

Running the DGE pipeline on LSF:

Prerequisites:

- 1) Download the DGE package and unpack it with tar
- 2) You'll need both python (I usually run it in 2.7) and bwa installed. bwa should be runnable by simply typing the command "bwa".

Usage is as follows:

```
usage: python /path/to/dge-prod/Scripts/run_DGE_analysis.py [-h]
                  [--short_lsf_queue SHORT_LSF_QUEUE]
                  [--long_lsf_queue LONG_LSF_QUEUE]
                  [--loose_barcodes] [--cleanup]
                  sample_map reference barcodes alignment_dir
                  analysis_dir
```

positional arguments:

|               |  |
|---------------|--|
| sample_map    | location of sample map file  |
| reference     | reference genome: Human Mouse Rat Chicken  |
| barcodes      | barcode plate: P1 P2 P3 P1P2 Trugrade_384_set1 <br>Trugrade_96_set1 Trugrade_96_set2 Trugrade_96_set3 <br>Trugrade_96_set4 |
| alignment_dir | directory to process alignments  |
| analysis_dir  | directory to calculate gene expression   |

optional arguments:

|                   |   |
|-------------------|---|
| -h, --help        | show this help message and exit                               |
| --short_lsf_queue | lsf_queue to run short jobs. default=hour                     |
| --long_lsf_queue  | lsf_queue to run long jobs. default=forest                    |
| --loose_barcodes  | allows well barcodes to have 1 mismatch                       |
| --cleanup         | removes resulting fq and sam files upon successful completion |

Most arguments should be self explanatory. The sample\_map should be a tab-delimited file where there's a line for each pair of fastqs like:

```
sample_id subsample_id /path/to/read1.fq /path/to/read2.fq
```

\*Please note: the pipeline expects read1 to have at least 16 called bases, where bases 1-6 represent the well barcode and bases 7-16 represent the UMI.

An example usage of the pipeline is:

```
python /path/to/dge-prod/Scripts/run_DGE_analysis.py --short_lsf_queue queue --long_lsf_queue
queue SampleMap Human Trugrade_384_set1 Alignment DGE
```

The word "queue" should be replaced with the name of your queue in LSF. At the Broad we have different queues for short and long-running jobs. If you don't, you can use the same queue for both settings.