2. Aims

2.1 Develop a robust workflow for analysing somatic mutations in normal tissue

Studying somatic mutations in normal tissue is still a new field. Previous studies have made use of targeted sequencing of a collection of known cancer drivers (Martincorena et al., 2015). For whole genome sequence data, a matched analysis has typically been favoured given the efficient removal of germline mutations that can be achieved. However, to capture the early embryonic mutations, an unmatched approach is necessary (Behjati et al., 2014). This comes at the cost of calling both germline and somatic mutations. Therefore, my first aim was to design a bioinformatics workflow that can confidently exclude germline variants from true somatic variants, with high sensitivity and specificity.

2.2 Characterise the landscape of somatic mutations in the normal pancreatic islets

Somatic mutations in healthy pancreatic islets have so far been investigated only with single-cell RNA sequencing (Enge et al., 2017). As such, this is limited to the exome and is affected by a high rate of errors introduced by the whole-genome amplification stage. This limits the insight into the mutational processes acting on them, particularly early in development. Using whole genome sequence data, I intend to obtain estimates for the mutational burden and use these to deduce what mutational processes the islets are subjected to. This is also of relevance to better understand the mutational processes that may be active in the normal cells that give rise to pancreatic neuroendocrine tumours.

2.3 Elucidate the early phylogeny of the pancreatic islets

By using somatic mutations, I hope to obtain new insights into the development of the pancreatic islets. The first question would be to determine clonality, confirming whether all cells in an islet derive from a single founder cell or lineage, or whether different lineages contribute to an islet. If islets are monoclonal or at least oligoclonal, dominated by one or a few major lineages, it might be possible to reconstruct an embryonic lineage tree (Figure 8). Integrating this with the spatial distribution of the islets would then provide a glimpse into the anatomical shaping of the pancreatic endocrine tissue, during early embryogenesis.

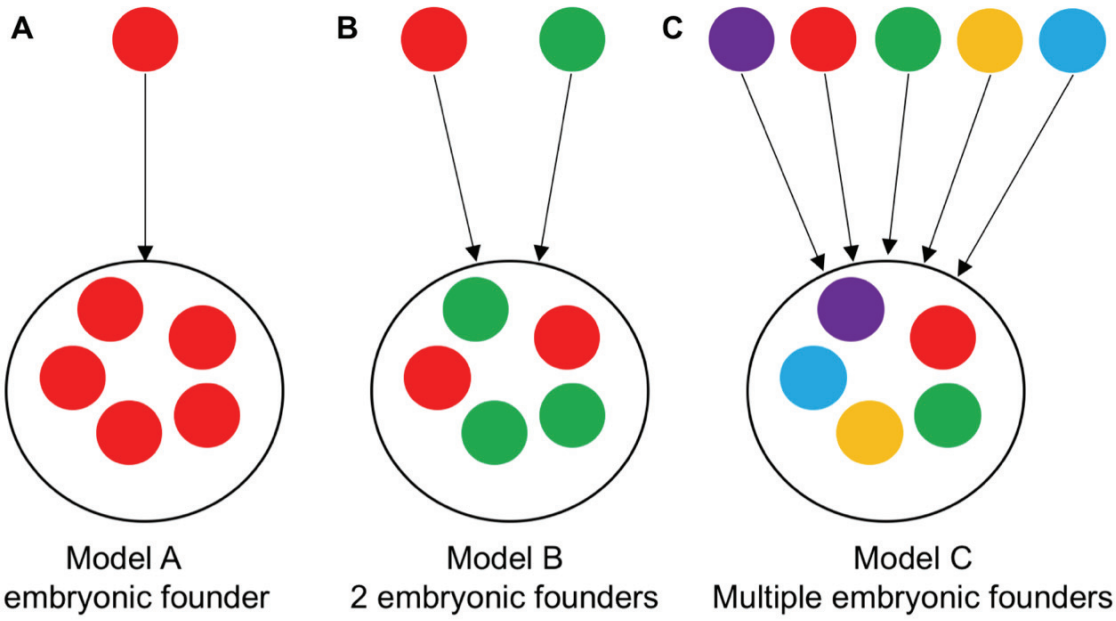

Figure 8 - Somatic mutations can be used to separate the embryonic lineages that contribute to the formation of the islets

(A) Model A shows a single cell founding an islet. The islet cells then go on to carry the same mutations as the founding cell. The VAF distribution of the entire islet would be centred on 0.5, in addition to the low VAF private mutations that have accumulated through life, in the individual islet cells. The MRCA of all the cells in the islet is the founding cell. Under this model, embryonic lineage trees can be obtained used standard phylogenetic methods such as maximum parsimony.

(B) Model B shows two different cells founding an islet. Each cell would carry ancestral mutations from either the red or green lineage plus their own private mutations. Mutations in each of the two founding lineages would have VAFs less than 0.5, but the sum should approach 0.5. The MRCA of the cells in the islet is the MRCA of the two founding cells. If each islet present has one or a few dominant lineages, embryonic lineage trees could be obtained using subclonal decomposition.

(C) Model C shows multiple cells founding an islet. Thus, the islet cells will share fewer ancestral mutations and have many more private mutations accounting for a smaller VAF each. If the contribution of different lineages does not vary across the islets, reconstructing the embryonic lineages under this scenario will not be possible.

2.4 Gain insight into the maintenance of the pancreatic islets through life

Some limited insights into the maintenance of the islet population throughout life might be obtained from the VAF distribution. If a large fraction of the cells of an islet was replenished by one, or a few, stem cells during adult life, this would be expected to manifest as peaks in the VAF distribution of the somatic mutations detected in the islet. A single stem cell could even take over the islet, much like those in the colonic crypts, and the resulting VAF distribution would show a single large peak at a high VAF. A similar distribution could also be obtained if islets are founded later in life by a single founding cell.

In contrast, if islets are maintained by the self-duplication of differentiated islet cells, or by a large number of slow cycling cells, or even if most islet cells are not replaced throughout life, clonal expansions within an islet would not be expected to reach detectable VAFs in adulthood. With this in mind, the VAF distributions may help differentiate between extreme models of tissue maintenance.