Building reference sequence:

|  |  |
| --- | --- |
| **Terminal Code** | **Results/Notes** |
| Installed bowtie2-2.2.6 | Installed in Documents |
| Installed RSEM with ebseq | Installed in Documents in bowtie2-2.2.6 folder |
| ./RSEM-1.3.3/rsem-prepare-reference --gff3 ref/17978mff.gff3 -gff3-genes-as-transcripts --hisat2-hca ref/17978mff.fasta ref/17978mff\_ref  ~/Documents/RNASeq/RSEM-1.3.3/rsem-prepare-reference --gff3 Genomes/plchrmff.gff3 -gff3-genes-as-transcripts –bowtie2 –-bowtie2-path ~/Documents/RNASeq/bowtie2-2.2.6 Genomes/genmff.fasta,Genomes/pAB3.fasta ref2/fullmff\_ref  ~/Documents/RNASeq/RSEM-1.3.3/rsem-prepare-reference --gff3 Genomes/plchrmff.gff3 -gff3-genes-as-transcripts --bowtie2 –-bowtie2-path ~/Documents/RNASeq/bowtie2-2.2.6 Genomes/A1S.fasta ref3/A1S\_ref | * 3596 transcripts extracted * 3747 transcripts extracted * Without –gff3-genes-as-transcripts, only 90 * Original genome indicates 3,798 genes & 3,663 CDS * Default is only mRNA * 3871 traanscripts extracted in A1S genome * Used bowtie2 2.4.5 (conda) |

Trimming data

|  |  |
| --- | --- |
| **Terminal Code or Long results** | **Quick Results/Notes** |
| RNAseqPhrB % pip3 install cutadapt |  |
| cutadapt --version | To check successful installation: 3.5 |
| RNAseqPhrB % ~/opt/anaconda3/bin/conda install -c bioconda fastqc | Pip3 did not work |
| fastqc --version | FastQC v0.11.5 |
| wget <https://github.com/FelixKrueger/TrimGalore/archive/0.6.6.tar.gz>  tar xvzf 0.6.6.tar.gz | Install TrimGalore & ensure accessible |
| mkdir Clean | Create directory for trimmed files |
| Gunzip \*fastq.gz | Unzip all files |
| Wc -lh  -rw-r--r--@ 1 hopekirby staff 1683581365 Jan 28 11:13 17978L1\_S1\_L001\_R1\_001.fastq  -rw-r--r--@ 1 hopekirby staff 1683771123 Jan 28 11:15 17978L1\_S1\_L001\_R2\_001.fastq  -rw-r--r--@ 1 hopekirby staff 1321278933 Jan 28 11:14 17978L2\_S2\_L001\_R1\_001.fastq  -rw-r--r--@ 1 hopekirby staff 1321488789 Jan 28 11:15 17978L2\_S2\_L001\_R2\_001.fastq  -rw-r--r-- 1 hopekirby staff 1509142903 Feb 3 15:47 17978WT\_S1\_L001\_R1\_001.fastq  -rw-r--r--@ 1 hopekirby staff 1509173521 Feb 3 15:41 17978WT\_S1\_L001\_R2\_001.fastq  -rw-r--r--@ 1 hopekirby staff 747340807 Jan 30 14:29 17978mutL1\_S3\_L001\_R1\_001.fastq  -rw-r--r--@ 1 hopekirby staff 747374519 Jan 30 14:30 17978mutL1\_S3\_L001\_R2\_001.fastq  -rw-r--r--@ 1 hopekirby staff 1355570440 Jan 30 14:30 17978mutL2\_S4\_L001\_R1\_001.fastq  -rw-r--r--@ 1 hopekirby staff 1355806768 Jan 30 14:32 17978mutL2\_S4\_L001\_R2\_001.fastq  **-rw-r--r--@ 1 hopekirby staff** 1810065559 **Feb 3 15:18 17978mut\_S2\_L001\_R1\_001.fastq**  -rw-r--r--@ 1 hopekirby staff 1810162421 Jan 29 22:48 17978mut\_S2\_L001\_R2\_001.fastq  -rw-r--r--@ 1 hopekirby staff 913424457 Jan 28 12:36 Abaumannii17978D1\_S1\_L001\_R1\_001.fastq  -rw-r--r--@ 1 hopekirby staff 913442701 Jan 28 12:36 Abaumannii17978D1\_S1\_L001\_R2\_001.fastq  -rw-r--r--@ 1 hopekirby staff 954165457 Jan 28 12:47 Abaumannii17978D2\_S2\_L001\_R1\_001.fastq  -rw-r--r--@ 1 hopekirby staff 954202389 Jan 28 12:48 Abaumannii17978D2\_S2\_L001\_R2\_001.fastq  -rw-r--r--@ 1 hopekirby staff 970662232 Jan 28 12:48 Abaumannii17978D3\_S3\_L001\_R1\_001.fastq  -rw-r--r--@ 1 hopekirby staff 970758052 Jan 28 12:48 Abaumannii17978D3\_S3\_L001\_R2\_001.fastq  -rw-r--r--@ 1 hopekirby staff 995466531 Jan 28 12:48 AbaumanniiBlsAD1\_S4\_L001\_R1\_001.fastq  -rw-r--r--@ 1 hopekirby staff 995535441 Jan 28 12:48 AbaumanniiBlsAD1\_S4\_L001\_R2\_001.fastq  -rw-r--r--@ 1 hopekirby staff 1877559369 Jan 28 11:17 BlsAD2\_S1\_L001\_R1\_001.fastq  -rw-r--r--@ 1 hopekirby staff 1877619547 Jan 28 11:15 BlsAD2\_S1\_L001\_R2\_001.fastq  -rw-r--r--@ 1 hopekirby staff 2222995453 Jan 28 11:17 BlsAD3\_S2\_L001\_R1\_001.fastq  -rw-r--r--@ 1 hopekirby staff 2222724175 Jan 28 11:13 BlsAD3\_S2\_L001\_R2\_001.fastq | To determine size of all files |
| ./TrimGalore-0.6.6/trim\_galore -q 20 --phred33 --fastqc --paired ./RawData/17978L1\_S1\_L001\_R1\_001.fastq ./RawData/17978L1\_S1\_L001\_R2\_001.fastq -o ./Clean/CWTL1\_24  ./TrimGalore-0.6.6/trim\_galore -q 20 --phred33 --fastqc --paired ./RawData/BlsAD3\_S2\_L001\_R1\_001.fastq ./RawData/BlsAD3\_S2\_L001\_R2\_001.fastq -o ./Clean/CBlsAD3\_24  Repeated for rest of files | Actually trimmed files  Run Parameters: Number of cores used for trimming: 1  Quality Phred score cutoff: 20  Quality encoding type selected: ASCII+33  Adapter sequence: 'AGATCGGAAGAGC' (Illumina TruSeq, Sanger iPCR; auto-detected)  Maximum trimming error rate: 0.1 (default)  Minimum required adapter overlap (stringency): 1 bp  Minimum required sequence length for both reads before a sequence pair gets removed: 20 bp |
| For some reason the fastqc was not activated so had to go to each folder and:  fastqc -f fastq \*.fq |  |

Run RSEM to calculate expression

|  |  |
| --- | --- |
| **Terminal Code or Long results** | **Quick Results/Notes** |
| mkdir exp | **Create a folder for the results** |
| conda install -c bioconda hisat2 (2-2.2.1)  conda install -c bioconda rsem  conda install -c bioconda samtools=1.3  rsem-calculate-expression -p 8 --paired-end --hisat2-hca --estimate-rspd --append-names --output-genome-bam ./Clean/CWTL1\_24/17978L1\_S1\_L001\_R1\_001\_val\_1.fq ./Clean/CWTL1\_24/17978L1\_S1\_L001\_R2\_001\_val\_2.fq ./ref/17978mff\_ref ./exp/WTL/WTL1  conda install -c bioconda ebseq  ~/Documents/RNASeq/RSEM-1.3.3/rsem-calculate-expression -p 8 –-paired-end –-bowtie2 –-estimate-rspd –-append-names –-output-genome-bam ./Clean/BlsA/BlsAD1.fq ./Clean/BlsA/BlsA1b.fq ./ref3/NZmff\_ref ./exp3/BlsAD/BlsAD1 | * - p tells RSEM to use 8 threads for input that is --paired-end * Align reads using Bowtie2 * Estimate-rspd to determine if sequencing bias (RSEM learns from data how the reads are distributed across a transcript to see if biases) * Append names tells RSEM to append the gene or transcript name to the results file |

Run EBseq for differential gene expression

|  |  |
| --- | --- |
| **Terminal Code or Long results** | **Quick Results/Notes** |
| Cd exp |  |
| rsem-generate-data-matrix WTL/WTL1.genes.results WTL/WTL2.genes.results WTL/WTL3.genes.results WTD/WTD1.genes.results WTD/WTD2.genes.results WTD/WTD3.genes.results > GeneMatWT.txt  rsem-generate-data-matrix BlsAL/BlsAL1.genes.results BlsAL/BlsAL2.genes.results BlsAL/BlsAL3.genes.results BlsAD/BlsAD1.genes.results BlsAD/BlsAD2.genes.results BlsAD/BlsAD3.genes.results > GeneMatBlsA2.txt | ../RSEM-1.3.3/rsem-generate-data-matrix BR1 BR2 BR1 BR2 and place matrix into GeneMat.txt  **Checked & looks good**  **Count matrix!** |
| rsem-run-ebseq GeneMatWT.txt 3,3 GeneMatWT.results 🡪 ERRORS SO INSTEAD WENT TO R & used EBseq-bioconductor! | Perform the differential analysis using ebseq by clarifying 3 biological replicates in each condition |

rsem-generate-data-matrix BlsAL/BlsAL1.genes.results BlsAL/BlsAL2.genes.results BlsAL/BlsAL3.genes.results BlsAD/BlsAD1.genes.results BlsAD/BlsAD2.genes.results BlsAD/BlsAD3.genes.results > GeneMatBlsA.txt

ON R

|  |  |
| --- | --- |
| WT | BlsA |
| 3805 transcripts test after transcripts with 100th quantile <0 removed | 3796 |
|  |  |
|  |  |
|  |  |
|  |  |