CRISPR Hub workshop

## Friday 29th November 2019

***Session 3 – CRISPR validation/analyses (hands-on session)***

In this workshop, you will perform methylation analysis of the *EBF3* gene promoter region.

DNA samples were isolated from small numbers of flow-sorted melanoma cells targeted for demethylation using Sun-Tag\_dCas9, TET1CD and gRNA system.

- NZM40 Baseline (untransfected; Control)

- NZM40 dCas9 + TET1CD + gRNA\_472-

- NZM40 dCas9 + iTET1CD + gRNA\_472- (inactive TET1CD protein; Control)

DNA was bisulfite-treated and then PCR amplified with *EBF3-*promoter specific primers, tagged with Illumina adaptors. After clean-up, PCR products were then amplified with Illumina index primers, such that each sample had a unique combination of indexes. After mixing and another clean-up, the samples were sequenced on an Illumina MiSeq next generation sequencer (250 bp, paired-end run).

**Results and analysis:**

After MiSeq sequencing, the sequencer will de-multiplex reads based on the index sequences, i.e. separate reads into samples, and output compressed sequence reads as ‘.fastq.gz’ files. As we have done paired-end sequencing, we have 2 ‘.fastq.gz’ files per sample. i.e. 6 files in total.

1. Check quality of sequence files with FastQC, for example:

**‘fastqc NZM40\_R1.fastq.gz’**

You should view the ‘.html’ file, particularly the Per base sequence quality plot

1. Use ‘PEAR’ joiner to join R1 and R2 ‘.fastq.gz’ files.

PEAR does not need files to be de-compressed. PEAR is an intuitive joiner – meaning that it will find overlapping regions and call bases based on quality score, so that good bases will be favoured over bad and quality should be better in the overlap region. Example:

**‘pear -f NZM40\_R1.fastq.gz -r NZM40\_R2.fastq.gz -o NZM40\_R1R2’**

Pear will output reads into several files, but we are only interested in the ‘.assembled.fastq’ file

1. Confirm/check quality of merged reads with FastQC, for example:

**‘fastqc NZM40\_R1R2.assembled.fastq’**

You can see that PEAR has improved quality of sequence over the join region by viewing the ‘.html’ file.

1. Remove Illumina adaptor sequences and bad quality reads with ‘trim\_galore’:

***‘trim\_galore NZM40\_R1R2.assembled.fastq’***

Two files are output: a report file and ‘\*.trimmed.fq’ file.

1. Convert fastq files to fasta format files, for example:

**‘fastq\_to\_fasta -i NZM40\_R1R2.assembled\_trimmed.fq -o NZM40\_R1R2.fasta’**

# option -i specifies the input file and option -o names the output file

1. Align read files (bisulfite-converted) with reference sequence using BiQ\_Analyzer, for example:

**java -jar BiQ\_Analyzer.jar -nogui -rseq EBF3\_seq.txt -bseq NZM40\_R1R2.fasta -outdir ./NZM40 -sortcgmeth**

#BiQ\_Analyzer is a JAVA program, so we execute within a JAVA environment; -nogui option runs command non-graphically; -rseq specifies the file that contains the reference sequence; -bseq specifies the read file to be analysed; -outdir specifies directory to which files will be written; -sortcgmeth will sort reads from highest methylation to lowest

1. Output files:

‘pearlNecklace’ files – condensed graphical view of methylation at each CpG site

‘heatmap’ – graphical view of CpG sites (horizontal) for all reads (vertical)

‘results.tsv’ – text file of results

‘summary.dat’ – summary of number of reads analysed, etc.

‘alignment.mfa’ – shows alignments for all reads.

1. Repeat for all 3 samples and then compare heatmaps for each. For NZM40\_TET1CD sample, all reads are demethylated. SunTag CRISPR methylation editing has worked!