miRquant 2.0: an Expanded Tool for Accurate Annotation and Quantification of MicroRNAs and their isomiRs from Small RNA-Sequencing Data

Matthew Kanke¹, Jeanette Baran-Gale¹, Jonathan Villanueva¹,², and Praveen Sethupathy¹,³,*

¹Department of Genetics,
²Post-baccalaurate Research Education Program,
³Lineberger Comprehensive Cancer Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Summary

Small non-coding RNAs, in particular microRNAs, are critical for normal physiology and are candidate biomarkers, regulators, and therapeutic targets for a wide variety of diseases. There is an ever-growing interest in the comprehensive and accurate annotation of microRNAs across diverse cell types, conditions, species, and disease states. High-throughput sequencing technology has emerged as the method of choice for profiling microRNAs. Specialized bioinformatic strategies are required to mine as much meaningful information as possible from the sequencing data to provide a comprehensive view of the microRNA landscape. Here we present miRquant 2.0, an expanded bioinformatics tool for accurate annotation and quantification of microRNAs and their isoforms (termed isomiRs) from small RNA-sequencing data. We anticipate that miRquant 2.0 will be useful for researchers interested not only in quantifying known microRNAs but also mining the rich well of additional information embedded in small RNA-sequencing data.

1 Introduction

Non-coding RNAs (ncRNAs) have long been recognized as important regulators of cellular function [1–4]. For example, ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) are essential for protein synthesis in all living organisms. However, recent advances in next-generation sequencing technology have helped unveil the prevalence and abundance of a diverse array of intracellular and extracellular ncRNAs [5], [6]. There are roughly three main classes of ncRNAs, which are somewhat arbitrarily differentiated by length: small (<40 nucleotides), intermediate (between 40 and 200 nucleotides), and long (>200 nucleotides). Increasingly well-studied examples of ncRNA types within these classes are microRNAs (miRNAs), cajal body specific RNAs (caRNAs), and long, intergenic non-coding RNAs (lincRNAs), respectively. The small RNA category [7] especially has experienced an explosion in the number of different types of ncRNAs, including, though not limited to, piwi-interacting small RNAs (piRNAs), tRNA-derived small RNAs (tDRs), Y-RNA derived small RNAs (yDRs), endogenous small interfering RNAs (endo-siRNAs), trans-acting small...
interfering RNAs (tasiRNAs), QDE-2 interacting small RNAs (qiRNAs), and termini-associated short RNAs (TASRs). The level of experimental support for the molecular and biological functions of these recently discovered small RNAs varies according to the type. For example, the roles of piRNAs [8] in germline development are much better understood than the proposed functions of yDRs [9] (also referred to as Y-RNA fragments) in tumor proliferation and metastasis. The most thoroughly studied class of small RNAs to date is miRNAs.

MicroRNAs are small RNAs (~18-24 nucleotides in length) that regulate intracellular gene expression across all eukaryotes [10]. Their roles in gene regulation are important in a wide variety of fundamental biological processes such as development, growth, and metabolism, as well as in many organ-specific functions. MicroRNAs often serve as fine tuners of gene expression to maintain intracellular and systemic homeostasis [11–13]. Defects in miRNA-mediated gene regulation have been implicated in many disorders, including Mendelian conditions such as deafness [14], a plethora of different complex diseases [15] including cancer [16], [17], and host-pathogen interactions [18–20]. Recently, miRNAs have emerged as candidate circulating biomarkers [21] and potential therapeutic targets [22], [23] for several diseases, suggesting that the study of miRNAs will not only advance understanding of basic biology and evolution, but also may promote the development of novel clinical strategies for diagnosis, prognosis, and treatment of complex diseases.

The biological and biomedical relevance of miRNAs has led to an ever-growing interest in high-throughput, accurate annotation and quantification of miRNAs in functionally distinct organ systems and cell types across many different species. The first major revolution in this research area occurred in the early 2000s with the use of microarray technology to detect miRNAs [24], [25]. Within the past ten years though, next-generation sequencing technology has replaced microarrays as the standard approach for high-throughput miRNA detection. Small RNA sequencing (smRNA-seq) approaches ameliorate many of the limitations of microarrays; however, they are not without their own challenges [26]. Some of the issues lie within the experimental realm; for example, mitigating sample contamination [27] or preparing small RNA libraries in a manner that faithfully captures all of the small RNAs present in the sample without bias [26]. Other issues pertain to the bioinformatic strategies for extracting accurate information out of the raw smRNA-seq data.

In 2013 we developed a bioinformatic strategy [28], which we now call miRquant, for accurate detection of miRNAs from smRNA-seq data, with an emphasis on separately annotating and quantifying functionally-distinct isoforms of canonical miRNAs, termed isomiRs. We have since expanded the tool significantly to miRquant 2.0. In addition to the quantification of miRNA and isomiR expression levels, we now provide additional detailed information on the quality of the sequencing data, genome mapping statistics, abundance of other types of small RNAs such as tDRs [29] and yDRs [9], prevalence of post-transcriptional modifications such as A-to-I edits and 3’ non-templated nucleotide additions, and correlation in miRNA profiles across multiple samples. Furthermore, miRquant 2.0 is now equipped to handle smRNA-seq data from human, mouse, and rat, and is currently being expanded to fruit fly. We make publicly available both the tool and a detailed tutorial for users to learn how to run the program and interpret the results. Below we provide an overview of the miRquant workflow and the diverse capabilities of the tool.

doi:10.2390/biecoll-jib-2016-307
2 Related works

miRquant 2.0 is an expansion of the original version of miRquant developed in 2013 [28]. The new version has added functionalities and improved performance (see above).

3 Architecture and implementation

The overall work flow of miRquant 2.0 is provided in Figure 1. Each of the major steps is outlined below. For further details, we refer the reader to the online manual and tutorial: https://github.com/Sethupathy-Lab/miRquant.

![Figure 1: Overview of miRquant 2.0. Major steps of the miRquant 2.0 workflow are shown. Cutadapt is used for trimming partial or full adapter sequences from the 3'-end of small RNA sequencing reads. Those reads that do not have a specified length of adapter sequence at the 3'-end, or that entirely consist of adapter sequences, are discarded. Bowtie and SHRiMP are used for the two-tier mapping process. Unaligned reads after the SHRiMP step are discarded. Annotation is provided for miRNAs, tRNA-derived RNAs (tDRs), and Y-RNA derived RNAs (yDRs). All other mapped loci are grouped into the unannotated category. Normalized expression levels are provided as reads per million mapped (RPMM), and as reads per million mapped to miRNAs (RPMMM) for miRNAs.](http://journal.imbio.de/)
3.1 3’ Adapter trimming

In smRNA-seq, the number of sequenced bases (generally ~50 on the HiSeq 2000 and more recent platforms) exceeds the length of the target RNA fragment. Therefore, part of the 3’ adapter is sequenced and requires removal prior to genome alignment. To accomplish this, miRquant 2.0 utilizes the adapter trimming program Cutadapt (version 1.12) [30]. Specifically, miRquant 2.0 requires a minimum of ten nucleotides overlap between the adapter sequence and the 3’-end of the sequencing read with less than 10% errors in the alignment. Reads that do not meet this criterion, or are <14 nucleotides in length following trimming, are discarded. Conversely, reads that meet this criterion are subject to trimming and retained for further analysis.

3.2 Read alignment – tier 1

Alignment of reads to the reference genome is accomplished by a two-tier process. In the first-tier, Bowtie (version 1.1.0) [31] is used to align only those reads that match perfectly to the genome. If a read aligns equally well to multiple genomic loci, it is proportionally assigned to each of the loci. Following this mapping step, miRquant 2.0 assembles contigs or ‘genomic windows’ that contain >=1 perfectly mapped reads. If genomic windows are within 65 nucleotides of each other, these windows are merged into larger ‘super genomic windows.’ This step is intended to ensure that all reads potentially related to a single pre-miRNA (roughly 65 to 110 nucleotides in length) are contained within the same window.

3.3 Read alignment – tier 2

In the second-tier, the SHRiMP aligner (version 2.2.2) [32] is used to align reads, with mismatches allowed, only to the genomic windows generated in tier 1. The types, locations, and number of mismatches allowed are defined in accordance with known patterns of post-transcriptional editing of miRNAs, most notably, A-to-I editing and 3’ non-templated nucleotide addition. For example, the number of mismatches allowed at the 3’-end of a miRNA-related read ranges from 0 to 3, depending on the length of the read (with read length as r; 0 if 14 <= r <= 15, 1 else if 16 <= r <= 19, 2 else if 20 <= r <= 23, 3 if r > 23). Similar to the Bowtie step, reads aligning equally well to multiple locations are proportionally assigned to each of those loci. The two-tier mapping process is intended to mitigate two common issues with bioinformatics analyses of smRNA-seq data:

(1) Computational tractability – Running the SHRiMP aligner on the entire genome/transcriptome would be computationally extremely expensive, particularly for short reads such as those from smRNA-seq

(2) Mapping ambiguity – Mapping short reads, particularly those for which several genomic mismatches are allowed in order to accommodate post-transcriptional editing of miRNAs, can lead to highly non-specific, ambiguous mapping. That is, some reads may map to numerous genomic locations because they happen to harbor similar enough stretches of sequence just by chance. Some of these loci may have little to do with miRNAs

These two concerns are effectively ameliorated by restricting the mappable space for SHRiMP to only the genomic windows generated in tier 1.
3.4 Annotation of microRNAs and other small RNAs

All reads aligned by the two-tier process outlined above are evaluated for overlaps with annotation libraries for miRNAs, tRNAs, and Y-RNAs, which are derived from miRBase [33], GtRNAdb [34], and GENCODE [35], respectively. For miRNAs, each read overlapping a known miRNA locus is compared to the start position of the reference miRNA listed in miRBase. If the start position of the read is identical to that of the reference miRNA, the read contributes one count toward the reference miRNA. If the start position of the read differs from the reference start location, the read contributes one count to an isomiR that is named to represent that offset (e.g., if the read start position is 2 nucleotides upstream of the reported start position for miR-25-3p, the read would be called miR-25-3p_+_2).

For tRNAs, each read overlapping a known tRNA locus in GtRNAdb is named to include the chromosome of origin, tRNA number, amino acid with which it associates, location within the tRNA to which the read maps, total length of the tRNA, and strand (chr#:tRNA#:amino acid and codon:start position in tRNA:end position in tRNA:tRNA length:strand). More detailed annotation and quantification of tDRs is available through other programs, such as tDRmapper [29] or MINTbase [36]. The same naming scheme is used for yDRs (Y-RNA loci are defined according to GENCODE [35] annotations). Reads for which no overlap occurs with miRNA, tRNA, or Y-RNA loci are named according to the genomic location of the start position and the strand (chr#:location#:strand).

3.5 miRquant 2.0 output

A hallmark feature of miRquant 2.0 is the extensive information provided on miRNAs, isomiRs, and other small RNAs detected in the smRNA-seq data (Figure 1, 2). This information is reported in separate files and these are summarized below.

3.5.1 Mapping statistics

For each sample, miRquant 2.0 reports tabular data that summarizes the quality of the sequencing data and the overall mapping statistics (Figure 2A). From these tables, users can determine the number of total reads generated, as well as the number and percentage of total reads that were successfully trimmed, trimmed reads that were mapped by Bowtie and SHRiMP, mapped reads that pertain to miRNAs, tRNAs, or Y-RNAs. These data inform the user of potential quality issues with the sample (RNA degradation, adapter dimerization during library prep, poor sequencing) as well as the relative abundance of different types of small RNAs represented in the data.

3.5.2 Cross-sample comparisons

miRquant 2.0 computes pair-wise Pearson correlations of miRNA expression profiles across all samples and carries out hierarchical clustering to plot the expression correlation heat map. Only miRNAs above a specified expression threshold are included in the correlation analysis. An example is provided in Figure 2B in which miRNA profiles are being compared across two different conditions with three replicates each. The color in each of the squares in the heat map corresponds to the extent of miRNA expression correlation between two particular samples. In this example, the heat map and dendrogram show that the replicates are grouped appropriately by condition, and that the two conditions cluster separately.
Figure 2: Summary of miRquant 2.0 output. Example output from running miRquant 2.0 on simulated data is shown. (A) Mapping statistics for two samples, A and B, are shown. The specific step in the miRquant 2.0 workflow at which a particular statistic is generated is indicated on the left. Sample A is an example of high-quality data, whereas B is an example of lower-quality data (e.g., lower rate of trimming). (B) miRNA expression correlation heat map for two different conditions with three replicates each is shown. The color of each square represents pair-wise Pearson correlation values. The dendrogram (relationship across samples) is determined by hierarchical clustering. (C) The length distribution of trimmed reads for a single sample is shown. The x-axis represents read length and the y-axis represents the percent of the total reads remaining following Cutadapt trimming. The x-axis represents read length and the y-axis represents the percent of the total reads remaining following Cutadapt trimming. tDRs, tRNA-derived RNAs; yDRs, Y-RNA derived RNAs. (D) Reads per million mapped (RPMM) data for two different samples. Examples of possible annotations (miRNAs, isomiRs, tDRs, yDRs, and unannotated) are provided in column 1. The RPMM values are provided in columns 2 (Sample A) and 3 (Sample B). Additional data provided by miRquant, such as qualitative and quantitative information on A-to-I editing and 3’ non-templated nucleotide additions, are not shown here.

3.5.3 Read length distributions

For each sample, miRquant 2.0 reports the length distributions of all trimmed reads (Figure 2C). A peak at ~18-24 nucleotides likely represents mature miRNAs, a subclass of tDRs known as tRNA fragments (tRFs), and possibly yDRs. A peak at ~30-33 nucleotides likely represents a subclass of tDRs referred to as tRNA halves (tRHs), and possibly yDRs. An elevated signal at other read sizes could indicate other known or novel small RNAs, but could also be suggestive of degradation.
3.5.4 Normalized expression

Read count normalization is performed in two ways. For a given miRNA/isomiR/tDR/yDR in a sample, the number of corresponding reads is divided by the total number of mapped reads and multiplied by 1 million. This is referred to as the reads per million mapped (RPMM). For miRNAs/isomiRs, an alternative method of normalization is employed. Specifically, the number of corresponding reads is divided by the total number of reads mapped to all miRNA/isomiR loci and multiplied by 1 million (Equation 1). This is referred to as reads per million mapped to miRNAs (RPMMM). The latter approach is generally favored for miRNAs. Notably, the types, locations, and frequency of potential post-transcriptional editing events are reported for every annotated miRNA/isomiR/tDR/yDR in the sample. Edits are separated into two groups: internal edits and 3’ non-templated additions.

\[ \text{RPMMM} = \frac{c_i}{c_{\text{tot}}} \cdot 1,000,000; \quad \text{Eq. 1} \]

Reads per million mapped to miRNAs, where \( c_i \) is the number of reads mapped to a specific miRNA or isomiR (precise 5’-start position) and \( c_{\text{tot}} \) is the total number of reads mapped to all miRNAs and isomiRs.

3.5.5 Programs required to run miRquant 2.0

In order to run miRquant 2.0 successfully on smRNA-seq data, users must have installed the following programs:

- python (https://www.python.org/)
- bedtools (http://bedtools.readthedocs.io/en/latest/)
- Bowtie (http://bowtie-bio.sourceforge.net/index.shtml)
- SHRiMP (https://github.com/compbio-UofT/shrimp)
- R (https://www.r-project.org/)

Further information on versions and installation can be found at:
https://github.com/Sethupathy-Lab/miRquant

4 Application

miRquant 2.0 is an analysis tool for small RNA sequencing datasets, allowing for the accurate annotation and quantification of miRNAs and their isoforms. miRquant 2.0 can be run in a linear fashion or can be split into multiple jobs that run in parallel. miRquant 2.0 can complete the analysis of an example mouse sample containing 20 million reads in 404.8 minutes (139.5, 264.7, and <1 minutes for 3’ adapter trimming/alignment, annotation, and final processing, respectfully). The run time is substantially decreased to 134.7 minutes when portions of miRquant 2.0 are run in parallel (91.6, 42.4, and <1 minutes for 3’ adapter trimming/alignment, annotation, and final processing, respectfully). In theory, when running miRquant 2.0 in parallel, overall run time for multiple samples will be the same as for a single sample. miRquant 2.0 can be downloaded at: https://github.com/Sethupathy-Lab/miRquant. Instructions for running the code and interpreting the output are provided at the same location.
5 Discussion

We have provided here a description of miRquant 2.0 – a tool for comprehensive annotation and quantification of miRNAs, isomiRs, and other small RNAs from smRNA-seq data. miRquant 2.0 also provides detailed information on the sequencing data quality, genome mapping statistics, relative abundance of different classes of small RNAs, prevalence of post-transcriptional modifications, and similarity of miRNA profiles across samples. We have successfully used a previous version of miRquant in several published studies involving both mouse and human smRNA-seq datasets [37–40]. We anticipate that version 2.0 will be even more useful for researchers interested not only in quantifying miRNAs, but also digging deeper into the rich well of additional information embedded in smRNA-seq data.

Acknowledgements

We would like to thank all of the members of the Sethupathy laboratory for helpful discussions and feedback on miRquant 2.0, as well as Karl Eckland and Jeff Roach for valuable input on the structure and implementation of miRquant 2.0. Funding for the development of miRquant 2.0 was provided in part by a grant from NIH/NIDDK (R01DK105965 awarded to P.S.), from the American Heart Association (Collaborative Science Award awarded to P.S.), and from the American Diabetes Association Pathway Program (1-16-ACE-47 awarded to P.S.).

References

[1] T. Caspersson and J. Schultz. Pentose nucleotides in the cytoplasm of growing tissues. Nature, 143(3623):602–3, 1939.
[2] F. H. Crick. On protein synthesis. presented at the Symp Soc Exp Biol, 1958, 12:8.
[3] R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, and A. Zamir. Structure of a Ribonucleic Acid. Science, 147(3664):1462–5, Mar. 1965.
[4] S. R. Eddy. Non–coding RNA genes and the modern RNA world. Nature Reviews Genetics, 2(12):919–929, 2001.
[5] M. Esteller. Non-coding RNAs in human disease. Nat Rev Genet, 12(12):861–74, Nov. 2011.
[6] M. B. Clark, A. Choudhary, M. A. Smith, R. J. Taft, and J. S. Mattick. The dark matter rises: the expanding world of regulatory RNAs. Essays In Biochemistry, 54:1–16, 2013.
[7] K. C. Vickers, L. A. Roteta, H. Hucheson-Dilks, L. Han, and Y. Guo. Mining diverse small RNA species in the deep transcriptome. Trends Biochem Sci, 40(1):4–7, Jan. 2015.
[8] E. M. Weick and E. A. Miska. piRNAs: from biogenesis to function. Development, 141(18):3458–71, Sep. 2014.
[9] S. K. Chakrabortty, A. Prakash, G. Nechooshtan, S. Hearn, and T. R. Gingeras. Extracellular vesicle-mediated transfer of processed and functional RNY5 RNA. RNA, 21(11):1966–79, Nov. 2015.
[10] D. P. Bartel. MicroRNAs: target recognition and regulatory functions. Cell, 136(2):215–33, Jan. 2009.
[11] S. Mukherji, M. S. Ebert, G. X. Zheng, J. S. Tsang, P. A. Sharp, and A. van Oudenaarden. MicroRNAs can generate thresholds in target gene expression. Nat Genet, 43(9):854–9, Aug. 2011.
[12] M. S. Ebert and P. A. Sharp. Roles for microRNAs in conferring robustness to biological processes. Cell, 149(3):515–24, Apr. 2012.

[13] J. M. Schniedel, S. L. Klemm, Y. Zheng, A. Sahay, N. Blüthgen, D. S. Marks, and A. van Oudenaarden. MicroRNA control of protein expression noise. Science, 348(6230):128–32, 2015.

[14] A. Mencia, S. Modamio-Hoybjor, N. Redshaw, M. Morin, F. Mayo-Merino, L. Olavarrrieta, L. A. Aguirre, I. del Castillo, K. P. Steel, T. Dalmay, F. Moreno, and M. A. Moreno-Pelayo. Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet, 41(5):609–13, May 2009.

[15] J. Couzin. MicroRNAs make big impression in disease after disease. Science, 319(5871):1782–4, Mar. 2008.

[16] A. Lujambio and S. W. Lowe. The microcosmos of cancer. Nature, 482(7385):347–55, Feb. 2012.

[17] M. Yousef, H. V. Trinh, and J. Allmer. Intersection of MicroRNA and gene regulatory networks and their implication in cancer. Curr Pharm Biotechnol, 15(5):445–54, 2014.

[18] R. L. Skalsky and B. R. Cullen. Viruses, microRNAs, and host interactions. Annu Rev Microbiol, 64:123–41, 2010.

[19] C. Maudet, M. Mano, and A. Eulalio. MicroRNAs in the interaction between host and bacterial pathogens. FEBS Lett, 588(22):4140–7, Nov. 2014.

[20] M. D. Sacar, C. Bagci, and J. Allmer. Computational prediction of microRNAs from Toxoplasma gondii potentially regulating the hosts’ gene expression. Genomics Proteomics Bioinformatics, 12(5):228–38, Oct. 2014.

[21] P. S. Mitchell, R. K. Parkin, E. M. Kroh, B. R. Fritz, S. K. Wyman, E. L. Pogosova-Agadjanyan, A. Peterson, J. Noteboom, K. C. O'Briant, A. Allen, D. W. Lin, N. Urban, C. W. Drescher, B. S. Knudsen, D. L. Stirewalt, R. Gentleman, R. L. Vessella, P. S. Nelson, D. B. Martin, and M. Tewari. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci U S A, 105(30):10513–8, Jul. 2008.

[22] Z. Li and T. M. Rana. Therapeutic targeting of microRNAs: current status and future challenges. Nature reviews Drug discovery, 13(8):622–638, 2014.

[23] P. Sethupathy. The Promise and Challenge of Therapeutic MicroRNA Silencing in Diabetes and Metabolic Diseases. Current diabetes reports, 16(1):47–53, Oct. 2004.

[24] J. Baran-Gale, C. L. Kurtz, M. R. Erdos, C. Sison, A. Young, E. E. Fannin, P. S. Chines, and P. Sethupathy. Addressing Bias in Small RNA Library Preparation for Sequencing: A New Protocol Recovers MicroRNAs that Evade Capture by Current Methods. Front Genet, 6:352, 2015.

[25] C. Bagci and J. Allmer. One Step Forward, Two Steps Back; Xeno-MicroRNAs Reported in Breast Milk Are Artifacts. PLoS One, 11(1):e0145065, 2016.

[26] J. Baran-Gale, E. E. Fannin, C. L. Kurtz, and P. Sethupathy. Beta cell 5’-shifted isomiRs are candidate regulatory hubs in type 2 diabetes. PLoS One, 8(9):e73240, 2013.

[27] S. R. Selitsky and P. Sethupathy. tDRmapper: challenges and solutions to mapping, naming, and quantifying tRNA-derived RNAs from human small RNA-sequencing data. BMC Bioinformatics, 16:354, Nov. 2015.

[28] M. Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, 17:10–12, May 2011.
[31] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 10(3):R25, 2009.

[32] M. David, M. Dzamba, D. Lister, L. Ilie, and M. Brudno. SHRiMP2: sensitive yet practical SHort Read Mapping. Bioinformatics, 27(7):1011–2, Apr. 2011.

[33] A. Kozomara and S. Griffiths-Jones. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res, 42(Database issue):D68-73, Jan. 2014.

[34] P. P. Chan and T. M. Lowe. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic acids research, 37(suppl 1):D93–D97, 2009.

[35] J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J. M. Rodriguez, I. Ezkurdia, J. van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, R. Guigo, and T. J. Hubbard. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res, 22(9):1760–74, Sep. 2012.

[36] V. Pliatsika, P. Loher, A. G. Telonis, and I. Rigoutsos. MINTbase: a framework for the interactive exploration of mitochondrial and nuclear tRNA fragments. Bioinformatics, 32(16):2481–9, Aug. 2016.

[37] C. L. Kurtz, B. C. Peck, E. E. Fannin, C. Beysen, J. Miao, S. R. Landstreet, S. Ding, V. Turaga, P. K. Lund, S. Turner, S. B. Biddinger, K. C. Vickers, and P. Sethupathy. MicroRNA-29 fine-tunes the expression of key FOXA2-ligated lipid metabolism genes and is dysregulated in animal models of insulin resistance and diabetes. Diabetes, 63(9):3141–8, Sep. 2014.

[38] B. C. Peck, M. Weiser, S. E. Lee, G. R. Gipson, V. B. Iyer, R. B. Sartor, H. H. Herfarth, M. D. Long, J. J. Hansen, K. L. Isacris, D. G. Trembath, R. Rahbar, T. S. Sadiq, T. S. Furey, P. Sethupathy, and S. Z. Sheikh. MicroRNAs Classify Different Disease Behavior Phenotypes of Crohn’s Disease and May Have Prognostic Utility. Inflamm Bowel Dis, 21(9):2178–87, Sep. 2015.

[39] S. R. Selitsky, T. A. Dinh, C. L. Toth, C. L. Kurtz, M. Honda, B. R. Struck, S. Kaneko, K. C. Vickers, S. M. Lemon, and P. Sethupathy. Transcriptomic Analysis of Chronic Hepatitis B and C and Liver Cancer Reveals MicroRNA-Mediated Control of Cholesterol Synthesis Programs. MBio, 6(6):e01500-15, Dec. 2015.

[40] J. Baran-Gale, J. E. Purvis, and P. Sethupathy. An integrative transcriptomics approach identifies miR-503 as a candidate master regulator of the estrogen response in MCF-7 breast cancer cells. RNA, 22(10):1592–603, Oct. 2016.