5. Discussion

5.1 LCM with an unmatched analysis may prove to be a reproducible workflow in other normal tissues

Capturing the landscape of somatic mutations within normal tissue is a growing field. At the Wellcome Sanger Institute (Cambridge, UK), a new pipeline using LCM of small areas of tissue and low-input DNA sequencing has been developed. Being able to take such precise biopsies, drawn freehand, makes few histological structures off limits. Whilst the spherical nature of the islets does pose an additional challenge in obtaining a complete three-dimensional sample, the use of multiple z-slices taken within the same islet, allows an approximation of the entire spherical islet to be made.

The study of somatic mutations has typically relied on matched data with the acknowledgement of the limitations this carries with regards to early embryonic mutations and phylogenetic reconstruction. However, the unmatched workflow presented here supersedes this matched approach. Germline mutations were confidently removed, identical somatic variants were called and, importantly, early embryonic mutations were recovered with minimal introduction of artefacts. Utilising this workflow, both prospectively and retrospectively, to similar data sets from other tissues could help decipher their somatic mutational profiles and early embryonic phylogenies. As the field grows and more normal tissues are investigated, the work here could prove pivotal in directing future somatic mutation research.

5.2 Novel insights have been gained into somatic mutations of the pancreatic islets

The almost unprecedented very low number of somatic mutations identified in each whole genome, as well as the pattern of mutation sharing across islets and across the pancreas and bladder, strongly suggests that many of the mutations detected in this analysis occurred during early embryogenesis. Signature analysis of the mutations detected confirmed that the majority of them can be assigned to intrinsic mutational processes without clear evidence of mutagen-induced mutations. This is perhaps to be expected for mutations of embryonic origin.

As expected given the very low mutational burden, there were few non-synonymous mutations identified. One notable variant in the *LIPI* gene stands out as a high VAF coding mutation (sample PD37726d_lo0056). Its high VAF in the absence of aberrant

copy number changes confirms that the mutation is present in approximately 90% of the islet cells (95% confidence interval: 56%-100%). Consistently, this islet appears to be mostly derived from a single branch of the phylogenetic tree. The high VAF of this variant could be consistent with it being an early passenger mutation present in an embryonic cell that gave rise to most cells in this islet, or with a later clonal expansion by drift or positive selection. Nevertheless, the low mutational burden in this islet, not dissimilar from other more polyclonal islets, suggests an early embryonic origin.

An interesting observation from this study is that the MRCA of all cells in the 32 islets of Langerhans also appears to be the MRCA of all cells in the ten matched bladder urothelium samples. This resembles the observation from a previous study in mice that showed the MRCA of the endoderm is also the MRCA of the ectoderm and mesoderm, with the suggestion this is likely to be the zygote (or at least the cell that gave rise to seemingly all cells in the adult) (Behjati et al., 2014). From the results presented here, it is unclear whether the MRCA of the pancreatic islets and bladder urothelium is the zygote, or whether it is simply the first cell that gave rise to all adult tissues.

5.3 The founding model of the pancreatic islets is still only partially understood

The presence of so many different embryonic lineages in each islet indicates that multiple embryonic founding cells of different lineages, come together to seed each islet. Given that many islets have all five branches of the second generation of the phylogenetic tree represented in their phylogeny, this suggests that at least five embryonic founding cells established these islets. This is consistent with a polyclonal founding model (Model C, Figure 8) for the pancreatic islets. However, the phylogeny of most islets suggests the existence of dominant lineages disproportionally contributing to each islet (Figure 34). This is in keeping with previous studies showing asymmetric contributions being made to adult tissues, from embryonic ancestors (Behjati et al., 2014; Ju et al., 2017).

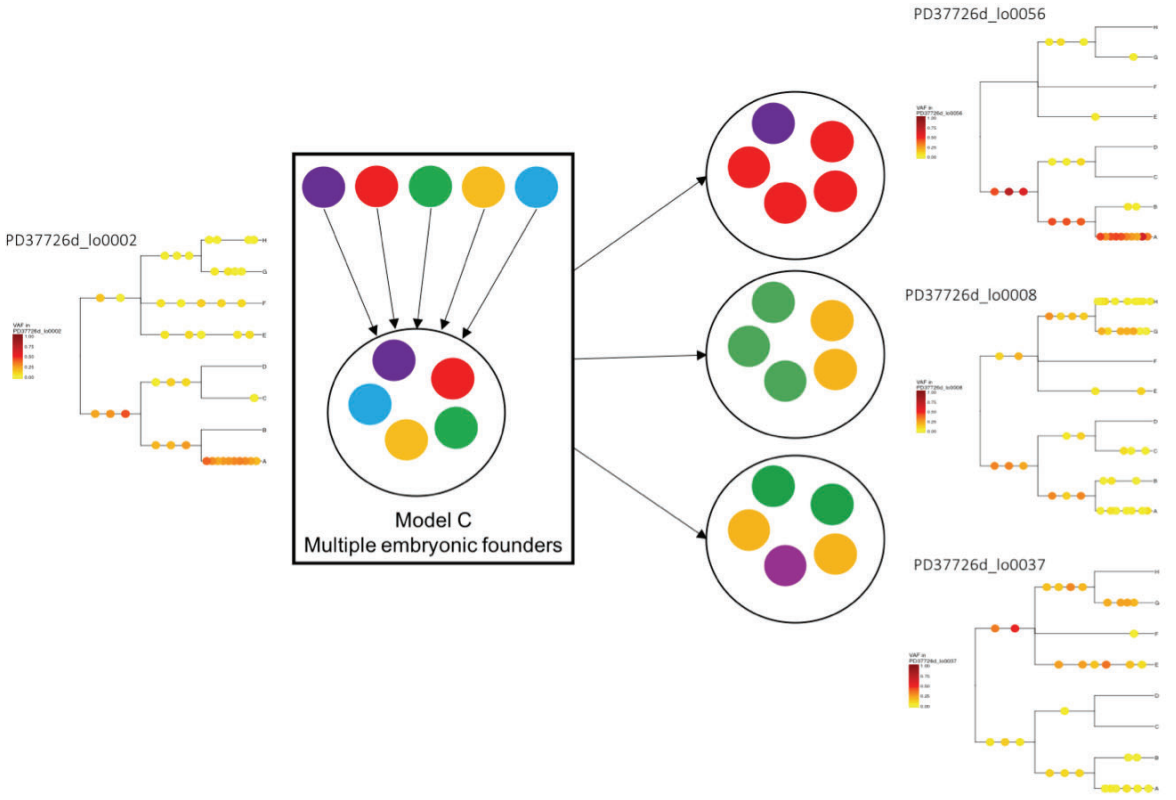

Figure 34 – Multiple embryonic founders may seed an islet

One interpretation of the data is that multiple founding cells (coloured circles) appear to make up the islets (black circle) (Model C, Figure 8). However, the proportion of lineages seeding the islet may vary. Some islets may be entirely polyclonal, whereby all five lineages occupy an equal proportion of the population (such as that represented in the box) or they may be oligoclonal, with examples shown on the right of the diagram. Indeed, some of the islets here appear to be more clonal than others, with PD37726d_lo0056 a notable example.

Since the pancreatic islets also contain other minor cell types, such as non-endocrine cells like endothelial cells, it remains to be shown whether the existence of a dominant lineage in many islets is due to a monoclonal origin of all endocrine cells, in a given islet, or whether endocrine cells in a single islet truly arise from multiple embryonic progenitors. Future studies combining information from adjacent areas of exocrine pancreas or additional phenotyping information from immunohistochemistry or single-cell RNA sequencing, may shed light on whether different lineages contribute to different cell types.

The results shown here are in stark contrast to the somatic mutations that can be detected in fast dividing clonal tissues like the colonic crypts. The patterns seen here in the islets instead suggest the islets are formed by a few founder cells early in development and do not undergo any subsequent clonal sweep later in life, as this would be expected to come with an increased burden of clonal mutations. This is consistent with the current belief that islets are maintained by the infrequent division of many cells, such as self-duplication of differentiated β-cells, and seem inconsistent with islets being replenished, or formed, in adulthood by one or a few cells (Bonner-Weir et al., 2012; Dor et al., 2004). Nevertheless, the findings here relate only to the 32 islets sampled.

Spatially, the founding of these islets appears to have been a non-random process, with islets nearby sharing similar ancestry. Further statistical methods would be necessary to quantify the extent of this with one option being a permutation approach with the islet positions and relatedness. The implication of this non-random distribution is that the same ancestral founding cells, occurring early in development, may seed regions of the pancreas during embryogenesis. Alternatively, once formed, islets may then divide into two identical islets through islet fission (Seymour et al., 2004). Several pairs of islets are seen that are close to each other with a high degree of resemblance in Figure 32 and these may represent examples of either the same founding cells or fission (Figure 35). Indeed, the two may not be mutually exclusive.

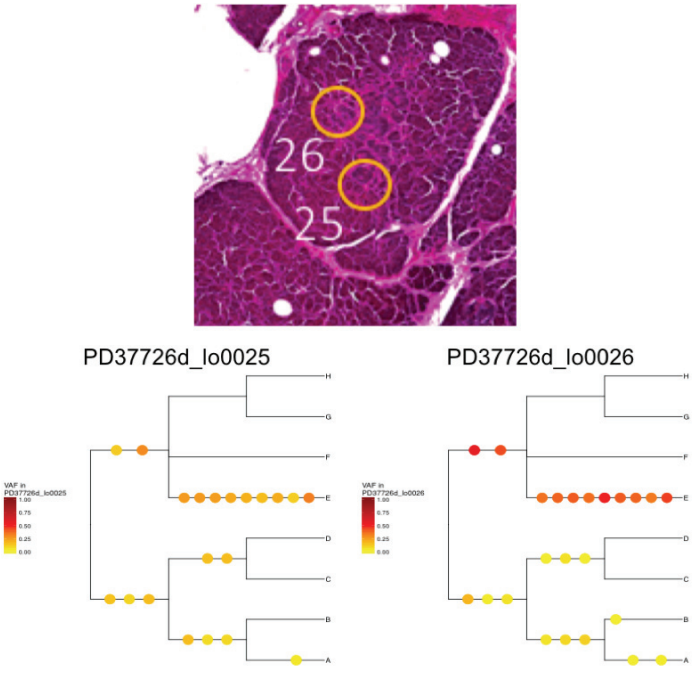

Figure 35 – Possible islet fission events are difficult to differentiate from similar founder populations

In light of their proximity to each other, samples PD37725d_lo0025 and PD37726d_lo0026 appear remarkably similar in their early phylogenetic ancestry. This may be due to an identical founder population or a fission event. Distinguishing the two will be difficult with the current method.

5.4 Limitations of the current methodology

There are several domains where improvements could be made. While LCM appears to be well-suited to the task, there are risks of contamination between plate wells. This could be from other samples during the cutting process, or when preparing the plate. In this study, this was closely monitored during the LCM stage through proper precautions and cleaning steps. The concordance of the duplicates and triplicates provides further support that there was little contamination of samples.

One key limitation was that all the islets were excised from a single biopsy, in one region of the pancreas, from a single patient. Whilst deep sampling of a single donor is necessary for phylogenetic reconstruction, multiple donors would be needed to make generalizable conclusions. To improve on this, many more islets would be required, from numerous biopsies across a range of donors.

Mutational signature analysis is a fast-growing field and has revealed numerous insights in cancer types. However, current knowledge of the mutational processes active in normal tissues is in its infancy. Whilst the same mutational processes may be occurring in normal tissue, there might be mutational signatures that are specific to certain normal tissues, and therefore not contribute sufficient numbers to malignancies for them to be detectable. By restricting the current analysis to signatures found in cancer genomes, the contribution of other signatures may have been overlooked. A *de novo* approach to mutational signature extraction may prove helpful in ascertaining the mutational processes specific to normal tissue, particularly when working with multiple patients. The R package HDP (https://github.com/nicolaroberts/hdp) (Roberts, 2018; Teh et al., 2006), could be used to do this. Still, one practical issue with *de novo* extraction is the need for a large number of mutations to work with, from

many different samples or patients with different contributions of these mutational signatures. Given the low mutational burden detected here, a *de novo* signature extraction would require a much larger number of islet whole genomes.

The minimum detectable VAF here was approximately 0.1, based on the WGS coverage and the CaVEMan default parameters. For a polyclonal tissue like the islets, the detectable variants were mostly ancestral mutations, most of which might have been present in the founder cells of an islet. Somatic mutations acquired through life by islets cells are unlikely to be present in a sufficient fraction of cells of an islet to be detectable. As a result, the mutational burden estimated in this study is expected to heavily underestimate the mutational burden of individual islet cells at the time of death of this donor, and instead likely represent the mutational burden of the founding cells.

A much greater coverage would enable the detection of more recent variants, and this would also be advantageous when looking at phylogenetic reconstruction. This would still be hindered by the polyclonal composition of the islets and the intrinsic errors introduced in sequencing. A possible solution to improve the recall of rare, or even private, somatic variants may be bottleneck sequencing (BotSeqS), whereby molecular barcoding combined with a dilution step prior to whole-genome library preparation, can dramatically increase the ability to identify those low VAF variants (Hoang et al., 2016). Additionally, the use of single-cell genomic and transcriptomic sequencing (G&T-seq) plus single-cell derived organoids, may play a role in the future somatic mutation workflow, particularly in polyclonal tissues (Enge et al., 2017; Jager et al., 2018; Macaulay et al., 2015).

Summarising, the sequencing of 32 pancreatic islet whole genomes, all from a single donor, has shown an unmatched analysis to be superior to a matched approach. The observed somatic mutational burden in these islets appears to be low and driven by intrinsic processes. Further, the pancreatic islets seem to be polyclonal units, established by multiple embryonic founders, with major and minor lineages. They do not appear to be maintained by a fast-dividing stem cell population and their spatial distribution is non-random, suggesting regions of the developing pancreas are seeded by the same populations of founding cells. Finally, the islets also appear to share a MRCA with the bladder urothelium, going back likely to the fertilised egg.