**Additional File 2**

**Generation of custom databases for precursor and mature miRNA**

The goal of creating a custom database of precursor and mature miRNA sequences for the PhenoGen website was to identify novel miRNA in heart, brain, and liver and to include strain-specific versions of both annotated and novel miRNA to improve mapping accuracy for these small RNA species.

Data Sets

The datasets for whole brain, liver, and heart left ventricle tissues were generated using 3-4 samples of each BN-Lx and SHR strains of rats. Each tissue was processed for both total RNA and small RNA. The small RNA (<200bp) fraction was used for sequencing on an Illumina HiSeq instrument with 1x50bp reads collected with 15-100 million reads per sample.

*Read processing*

Initially, reads from sequencing of small RNAs were trimmed for adapter sequence and read quality with cutadapt (v1.9.1; Martin 2011) removing reads with any lengths shorter than 18 nucleotides (nt). Reads that align to rRNA using Bowtie2 (v2.3.0; Langmead and Salzberg 2012) were removed.

*Identification of precursor miRNA sequences*

The remaining reads were aligned to the appropriate strain-specific genome (http://phenogen.org) using Tophat2 (v2.1.0; Kim et al. 2013). MirDeep (v2.0.0.7; Friedländer et al. 2012) was used with genome-aligned reads to predict precursor and mature miRNAs based on genomic sequence and read coverage. MirDeep was applied to each strain within each tissue separately, i.e., 6 MirDeep data sets were generated. These six data sets were merged along with precursor miRNA annotated in Ensembl and/or RefSeq based on physical location of the genome that encodes the precursor miRNA.

Because SNPs within a miRNA (precursor or mature) can dramatically influence alignment rates when read lengths are approximately 20 nt long, we generated a database of precursor miRNA sequences from the locations mentioned previously that incorporate SNPs within the sequence. For example, if a SNP existed within a precursor miRNA for the SHR strain but not the BN-Lx strain, two versions of that precursor miRNA would be included in the database, one with the SNP and one with the reference sequence.

*Identification of mature miRNA sequences*

To identify mature miRNA sequences, we identified the 18-27 nt region within the precursor miRNA with read coverage greater than the median base coverage across the entire precursor miRNA. Read coverage for each base was calculated by aligning reads directly to the precursor miRNA sequences described above using STAR (v2.7.3a; Dobin et al. 2013) and bedtools (b2.28.0; Quinlan and Hall 2010). Mature miRNAs were identified in this manner for each strain within each tissue separately. Like the precursor miRNA, the six data sets were merged along with mature miRNA annotated in Ensembl and/or RefSeq based on the physical location of the genome that encodes the mature miRNA.

To generate mature miRNA sequences suitable for mapping reads to, we retrieved the sequence of each mature miRNA based on its strain-specific genome. Because the same mature miRNA can be derived from multiple different precursor miRNAs transcribed from different regions of the genome, it was also important, at this step, to eliminate duplicated mature miRNA sequences. Keeping duplicated mature miRNA sequences in our database would have increased ambiguity of read alignment. This final set of unique mature miRNA sequences was used for read alignment and miRNA quantitation.

*References*

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. “STAR: Ultrafast Universal RNA-Seq Aligner.” *Bioinformatics (Oxford, England)* 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Friedländer, Marc R., Sebastian D. Mackowiak, Na Li, Wei Chen, and Nikolaus Rajewsky. 2012. “MiRDeep2 Accurately Identifies Known and Hundreds of Novel MicroRNA Genes in Seven Animal Clades.” *Nucleic Acids Research* 40 (1): 37–52. https://doi.org/10.1093/nar/gkr688.

Kim, Daehwan, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L. Salzberg. 2013. “TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions.” *Genome Biology* 14 (4): R36. https://doi.org/10.1186/gb-2013-14-4-r36.

Langmead, Ben, and Steven L. Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Martin, Marcel. 2011. “Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads.” *EMBnet.Journal* 17 (1): 10–12. https://doi.org/10.14806/ej.17.1.200.

Quinlan, Aaron R., and Ira M. Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics (Oxford, England)* 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.