Help on completing this project proposal document can be found in the “working with us “

section of the BABS website <https://bioinformatics.thecrick.org/babs/working-with-us/>

|  |  |
| --- | --- |
| Budget code | 10309 |
| Time estimate | To be agreed on after project discussions |
| Analysis goals | **Please provide an outline of the analysis goals:**    ATAC-seq: differential accessibility analysis between genotypes using as input the provided set of peaks from wild-type neurons. |
| Analysis details | **Please provide details of the data analysis required.**  An initial ATAC-seq experiment (Lims ID PM19276, run 200120\_K00102\_0447\_AHFVFTBBXY) was carried out and Harshil Patel performed the peak calling and differential accessibility analysis. From this experiment, we were particularly interested in the following comparisons:   1. NP\_WT vs NRS\_WT 2. NRS\_ASCL1\_KO vs NRS\_WT   A second ATAC-seq experiment (Lims ID PM21132, run 210723\_A01366\_0037\_BH2NLTDMXY) was carried out and Miriam Iloriansopena performed the peak calling and differential accessibility analysis. From this experiment, we were particularly interested in the following comparisons:   1. 48h\_DMSO vs WT 2. 48h\_BRM014 vs 48h\_DMSO 3. 48h\_BRM014 vs WT   While some of the samples were unique to each of the two experiments, the WT samples come from the same cells (human iPSC-derived neuronal cells at day 24 post neural induction). However, we noticed approximately 55,000 accessible regions in the consensus set of peaks in WT cells from experiment 1, and only approximately 33,000 accessible regions in the consensus set of peaks in WT cells from experiment 2. Since ultimately we are interested in the regions which loose accessibility in both ASCL1\_KO and BMR014 mutants relative to the WT cells, we find this discrepancy in their relative WT controls unexpected. Having a closer look at specific genomic regions, we concluded that, most likely, this discrepancy comes from the different signal to noise ratios in the two experiments. More specifically, two of the three wild-type replicates we provided as wild-type samples in the second experiment seem to have a lower signal to noise ration when compared to the third one from the same experiment, or with the other two wild-type replicates from the first experiment.  Therefore, we would like to generate a consensus set of peaks from the two experiments and use them for the downstream differential accessibility analysis. To be able to carry on the analysis described above, I swiched to “—min\_reps\_consensus 1” for both experiments (since most likely the discrepancy between the two experiments came from the weak signal to noise ratio in the two replicates from the second experiment). I then overlapped the wild-type peaks from both experiments (1bp overlap) and generate what we will now call the consensus set of peaks for the wild-type samples (called “wt\_nrs\_consensus\_peaks.txt”).  After further discussions with Miram Llorian Sopena, I also used the same criteria (“—min\_reps\_consensus 1”) to generate new consensus sets of peaks for all the other genotypes.  Using the provided sets of consensus peaks and specifically only for the peaks identified in the wt neurons (“wt\_nrs\_consensus\_peaks.txt”), I would like to ask for differential accessibility analysis for the following comparisons:   1. – wt progenitors (“wt\_np\_consensus\_peaks.txt”) vs wt neurons (“wt\_nrs\_consensus\_peaks.txt”)   – ascl1ko neurons (“ascl1ko\_nrs\_consensus\_peaks.txt) vs wt neurons (“wt\_nrs\_consensus\_peaks.txt”)  Please find the IDs corresponding to the raw data of these samples highlighted in the attached PM19276 Excel spreadsheet.   1. – 48h dmso neurons (“48h\_dmso\_nrs\_consensus\_peaks.txt”) vs wt neurons (“wt\_nrs\_consensus\_peaks.txt”)   – 48h brm014 neurons (“48h\_brm014\_nrs\_consensus\_peaks.txt”) vs wt neurons (“wt\_nrs\_consensus\_peaks.txt”)  – 48h brm014 neurons  (“48h\_brm014\_nrs\_consensus\_peaks.txt”) vs 48h dmso  neurons (“48h\_dmso\_nrs\_consensus\_peaks.txt”)  Please find the IDs corresponding to the raw data of these samples highlighted in the attached PM21132 Excel spreadsheet. |

Please note that the number of hours quoted to carry out the analysis is an estimate and represents the maximum amount of time you are willing to commit to the work. We will contact you if this analysis time is reached and we can decide together whether to commit more time to the work. You will only be charged for the time taken to carry out the analysis.

We work in line with the [Crick's authorship policy](https://intranet.crick.ac.uk/our-crick/research-integrity/pages/publication-authorship) and our expectation is significant contributions to your research are recognised by authorship. If you have any questions or concerns regarding authorship please discuss them with us as early as possible.