## 1. Introduction

1.1 Somatic mutations are acquired throughout life

Mutations are changes in the deoxyribonucleic acid (DNA) sequence. These can occur in the germline or the soma, to differing effects. Germline mutations are those mutations present in the haploid genomes of the gametes, that go on to fuse at conception. As a result, these mutations are inherited from the parental generation and are present in all the descendant cells of the totipotent zygote. Somatic mutations are those that occur any time after zygote formation and are not inherited from parental DNA. From the moment of conception, the zygote is under mutagenic pressure from intrinsic mutational process, one example being DNA replication errors. As the first cells undergo rounds of cleavage, each division is an opportunity for further mutations to occur. With time and exposure, extrinsic mutational pressures such as ultraviolet (UV) radiation and tobacco smoking, come to play a part in the development of somatic mutations (Stratton, Campbell, & Futreal, 2009).

Depending on their functional impact, somatic mutations can be classified as drivers or passengers (Figure 1) (Stratton et al., 2009). Drivers are a very small minority of somatic mutations that provide a phenotypic benefit to the cell. Often, these drivers are non-synonymous coding mutations, although some in non-coding regions of the genome can also act as drivers (Horn et al., 2013; Huang et al., 2013). The phenotypes bestowed upon the cell overlap with the hallmarks of cancer and include sustained proliferative signalling, enabling replicative immortality and resisting cell death (Hanahan & Weinberg, 2011). Due to the conferred growth advantage, cells with drivers have a relative gain of fitness over their neighbours, leading to positive selection and clonal expansions, by Darwinian evolution. This represents a critical step in carcinogenesis, as these drivers become causally implicated in the emergence of a future tumour (Stratton et al., 2009). Passengers on the other hand do not result in a growth advantage for the cell. These mutations include nearly all non-coding variants and the vast majority of coding mutations in genes not implicated in cancer. Passengers essentially "hitchhike" with those drivers that power a clonal expansion, as they too are present in the genome that is being positively selected for (Stratton et al., 2009).
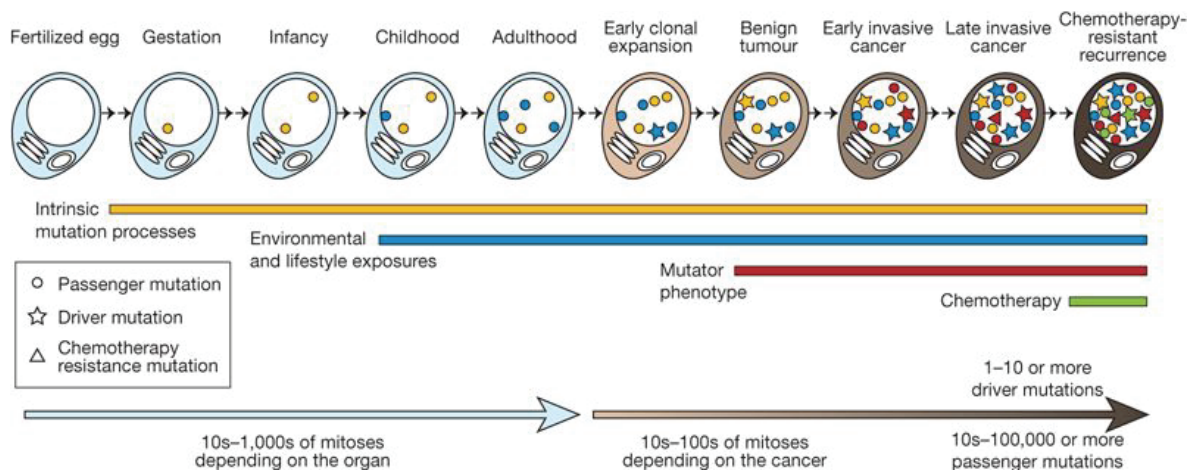
Figure 1 – Somatic mutations occur throughout life

Driver and passenger mutations accumulate throughout life. Drivers lead to clonal expansions that can emerge as tumours. There are several different mutagenic processes that contribute to these somatic mutations (Stratton et al., 2009).

1.2 The somatic mutational burden in normal tissues is comparable to some tumours

The mutational burden is the observed number of mutations that have been progressively acquired by a cell population. With increasing numbers of sequenced cancer genomes, the mutational burden across cancer types has become increasingly well-studied. Across 2,583 donors in the Pan-Cancer Analysis of Whole Genomes network (PCAWG), 43,778,859 single nucleotide variants (SNVs) have been detected across an array of tumour types, revealing an unprecedented insight into the typical numbers of mutations per tumour. These mutational burdens range from 10,000 to 100,000, in some of the most highly mutated cancer types such as UV-associated melanoma and tobacco-induced squamous cell lung cancer, to as low as 100 mutations per genome, in some bone and brain cancers (Campbell, Getz, Stuart, Korbel, & Stein, 2017).

Understanding the somatic mutations that arise early in cancer development, perhaps before the "mutator" phenotype exists, requires an in-depth look at healthy tissue. Prior to neoplastic transformation, healthy tissue would be expected to be harbouring somatic mutations and possibly the first driver, that can eventually lead to cancer (Stratton et al., 2009). Using deep, targeted sequencing of 74 known cancer genes, in 234 biopsies of healthy skin samples, Martincorena et al., (2015) revealed mutational

burdens averaging two to six base substitutions per megabase. This equates to genomes of a normal skin biopsy harbouring up to 30,000 mutations. Remarkably, in these histologically normal skin samples, the mutational burdens across all four patients were comparable to those seen in some skin cancers and several solid tissue malignancies (Martincorena et al., 2015).

1.3 The mutational signatures in normal tissues and cancer give insight into mutational processes

Elucidating the processes that drive the accumulation of somatic mutations provides insights into carcinogenesis. Mutations can be classified according to different features with different mutational processes often inducing distinct patterns of somatic mutations. These patterns can be used as fingerprints, or signatures, of mutational activity. The key features involved in modelling signatures were set out by Alexandrov et al., (2013):

1. The type of mutations observed, such as single base substitutions, insertions/deletions or chromosomal rearrangements;
2. The local sequence context, such as the bases that precede and follow a base substitution;
3. The location of the mutations throughout the genome, such as in particular regions susceptible to a certain mutagenic process or spatial clustering of mutations;
4. DNA damage repair mechanism involvement, as this leaves tell-tale marks on the DNA sequence and contributes to mutagenesis itself.

With increasing amounts of data and new analytic methods, the list of mutational signatures has continued to grow, from an initial 22, to the COSMIC-30 and recently the PCAWG-65 (Alexandrov et al., 2018; Forbes et al., 2017; Nik-Zainal, Alexandrov, et al., 2012). The PCAWG-65 mutational signatures are based on 84,729,690 somatic mutations, with 49 of these signatures relating to single base substitutions (SBS) (Alexandrov et al., 2018). Many different aetiologies, occurring in numerous cancer types, have been assigned a mutational signature including smoking tobacco and defective DNA damage repair due to BRCA 1/2 mutations (Alexandrov et al., 2018).

Despite these signatures having been defined in cancers, normal tissue also displays evidence of mutational signatures. For example, normal sun-exposed skin displays a high burden of C>T mutations at dipyrimidine sites caused by transcription-coupled repair of UV-induced DNA damage (Martincorena et al., 2015). This results in the high prevalence of the mutational signature SBS7a being found in histologically normal tissue, having previously been well-documented in UV-associated melanoma, as well as head and neck squamous cell carcinoma (Figure 2) (Alexandrov et al., 2018). The implication being that analysis of mutational signatures in normal tissues can shed light on the mutational processes driving precancer evolution.
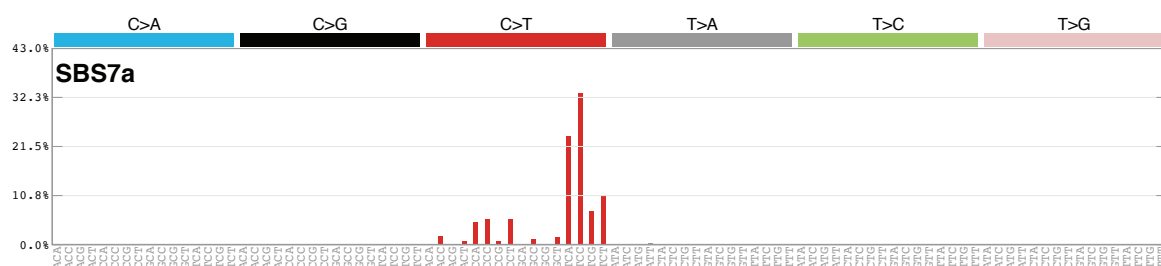

Figure 2 – Mutational signatures each have a unique profile of mutations

A 96-trinucleotide bar plot showing mutational signature SBS7a, one of the PCAWG-65 (Alexandrov et al., 2018). Each single base substitution is shown in the context of the pyrimidine bases involved, with the 5' and 3' bases included to make a trinucleotide. SBS7a is associated with UV light shows an excess of C>T substitutions, particularly in the context of TpCpA and TpCpC (Alexandrov et al., 2018).

1.4 The clonality of a tissue sample can be estimated using the somatic mutations present

The fraction of DNA molecules, within a sample, that harbour a given mutation is termed the Variant Allele Frequency (VAF). For example, inherited germline heterozygous mutations present in all diploid cells of the body will show VAFs around 0.5. This is because one of the two copies of the genome in every cell contains the mutant allele (Figure 3A). In contrast, somatic mutations occur once the zygote has been formed and are only present in a fraction of all somatic cells in an individual (Figure 3B). Somatic mutations occurring in the first few divisions of the embryo can appear in a considerable fraction of cells in the adult and are often termed mosaic

mutations, while late occurring mutations are typically constrained to small clones within a tissue.
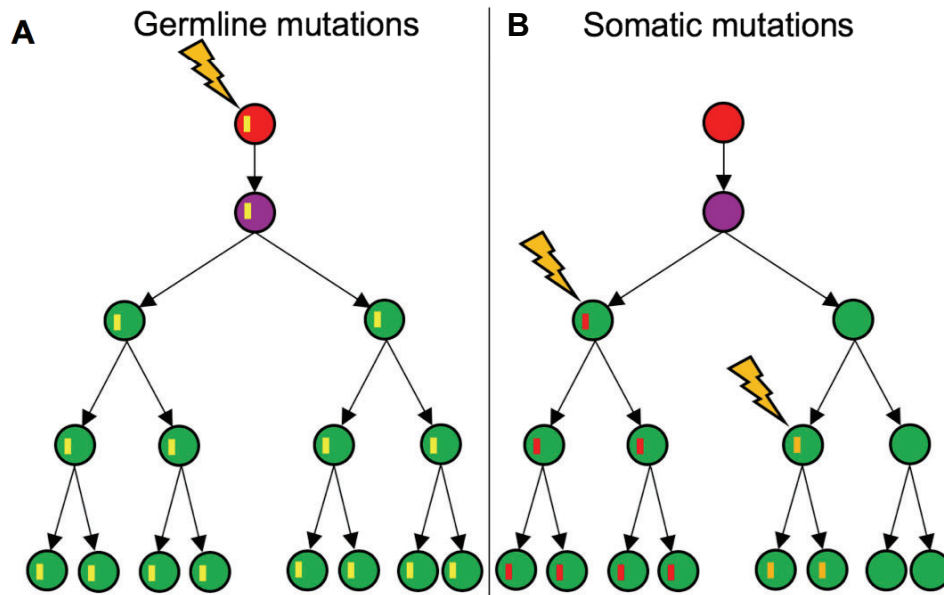
Figure 3 – The VAF of a mutation can be used to assess clonality

Phylogenetic trees displaying the ancestry of a cell population. Mutations (yellow lightning strike) are passed onto progeny. The parental lineage is shown by the red circle and the fertilised egg is the purple cell. Somatic descendants are shown by the green circles. Heterozygous variants are shown by the coloured bars in circles.

(A) A heterozygous germline mutation results in a VAF of 0.5 as they are present in all cells.

(B) Heterozygous somatic mutations will produce a variety of VAFs, according to how early they occur in development and to what degree a tissue is composed of lineages carrying the somatic variant. The exception to this is a somatic mutation occurring in the fertilised egg itself, which would give a VAF of 0.5.

The clonality of a sample can be studied by analysing the VAF distribution of all the mutations within a sample. Heterozygous variants in a clonal sample, one where all cells carry the same mutations and are thus closely related, would be expected to show binomial variation around a VAF of 0.5, assuming a diploid genome. Colonic crypts are a well-known example of a clonal tissue. Although each crypt contains multiple stem cells, by mere drift, single stem cells frequently take over a crypt

(Snippert et al., 2010). This leads to all cells of a crypt recently deriving from the same stem cell and manifests as a VAF distribution centred around 0.5. Contrasting this, a polyclonal sample would have fewer mutations at a high VAF as different cells carry different mutations, each representing a small proportion of the sample. Within the sample VAF distribution, these subclonal populations may then be represented as multiple peaks each with mean VAFs less than 0.5 (Figure 4) (Nik-Zainal, Van Loo, et al., 2012).
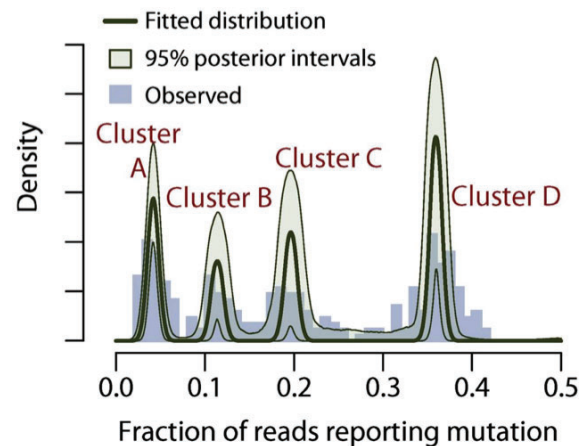
Figure 4 – The VAF distribution reflects the clonality of a sample

A polyclonal sample has numerous clusters of VAFs that represent subclones in the cell population (Nik-Zainal, Van Loo, et al., 2012). A Bayesian Dirichlet process has been used to model the VAF distributions with 95% posterior confidence intervals displayed in green. Four subclones are present, with cluster D being the dominant lineage, as it has the highest VAF (Nik-Zainal, Van Loo, et al., 2012).
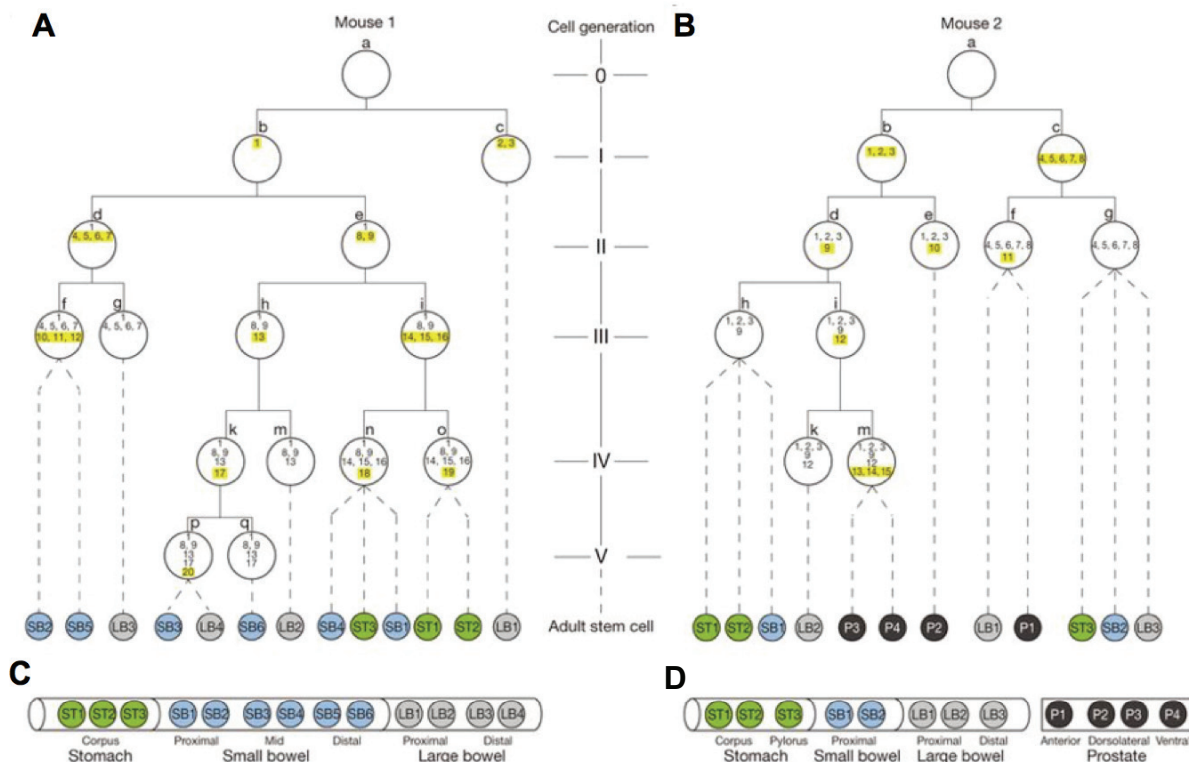
1.5 Phylogenetic tree reconstruction of early embryogenesis has been demonstrated in clonal organoids

Utilising somatic mutations to reconstruct a phylogenetic tree, in a clonal tissue, has previously been demonstrated in mice (Behjati et al., 2014). Clonal organoids derived from the stomach, small and large bowel and the tail of two mice, were sent off for whole-genome sequencing and the variant caller, CaVEMan (Cancer Variants through Expectation Maximization) was used to identify somatic mutations (Behjati et al., 2014; Nik-Zainal, Van Loo, et al., 2012). Initially, this variant calling was performed with a matched tail sample to ensure efficient removal of germline variants. A subsequent unmatched run captured the entire complement of germline and detectable somatic

mutations and by comparing this to the matched run, the germline mutations could again be removed, leaving behind those variants exclusive to the unmatched run. After capillary sequencing of these exclusive variants, 35 were confirmed. As they are shared between the organoids and the matched tail sample, they likely occur early in embryonic development.

Maximum parsimony was then used to reconstruct a phylogenetic tree detailing the hypothetical order of mutation acquisition (Figure 5) (Behjati et al., 2014). This totalled 23 cell divisions across two trees, one from each mouse. Both were resolved to a single ancestral origin and although this first cell may be the zygote, the possibility of silent cell divisions and lack of statistical power in distinguishing real differences in read counts, means it isn't certain that this is the case. The mutation rate in early embryogenesis was estimated at 1.5 mutations per cell division. Importantly, the reconstructed tree represents the earliest divisions in the embryo and pre-dates gastrulation, confirming that germ layers are polyphyletic in origin, formed by the spatial aggregation of cells from different lineages (Behjati et al., 2014).

Figure 5 – Phylogenetic tree reconstruction of different tissues in mice

(A, B) Reconstructed phylogenetic trees for the two mice studied (Behjati et al., 2014). The numbers inside each node on the tree represent a unique mutation and by tracing the branches of the tree, the mutations can be seen accumulating in the most recent generations.

(C, D) The coloured circles at the tips of each tree branches indicate the tissue to which the lineage ultimately contributes.
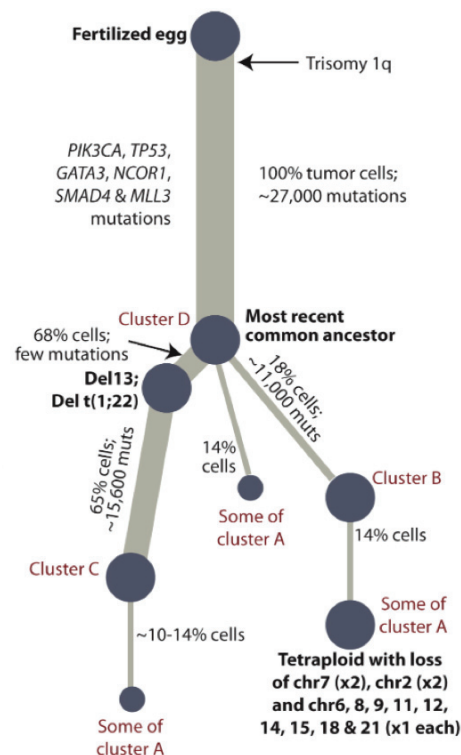
1.6 Phylogenetic reconstruction with subclonal tissues requires a framework to identify cell populations

While single-cell derived clones, such as organoids, enable an easy reconstruction of phylogenetic trees, standard phylogenetic methods are not suitable when sequencing polyclonal populations of cells. In order to reconstruct phylogenetic trees from polyclonal populations, new methods had to be developed that first group mutations in discrete subclones and then build trees of the subclones. For example, Bayesian Dirichlet processes have been used in studies of breast and prostate cancer (Gundem et al., 2015; Nik-Zainal, Van Loo, et al., 2012). The premise of this is that by clustering variants together based on their VAFs, distinct subclones can be defined within the population.

This was demonstrated with 21 breast cancer genomes, whereby whole-genome sequencing, copy number analysis and somatic variant calling with CaVEMan produced a list of variant calls in each sample (Nik-Zainal, Van Loo, et al., 2012). Applying the Bayesian Dirichlet process based on the coverage of the variant read site and the VAF, the clustering of mutations reveal distinct subpopulations. Amongst these subpopulations in each sample, there was a dominant lineage, which accounted for more than half of the sample (Nik-Zainal, Van Loo, et al., 2012). Given the high numbers of shared mutations between clusters, these different populations appear to co-exist for a significant portion of their life history, before diverging into separate subclones. Applying the pigeonhole principle to these subclones enabled the order of mutation acquisition to be inferred and as such, a phylogenetic tree was reconstructed for each sample, the origin of which is the most recent common ancestor (MRCA) of all the identified subpopulations (Figure 6) (Nik-Zainal, Van Loo, et al., 2012). In this

way, somatic mutations, copy number information, mutation phasing and clustering methods (such as the Bayesian Dirichlet process) can be used for phylogenetic reconstruction in a single polyclonal sample.

Figure 6 – Phylogenetic reconstruction of a single breast cancer sample
The somatic mutations identified within a tumour population can then be clustered together based on their shared variants with the Bayesian Dirichlet process. Applying the pigeonhole principle can then place these clusters in sequential order, tracing their phylogenetic lineage back to the MRCA (Nik-Zainal, Van Loo, et al., 2012).

While basic phylogenetic trees depicting the relationship between a few subclones can be inferred from a single sample, phylogenetic reconstruction from polyclonal samples is greatly helped by sequencing multiple related samples, such as sequencing multiple regions of a tumour. This is exemplified by a study that performed whole-genome sequencing and somatic variant calling in 51 tumour samples, obtained from ten patients with metastatic prostate cancer (Gundem et al., 2015). An n-dimensional Bayesian Dirichlet process was applied, enabling the identification of clonal and subclonal populations within each sample per patient. By retracing the phylogeny of

multiple tumour samples from the same patient, some including both the primary and secondary tumours, remarkable insights were gained into the metastatic process. Minor subclonal populations appeared to be responsible for the initiation of metastasis and in several cases, multiple subclonal populations from the same tumour appeared to independently achieve metastatic potential (Gundem et al., 2015). Furthermore, not only do metastases appear to *de novo* seed new metastases, but multiple metastases can seed a new metastatic deposit, forming a polyclonal foundation (Gundem et al., 2015). This rapidly diversifies the tumour populations in each secondary tumour and provides a new spatial dimension to the evolutionary history of cancer

Summarising, phylogenetic reconstruction requires accurate somatic variant calling, particularly for the reconstruction of early embryonic lineage trees, in which early branches can be supported by one or a few variants. The Bayesian Dirichlet process provides a framework within which the subclonal populations of a sample, and those between samples, can be identified. The relationship between samples can then be inferred using these clusters while the pigeonhole principle allows the deduction of the sequence in which these populations arose. By applying these principles to normal tissues, novel insights into embryological development, tissue maintenance and carcinogenesis can be sought.

1.7 The pancreatic islets perform endocrine functions

The pancreas is a glandular organ situated in the upper region of the abdomen, with dual exocrine and endocrine functions. The exocrine tissue forms the majority of the parenchyma of this organ and consists of acini and ducts (Figure 7). The acini produce and secrete pancreatic juice, an alkaline solution rich in digestive enzymes, into the branched ductal network which then drains into the duct of Wirsung and into the duodenum via the ampulla of Vater (Horiguchi & Kamisawa, 2010).

In contrast, the islets of Langerhans, or pancreatic islets, are spherical micro-organs distributed throughout the parenchyma of the pancreas that undertake numerous endocrine functions (Figure 7). Accounting for just over 2 cm^3 of tissue in an average adult human, the pancreatic islets are a mosaic of several cell types including α-cells, β-cells, δ-cells, ϵ-cells and PP-cells (Ionescu-Tirgoviste et al., 2015). The most

common cell type within the islet are β-cells, accounting for 60%, followed by α-cells making up 30% and the remaining 10% being δ-cells, ε-cells and pancreatic polypeptide cells (PP-cells) (Cabrera et al., 2006; Ionescu-Tirgoviste et al., 2015). Fundamental to glucose homeostasis, the β- and α-cells are locked in negative feedback pathways. In the presence of glucose, β-cells produce insulin, a peptide hormone with the primary aim of empowering tissues to utilise the glucose from the bloodstream. In contrast, glucagon from the α-cells acts to increase blood glucose levels from stores in the muscle and liver. Together these opposing functions form a tightly regulated system that promotes euglycaemia. It is the dysregulation of this homeostasis that results in diabetes mellitus (Zheng, Ley, & Hu, 2018).
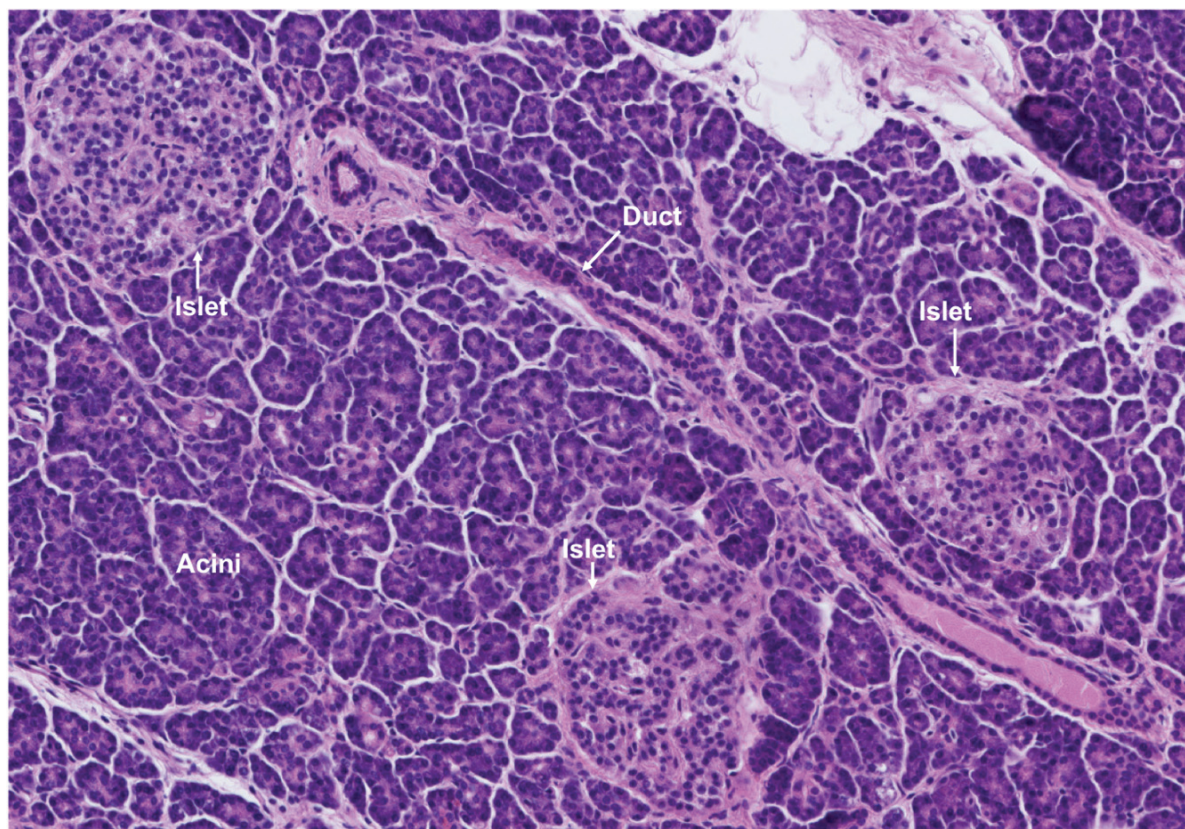
Figure 7 - An overview of pancreatic histology

Pancreatic section (5 μm thickness), taken from patient 290B. Exocrine tissues are the acini and ducts whereas islets are the endocrine component. Whilst exocrine and endocrine tissues are in close proximity, they greatly differ in their functions.

<u>1.8 Primitive islets develop early in foetal development and numbers peak in the post-natal period</u>

The pancreas is first evident as a developing endodermal embryological structure at human gestational day 26 (Piper et al., 2004). By 47 days post-conception, cells expressing the PDX1 transcription factor first appear (Jennings et al., 2013). These progenitors are capable of becoming either a ductal or endocrine cell. The expression of NGN-3 at weeks 7 and 8 diverts these PDX1 positive cells to the endocrine lineage and the first insulin expressing cells are seen around this time (Jennings et al., 2013). This marks a clear difference to the murine model whereby the initial hormone expressed is glucagon (Jennings et al., 2013; Rall, Pictet, Williams, & Rutter, 1973).

Over two more weeks of gestation, the other endocrine cell types emerge, although the β-cells remain the most prevalent endocrine cell during the first trimester (Piper et al., 2004). Key transcription factors that play a role in this continued endocrine development, through weeks 9 to 21, include NKX2.2, NKX6.1, ISLET1, NEUROD1 PAX4 and 6 (Jennings et al., 2013; Lyttle et al., 2008; Sarkar et al., 2008). In contrast, loss of SOX9 expression appears linked with the differentiation of progenitor cells into foetal β-cells (Jennings et al., 2013).

Clusters of endocrine cells first appear around week 12 of gestation. By week 14, these primitive islets develop a vascular network (Jeon, Correa-Medina, Ricordi, Edlund, & Diez, 2009; Piper et al., 2004). These clusters consist initially of more β-cells than α-cells, but this ratio balances out by week 16, remaining at 1:1 until birth (Gregg et al., 2012; Jeon et al., 2009; Riedel et al., 2012). It is the α-cells and δ-cells that show a greater proliferative index compared to the β-cells in the remaining pre-term period (Jeon et al., 2009; Sarkar et al., 2008).

In the neonatal period, β-cell numbers increase compared to the static α-cell population (Gregg et al., 2012). Whilst β-cell neogenesis, where ductal progenitors differentiate into β-cells, is more common in the pre-natal developing pancreas, this does not play a prominent role in the post-natal period. Instead, proliferation of existing β-cells accelerates, reaching a peak of 2% before ceasing by two years old (Gregg et al., 2012). This period also involves many immature islets taking on a more familiar

architecture, seen in the adult pancreas, and by six months all islets have reached this point (Gregg et al., 2012). This coincides with the nutritional shift that occurs at weaning and microRNAs, such as miR-17-5p and miR-29-3b, have been shown to play a key role at this critical time (Jacovetti, Matkovich, Rodriguez-Trejo, Guay, & Regazzi, 2015).

1.9 There are physiological, and pathological, causes for β-cell proliferation

After post-natal proliferation ceases, the proportion of β-cells proliferating, at any one time, drops to approximately 0.5-1% and continues to decrease further with age (Gregg et al., 2012). Even following a loss of endocrine tissue from a partial pancreatectomy, little evidence of β-cell proliferation has been observed (Menge et al., 2008). Human studies using *in vivo* thymidine analogue incorporation combined with radiocarbon, and lipofuscin accumulation, have both supported this, with the suggestion that final β-cell populations are defined before age 30 with little activity afterwards (Cnop et al., 2010; Perl et al., 2010). Only in rare, sporadic cases has β-cell neogenesis been observed in specimens obtained from donors older than five-years-old (Gregg et al., 2012). In light of this, the vast majority of β-cells appear to remain in a quiescent state through life.

The primary pathological cause of β-cell proliferation in adulthood is seen in diabetes mellitus type 2. Recognised by the World Health Organization as an important public health problem, estimates in 2015 put the age-standardised global prevalence of diabetes mellitus, both type 1 (DM1) and type 2 (DM2), at one in eleven adults (World Health Organization, 2016; Zheng et al., 2018). The majority of these are believed to be patients with DM2 (World Health Organization, 2016; Zheng et al., 2018). Primarily a disease driven by insulin resistance in the liver, muscles and islet cells, the dysregulation of glucose homeostasis that occurs in DM2 triggers compensatory β-cell hyperplasia (DeFronzo & Tripathy, 2009; El Ouaamari et al., 2016; Escribano et al., 2009).

While the initial compensatory proliferation and associated increased insulin secretion can help cope with the insulin resistance, the hyperinsulinaemia that results actually drives further insulin resistance and glucose production. This positive feedback

eventually overwhelms the compensatory mechanisms and hyperglycaemia prevails (Zheng et al., 2018). This produces the clinical symptoms often associated with DM2 including polyuria, polydipsia and fatigue. With DM2 established, significant changes then follow in the pancreatic endocrine tissue resulting in a reduced β-cell mass, altered β:α-cell ratios, co-expression of endocrine hormones and loss of β-cell identity (Butler et al., 2003; Enge et al., 2017; Mezza et al., 2014; Spijker et al., 2015). Disease progression often mandates insulin replacement therapy and significant macro- and micro-vascular complications become increasingly more prevalent (Fowler, 2008).

A more physiological cause of β-cell proliferation is seen in pregnancy. The introduction of placental lactogens and growth hormones drives hepatic gluconeogenesis and lipolysis, leading to hyperglycaemia and insulin resistance (Beck & Daughaday, 1967; Rieck & Kaestner, 2010; Sorenson & Brelje, 1997). In response, a 1.4-2.4-fold increase in β-cell mass has been demonstrated (Butler et al., 2010; Van Assche, Aerts, & De Prins, 1978). While in many women this is sufficient and entirely normal, if the insulin resistance is too great and there are other risk factors present, gestational diabetes can arise. The specifics of how the β-cells proliferate remains unclear, with both self-duplication of β-cells and islet cell neogenesis being hypothesised (Butler et al., 2010; Van Assche et al., 1978).

1.10 The maintenance of the pancreatic islets, through adulthood, is unclear
The maintenance of adult pancreatic endocrine tissue has been studied extensively, albeit mainly in model organisms, with numerous hypotheses generated. Mechanisms suggested for islet cell maintenance include self-duplication of existing differentiated β-cells, neogenesis of new islets through transdifferentiation of ductal cells and progenitor/stem cell replenishment.

Originally proposed many decades ago by Messier and Leblond (1960), self-duplication has been best demonstrated using a Cre/lox pulse-chase system in adult mice (Dor, Brown, Martinez, & Melton, 2004). β-cells were labelled and following the chase, the fraction of β-cells per islet was assessed. Over 12 months, self-duplication of pre-existing β-cells should not alter this fraction whereas stem-cell and progenitor renewal would. The results revealed little change in the fraction, but an increase in

endocrine tissue mass. This indicated self-duplication to be the main proliferative pathway. Given that the β-cells were observed to increase in number, this also challenged the notion that islet cells were post-mitotic (Dor et al., 2004). This has been supported by subsequent studies confirming that all β-cells retain the capacity to self-duplicate and that each cell appears to contribute equally to the maintenance of the islet (Brennand, Huangfu, & Melton, 2007).

However, there remain aspects of pancreatic islet maintenance that cast doubt on self-duplication being the only mechanism for islet proliferation. These are mostly focused on the potential that stem cells and progenitors have to differentiate into β-cells. Several different candidates have been suggested to exist, with the locations harbouring these stem cells and progenitors including the pancreatic ductal epithelium and the islet itself (Bonner-Weir, Baxter, Schuppin, & Smith, 1993; Zulewski et al., 2001). One such example supporting a progenitor hypothesis, involved the xenografting of human embryonic pancreases, with PDX1+ and Ngn-3+ progenitors, into immunocompromised mice. Whilst the PDX1+ progenitors differentiated into β-cells, the Ngn-3+ progenitors did not, suggesting differentiated endocrine cells were unable to self-replicate (Castaing, Duvillie, Quemeneur, Basmaciogullari, & Scharfmann, 2005; Castaing et al., 2001).

The definitive existence of pancreatic islet stem cells is proving difficult to confirm, with recent forays into the single-cell transcriptomics of pancreatic islets, failing to identify a single stem cell lineage (Muraro et al., 2016). This does not completely rule out the stem cell theory, as there may exist multiple, different stem cell populations that contribute to islet maintenance. However, given that these stem cells appear to be extremely rare within the islet, isolating even one population with single cell transcriptomics will require far larger data sets (Andrews & Hemberg, 2018).

Transdifferentiation of non-endocrine cells, such as the pancreatic ducts and acini, into endocrine cells, has also been suggested, particularly under injury. Given the translational potential of islet neogenesis for the treatment of diabetes mellitus, this hypothesis has garnered much attention. Researchers have transformed *in vitro* acini, islet cell precursors and even splenocytes into functioning β-cells (Guz, Nasir, &

Teitelman, 2001; Kodama, Kuhtreiber, Fujimura, Dale, & Faustman, 2003; Lipsett & Finegood, 2002; Socorro et al., 2017). Recent work has revealed peripheral regions of islets harbour immature β-cells that appear to be descended from nearby α-cells, the implication being that transdifferentiation may occur within the pool of different endocrine cells themselves (Chakravarthy et al., 2017; van der Meulen et al., 2017).

Finally, whether an entire islet unit can duplicate itself, in a "fission" event is debated. Fission has been well-proven in the colonic crypts, both in the post-natal period and in adulthood, as has crypt fusion (Bjerknes, 1986; Bruens, Ellenbroek, van Rheenen, & Snippert, 2017; Cheng & Bjerknes, 1985; Clarke, 1972). In the pancreatic islets, fission has been postulated using X-inactivation mosaic mice with lacZ insertion, on the X-chromosome (Seymour, Bennett, & Slack, 2004). Islets were identified that appeared to have an irregular morphology, whereby two small masses of endocrine cells appear to be linked by an isthmus of α-cells. These were named "dumb-bell" islets (Seymour et al., 2004). By comparing the X-inactivation status, and hence lacZ expression, of the masses on either side of the isthmus, it was deduced that the masses on either side of the isthmus were more related to each other than two randomly selected nearby islets were (Seymour et al., 2004). Further, comparing distinct islets to each other revealed this same measure of similarity decreased as the distance increased between them. The conclusions drawn were that these dumb-bell islets were in a state of fission, rather than fusion (Seymour et al., 2004).

1.11 Summary

Identifying somatic mutations has proved successful in cancer and the stage is set for studying normal tissue. The pancreatic islets represent a high-priority normal tissue to investigate, given the scale of the health burden that DM2 poses. While efforts have been made with single-cell RNA sequencing to decipher the somatic mutational landscape of the pancreatic islets, these methods continue to be burdened by a high false discovery rate (Enge et al., 2017). By establishing a workflow using whole-genome sequencing and laser capture microdissection (LCM), the somatic mutational profile of the islets can be examined and key questions regarding the development and maintenance of the pancreatic endocrine tissue can hopefully be answered, opening up the possibility of translational benefits in pancreatic islet disease.