**Experimental Plan: EBV project**

**I. Total RNA Libraries** *(Used ERCC mix 1 & 2 according to ERCC guidelines—can’t remember which individual libraries got mix 1 or mix 2, but info is in my lab notebook #3 if you need it)*

All of these libraries are made from the same base cell line: RPE = human retinal pigmented epithelial cells immortalized with hTERT. The only difference among these four sets of lines is the vector added to them, as follows:

|  |  |
| --- | --- |
| **Cell Line Descriptor** | **Name of sample** |
| **RPE w/nothing added at all** (negative control for basal cell line expression profiles) | RPE-1 |
| RPE-2 |
| RPE-3 |
| **RPE w/ vector- Empty Construct only** (negative control for background effects of the transfection or vector only) | miA-1 |
| miC-1 |
| miD-1 |
| **RPE w/ vector+ EBER2-smRNA-1** (this is the small RNA we found within EBER2) | shA-5 |
| shD-2 |
| shE-3 |
| **RPE w/ vector+ EBER2 full-length** (this is the full length EBER2) | RE-8 |
| RE-10 |
| RE-12 |

The comparison here to do as follows:

FIRST – process the libraries – trim and QC clean, get rid of adaptors, etc

Map to hg38 genome (AND map to CHM13, the new T2T genome) – use Hisat2

Also need to map to EBV.

We need to know what genes are affected AND what repeats and noncoding RNAs may be affected. So, map two ways to see if there is a difference - calling a single map location or calling multiple mappers. We do expect some genes to go up and or down, but also many to turn on or turn off.

Then perform DESEQ2 analyses

**1. RPE w/nothing** vs **RPE w/ vector- Empty =>>** tells us how vector and transfection method impacts transcription

**2.** **RPE w/ vector+ EBER2 full-length** vs **RPE w/nothing; RPE w/ vector+ EBER2 full-length** vs **RPE w/vector-Empty** (or a single comparison of **RPE w/ vector+ EBER2 full-length** vs **RPE w/nothing+RPE w/vector-Empty** (merged after comparison?)**=>>** tell us how the EBER2 noncoding RNA from EBV affects transcription in the *absence* of the full virus.

**3. RPE w/ vector+ EBER2-smRNA-1** vs **RPE w/nothing; RPE w/ vector+ EBER2-smRNA-1** vs **RPE w/vector-Empty** (or a single comparison of **RPE w/ vector+ EBER2-smRNA-1** vs **RPE w/nothing+RPE w/vector-Empty** (merged after comparison?)**=>>** tell us how the EBER2 smRNA (ONLY) from EBV affects transcription in the *absence* of the full virus.

**4. RPE w/ vector+ EBER2-smRNA-1** vs **RPE w/ vector+ EBER2 full-length =>>** tell us how the EBER2 smRNA (ONLY) *differentially* affects transcription in the *absence* of the full virus. This should help us figure out if the effects of EBER2 are really the small RNA only or if the small RNA affects cells in a specific (likely subset of EBER2 full length) way.

**II. Small RNA Libraries (all size selected, so use Bowtie)** *(Used ExiSEQ smRNA spike-ins—only one mix used for all libraries—we need to find the .fasta file for spike-in sequences and pdf of protocol which i followed when adding to samples)*

This is a very different analysis. There are FOUR cell lines here, all from different tissues and individuals (all human). **Daudi** = Burkett's Lymphoma (an EBV+ cancer), **Hsa** = lymphoblast cells (immortalized by the addition of EBV in the lab), **SNU** = gastric carcinoma (an EBV+ cancer), **RE-8** = the RPE cell line from above with the vector containing the full length EBER2 only (no EBV infection). In EACH of these four lines, a single protein from the list of proteins involved in small RNA biogenesis were knocked down to see what effect, if any, the loss of the protein had on the processing of EBER2 into a small RNA, along with other small RNAs in the virus and/or genome. The lines are as follows:

|  |  |
| --- | --- |
| **Cell Line Descriptor** | **Protein KD (knockdown)** |
| **Daudi** = Burkett's Lymphoma (an EBV+ cancer) | no treatment (neg control) |
| Dicer |
| Drosha |
| Ago1 |
| Ago2 |
| Ago3 |
| Ago4 |
| La |
| **Hsa** = lymphoblast cells EBV+ (immortalized by the addition of EBV in the lab) | no treatment (neg control) |
| Dicer |
| Drosha |
| Ago1 |
| Ago2 |
| Ago3 |
| Ago4 |
| La |
| **SNU** = gastric carcinoma (an EBV+ cancer) | no treatment (neg control) |
| Dicer |
| Drosha |
| Ago1 |
| Ago2 |
| Ago3 |
| Ago4 |
| La |
| **RE-8** = the RPE cell line from above with the vector containing the full length EBER2 only (no EBV infection) | no treatment (neg control) |
| Dicer |
| Drosha |
| Ago1 |
| Ago2 |
| Ago3 |
| Ago4 |
| La |

For these experiments, we need to know the small RNAs in the cell affected by each KD (ie. map to hg38) AND the small RNAs derived from EBV that are affected be each KD (i.e. map to EBV genome). For analyses: within each of the four lines, compare each KD to the no treatment and to each other KD (what is the same, what is different). We need to intersect these mapped reads with known miRNAs (I think you have to use hg38 for this) and other smRNAs and focus on these first (it will be a smaller dataset). For EBV you will have to just map to the whole thing I think - many of the small RNAs are not annotated as such. THEN we should compare across the lines - how does KD of La in SNU affect small RNA processing vs in RE vs Daudi vs Hsa? (for example). The best place to start in this comparison is to look at the EBV genome mapping first; which small RNAs from the EBV genome change across these datasets? And specifically, does ours small RNA from EBER 2 change in any of the KDs in any of the cell lines (or is there a differential effect? some lines may lose more of the EBER2 small RNA than others with the same KD)? The spike ins will be important in this analysis as we want to quantify the changes in expression (a proxy for measuring small RNA processing) of small RNA in each sample.