Chapter 6

6. Overall summary and future work

6.1. Overall summary

Healthy humans have approximately $3x10^9$ B-cells in the peripheral blood and this population encompasses the repertoire of distinct B-cells expressing different B-cell receptors (BCRs) necessary to bind diverse antigens and produce an effective humoral immune response. B-cells are dynamic populations of immune cells that evolve over time. The aim of this thesis was to investigate B-cell population diversity and dynamics in health and disease using the sequence diversity and population structure of the B-cell BCR repertoire. This required the development of novel, robust, sensitive and reproducible high-throughput B-cell receptor sequencing methods.

This thesis demonstrates that human BCR repertoire diversity can be interpreted through full V-D-J genotype diversity using networks. BCR sequences can be organised into networks based on sequence diversity, with differences in network connectivity providing clinically useful B-cell repertoire structure information. An important result of this framework is the ability to determine how B-cell repertoire structures differ between health and disease. Samples from clonal B-cell populations, such as from CLL, B-ALL and other clonal blood disorders, can readily be distinguished from healthy samples by an increase in BCR clonality and decrease in BCR diversity. For example, different features of the B-cell repertoire can be used to distinguish between patients with B-ALL patients with high leukaemic cell loads, B-ALL patients with low or undetectable levels of leukaemic cell loads and healthy individuals, and can be used as a sample classifier. Interestingly, the B-ALL samples remain largely distinct from healthy B-cell repertoires even after years of undetectable disease, suggesting a long-term B-cell repertoire impact of either the disease or, more likely, the anti-leukaemic therapy. Similarly, even though CLL therapy by Chlorambucil results in significant reduction in peripheral blood B-cell clonality, CLL patient samples remain distinct from equivalent samples from healthy individuals. Long-term effects of B-cell depletion therapy have been observed in previous studies, where SHM rate is reduced in rheumatoid arthritis patients even 6 years after

rituximab treatment (Dorner et al., 2010, Muhammad et al., 2009, Stolz and Schuler, 2009). There was variation between the diversity measures of the BCR repertoires between the healthy individuals, thus indicating a range representing healthy B-cell clonality and diversity. A larger-scaled assessment of primary immune responses compared to early stage leukaemias could provide clinically important diagnostic or prognostic information to patients.

The utility of this method can extended to clinical monitoring of disease, MRD and relapse. Unparalleled sensitivity of BCR sequencing for detecting MRD was demonstrated here compared to conventional clinical methods, where detection of leukaemic BCR RNA is greater than 1 in 10^7 RNA molecules, which is increased 13.57-fold by using only a single IgHV-specific primer corresponding to the specific BCR of interest. In practice, when there is prior knowledge of a BCR of interest, such as in leukaemia, the limit of detection is dependent on the number of cells sampled and the sequencing depth. However, the limit of *de novo* detection of malignant clonality is at least 1 in 100 dilution of CLL or ALL cells into healthy blood. When the clone of interest is small (i.e. less than 1 in 100 cells), diversity measures alone cannot directly be used to distinguish from healthy samples.. Therefore, detection of MRD is achievable as long as the clinical sample contains malignant cells, sequencing is performed at an adequate depth and the malignant clonal sequence is known *a priori*. In addition to increased sensitivity, the ability to detect multiple subclones in leukaemias by BCR sequencing highlights its advantages over qPCR methods, thought to occur between 1.38-2.70% in CLL (Plevova et al., 2014, Kern et al., 2014), 19.35-27% in ALL (Beishuizen et al., 1991, Kitchingman et al., 1986), and 10% in lymphomas (Sklar et al., 1984). This is particularly relevant in diseases where B-cell clones can undergo secondary rearrangements or in cases of two independent B-cell malignancies (Boyd et al., 2009, Bashford-Rogers et al., 2013). Enlarged clusters representing BCRs with different IgHV-D-J gene combinations may be due to either the expansion of two distinct malignant B-cell transformations, or separate antigen-stimulated B-cell clonal expansion unrelated to the malignancy. The presence of more than one BCR clonal expansion has unknown clinical implications in CLL and B-ALL, but with the risk of secondary malignancies in these patients, monitoring these bi-clonal B-cell disorders is of great clinical importance.

The utility of these methods could extend further to autoimmunity, immunodeficiency, response to infection and vaccination, thus potentially improving the understanding and clinical practices of a vast realm of diseases.

6.2. Future work

Using this thesis as a framework for immune repertoire analysis, it is apparent that there are many biological and clinical applications to the methods described here. The utility of these methods extend beyond malignancy to autoimmunity, immunodeficiency, response to infection and vaccination. However, this section will cover directions of future work directly derived from the findings in this thesis.

Firstly the full human allelic variation in the heavy and light Ig V, (D) and J genes is still unknown, where population differences in gene sequences may result in differential susceptibility of diseases. Biases in immunoglobulin gene recombination patterns have been shown to affect influenza susceptibility, where a polymorphism in the recombination signal sequence of IgKV locus in the Navajo population is associated with increased influenza susceptibility. This polymorphism reduces recombination of a commonly used IgKV gene by about 4.5-fold (Feeney et al., 1996). Therefore, future work should include determining the association between allelic variation in the heavy and light IgV, (D) and J genes and corresponding promoter regions. This could potentially be achieved through the analysis of the immunoglobulin loci of large-scale datasets, such as exome or whole genome sequencing of large numbers of individuals from the UK10K and 1000 Genome datasets and by a large scale analysis of IgH and IgL productive rearrangement frequencies in the peripheral blood of diverse populations of people.

Further experiments should include determining the B-cell repertoire differences between different anatomical locations within an individual, such as between lymph nodes and peripheral blood. Model systems, such as mice, can be used to investigate the development and spatial structure of immune responses during vaccination or infectious challenge. However, the availability of some anatomical regions from humans is limited, for example, bone marrow biopsies are typically only taken in individuals with blood abnormalities, such as anaemia, leukopenia, thrombocytopenia and leukaemia. However, even these samples, when paired with peripheral blood, could give valuable information on the spatial arrangement of specific B-cell populations. For example, this thesis has shown important potential uses of BCR sequencing in monitoring disease during therapy (Chapter 4) and minimal residual disease detection (in Chapter 5). However, it is of great clinical benefit to defining optimal anatomical locations for detecting minimal residual B-cell

and, potentially, T-cell populations during leukaemia therapy. In particular, it would be clinically useful to determine if single or multiple peripheral blood draws are more effective at sampling malignant B-cells for detection of MRD compared to bone marrow biopsies in B-ALL. Additionally, the mode of relapse remains a question in many leukaemias, such as where B-ALL MRD cells reside during therapy and what circumstances lead to relapse, particularly to certain anatomical sites such as CSF. Therefore, multiple sampling of different anatomical sites may give information on which regions are less readily accessible to therapy and potential reservoirs of cancer cells.

Multiple B-cell clonal expansions were observed in some of the CLL and B-ALL patients in this thesis, which opens the question of whether these clones are distinct malignant B-cell transformations, or separate antigen-stimulated B-cell clonal expansion unrelated to the malignancy. This may be answered by two different approaches. Firstly, BCR sequencing of longitudinal samples from these patients may be used to determine whether there is reduction in the sizes of any of the clones in the absence of therapy, suggesting antigen-driven clonal expansion and subsequent reduction. Secondly, single-cell whole-genome or exome sequencing may be performed on cells from the expanded clones to determine whether there are shared genomic features or aberrations that may be indicative of single or multiple malignant expansions. An alternative to this would be single-cell transcriptomic analysis, which would give information on the differences between cells from different clonal expansions on an RNA-expression level, potentially shedding light on the similar or different processes that have led to their clonal growth.

Previous studies have shown that non-B- and non-T-cell malignancies are often marked by profound defects in B-cell and T-cell function, such as in melanoma and solid tumors in mice (Baitsch et al., 2011, Ahmadzadeh et al., 2009, Sakuishi et al., 2010)). B- and T-cell exhaustion prevents optimal control of infection and malignancy, and therefore understanding the structure and dynamics of the normal B- and T-cell repertoires in patients with different malignancies may help identify common underlying principles of immune-dysfunction, to assess potential for diagnostic or prognostic marker for disease development and to identify therapeutic opportunities. This is achievable by cell sorting of activated memory B-cells or plasma cells, paired heavy and light chain sequencing and screening for reactivity against the malignant cell populations of interest. Next, the question of which B-cell

subsets produce these anti-malignant BCRs could be addressed by BCR high-throughput sequencing of flow sorted B-cell populations. This may determine whether anti-malignant B-cells are indeed prone to immunological exhaustion, and the dynamics of such a process may be determined using longitudinal samples. Importantly, an exhaustive phenotype of B-cells that specifically bind malignant cells may be a useful biomarker for relapse risk.