Chapter 1

Introduction

1.1 CLL

1.1.1 Disease characteristics

Chronic lymphocytic leukaemia (CLL) is a malignancy of mature B cells, characterised by the progressive accumulation of malignant lymphocytes in the blood, bone marrow and lymph nodes (Thomas J. Kipps et al. 2017). The malignant cells can be distinguished by expression of CD5, CD19, and CD23, and by lower levels of membrane IgM, IgD, and CD79B (Matutes et al. 1994; Moreau et al. 1997; Chiorazzi, Rai, and Ferrarini 2005) (Figure 1.1).

The disease pathogenesis is driven by multiple factors, including molecular features, signalling via the B cell receptor (BCR) and interactions with non-neoplastic cells within the lymphoid tissues (known as the tumour microenvironment) (Thomas J. Kipps et al. 2017). Whilst the majority of CLL cells are in a resting state (B. T. Messmer et al. 2005; Defoiche et al. 2008), evidence suggests that CLL cells migrate towards lymph nodes where they form proliferation centres (Granziero et al. 2001; B. T. Messmer et al. 2005), similar to germinal centres in healthy lymph nodes, which can show a daily birth rate of up to 3.3% of the tumour (Herndon et al. 2017).

The disease is the most common leukaemia in the West, accounting for 37% of leukaemia cases and ~19,000 newly detected cancers in the US in 2016 (Thomas J. Kipps et al. 2017; Dubois et al. 2020). The risk of developing CLL is twice as high for men than women, and more likely with increasing age (Siegel et al. 2012; Nabhan et al. 2014; Li et al. 2015; Pulte et al. 2015). Chemotherapy (Robak 2005; Chang and Kahl 2012;

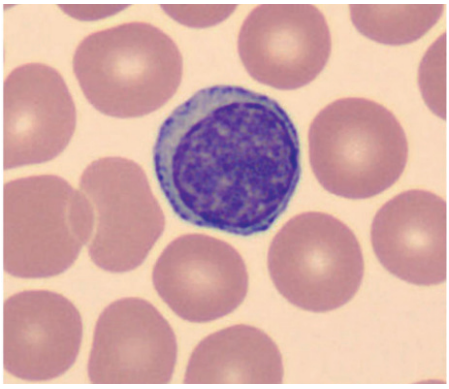

Figure 1.1: Wright–Giemsa-stained blood smear of a CLL malignant B cell, depicting typical morphology. *Figure from Thomas J. Kipps et al. (2017)* .

Lukenbill and Kalaycio 2013) and chemoimmunotherapy (M. Hallek et al. 2010; Goede et al. 2014; Hillmen et al. 2015) has been the mainstay of therapy for many years. More recently, the central role of BCR signalling in disease pathogenesis has been appreciated and new drugs targeting this pathway have improved patient outcomes. The overall 5-years relative survival for CLL is 84% (Miller et al. 2019).

In spite of this, CLL is widely considered to be incurable (Bosch and Dalla-Favera 2019), and treatment regimens remain harsh. Moreover, the disease is characterised by its clinical heterogeneity (Miller et al. 2019) and there is significant variation in survival amongst patients. Some patients harbour indolent disease, and can be followed by a watch-and-wait approach, sometimes going decades without requiring treatment. Others may need immediate treatment and survival can be short (Miller et al. 2019). There is a clinical need to understand the underlying biology of the disease, with a view to understanding the causes of this clinical heterogeneity and to improving patient outcomes and experiences.

1.1.2 Cell of Origin

Identifying the cell of origin of a cancer can help explain disease pathogenesis, and understand the different subtypes, and their associated prognosis. Various studies have established mature CD5+ B cells as the cell of origin in CLL (**Campo2017?**; Bosch and Dalla-Favera 2019).

In healthy tissues, these mature B cells are derived from haematopoietic stem cells in the bone marrow (Kondo 2010; Fischer et al. 2020). These stem cells develop in multi-

ple stages, where each stage is defined by rearrangements within the immunoglobulin heavy chain and light chain loci (Pelanda and Torres 2012) that encode components of the BCR. The aim of this process is to generate a large BCR repertoire, capable of recognising a range of foreign antigens.

Developing B cells in the bone marrow undergo positive and negative selection, ensuring that each BCR binds effectively to foreign antigen, whilst eliminating those that bind strongly to self-antigen (Lebien and Tedder 2008; Mårtensson et al. 2010). These immature B cells then migrate to the lymph nodes and spleen where they differentiate into mature cells that are considered to be antigen "naive"(Chung, Silverman, and Monroe 2003).

B cells are later activated when they encounter their respective antigen. When activated, the B cells form germinal centres, which are specialised microenvironments within the lymph node that facilitate extensive proliferation. Here they undergo a process called affinity maturation, in which the loci encoding components of the BCR undergo somatic hypermutation to optimise BCR antigen specificity (Shlomchik and Weisel 2012). This process generates short-lived plasmablasts which provide immediate protection, along with plasma cells and memory B cells procuring longer-term immunity (Nutt et al. 2015).

1.1.3 IGHV status

The two major subtypes of CLL can be defined by whether the B cell of origin has undergone this process of somatic hypermutation within a germinal centre. This is reflected in the degree of mutation of the immunoglobulin heavy-chain variable region (IGHV) genes. Evidence suggests that IGHV-mutated (IGHV-M) CLLs (60% of cases) are derived from antigen-experienced B cells. On the other hand, there is continued debate as to whether IGHV-unmutated (IGHV-U) CLLs derivate from naive B cells that have yet to transit through the germinal centre, or germinal-centre, antigen-experienced B cells (Klein et al. 2001).

The two subtypes show distinct clinical and molecular properties. In contrast to IGHV-M CLLs, IGHV-U CLLs have a higher proportion of high-risk genetic lesions, more commonly undergo clonal evolution and consequently have a shorter time to first treatment (TTFT) and less favourable overall survival (OS) (Hamblin et al. 1999; Damle et al. 1999; Landau et al. 2013; Xose S. Puente et al. 2015; Bosch and Dalla-Favera 2019). Their differential antigen experience also affects how they respond to signals from the

microenvironment, including stimulation of the BCR.

1.1.4 The central role of BCR signalling

The BCR pathway is central to the process of selection, development, proliferation and survival of CLL clones (Chiorazzi and Ferrarini 2003; Stevenson and Caligaris-Cappio 2004; Agathangelidis et al. 2012; Jan A. Burger and Chiorazzi 2013; Iacovelli et al. 2015; Dubois et al. 2020). Evidence of the importance of this pathway comes from the observation that BCRs of CLL cells show a biased selection of IGHV and IGLVκ/λ genes. This generates BCRs that are remarkably similar across patients (T. J. Kipps et al. 1989; Fais et al. 1998; Widhopf et al. 2004; B. T. Messmer et al. 2004; K. Stamatopoulos et al. 2007), suggesting that BCR binding to certain antigens may drive selection and proliferation of CLL clones. Other compelling studies have shown that BCR signalling genes are upregulated in CLL cells taken from lymph nodes (**Herishanu2011?**) and that cells from IGHV-U CLL patients (with poorer outcomes), also show activation of BCR-related genes (Rosenwald et al. 2001). The success of BCR inhibitors in the clinic (John C. Byrd et al. 2013), also underlines the critical importance of this pathway in CLL.

Figure 1.2 shows a schematic of the BCR pathway itself. The BCR is a multimeric complex consisting of surface immunoglobulin (Ig), which recognises antigen, plus the Ig-α/Ig-β hetero-dimers, known as CD79A and CD79B. The BCR may bind to external antigens within the microenvironment (Binder et al. 2010) or intra-BCR self-antigens (**Minden?**), which in turn recruits SYK and the Src kinase LYN. These kinases phosphorylate motifs located on the cytoplasmic tails of CD79A and CD79B, initiating a signalling cascade involving a number of proteins and pathways. These include the proteins BTK (Herman et al. 2011) and PI3K (**Longo2006?**) which activate a number of downstream pathways and players including PLCγ 2, calcium signalling, PKC, NFκB signalling, ERK and MAPKs, and nuclear transcription (Jan A. Burger and Chiorazzi 2013; Dubois et al. 2020).

There are two types of BCR signalling in healthy and CLL B cells: ligand-dependent "active" signalling that relies on antigen-binding, and ligand-independent "tonic" signalling (**Lam1997?**; Kraus et al. 2004). Whilst active signalling engages the entire signalling cascade described above, tonic signalling activates only a subset of these. Kraus et al. (2004) showed that tonic signalling is important for prolonged B cell survival whereby PI3K signalling is thought to play a key role in delivering survival signals (**Srinivasan2009?**; **Pula2019?**). Both these modes of BCR signalling are believed to

impact of the survival and growth of the tumour, although the dominant mode remains a matter of debate (Jan A. Burger and Chiorazzi 2013).

The different clinical properties of IGHV-M and IGHV-U cell is believed to be determined by in part by their differential response to BCR stimulation, which is influenced by their cell of origin (Jan A. Burger and Chiorazzi 2013). IGHV-U CLLs have not undergone somatic hypermutation and consequently express low-affinity BCRs that are frequently activated by numerous antigens and auto-antigens in the microenvironment (Borche et al. 1990; Bröker et al. 1988; Sthoeger et al. 1989; Hervé et al. 2005; Myhrinder et al. 2008; Chu et al. 2008; Binder et al. 2010; Krysov et al. 2010; Kostareli et al. 2012; **Hoogeboom2012?**). In contrast, IGHV-M BCRs only recognise highly specific antigens, which either occur infrequently or induce anergy due to high-affinity binding (Chiorazzi and Ferrarini 2003; Stevenson and Caligaris-Cappio 2004; Chiorazzi, Rai, and Ferrarini 2005; Jan A. Burger and Chiorazzi 2013). IGHV-M CLL clones therefore are more stable and expand at a slower rate.

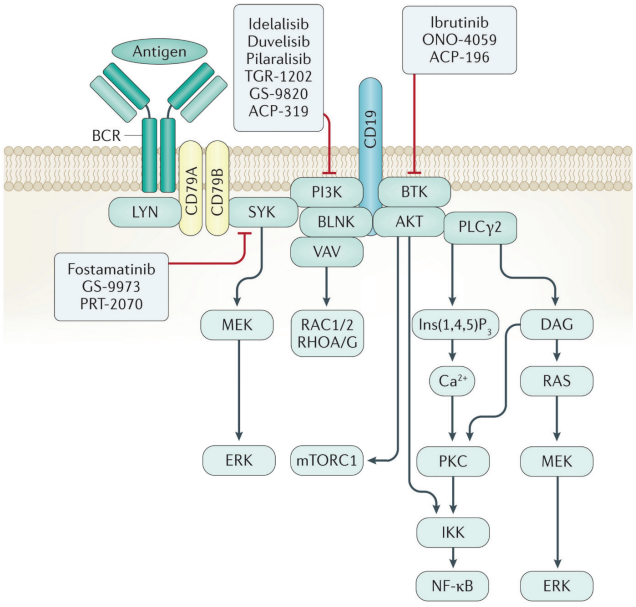

Figure 1.2: Graphical depiction of the BCR pathway. *Figure originally published in Thomas J. Kipps et al. (2017)*

1.1.5 Recurrent genetic features in CLL

The genomic landscape of CLL has been thoroughly characterised in a number of studies, including two seminal papers involving >500 CLL samples (Landau et al. 2015;

Xose S. Puente et al. 2015). Such studies have indicated that CLL shows a lower mutational load than other lymphoid neoplasms (Vogelstein et al. 2013; Alexandrov et al. 2013; Bosch and Dalla-Favera 2019), in which a relatively large number of genes are rarely mutated (Fabbri and Dalla-Favera 2016). A small number of driver genes are mutated in a significant proportion of cases, though a common genetic event which can accounts for most cases of CLL has not been identified (Fabbri and Dalla-Favera 2016). These genetic alterations encompass chromosomal alterations, mutations, alterations in miRNA expression and epigenetic modifications (Thomas J. Kipps et al. 2017).

Somatic Mutations

There are many recurrent somatic mutations in CLL, and these centre on several major pathways and functions that are frequently altered in patients (Figure 1.3). (Xose S. Puente et al. 2015; Fabbri and Dalla-Favera 2016; Thomas J. Kipps et al. 2017). These pathways include Notch signalling, DNA damage response, RNA processing, NFκB signalling, BCR signalling, WNT signalling and chromatin modification (Thomas J. Kipps et al. 2017; Bosch and Dalla-Favera 2019).

Within these pathways, most mutations are rare and only a few occur at a frequency >5% (Bosch and Dalla-Favera 2019). One landmark study indicated the most frequent mutations occur in *NOTCH1* (12.6% of patients), *ATM* (11%), *BIRC3* (8.8%) and *SF3B1* (8.6%), although these frequencies are dependent disease stage and treatment status (Xose S. Puente et al. 2015). The functional role and prognostic importance of a number of these putative driver mutations has been established, and are discussed below.

Notch signalling The Notch pathway activates genes required for proliferation, metabolism and survival, including *MYC*, via the activation of the NOTCH1 transmembrane receptor (Guruharsha, Kankel, and Artavanis-Tsakonas 2012). Mutations in *NOTCH1* are very common in CLL (~4-20%) (Fabbri et al. 2011; Xose S. Puente et al. 2011; Landau et al. 2013, 2015; Xose S. Puente et al. 2015). A number of other recurrent mutations also centre on Notch deregulation, indicating that this pathway is disrupted in many CLL cases (Fabbri and Dalla-Favera 2016). *NOTCH1* mutations occur more often in IGVH-U CLLs and are associated with a less favourable OS (Fabbri et al. 2011). *NOTCH1* mutants also respond less-well to anti-CD20-based therapies, due to decreased surface expression of CD20 (**Fabbri2001?**; Rossi et al. 2012).

DNA Damage Response A number of frequently occurring mutations disrupt the DNA

damage response, the most clinically important of which is *TP53*. *TP53* is known as the "guardian of the genome," owing to its role as a tumour suppressor gene protecting genome integrity by preventing mutation. Despite being a tumour suppressor, *TP53* mutations can have a dominant negative effect on function, such that loss of genomic stability occurs even when a single allele is mutated (Zenz et al. 2008). *TP53* is an important prognostic marker, as dysregulation of its function is associated with resistance to DNA-damaging agents (chemotherapy and radiotherapy) (Zenz et al. 2008; **Dicker2009?**; Rossi et al. 2009), and patients who present with mutations at diagnosis often show shorter TTFT and a less favourable OS (H. Döhner et al. 2000; Zenz et al. 2008; Rossi et al. 2013; Xose S. Puente et al. 2015).

Upstream of *TP53*, the tumour suppressor *ATM* is also frequently mutated in CLL (Bosch and Dalla-Favera 2019). *ATM* activates the DNA damage response upon recognition of DNA double strand breaks (Austen et al. 2005; Shiloh and Ziv 2013). Similarly to *TP53* mutants, *ATM*-disrupted CLLs also show genomic instability and these mutations are associated with shorter TTT and OS, and chemoresistance (Austen et al. 2005; Stankovic and Skowronska 2014).

POT1 is also involved in genomic stability, and is mutated in around 3 – 7% of cases (Bosch and Dalla-Favera 2019; **Ramsay2012?**). *POT1* mutations disrupt telomere protection, leading to a increased in structural aberrations and chromosomal breaks (**Ramsay2012?**). Mutations in *POT1* are associated with IGHV-U CLL and advanced clinical stage (**Ramsay2012?**).

RNA processing 30% of CLL patients harbour mutations disrupting RNA processing and the spliceosome machinery (Xose S. Puente et al. 2011; Wang et al. 2011; Quesada et al. 2012; Xose S. Puente et al. 2015; Fabbri and Dalla-Favera 2016). *SF3B1* mutations are the most frequent to occur, and are found in 10% of cases, usually in IGHV-U patients (Xose S. Puente et al. 2011; Wang et al. 2011; Quesada et al. 2012; Xose S. Puente et al. 2015; Fabbri and Dalla-Favera 2016). The *SF3B1* gene encodes part of the U2 snRNP complex which is involved in RNA splicing (Shin and Manley 2004), though the functional implication of *SF3B1* mutations are yet to be established, and many transcripts show abnormal splicing in*SF3B1* cases (Quesada et al. 2012). The presence of *SF3B1* mutations is associated with a decreased TTFT and unfavourable OS (Bosch and Dalla-Favera 2019).

NFκB signalling A wide range of mutations across various pathways converge on the activation of NFκB (Fabbri et al. 2011; Xose S. Puente et al. 2011; Wang et al. 2011;

Quesada et al. 2012; Xose S. Puente et al. 2015; Landau et al. 2015), including *BIRC3* and *MYD88*. For example, certain mutations in *MYD88* result in increased binding to IRAK1 and higher activation of NFκB (Xose S. Puente et al. 2011). However, the role and prognostic importance of NFκB activation in CLL is still unclear (Bosch and Dalla-Favera 2019).

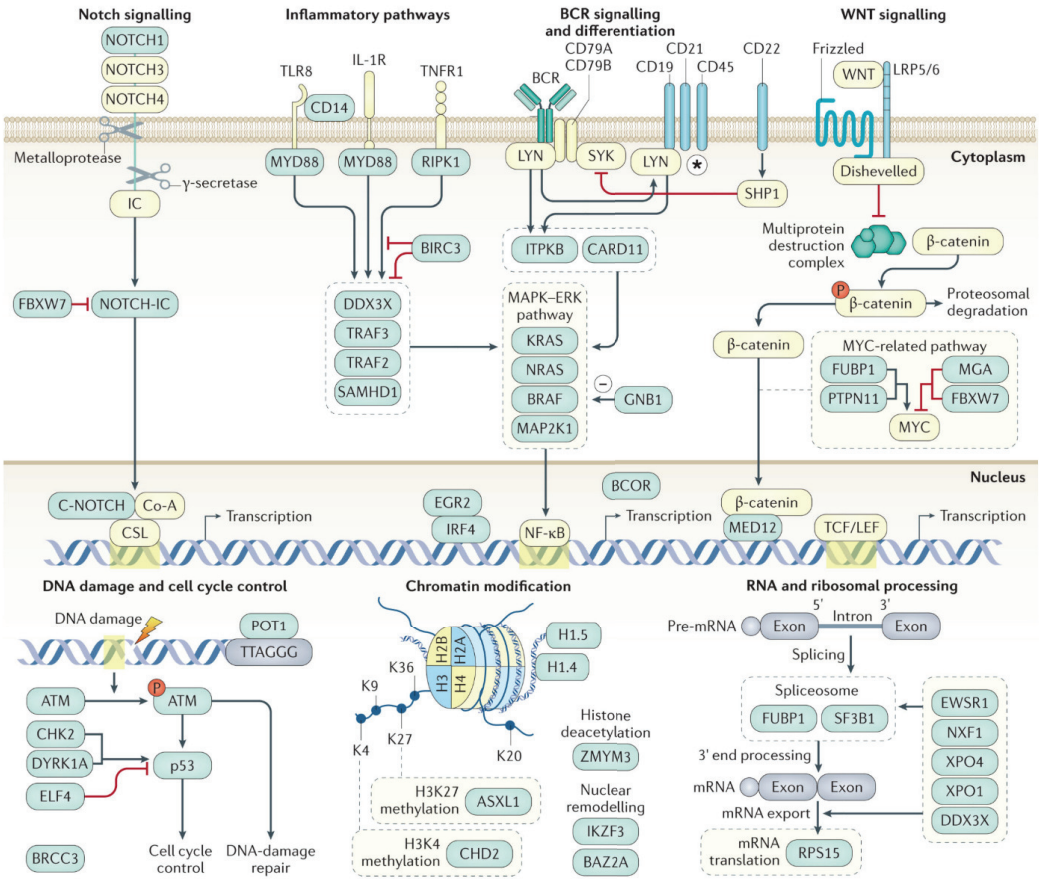

Figure 1.3: Graphical depiction of commonly mutated genes in CLL, grouped into cellular pathways (blue boxes). Minus sign indicates negative regulation. _Figure originally published in (**Kipps2017_?**)

Structural Aberrations

In addition to the aforementioned mutations, a number of common structural aberrations confer similar disruption to normal B cell function. The key structural aberrations and their prognostic value were set out in a landmark paper by H. Döhner et al. (2000), as follows.

del(11q) Deletions of chromosome 11q (del(11q)) are fairly common (~10%) in CLL, and are believed to target the *ATM* gene in the 11q22-23 region (H. Döhner et al. 2000; Austen et al. 2005; Bosch and Dalla-Favera 2019). del(11q) is usually monoallelic, but can also be associated with mutations in the remaining ATM allele (~30% of cases) (Austen et al. 2005). In certain instances, the deleted region does not include *ATM* but rather *BIRC3*, a negative regulator of the NFκB pathway (Rossi et al. 2012). Patients with del(11q) or *ATM* lesions have a shorter TTFT and OS, especially if the lesion is biallelic (Austen et al. 2005; Skowronska et al. 2012; Stankovic and Skowronska 2014; Nadeu et al. 2016).

del(17p) Deletion of chromosomal region 17p13 (del(17p)) is found in 1 – 20% of cases, depending on the stage of the disease and most common in chemo-refractory cases (H. Döhner et al. 2000; Zenz et al. 2008, 2010; Stilgenbauer et al. 2014). The target of this lesion is thought to be *TP53*; the deleted region consistently includes the *TP53* locus (Hartmut Döhner et al. 1995), and around 80% del(17p) cases also have missense mutations in the second *TP53* allele (Zenz et al. 2008; Gonzalez et al. 2011; Trbusek et al. 2011; **Rossi2014?**). Del(17p) CLLs show increased genomic instability (L. Yu et al. 2017), resistance to chemotherapy and, correspondingly, a shorter TTFT and a less favourable OS (H. Döhner et al. 2000; Zenz et al. 2008; Rossi et al. 2013; Xose S. Puente et al. 2015).

del(13q) Deletion in the 13q14 region (del(13q)) is the most common genetic lesion in CLL (~50–60% of patients) (H. Döhner et al. 2000). Experiments to determine the minimal deleted region identified that this invariably contains *DLEU1* and *DLEU2*, two long non-coding RNA genes, and the microRNA gene cluster *MIR15A–MIR16-1* (Kalachikov et al. 1997; Migliazza et al. 2001; Calin et al. 2002; **Hammarsund2004?**; Palamarchuk et al. 2010). *in vitro* studies have demonstrated the role of these genes in regulation of the cell cycle and apoptosis (Cimmino et al. 2005; **Klein2010?**; Bosch and Dalla-Favera 2019). In some CLL cases, del(13q) is the sole genetic abnormality, indicating that this lesion may be involved in early CLL development (H. Döhner et al. 2000; **Klein2010?**; Landau et al. 2015). Moreover, conditional deletion of the equivalent minimal deleted region in mice recapitulated CLL initiation and progression, and these mice developed clonal lymphoproliferations (Migliazza et al. 2001; **Klein2010?**). del(13q) patients have the best prognosis, with prolonged TTFT, and OS compared to patients with other lesions (H. Döhner et al. 2000; Rossi et al. 2013).

The incompletely understood role of trisomy12

Complete duplication of chromosome 12 (trisomy 12) is observed in ~15% of CLL patients at diagnosis (H. Döhner et al. 2000). Despite its recurrence, there is currently no functional explanation for this lesion (Bosch and Dalla-Favera 2019), although a number of features have been observed. Trisomy 12 is more common in IGHV-M patients than IGHV-U (Hamblin et al. 1999), and is thought to confer an abnormal cellular morphology (Bosch and Dalla-Favera 2019). Previous work in our lab has also demonstrated that trisomy 12 CLLs show a specific signalling signature and distinct transcriptomic (Dietrich et al. 2017) and proteomic profiles (Herbst 2020), including differential expression of genes within the BCR, PI3K, AKT, and mTOR signaling and chemokine signaling pathways. Moreover, trisomy 12 CLLs show higher sensitivity to BCR inhibitors, indicating that BCR signalling may be amplified in these cases (Dietrich et al. 2017).

Traditionally trisomy 12 has been classified as an intermediate-risk lesion (H. Döhner et al. 2000): these cases have a higher proliferative capacity, but are more treatable with chemotherapeutics and BCR inhibitors. However, *NOTCH1* mutations are frequently observed in trisomy 12 cases, and this is associated with poorer outcomes (Balatti et al. 2012; Del Giudice et al. 2012).

Epigenetic alterations

In addition to genetic lesions, the epigenome is also modified in CLL and samples typically show global hypomethylation combined with local hypermethylation (Wahlfors et al. 1992; Cahill et al. 2013; Ziller et al. 2013; Thomas J. Kipps et al. 2017; Bosch and Dalla-Favera 2019). Studies investigating the epigenome in CLL have proven revealing, in particular, higher levels of intra-sample methylation heterogeneity have been shown to be associated with high-risk genetic lesions and poorer prognosis (Landau et al. 2014).

Moreover, methylation signatures have been used to classify distinct clinical CLL subgroups (Kulis et al. 2012; Bhoi et al. 2016), as they are useful to trace the cell of origin. For example, CLL cells from distinct patients originate from many different B cell maturation states, possibly reflecting the biological and phenotypic heterogeneity of CLL (Oakes et al. 2016). Accordingly, IGHV-U CLLs have a distinct methylation signature to IGHV-M CLLs, and these patterns correspond approximately to those of pre-germinal centre or post-germinal centre memory B cells, respectively (Kulis et al. 2012; Oakes et al. 2016). Epigenetic studies have also revealed the relationship between certain

genetic lesions and specific epigenetic signatures, for example *MYD88* mutations and trisomy 12 (Beekman et al. 2018). The CLL epigenome can also be modulated by drugs and thus is of increasing clinical interest (Timp and Feinberg 2013; Beekman et al. 2018; Gaiti et al. 2019).

1.2 Therapies in CLL

Extensive work to uncover the molecular drivers of CLL has led to the development of several therapeutic strategies. CLL represents a successful example of how developing a complex understanding of the biological characteristics of a disease can lead to significantly improved patient outcomes (Yosifov et al. 2019). Treatment of CLL patients can be via chemotherapy, chemoimmunotherapy or targeted therapies that inhibit specific pathways (Thomas J. Kipps et al. 2017; Jan A. Burger 2020). Additionally, allogeneic stem cell transplantation is increasingly considered as an alternative option in relapsed or refractory patients (Thomas J. Kipps et al. 2017).

1.2.1 Chemotherapy and chemoimmunotherapy

Chemotherapy has been the standard of care for CLL for many decades, either with purine analogues such as fludarabine or alkylating agents such as chlorambucil (Robak 2005; Lukenbill and Kalaycio 2013). Chemoimmunotherapy has also benefited many patients owing to the rapid improvements to monoclonal antibody technology and the development of anti-CD20 treatments, such as rituximab (Yosifov et al. 2019; Robak et al. 2010). However, patients with higher risk lesions such as *TP53* and del(17p) do not respond well to chemoimmunotherapy, and require alternative therapeutic options (Zenz et al. 2010).

1.2.2 Targeted therapies

More recently, the importance of BCR signalling and upregulation of anti-apoptotic proteins in CLL expansion has been increasingly realised and led to the development of therapies targeting these pathways. Drugs targeting BCR signalling and BCL-2 have changed the treatment landscape dramatically (Scheffold and Stilgenbauer 2020). Three main drug classes that target BCR signalling have been developed for CLL: BTK inhibitors, PI3K inhibitors and SYK inhibitors (De Rooij et al. 2012; Thomas J. Kipps et al. 2017).

Ibrutinib is a BTK inhibitor approved for use as an initial therapy and for patients who are refractory to chemoimmunotherapy (John C. Byrd et al. 2013; John C. Byrd et al. 2014). Despite such success, complete remission is rare and many patients continue to harbour minimal residual disease within the bone marrow (John C. Byrd et al. 2013), requiring continued therapy for years (John C. Byrd et al. 2013; Woyach and Johnson 2015). Resistance can also occur via the acquisition of mutations in BTK or PLCγ 2 (Woyach and Johnson 2015) genes. Treatment initiation with ibrutinib is associated with a concomitant increase in the absolute lymphocyte count in the blood (Woyach et al. 2014), thought to caused by the inhibition of chemokine receptor signalling leading to the release of malignant B cells from the lymph nodes into the peripheral blood.

Idelalisib also acts to inhibits BCR signalling and chemokine signalling (Hoellenriegel et al. 2011), via inhibition of PI3K. Whilst the drug is highly efficacious in CLL (Furman et al. 2014; Brown et al. 2014), idelalisib demonstrates more toxicities and lower efficacy than BTK inhibitors and thus is generally used as an alternative therapy in patients for whom BTK inhibitors are unsuitable (Ghia et al. 2020; Jan A. Burger 2020).

Other drugs targeting SYK, downstream of BTK, have also shown promise in Phase I/II clinical trials (Friedberg et al. 2010), though none are licenced as yet.

Aside from BCR inhibitors, BCL-2 inhibitors such as venetoclax, have also shown good efficacy in CLL. Venetoclax is thought to induce apoptosis in CLL cells, by acting as a BH3 mimetic interfering with the ability of BCL-2 to sequester BIM (Moore et al. 2007). This makes venetoclax an attractive alternative in patients with relapsed or refractory disease (Roberts et al. 2016), or who have del(17p) i.e. loss of *TP53* (Stilgenbauer et al. 2016). Complete remissions are observed in ~20% patients (Roberts et al. 2016) which is higher than with other targeted therapies, and minimal residual disease in the bone marrow is less common (Roberts et al. 2016).

1.2.3 Management algorithm

The decision to initiate therapy is guided by the stage of disease, evidence for rapid disease progression or disease-related symptoms (Michael Hallek et al. 2008; Thomas J. Kipps et al. 2017). Disease stage is determined by an index such as the Rai staging system, which categorises patients based on disease severity (Thomas J. Kipps et al. 2017). The choice of therapy accounts for certain mutations, in particular the presence of *TP53* or del(17p) mutations, the age of the patient and the objective of therapy. IGHV status is also increasingly used for patient stratification (Figure 1.4, (Thomas J. Kipps

et al. 2017)).

CLL is a manageable disease, with a well-established arsenal of treatments and associated management algorithm. However, treatments are harsh, and the disease is still considered incurable (Bosch and Dalla-Favera 2019). Many cases develop resistance to therapy, both via acquired mutations and through survival signals provided by the microenvironment (see section 1.3.4) and minimal residual disease within the bone marrow is common. Collectively, these lead to relapse or required prolonged therapy and its associated toxicities. There remains a clinical need to improve patient outcomes and experience.

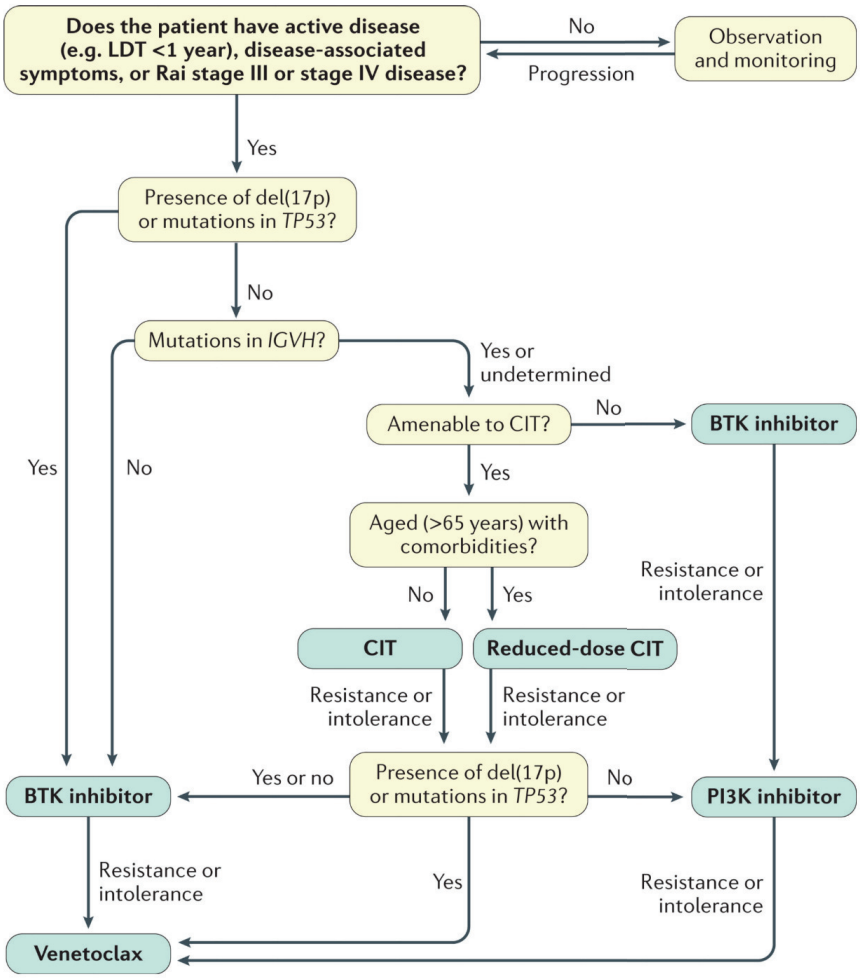

Figure 1.4: Management algorithm for patients with CLL. CIT (chemoimmunotherapy), LDT (lymphocyte doubling time). *Figure originally published in Thomas J. Kipps et al. (2017)* .

1.3 The tumour microenvironment in CLL

1.3.1 The role of the tumour microenvironment in CLL

In addition to genetic aberrations, the tumour microenvironment is an important driver of disease pathogenesis in CLL (**tenHacken2016?**). The term microenvironment encompasses the set of non-neoplastic cells within the lymphoid tissues, including the bone marrow and lymph nodes, that provide survival signals to the tumour, leading to clonal expansion and drug resistance (Jan A. Burger and Gribben 2014). Malignant B cells engage in a dialogue with the non-neoplastic cells, via cell-cell contacts and soluble factors, including chemokines, integrins, cytokines and survival factors, centring on a number of important pathways including BCR signalling and tissue homing chemokine receptors (**tenHacken2016?**).

The importance of the microenvironment in CLL pathogenesis was first recognised in studies that showed CLL cells rapidly undergo apoptosis *in vitro*, whilst their survival can be extended by stimulation or by co-culture with nurselike cells (NLCs) or mesenchymal bone marrow stromal cells (BMSCs) (Collins et al. 1989; Jan A. Burger et al. 2000; Kurtova et al. 2009; **DeaglioMalavasi2009?**; Purroy et al. 2015). This observation indicated that the ability of CLL to progressively accumulate *in vivo* may be highly dependent on external stimulation, rather than some cell-intrinsic feature of the tumour. Building on these observations, further *in vitro* studies have shown a number of cell types and soluble factors belonging to the microenvironment are also capable of protecting the tumour cells from drugs and chemotherapeutic agents. Many CLL patients continue to harbour minimal residual disease (MRD),in which a fraction of the malignant cells remain whilst the patient is in remission and eventually lead to relapse (Hayden et al. 2012). It is believed that the tumour microenvironment provides a sanctuary for the malignant B cells to shield from the effects of therapy (Dubois et al. 2020).

1.3.2 Components of the tumour microenvironment

Lymph Nodes Over the last two decades, significant progress has been made in unravelling this complex cross-talk and many of the important cellular and molecular components have been defined and studied. Malignant B cells circulate through the blood in a resting state, and follow chemokine gradients towards the lymph nodes to form “proliferation centres” (Figure 1.5), similar to germinal centres (**Herishanu2011?**). Studies using detuerated water labelling have shown that up to 3% of the clone is actively proliferating within the lymph node (B. T. Messmer et al. 2005; Herndon et al. 2017).

Cross-talk with the non-neoplastic cells in the lymph node shapes the transcriptomic profile of the malignant B cells, and leads to upregulation of the BCR pathway Mittal et al. (2014), a central driver of CLL pathogenesis.

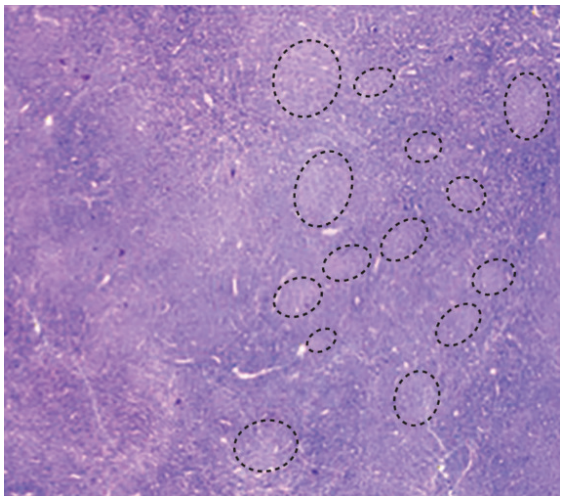

Figure 1.5: Haemotoxylin and eosin stain of CLL-infiltrated lymph node tissue section, showing pale-staining profileration centres (circled). _Figure originally published in (**Kipps2017_?**).

Bone marrow The bone marrow is also known to be important, and several studies have shown *ex vivo* BMSCs to protect CLL cells against the drug toxicity (Kay et al. 2007; Kurtova et al. 2009). However, gene expression changes are less pronounced within the bone marrow compared to the lymph node (**Herishanu2011?**).

Cellular components Within these compartments, the CLL cells engage in a dialogue with mesenchymal stromal cells (MSCs), NLCs and follicular dendritic cells (FDCs), in concert with T cells, natural killer cells (NK cells) and components of the extracellular matrix (**tenHacken2016?**). NLCs are of monocytic origin: their critical role in CLL was first demonstrated by the observation that peripheral blood-derived monocytes differentiate into NLCs, and that these cells prolong CLL cell survival *ex vivo* (Jan A. Burger et al. 2000). They are also found within the lymphoid tissues of CLL patients (Tsukada et al. 2002; Bürkle et al. 2007). MSCs, which include BMSCs, are frequently observed within the secondary lymphatic tissues of CLL patients. Many studies have demonstrated the ability of these cells to inhibit spontaneous and drug-induced apoptosis in *in vitro* CLL co-cultures Kurtova et al. (2009).

FDCs are important for tissue homing and retention of CLL cells within tissues. In

healthy tissues, they are usually found within germinal centres (Allen and Cyster 2008), and present unprocessed antigen to B cells. In CLL they play an important role within the secondary lymphoid organs, having a protective effect on CLL via cytokine secreton, adhesion molecules and the activation of BCR signalling (Dubois et al. 2020). CLL co-culture with FDCs leads to inhibition of spontaneous apoptosis and upregulation of anti-apoptotic MCL (Pedersen et al. 2002).

The T cell compartment is also altered in CLL, first described by Scrivener et al. (2003). T cells have been observed to have pro-tumour and anti-tumour behaviour. On the one hand, higher numbers of CD4+ T-helper (Th) cells are seen in CLL patient blood samples (Palma et al. 2017; **Catakovic2017?**; Elston et al. 2020), and in line with this, Th cell cytokines have been shown to provide pro-survival signals *in vitro*, for example IL4 from Th2 cells (Dancescu et al. 1992; Bhattacharya et al. 2015; Aguilar-Hernandez et al., n.d.). Activated CD4+ T-cells in murine xenograft models (Bagnara et al. 2011; Os et al. 2013) of CLL have also been shown to increase survival and growth of the tumour. On the other hand, there is evidence of increased antigen-experienced CD8+ T cells in CLL, which control tumour growth in a CLL mouse model (Roessner and Seiffert 2020; Grioni et al. 2021).

1.3.3 Microenvironmental pathways

Cross-talk between these non-neoplastic cells and the malignant B cells can occur directly, via cell-cell contacts and adhesion molecules, indirectly, via soluble factors that bind to receptors on the CLL cells, or through the exchange of material held in extracellular vesicles (Guarini et al. 2008; Oppezzo and Dighiero 2013; Crompot et al. 2017). Collectively these induce pathway activation (most importantly BCR and NFκB (**Herishanu2011?**)) and gene expression modifications with the CLL cells, leading to chemotaxis, homing to lymphoid tissues and survival of the tumour cells (Dubois et al. 2020). Figure 1.6 depicts an overview of this cross-talk.

Cell-cell contacts The importance of direct contact between cells became clear with observations that the ability of MSCs to provide efficient rescue from spontaneous and drug-induced apoptosis is dependent on direct contact and can be blocked by separation through a filter (Lagneaux et al. 1998; Jan A. Burger et al. 2000; Kay et al. 2007; Kurtova et al. 2009; Ding et al. 2009). Likewise, FDCs also operate through direct contact, as evidenced by the observation that contact with HK cells (an FDC cell line) protects against CLL cells from apoptosis (Pedersen et al. 2002).

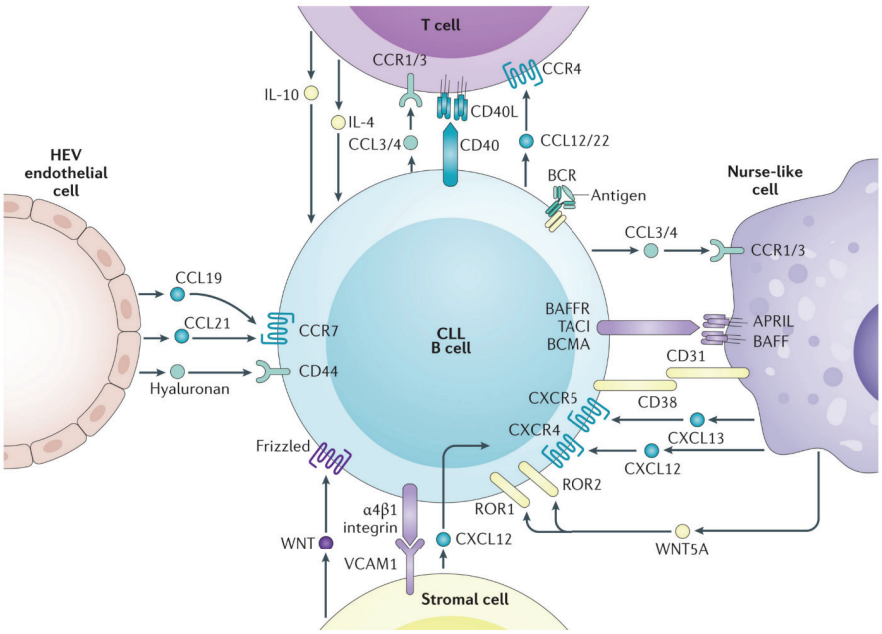

Figure 1.6: Graphic summarising soluble factors and cell-cell contacts involved in cross-talk between CLL cells and non-neoplastic cells of the tumour microenvironment *Figure originally published in Thomas J. Kipps et al. (2017) .*

This direct contact operates through a number of receptor-receptor interactions. For example, CLL-stromal cell binding involves β 1 integrin (ITGB1, or CD29) and β 2 integrin (ITGB2, or CD18) (**Lagneaux1999?**; **Lee2001?**). VLA-4 is also an important integrin for retention of CLL cells within the lymph nodes and bone marrow, by interacting with its ligand VCAM-1 (or CD106) on stromal cells (J. A. Burger et al. 2001).

These cell-cell interactions then lead to pathway activation (including BCR (Basile Stamatopoulos et al. 2015) and TLR (Schulz et al. 2011)), gene expression changes and epigenetic changes (Vangapandu et al. 2017; Xu et al. 2018) within the tumour cells. For example, contact between CLL B cells and MSCs alters the transcriptomic profile of the cells (Schulz et al. 2011; Mangolini et al. 2018), leading to increased expression of anti-apoptotic proteins such as BCL2 (**Nwabo2012?**; **Patel2013?**), BCL-XL (**Patel2013?**; Amigo-Jiménez et al. 2015), MCL1(Kurtova et al. 2009; Amigo-Jiménez et al. 2015), and β-catenin (Mangolini et al. 2018).

Soluble Factors NLCs, MSCs, FDCs and T cells also secrete soluble factors that have a protective effect on the tumour. For example, MSCs secrete a number of cytokines. One of the most widely studied is SDF1-α (or CXCL12), which interacts with CXCR4 on

CLL cells (Jan A. Burger, Burger, and Kipps 1999; J. A. Burger et al. 2001; Kay et al. 2007), stimulating the PI3K (M. Burger et al. 2005), STAT3 (Jan A. Burger, Burger, and Kipps 1999), and p44/42 MAPK (Jan A. Burger et al. 2000) pathways which activates BTK (**Montresor2011?**), ERK (D. Messmer et al. 2011), and AKT (O'Hayre et al. 2010). FDCs on the other hand secrete B cell-activating factor (BAFF), which has been shown to increase survival of CLL cells through the activation of canonical NFκB signalling (Nishio et al. 2005). A number of Th T-cell-derived cytokines have also been shown to increase CLL viability *in vitro*, including IFNγ (**Buschle2003?**), IL15 (Trentin et al. 1996), IL21 (Totero et al. 2006; Pascutti et al. 2013), IL4 (Dancescu et al. 1992; Bhattacharya et al. 2015; Aguilar-Hernandez et al., n.d.), IL2 (Decker et al. 2010) and CD40L (Kitada et al. 1999; Pascutti et al. 2013; Bhattacharya et al. 2015). Certain soluble factors can also increase apoptosis and act against the tumour, including TGFβ (Lotz, Ranheim, and Kipps 1994).

1.3.4 The influence of the microenvironment on drug response

In addition to their effect of spontaneous apoptosis, the cell types and soluble factors outlined above have also been shown to impact on drug-induced apoptosis *in vitro*. For example, NLCs and stromal cells have been shown to meditate ibrutinib resistance (Cheng et al. 2014; Boissard et al. 2015; Guo et al. 2017), and the chemotherapeutics fludarabine, oxaliplatin, chlorambucil, cyclophosphamide and doxorubicine show reduced efficacy in stromal cell co-cultures (Kay et al. 2007; Kurtova et al. 2009; Mraz et al. 2011; Zhang et al. 2012). A number of soluble factors also induce resistance to drugs *in vitro*. These include decreased efficacy of fludarabine and venetoclax in the presence of TLR stimulation (**Fonte2013?**; Kallesh D. Jayappa et al. 2017) and reduced sensitivity to ibrutinib in the presence of IL4 (Aguilar-Hernandez et al., n.d.) and BAFF (McWilliams et al. 2019).

Evidence of microenvironmentally-induced drug resistance *in vivo* is less prevalent, although there is widespread consensus that the microenvironment, in particular the lymph node, plays an important role in patient outcomes. Low rates of complete response and the inevitability of relapse in CLL have implicated the protective niche in enabling MRD (O'Brien and Kay 2011; Hayden et al. 2012). A number of studies have shown enlarged lymph nodes are associated with MRD (Moreton et al. 2005), in particular, incomplete response to ibrutinib is associated with persistently enlarged lymph nodes (Ahn et al. 2018). MSCs have also been shown to protect CLL cells taken from patients before and after *in vivo* fludarabine therapy (Trimarco et al. 2015).

In light of this, an important goal in CLL research is to develop strategies to overcome microenvironmentally-induced drug resistance. Targeting microenvironmental signalling in the lymph node tissue could be key to achieving long term remission and cure (Hayden et al. 2012) and thus there is a need for combinatorial therapies that aim to eliminate CLL cells in the lymph node and reduce CLL load in the peripheral blood. For example, Guo et al. (2017) have proposed cerdulatinib as a potential CLL therapy. Cerdulatinib is a dual inhibitor of the BCR pathway and the JAK-STAT pathway, capable of inducing cell death whilst also inhibiting the protective effects from the microenvironment.

The development of rational strategies to target the microenvironment requires a more comprehensive understanding of drug – microenvironment interactions, and how these interplay with molecular features. Some studies have worked in this direction, for example Kallesh D. Jayappa et al. (2018) tested the impact of several agonists on ibrutinib and venetoclax. In a larger scale approach, (**Giminez2020?**) applied machine learning to identify drugs targeting proteins involved in microenvironmental signalling, and later screened these drugs in combination with venetoclax and ibrutinib, for activity against CLL in the presence of stromal cells.

Previous work in our lab probed drug activity in CLL peripheral blood mononuclear cell (PBMC) samples in the context of BMSC co-culture in a large-scale screen of 81 CLL patients (Herbst 2020). Similar larger scale systematic studies of drug – microenvironment interactions, particularly in the context of molecular features, are required.

1.3.5 Modelling the tumour microenvironment

A major goal of current research in CLL is to unravel the complexity of CLL-microenvironment cross-talk and its role in drug response. Many studies have applied a range of strategies to mimic the pro-survival effect of the microenvironment, each with their own advantages (Crassini et al. 2017; Scielzo and Ghia 2020).

Stimulation with soluble factors One such strategy is to stimulate individual pathways in CLL samples *ex vivo* in order to elucidate their impact on CLL survival and drug response. For example, studies of BCR, TLR, CD40L and interleukin stimulation (Muzio et al. 2009; Crassini et al. 2017; Scielzo and Ghia 2020) have proven critical in demonstrating the marked effect each of these have on CLL survival and the key downstream pathways involved (in particular NFκB and MAPK) (Crassini et al. 2017). This strategy allows a direct understanding of cause and effect, with the caveat that accurately

mimicking cytokine concentrations *in vitro* is challenging, and that the activity of certain stimuli may be altered in the absence of other signals.

Co-culture Stimulation studies omit the impact of cell-cell contacts and thus co-culturing CLL cells with cell lines can provide a more complete picture. Various co-culture systems have been developed in order to mimic different components of the microenvironment, including stromal cells, T cells, endothelial cells, NLCs and FDCs (Panayiotidis et al. 1996; Lagneaux et al. 1998; Pedersen et al. 2002; Kurtova et al. 2009; B. Stamatopoulos et al. 2010; Basile Stamatopoulos et al. 2012; Asslaber et al. 2013; Hamilton et al. 2012; Crassini et al. 2017).

3D Models In recent years, interest has developed in the use of 3D culture systems, to create yet more *in vivo*-like models of the microenvironment (Jensen and Teng 2020; Scielzo and Ghia 2020). Static 3D approaches involve the use of scaffolds or the generation of spheroids, recapitulating the complexity of the protective niche to a greater degree (Farinello et al. 2018; Scielzo and Ghia 2020). More ambitious still is the development of dynamic 3D cultures, through the use of bioreactors and microfluidics. These systems attempt to capture the influences of gravity, flow and mechanical stresses, to study the phenotypic changes that occur as CLL cells traffic through and communicate with the non-neoplastic tissues (Walsby et al. 2014; Scielzo and Ghia 2020).

in vivo murine models Murine models to investigate CLL-microenvironment interactions are also possible (D. Lu et al. 2004; Enzler et al. 2009; **Herishanu2011?**; Fedorchenko et al. 2013; Simonetti et al. 2014; Crassini et al. 2017), though the value of these models can be hampered by species-specific biological differences (Simonetti et al. 2014) and *in vitro* modelling is often a more accurate approach (Crassini et al. 2017).

Of the many strategies to model the microenvironment, the reductionist approach of stimulating individual pathways is a useful tool to demonstrate direct causal relationships between signal and response. So far, most of these studies have investigated individual stimuli. Larger scale systematic studies of stimuli in other lymphomas have proven successful, such as work by Carey et al. (2017) to functionally screen many immune stimuli in Acute Myeloid Leukaemia (AML), suggesting that similar approaches could be valuable in CLL.

Moreover, most of these studies have been performed in smaller patient cohorts, omitting the influence of the molecular heterogeneity of CLL. Indeed, integrative studies of the interplay between external stimuli and cell-intrinsic features in CLL are lacking. A

few studies have identified interactions between genetic features and the microenvironment, for example, Martínez-Trillos et al. (2016) have established a link between MYD88 mutations and TLR response, and Chatzouli et al. (2014) demonstrated a link between IGHV status and the response to TLR activation. In addition, Mansouri et al. (2016) have discussed the convergence of mutations and external signals on the NFκB pathway.

The importance of interplay between microenvironment and molecular features in CLL survival and drug response is abundantly clear. However, a systematic study of the integrative influence of mutations and signals, particularly in the context of drug response, is missing in CLL.

1.4 Background to the approaches used in this thesis

This thesis explores drug - microenvironment - gene interplay in CLL through the analysis of *ex vivo* perturbation assays combined with multi-omic profiling of patient samples. Background information on these experimental approaches, and associated data analysis, is outlined below.

1.4.1 *Ex-vivo* drug pertubation screens

Drug perturbation screens have been invaluable in identifying pathway dependencies, biomarkers and potential therapies in CLL (Bosch and Dalla-Favera 2019). Drug perturbation screens are usually performed in microtiter plates that contain a grid of wells suitable for performing an array of pharmacological or genetic experiments (Letai 2017). Tumour cells, either cell lines or primary samples, can be deposited in each well to test their sensitivity to a set of compounds of interest. Tumour cells are incubated with each of the compounds, commonly dissolved in an aqueous solution of dimethyl sulfoxide (DMSO). After a set amount of time has passed to allow the cells to respond to the compound, the effect of each compound is measured. This “read-out” can take a number of forms. For example the morphology of the cells can imaged (Snijder et al. 2017; Herbst 2020), or the cell viability can be measured via the number of cells or the level of adenosine triphosphate (ATP) in the well (Dietrich et al. 2017).

High-throughput drug screens of cancer cell lines have been widely used to link drug responses to molecular features (Barretina et al. 2012; Basu et al. 2013; Garnett et al. 2012; Iorio et al. 2016). However, cell lines do not capture the genetic heterogeneity of

a cancer (Goodspeed et al. 2016), and thus drug screening of primary samples can be more valuable (Dietrich et al. 2017; Tyner et al. 2013; Pemovska et al. 2013; Snijder et al. 2017). In the case of CLL, screening primary samples has the caveat that CLL cells do not proliferate *ex vivo* (**Collins1987?**) and thus read-outs need to be taken on the basis of the rate of apoptosis relative to controls, rather than proliferation rate.

A few *ex vivo* perturbation screens have also investigated the impact of microenvironmental stimulation on cancer biology, for example Carey et al. (2017)'s functional screen of 94 cytokines in primary AML samples. Studies of stimuli are much rarer despite the well-recognised role of the microenvironment across haematological malignancies and other cancers.

Drug perturbation assays are also suited to combinatorial approaches, most commonly to test the efficacy of pairs of drugs in order to identify synergistic combinations (Axelrod et al. 2014; Lukas et al. 2020). Combinatorial screening to test drug efficacy in the context of microenvironmental stimulation is also possible, though rare.

1.4.2 Multiomics datasets to study CLL

Multi-omics profiling of samples, in combination with *ex vivo* perturbation screening, is a powerful approach to link cell phenotypes with molecular features in cancer:

The concept of systems biology Systems biology considers biological entities as a set of complex molecular and environmental components that each interact to shape the functional phenotype of the system as a whole (Anda-Jáuregui and Hernández-Lemus 2020). In cancer, disease pathogenesis and drug response is determined by complex interactions between mutations, epigenetic alterations, gene expression, metabolic abnormalities, and aberrant signalling functions (Anda-Jáuregui and Hernández-Lemus 2020). Thus, the study of tumour biology, and indeed CLL biology, requires integrative methodologies and analyses to decipher this complex network (Du and Elemento 2015).

Enter multi-omics, an approach to studying biological systems that utilises multiple "omic" layers to study biological entities (Anda-Jáuregui and Hernández-Lemus 2020; Menyhárt and Gyrffy 2021). These "omic" layers can encompass next-generation sequencing techniques, including DNA sequencing (DNAseq) and RNA sequencing (RNAseq) and high-throughput proteomics and metabolomics, along with newer single cell technologies and other sequence-based approaches such as ChIPseq (chromatin immunoprecipitation sequencing) and ATACseq (Assay for Transposase-Accessible Chromatin

using sequencing) (Anda-Jáuregui and Hernández-Lemus 2020). A number of methods have been built to integrate these diverse data types, calling on tools from statistics, probability, machine learning and network analysis (**Hernandez2013?**; Hernández-Lemus 2014; Argelaguet et al. 2018; Anda-Jáuregui and Hernández-Lemus 2020).

In CLL, studies have profiled each of these layers independently, including the genomic (Landau et al. 2015; Xose S. Puente et al. 2015), transcriptomic (Ferreira et al. 2014; Zenz et al. 2019), epigenomic (Rendeiro et al. 2016; Beekman et al. 2018; Mallm et al. 2019; **Rendeiro2020?**) and proteomic landscapes (Herbst 2020; Meier-Abt et al. 2021).

Building on these studies, multi-omics approaches have the power to identify causal relationships between phenotypic layers of CLL, and thus have led to important biological insights and clinical perspectives (Dietrich et al. 2017; Berest et al. 2019; Lipsky et al. 2020; Herbst 2020; J. Lu et al. 2021). For example, a recent study in our lab integrated multiple data types to identify a novel biological axis, termed CLL proliferative drive, which is strongly associated with disease outcome (J. Lu et al. 2021). Other multi-omics studies have to identified markers of drug response (Dietrich et al. 2017). Integration of ATACseq and RNAseq has also been used to determine differences in transcription factor (TF) activity in CLL between the two major subtypes (IGHV-M and IGHV-U) (Berest et al. 2019). These studies highlight the importance of integrative approaches to gain further insights into CLL pathogenesis, especially with a view towards more personalised treatment strategies for patients.

1.4.3 Mathematical modelling

A number of mathematical tools are valuable in analysing complex multi-omics datasets. In this thesis, linear regression with and without lasso penalisation is used extensively to model the impact of the microenvironment and molecular features on drug response and a basic background to these approaches is outlined below.

Basic linear regression Linear regression involves fitting a linear model to a dataset, to model the process that generated the data. For example, equation (1.1) describes a basic model to map the values of X, to the values of Y:

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{1.1}$$

This model specifies two components: the linear predictor $\beta_0 + \beta_1 X$ and the error ϵ. The linear predictor can be compared to the equation for a straight line ($Y = mX + c$), where

β_0 represents the intercept (c), and β_1, the gradient (m). The error ϵ can be modelled by sampling from a normal distribution e.g. $N(0, \sigma^2)$, a normal distribution with mean zero and variance σ^2.

Thus, we can model a variable Y via an expected value derived from the X independent variable(s), plus a random value derived from normal distribution with specified variance (Walker 2018; Huber and Holmes 2019).

Multiple Linear Regression and Interaction Effects In certain cases, multiple independent variables may be predictors of the value of Y. In these cases, multiple linear regression is required, involving more than one explanatory variable. The basic model, for two independent variables X_1 and X_2, is denoted as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \tag{1.2}$$

In some cases, the effect of an independent variable on Y may depend on the on the value of another independent variable. For example, the effect of a drug on the viability of a CLL cell may depend on whether *TP53* is mutated. Such "interactions" between independent variables can also be accounted for within linear models. These interactions are denoted as the product of two or more independent variables:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon \tag{1.3}$$

Here $X_1 X_2$ represents the interaction, and β_3 is the associated regression coefficient. Higher-order interactions with more terms are also possible.

Generalised Linear Models Not all dependent variables (Y) can be assumed to derive from sampling a normal distribution, as in equation (1.1). For example, if Y is binary and takes only the values 0 or 1, a Bernoulli distribution is more appropriate. Generalised linear modelling builds on linear regression such that the response variable can have an error distribution other than the normal distribution (Nelder and Wedderburn 1972). A number of distributions are possible, including the binomial, Poisson and gamma distributions. Where Y is categorical, binomial or multinomial distributions are valuable; for count data, the Poisson distribution is often used.

Lasso Regularisation When fitting such models to a particular dataset, it is important to avoid overfitting the data. Overfitting occurs when a model conforms too precisely to the dataset in hand. The model may not be representative of other datasets, and thus any predictions made using model are not reliable. Regularisation is an important tool to minimise such issues (Kumar n.d.).

Regularisation adds a penalty term to the best fit model. This reduces the influences of dependent variables on the value of Y, by compressing the coefficients. This often acts to reduce the number of predictors and to generate a lesser variance with the test dataset. There are two main methods for this, named L1 Lasso Regression and L2 Ridge Regression. The models described in this thesis use Lasso Regression, which is more interpretable than Ridge Regression.

Lasso regularisation shrinks coefficients towards a central point, by adding a penalty to each coefficient, equal to the absolute value of its magnitude. This shrinkage approach means than some coefficients are reduced to 0 and are eliminated from the model, generating models that are sparse and have fewer parameters (Tibshirani 1996; Kumar n.d.).

This can make models simpler to interpret and can be useful in cases where the dependent variables are highly correlated, as in these cases only one of the correlated variables will usually be assigned a coefficient. This also has the caveat that the correlated features may each independently effect Y, but only one of these variables will be deemed important by the model. It is also often impossible to determine whether one or all the correlated variables are truly influencing Y. Thus, it is important to bear in mind that whilst modelling approaches have proven invaluable in advancing our understanding of biological processes, careful interpretation is required.

1.4.4 CLL as a model for studying tumour biology

CLL represents a valuable model system in cancer and studies of CLL can offer proof-of-principle for the application of new approaches in other entities. Primary PMBC samples are relatively simple to obtain, as CLL is the most common leukaemia and biopsies are performed by taking blood samples rather than intrusive operations. In addition, multiple biopsies cab be taken over the course of a patient's monitoring and therapy.

For example, the work by J. Lu et al. (2021) to decipher a new multi-omic marker of disease aggression not only represents an important advance in our understanding of CLL drive, it also demonstrates an integrative approach to the study of cancer which could be useful in other cancers, particular where the heterogeneity of outcome remains unexplained.

Moreover, CLL is widely viewed as the prototypic disease for studying the integrative role of cell-intrinsic and cell-extrinsic features in disease initiation, expansion and drug

response (Mansouri et al. 2016; Srinivasan et al. 2020; **Opezzo2021?**). Thus, studies seeking to integrate the impact of cell-intrinsic and cell-extrinsic features on CLL survival and drug response are important in the study of CLL and beyond.