Chapter 6

Discussion

Collectively, this work has shed light on the landscape of clonal haematopoiesis in three distinct settings: in the years preceding a diagnosis of either AML (Chapter 3) or a lymphoid malignancy (Chapter 4) and following intensive cytotoxic therapy for a childhood cancer (Chapter 5). In this discussion I will highlight common themes emerging from the results of the preceding chapters and provide an overview of further questions and areas for methods development.

1. Overview of emerging concepts

1.1 Key points:

- CH in individuals who years later develop a haematological malignancy is characterised by a different genetic landscape compared to CH in the general population, not merely by a higher mutation burden.
- Predictive models incorporating genetic and demographic variables identify most individuals with CH at high risk of progression to AML. Mutations in *TP53* and *U2AF1* are associated with a higher risk of AML progression than somatic events in the most frequently mutated CH genes.
- Clones harbouring *DNMT3A* or *TET2* mutations confer similar risks of progressing to AML versus a lymphoid neoplasm.
- Readily available clinical information improves CH risk-stratification. Higher RDW helps discriminate indolent CH from pre-AML. Lower cholesterol is reaffirmed as a likely biomarker of both lymphoid and myeloid malignancy risk.

- This work adds to the preliminary evidence suggesting that evolution of childhood t-MN may frequently be detectable early in treatment for the primary malignancy. For childhood cancer patients, the relative rarity of CH, heavy burden of t-MN and survival advantage of prompt HSCT highlight this patient group as a top priority for further study of the clinical utility of CH screening.

1.2 The mutational spectrum of premalignant CH

The prevalence, number of driver mutations and clone sizes all tended, unsurprisingly (Genovese et al., 2014; Jaiswal et al., 2014), to be markedly higher among individuals who later developed a blood cancer. However, there were also significant differences in the genetic landscape of CH in these different contexts. Within the pre-AML cohort, the spectrum of CH drivers overlapped with that seen in the general population, but was enriched for spliceosome mutations in younger individuals. By contrast, the mutational landscape preceding lymphoid cancer diagnosis was remarkably diverse, with a long 'tail' of driver mutations in genes seldom if ever implicated in CH in the general population but highly associated with lymphoid neoplasms.

1.3 CH as a biomarker of blood cancer risk irrespective of phylogenetic relationship with future malignancy

Several findings reported here add to the growing evidence that CH is a risk factor for haematological malignancy even when not related to the future neoplastic clone. Models estimating future AML or LN risk demonstrated that the number, clone size and specific genes mutated all carried predictive value. Although the power to discern gene-level risk for the pre-LN cohort was limited by the large number of infrequently mutated genes, a key finding from the LN predictive models was that *DNMT3A* and *TET2* mutations were robustly predictive of future LN risk, and that hazard ratios were equivalent to those observed for AML progression. Given that *DNMT3A* and *TET2* are much less frequently implicated as drivers in lymphoid compared to myeloid cancers, this finding suggests that CH can be a biomarker of blood cancer risk independent of the relationship between the CH clone and future malignancy. This is in keeping with observations that CH is a biomarker for t-MN risk in adult cancer patients, despite that the antecedent CH and future t-MN are often phylogenetically unrelated (Gibson et al., 2017; Gillis et al., 2017; Takahashi et al., 2017). Equally, a recent

study suggests that *de novo* AML frequently arises from one out of many co-existing independent CH clones detectable pre-treatment (Wong et al., 2015a).

The time-course experiment data in both Chapters 3 and 4 provide further insight into the relationship between CH and future malignancy risk. Variable clonal growth trajectories were observed in premalignant cases and controls. Many clones regressed over time, including some harbouring high VAF canonical hotspot mutations, e.g., *DNMT3A* p.R882H. Hence the cell-intrinsic self-renewal advantage conferred by such mutations (Brunetti et al., 2017) does not necessarily induce inexorable clonal expansion over time, despite that they collectively confer higher leukaemia risk. Among the few pre-AML for whom diagnostic or peri-diagnostic specimens were available, most clones, though not all, expanded and appeared likely to contribute to the AML. The pre-LN serial sampling data offers even more compelling evidence that mutations unrelated to the future cancer are *bona fide* biomarkers of malignant transformation risk. Comparing pre-LN cases to controls revealed that *DNMT3A* mutations were present at significantly higher VAF in pre-LN cases. Nevertheless, even large (VAF>5%) *DNMT3A*-mutated clones often declined in size leading up to cancer diagnosis, frequently coinciding with the appearance of new, LN-associated drivers. Hence it is likely that most of the predictive power of *DNMT3A* mutations does not stem from their direct contribution to LN evolution.

Collectively, these experiments, in conjunction with the aforementioned t-MN studies, strongly suggests that CH unrelated to the future malignant clones is nevertheless a biomarker of malignant transformation risk. There are several non-mutually exclusive potential explanations for this observation. It is possible that the HSC mutation rate, and hence the likelihood of serial acquisition of drivers in any given clone, tends to be higher among individuals who develop a cancer, and the presence of multiple detectable clones is a surrogate marker of the higher mutation rate. However, the mutation burden and signatures in AML compared to normal HSCs of the same age argue against this as a universally active mechanism (Alexandrov et al., 2013; Welch et al., 2012). Alternatively, CH may be a surrogate marker of the presence/intensity of selection pressures that influence the fitness advantage conferred by particular driver mutations. Studies of CH in the context of aplastic anaemia (Yoshizato et al., 2015) and cytotoxic therapies (Gibson et al., 2017; Hsu et al., 2018; Kahn et al., 2018) provide strong evidence that extrinsic selective pressures can dramatically increase the prevalence of CH, shape the genetic landscape, and increase the malignant

transformation risk. By extension, it is conceivable that the same may be true for diverse subtler extrinsic selection pressures, e.g., arising from variable ageing processes, environmental exposures, or inter-individual genetic variation. For example, physiological ageing processes occur at different rates in different individuals (Andersen et al., 2012; Finkel et al., 2007; Lopez-Otin et al., 2013). It is conceivable that age-associated increases in endogenous genotoxic stress (Rossi et al., 2007) and declines in HSC self-renewal capacity (Flach et al., 2014; Geiger et al., 2013) occur earlier or more severely in some individuals. This in turn could confer selective advantage on many mutated HSCs, increasing the number of detectable clones in younger age groups and the probability of any one of the clones acquiring additional oncogenic hits. These questions warrant further investigation, as discussed below.

2. Further questions and methodological challenges

2.1 To what extent is mutation acquisition a rate-limiting step in CH evolution?

Understanding the relative importance of mutation acquisition and extrinsic selective pressures in CH pathogenesis is an important gap in knowledge, not least for informing any future intervention strategies. For certain genes, e.g., *TP53*, very sensitive sequencing assays have demonstrated that driver mutations are common in older individuals at extremely low VAF and tend to be stable over time in the absence of any environmental selective pressures which increase mutated HSC fitness advantage (Wong et al., 2015b). By contrast, the exponential increase in the prevalence of CH harbouring spliceosome gene mutations observed in individuals aged >70 years (McKerrell et al., 2015) is poorly understood. It is possible that this phenomenon reflects ageing-associated changes in the haematopoietic niche (McKerrell and Vassiliou, 2015). For instance, spliceosome mutations may generate neoantigens that elicit a stronger immune response in younger individuals (McKerrell and Vassiliou, 2015). However, this speculation has yet to be supported by experiments demonstrating low-level persistence of rare HSCs carrying spliceosome mutations in younger individuals. The sensitivity of error-corrected sequencing assays has been a major obstacle to this type of experiment (Kennedy et al., 2014; Schmitt et al., 2012). In particular, sensitivity is hindered by target pulldown efficiency and stochastic molecular sampling, issues which can

be partially circumvented by multiple target enrichment steps and using a limited number of cells as starting material (Schmitt et al., 2015). However, novel methods of increasing sensitive, accurate detection of specific mutations (Nachmanson et al., 2018; Newman et al., 2016) could be applied to the detection of canonical spliceosome gene hotspot driver variants in younger cohorts.

2.2 Haematopoiesis and ageing in health and disease

To what extent do the number, mutation rate, and clonal dynamics of haematopoietic stem and progenitor cells vary between individuals? These are pertinent questions for understanding the increased CH burden seen in individuals who later develop a haematological malignancy, as discussed above. Two recent studies have used somatic mutations to study clonal dynamics in native human haematopoiesis (Lee-Six et al., 2018; Osorio et al., 2018). Based on this work, it is likely that there are circa 50,000-200,000 HSCs contributing to haematopoiesis, dividing roughly every 2-20 months with around 14 mutations introduced per cell division (Lee-Six et al., 2018; Osorio et al., 2018). Applying similar approaches, potentially with superimposed phenotypic information, to many individuals across the age range and in disease/cancer-predisposition states will likely give valuable insights into haematopoietic ageing and CH pathogenesis.

2.3 Refining CH detection methods

The definition and terminology used to describe CH has evolved rapidly and sometimes included VAF cut-offs (Bejar, 2017). However, the latter have been decided based on technical limitations rather than mature understanding of what constitutes clinically significant CH (Bejar, 2017; Steensma et al., 2015). In future, cheaper sequencing should enable comprehensive assays to detect subclonal cancer-associated structural events in addition to point mutations. Novel sequencing methods for detecting rare somatic mutations, notably bottleneck sequencing (BotSeq), may enable broader screens for genes under positive selection in CH (Hoang et al., 2016). Briefly, BotSeq combines molecular barcoding with a subsequent dilution step, permitting highly accurate detection of rare mutations across the entire genome without the need to achieve prohibitively expensive sequencing depth. It is conceivable, though currently entirely speculative, that transcriptional or methylation-

based signal may also be amenable to identifying and characterising CH and may warrant exploration in tandem with future studies of genomically-defined CH.

2.4 Prospective longitudinal studies of CH and potential intervention strategies

An important next step will be to establish large prospective longitudinal studies enabling validation and refinement of combined genomic-clinical CH risk prediction models. Ideally such studies will examine multiple clinically relevant sequelae of CH and permit identification of high-risk groups that might benefit from intervention. The nature of potential interventions is speculative at present. An increasing arsenal of targeted therapies active against recurrent cancer-associated CH mutations, including those in splicing genes (Lee et al., 2016), *JAK2* (Van den Neste et al., 2018; Vannucchi and Harrison, 2016) and *IDH1/2* (Döhner et al., 2015), may warrant investigation in high-risk CH. Moreover, two recent studies suggest that a much less costly option, ascorbic acid (vitamin C), helps restore *TET2* function in HSCs and stall leukaemia progression (Agathocleous et al., 2017; Cimmino et al., 2017). Lastly, this work further corroborates a long-recognised connection between hypocholesterolaemia and haematological malignancies. Lower HDL and LDL were both risk factors for AML in the clinical risk prediction model discussed in Chapter 3. Lower HDL was associated with a higher risk of developing a lymphoid neoplasm (Chapter 4). The latter result corroborates previous work identifying low HDL as a biomarker of future lymphoma risk years prior to diagnosis (Matsuo et al., 2017). Hypocholesterolaemia is common among blood and solid cancer patients and is inversely correlated with cancer cell LDL-/HLD-receptor activity (Ho et al., 1978; Vitols et al., 1985; Vitols et al., 1984; Vitols et al., 1992). A mendelian randomisation study by Benn et al. found that the correlation between low LDL and cancer was absent in individuals with genetic predisposition to hypocholesterolaemia, suggesting a causal link (Benn et al., 2011), though this remains contentious (Pirro et al., 2018). Pharmacologic agents targeting HDL uptake receptors and other targets involved in cholesterol metabolism have shown early evidence of therapeutic potential in several haematological malignancies (Crusz and Balkwill, 2015; McMahon et al., 2017; Pandyra et al., 2014). Interestingly, statin treatment is associated with a significant relative risk reduction for several solid tumours as well as cardiovascular disease (Demierre et al., 2005; Poynter et al., 2005). The molecular mechanisms underpinning these observations are poorly

understood and may involve pleiotropic effects on multiple processes relevant to oncogenesis, including angiogenesis and inflammation (Crusz and Balkwill, 2015; Demierre et al., 2005; Hanahan and Weinberg, 2011). Collectively, these observations suggest that existing agents targeting cholesterol metabolism (Pandyra et al., 2014) warrant investigation as potential strategies for mitigating cardiovascular disease and cancer risks associated with CH.

In summary, the degree to which clones at high risk of malignant transformation - in blood and other tissues - can be reliably distinguished from their indolent counterparts is an important biological question with compelling clinical ramifications. This dissertation has explored the ability of genetic and clinical factors to identify individuals at high risk of AML and other haematological malignancies. Understanding the selective pressures and cell-intrinsic mechanisms governing clonal fate is the next important step in developing strategies to predict and prevent progression to overt malignancy.