frontiers in Genetics

METHODS
published: 05 October 2018
doi: 10.3389/fgene.2018.00461

TEsmall Identifies Small RNAs Associated With Targeted Inhibitor Resistance in Melanoma

Kathryn O'Neill[1†], Wen-Wei Liao[2†], Ami Patel[3] and Molly Gale Hammell[1]*

[1] Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, United States, [2] McDonnell Genome Institute, School of Medicine, Washington University in St. Louis, St. Louis, MO, United States, [3] Icahn School of Medicine at Mount Sinai, New York City, NY, United States

MicroRNAs (miRNAs) are small 21–22 nt RNAs that act to regulate the expression of mRNA target genes through direct binding to mRNA targets. While miRNAs typically dominate small RNA (sRNA) 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 sRNA classes remains unclear, largely because these sRNAs are typically analyzed separately. Here, we present TEsmall, a tool that allows for the simultaneous processing and analysis of sRNAs 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 analysis. Several interactive charts are also produced to look at overall distributions in length and annotation classes. We next applied the TEsmall pipeline to sRNA 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 sRNA 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.

Keywords: NGS software tools, small RNA and microRNA, transposon activity, melanoma, computational biology

OPEN ACCESS

Edited by:
Hervé Seitz,
Centre national de la recherche scientifique (CNRS), France

Reviewed by:
Marius van den Beek,
Institut Curie, France
Elsa Bernard,
Memorial Sloan Kettering Cancer Center, United States

****Correspondence:***
Molly Gale Hammell
mhammell@cshl.edu

† These authors have contributed equally to the work

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *29 June 2018*
Accepted: *20 September 2018*
Published: *05 October 2018*

Citation:
O'Neill K, Liao W-W, Patel A and Hammell MG (2018) TEsmall Identifies Small RNAs Associated With Targeted Inhibitor Resistance in Melanoma. Front. Genet. 9:461. doi: 10.3389/fgene.2018.00461

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. While typically derived from short hairpin RNA precursors located in both intergenic and intronic regions, miRNAs can also be processed from other ncRNAs including tRNAs and spliced intron lariats (Schorn et al., 2017;

Bartel, 2018). In addition to governing development, small RNAs (sRNAs) play a critical role in repressing transcripts derived from repetitive regions of the genome. In animals, siRNAs and piRNAs function to repress transposons in somatic cells, and the germline, respectively (Ghildiyal et al., 2008; Castañeda et al., 2011). 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 DESeq2 (Love et al., 2014).

We have tested the ability of TEsmall to characterize the expression profiles of sRNAs 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 tumors (Hodis et al., 2012), 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 6 months or less, as the tumors rapidly develop resistance to these targeted therapies (Villanueva et al., 2010). While some tumors resistant to BRAF inhibitors acquire additional genetic lesions that elevate MAPK or AKT signaling (Alcala and Flaherty, 2012), many therapy-resistant cell lines establish resistance without a clearly understood mechanism of resistance (Gatenby and Brown, 2018). Changes to sRNA 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 sRNA alterations might contribute to the development of resistance to BRAF inhibitors in 451Lu melanoma cells that carry BRAFV600E mutations, we undertook a sRNA 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 1 μM 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. (2010) 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 $\sim 1 \times 10$ (Hodis et al., 2012) 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 sRNA sequencing libraries. The sRNA 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 6 nt Illumina barcodes, such that six libraries could be sequenced per lane on an Illumina Genome Analyzer IIx. The reads were sequenced as single-end 50 bp 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 **Supplementary File S1**.

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 sRNA. 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 as described in the Thermo Fisher Scientific TaqMan MicroRNA Assay protocol, available from the manufacturer. Briefly, 10 ng of total RNA was used as the input to a microRNA specific reverse transcription (RT) assay using the TaqMan MicroRNA Reverse Transcription Kit and an RT primer specific to the miRNA of interest. Next, qPCR amplification was performed using 1.33 μL of the output from the RT reaction, TaqMan Universal PCR Master Mix II no UNG, and the TaqMan Small RNA Assay Kit specific to the miRNA of interest. Each sRNA specific assay was performed with input RNA from two biological replicates of the 451Lu-Par and 451Lu-BR cell lines, with three technical replicates per biological replicate, on a Thermo Fisher StepOnePlus Real-Time PCR System. 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 $2^{\Delta-\Delta-Ct}$, 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 package (Martin, 2011). 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 two mismatches and filters a single alignment per read specified by the option: bowtie -v 2 -k 1 using bowtie (v1.2.1) (Langmead, 2010). sRNA 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 is annotated via a sequential decision tree, as follows. The alignments are distributed 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 annotated piRNA clusters. This process is depicted in **Supplementary Figure S1**. This annotation class priority can be re-ordered by the user to suit the application and user preferences. An HTML output file is then created using python based Bokeh tools (Bokeh Development Team, 2014) to visualize the abundance distributions, length distributions, and mapping logs of all sRNAs in the dataset (**Figure 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 DESeq2 (Love et al., 2014) for differential analysis. The abundance calculations for these counts files are $1/n$ normalized at the end of this annotation process, where n represents the number of alignments per read, to ensure no double-counting of multimappers. Additional packages employed within the TEsmall workflow include bedtools (Quinlan and Hall, 2010), pandas (McKinney, 2010), samtools (Li et al., 2009), pybedtools (Dale et al., 2011), and scipy (Jones et al., 2001).

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

Annotation files for the human (hg19) and fly (dm3) genomes can be found at the following location: http://labshare.cshl.edu/shares/mhammelllab/www-data/TEsmall/.

Instructions for creating annotation files for additional genomes can be found in **Supplementary File S4**.

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

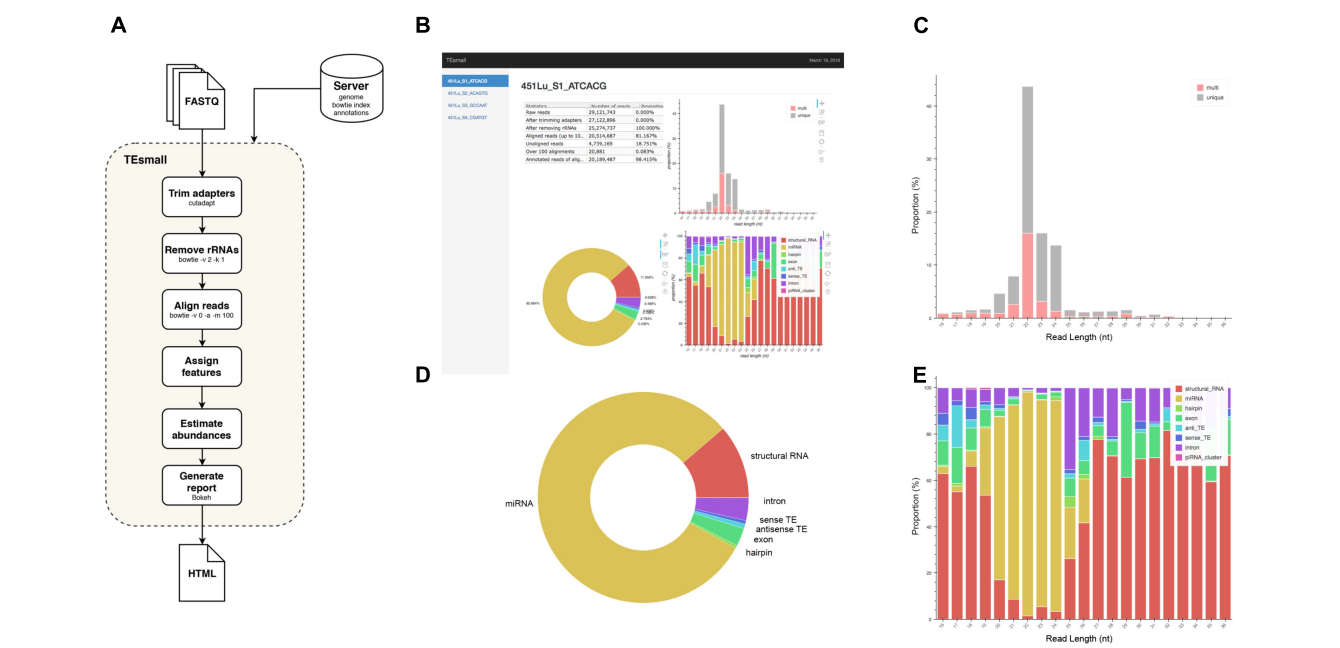

FIGURE 1 | Flow chart and output HTML of TEsmall. **(A)** Flow chart of TEsmall's treatment of input high-throughput sequencing data, input genome indices, and output. **(B)** Screenshot of HTML output file for one sample. **(C)** Bar plot depicting size distribution of unique and multimapper reads.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

statistical analysis. For downstream visualization, the counts were normalized by the built-in variance stabilizing transformation (VST) method in DESeq2. sRNAs with an adjusted p-value < 0.05 were considered statistically significant. The full DESeq2 output file is given in **Supplementary File S2**.

Visualization

Figures were produced using the R packages ggplot2 (Wickham, 2009), gplots (Warnes et al., 2016) and GenomicRanges (Lawrence et al., 2013) for scatterplots, heatmaps, and wiggleplots, respectively. Python package matplotlib (Hunter, 2007) was used for all barplots. RNA secondary structures were rendered using the forna webtool (Kerpedjiev et al., 2015), 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 (Gruber et al., 2008).

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 sRNAs on a common scale in a single pipeline. This serves as a novel improvement to currently available software such as mirDeep2 (Friedländer et al., 2012) and piPipes (Han et al., 2015), which are optimized for the analysis of miRNAs and piRNAs, respectively, but are not equipped to evaluate both types of sRNAs together. TEsmall is also designed so that its output is optimally formatted for downstream differential analysis with statistical modeling software, such as DEseq2 (Anders and Huber, 2010). A flowchart describing the entire TEsmall workflow is given in **Figure 1A**, with example output charts given in **Figures 1B–E**. Specifically, in the first module of TEsmall, raw sRNA 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 bowtie (Langmead, 2010) 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 sRNA 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 outputted 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 (**Figure 1B**) summarizing distribution of read lengths (**Figure 1C**), proportion of reads assigned to each sRNA type (**Figure 1D**), and distribution of reads of a particular size to each of the sRNA categories (**Figure 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 in two biological replicates of parental and BRAF inhibitor resistant cell lines. Resistant lines were derived through exposure of 451Lu patient derived parental cell lines to increasing concentrations of vemurafenib up to 1 μM. Resistant clones were selected and expanded before exposure to an increase in vemurafenib. Cells were otherwise treated as described in Villanueva et al. (2010) 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 (**Figure 2**). Overall, there appears to be a trend toward lower expression of many sRNA classes in the 451Lu BRAF resistant samples, with more down-regulated than up-regulated species for most classes of sRNAs (**Figures 2**, **3B**). However, upon testing by a two tailed Welch's t-test only structural RNAs were shown to be significantly downregulated as a class in BRAF inhibitor resistant libraries ($P < 3.52 \times 10^{-13}$). Upon filtering for the most abundant (base mean across all replicates > 500) and significantly differentially expressed transcripts (multiple-hypothesis testing adjusted p-value < 0.05), trends of lower sRNA expression in the BRAF resistant samples was still seen for many intronic, exonic, and transposon mapped sRNA species (**Figure 3**). Interestingly, miRNAs show an even distribution of species with negative and positive log fold changes, and since miRNAs

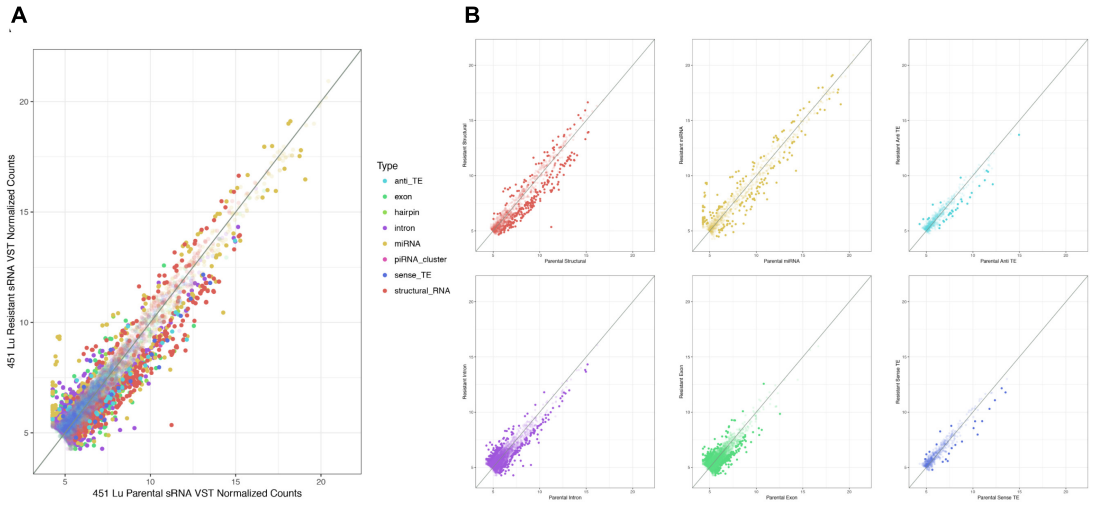

FIGURE 2 | Scatterplots depicting 451Lu BRAF resistant versus parental RNA: VST normalized counts. **(A)** Overlay of all subtype scatterplots. **(B)** RNA subtype specific scatterplots. Transparent points represent RNA species with adjusted p-value < 0.05.

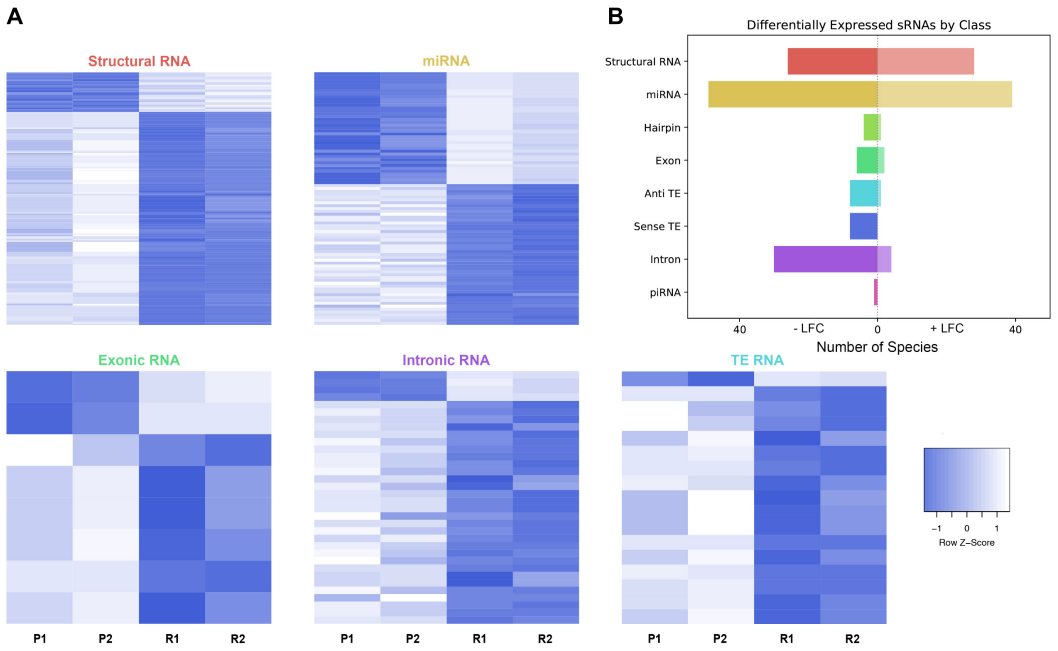

FIGURE 3 | Differential expression analysis of highly abundant sRNAs. **(A)** Heatmaps depicting all significantly differentially expressed sRNAs per subtype. All heatmaps are row-scaled to lie between –1 and 1, based on the VST normalized count values. P1, P2, R1, and R2 represent 451Lu parental replicates 1 and 2 and 451Lu BRAF Resistant replicates 1 and 2, respectively. **(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 for duplicate tRNAs.

were the most abundantly sequenced sRNAs in the libraries, this rules out a normalization issue as the explanation for down-regulated sRNAs in the other classes. It may be of interest that, after filtering for significance ($P < 0.05$) and abundances greater than 500 reads per million mapped (RPM), structural RNAs with a negative log fold change are almost exclusively tRNAs 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 **Supplementary File S2**.

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 de-repression of Notch in normal human keratinocytes and murine epidermis (Nagosa et al., 2017). 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 pathways (Pinnix and Herlyn, 2007). Our data showed an approximate fivefold increase in miR-184 expression in BRAF inhibitor resistant cells relative to parental (**Figure 4A** and **Supplementary File S2**), consistent with a model where MAPK pathway activation provides a mechanism for BRAF inhibitor resistance (Villanueva et al., 2010, 2011). 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 (Van Allen et al., 2014). 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 (Golan et al., 2015). 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 (Leite et al., 2013).

To validate the expression profiles from our TEsmall 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 sRNA-seq dataset (**Figure 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. (2018) following induction of BRAF inhibitor resistance.

Upon further investigation into individual RNA species from different subtypes for follow up, we encountered an interesting and novel 21 nucleotide sRNA derived from the sixth intron in the vimentin gene (VIM) (**Figure 4B**). Intronic microRNAs can either derive from their own precursor transcripts dubbed pri-miRNAs or can alternately derive from short spliced introns with internal hairpin structures dubbed miRtrons (Okamura et al., 2007). MiRtrons are typically generated via the splicing machinery from short introns (∼100 nt) and subsequently processed by DICER in the cytoplasm, bypassing the canonical nuclear DROSHA processing steps. This is in contrast to intronically located miRNAs that derive from their own precursor pri-miRNAs and are dependent upon both DROSHA and DICER for processing. As visible in the minimum free energy secondary structure prediction by RNAfold, the candidate miRNA of interest is located in a stem loop structure which appears conducive to processing by DICER (**Figure 4C**). The length of this mature sRNA (21 nt), the short length of its host intron (350 nt), its abundance as a single RNA species, and its secondary structure within VIM intron 6 could all be consistent with a miRNA derived from either an intronic pri-miRNA or as a miRtron. 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 miRNA, suggesting this candidate VIM miRNA might represent a novel sRNA with particular abundance in melanoma cells.

In addition to miRNAs, TEsmall recognizes several other types of sRNAs. It has been previously reported that tRNA derived sRNA 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 (Mak and Kleiman, 1997; Schorn et al., 2017). Through TEsmall, 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 (**Figure 2**). Upon further investigation we were able to determine that a subset of these reads

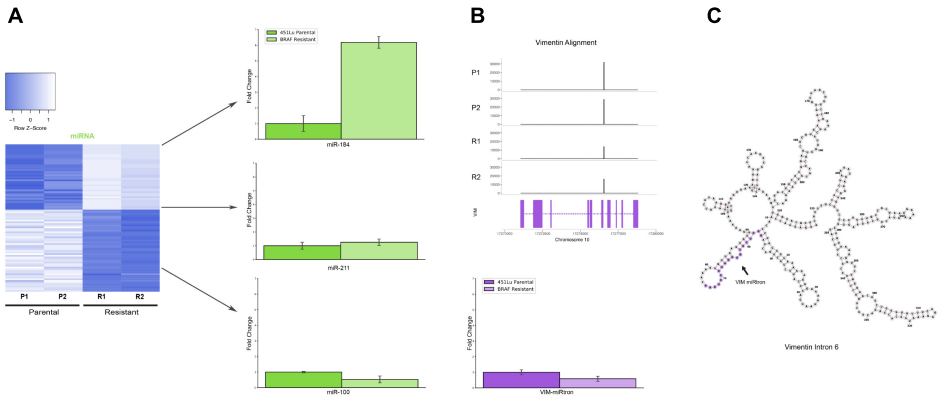

FIGURE 4 | Detailed analysis miRNAs of interest. **(A)** qPCR representing fold change of miRNAs miR-184, miR-211, and miR-100 in 451Lu BRAF Resistant samples relative to 451Lu parental expression levels across replicates. **(B)** qPCR representing log fold change of the VIM miRNA in 451Lu BRAF Resistant vs. Parental samples and BAM gene alignment tracks across samples. **(C)** RNAfold predicted secondary structure of VIM intron 6 with the candidate miRNA highlighted in purple.

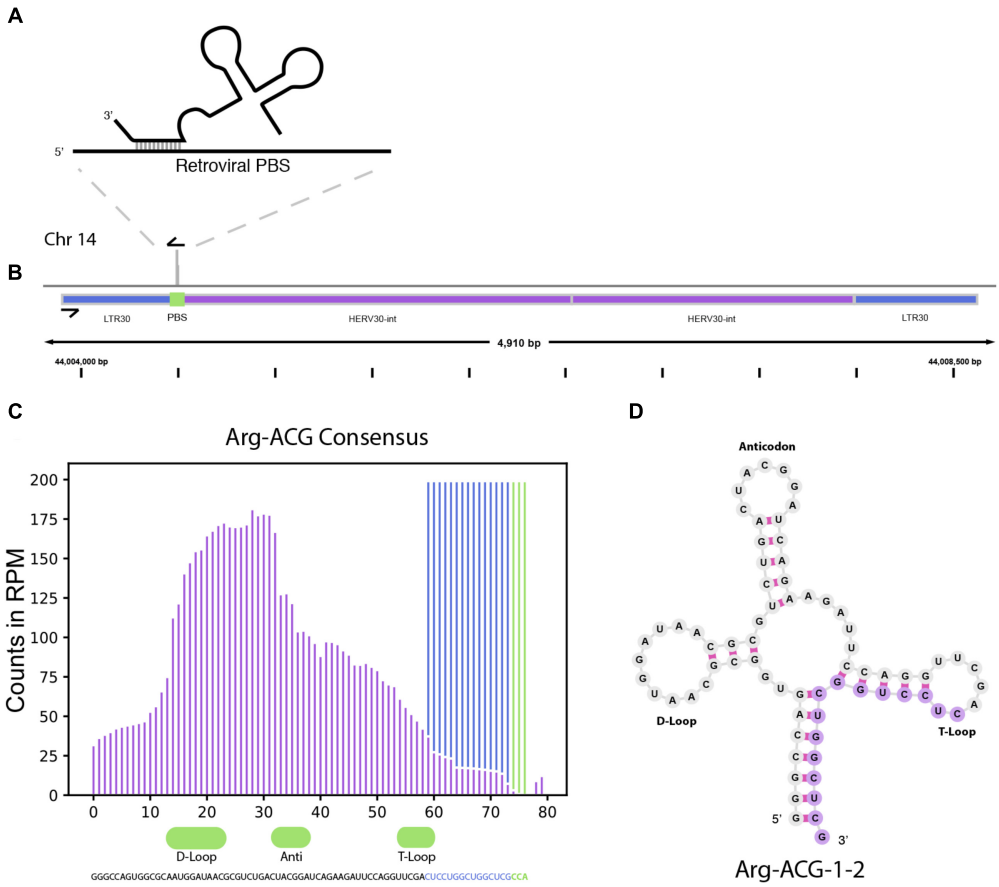

FIGURE 5 | tRNA and TE interaction through primer binding sites. **(A)** Diagram of primer binding by tRNA to facilitate retroviral reverse transcription. **(B)** Read alignment track of Arg-CGY family derived 18 nt CCA tailed fragment to HERV30 PBS. **(C)** Consensus histogram of reads distributed from Arg-CGY tRNAs, and derived 3′ tRFs. **(D)** Secondary structure of Arg-CGY family member Arg-ACG-1-2 with highlighted 15 nt CCA (-) fragment.

were tRFs derived from the Arg-CGY family of tRNAs (**Figure 5B**). These candidate tRFs mapped to a subset of HERVs including HERV3, HERV30, MER51, and others (see **Supplementary File S3**). This is consistent with previous literature showing HERV-R type retrotransposons are primed by Arg tRNAs (**Figure 5A**) and that tRNA derived fragments can occupy retroviral PBSs to suppress transposon activity (Mak and Kleiman, 1997; Schorn et al., 2017). It is important to note that the Arg-CGY sRNAs reported by TEsmall are consistent with the tRFs previously described in Schorn et al. (2017) as they are 18 nt CCA-appended fragments originating from the 3′ T-arm of tRNAs. This is shown graphically in **Figure 5**, where the pileup of reads at an example HERV PBS locus can be seen in **Figure 5B**, and the pileup of these same reads at the originating tRNA locus can be seen **Figures 5C,D**. In the tRNA profiles, other tRNA fragments including tRNA derived stress-induced RNAs (tiRNAs) can be seen outside of the tRF generating 3′ end, but these do not predominantly accumulate as a single abundant sRNA species.

In addition to the miRNAs and tRFs highlighted above, several additional species of sRNAs were reported by TEsmall 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 sRNAs can be found in **Supplementary File S2**.

Comparison of TEsmall With sRNA Analysis Software

Several software packages exist to characterize sRNA data for expression profiling analysis. However, programs designed for this purpose such as miRDeep2 (Friedländer et al., 2012), ShortStack (Axtell, 2013), Chimira (Vitsios and Enright, 2015), sRNAtoolbox (Rueda et al., 2015), and Oasis 2 (Rahman et al., 2018) typically focus on a particular category of sRNA, predominantly miRNAs. Several packages also consider multiple sRNA types including piPipes (Han et al., 2015), omiRas (Müller et al., 2013), and unitas (Gebert et al., 2017) which include analysis of other non-coding RNAs. However, the output formats of these packages do not lend themselves to easy application of statistical analysis tools like DEseq2 for downstream use

and manipulation. 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 tRFs. PiPipes provides 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. TEsmall, which does not perform piRNA-specific ping-pong analysis, provides a complementary package that is designed to be a general purpose sRNA analysis suite to identify and analyze many types of sRNAs concurrently, presenting the output in a format intended for expression profiling analysis. As piPipes output was not directly comparable to TEsmall output, we performed quantitative comparisons between TEsmall and miRDeep2 output, as seen in **Supplementary Figure S2**. TEsmall and miRDeep2 preformed comparably with differences originating from higher stringency in TEsmall annotation. This stringency caused fewer reads to be mapped to miRNAs by the TEsmall pipeline. Some reads were attributed by TEsmall to rRNAs and discarded, while others were not attributed to the respective miRNA loci due to mismatched nucleotides within the miRNA. Users interested in the possibility of A-I editing, or other sources of mismatched alignments, may optionally choose to allow mismatches during TEsmall alignment to capture these reads. Output of TEsmall and miRDeep2 annotation was found to be highly comparable with Pearson correlation coefficients of DEseq2 normalized miRNA counts between parental and resistant libraries of 0.882 and 0.910, respectively. Following differential expression analysis by DEseq2 of the TEsmall and miRDeep2 outputs, the Pearson correlation coefficient of $\log_2$ fold change values was 0.867. Finally, the novel VIM-encoded miRNA was not captured in the miRDeep2 output. This analysis supports TEsmall as comparable to class-specific sRNA expression analysis packages, while providing information on a wider source of sRNAs.

DISCUSSION

TEsmall 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 (Friedländer et al., 2012; Axtell, 2013) or piRNAs (Han et al., 2015) 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 TEsmall is formatted for direct integration into downstream analysis pipelines. TEsmall's output files are compatible with statistical analysis software like DESeq2 and efficient heatmap generation. In addition to requiring little data preprocessing, TEsmall outputs an aesthetic HTML file of charts (**Figure 1B**) which allows for fast and effortless assessment of library quality, sRNA composition, and size distribution. TEsmall 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 TEsmall 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 **Supplementary File S2**. 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, TEsmall detected several novel classes of sRNAs which would not have been found using packages designed for miRNA analysis. TEsmall allows the user to investigate tRNA derived fragments which have been shown to play a critical role in LTR retro-transposon suppression (Schorn et al., 2017). In the melanoma dataset, we identified a novel candidate tRF that appears to derive from ARG-tRNAs and to potentially regulate several HERV-R type LTR elements through occupancy of the PBS. 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 that small non-coding RNAs of different subtypes types 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 (Tam et al., 2008). 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 (Hakim et al., 2008; Zlotorynski, 2014). 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 TEsmall, an easy to use package with aesthetic output designed for the concurrent expression analysis of multiple sRNA subtypes.

AUTHOR CONTRIBUTIONS

KO and MH analyzed the data, generated the figures, and wrote the manuscript. MH and AP designed the experiments and generated the data. W-WL wrote the code for the TEsmall pipeline, packaged the code for GitHub, and contributed to the writing of the manuscript.

ACKNOWLEDGMENTS

MH is a scholar of the Rita Allen Foundation. KO acknowledges funding from the NIH training grant 2T32GM065094-16. MH and W-WL were supported by a grant from the NIH/NINDS: 5R21NS088449. We would also like to acknowledge the Meenhard Herlyn lab for help, advice, and for the kind gift of the 451Lu cells.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00461/full#supplementary-material

FIGURE S1 | Flow chart of the default sequential decision tree used by TEsmall to assign annotations. Alignments are assigned to each category in the indicated order and, if annotated, are removed from the pool before preceding to the next annotation category. Users may opt to re-order the priority table.

FIGURE S2 | Scatterplots comparing TEsmall and miRDeep2 miRNA abundance quantification. Mean abundances between biological replicates **(A,B)** and fold change between conditions **(C)** were calculated with DEseq2 on the count tables output by each software package. Low abundance miRNAs with fewer than 2,000 counts across all samples are marked as transparent. Shown in pink are the miRNAs validated by qPCR in **Figure 4**. **(A)** Log scaled comparison of 451Lu-Par normalized miRNA counts of TEsmall versus miRDeep2, with a correlation coefficient of $r = 0.882$. **(B)** Log scaled comparison of 451Lu-BR miRNA counts of TEsmall versus miRDeep2, with a correlation coefficient of $r = 0.910$. **(C)** Comparison of $\log_2$ fold change as reported by TEsmall and miRDeep2, with a correlation coefficient of $r = 0.867$.

FILE S1 | Table of library composition and mapping rates across all replicate small RNA libraries.

FILE S2 | Table including the raw output from DEseq2 and several worksheets of filtered differentially expressed sRNAs used in constructing the heatmaps in **Figure 3**.

FILE S3 | List of long terminal repeat flanked transposable elements mapped to the Arg-ACG 3′ tRF and their chromosomal start site in the hg19 genome.

FILE S4 | Instructions for creating new annotation files for additional species. Instructions for customizing the path for stored annotation files.

REFERENCES

Alcala, A. M., and Flaherty, K. T. (2012). BRAF inhibitors for the treatment of metastatic melanoma: clinical trials and mechanisms of resistance. *Clin. Cancer Res.* 18, 33–40. doi: 10.1158/1078-0432.CCR-11-0997

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, 1–12. doi: 10.1186/gb-2010-11-10-r106

Axtell, M. J. (2013). ShortStack: comprehensive annotation and quantification of small RNA genes ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA* 19, 740–751. doi: 10.1261/rna.035279.112

Bartel, D. P. (2018). Metazoan MicroRNAs. *Cell* 173, 20–51. doi: 10.1016/j.cell.2018.03.006

Bokeh Development Team (2014). *Bokeh: Python Library for Interactive Visualization*. Wichita, KS: Bokeh Development Team.

Castañeda, J., Genzor, P., and Bortvin, A. (2011). piRNAs, transposon silencing, and germline genome integrity. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* 714, 95–104. doi: 10.1016/j.mrfmmm.2011.05.002

Dale, R. K., Pedersen, B. S., and Quinlan, A. R. (2011). Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* 27, 3423–3424. doi: 10.1093/bioinformatics/btr539

Díaz-Martínez, M., Benito-Jardon, L., Alonso, L., Koetz-Ploch, L., Hernando, E., and Teixido, J. (2018). miR-204-5p and miR-211-5p contribute to BRAF inhibitor resistance in melanoma. *Cancer Res.* 78, 1017–1030. doi: 10.1158/0008-5472.CAN-17-1318

Friedländer, M. R., MacKowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52. doi: 10.1093/nar/gkr688

Gatenby, R., and Brown, J. (2018). The evolution and ecology of resistance in cancer therapy. *Cold Spring Harb. Perspect. Med.* 8, 1–12. doi: 10.1101/cshperspect.a033415

Gebert, D., Hewel, C., and Rosenkranz, D. (2017). Unitas: the universal tool for annotation of small RNAs. *BMC Genomics* 18:644. doi: 10.1186/s12864-017-4031-9

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., and Lee, S. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* 320, 1077–1081. doi: 10.1126/science.1157396

Golan, T., Messer, A. R., Amitai-Lange, A., Melamed, Z., Ohana, R., Bell, R. E., et al. (2015). Interactions of melanoma cells with distal keratinocytes trigger metastasis via notch signaling inhibition of MITF. *Mol. Cell* 59, 664–676. doi: 10.1016/j.molcel.2015.06.028

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. (2008). The Vienna RNA websuite. *Nucleic Acids Res.* 36, 70–74. doi: 10.1093/nar/gkn188

Hakim, S. T., Alsayari, M., McLean, D. C., Saleem, S., Addanki, K. C., and Aggarwal, M. (2008). A large number of the human microRNAs target lentiviruses, retroviruses, and endogenous retroviruses. *Biochem. Biophys. Res. Commun.* 369, 357–362. doi: 10.1016/j.bbrc.2008.02.025

Han, B. W., Wang, W., Zamore, P. D., and Weng, Z. (2015). PiPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome-and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics* 31, 593–595. doi: 10.1093/bioinformatics/btu647

Hodis, E., Watson, I. R., Kryukov, G. V., Arold, S. T., Imielinski, M., and Theurillat, J. (2012). A landscape of driver mutations in melanoma. *Cell* 150, 251–263. doi: 10.1016/j.cell.2012.06.024

Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55

Jones, E., Oliphant, E., and Peterson, P. (2001). *SciPy: Open Source Scientific Tools for Python.* Available at: http://www.scipy.org/

Kerpedjiev, P., Hammer, S., and Hofacker, I. L. (2015). Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. *Bioinformatics* 31, 3377–3379. doi: 10.1093/bioinformatics/btv372

Langmead, B. (2010). Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinforma* CHAPTER 11:Unit 11.7. doi: 10.1002/0471250953.bi1107s32

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., and Gentleman, R. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9:e1003118. doi: 10.1371/journal.pcbi.1003118

Leite, K., Morais, D., Reis, S., Viana, N., Moura, C., Florez, M., et al. (2013). MicroRNA 100: a context dependent miRNA in prostate cancer. *Clinics* 68, 797–802. doi: 10.6061/clinics/2013(06)12

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21. doi: 10.1186/s13059-014-0550-8

Mak, J., and Kleiman, L. (1997). Primer tRNAs for reverse transcription. *J. Virol.* 71, 8087–8095.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200

McKinney, W. (2010). "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, Austin, TX.

Müller, S., Rycak, L., Winter, P., Kahl, G., Koch, I., and Rotter, B. (2013). omiRas: a Web server for differential expression analysis of miRNAs derived from small RNA-Seq data. *Bioinformatics* 29, 2651–2652. doi: 10.1093/bioinformatics/btt457

Nagosa, S., Leesch, F., Putin, D., Bhattacharya, S., Altshuler, A., Serror, L., et al. (2017). microRNA-184 induces a commitment switch to epidermal differentiation. *Stem Cell Rep.* 9, 1991–2004. doi: 10.1016/j.stemcr.2017.10.030

Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., and Lai, E. C. (2007). The mirtron pathway generates microRNA-Class regulatory RNAs in *Drosophila*. *Cell* 130, 89–100. doi: 10.1016/j.cell.2007.06.028

Pinnix, C. C., and Herlyn, M. (2007). The many faces of Notch signaling in skin-derived cells. *Pigment Cell Res.* 20, 458–465. doi: 10.1111/j.1600-0749.2007.00410.x

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Rahman, R. U., Gautam, A., Bethune, J., Sattar, A., Fiosins, M., Magruder, D. S., et al. (2018). Oasis 2: improved online analysis of small RNA-seq data. *BMC Bioinformatics* 19:54. doi: 10.1186/s12859-018-2047-z

Rueda, A., Barturen, G., Lebrón, R., Gómez-Martín, C., Alganza, Á, Oliver, J. L., et al. (2015). SRNAtoolbox: an integrated collection of small RNA research tools. *Nucleic Acids Res.* 43, W467–W473. doi: 10.1093/nar/gkv555

Schorn, A. J., Gutbrod, M. J., LeBlanc, C., and Martienssen, R. (2017). LTR-Retrotransposon control by tRNA-Derived small RNAs. *Cell* 170, 61–71. doi: 10.1016/j.cell.2017.06.013

Tam, O. H., Aravin, A. A., Stein, P., Girard, A., Murchison, E. P., Cheloufi, S., et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453, 534–538. doi: 10.1038/nature06904

Van Allen, E. M., Wagle, N., Sucker, A., Treacy, D. J., Johannessen, C. M., Goetz, E. M., et al. (2014). The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. *Cancer Discov.* 4, 94–109. doi: 10.1158/2159-8290.CD-13-0617

Villanueva, J., Vultur, A., and Herlyn, M. (2011). Resistance to BRAF inhibitors: unraveling mechanisms and future treatment options. *Cancer Res.* 71, 7137–7141. doi: 10.1158/0008-5472.CAN-11-1243

Villanueva, J., Vultur, A., Lee, J. T., Somasundaram, R., Cipolla, A. K., Wubbenhorst, B., et al. (2010). 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* 18, 683–695. doi: 10.1016/j.ccr.2010.11.023

Vitsios, D. M., and Enright, A. J. (2015). Chimira: analysis of small RNA sequencing data and microRNA modifications. *Bioinformatics* 31, 3365–3367. doi: 10.1093/bioinformatics/btv380

Warnes, G. R., Bolker, B., and Lumley, T. (2016). *gplots: Various R programming tools for plotting data. R package version 2.6.0 n.d.*

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag. doi: 10.1007/978-0-387-98141-3

Zlotorynski, E. (2014). RNA interference: MicroRNAs suppress transposons. *Nat. Rev. Mol. Cell Biol.* 15, 298–299. doi: 10.1038/nrm3788

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 O'Neill, Liao, Patel and Hammell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.