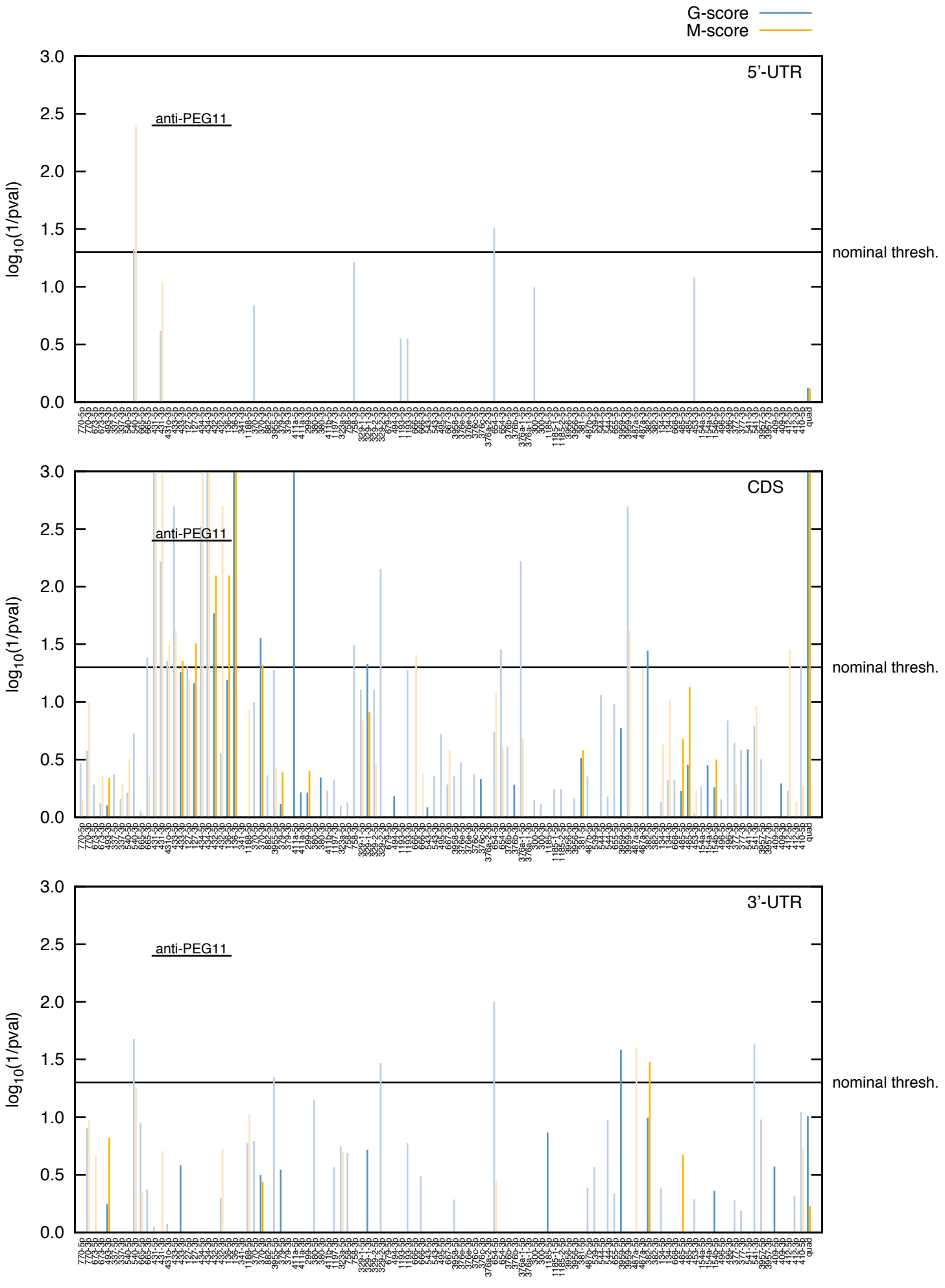
Suppl. Fig. 14


chromosome
Un
X
25
20
15
10
5
DLK1-GTL2 domain
log10(1/pval)
0
1
2
3
4
5
miRNA relative position
0
50
100
150
200
250

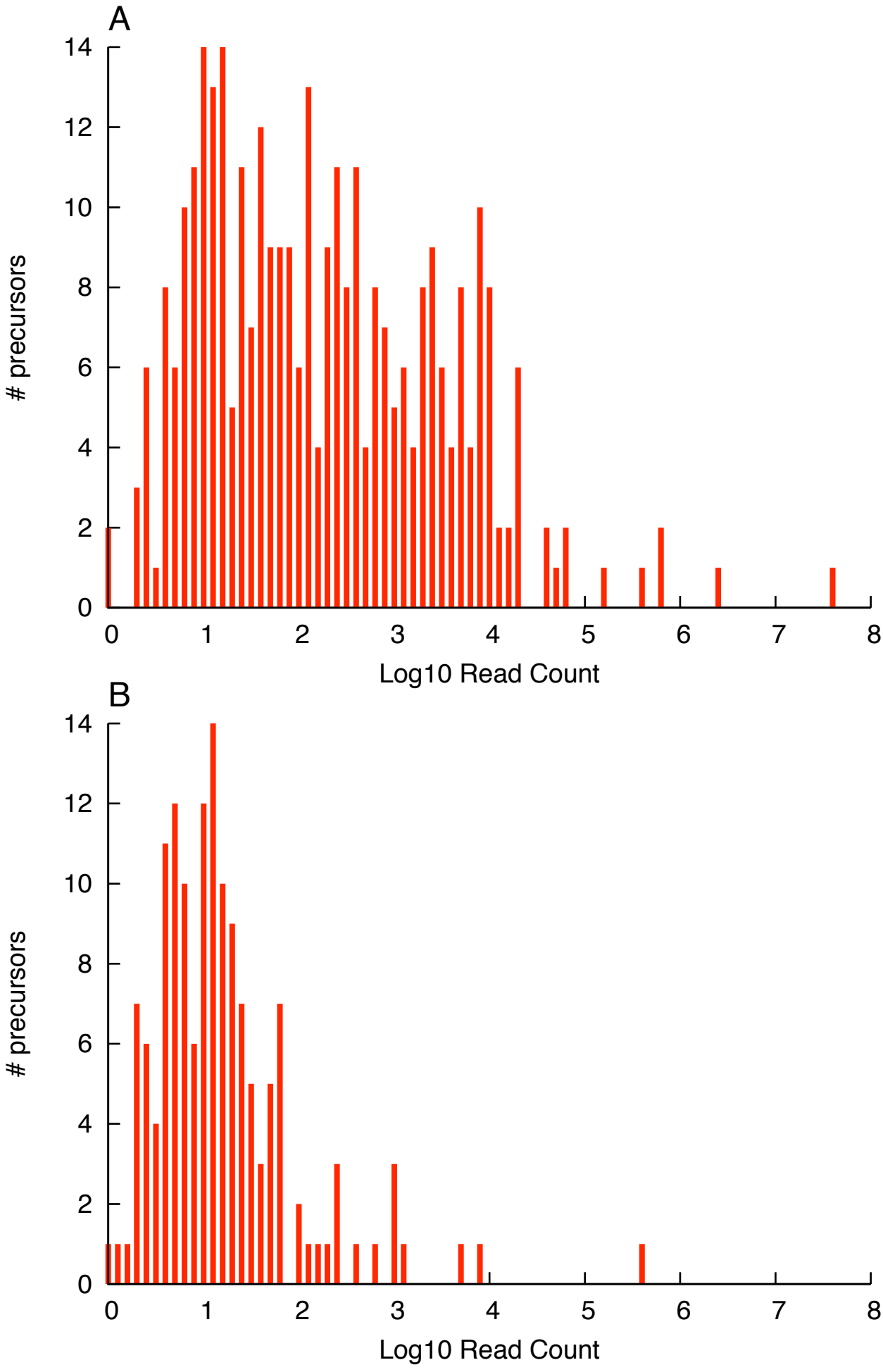
Suppl. Fig. 15A



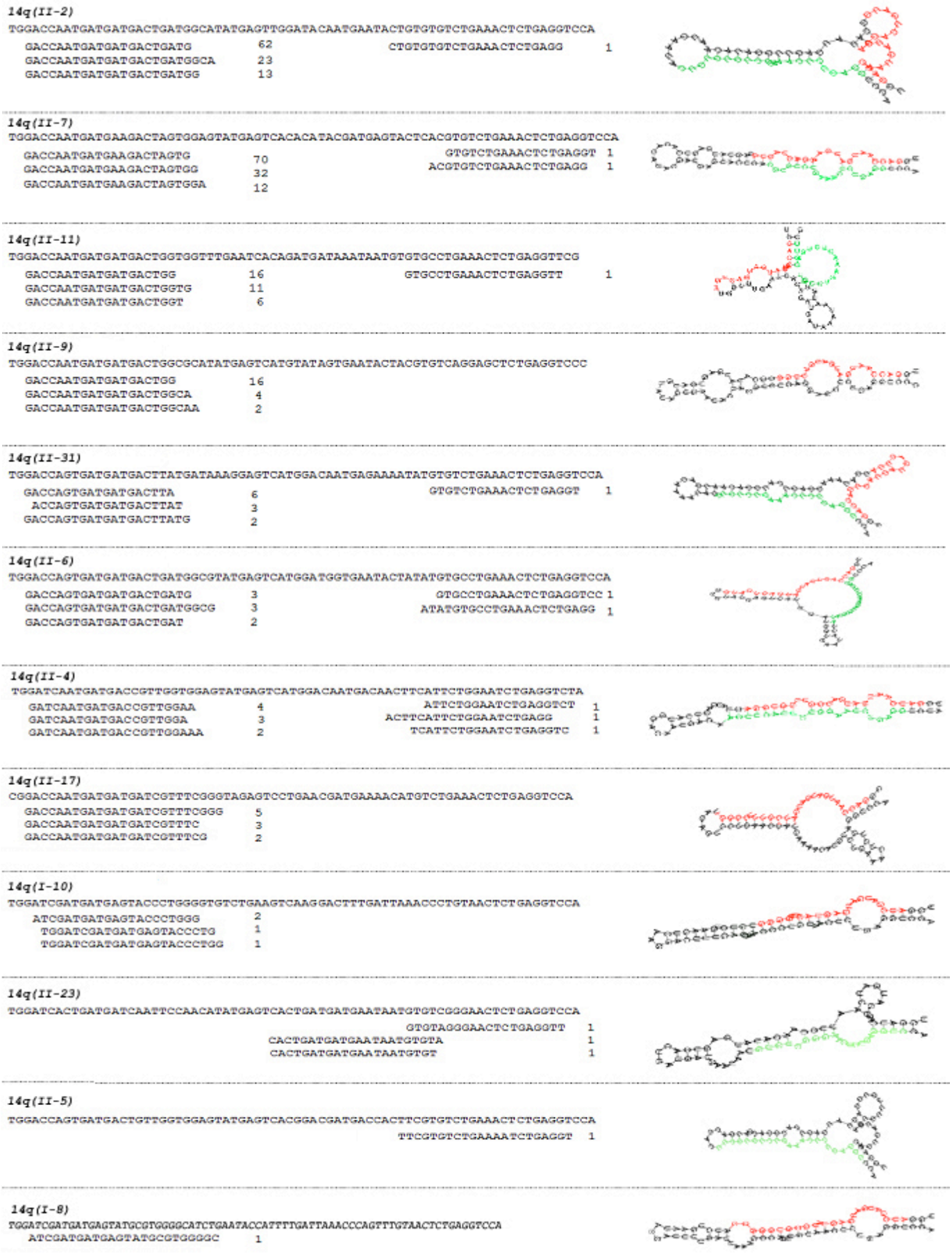
Suppl. Fig. 15B

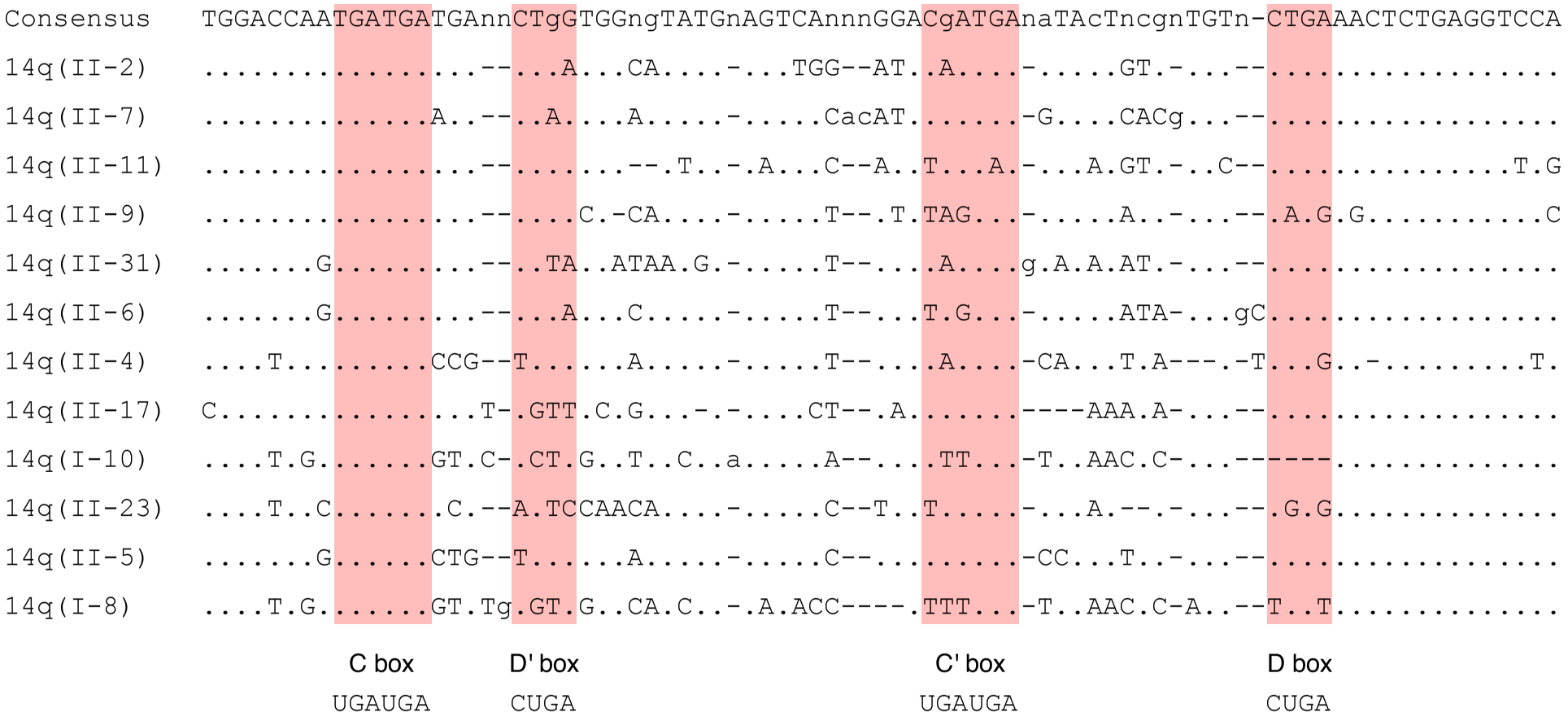


Suppl. Fig. 16

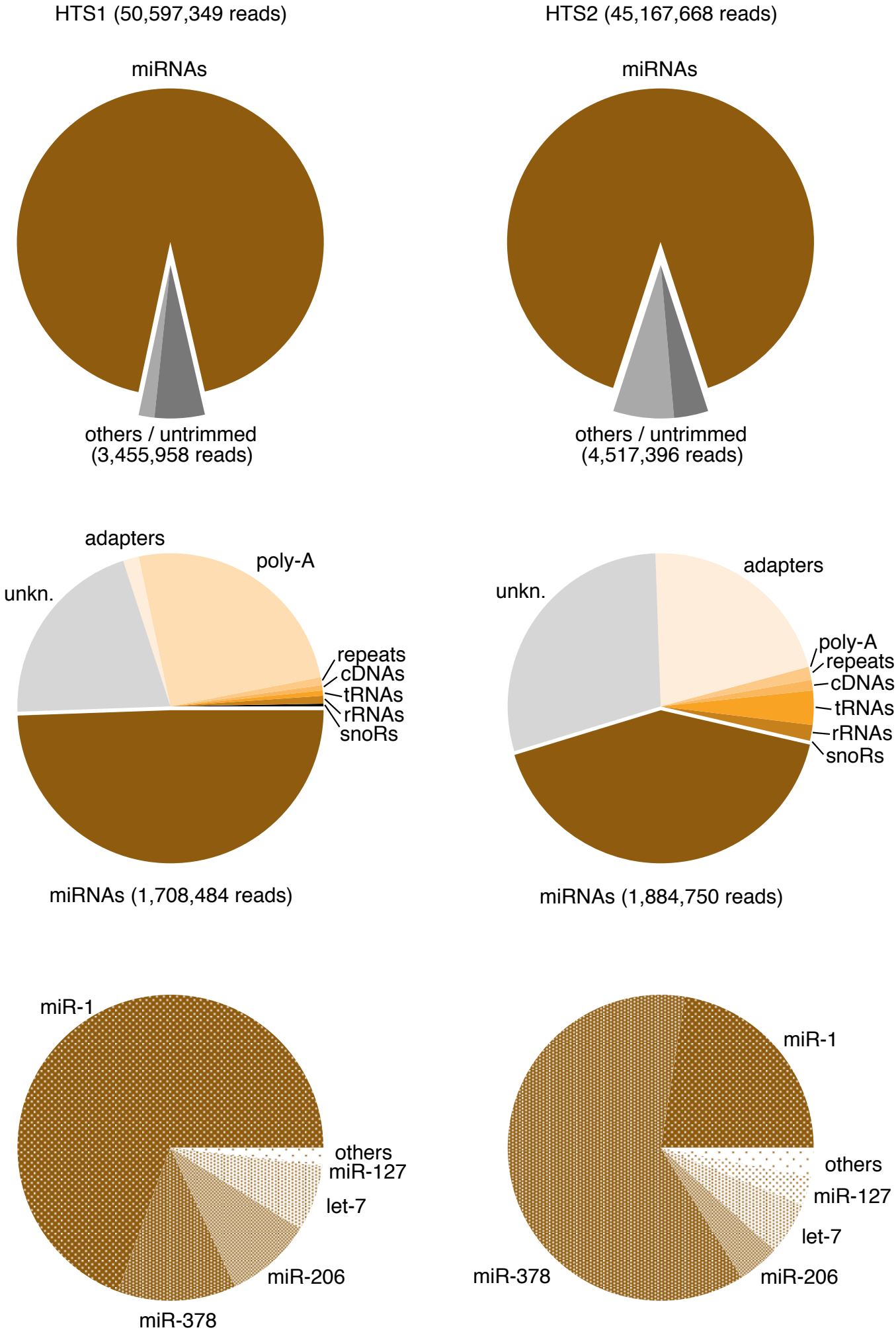

A
precursors
Log10 Read Count
B
precursors
Log10 Read Count

Suppl. Fig. 17A




Consensus TGGACCAATGATGATGAnnCTgGTGGngTATGnAGTCAnnnGGACgATGAnaTAcTncgnTGTn-CTGAAACTCTGAGGTCCA
14q(II-2)
14q(II-7)
14q(II-11)
14q(II-9)
14q(II-31)
14q(II-6)
14q(II-4)
14q(II-17)
14q(I-10)
14q(II-23)
14q(II-5)
14q(I-8)
C box
UGAUGA
D' box
CUGA
C' box
UGAUGA
D box
CUGA

Suppl. Fig. 18


HTS1 (50,597,349 reads)
miRNAs
others / untrimmed
(3,455,958 reads)
HTS2 (45,167,668 reads)
miRNAs
others / untrimmed
(4,517,396 reads)
adapters
poly-A
unkn.
repeats
cDNAs
tRNAs
rRNAs
snoRs
miRNAs (1,708,484 reads)
adapters
unkn.
poly-A
repeats
cDNAs
tRNAs
rRNAs
snoRs
miRNAs (1,884,750 reads)
miR-1
others
miR-127
let-7
miR-206
miR-378
miR-1
others
miR-127
let-7
miR-378
miR-206

Supplemental Table 2A: Gene ontology terms enriched amongst mammalian genes with conserved target sites for miRNAs in the *DLK1-GTL2* domain.

GO	GO ID	GO TERM NAME	N° of genes	p_{NOM}	p_{BONF}
SLIM	GO:0005488	binding	2070	$5x10^{-4}$	0.027
	GO:0005515	protein binding	1530	$2x10^{-4}$	0.011
	GO:0003676	nucleic acid binding	640	$\leq 1x10^{-4}$	0.005
	GO:0030528	transcription regulator activity	387	$\leq 1x10^{-4}$	0.005
	GO:0050789	regulation of biological process	1320	$\leq 1x10^{-4}$	0.005
	GO:0005634	nucleus	964	$\leq 1x10^{-4}$	0.005
	GO:0007275	multicellular organismal development	631	$\leq 1x10^{-4}$	0.005
	GO:0030154	cell differentiation	363	$\leq 1x10^{-4}$	0.005
WHOLE	GO:0000932	cytoplasmic mRNA processing body	15	$9.5x10^{-5}$	0.363
	GO:0003723	RNA binding	126	$6.5x10^{-5}$	0.215
	GO:0006511	ubiquitin-dependent protein catabolic process	45	$3.5x10^{-5}$	0.129
	GO:0051246	regulation of protein metabolic process	21	$2x10^{-5}$	0.098
	GO:0007399	nervous system development	110	$1.5x10^{-5}$	0.098
	GO:0005634	nucleus	932	$\leq 5x10^{-6}$	0.034
	GO:0045449	regulation of transcription	242	$\leq 5x10^{-6}$	0.034
	GO:0043565	sequence-specific DNA binding	140	$\leq 5x10^{-6}$	0.034

GO: utilized version of gene ontology graph. p_{NOM}: nominal significance. p_{BONF}: significance after correction for the total number of GO terms (whole: 6930; slim: 55).