**Bintu Lab EM-Seq Pipeline**

**Original code by Ben Doughty, adapted by Abby Thurm (w/ code from Connor Ludwig)**

**Pipeline and notes by Abby Thurm**

Part I – Alignment

For methylation-specific sequencing, we’ll use bwa-meth to align. Create a fasta file of the amplicon or genome you wish to align to (if genome, use primary\_assembly.fa from whatever genome you’ve built and gotten from UCSC; if amplicon, copy from Benchling OR create fasta in following format in text editor saved with extension.fasta):  
>seq\_name  
{sequence\_here}

Clone bwa-meth from Github – git clone bwa-meth – in this example, we have cloned it into a folder called bwa-meth-master.

Create a bwa-meth index for your amplicon of choice using the following code:

python ~/win\_f/Abby/EM-Seq/bwa-meth-master/bwameth.py index {your\_fasta}

Bwameth will output a bunch of .c2t files, etc that it will use to align your converted reads. Leave these in the same directory where you’ve made and stored your fasta.

Download raw data from your sequencing run and run bcl2fastq (remember to flip your i5 if on the Miseq!)/get folders from your sequencing company. Create a directory for each sample with the same root as the files; move the paired fastqs into each directory.

Align your reads using bwa-meth – the script is taken from C&R/RNA-Seq processing pipelines and batches all files together.

python batch\_bwameth\_alignment.py {path to fastq directory} {path to amplicon}

You should now have indexed .bam files in the bam/ folder that can be used for downstream processing.

Part II – Single-molecule methylation analysis

We’ll now use code from Ben in Greenleaf lab (from Georgi in Greenleaf lab, etc…) to process the methylation data into matrices that will be used to 1. Produce single-molecule matrices that can be used for bulk methylation analysis later and 2. Produce plots of single molecules for visual analysis.

NOTE: you do NOT need to ever touch the script entitled dSMF…/py or heatmap.py after you hard-code their location into the singlemol\_heatmaps.py script. I’ve hard-coded their location into this script and it will run from wherever you need your new files to be stored. Pass in your parent directory (i.e. the directory that has your folder of bams and sams, NOT your fastq directory). Pass in the LITERAL string (starting with ~/win\_x/xxxx) of your file name to get the path.

If you have NOT yet made a .bed file of your amplicon (amplicon\_name\t0\tamplicon\_len), the script should do that for you if you add a third arg ‘bed’ when you call the script below.

python singlemol\_heatmaps.py {path to parent directory} {path to amplicon}

You should now have a folder called heatmaps/ that contains MATRIX files (text files you can process later for bulk analysis) and single-molecule heatmap pictures.