

TEsmall identifies small RNAs associated with targeted inhibitor resistance in melanoma

Kathryn O'Neill^{1*}, Wen-Wei Liao^{2*}, Ami Patel³ and Molly Gale Hammell^{1**}

¹ Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

² Washington University School of Medicine, St. Louis, MO 63110

³ Mount Sinai Health System, New York, NY 10029

*These authors contributed equally to the work

** Correspondence should be addressed to mhammell@cshl.edu

MicroRNAs (miRNAs) are small 21-22nt RNAs that act to regulate the expression of mRNA target genes through direct binding to mRNA targets. While miRNAs typically dominate small RNA transcriptomes, many other classes are present including tRNAs, snoRNAs, snRNAs, Y-RNAs, piRNAs, and siRNAs. Interactions between processing machinery and targeting networks of these various small RNA classes remains unclear, largely because these small RNAs are typically analyzed separately. Here we present TEsmall, a tool that allows for the simultaneous processing and analysis of small RNAs from each annotated class in a single integrated workflow. The pipeline begins with raw fastq reads and proceeds all the way to producing count tables formatted for differential expression. Several interactive charts are also produced to look at overall distributions in length and annotation classes. We next applied the TEsmall pipeline to small RNA libraries generated from melanoma cells responding to targeted inhibitors of the MAPK pathway. Targeted oncogene inhibitors have emerged as way to tailor cancer therapies to the particular mutations present in a given tumor. While these targeted strategies are typically effective for short intervals, the emergence of resistance is extremely common, limiting the effectiveness of single-agent therapeutics and driving the need for a better understanding of resistance mechanisms. Using TEsmall, we identified several microRNAs and other small RNA classes that are enriched in inhibitor resistant melanoma cells in multiple melanoma cell lines and may be able to serve as markers of resistant populations more generally.

Introduction

microRNAs (miRNAs) are small 21-22 nucleotide RNA molecules which have been shown to play a critical role in metazoan development and gene regulation, and while canonically derived from short hairpin RNA precursors, have been shown to originate from a variety of sources including tRNAs and introns.^{1,2} In addition to governing development, sRNAs play a critical role in repressing transcripts derived from repetitive regions of the genome like transposons. In animals, siRNAs and piRNAs function to repress transposons in somatic cells, and the germline respectively^{3,4}. Identification of miRNAs and siRNAs which originate from non-canonical regions of the

genome is more challenging with few programs designed to detect sRNAs from all classes in both unique and repetitive genomic loci. It is for this reason we present TEsmall, a package designed specifically for the simultaneous analysis of sRNAs derived from a variety of genomic features. In particular, this package facilitates the discovery of intriguing biological phenomena otherwise masked by insufficient annotation of repetitive genomic elements, such as siRNAs, and allows these elements to be easily incorporated into downstream differential analysis through packages like DESeq2⁵.

We have tested the ability of TEsmall to characterize the expression profiles of small RNAs from a variety of classes in the context of melanoma cell lines responding to targeted inhibitors of the BRAF oncogene. The genetic basis of melanoma development is fairly well understood, with activating mutations in the oncogene BRAF occurring in a majority of melanoma patient tumors⁶, which also harbor hundreds of secondary mutations of unknown impact. Specific inhibitors that target activated BRAF as well as the downstream MAPK/ERK signaling pathway have been developed, which dramatically reduce the growth of melanoma cells in patients. However, the effects of these drugs typically extend patient lifespan for six months or less, as the tumors rapidly develop resistance to these targeted therapies⁷. While some tumors resistant to BRAF inhibitors acquire additional genetic lesions that elevate MAPK or AKT signaling⁸, many therapy-resistant cell lines establish resistance without a clearly understood mechanism of resistance⁹. Changes to small RNA profiles in melanoma cells responding to targeted inhibitors is an especially poorly understood subset of the genomic and transcriptomic changes that occur. To understand how small RNA alterations might contribute to the development of resistance to BRAF inhibitors in 451Lu melanoma cells that carry BRAFV600E mutations, we undertook a small RNA sequencing study of cells before and after the establishment of BRAF inhibitor resistance.

Materials and Methods

Melanoma cell culture

In this dataset, 451 Lu patient derived melanoma cell lines were used to explore the sRNA profiles of cells that are either sensitive or resistant to small molecule inhibitors of the BRAF kinase. Specifically, the melanoma patient derived 451Lu-Par cells are grown in standard growth media (DMEM with 10% FBS), while the 451Lu-BR cells are grown in standard growth medium supplemented with a 1uM concentration of the BRAF inhibitor vemurafenib. Both cell lines are adherent cells grown in standard 2D cell culture. The derivation of BRAF inhibitor resistance in these cells lines is described by Villanueva et al.⁷ and the cell lines are available from Rockland for both 451Lu cells (cat: 451Lu-01-0001) and 451Lu-BR cells (cat: 451Lu BR-01-0001).

Small RNA sequencing libraries

Total RNA was extracted using the Ambion PureLink RNA Mini Kit to extract up to 2 µg of total RNA from $\sim 1 \times 10^6$ melanoma cells from either the 451Lu-Par or 451Lu-

BR lines. Following Bioanalyzer verification of RIN numbers at or above ~9, the RNA extracts were next used to create small RNA sequencing libraries. The small RNA sequencing libraries were prepared with the Illumina TruSeq Small RNA Library Preparation Kits using an input of 1.2 µg total RNA and following the manufacturer's protocols as described, using 15 PCR cycles to reduce the likelihood of PCR amplification artifacts. The libraries were pooled and indexed with 6nt Illumina barcodes, such that 6 libraries could be sequenced per lane on an Illumina Genome Analyzer IIx. The reads were sequenced as single-end 50bp reads, to a depth of approximately 35 million reads per library. The dataset is available through GEO at the following accession number: GSE116134. A table of sequenced and mapped read counts for each library is presented in Supplemental File 1.

qPCR Validation

Taqman qPCR assays were used to validate the analysis results of TEsmall for a subset of microRNAs. Specifically, standard Taqman qPCR probes were obtained for the following microRNAs: miR-100, miR-184, and miR-211. Control probes were obtained for RNU58 and the U6 small RNA. Custom Taqman probes were obtained for the predicted mature sequence of the novel candidate miRtron derived from the VIM intron 6 locus. The Taqman protocol was followed exactly as described, using 1 µg of total RNA as the input from two biological replicates of the 451Lu-Par and 451Lu-BR cell lines, and using 3 technical replicates per biological replicate. The "Comparative Ct" analysis method described in the manufacturer's protocol was used for calculating fold change, standard deviation, and t-test based P-values. Briefly, the three technical replicates for each probe were combined to create a mean Ct value per probe per sample. The average of the Ct values from the two control probes in each sample was then subtracted from each microRNA Ct value to create a normalized " Δ -Ct" value for each microRNA in each sample. Following averaging of Δ -Ct values between the two biological replicates in each condition, the Δ - Δ -Ct value was calculated as the difference in mean Δ -Ct values for the same microRNA across conditions. Fold change represents $2^{\Delta\Delta\text{-Ct}}$, with errors on each Ct value combined quadratically.

TEsmall module

TEsmall functions by accepting raw input in FASTQ file format from next generation sequencing platforms in conjunction with genomic annotation sets via an online server. Adapters associated with siRNA library preparation are trimmed by TEsmall through the cutadapt package¹⁰. In order to remove degradation products from abundant ribosomal RNAs, rRNA derived reads are next filtered from the data before proceeding to analysis. This mapping step allows for up to 2 mismatches and filters a single alignment per read specified by the option: bowtie -v 2 -k 1 using bowtie (v1.2.1)¹¹. Small RNA reads remaining after rRNA filtering are then aligned more stringently, disallowing mismatches, option: bowtie -v 0 -a -m 100. All alignments in this step which map to fewer than 100 genomic loci are reported allowing for the classification of multimapper reads common to sRNA data, in particular structural RNAs like tRNAs and transposable element targeting siRNAs. Following alignment to the

genome, each alignment annotated via a sequential decision tree, as follows. The reads are aligned to each annotation category in order, then removed from the pool of alignments in order to facilitate priority annotation of, for example, intronic microRNA reads that should properly be annotated as microRNAs rather than intronic RNAs. The default order is: structural RNAs, miRNAs and hairpins, exons, sense transposons, antisense transposons, introns, and ultimately piRNAs. This ordering, can be re-ordered by the user to suit the application and user preferences about annotation priority. An HTML output file is then created using python based Bokeh tools¹² to visualize the abundance distributions, length distributions, and mapping logs of all small RNAs in the dataset (Fig. 1). In conjunction with this HTML output, TEsmall compiles multiple flat text output files, including a counts file that is structured to be directly compatible with DESeq2⁵ for differential analysis. The abundance calculations for these counts files are 1/n normalized, where n represents the number of alignments per read, to ensure no double-counting of multimappers.

The TEsmall code is available open-source from GitHub at the following location: <https://github.com/mhammell-laboratory/tesmall>.

Differential Analysis with DESeq2

The counts file produced by TEsmall were subsequently imported into DESeq2 (v1.18.1) to perform differential analysis between 451Lu-PAR and 451Lu-BR cell lines, as follows. The counts file was filtered to remove low abundance species (< 20 counts across all libraries) and increase the sensitivity of DESeq2. Normalization of the counts for differential analysis was performed using the default DESeq2 method during statistical analysis. For downstream visualization, the counts were normalized by the built-in variance stabilizing transformation (VST) method in DESeq2. Small RNAs with an adjusted P-value < 0.05 were considered statistically significant. The full DESeq2 output file is given in Supplemental File 2.

Visualization

Figures were produced using the R packages ggplot2¹³, gplots¹⁴ and GenomicRanges¹⁵ for scatterplots, heatmaps, and wiggleplots respectively. Python package matplotlib¹⁶ was used for all barplots. RNA secondary structures were rendered using the forna webtool¹⁷, secondary structure for the Arg-ACG-1-2 tRNA was pulled from the UCSC GtRNAdb tRNA covariance model, and structure of vimentin intron 6 was predicted using RNAfold's minimum free energy model.¹⁸

Results

TEsmall Workflow

TEsmall is a package specifically designed to identify sRNAs derived from a variety of genomic features simultaneously, such that users can evaluate the relative

abundances and profiles of many source of small RNAs on a common scale in a single pipeline. This serves as a novel improvement to currently available software such as mirDeep2¹⁹ and piPipes²⁰, which are optimized for the analysis of miRNAs and piRNAs respectively, but are not equipped to evaluate both types of small RNAs together. TESmall is also designed so that its output is optimally formatted for downstream differential analysis with statistical modeling software, such as DESeq2²¹. A flowchart describing the entire TESmall workflow is given in Figure 1A, with example output charts given in panels 1B-E. Specifically, in the first module of TESmall, raw small RNA sequencing reads from Illumina NGS sequencing platforms serve as the input without the need to pre-process the data before beginning analysis. TESmall first trims adapter contaminants from the reads and then filters the reads for appropriate size ranges, with a default of 16-36 nucleotides in length. The next module of TESmall removes contaminating ribosomal RNA fragments by mapping with bowtie¹¹ to a library of rRNAs for the specified genome. Removal of rRNA reads is critical as rRNA degradation products are a major source of contamination for sRNA data. Remaining reads are mapped to the genome, with a default mapping strategy optimized for repetitive regions with up to 100 alignments per read, though this may be altered by the user. The reads are next sequentially annotated to several small RNA classes and genomic features, with a decision tree implemented to prioritize annotation categories. This has the goal of attributing reads mapping to intronic microRNAs as “microRNA” reads, for example, rather than annotating these reads as having an intronic source. Following annotation of each read, aggregate abundances are calculated for each sequencing library and output as a counts table suitable for downstream differential expression analysis. Importantly, any multimapper reads in these counts tables are weighted according to the number of genomic loci from which they derive ($1/n$ where n is the number of alignments) to avoid any double-counting of multimapper reads in the counts tables. In addition to an output file including all raw count data per sample, RNA species ID, and type classification, TESmall provides an aesthetic output HTML (Fig. 1B) summarizing distribution of read lengths (Fig. 1C), proportion of reads assigned to each sRNA type (Fig. 1D), and distribution of reads of a particular size to each of the sRNA categories (Fig. 1E). In addition to these summary plots, TESmall presents a table with summary statistics of read proportion, raw input and trimmed read counts to quickly assess any potential biases in library preparation that may affect downstream normalization.

Application of TESmall to melanoma sRNA profiles

As described, drug resistance is a known hurdle in the treatment of melanoma, driving the need for a better understanding of how cells develop resistance. We have chosen to investigate the alterations in sRNA profiles, as one marker of cellular state. To investigate the effect of BRAF kinase drug resistance on sRNA composition in patient derived melanoma cell lines, we performed differential expression analysis following classification by TESmall. Resistant lines were derived through exposure of 451Lu patient derived parental cell lines to increasing concentrations of vemurafenib up to 1 μ M. Resistant clones were selected and expanded before exposure to an increase in vemurafenib. Cells were otherwise treated as described in Villanueva et al.⁷ Raw count data was normalized as described in Materials and Methods by DESeq2. All VST

normalized counts were averaged between parental or resistant replicates and plotted against each other to visualize trends of expression across sRNA subtypes without filtering for significant or abundant transcripts, significantly differentially expressed transcripts are represented by solid coloring (Fig. 2). Overall, there appears to be a trend towards lower expression of many sRNA classes in the 451Lu BRAF resistant samples (Fig. 2 and Fig. 3B). Upon filtering for the most abundant (base mean across all replicates > 500) and significantly differentially expressed transcripts (multiple-hypothesis testing adjusted p-value < .05), trends of lower sRNA expression in the BRAF resistant samples was still seen for many intronic, exonic, and transposon mapped sRNA species (Fig. 3). Interestingly, miRNAs show an even distribution of species with negative and positive log fold changes, and since miRNAs were the most abundantly sequenced small RNAs in the libraries, this rules out a normalization issue as the explanation for down-regulated small RNAs in the other classes. It may be of interest that, after filtering for significance and abundance, structural RNAs with a negative log fold change are exclusively tRNAs with the exception of vaultRNA 2-1, and those with a positive log fold change are almost exclusively snoRNAs. Details of the particular sRNAs differentially expressed in each of these classes are given in Supplemental File 2.

Type specific analysis of sRNA species and Validation with qPCR

Several miRNAs which are significantly differentially expressed in our dataset have been previously described in the literature as playing critical roles in melanoma progression or epidermal differentiation. This includes the miRNAs miR-184, miR-211, and miR-100. In other contexts, miR-184 has been shown to arrest epidermal differentiation through derepression of Notch in normal human keratinocytes and murine epidermis²². While expression of Notch in keratinocytes is known to have a tumor suppressive phenotype, its expression has the opposite effect in melanocytes through upregulation of the PI3K/Akt and MAPK pathways²³. Our data showed an approximate 5-fold increase in miR-184 expression in BRAF inhibitor resistant cells in comparison to parental (Fig. 4A and Supplemental File 2), consistent with a model where MAPK pathway activation provides a mechanism for BRAF inhibitor resistance.^{7,24} It has also been shown that BRAF inhibitor resistance can be mediated by regulatory escape of the transcription factor MITF from the MAPK pathway, where MITF overexpression itself conferred resistance in several melanoma cell lines.²⁵ Consistent with a high MITF state, our data shows a significant upregulation of miR-211, derived from the MITF activated gene melastatin, and a significant downregulation of miR-222, known to be inversely correlated with MITF expression.²⁶ Finally, miR-100 was also shown to be significantly downregulated in our data; this was a miRNA of interest as it has been implicated in prostate cancer as a repressor of the oncogene mTOR.²⁷

To validate the expression profiles from our TEs small based differential expression analysis, we performed qPCR on several miRNAs of interest including miRNAs miR-184, miR-211, and miR-100, all of which recapitulated the trend observed in our small RNA-seq dataset (Fig. 4A). Reassuringly, the expression alterations of miR-

204 and miR-211 seen in our data were also seen in an alternative melanoma derived cell line A375 in Díaz-Martínez et al. following induction of BRAF inhibitor resistance.²⁸

Upon further investigation into individual RNA species from different subtypes for follow up, we encountered an interesting and novel 21 nucleotide sRNA associated with the sixth intron in the vimentin gene (VIM) (Fig. 4B). Intronic microRNAs that derive from short spliced introns with internal hairpin structures have been observed previously and dubbed miRtrons.²⁹ MiRtrons are typically generated via the splicing machinery and subsequently processed by DICER in the cytoplasm, bypassing the canonical nuclear Drosha processing steps. As visible in the minimum free energy secondary structure prediction by RNAfold, the candidate miRtron of interest is located in a stem loop structure which appears conducive to processing by DICER (Fig. 4C). The length of this mature sRNA, its abundance as a single RNA species, and its secondary structure within VIM intron 6 are all consistent with miRtronic miRNAs. This is particularly interesting as VIM is a known marker for the epithelial to mesenchymal transition and is well expressed in many cell types, but has not previously been shown to harbor a miRtron, suggesting this VIM miRtron might represent a novel miRNA with particular abundance in melanoma cells.

In addition to miRNAs, TEs_{small} recognizes several other types of sRNAs. It has been previously reported that tRNA derived small RNA molecules (tRFs) can silence LTR retrotransposable elements through occupation of the primer binding site (PBS) as an adaptation of the role of tRNAs as retroviral primers.^{2,30} Through TEs_{small}, one is able to detect reads associated with tRNAs and transposable elements in the same pipeline facilitating observation of phenomena such as these. In our analysis, several species of sRNAs mapping antisense to transposable elements were significantly depleted in BRAF resistant cell lines compared to parental (Fig. 2). Upon further investigation we were able to determine that a subset of these reads were tiRNAs derived from the Arg-CGY family of tRNAs (Fig. 5B). These candidate tiRNAs mapped to a subset of HERVs including HERV3, HERV30, MER51, and others (please see Supplemental File 3). This is consistent with previous literature showing HERV-R retrotransposons are primed by Arg tRNAs (Fig. 5A) and tRNA derived fragments (tRFs) can occupy retroviral primer binding sites to suppress transposon activity.^{2,30} It is important to note that the Arg-CGY sRNAs reported by TEs_{small} are consistent with the tRFs previously described in Schorn et al. as they are 18nt CCA-appended fragments originating from the 3' T-arm of tRNAs.² This is shown graphically in Fig. 5, where the pileup of reads at an example HERV PBS locus can be seen in Fig. 5B, and the pileup of these same reads at the originating tRNA locus can be seen Fig. 5C-5D. In the tRNA profiles, other tRNA fragments can be seen outside of the tiRNA generating 3' end, but these do not predominantly accumulate as a single abundant sRNA species.

In addition to the miRNAs and tiRNAs highlighted above, several additional species of small RNAs were reported by TEs_{small} as differentially expressed in the 451Lu BR cells including: siRNAs mapping to transposable element loci, exonic loci, and a variety of structural RNA classes. The structural RNA group included snoRNAs,

snRNAs, tRNA fragments, and a vault RNA. The full list of differentially expressed small RNAs can be found in Supplemental File 2.

Comparison of TEs_{small} with sRNA Analysis Software

Several software packages exist to characterize small RNA data for expression profiling analysis. However, programs designed for this purpose such as miRDeep2¹⁹ and ShortStack³¹ typically focus on a particular category of sRNA, predominantly miRNAs. One package that considers multiple sRNA types is piPipes²⁰, a software suite designed for both piRNA and siRNA analysis. While piPipes functions well to annotate and characterize piRNAs by read pileups associated with the ping-pong cycle of piRNAs, it is not particularly suited for annotation of sRNAs from other types of genomic loci, such as miRNAs, siRNAs, and tiRNAs. PiPipes will provide plots of read distribution across lists of transposable elements and piRNA clusters, however, one cannot access tables of these counts with associated TE annotation, suitable for differential expression analysis. While piRNAs are annotated with their respective piRNA clusters, siRNAs are assigned a chromosomal coordinate providing some difficulty in determining patterns in the sources or targets of these reads. It is also of import that intron derived miRNAs like the VIM miRtron were not captured, as there is no mechanism by which to assign siRNA reads beyond mapping the chromosomal coordinates associated to preloaded annotation sets associated with TEs and piRNA clusters. TEs_{small}, which does not perform piRNA-specific ping-pong analysis, provides a complementary package that is designed to be a general purpose small RNA analysis suite that can identify and analyze many types of sRNAs concurrently, presenting the output in a format intended for expression profiling analysis.

Discussion

TEs_{small} is a software package with novel functionality in that it allows the user to simultaneously map and annotate many types of sRNAs including structural RNAs, miRNAs, siRNAs, and piRNAs. This allows one to compare trends in expression between all sRNA types and investigate the cross-talk between distinct sRNA regulatory pathways. Other packages released to date focus on individual sRNA types like miRNAs^{19,31} or piRNAs²⁰ and while optimized for these applications, are not adapted for comparison across sRNA categories. In addition to handling multiple classes of sRNAs, the output of TEs_{small} is formatted for direct integration into downstream analysis pipelines. TEs_{small}'s output files are compatible with statistical analysis software like DESeq2 and efficient heatmap generation. In addition to requiring little data preprocessing, TEs_{small} outputs an aesthetic HTML file of charts (Fig. 1B) which allows for fast and effortless assessment of library quality, sRNA composition, and size distribution. TEs_{small} can also be expanded to function for any novel sRNA species provided the appropriate annotation files are available, allowing it to serve as a powerful tool to study RNA biology in many organisms.

We applied TEs_{small} to a novel dataset in which we compared the effects of BRAF inhibitor resistance on sRNA abundance in melanoma derived cell lines. In this

analysis we found several microRNAs whose expression was altered in BRAF inhibitor resistant cells in comparison to parental lines. A table of these hits can be found in Supplemental File 2. Among these candidates, we experimentally validated changes in expression of miRNAs miR-184, -211, and -100. Of particular interest is the novel Vimentin derived miRtron candidate, whose expression pattern was also experimentally validated. Close examination of the characteristic read pile up associated with the VIM miRtron, and secondary structure of intron 6 are all consistent with miRtron processing pathways. Further investigation will be required to determine if this is a true miRtron formed through an intermediate spliceosome derived lariat independent of the Drosha microprocessor subunit, or is instead a canonical Drosha-dependent miRNA.

In addition to revealing miRNAs previously described in the literature, TEs_{small} detected several novel classes of small RNAs which would not have been found using packages designed for miRNA analysis. TEs_{small} allows the user to investigate tRNA derived fragments which have been shown to play a critical role in LTR retro-transposon suppression². In the melanoma dataset, we identified a novel candidate tiRNA that appears to derive from ARG-tRNAs and to potentially regulate several HERV-R type LTR elements through occupancy of the primer binding site. Other types of siRNAs that regulate transposon expression were also shown to be differentially expressed in these datasets, suggesting the possibility that transposon-derived transcripts are altered in these BRAF inhibitor resistant melanoma cells.

It is well known small noncoding RNAs of different subtypes work in conjunction to regulate cellular processes through complex networks, particularly in the realm of transposon silencing. piRNAs known to regulate transposon expression in the germline have been found to work in cooperation with siRNAs to perform this task.³² In plants, miRNAs have been shown to play a role in transposon silencing by serving as an intermediate to form 21 nucleotide siRNAs via RNA dependent RNA polymerase and while the mechanism would be disparate from plants, hints of miRNAs facilitating transposon silencing have been seen in animals as endogenous and introduced retroviral elements with homologous regions to miRNAs have lower genomic activity.^{33,34} Current sRNA analysis packages are specific to one or two types of sRNAs making it easy to overlook biologically interesting patterns of interaction between sRNA classes. For this reason, we have created TEs_{small}, an easy to use package with aesthetic output designed for the concurrent expression analysis of multiple sRNA subtypes.

Figures Legends

Figure 1. Flow chart and output HTML of TEsmall. (A) Flow chart of TEsmall's treatment of input high-throughput sequencing data, input genome indices, and output. (B) Screenshot of HTML output file for one sample (C) Bar plot depicting size distribution of unique and multimapper reads. (D) Circle plot depicting distribution of reads to each subtype. (E) Bar plot depicting proportion of subtypes across read length.

Figure 2. Scatterplots depicting BRAF resistant versus parental RNA VST normalized counts. (A) Overlay of all subtype scatterplots. (B) RNA subtype specific scatterplots. Transparent points represent RNA species with adjusted p-value < .05.

Figure 3. Differential expression of highly abundant transcripts. (A) Heatmaps depicting all significantly differentially expressed genes per subtype, all heatmaps are normalized to between -1 and 1 by row. (B) Bar plot depicting number of abundant and significantly differentially expressed species with a negative or positive log fold change per subtype. Structural RNA species are collapsed on duplicate tRNAs. Abundance is determined by a base mean between samples of greater than 500 counts. P1, P2, R1, and R2 represent parental replicates 1 and 2 and BRAF resistant replicates 1 and 2 respectively.

Figure 4. Detailed analysis miRNAs of interest. (A) qPCR representing log fold change of miRNAs 184, 211, and 100 in respect to parental expression levels across 3 replicates. (B) qPCR representing log fold change with respect to parental cell line and BAM gene alignment tracks across samples of the VIM miRtron C) RNAfold predicted secondary structure of VIM intron 6 with miRtron highlighted in purple.

Figure 5. tRNA and TE interaction through primer binding sites. (A) Diagram of primer binding by tRNA to facilitate retroviral reverse transcription. (B) Read alignment track of Arg-CGY family derived 18 nt CCA tailed fragment to HERV30 PBS. (C) Consensus histogram of reads distributed from Arg-CGY tRNAs, and derived tiRNAs D) Secondary structure of Arg-CGY family member Arg-ACG-1-2 with highlighted 15 nt CCA (-) fragment.

Acknowledgements

MGH is a scholar of the Rita Allen Foundation. KO acknowledges funding from the NIH training grant 2T32GM065094-16. We would also like to acknowledge the Meenhard Herlyn lab for help, advice, and for the kind gift of the 451Lu cells.

Author Contributions Statement

Authors KO and MGH analyzed the data, generated the figures and wrote the manuscript. MGH and AP designed the experiments and generated the data. WWL wrote the code for the TEs_{small} pipeline, packaged the code for GitHub, and contributed to the writing of the manuscript.

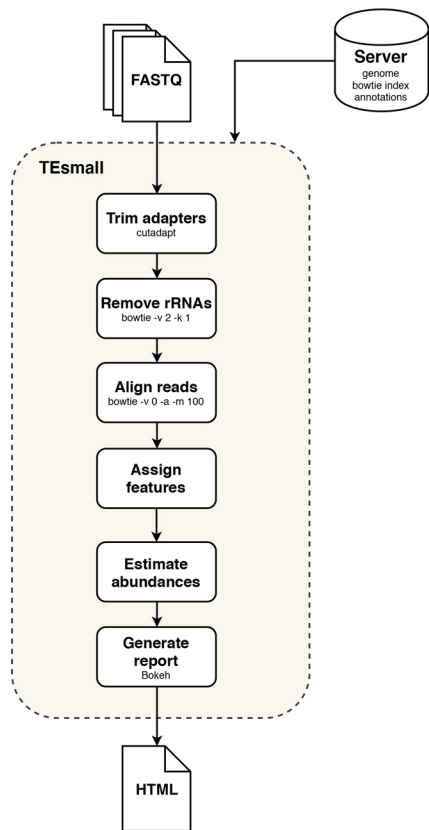
References

- 1 Bartel DP. Metazoan MicroRNAs. *Cell* 2018;**173**:20–51. <https://doi.org/10.1016/j.cell.2018.03.006>.
- 2 Schorn AJ, Gutbrod MJ, LeBlanc C, Martienssen R. LTR-Retrotransposon Control by tRNA-Derived Small RNAs. *Cell* 2017;**170**:61–71. <https://doi.org/10.1016/j.cell.2017.06.013>.
- 3 Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, *et al.* Endogenous siRNAs Derived from Transposons and mRNAs in Drosophila Somatic Cells. *Science (80-)* 2008;**320**:1077–81. <https://doi.org/10.1126/science.1157396>.
- 4 Castañeda J, Genzor P, Bortvin A. piRNAs, transposon silencing, and germline genome integrity. *Mutat Res - Fundam Mol Mech Mutagen* 2011;**714**:95–104. <https://doi.org/10.1016/j.mrfmmm.2011.05.002>.
- 5 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:1–21. <https://doi.org/10.1186/s13059-014-0550-8>.
- 6 Hodis E, Watson IR, Kryukov G V, Arold ST, Imielinski M, Theurillat J, *et al.* A Landscape of Driver Mutations in Melanoma. *Cell* 2012;**150**:251–63. <https://doi.org/10.1016/j.cell.2012.06.024>.
- 7 Villanueva J, Vultur A, Lee JT, Somasundaram R, Cipolla AK, Wubbenhorst B, *et al.* Acquired resistance to BRAF inhibitors mediated by a RAF kinase switch in melanoma can be overcome by co-targeting MEK and IGF-1R/PI3K. *Cancer Cell* 2010;**18**:683–95. <https://doi.org/10.1016/j.ccr.2010.11.023>.
- 8 Alcala AM, Flaherty KT. BRAF Inhibitors for the Treatment of Metastatic Melanoma : Clinical Trials and Mechanisms of Resistance. *Clin Cancer Res* 2012;**18**:33–40. <https://doi.org/10.1158/1078-0432.CCR-11-0997>.
- 9 Gatenby R, Brown J. The Evolution and Ecology of Resistance in Cancer Therapy. *Cold Spring Harb Perspect Med* 2018;**8**:1–12. <https://doi.org/10.1101/cshperspect.a033415>.
- 10 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing

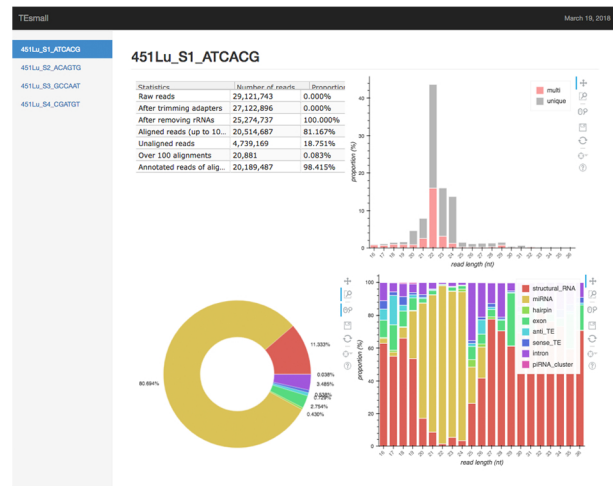
- reads. *EMBNet.journal* 2011;**17**:10–2. <https://doi.org/10.14806/ej.17.1.200>.
- 11 Langmead B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinforma* 2010;1–24. <https://doi.org/10.1002/0471250953.bi1107s32>.
- 12 Bokeh Development Team. Bokeh: Python library for interactive visualization 2014.
- 13 Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2009.
- 14 Warnes GR, Bolker B, Lumley T. gplots: Various R programming tools for plotting data. R package version 2.6.0 n.d.
- 15 Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, *et al*. Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* 2013;**9**:1–10. <https://doi.org/10.1371/journal.pcbi.1003118>.
- 16 Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng* 2007;**9**:90–5. <https://doi.org/10.1109/MCSE.2007.55>.
- 17 Kerpedjiev P, Hammer S, Hofacker IL. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics* 2015;**31**:3377–9. <https://doi.org/10.1093/bioinformatics/btv372>.
- 18 Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Res* 2008;**36**:70–4. <https://doi.org/10.1093/nar/gkn188>.
- 19 Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* 2012;**40**:37–52. <https://doi.org/10.1093/nar/gkr688>.
- 20 Han BW, Wang W, Zamore PD, Weng Z. PiPipes: A set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome-and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics* 2015;**31**:593–5. <https://doi.org/10.1093/bioinformatics/btu647>.
- 21 Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010;**11**:1–12.
- 22 Nagosa S, Leesch F, Putin D, Bhattacharya S, Altshuler A, Serror L, *et al*. microRNA-184 Induces a Commitment Switch to Epidermal Differentiation. *Stem Cell Reports* 2017;**9**:1991–2004. <https://doi.org/10.1016/j.stemcr.2017.10.030>.
- 23 Pinnix CC, Herlyn M. The many faces of Notch signaling in skin-derived cells. *Pigment Cell Res* 2007;**20**:458–65. <https://doi.org/10.1111/j.1600-0749.2007.00410.x>.
- 24 Villanueva J, Vultur A, Herlyn M. Resistance to BRAF Inhibitors : Unraveling Mechanisms and Future Treatment Options. *Cancer Res* 2011;**71**:7137–41. <https://doi.org/10.1158/0008-5472.CAN-11-1243>.
- 25 Van Allen EM, Wagle N, Sucker A, Treacy DJ, Johannessen CM, Goetz EM, *et al*. The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. *Cancer Discov* 2014;**4**:94–109. <https://doi.org/10.1158/2159-8290.cd-13-0617>.
- 26 Golan T, Messer AR, Amitai-Lange A, Melamed Z, Ohana R, Bell RE, *et al*. Interactions of Melanoma Cells with Distal Keratinocytes Trigger Metastasis via Notch Signaling Inhibition of MITF. *Mol Cell* 2015;**59**:664–76. <https://doi.org/10.1016/j.molcel.2015.06.028>.
- 27 Leite K, Morais D, Reis S, Viana N, Moura C, Florez M, *et al*. MicroRNA 100: a context dependent miRNA in prostate cancer. *Clinics* 2013;**68**:797–802.

- [https://doi.org/10.6061/clinics/2013\(06\)12](https://doi.org/10.6061/clinics/2013(06)12).
- 28 Díaz-Martínez M, Benito-Jardon L, Alonso L, Koetz-Ploch L, Hernando E, Teixido J. miR-204-5p and miR-211-5p contribute to BRAF inhibitor resistance in melanoma. *Cancer Res* 2018;**78**:1017–30. <https://doi.org/10.1158/0008-5472.CAN-17-1318>.
- 29 Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in Drosophila. *Cell* 2007;**130**:89–100. <https://doi.org/10.1016/j.cell.2007.06.028>.
- 30 Mak J, Kleiman L. Primer tRNAs for reverse transcription. *J Virol* 1997;**71**:8087–95.
- 31 Axtell MJ, ShortStack : Comprehensive annotation and quantification of small RNA genes *RNA* 2013;740–51. <https://doi.org/10.1261/rna.035279.112>.
- 32 Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, *et al*. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 2008;**453**:534–8. <https://doi.org/10.1038/nature06904>.
- 33 Zlotorynski E. RNA interference: MicroRNAs suppress transposons. *Nat Rev Mol Cell Biol* 2014;**15**:298–9. <https://doi.org/10.1038/nrm3788>.
- 34 Hakim ST, Alsayari M, McLean DC, Saleem S, Addanki KC, Aggarwal M, *et al*. A large number of the human microRNAs target lentiviruses, retroviruses, and endogenous retroviruses. *Biochem Biophys Res Commun* 2008;**369**:357–62. <https://doi.org/10.1016/j.bbrc.2008.02.025>.

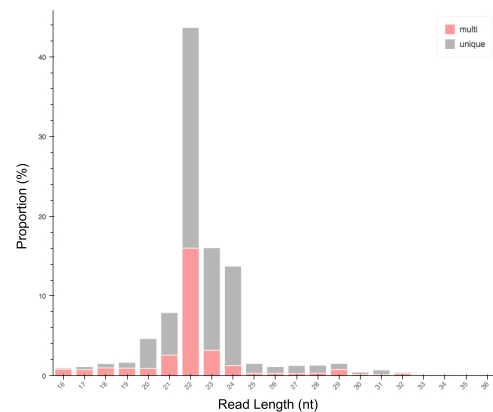
A



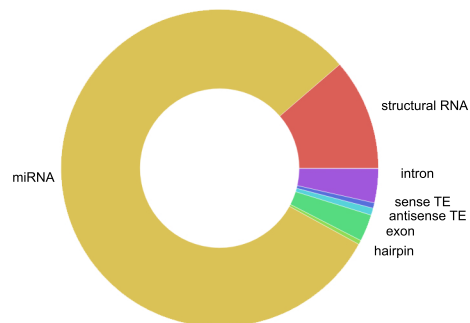
B



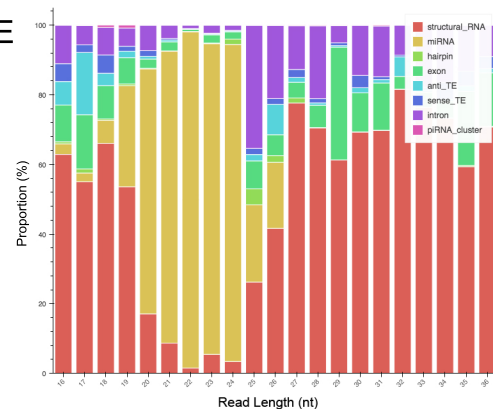
C



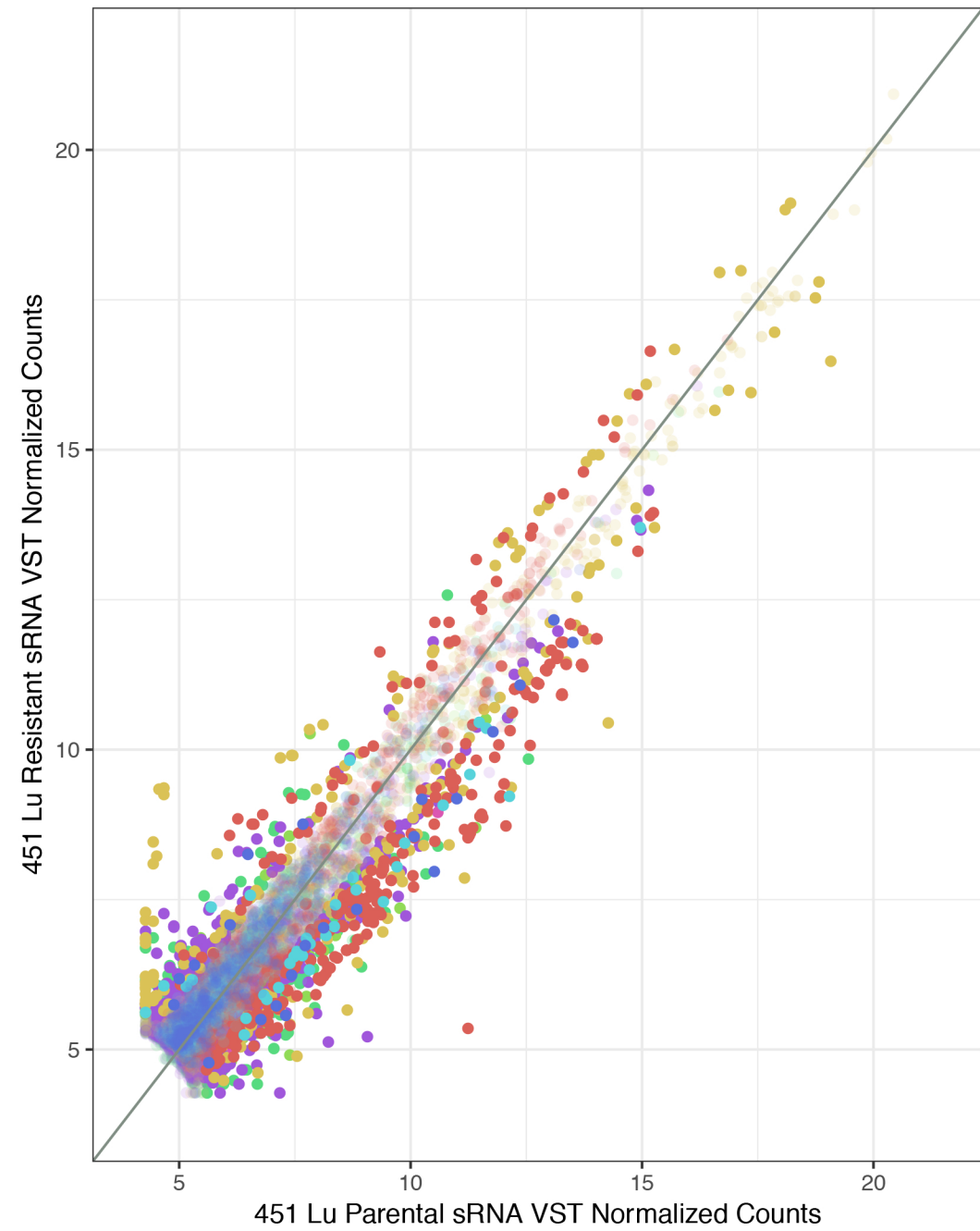
D



E



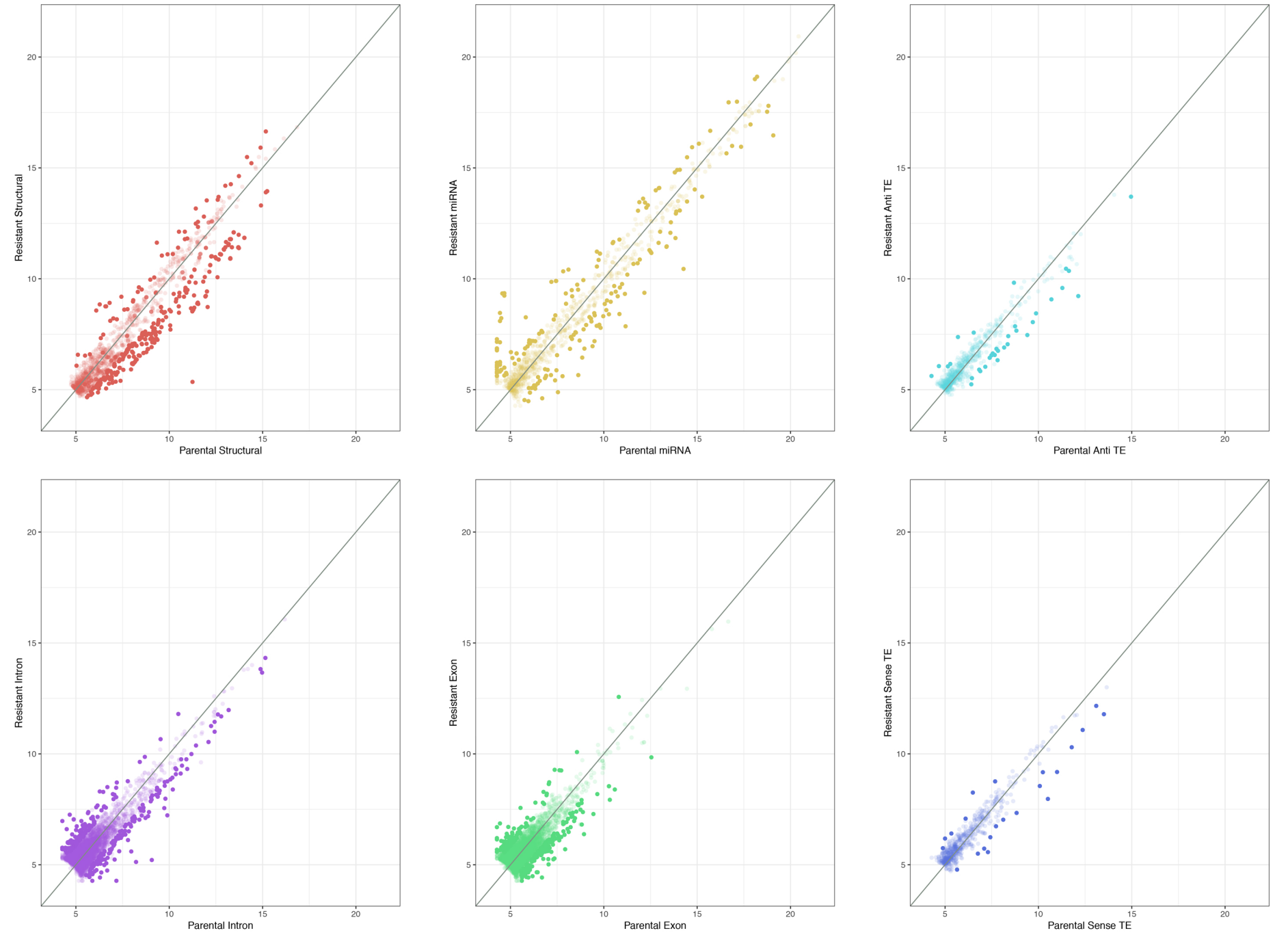
A



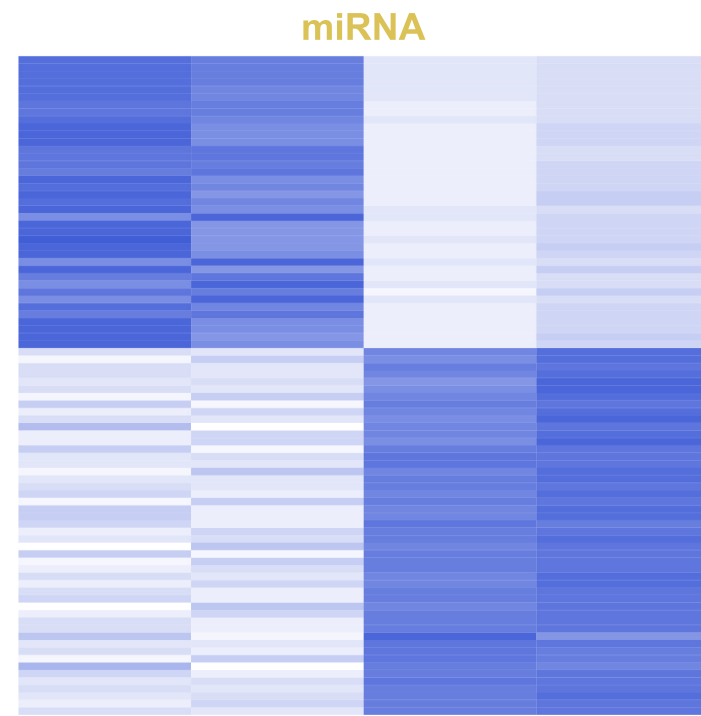
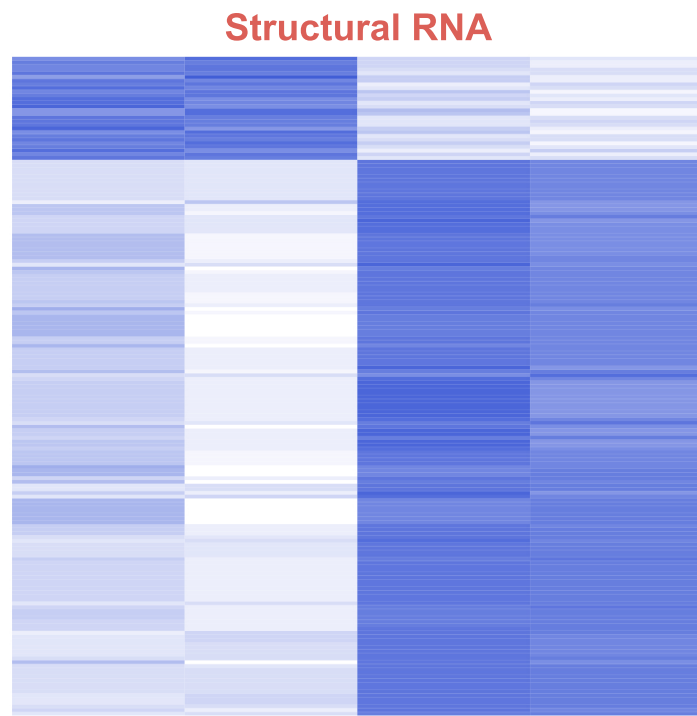
Type

- anti_TE
- exon
- hairpin
- intron
- miRNA
- piRNA_cluster
- sense_TE
- structural_RNA

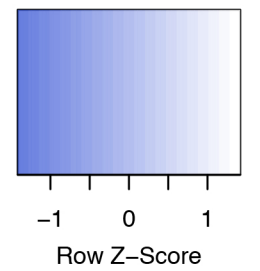
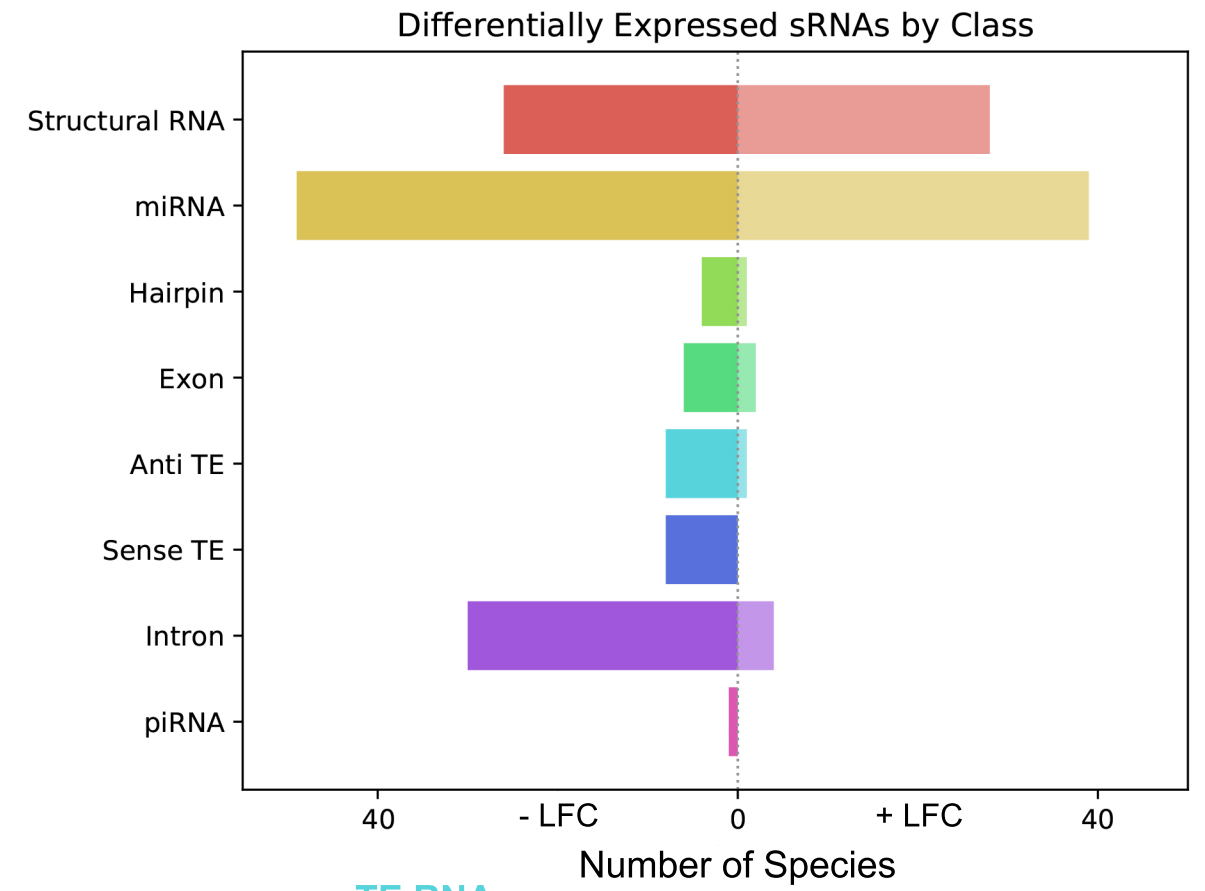
B



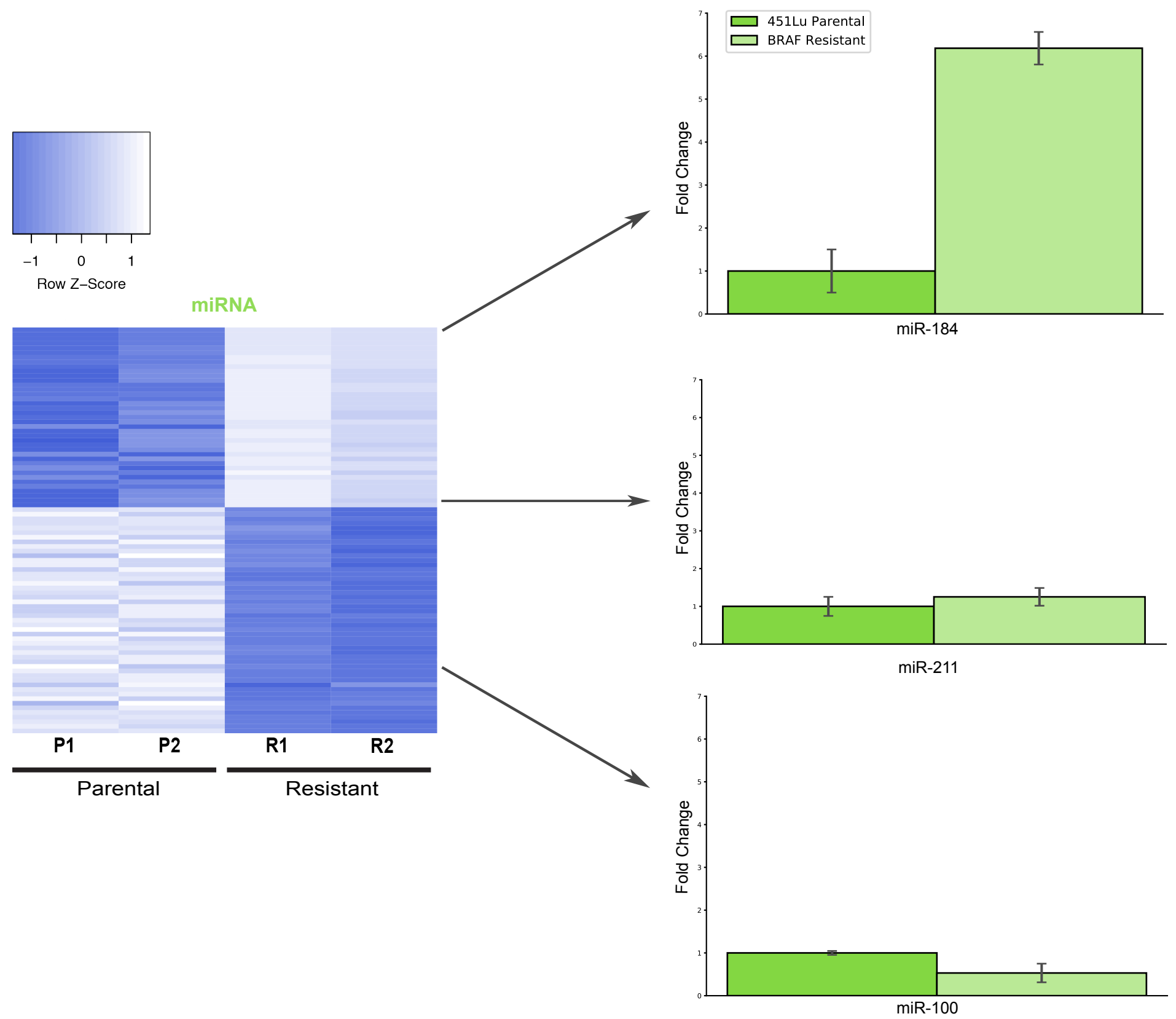
A



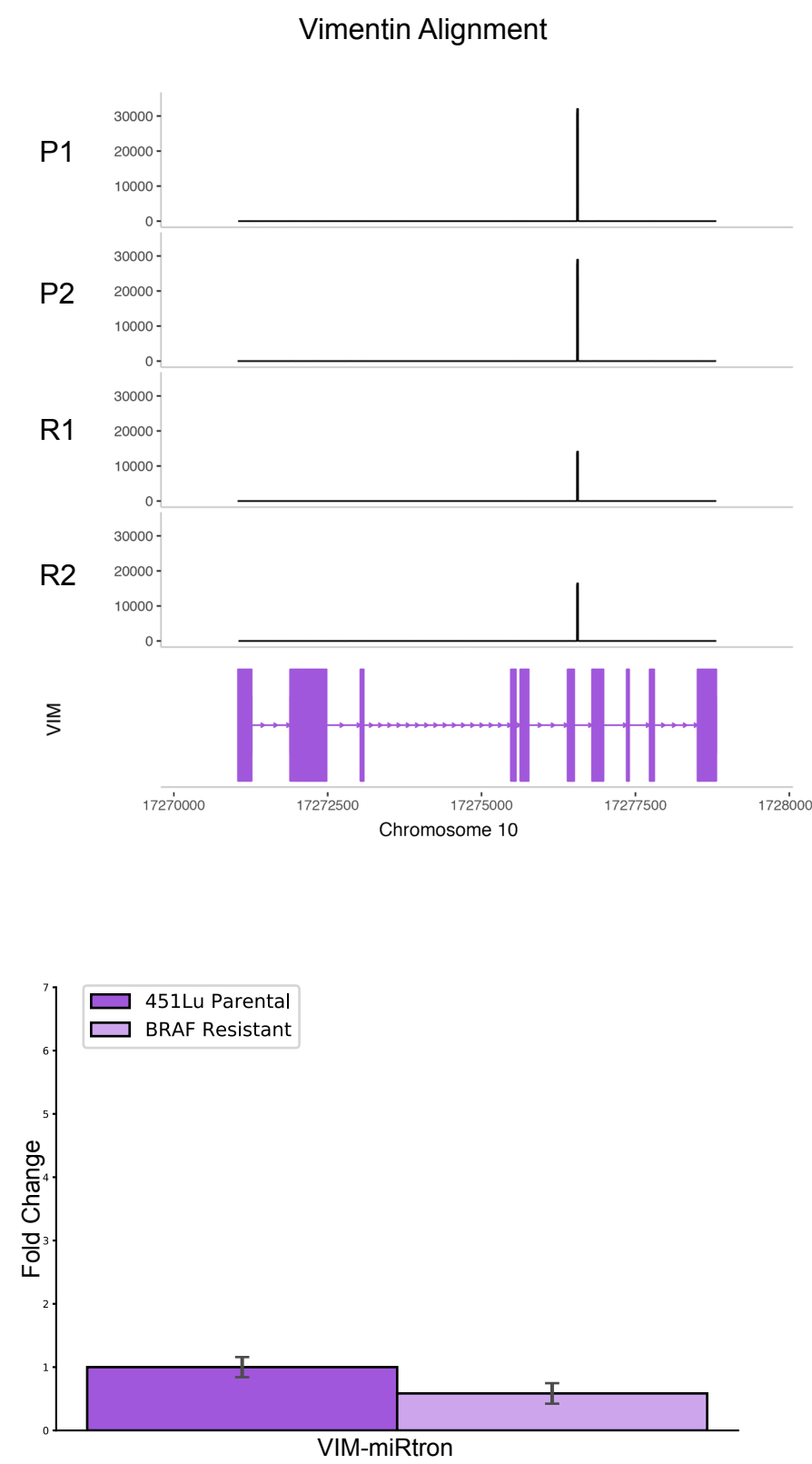
B



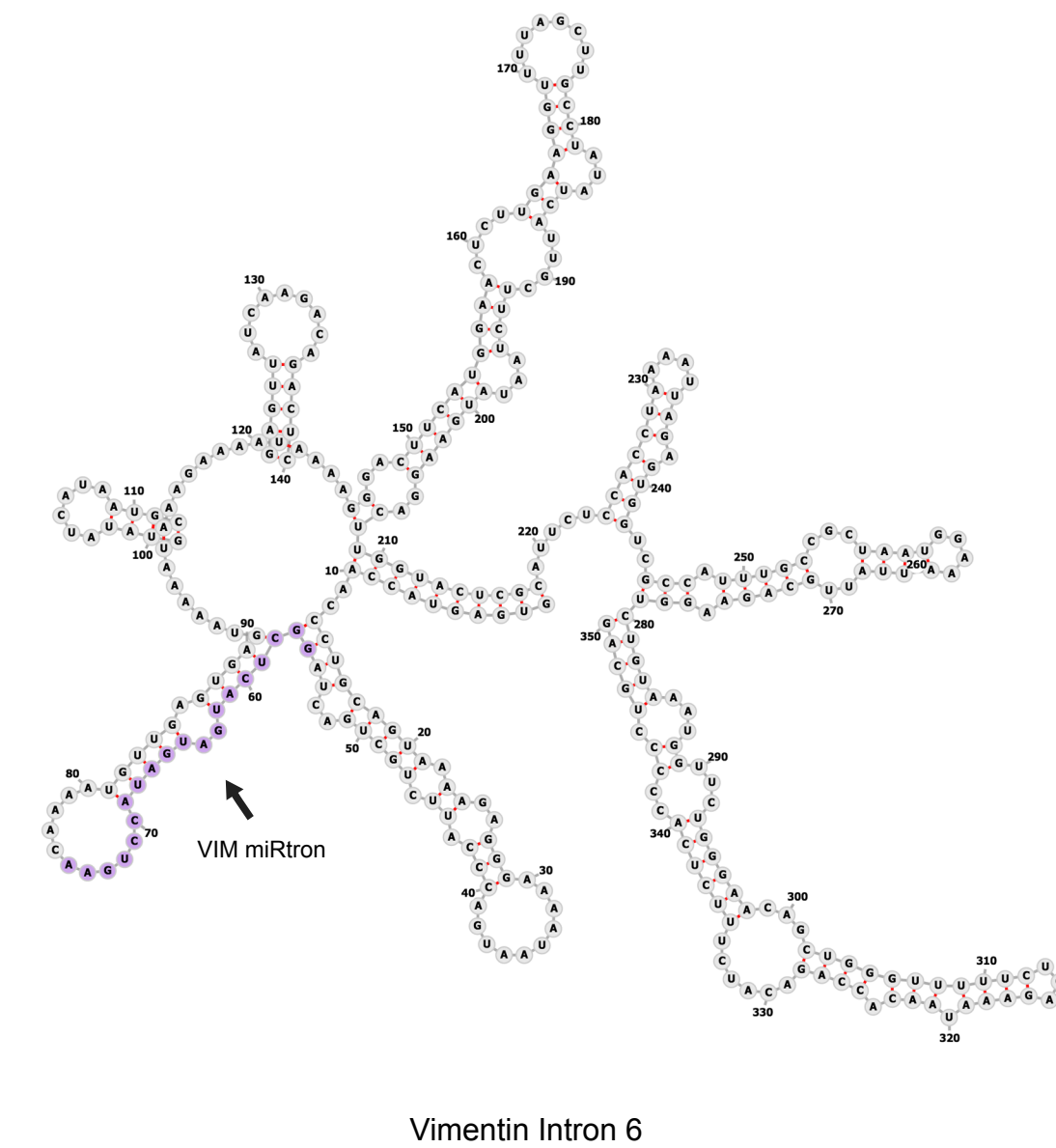
A



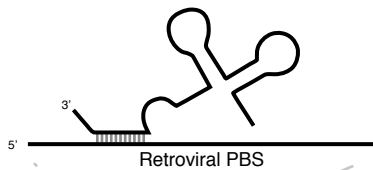
B



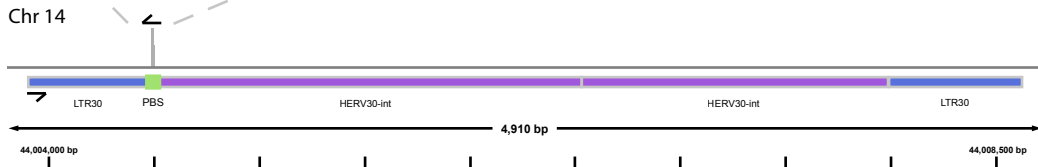
C



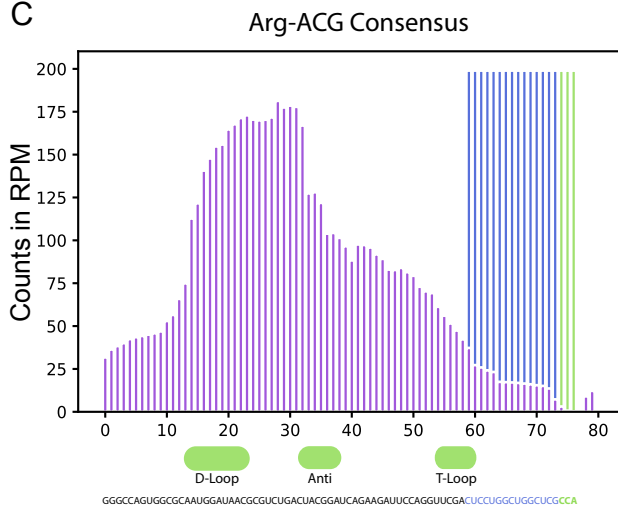
A



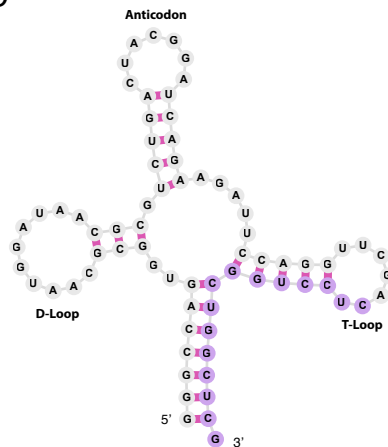
B



C



D



Arg-ACG-1-2