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	CC2186
Time estimate	To be agreed on after project discussions
Analysis goals	**Please provide an outline of the analysis goals:** The expected outcomes will be: (a) polished figures (in PDF format) for inclusion in a manuscript, alongside (b) any data analysis tables, and (c) data deposited in accessible server system, eg GEO. Ideally, sharing of R code and intermediate data would be hugely helpful as the project progresses, so we can work on things together.

Budget code	CC2186

Analysis details	**Please provide details of the data analysis required.**
	Dataset
	Chris Cooke (PhD in Moris lab) has been working on characterising a population of primordial germ cell like cells (PGCLCs) in mouse gastruloids. He has generated some pilot 10x data and has frozen cells from a timecourse of FACS sorted cells that we plan to 10x sequence as soon as possible. I have already done some basic analysis of the pilot dataset, but am awaiting the additional populations to undergo sequencing.
	Biologically, we are interested to know what these sorted populations of cells are, whether they are equivalent to embryonic PGCs, and particularly, which timepoint they are most similar to. As such, the analysis revolves around initial basic exploration of the timecourse, comparison to similar (smaller) datasets from the same in vitro model, and then comparison to embryonic datasets from different stages of development, as described below.
	Full descriptions of the sorted populations are provided in attached spreadsheet.
	Done so far:
	1. Pilot dataset analysis
	   a. Initial QC check
	   b. Cellranger aggregation of populations together
	   c. Identification of sorted populations (by biological cell type)
	To do:
	1. Compare pilot dataset to existing in vitro datasets (see Point 4 below)
	2. Examine timecourse dataset
	   a. QC of new datasets
	   b. Confirm identity of sorted populations (across timepoints) using *a priori* marker gene lists [NM to provide]
	   c. How are sorted populations related to one

Budget code	CC2186

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 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.