DISCUSSION

In this dissertation, I have presented two case studies of somatic evolution. One investigates clonal dynamics, and the other somatic mutations, using blood and colon respectively as the model tissues. Here, I will discuss how this work informs our understanding of somatic evolution, drawing out future perspectives for how their extension, and the development of the field more broadly, might change the way in which we manage diseases that stem from natural selection within a multicellular organism.

1. Discussion of findings

1.a. The clonal dynamics of blood

Our study is the first to investigate haematopoiesis in an unperturbed healthy human using spontaneous somatic mutations. Our method is analogous to lineage tracing experiments in animal models, but we were able to use thousands of nested markers, detecting clones that arose at different times throughout life. We provide proof-of-concept that somatic mutations are a powerful tool to dissect the stem cell dynamics of normal tissues, as they have been used to study population dynamics and cancer evolution.

There is no previous *ab initio* estimate of the number of blood stem cells in a human in steady state. The value that we derive of 97,000 (90% CI 45,000−215,000) stem cells contributing to granulopoiesis will require validation by studying a large number of individuals. A more precise estimate could be obtained by whole genome sequencing a larger number of cells and, probably more importantly, developing a method of error corrected sequencing that would enable us to detect rarer mutations with high confidence. Methods related to duplex sequencing (Schmitt et al. 2012, Kennedy 2014, Hoang et al. 2016) may provide such an opportunity.

Quantification of the number of haematopoietic stem cells is an important biological insight and is also crucial to our understanding of the development of malignancy. As stem cells are believed to be at risk of transformation (General Introduction) each one may represent a potential cancer. In their discussion of cancer risk, Tomasetti and Vogelstein estimate a number of blood stem cells that is two orders of magnitude greater than ours (Tomasetti and Vogelstein 2015, General Introduction). Other factors, such as microenvironmental changes (General Introduction), are likely to be involved in determining the probabilities of mutant clones arising and outcompeting their neighbours, but our estimate provides a step on the road to a quantitative model of cancer development.

It remains to be shown precisely how the cells that we define as stem cells based on their long-term self-renewal *in vivo* map onto those at risk of malignant transformation. Our observation of granulocyte-B cell restricted clones in adults indicates that the cells that we are counting have both lymphoid and myeloid potential. Our study could be expanded to all phenotypically distinct cell types, resolving the clonal contribution of stem cells to each. We would, in theory, be able to build a quantitative map of the differentiation hierarchy of blood, calculating the number of stem cells responsible for maintaining each cell type. The gamut of haematological malignancies could be investigated, revealing the clone that gave rise to the neoplasm, and the lineage restriction of the line-of-descent that predates transformation. As clonal haematopoiesis is a common feature of ageing it could be followed prospectively, tracking its emergence from a normal clone through longitudinal blood draws. The method could further be applied to bone marrow failure syndromes, inflammatory conditions, and ageing.

1.b. Somatic mutations in normal colon

1.b.i. Driver mutations

We investigated the landscape of somatic mutations in hundreds of normal cells from 42 people. Due to the large number of crypts investigated, we were able to capture infrequent events, such as driver mutations. Despite sequencing over 1,000 crypts, driver mutations are so rare that assaying even more crypts will be necessary to develop an accurate estimate of the frequency of

each individual mutation in normal colon. Studies in mice and humans have begun to quantify the selective advantage that mutations confer in competition within the crypt and between crypts (Results Chapter 2, section I.5.). The laser capture microdissection method developed here would allow these studies to be extended to all driver mutations in humans. Ultimately, the selective advantage of every mutation, in different genetic and microenvironmental contexts, would allow us to chart all the possible paths to cancer and their respective probabilities, indicating the best points to intervene. It might be that early 'clonal purging' is the best approach. For example, the presence of *PIK3CA* mutations in normal colorectal mucosa, combined with a recent successful trial of a PIK3CA inhibitor in patients with an overgrowth syndrome as a result of an early embryonic somatic mosaic *PIK3CA* mutation (Venot et al. 2018), indicates that it would be possible to intervene pharmacologically in somatic evolution in the normal colon. The *ERBB2* and *ERBB3* mutations that we found may also represent druggable targets.

Caution should be exercised, however, as not all mutations that confer a selective advantage in normal mucosa need promote cancer development. Though *STAG2* loss confers a selective advantage in normal colon (Results Chapter 2 section I.5. and R.4.), it is rare in colon cancers. Similarly, in the skin and oesophagus, *NOTCH* family mutations are, if anything, more frequent in normal epithelia than in the cancers that derive from them (Martincorena et al. 2015, Inigo Martincorena personal communication). It seems, therefore, that some mutations provide a selective advantage without conferring the hallmarks of cancer (Hanahan and Weinberg 2000). In the colon, where the progression to malignancy is well described, a distinction could be drawn between potentially innocuous mutations and those that actually cause cancer by repeating the experiment along the adenoma carcinoma sequence.

Although the frequency of driver mutations *per colon* is high, the frequency of driver mutations *per crypt* is low, certainly in comparison to skin (30% of cells, (Martincorena et al. 2015)), oesophagus (40% of cells, Inigo Martincorena personal communication). This is despite the fact that the incidence of colon cancer is similar to that of oesophageal (5% and 1% lifetime risk respectively). We speculate that the difference in driver frequencies between these tissues is due in part to the modular structure of glandular epithelium, which determines the strength of selection within the crypt and prevents clonal expansions beyond it (General Introduction). Indeed, it seems that some colorectal cancer driver mutations promote crypt fission (Results Chapter 2 section I.5.b.), allowing mutant stem cells to escape the constraint of the gland. Two of the 14

driver mutations that we observed were associated with crypt fission events. Recently, somatic driver mutations were found at a high frequency in endometrial glands (Suda et al. 2018). *PIK3CA* mutations were reported in 20% of glands, although not all of these need be driver mutations. Endometrium may represent an unusual glandular tissue in that it is shed and regenerated on a monthly basis during reproductive life. This might provide opportunities for glands harbouring mutant stem cells to fission and colonise the epithelium more effectively than in the colon. The larger size of endometrial glands, if coupled to a larger number of stem cells per gland, may also provide greater opportunities for stem cells that have acquired driver mutations to colonise the gland. Further work will be required to establish the frequency of driver mutations in other glandular tissues and their basis in terms of the rate of acquisition of somatic mutations, their selective advantage, and the constraints on clonal expansion.

1.b.ii. Mutational signatures

We observe a surprising diversity of mutational processes active across the normal colons of different people. Many of these processes had previously been reported, but the size of our cohort and the low complexity of normal genomes allowed us to investigate how their rates vary across and within different people and reveal differences between sites in the colon. Intriguingly, rates of common mutational processes were higher on the right side of the colon than on the left, which is the opposite trend to cancer incidence. We also confirm the finding that mutation rates are similar between the colon and the ileum (Blokzijl et al. 2016), despite the former developing cancers more frequently. Part of this discrepancy may be due to the number of stem cells per crypt and differences in selection pressures.

We found signatures that have not been detected in the PCAWG analysis of tens of thousands of cancers (Alexandrov et al. 2018). These novel signatures may not have been described in cancers simply because they are specific to the colon and fewer colorectal cancer genomes have been sequenced than we have sequenced normal crypts. Alternatively, these processes may be present in cancer genomes, but obscured by the panoply of additional mutational processes. Some novel signatures have an explanation, such as SBSN1 which is found in a patient who had been exposed to a cocktail of chemotherapies. It is sobering that chemotherapy could

cause so many mutations all around the colon. Nevertheless, during this patient's colonoscopy, which occurred 17 years after his first chemotherapy, only one adenocarcinoma and one adenoma were found, which is not a particularly rare finding in a 66 year-old man. Furthermore, although the increase in mutations is likely to have affected many other tissues too, this individual survived for 19 years after his first chemotherapy. Some signatures were of unknown cause, such as SBSN2 and SBSN3. These seem to be active early in life, and their features suggest that they are caused by extrinsic mutational processes. They warrant further investigation as they may represent preventable exposures that account for thousands of somatic mutations in a high proportion of people.

A comparison of the mutational processes in tumours and normal tissues reveals that the process of transformation is associated with both an increase in the rate of normal mutational processes and the acquisition of additional ones, even in non-hypermutated cancers. Ostensibly, they allow an accelerated evolution of the cancer and contribute to the subclonal diversity within the tumour that allows some cells to survive chemotherapy. The origin of many of these aberrant mutational processes is unknown; one can hope that the identification of their aetiologies might indicate ways to intervene in tumour evolution.

2. A comparison of the somatic evolution of blood and colon

A brief comparison of blood and colon highlights the diversity of the forces that govern somatic evolution across the human body.

Stem cell numbers appear to be quite different between the two tissues. Assuming seven functional stem cells per crypt and 15 million crypts per colon, there are approximately 100 million stem cells per human colon. This is 1,000 times larger than our estimate of the number of active blood stem cells. This is a striking difference given that the number of mature blood cells produced per day is larger than the number of colonocytes. Presumably, the optimal number of stem cells per tissue has been determined over the course of species evolution in part by balancing the requirement to produce large numbers of differentiated cells against the risk of malignancy. Why, then, should there be so many more colorectal stem cells than blood stem cells? We speculate that

the colon can have far more stem cells than blood without a substantially increased cancer risk[5] because of the organisation of colonic stem cells into crypts, which reduces the effect of mutations that confer a selective advantage (General introduction, section 3.c.). We might predict from this that most mutations that are known to cause colorectal cancers confer a stronger selective advantage than those that are known to cause blood cancers, since even a weak driver mutation can, over time, sweep through the relatively unstructured blood stem cell pool. The presence of common and indolent blood clones that rarely (or, at least, slowly) progress to malignancy supports this view. Recent evidence indicates that clonal haematopoiesis is likely to be responsible for age-related morbidities beyond cancer (General Introduction, section 4). The very liquidity of blood and the recirculation of blood stem cells, which must limit architectural controls on clonal competition, can perhaps be considered to be partly responsible for these diseases.

Mutation burdens and processes are different across the two tissues, as a colonic stem cell has approximately three times as many mutations as a blood stem cell of the same age. Although only 140 blood stem cells from one individual were sequenced, it seems that there is more variability in the mutational processes and mutation burden across the colon than in blood. Some of the sporadic mutational processes in the colon are likely to be caused by mutagens, a likely corollary of the exposure of intestinal stem cells to the luminal contents of microflora and dietary carcinogens. Blood stem cells, in contrast, are maintained in a more homogeneous environment.

Both variation and selection are therefore different in blood and colon, and are likely to vary across all tissues. Indeed, cancers from different organs have a different incidence, repertoire of mutational signatures, and predilection for particular driver mutations. Every tissue will have to be studied in order to build up a holistic view of somatic evolution across the human body.

3. Perspective on managing the diseases of somatic evolution

Somatic evolution is inevitable in a multicellular organism with a long lifespan and imperfect DNA replication and repair. Nonetheless, an understanding of the forces of variation

[5] Colorectal cancer incidence is approximately four times that of all leukaemias combined (Cancer Research UK), but this excess is probably largely a result of recent dietary changes (chapter 2, section I.3.), and the disease mostly affects the elderly. Throughout the course of much of human evolution it seems probable that colorectal cancers were not a significantly more frequent cause of premature death than leukaemias.

and selection in each tissue provides clues as to how we can manage the diseases that stem from it.

First, the description of mutational signatures across tissues and the identification of their causes may provide opportunities to decrease the mutation rate. Extrinsic mutagens may be avoided, and protective drugs may be discovered. For example, low-dose aspirin decreases the risk of bowel cancer (Brenner et al. 2014), and it has been suggested that this effect of non-steroidals on cancer incidence may be through reducing the mutation rate (Kostadinov et al. 2013). Studying the mutational signatures of normal colonic crypts of those who have and have not taken aspirin would reveal whether or not this was the case. Some mutational processes, such as polymerase slippage in normal DNA replication, will always, however, be unavoidable. Measuring a person's normal mutation rate may allow patient stratification: those with higher mutation rates might receive more regular screening tests than those with lower ones.

Second, clonal dynamics might be monitored. It has recently been shown that it is possible to predict which patients with clonal haematopoiesis are at high risk of progression to acute myeloid leukaemia (Abelson et al. 2018). By developing a good understanding of healthy clonal dynamics, it will be possible to detect a departure from normality and quantify the risk of progression to malignancy or another disease. This would be easiest in blood, but one can imagine that a similar approach might be possible for other tissues, such as through circulating DNA (Wan et al. 2017) or other methods of randomly sampling epithelium (Lao-Sirieix and Fitzgerald 2012). This could inform risk stratification and treatment.

Finally, as discussed above, judicious 'clonal purging' of mutations associated with a high risk of progression might remove clones early, before they can grow large enough that they contain sufficient genetic diversity to be resistant to drugs.

4. Conclusion

> It is interesting to contemplate an entangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent on each other in so complex a manner have all been produced by laws acting around us. These laws, taken in the largest sense, being Growth with Reproduction; Inheritance which is almost implied by reproduction; Variability from the indirect and direct action of the external conditions of life, and from use and disuse; a Ratio of Increase so high as to lead to a Struggle for Life, and as a consequence to Natural Selection, entailing Divergence of Character and the Extinction of less-improved forms.
>
> (Charles Darwin, *On the Origin of Species*)

A slide of cancer histology presents a comparable richness and diversity to Darwin's entangled bank, but on a different scale: cells acquire mutations, creating diversity, and compete with one other. Cancers emerge as a result of this struggle for existence. Darwinian theory can be applied to somatic tissues: this requires the integration of the fields of cancer genomics, evolutionary biology, population genetics, and cell and developmental biology. It is to be hoped that such an approach offers opportunities to extend our understanding of health and disease.