MEDICAL UNIVERSITY
OF VIENNA

Stratification and monitoring of chronic lymphocytic leukemia with high-dimensional molecular data and computational methods

Doctoral thesis at the Medical University of Vienna
for obtaining the academic degree

Doctor of Philosophy

Submitted by
André Figueiredo Rendeiro

Supervisor:
Christoph Bock
CeMM – Research Center for Molecular Medicine
of the Austrian Academy of Sciences
Lazarettgasse 14, AKH BT 25.3
1090 Vienna, Austria

Vienna, 09/2019

Declaration

The work presented in this thesis has been performed at the CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, in the laboratory of Christoph Bock. The author declares to have written all sections of this thesis.

This doctoral thesis focuses on work that has been presented in two journal papers as outlined below (a full list of publications arising from this work is provided as part of the CV on page 110):

Manuscript 1, included in the results chapter, was published in Nature Communications:
André F. Rendeiro, Christian Schmidl, Jonathan C. Strefford, Renata Walewska, Zadie Davis, Matthias Farlik, David Oscier & Christoph Bock (2016). *Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks*. Nature Communications (2016). doi:10.1038/ncomms11938

Manuscript 2, included in the results chapter, was published in Nature Communications:
André F. Rendeiro, Thomas Krausgruber, Nikolaus Fortelny, Fangwen Zhao, Thomas Penz, Matthias Farlik, Linda C. Schuster, Amelie Nemc, Szabolcs Tasnády, Marienn Réti, Zoltán Mátrai, Donat Alpar, Csaba Bödör, Christian Schmidl, Christoph Bock. *Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib drug response in chronic lymphocytic leukemia*. Nature Communications (2020). doi:10.1101/597005

Table of Contents

Abstract

Chronic lymphocytic leukemia (CLL) is a lymphoproliferative disease of B cells affecting mostly elderly individuals. While great improvement in the treatment of CLL has been made through development of targeted therapy and close patient monitoring, improved methods for stratifying patients by relative risk and necessity of treatment are still needed.

Biochemical assays powered by next-generation sequencing for chromatin and transcription profiling in bulk samples and single cells are now amenable to application in primary human material, providing an opportunity to profile large numbers of patients. These methods generate rich, high-dimensional data that can be used for patient stratification, treatment recommendation, and disease monitoring, while at the same time providing insights into the disease mechanisms.

In the course of this thesis work, we explored the value of novel high-dimensional assays measuring different layers of cellular regulation to provide insights into patient stratification and disease monitoring in two cohorts of CLL patients.

Profiling the genome-wide chromatin accessibility of a large cohort of CLL patients revealed a dynamic regulatory landscape dominated by the differentiation state of the cell-of-origin giving rise to the malignant cells in each patient. Inference of the underlying gene regulatory networks of the two major groups of CLL cells uncovered key transcription regulators in CLL. Furthermore, chromatin accessibility data were readily amenable to machine learning-powered classification of patient samples with high accuracy, attesting to the usefulness of this data type for patient stratification.

To assess how useful high-dimensional data is for disease monitoring during treatment, we investigated a second cohort of seven CLL patients starting ibrutinib therapy. Assembling a longitudinal dataset of immunophenotypes, transcription at single-cell level, and chromatin accessibility in eight time points, we were able to characterize the biological changes induced by ibrutinib across several immune cell types. While each cell type was affected differently, we identified a conserved signature of ibrutinib effect on lymphocytes characterized by a quiescent-like state as therapy progressed. This signature allowed us to monitor the patient's molecular response to treatment and was validated in an independent cohort for which bulk transcriptome profiles were available. Furthermore, using machine learning algorithms leveraging the vast single-cell dataset, we developed a method for predicting the speed of response to ibrutinib treatment for individual patients.

The work in this thesis demonstrates the power of modern high-dimensional assays for genome-centric, data-driven, personalized treatment and biological understanding of leukemia.

Zusammenfassung

Die chronische lymphatische Leukämie (CLL) ist eine lymphoproliferative Erkrankung der B-Zellen, an der vor allem ältere Patienten erkranken. Obwohl durch die Entwicklung gezielter Therapien und eine genauere Überwachung der Patienten zuletzt große Erfolge in der Behandlung der CLL erzielt werden konnten, werden dringend bessere Methoden zur Abschätzung des Risikos und Behandlungserfolges benötigt.

Molekularbiologische Methoden basierend auf DNA Sequenzierung erlauben die Analyse des Epigenoms und Transkriptoms in großen Patientenkohorten. Diese Methoden generieren umfangreiche, hochdimensionale Daten, die zur Stratifizierung von Patienten, zur Empfehlung bestimmter Behandlungen und zur Überwachung des Krankheitsverlaufes herangezogen werden können, aber auch einen genauen Einblick in die Mechanismen der Erkrankung bieten.

In dieser Dissertation untersuchen wir mittels neuartiger Verfahren verschiedene Ebenen der Genregulation, und erproben deren Anwendung zur Stratifizierung und Überwachung von Patienten in zwei Patientenkohorten mit CLL.

Genomweite Messungen der Chromatinregulation in einer großen CLL Kohorte zeigten eine sehr dynamische Regulation, dominiert vom Differenzierungsstatus der ursprünglichen Tumorzellen in jedem Patienten. Die zugrunde liegenden genregulatorischen Netzwerke in den zwei Hauptgruppen der CLL werden dominiert von einer Gruppe Transkriptionsfaktoren. Wir konnten zeigen, dass epigenetische Daten sich gut zur Klassifikation und Stratifizierung von Patienten durch maschinelles Lernen eignen.

Um die Eignung von hochdimensionellen Daten zur Überwachung der Erkrankung während der Behandlung zu prüfen, verfolgten wir eine zweite Patientenkohorte von insgesamt sieben Patienten während der Therapie mit Ibrutinib. Unsere Verlaufsstudie erfasst über acht Zeitpunkte detaillierte Immunphänotypen, Einzelzell-Transkriptome und Chromatin-Daten. Wir konnten durch Ibrutinib verursachte biologische Veränderungen in verschiedenen Immunzellen charakterisieren. Obwohl jeder Zelltyp unterschiedlich betroffen war, konnten wir im Laufe der Behandlung eine allgemeine Signatur von Ibrutinib auf Lymphozyten identifizieren, die an ruhende Zellen erinnert. Diese Signatur erlaubte uns das Ansprechen der Patienten auf die Therapie zu messen, und konnte in einer unabhängigen Patientengruppe validiert werden. Außerdem entwickelten wir Algorithmen welche die Dauer bis zum Ansprechen der Patienten vorhersagen können.

Diese Dissertation demonstriert eindrucksvoll die Anwendung hochdimensionaler Methoden zur personalisierten Therapie der Leukämie und Grundlagenforschung.

Abbreviations

ADCC: Antibody-dependent cellular cytotoxicity

AKT: RAC-alpha serine/threonine-protein kinase

amp2p: Amplification of chromosome 2p

APC: Antigen-presenting cell

BCR: B cell receptor

BTK: Bruton Tyrosine Kinase

CAR-T: Chimeric antigen receptor T cell

CD: Cluster of differentiation

CDC: Complement-dependent cytotoxicity

CLL-IPI: CLL international prognostic index

CLL: Chronic lymphocytic leukaemia

CMB: CARD11-MALT1-BCL10 signalosome

DAG: Diacylglycerol

del11q: Deletion of chromosome 11q

del13q14: Deletion of chromosome 13q14

del17p: Deletion of chromosome 17p

DLBCL: Diffuse large B cell lymphoma

DNA: Deoxyribonucleic acid

FCR: Fludarabine, cyclophosphamide, rituximab

FFPE: Formalin-fixed paraffin-embedded

FISH: fluorescent in situ hybridization

FO: Follicular B cell

GWAS: genome-wide association study

HSC: Haematopoietic stem cell

Ig: Immunoglobulin

IGHV: Immunoglobulin heavy chain variable region

ITAM: Immunoreceptor tyrosine-based activation motif

mTOR: Mechanistic target of rapamycin

MZ: Marginal zone B cell

NFKB: Nuclear factor kappa-B

NFKBIA: NFKB Inhibitor Alpha protein

OS: overall survival

PD-1: Programmed cell death protein 1

PFS: progression-free survival

PIK3: Phosphoinositide 3-kinase

PIP2: Phosphatidylinositol 4,5-bisphosphate (3,4,5)-trisphosphate

PIP3: Phosphatidylinositol (3,4,5)-trisphosphate

PKCB: Protein Kinase Cbeta

PLCG2: 1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

PTEN: Phosphatase and tensin homolog

RNA: Ribonucleic acid

SNP: single nucleotide polymorphisms

Acknowledgments

I would like to express my gratitude to many people who supported me and contributed to the work presented in this thesis.

I am extremely grateful to my advisor Christoph Bock, for the opportunity and for the great freedom, trust, and responsibility he has deposited in me.

To the members of the Bock lab in the past five years, who have inspired and mentored me. In particular Christian, Paul, and Thomas, with whom I partnered closely, but also Florian, Lukas, Michi, Nathan and Nikolaus to whom I look up to with great admiration. To Amelie, Daria, Evgeniia, Fangwen, Johanna, Linda, Matthias, Martin, Peter S., Peter T., Stefan, Thorina, Victoria, and Vitalii – a great thank you for all the help, and for sharing these amazing times in Vienna.

I would like to thank all at CeMM who have contributed to its stimulating environment that has made me a critical scientist and grow as a person.

I would like to specially thank Hatoon, Rico and Tamara, my closest companions in this journey for their constant support and unconditional friendship.

A final big thank you to my parents and Hermine for all the support.

Introduction

Chronic lymphocytic leukemia

Epidemiology, etiology and diagnosis of CLL

Chronic lymphocytic leukemia (CLL) is a mostly incurable disease of blood cells. Being one of the most common leukemias in Western countries (Rozman & Montserrat, 1995), it affects between 2 to 7 persons per 100,000 (Lenartova *et al*, 2016; Hao *et al*, 2019), constituting more than 1% of all cancers (Siegel *et al*, 2018). The median age of patients at the time of CLL diagnosis is 70 years, with only about 10% of patients diagnosed before 55 years of age. The risk of CLL developing in males is twice that of female individuals (Montserrat *et al*, 1991).

CLL patients present with an accumulation of monoclonal B cell lymphocytes (lymphocytosis) due to over-proliferation (Montserrat *et al*, 1991; Linet *et al*, 2007). These cells can accumulate in the bone marrow, spleen, liver and peripheral lymphoid organs, where they overcrowd other cells that constitute the normal hematopoietic and immune system, preventing their normal differentiation and function (Montserrat *et al*, 1991; Rozman & Montserrat, 1995; Linet *et al*, 2007). This is the root cause of most symptoms that CLL patients display. CLL cells are generally small mature lymphocytes with a particularly highly skewed ratio of nucleus-to-cytoplasm (Criel *et al*, 1997). CLL cells often form compact proliferation centres together with other cells such as prolymphocytes and stromal cells in infiltrated tissues (Herishanu *et al*, 2011; Kipps *et al*, 2017).

Other leukemias and lymphoproliferative disorders exist with similar global phenotypes and symptoms, such as Hairy cell leukemia, Mantle cell lymphoma, and Prolymphocytic leukemia (Campo *et al*, 2011). A differential diagnosis can be achieved upon observing the morphology of the cells in a blood smear and the expression levels of specific cell surface markers. The high expression of CD19, CD5, CD23, and low expression of CD20 and immunoglobulins on the cell surface is sufficient for the diagnosis of CLL (Rawstron *et al*, 2018; Kipps *et al*, 2017; Hallek *et al*, 2018).

In addition the above separate disease entities, Monoclonal B-cell lymphocytosis (MBL), and Small Lymphocytic Lymphoma (SLL) share the broad phenotype of CLL cells and are considered to be the same disease as CLL, albeith with different stages. MBL is considered a pre-malignant stage of CLL with overall less degree of lymphocytosis, while SLL tends to manifest preferencially in lymph nodes. Nonetheless, from a biological view they are the same disease (Hallek *et al*, 2018).

Cellular signaling in B cells and the pathobiology of CLL

Central to the pathophysiology of CLL is the B cell receptor (BCR) and the cellular signaling associated with it (Burger & Chiorazzi, 2013). The BCR of normal B cells is formed by a heterodimer between an immunoglobulin molecule capable of antigen binding, and an intracellular domain that effects the cellular signaling (Wang & Clark, 2003; Burger & Chiorazzi, 2013). The extracellular part is formed of a heterodimer of two heavy chain and two light chain immunoglobulin molecules. Each of the heavy or light chains comprises three parts referred to as V, D, and J. These are combined in a manner unique to each cell from a number of genes, in a process called VDJ recombination, which is unique to the immune system (Depoil *et al*, 2008). One of the sections of the heavy chain immunoglobulin genes is always one of five isotypes, namely IgG, IgM, IgA, IgE, and IgD. This process allows the body to produce cells with a myriad of receptor combinations capable of detecting an astronomical number of antigens (Depoil *et al*, 2008; Schroeder & Cavacini, 2010). Rearrangement of the heavy chain immunoglobulins is one of the first steps in the development of a B cell. The intracellular part of the BCR is constituted by the CD79a and CD79b heterodimer (also known as Ig-α/Ig-β) (Fu *et al*, 1974; Jondal, 1974; Radaev *et al*, 2010), which contain a small transmembrane moiety and immunoreceptor tyrosine-based activation motif (ITAM) which induces the signaling cascade once the receptor recognizes an antigen (Nel *et al*, 1984; Radaev *et al*, 2010).

Activation of the BCR can induce several signaling pathways. Among the most prominent is the IKK/Nuclear Factor KB (NFKB) transcription factor pathway. In this case, upon BCR receptor cross-linking after antigen binding, the CD79 dimer is phosphorylated which in turn activates the Src family kinase SYK, and the Bruton Tyrosine Kinase (BTK) subsequently (Packard & Cambier, 2013). BTK in turn phosphorilates 1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2 (PLCG2) and Phosphoinositide 3-kinase (PIK3). These last two proteins work to move a gradient of Phosphatidylinositol 4,5-bisphosphate (PIP2) to the secondary messengers Phosphatidylinositol (3,4,5)-trisphosphate (PIP3) and subsequently into diacylglycerol (DAG) in the cellular membrane (Koyasu, 2003). Following this cascade, Protein Kinase C beta (PKCB) is activated by the increased DAG concentration and in turn activates the CARD11-MALT1-BCL10 signalosome (CMB). This important signaling complex activates the IκB kinase, itself composed of IKKa (aka IKK1 or CHUK) and IKKb (aka IKK2 or IKBKB) and IKKy (also known as NEMO or IKBKG). IKKb then phosphorylates the NFKB Inhibitor Alpha protein (NFKBIA) which forms a complex with Nuclear factor NF-kappa-B p105 subunit (NFKB) and the nuclear factor NF-kappa-B p65 subunit (encoded by the RELA gene) (Kaisho *et al*, 2001). The relevant activity of NFKBIA is to sequester p65 in the cytoplasm. Proteosomal degradation of NFKBIA occurs upon its phosphorylation, which frees the p65-p105 dimer. When free, p105 is post-translationally

shortened to its N-terminus region making the p50 protein (Fan & Maniatis, 1991). Now, in the p50-p65 form, the complex translocates to the nucleus where binding to sequence-specific DNA motifs can activate transcription. NFKB target genes are generally associated with cell survival and cell proliferation.

An important molecule for BCR signaling is the co-receptor Cluster of Differentiation 19 (CD19). On the one hand CD19 has a crucial function in BCR signaling by localizing cytoplasmatic signaling proteins involved in the cascade to the cellular membrane in the vicinity of CD79 (Depoil *et al*, 2008); on the other hand, it also has a regulatory function on BCR signaling when in complex with the complement-binding receptor CD21, by lowering the threshold of BCR activation upon antigen binding (Fearon *et al*, 2000). In addition to NFKB transcription factor activation, BCR signaling is also tightly linked to activation of cellular survival and proliferation through activation of the PI3K-AKT-mTOR signaling axis. Since BCR signaling is conveyed through PIK3 kinases and the balance of PIP2 to PIP3, another important kinase that is consequentially activated is the AKT Serine/Threonine Kinase (AKT) (Pogue *et al*, 2000). AKT has several targets, among which the most prominent is mechanistic target of rapamycin (mTOR), which in turn participates in numerous signaling pathways, among which activation of the cell cycle inducing the BRAF-MAPK-MEK-ERK pathway.

In CLL cells, BCR signaling and several of the pro-proliferative downstream pathways such as NFKB, PI3K-AKT-mTOR and MAPK-MEK-ERK signaling are overactive. At the same time, signaling provided by tumor suppressors p53 and PTEN is suppressed, and several BCL2-like proteins are overexpressed. This results in the malignant, over-proliferative phenotype of B cells. There is considerable heterogeneity in both the amount of BCR signaling and the downstream efects of that cascade on CLL cells between different patients. In CLL patients, different levels of BCR signaling can cause either increased cell activation or cell anergy – a state of cellular lethargy caused by chronic surface receptor activation without proper T helper cell engagement. B cell activation is more common in CLL cells that express unmutated immunoglobulin heavy chain variable region genes (IGHV), whereas anergy is more often observed in cases expressing mutated IGHV. Anergic CLL cells have less proliferative capacity in response to BCR signaling compared with activated cells, and this accounts in part for a more indolent disease in patients with mutated IGHV genes. This interaction between cellular anergy/activation and IGHV gene mutation appears to be due to the stage of B cell differentiation of the cell from which the malignant phenotype originated.

Normal development of B cells and its influence in CLL

B cells are formed through cellular differentiation within the hematopoietic lineage, which starts with hematopoietic stem cells (HSC) in the bone marrow. HSCs differentiate throughout a human's life into multipotent progenitors and lymphoid progenitor cells – the progenitor of B and T cells (Corces *et al*, 2016; Farlik *et al*, 2016; Buenrostro *et al*, 2018). Commitment of a progenitor to the B cell lineage elicits or is concurrent with the process of somatic hypermutation of immunoglobulin genes, and the arrangement of the immunoglobulin genes of variable (V), diversity (D), and joining (J) segments (collectively also called VDJ recombination), which with the constant (C) segment part make the final immunoglobulin gene to be expressed by the B cell (Akira *et al*, 1987; Bassing *et al*, 2002). This process allows the creation of an enormous diversity of antibody conformations used by the immune system to bind arbitrary foreign antigens.

From this point, B cell development in the bone marrow is dependent on two key selection events: a) antigen-independent positive selection for the binding of the pre-BRC and BCR receptors and subsequent signaling (Bannish *et al*, 2001; Yam-Puc *et al*, 2018); b) antigen-dependent negative selection of B cells whose BCR strongly binds self-antigens (Nemazee, 2017; Yam-Puc *et al*, 2018). To complete maturation, immature B cells migrate to the spleen (transitional T1 B cells), where they further mature (transitional T2 B cells). Dependent on a complex balance of signaling stimulus, transitional B cells can become marginal zone (MZ) B cells or follicular (FO) B cells – completing their maturation and collectively called naive B cells (Yam-Puc *et al*, 2018). While MZ B cells remain mostly in the spleen, FO B cells are circulating between secondary and tertiary lymphoid organs and primary and secondary lymphoid follicles.

When B cells encounter and strongly bind an antigen, a new chapter in B cell development starts. Antigen binding in FO B cells when supported by adequate T cell help, elicits a strong activation of B cells which causes proliferation and clonal expansion of the specific antigen-recognizing cell (Parker, 1993; Yam-Puc *et al*, 2018). These cells will undergo somatic hypermutation of the variable immunoglobulin genes to increase affinity for the antigen, in a process that is dependent of the activation-induced cytidine deaminase (AID) for hypermutation and on several DNA repair pathways for the proper reparation of DNA breaks (Li, 2004). At the same time, activated B cells may also undergo a switch in the class of immunoglobulin being expressed. Since naive B cells express membrane bound IgM or IgD variants, immunoglobulin class switching allows for example the secretion of antibody to the serum and elicit an immunological action independent of direct cellular contact. B cells which specialize in antibody production develop further and are called plasma cells (also known as effector B cells), whereas a smaller fraction of activated B cells undergoes a different maturation which requires tight interactions with helper T cells, but finally reside in the germinal centres as long-lived memory B cells.

In addition to binding antigens and producing antibodies, B cells are also professional antigen-presenting cells (APCs) and cytokine secreters (Kambayashi & Laufer, 2014). The large spectrum of immunological functions that B cells display, together with their relatively complex differentiation process and relative high proliferation rate compared with most other tissues, are likely factors contributing to the fairly frequent occurrence of hematopoietic malignancies such as CLL in humans. For CLL in particular, the increasing knowledge of B cell development has brought to light the interaction of B cell development with the pathogenesis and aggressiveness of CLL, as demonstrated by the prognostic value of the mutation status of immunoglobulin genes in CLL – a proxy for the stage of B cell differentiation and a routinely measured clinical variable during patient prognostication.

Genetic and epigenetic factors of CLL

Germline genetic variation mutation contributing to CLL

A commonly used tool to identify genetic loci associated with a phenotypic trait are genome-wide association studies (GWAS). In these studies, the genotype of large numbers of donors carrying a trait of interest (e.g. a disease such as CLL) are compared with the ones from donors not carrying that trait (i.e. healthy donors) (Gibson, 2018). This allows the discovery of loci containing mostly single nucleotide polymorphisms (SNP) inherited through the germline that are associated with the trait. The procedure is usually performed in a discovery cohort and a replication cohort in order to assess the replicability of the findings. There have been at least eight GWAS studies for CLL, employing between 700 and 20,000 participants, which cumulatively yielded 101 unique associations of genomic loci containing single nucleotide polymorphisms (SNP), all studies recently jointly analysed in a meta study (Berndt *et al*, 2016). These include genes with relevance for B cell biology such as IRF4, but the discovered associations have provided little new insights into with the development and biological basis of CLL.

Somatic mutation of CLL

There is abundant evidence that somatic genetic alterations influence the development of CLL. One of the earliest discovered such factors were large chromosomal aberrations present in CLL cells (Autio *et al*, 1979; Gahrton *et al*, 1987; Ross & Stockdill, 1987; Juliusson *et al*, 1990). These manifest in alterations to the diploid number of somatic chromosomes (Gahrton *et al*, 1987), or in alterations to their structure, such as loss of an arm (Ross & Stockdill, 1987) or translocations between or within them (Autio *et al*, 1979). In approximate order of frequency, the most common chromosomal aberrations in CLL are deletions of chromosome 13q14 (del13q), chromosome 11q (del11q), trisomy of chromosome 12, amplification of chromosome 2p (amp2p), and deletion of chromosome 17p (del17p) (Landau *et al*, 2015). While varying in length and abundance within the

CLL cells, it is estimated that approximately 80% of CLL patients carry at least one of these major chromosomal aberrations (Landau *et al*, 2015). It is thought that these aberrations facilitate or promote the development of the malignant cells in part due to the genes encoded in those chromosome parts being of relevance: the del13q region encompasses the cluster of microRNA (miR) genes 15 and 16, which are known to induce apoptosis by targeting the BCL2 transcript (Cimmino *et al*, 2005); del17p overlaps the master tumor suppressor gene TP53 (Campo *et al*, 2018); del11q often causes absence of the ATM gene, which encodes a protein with DNA-repair function (Guarini et al, 2012); while the link between trisomy 12 and CLL pathogenesis is less more difficult to explain. Due to the recurrence and importance of such chromosomal aberrations, fluorescent in situ hybridization (FISH) assays have been developed and are routinely employed in routine clinical diagnostic of CLL (Hallek *et al*, 2018).

With the advent of affordable massively parallel sequencing techniques, a wider repertoire of somatic mutations was identified (Puente *et al*, 2011; Wang *et al*, 2011; Quesada *et al*, 2012; Landau *et al*, 2013; Ramsay *et al*, 2013; Baliakas *et al*, 2014; Puente *et al*, 2015; Landau *et al*, 2015). Genes that are recurrently mutated in CLL have a role in the processing of messenger and ribosomal RNA (SF3B1, XPO1, DDX3X, EWSR1, NXF1, XPO4, FUBP1, RPS15), chromatin modification and regulation (ASXL1, BAZ2A, CHD2, IKZF3, HIST1H1B, HIST1H1E, ZMYM3, MED12), DNA damage and cell cycle control (ATM, TP53, POT1, ELF4, CHEK2, DYRK1A, ELF4, BRCC3), and important signaling pathways (NOTCH1, FBXW7, MYD88, RIPK1, SAMHD1, KRAS, NRAS, TRAF2, TRAF3, EGR2, IRF4, BCOR, BRAF, MAP2K1, ITPKB, CARD11, GNB1, FUBP1, MGA, PTPN11, FBXW7).

While several of these mutations have no or little independent influence on progression-free survival (PFS) or overall survival (OS) (Landau *et al*, 2015), the functional consequences of some have been elucidated. It has been shown for example that activating mutations in Wnt pathway genes such as WNT1, CHD8, BRD7 and BCL9, reduce the viability of primary CLL cells (Wang *et al*, 2014), while mutations in the POT1 gene have caused its protein to be unable to bind telomeric regions of chromosomes, resulting in chromosomal abnormalities (Ramsay *et al*, 2013). Other examples include mutation or deletion of MGA – a regulator of the oncogene MYC (Chigrinova *et al*, 2013; Puente *et al*, 2015; Landau *et al*, 2015), mutations in IKZF3 (Puente *et al*, 2015; Landau *et al*, 2015) – an important transcription factor in B cells. Nonetheless, for most cases, precisely how the recurrent mutation of most of genes contributes to growth advantages in CLL cells during the formation of malignancy remains to be discovered.

In addition, the relative abundance of the mutations within the pool of sampled CLL cells and its quantification seems variable across different cohorts analyzed with exome and whole-genome

sequencing. Nonetheless, SF3B1, ATM and NOTCH1 mutations are generally the most abundant and all other mutations are present in frequencies below 10% (Landau *et al*, 2013; Baliakas *et al*, 2014; Puente *et al*, 2015; Landau *et al*, 2015).

With the increasing usage of whole genome sequencing recurrent somatic mutations in non-coding regions occurring in CLL cells are being discovered (Spina & Rossi, 2019). Early examples include somatic mutations in an intron of the PAX5 gene, which overlaps a regulatory element which is a putative enhancer element of the PAX5 gene – an important transcription factor for the development and survival of B cells (Puente *et al*, 2015). Another example of non-coding mutations in CLL is a mutation in the 3' untranslated region of the NOTCH1 gene. This mutation confers to carrying patients reduced overall survival, mimicking the impact of activating mutations in the coding region of NOTCH1 that alter the protein's amino-acid sequence (Puente *et al*, 2015).

While the genetic landscape of the disease has been the initial focus, genetic variation within a patient has recently become the subject of increased scrutiny (Ouillette *et al*, 2013; Landau *et al*, 2013). It is commonly assumed that all CLL cells within a patient come from a single cell that developed the malignant phenotype, gaining proliferative or survival advantage over similar other cells (Ouillette *et al*, 2013; Gruber *et al*, 2019). These cells are subject to pressure and constrain of immune, nutritional, environmental, and therapeutic origin, and therefore under the influence of natural selection (Quail & Joyce, 2013; McGranahan & Swanton, 2017). This means that cells that accumulate new mutations are more likely to have an evolutionary advantage, which gives rise to an evolutionary process with a clonal branching pattern.

Next generation sequencing provides a quantitative way to probe into the somatic genetic variability between CLL cells in the same patient and therefore a way to reconstruct the clonal structure of these cells (Landau *et al*, 2013, 2015). In addition to identifying the various clones present at the given time of sampling, it is also possible to infer hierarchical relationships between the present mutations. For example, deletion of chromosome 13, trisomy of chromosome 12 or mutations in MYD88 are often found in a large majority of CLL cells of a patient, whereas others such as NOTCH1 or deletion of chromosome 17p tend to occur in subclones (Landau *et al*, 2013). By observing these frequencies of these mutations and using parsimonious reasoning, one can infer that MYD88 mutations are much more likely to have arisen earlier, and that the NOTCH1 and del17p appeared later in the development of most CLLs (Landau *et al*, 2013, 2015). Due to this, the former mutations are thought to be more important in the start of the leukemic process.

Shifts in clonal structure of the CLL cells of a patient have also been observed along disease progression (Braggio *et al*, 2012), following treatment and are under great scrutiny for their likely role in the acquisition of resistance to treatment (Landau *et al*, 2013, 2017). For example,

prolonged use of PI3K and BTK inhibitors has been shown to create genomic instability in B cells by deregulating AID and increasing the rate of mutations and translocations (Compagno *et al*, 2017). Another example is when sub-clones harboring specific mutations causing direct resistance to the BTK inhibitor ibrutinib have been detected after treatment (Burger *et al*, 2016; Landau *et al*, 2017).

Another factor that has been shown to contribute to clonal diversity particularly in CLL is the epigenetic state within single cells (Landau *et al*, 2014; Oakes *et al*, 2014; Pastore *et al*, 2019; Gaiti *et al*, 2019). Taking advantage of measurements of DNA methylation using sequencing of bisulfite-converted DNA, it was observed that the fraction of contiguous CG di-nucleotides in the genome that have non-matching methylation pattern reflects the state of order of the epigenome and low ordering is associated with adverse clinical outcomes (Landau *et al*, 2014; Oakes *et al*, 2014). More recently, joint inspection of the DNA methylation and transcriptional landscapes in single cells showed that an inherently disorganized DNA methylation lanscape gives rise to disordered transcription (Pastore *et al*, 2019).

Genetic regulation of CLL

The first genome-scale dissection of transcription in CLL used gene expression microarrays in order to quantify the expression of CLL patient cells in an unprecedentedly unbiased manner (Klein *et al*, 2001; Rosenwald *et al*, 2001). In these two studies the authors found that a considerable portion of the transcriptome of CLL cells was different from healthy naive or memory B cells circulating in the periphery or present in germinal centres, and different from cells from non-Hodgkin lymphoma, follicular lymphoma, and diffuse large B cell lymphoma. Contributing to these differences were genes that likely have important functions in the formation or development of the malignancy such as over-expression of the anti-apoptotic BCL2, the immune checkpoint receptor CTLA4, the surface molecule CD5, the interleukin 2 receptor, and the immune-supressor cytokine TGF-Beta. Furthermore, cells from CLL patients of different IGHV mutation status were readily distinguishable, indetifying or confirming important genes used later in the stratification of CLL patients such as ZAP70 (Claus *et al*, 2014). These early studies ask many of the still relevant questions about CLL nowadays and in fact contain a roadmap for tackling challenging aspects on the biology of cancer which come to be the norm a decade later: apply modern biochemical assays on primary human tissues in a systematic manner in order to gain a systems-level perspective of the biological question at hand (Liu *et al*, 2018).

At the same time that massively parallel sequencing was illuminating the landscape of somatic mutations of CLL and many other cancers, it also allowed new insights into the regulation of the genome. DNA methylation in particular is a very appealing aspect of DNA regulation to study. This

is because DNA methylation is a key mechanism regulating gene expression and responsible for the maintenance of a certain state of genomic regulation that encodes cellular identity, lineage, and differentiation potential (Barrero *et al*, 2010). Profiling DNA methylation is also attractive from a practical point of view, since it allows the use of previously collected samples from routine clinical follow-up or even from preserved material as formalin-fixed paraffin-embedded (FFPE) in a retrospective fashion.

Following on early studies of DNA methylation in CLL cells focused on only a few CLL candidate markers with the intent of developing biomarkers for CLL stratification (Corcoran *et al*, 2005; Chantepie *et al*, 2010; Amin *et al*, 2012; Claus *et al*, 2014), massively parallel sequencing was employed to sequence the whole genome DNA methylation landscape of three types of B cells from healthy donors (Kulis *et al*, 2015). Comparing different populations of healthy B cells, the authors were able to observe a major shift in the genome-wide levels of DNA methylation: memory B cells have a massively hypo-methylated genome when compared with naive B cells. This observation stems likely from an event of major hypo-methylation that takes place in germinal centres during B cell maturation and remains in memory cells, happening prior to the immunoglobulin class switching event since it is present in both class-switched and non-class-switched memory B cells.

The implications of such a massive DNA demethylation event in B cells can also be seen in CLL cells (Kulis *et al*, 2012, 2015). Overlaying the whole-genome DNA methylation sequencing data with data from a DNA methylation microarray from CLL patients, the authors show that CLL cells likely originate from at least two distinct ends of the differentiation process of B cells that correspond broadly with the status of IGHV gene mutation. These results confirmed the earlier gene expression based studies, but shine an unprecedented light on the issue, tying B cell differentiation with the phenotype of CLL cells. On the other hand, it also demonstrated the value of understanding the epigenomic regulation of cancer compared to transcriptome sequencing, which was performed on the same matched samples but could not distinguish cells expressing unmutated from mutated IGHV genes as easily, something that was observed later (Ferreira *et al*, 2014).

The observation that CLL cells from different patients can have origin in different B cell subtypes as revealed from their DNA methylation profiles was further expanded in a subsequent paper where additional B cell populations and CLL cells from more than four hundred patients were profiled for DNA methylation (Oakes *et al*, 2016). By establishing a hierarchy of differentiation using the DNA methylation data from the healthy B cells and positioning the CLL cells along this trajectory, large differences between CLL samples was likely attributed to different stages in B cell development

from which the original malignant clone transformed rather than massive malignant reprogramming of the epigenome in CLL patients.

Survival and prognosis of CLL patients

In order to stratify the patients based on relative risk, two staging systems have been developed for CLL. The current Rai classification system comprises three stages: lymphocytosis in the blood and or bone marrow (low-risk); lymphocytosis and enlarged lymph nodes and either spleno- or hepatomegaly (intermediate); anemia or thrombocytopenia due to the disease (high-risk) (Rai *et al*, 1975). The Binet staging system is based on hemoglogin and platelet levels as well as on the number of locations with leukemic cells infiltrated. The first two stages (A and B) are both defined as having hemoglobin levels above 10 g/dL and platelets greater than $100x10^9$/L but either two or three affected areas respectively. Stage C requires that either the hemoglobin or platelet level is below these thresholds (Binet *et al*, 1977). These two systems require only routine examination and widely available blood tests and are therefore widely used (Hallek *et al*, 2018).

There are additionalprognostic factors that can be used in combination as a prognostication score for more refined risk stratification (The International CLL-IPI working Group, 2016; Hallek *et al*, 2018). The widely accepted CLL international prognostic index (CLL-IPI) combines the clinical stage, age of the patient, the IGHV mutation status of the CLL cells, serum B2 microglobulin levels, and TP53 mutations or chromosome 17p deletions of the CLL cells (The International CLL-IPI working Group, 2016). A weighted combination of these factors have been shown to have independent prognostic value in large cohorts of CLL patients across the whole spectrum of disease progression and treatment. This score eventually classifies patients in four levels of risk depending on the presence of each of these factors (The International CLL-IPI working Group, 2016).

The decision to treat a patient should not however be based on the CLL-IPI score since it has not been developed and evaluated for that purpose, but only for overall hazard (i.e. inverse of overall survival). In addition, the development of the CLL-IPI is conditioned by the therapy administered to the patients under study and certain therapies can render the independent prognostic value of a marker (Eichhorst & Hallek, 2016). For example, the poor prognosis of 11q deletion is diminished if chemotherapy achieves sufficient depletion of CLL cells (Eichhorst & Hallek, 2016). With new therapies being adopted, this is likely to continue being the case. As an example, since it has been shown that TP53 mutation is not a factor conditioning response to ibrutinib therapy (Farooqui *et al*, 2015), the prognostic score is likely to be inaccurate to estimate risk in patients undergoing ibrutinib therapy, since it uses TP53 mutation (or deletion of the small arm of chromosome 17,

where TP53 is located) for risk calculation (Brander *et al*, 2016). This illustrates the need for biomarkers and prognostic scores independent of categorical factors.

Treatment of CLL

The clinical course of CLL patients can vary widely, with patients in the lowest risk group with a 70-80% overall survival five years after diagnosis, whereas the highest risk group has a 50% survival rate over two years (The International CLL-IPI working Group, 2016). Treatment is not recommended for CLL patients in early-stage, low-risk disease (Binet stage A or B; Rai stage 0-II) which are asymptomatic. For such patients a "watch-and-wait" approach is employed, highlighting the need for accurate and robust metrics and biomarkers to stratify patients depending on the need for treatment. Moreover, when the patient meets the criteria for treatment, a careful choice of treatment should take the patient's specific symptoms, the genetic makeup of the CLL cells, the patient's previous clinical history, and the clinical indication of drugs into account.

Chemotherapy

The oldest class of drugs approved for the clinical treatment of CLL are small molecules that form the large family of chemotherapeutic drugs (Fenn & Udelsman, 2011). These molecules generally work by impairing cell proliferation, either by targeting nucleic acids or their metabolism. Broad classes of chemotherapeutic drugs are antimetabolites, alkylating agents and alkaloids. Antimetabolites, including purine analogs such as fludarabine, pentostatin, and cladribine mimic the adenosine nucleoside and inhibit ribonucleotide reductases, adenosine deaminases or DNA polymerases, thereby interfering with the process of producing or depositing adenosine in newly synthesized DNA (Swift & Golsteyn, 2014; Parker, 2009). Alkylating agents such as chlorambucil, bendamustine, and cyclophosphamide, cause covalent changes in DNA bases such as intra- and inter-strand cross-links, causing damage to DNA which can prove fatal to the cell if not repaired (Swift & Golsteyn, 2014). Alkaloid drugs, such as vinblastine and vincristine, target the cytoskeletal proteins such as tubulin, inhibiting their polymerization, which is essential for cell cycle progression (Mukhtar *et al*, 2014).

The general strategy behind the mode of action of chemotherapeutic drugs, leverages that cancer cells divide more often than most healthy cell types. Interfering with the DNA synthesis or replication processes should therefore target the treatment primarily to cancer cells (Mukhtar *et al*, 2014). While useful in many cancer types, this approach suffers from the disadvantage that the biological functions they try to interfere with are also used by healthy cells and often constitute essential cellular functions. This can result in generalized cytotoxicity, leading to extensive side effects for the patient (Swift & Golsteyn, 2014; Parikh, 2018). In addition, cancer cells can either be resistant or evolve resistance to chemotherapeutic agents for example by up-regulating anti-

apoptotic proteins, thereby circumventing programmed cell death upon the accumulation of DNA lesions (Woyach & Johnson, 2015).

Targeted treatment

In order to decrease the general cytotoxic effect of chemotherapeutic drugs by increasing specificity for cancer cells, a set of drugs aiming to target specific vulnerabilities of cancer cells has been developed over the recent years (Zhang *et al*, 2009). The growing understanding of cancer at the molecular level, in particular the identification of disease driver oncogenes and genes conferring resistance to treatment have led to rational development of drugs that target specifically used or altered key players in a specific cancer or cancer subtype (Jain & O'Brien, 2015; Woyach & Johnson, 2015). Following this paradigm, several targeted therapies that exploit the specific characteristics of CLL cells have been approved, and these currently constitute some of the most powerful types of cancer treatment (Woyach & Johnson, 2015).

Since CLL cells are phenotypically very similar to B cells, several monoclonal antibodies have been developed and approved that target cell surface molecules which are primarily expressed in B cells only (Mavromatis & Cheson, 2003; Jaglowski *et al*, 2010). The mode of action of monoclonal antibodies is either by causing Complement-dependent cytotoxicity (CDC) – activation of the complement cascade resulting in a membrane attack complex which lyses the target cells – or by Antibody-dependent cellular cytotoxicity (ADCC) – recognition of the antibody by Natural Killer cells and subsequent activation leading to the release of cytotoxic factors that cause death of the target cell (Jaglowski *et al*, 2010). The CD20 protein is an example of such a target surface molecule, with three monoclonal antibodies approved for use in CLL: Rituximab (Reff *et al*, 1994), Ofatumumab (Coiffier *et al*, 2007) and Obinutuzumab (Bologna *et al*, 2011). While the three antibodies bind different epitopes of the CD20 protein, all are capable of causing death of the target cells by both CDC and ADCC, albeit to different levels. Another interesting target protein is the sperm cell- and lymphocyte-specific CD52 surface protein, which is targeted by Alemtuzumab (Byrd *et al*, 2004), a monoclonal antibody which acts primary through ADCC.

Another vulnerability of CLL cells that can be exploited is the expression of specific proteins that make part of signalling cascades active in lymphocytes in general or B and CLL cells specifically.

The BCL2 inhibitor Venetoclax is indicated for refractory or relapsed CLL as a single agent where it has a 79% and 20% overall and complete response rate, respectively (Roberts *et al*, 2016). Of note, Venetoclax seems successful even with high-risk patients such as carriers of 17p deletion and unmutated IGHV genes. Since BCL2 over-expression is of importance to CLL cell survival by preventing apoptosis (Tzifi *et al*, 2012), its inhibition is an attractive therapeutic option.

Protein Tyrosine Kinases are a family of proteins that often participate in cellular signaling for which there are various small molecular inhibitors available due to the attractive binding pockets used (Rosenthal, 2017; Zhang *et al*, 2009). Drugs of this class constitute some of the most earlier and successful targeted treatments of leukaemia (Druker & Lydon, 2000). For CLL, inhibitors of Phosphoinositide 3-kinases (PI3Ks) complex Idelalisib (Furman *et al*, 2014; O'Brien *et al*, 2015) and Duvelisib (Flinn *et al*, 2018) are available for the treatment of relapsed disease. Moreover, Bruton's Tyrosine Kinase (BTK) inhibitor Ibrutinib has been approved for the treatment of relapsed CLL as well as first-line therapy for patients which require treatment and have been recently diagnosed (Farooqui *et al*, 2015; O'Brien *et al*, 2018; Burger *et al*, 2014). BTK acts at the centre of the signalling cascade of the B cell receptor, where it conducts signalling that eventually manifests largely as pro-proliferative, pro-inflammatory and therefore contributing to the malignant phenotype (Forconi *et al*, 2010). By inhibiting this protein in a covalent, permanent way, ibrutinib effectively stops BCR signalling within days of administration (Herman *et al*, 2014). Recent studies aim to enhance the efficacy of ibrutinib by combining it with approved drugs for CLL (Burger *et al*, 2019).

An important aspect of most cancer treatments, particularly of targeted treatments, is the acquisition of resistance to treatment over time (Zhang *et al*, 2009). Due to the nature of targeted treatments to focus on one (or a few) specific molecular entities, the acquisition of genetic mutation conferring resistance to the treatment is a clinical reality. To combat this, targeted treatments are often administered in combinations. One reason is that the usage of a combination can create an additive or synergistic effects regarding killing of cancer cells, but also that a single mutation that could confer resistance to all the effects of the combination treatment is extremely unlikely (Dancey & Chen, 2006; Zhang *et al*, 2009). Successful drug combinations in CLL include Fludarabine, Cyclophosphamide and Rituximab (FCR) (Tam *et al*, 2008) or Ibrutinib and Obinutuzumab (Moreno *et al*, 2019).

Steroids as co-adjuvants and palliative care

Steroid drugs such as the glucocorticoid prednisone and corticosteroids methylprednisolone and dexamethasone are also an important part of treatment for CLL patients. Administered in high doses to high-risk patients relapsing from chemotherapy agents (e.g. patients with TP53 mutation treated with purine analogs), they can cause quick remission (Burger & Montserrat, 2013). This is however not a viable long-term solution, but such treatment regiments are often used prior to allotransplantation to increase the odds of success. Another common use of steroid drugs in CLL is as palliative care. This can be administered as relief for pain and physical stress caused by the accumulation of CLL cells or administered jointly with chemotherapeutic drugs to induce some relief from side-effects (Pufall, 2015).

Immunotherapy

An emerging type of therapy for cancer are treatments which help the patient's own immune system fight the malignant cells – while diverse approaches are being pursued in this area, the general class of treatment is called Immunotherapy.

One major class of immunotherapy drugs are immune checkpoint inhibitors (Pardoll, 2012). These drugs work by blocking cell surface receptors with immunological activity such as the Programmed cell death protein 1 (PD-1) and its ligand PD-L1. Interaction between PD-1 and PD-L1 is an immune checkpoint that guards against immune cell activation and thereby prevents autoimmunity (Darvin *et al*, 2018). However, several cancer types hijack this as a way of suppressing the action of the immune system (Park *et al*, 2018). While immune checkpoint inhibitors are rapidly developing and successful anti-cancer therapies, none has yet been approved for clinical treatment of CLL.

Another class of treatments that also rely on the immune system are chimeric antigen receptor T cells (CAR-T cells). This approach uses T cells that are artificially engineered to express a receptor that targets a specific protein epitope (Eshhar *et al*, 1993; Turtle *et al*, 2016; Park *et al*, 2016; Abken *et al*, 2012). In CLL, CAR-T cell immunotherapy has had some early success, with high short-persistent and activity of the cells in the recipients (Brentjens *et al*, 2011) or in punctual cases an effective curative outcome (Bagg & June, 2016). Despite this, acute toxicity due to strong cytokine release as well as the high costs and logistics involved in the application of the therapy have so far prevented the broad adoption of CAR-T cell therapy for CLL. In addition, mechanisms of resistance and factors contributing to the success of CAR-T therapy in CLL have been described (Ninomiya *et al*, 2015; Fraietta *et al*, 2018), revealing that no unique therapeutic option for CLL is likely to work as a "silver bullet" for all patients. Rather, close monitoring, combination therapies, personalization of drug indications, and continuous research are the path to follow for ever-improving clinical care of CLL patients.

Aims of this thesis

The main aims of this thesis were: 1) to evaluate the practical feasibility and scientific value of epigenome and single-cell transcriptome profiling in primary human CLL samples in large scale; 2) to understand the regulatory basis of CLL during its progression and treatment; 3) to explore the use of high-dimensional analysis methods including machine learning for patient stratification and disease monitoring; 4) to discover biologically meaningful multivariate biomarkers of disease subtypes and response to treatment that can be used for differential diagnosis and treatment monitoring of CLL.

Results

Manuscript #1

The following section contains the manuscript entitled “Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks”, which has been published in Nature Communications:

André F. Rendeiro, Christian Schmidl, Jonathan C. Strefford, Renata Walewska, Zadie Davis, Matthias Farlik, David Oscier & Christoph Bock (2016). *Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks*. **Nature Communications** (2016). doi:10.1038/ncomms11938

nature COMMUNICATIONS

ARTICLE

Received 21 Dec 2015 | Accepted 16 May 2016 | Published 27 Jun 2016

DOI: 10.1038/ncomms11938 OPEN

Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks

André F. Rendeiro[1,*], Christian Schmidl[1,*], Jonathan C. Strefford[2,*], Renata Walewska[3], Zadie Davis[3], Matthias Farlik[1], David Oscier[3] & Christoph Bock[1,4,5]

Chronic lymphocytic leukaemia (CLL) is characterized by substantial clinical heterogeneity, despite relatively few genetic alterations. To provide a basis for studying epigenome deregulation in CLL, here we present genome-wide chromatin accessibility maps for 88 CLL samples from 55 patients measured by the ATAC-seq assay. We also performed ChIPmentation and RNA-seq profiling for ten representative samples. Based on the resulting data set, we devised and applied a bioinformatic method that links chromatin profiles to clinical annotations. Our analysis identified sample-specific variation on top of a shared core of CLL regulatory regions. *IGHV* mutation status—which distinguishes the two major subtypes of CLL—was accurately predicted by the chromatin profiles and gene regulatory networks inferred for *IGHV*-mutated versus *IGHV*-unmutated samples identified characteristic differences between these two disease subtypes. In summary, we discovered widespread heterogeneity in the chromatin landscape of CLL, established a community resource for studying epigenome deregulation in leukaemia and demonstrated the feasibility of large-scale chromatin accessibility mapping in cancer cohorts and clinical research.

[1] CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Lazarettgasse 14, 1090 Vienna, Austria. [2] Faculty of Medicine, Cancer Sciences, University of Southampton, Southampton SO17 1BJ, UK. [3] Department of Molecular Pathology, Royal Bournemouth Hospital, Bournemouth BH7 7DW, UK. [4] Department of Laboratory Medicine, Medical University of Vienna, 1090 Vienna, Austria. [5] Max Planck Institute for Informatics, 66123 Saarbrücken, Germany. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to C.B. (email: cbock@cemm.oeaw.ac.at).

Chronic lymphocytic leukaemia (CLL) is the most common type of leukaemia in the Western world[1]. It is characterized by a remarkable clinical heterogeneity, with some patients pursuing an indolent course, whereas others progress rapidly and require early treatment. The diverse clinical course of CLL patients, in particular those that initially present with low disease burden, fuels interest in prognostic biomarkers and personalized therapies[2]. Current clinical biomarkers for CLL include mutational status of the *IGHV* genes[3,4], *IGHV* gene family usage[5], stereotyped B-cell receptors[6,7], serum markers[8,9], chromosomal aberrations[10,11] and somatic mutations[12–14]. Most notably, *IGHV* mutation status distinguishes between a less aggressive form of CLL with mutated *IGHV* genes (mCLL) and a more aggressive form with unmutated *IGHV* genes (uCLL). Several surrogate biomarkers of *IGHV* mutation status have been described. For example, high levels of *ZAP70* expression appear to be associated with uCLL[15]. In addition to these focused biomarkers, transcriptome profiling has been used to define broader molecular signatures that may improve disease stratification independent of *IGHV* mutation status[16].

Recent genome and exome sequencing projects have identified additional genes that are recurrently mutated in CLL[17,18], some of which have prognostic significance. Nevertheless, CLL samples carry relatively few genetic aberrations compared with other adult cancers[19], and some patients develop progressive disease despite being classified as 'low risk' based on genetic markers, suggesting that non-genetic factors are relevant for CLL aetiology and outcome. Several lines of evidence point to a role of epigenome deregulation in CLL pathogenesis: first, somatic mutations have been observed in non-coding regions of the genome, where they appear to induce deregulation of relevant cancer genes[18]. Second, chromatin remodelling proteins such as *ARID1A* and *CHD2* are recurrently mutated in CLL[17,18], indicating causal links between chromatin deregulation and CLL. Third, aberrant DNA methylation was observed in all studied CLL patients[20–22], correlated with *IGHV* mutation status and identified a new subtype (iCLL) that appears to be an intermediate between mCLL and uCLL[20,23].

Although prior studies of epigenome deregulation in primary cancer samples have focused almost exclusively on DNA methylation[24], recent technological advances now make it possible to map chromatin landscapes in large patient cohorts. Most notably, the assay for transposase-accessible chromatin using sequencing (ATAC-seq) facilitates open chromatin mapping in scarce clinical samples[25] and ChIPmentation provides a streamlined, low-input workflow for genome-wide mapping of histone marks and transcription factors[26]. These two assays use a hyperactive variant of the prokaryotic Tn5 transposase, which integrates DNA sequencing adapters preferentially in genomic regions with accessible chromatin. ATAC-seq profiles are similar to those of DNase-seq, sharing the ability to detect footprints of transcription factor binding in the chromatin accessibility landscape[27]. ChIPmentation closely recapitulates the results obtained by more classical chromatin immunoprecipitation followed by sequencing protocols[26]. Both assays work well on scarce patient samples, and they enable fast sample processing on timescales that would be compatible with routine clinical diagnostics.

To establish the feasibility of large-scale chromatin analysis in primary cancer samples and to provide a basis for dissecting regulatory heterogeneity in CLL, we performed chromatin accessibility mapping using the ATAC-seq assay on a cohort of 88 primary CLL samples derived from 55 patients. Furthermore, for ten of these samples we established histone profiles using ChIPmentation for three histone marks (H3K4me1, H3K27ac and H3K27me3) and transcriptome profiles using RNA sequencing (RNA-seq). We also developed a bioinformatic method for linking these chromatin profiles to clinical annotations and molecular diagnostics data, and we performed an initial analysis of gene regulatory networks that underlie the major disease subtypes of CLL. In summary, this study provides a publicly available reference data set and a rich source of testable hypotheses for dissecting CLL biology and pathogenesis.

Results

Chromatin accessibility maps for 88 CLL samples. To map the chromatin accessibility landscape of CLL (Fig. 1a), we performed ATAC-seq on 88 purified lymphocyte samples obtained from the peripheral blood of 55 CLL patients. These patients were managed at a single medical centre, and they collectively represent the spectrum of clinical phenotypes that are commonly observed in CLL (Supplementary Data 1). Their average age at sample collection was 73 years, and 8% of patients were sampled at relapse following initial or subsequent therapy. The majority of samples (58%) had been classified as *IGHV*-mutated as part of routine clinical diagnostics (Supplementary Fig. 1 and Supplementary Data 1).

All samples selected for ATAC-seq library preparation contained at least 80% leukaemic cells. The ATAC-seq libraries were sequenced with an average of 25.4 million fragments, resulting in a data set comprising a total of 2.2 billion sequenced fragments (Supplementary Data 2). Data quality was high in all cases, with mitochondrial read rates in the expected range for ATAC-seq (mean: 38.3%; s.d.: 9.3%) and the characteristic patterns of nucleosome phasing derived from paired-end data (Supplementary Fig. 2).

The individual samples were sequenced with sufficient depth to recover the majority of chromatin-accessible regions that are detectable in each sample (Supplementary Fig. 3). Moreover, by combining data across all 88 samples we approached cohort-level saturation in terms of unique chromatin-accessible regions (Fig. 1b), indicating that our cohort is sufficiently large to identify most regulatory regions commonly accessible in CLL samples.

As illustrated for the *BLK* gene locus (Fig. 1c), our ATAC-seq data set can be aggregated into a comprehensive map of chromatin accessibility in CLL. This map comprises 112,298 candidate regulatory regions, of which 11.6% are constitutively open across essentially all CLL samples, whereas 59.1% are open in a sizable proportion of samples (5–95% of samples) and 29.3% are unique to only one or very few samples (Supplementary Fig. 4a). All data are available for interactive browsing and download from the Supplementary Website (http://cll-chromatin.computational-epigenetics.org/).

Chromatin-accessible regions in CLL are widely distributed throughout the genome, with moderate enrichment at genes and promoters (Fig. 1d and Supplementary Fig. 4b). We also compared the CLL-accessible regions with epigenome segmentations for CD19+ B cells (Fig. 1e and Supplementary Fig. 4c), a related cell type for which comprehensive reference epigenome data are publicly available[28]. Strong enrichment was observed for regions that are classified as transcription start sites or as enhancer elements in the B cells, indicative of a globally similar chromatin accessibility landscape between B cells and CLL. Nevertheless, a sizable fraction of CLL-accessible regions carried quiescent or repressive chromatin in B cells, which is the expected pattern for regulatory elements that are subject to CLL-specific activation.

Heterogeneity in the CLL chromatin accessibility landscape. Although the number of constitutively accessible regions in our

cohort was relatively low (11.6%, Supplementary Fig. 4a), we still observed high consistency between individual samples and, any two samples in our data set shared 70–98% of their chromatin-accessible regions (Supplementary Fig. 5a). Conversely, we also observed robust differences in the ATAC-seq signal intensity between samples. To facilitate gene-by-gene investigation of this heterogeneity, we established the 'chromatin accessibility corridor' as a means of aggregating the cohort-level variation into a single intuitive genome browser track (Fig. 2a and Supplementary Website). As illustrated by the *PAX5* and *BCL6* gene loci, even where the locations of chromatin accessible regions are shared across most samples, substantial differences in the ATAC-seq intensity levels were observed (Fig. 2a).

For a more systematic investigation of chromatin heterogeneity in CLL, we calculated the cohort-level variance for each of the 112,298 regions in the CLL consensus map and linked these regions to nearby genes that they may regulate (see Methods for details). Promoters of genes with a known role in B-cell biology and/or CLL pathogenesis showed significantly reduced variability ($P<10^{-5}$, Kolmogorov–Smirnov test; Supplementary Fig. 5b), which was not due to differential representation of CpG islands among the promoters of the gene sets ($P=0.49$, Fisher's exact

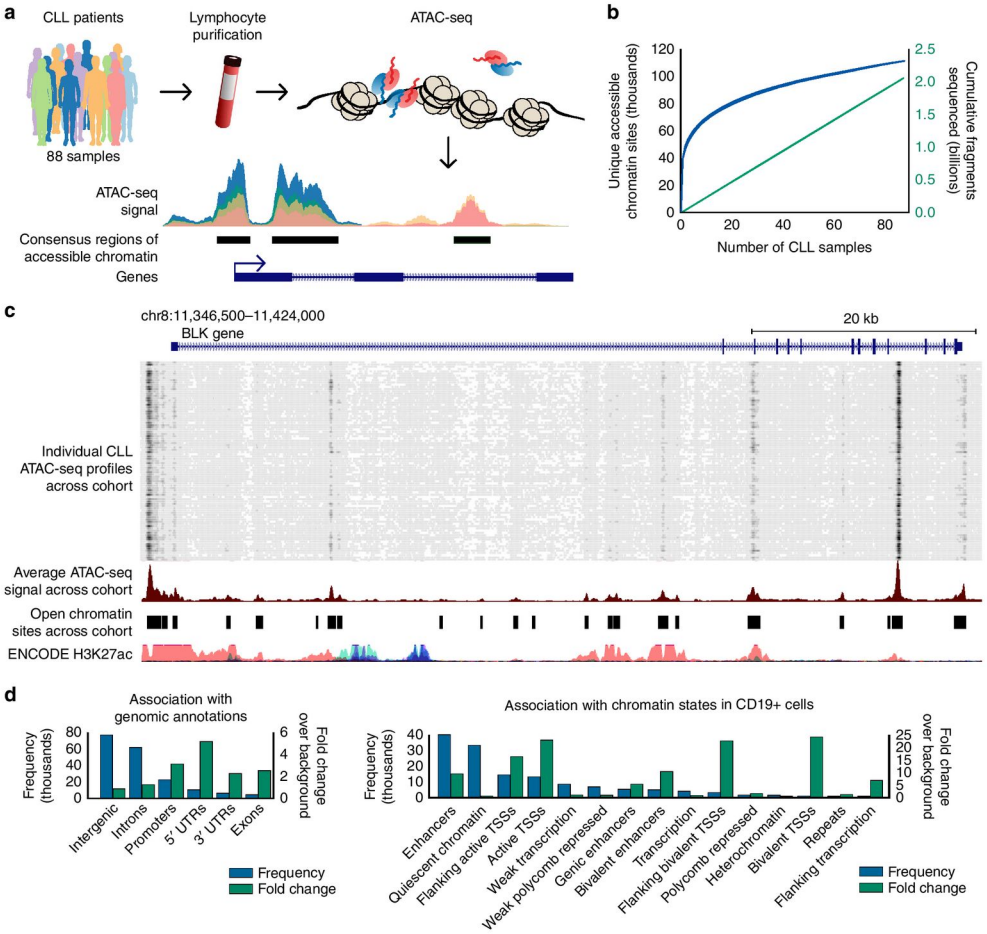


Figure 1 | The chromatin accessibility landscape of CLL. (**a**) ATAC-seq profiling and analysis workflow for establishing patient-specific and cohort-level maps of chromatin accessibility in CLL. (**b**) Saturation analysis showing the number of unique chromatin-accessible regions detected across 88 samples and with a total sequencing depth of 2.2 billion ATAC-seq fragments. The narrow blue and green corridors indicate 95% confidence intervals for samples added in random order (1,000 iterations). (**c**) Genome browser plot showing ATAC-seq signal intensity for 88 individual CLL samples (top), average signal intensity across the cohort and cohort-level peak calls (centre) and reference data from the ENCODE project (bottom). Interactive genome browser tracks are available from the Supplementary Website: http://cll-chromatin.computational-epigenetics.org/. (**d**) Absolute (frequency) and relative (fold change) co-localization of unique chromatin-accessible regions in CLL with gene annotations (left) and chromatin state segmentations for CD19+ B cells from the Roadmap Epigenomics project (right).

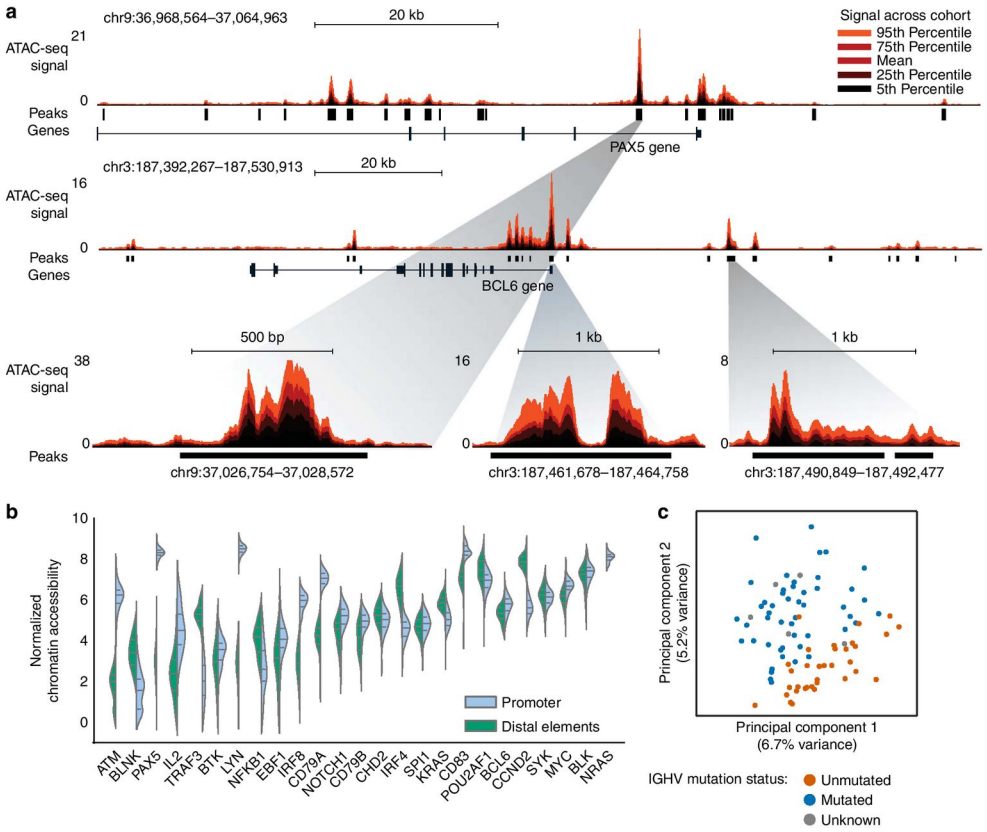


Figure 2 | Heterogeneity in the chromatin accessibility landscape of CLL. (**a**) Genome browser plot showing ATAC-seq signal intensity across the CLL cohort in the vicinity of two genes with a known role in B-cell biology (*PAX5* and *BCL6*). This cohort-level track uses colour-coded percentiles to visualize the observed heterogeneity between samples. The bottom row zooms in on the chromatin accessibility landscape at three specific regulatory regions. (**b**) Violin plots showing the cohort-wide distribution of chromatin accessibility at promoters (chromatin-accessible regions located within 2,500 bp from the transcription start site) and putative enhancers of genes with a known role in B-cell biology and/or CLL pathogenesis. (**c**) Unsupervised principal component analysis based on the chromatin accessibility for all 88 samples at each of the 112,298 chromatin-accessible regions in the CLL cohort. Samples are colour coded according to their *IGHV* mutation status, using <98% germline homology as threshold for classifying a sample as mutated.

test). For distal enhancer elements we did not observe any clear differences in heterogeneity between genes with and without a link to B cells and CLL ($P=0.08$, Kolmogorov–Smirnov test).

Beyond these global trends, the variance and distribution of chromatin accessibility across samples was highly gene specific (Fig. 2b and Supplementary Fig. 5c), as illustrated by CLL-linked genes including B-cell surface markers (*CD19*), B-cell receptor signalling components (*CD79A/B*, *LYN* and *BTK*), common oncogenes (*MYCN*, *KRAS* and *NRAS*) and genes that are recurrently mutated in CLL (*NOTCH1*, *SF3BP1*, *XPO1* and *CDKN1B*)[17,18,29].

Unsupervised principal component analysis clearly identified *IGHV* mutation status as the major source of heterogeneity in chromatin accessibility among CLL samples (Fig. 2c and Supplementary Fig. 6). However, the first two principal components explained only 6.8 and 5.2% of the total variance in the chromatin accessibility data set, suggesting that many other factors contribute to the observed differences between samples.

The most direct way by which differences in chromatin accessibility may influence disease course would be through differential regulation of CLL-relevant genes. Therefore, to systematically assess the link between chromatin accessibility and gene expression in our cohort, we performed RNA-seq on ten CLL samples with matched ATAC-seq data. A weak positive correlation was observed between chromatin accessibility and gene expression (Pearson's $r=0.33$; Supplementary Fig. 7a), which was highly dependent on the distance of the chromatin-accessible region to the nearest transcription start site (Supplementary Fig. 7b).

For chromatin-accessible regions in the vicinity of genes that RNA-seq identified as differentially expressed between *IGHV*-mutated (mCLL) and *IGHV*-unmutated (uCLL) samples (Supplementary Data 3), we observed significant differences in chromatin accessibility, which provided partial separation of the two disease subtypes (Supplementary Fig. 7c). A more pronounced separation was observed when we focused our

analysis on those regions that had been identified as differentially methylated between mCLL and uCLL in a prior study of DNA methylation in CLL[20] (Supplementary Fig. 7d).

Finally, we assessed whether patterns of differential variability between mCLL and uCLL (that is, higher levels of heterogeneity in one or the other subtype) may provide insights into the biology of these two disease subtypes. We identified 389 regions that showed a higher degree of variability among mCLL samples, whereas 581 regions were more variable among uCLL samples (Supplementary Fig. 8a)—consistent with prior results showing higher gene expression variability among uCLL samples[30]. These differentially variable regions were distributed across a broad range of ATAC-seq intensity values and were not a side effect of differences in average chromatin accessibility (Supplementary Fig. 8b). Genomic region enrichment analysis using the LOLA software[31] found mCLL-variable regions enriched for B-cell-specific transcription factor binding (*ATF2*, *BATF*, *BCL6*, *NFKB* and *RUNX3*) and active histone marks (Supplementary Fig. 8c). In contrast, uCLL-variable regions were strongly associated with the cohesin complex, including binding sites for *CTCF*, *RAD21* and *SMC3*.

Disease subtype-specific patterns of chromatin accessibility. To link the CLL chromatin accessibility landscape to clinical annotations and molecular diagnostics data (most notably to the *IGHV* mutation status that distinguishes between mCLL and uCLL), we devised a machine learning-based method that derives subtype-specific signatures directly from the data (Fig. 3a). Random forest classifiers were trained to predict whether a sample is *IGHV*-mutated or *IGHV*-unmutated, based on the chromatin accessibility values for all 112,298 regions in the CLL consensus map. We evaluated the performance of the resulting classifier by leave-one-out cross-validation and observed excellent prediction accuracy with a receiver operating characteristic (ROC) area under curve of 0.96 (Fig. 3b), which corresponds to a sensitivity of 95.6% at a specificity of 88.2%. To confirm that this cross-validated test set performance was not inflated by any form of overtraining, we repeated the same predictions one thousand times with randomly shuffled class labels. Much lower ROC area under curve values were observed in all cases, and their mean was very close to the theoretical expectation of 0.5 (Fig. 3b).

Next, we extracted the most predictive regions from the trained classifiers, giving rise to data-driven chromatin signatures that discriminate between mCLL and uCLL (Supplementary Data 4). Hierarchical clustering categorized these regions into 719 with increased chromatin accessibility in *IGHV*-mutated samples ('mCLL regions', cluster 1 in Fig. 3c) and 764 regions with increased chromatin accessibility in *IGHV*-unmutated samples ('uCLL regions', cluster 2 in Fig. 3c). More than half (51%) of these machine learning-based signature regions overlapped with statistically significant differential ATAC-seq peaks between *IGHV*-mutated and *IGHV*-unmutated samples (Supplementary Fig. 9a and Supplementary Data 4, see Methods for details). The remaining regions contributed to accurate prediction of CLL subtypes as part of a broader signature, even though they did not by themselves reach the stringent thresholds of the differential peak analysis (Supplementary Fig. 9b).

To test whether these subtype-specific chromatin signatures reflected more general differences in the gene regulatory landscape of CLL, we compared RNA-seq profiles and ChIPmentation maps for three histone marks (H3K4me1, H3K27ac and H3K27me3) between five *IGHV*-mutated and five *IGHV*-unmutated samples. We found that the genes in the vicinity of the signature regions were on average more highly expressed in the cell type showing higher chromatin accessibility (Fig. 3d and Supplementary Fig. 10), although only a small percentage of these genes were significantly differentially expressed between mCLL and uCLL samples based on our RNA-seq data (0.8% and 6.3%, respectively). Moreover, the ChIPmentation profiles were consistently associated with the differences in chromatin accessibility. Higher levels of the active H3K27ac mark as compared with repressive H3K27me3 were found in mCLL samples and mCLL-specific regions, and vice versa for uCLL (Fig. 3e). This observation is illustrated by the *ZNF667* promoter and an enhancer at the *ZBTB20* locus (Fig. 3f), two genes that have been identified as predictors of time to treatment and overall survival in CLL[32,33].

Between individual samples we observed both qualitative (that is, the presence or absence of a peak) and quantitative (that is, different peak height) differences in chromatin accessibility, as illustrated by several genes with a known role in CLL (Supplementary Figs 11 and 12). For example, the expression ratio between *ADAM29* and *LPL* has been shown to have prognostic value in CLL[34] and our data set identifies an mCLL-specific chromatin-accessible region within the *ADAM29* locus (Supplementary Fig. 11) as well as a uCLL-specific chromatin-accessible region overlapping with the *LPL* promoter (Supplementary Fig. 12), which may provide a regulatory basis for the previously described association. *CD83*, which has been associated with treatment-free survival[35], is another example of a gene locus containing an mCLL-specific chromatin-accessible region (Supplementary Fig. 11). In contrast, uCLL-specific regions were identified in the gene loci encoding the CLL-linked transcription factor CREBBP[18] and the surface protein CD38, which has been extensively validated as a prognostic factor in CLL[36] (Supplementary Fig. 12).

To gain insight into the more general biological characteristics of the mCLL and uCLL signature regions, we performed genomic region set analysis using LOLA[31] (Fig. 3g), and we observed that the mCLL regions were enriched for active promoter and enhancer regions (marked by H3K4me1 and H3K27ac) in lymphocyte-derived cell lines (SU-DHL-5, JVM-2, GM12878 and KARPAS-422), as well as binding sites of relevant transcription factors (BATF, BCL6 and BLC3). In contrast, the uCLL regions were enriched for H3K4me1-marked promoter/enhancer regions in CD38-negative naive B cells, reflecting the postulated naive B-cell origin of these CLL cells[37]. The uCLL regions were also enriched for transcribed regions (H3K36me3) in naive B cells and in B-cell-derived cell lines such as the BL-2 cell line, which has not undergone class-switch recombination.

We also performed motif enrichment analysis for the mCLL and uCLL signature region sets and, we observed significant enrichment relative to a random background model but no clear-cut differences when comparing the two region sets directly with each other (which is expected given the low statistical power of such an analysis). Nevertheless, when we linked chromatin-accessible regions to co-localized genes, we observed strong differences in the enrichment for cellular signalling pathways (Fig. 3h). The mCLL regions were associated with pathways having an established role in normal lymphocytes (CTLA4 inhibitory signaling, high-affinity IgE receptor signalling, Fc epsilon signalling and Fc gamma receptor signalling), whereas the uCLL regions were associated with cancer-associated pathways such as NOTCH signalling and fibroblast growth factor receptor signalling. All of these enrichment analyses were validated based on the statistically significant differential ATAC-seq peaks between *IGHV*-mutated and *IGHV*-unmutated samples, which gave rise to highly similar results (Supplementary Fig. 13).

Finally, we investigated whether a third CLL subtype—termed *IGHV* intermediate (iCLL)—could be detected in our data set,

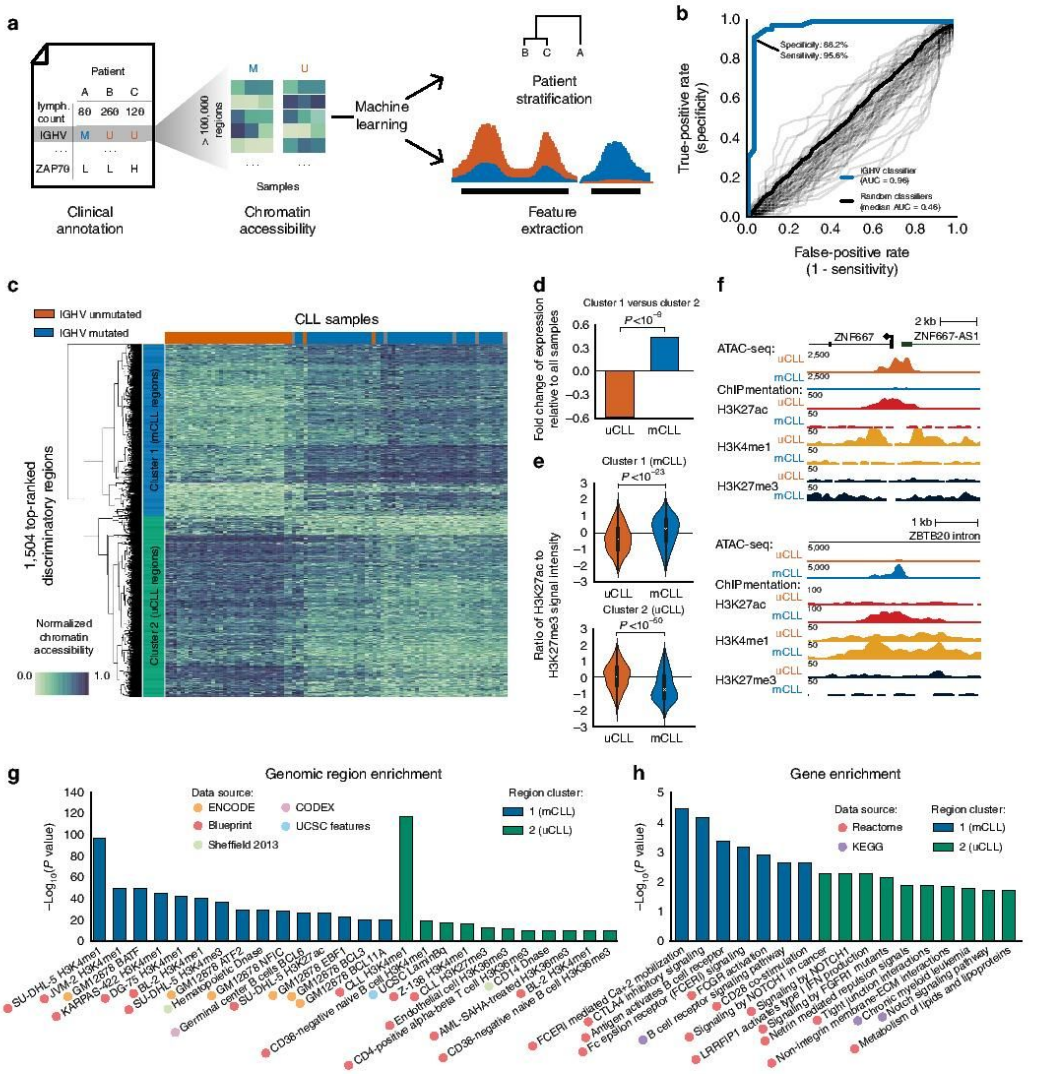


Figure 3 | Disease subtype-specific patterns of chromatin accessibility. (**a**) Methodology for deriving disease subtype-specific patterns of chromatin accessibility: a machine learning algorithm is trained to distinguish between different sample sets (here *IGHV*-mutated versus *IGHV*-unmutated), the prediction performance is evaluated by cross-validation, and the most predictive features are obtained by feature extraction from the trained classifiers. (**b**) ROC curve summarizing the test set prediction performance (estimated by leave-one-out cross-validation) of a random forest classifier that uses the ATAC-seq data set to distinguish between *IGHV*-mutated and *IGHV*-unmutated samples. 'AUC' refers to the ROC area under curve as a measure of prediction performance, and sensitivity/specificity values are shown for the point on the ROC curve that is closest to the top left corner. The grey lines indicate the performance of 1,000 classifiers trained and evaluated in the same way but based on randomly shuffled class labels. (**c**) Clustered heatmap based on the most predictive regions extracted from the cross-validated classifiers. (**d**) Ratio of expression levels for genes linked to mCLL-accessible regions versus genes linked to uCLL-accessible regions. (**e**) Ratio between ChIPmentation signal for active chromatin (H3K27ac) and repressive chromatin (H3K27me3) at mCLL-linked and uCLL-linked regions. (**f**) Genome browser plots showing ATAC-seq and ChIPmentation profiles for gene loci with a known role in CLL (*ZNF667* and *ZBTB20*). (**g**) Most highly enriched region sets for mCLL (blue) and uCLL (green) associated regions. (**h**) Most highly enriched pathways among genes linked to mCLL (blue) and uCLL (green) regions.

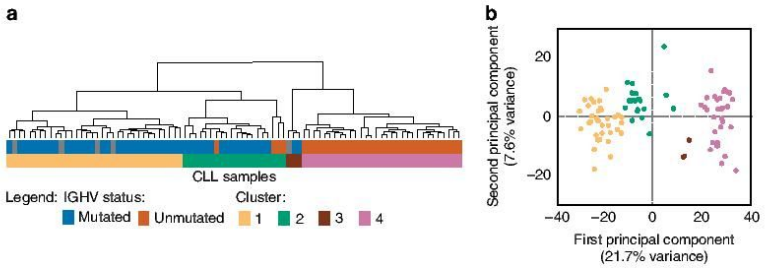


Figure 4 | Patient stratification into CLL subtypes based on chromatin accessibility. (**a**) Hierarchical clustering of all CLL samples based the classifiers' most predictive regions. Cluster 1 corresponds to mCLL, cluster 4 to uCLL, and the clusters 2 and 3 correspond to iCLL. Samples are coloured by *IGHV* mutation status (top) and cluster assignment (bottom). (**b**) Principal component analysis for the same data as in **a**, showing the first two principal components and their explained variance.

as it was recently proposed based on DNA methylation data[20,23]. Clustering all samples based on the *IGHV* mutation signature regions, we indeed observed two intermediate clusters, the larger one comprising 20 samples from 14 patients (Fig. 4a, green) and the smaller one comprising 3 samples from 2 patients (Fig. 4a, brown). Most but not all of these iCLL samples were classified as *IGHV*-mutated based on the available molecular diagnostics data (Supplementary Fig. 14). Principal component analysis provided further evidence that the iCLL samples fall between mCLL and uCLL samples based on their ATAC-seq profiles (Fig. 4b). Their intermediate character was also supported by the RNA-seq and ChIPmentation data, where the iCLL group showed patterns that consistently ranged between those of the mCLL and uCLL groups (Supplementary Fig. 15).

Gene regulatory networks in mCLL and uCLL disease subtypes. In addition to providing chromatin accessibility maps, ATAC-seq can also detect transcription factor binding based on characteristic chromatin footprints[25]. Using this property of our data, we inferred chromatin-based gene regulatory networks for CLL and its two major disease subtypes (Fig. 5a). To that end, we pooled all ATAC-seq data across the analysed samples, identified footprints for 366 transcription factors with high-quality motifs in the JASPAR database[38] and linked these regulatory elements to their putative target genes (see Methods for details). The quality of the observed footprints was comparable to those in publicly available DNase-seq data for CD19+ B cells (Supplementary Fig. 16), although we observed some deviations between the two assays that are likely due to the different sequence specificity of the Tn5 enzyme as opposed to the DNase I enzyme.

We first inferred a pan-CLL gene regulatory network using ATAC-seq data from all samples (Supplementary Fig. 17). The resulting network was dominated by highly connected transcription factors, including broadly activating factors (SP1/2/3), the insulator protein CTCF and regulators of biological processes such as cell proliferation (EGR), cell cycle (E2F) and B-cell maturation (SPI1 and PAX5). This pan-CLL network was structurally similar to a network for CD19+ B cells that we inferred from publicly available DNase-seq data using the same bioinformatic method (Supplementary Fig. 18), and in the absence of a large chromatin accessibility data set of B cells from healthy individuals it is not possible to conclusively identify the CLL-specific parts of our network.

Second, to investigate regulatory differences between CLL subtypes, we inferred gene regulatory networks separately for mCLL and uCLL samples (Supplementary Fig. 19) and identified the most differentially connected genes between the two (Fig. 5b). Genes that were more highly connected in the mCLL network coded for the transcription factors ZNF354C and ELF5, the metallopeptidase ADAM29 and the membrane protein CD22. In contrast, the BMP receptor CRIM1, the transcription factors MECOM and PAX9, the fibroblast growth factor signalling receptor FGFR1 and the membrane protein CD9 were more highly connected in the uCLL network (Fig. 5c). The more highly connected genes in either subtype also showed higher levels of H3K4me1 and H3K27ac in their regulatory elements in samples of the corresponding subtype (Supplementary Fig. 20a,b).

When we restricted our analysis to genes with a known role in B-cell biology and/or CLL pathogenesis (Fig. 5d), we observed a highly specific association of *CD22* (which codes for an inhibitory receptor involved in B-cell receptor signalling) with mCLL, whereas *CD38* and *ZAP70* were preferentially associated with uCLL. Focusing on *CD22* and *PAX9* as two high-ranking genes in our analysis, we plotted the sub-networks of their direct neighbours and observed characteristic differences between the gene regulatory networks for mCLL and uCLL (Supplementary Fig. 20c). Many of the subtype-specific genes identified by the regulatory network also showed locus-specific differences in their ChIPmentation profiles (Supplementary Fig. 20d). Altogether, our results confirm that ATAC-seq profiles are useful for identifying epigenome differences in clinical samples, and they illustrate how this data set can be used for deriving testable hypotheses about the regulatory basis of CLL.

Discussion

By ATAC-seq profiling on a large set of primary CLL samples, we have established a detailed map of the chromatin accessibility landscape in CLL. The ATAC-seq data were complemented by RNA-seq profiles and ChIPmentation for three histone marks, performed in ten representative samples covering three disease subtypes (mCLL, uCLL and iCLL). To our knowledge, this data set is currently the largest catalogue of chromatin accessibility maps for any cancer type, demonstrating the feasibility of chromatin profiling in large cohorts of primary cancer samples and validating a broadly applicable bioinformatics workflow for analysing such data.

The large number of patient samples included in this study allowed us to dissect the role of epigenome variability as a potential contributor to cancer heterogeneity[39]. We found that variability between samples was common in our data set, both in the form of qualitative (that is, the presence or absence of a peak) and quantitative (that is, different peak height) differences between individual samples. In the absence of a reference data set with chromatin accessibility maps for normal B cells from a

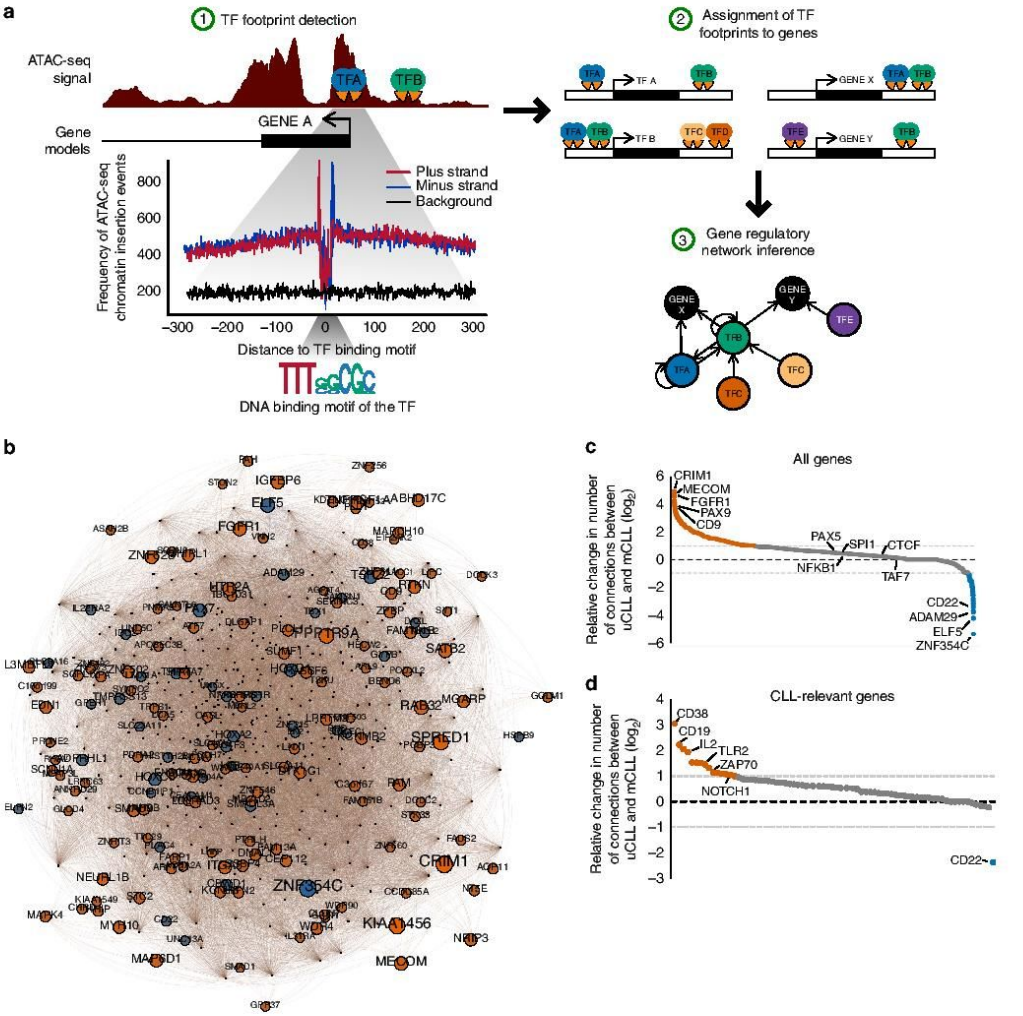


Figure 5 | Gene regulatory networks underlying the mCLL and uCLL disease subtypes. (**a**) Methodology for deriving gene regulatory networks from ATAC-seq data using transcription factor (TF) footprinting, mapping of transcription factor-binding footprints to co-localized genes and regulatory network inference. (**b**) CLL gene regulatory network derived from the data of all 88 samples, showing the most differentially connected genes between uCLL and mCLL (the full network is shown in Supplementary Fig. 17). Node size reflects the total number of connections of each node, and colours denote the subtype-specific network in which the nodes are more highly connected (mCLL: blue; uCLL: orange). (**c**) Relative change in the number of connections between the mCLL and uCLL networks, showing all genes. (**d**) Relative change in the number of connections between the mCLL and uCLL networks, focusing on genes with a known role in B-cell biology and/or CLL pathogenesis.

large number of healthy donors, it remains unclear whether or not the observed heterogeneity in CLL constitutes a major increase over the expected heterogeneity in a genetically diverse cohort. Nevertheless, significantly reduced heterogeneity at the promoters of genes involved in B-cell biology and/or CLL pathogenesis suggest a functional role of the observed inter-individual differences. Overall, our data support the existence of a core regulatory landscape shared by most or all CLL samples, which is complemented by sample-specific subsets of a substantially larger number of CLL-associated regulatory regions.

IGHV mutation status was the single biggest contributor to sample-specific differences in chromatin accessibility, although it explained only 5–10% of the observed variance in our data set. Based on the ATAC-seq profiles we were able to distinguish with excellent accuracy between *IGHV*-mutated mCLL and *IGHV*-unmutated uCLL. Our analysis also suggested the existence of one (or possibly two) intermediate type (iCLL), consistent with a recent report that used DNA methylation analysis of a large CLL cohort to identify novel CLL subtypes[20]. Chromatin accessibility and DNA methylation both appear to separate

better between these disease subtypes than gene expression data, suggesting that the biological differences between the major subtypes of CLL are primarily encoded in the epigenome and possibly reflect patterns retained from a subtype-specific cell-of-origin.

Combining data across samples provided sufficient sequencing depth for footprinting analysis of transcription factor binding, allowing us to infer gene regulatory networks from the data and to compare them between mCLL and uCLL. Although genomic footprinting has its limitations[40], the resulting network models give rise to predictions that can provide a starting point for further experimental dissection of the transcription regulatory landscape of CLL. For example, mCLL-associated regions were enriched for transcription factors that are active in mature B cells and involved in memory B-cell differentiation (BATF and BCL6), whereas the uCLL group was enriched for regulatory regions that are active in other haematopoietic cell types, indicative of a less differentiated cell state. Moreover, pathways that may boost proliferation, such as NOTCH signalling[41] and interferon signalling[42], were specifically observed in the more aggressive subtype (uCLL), whereas enrichment of inhibitory signalling by CTLA4 may contribute to the more indolent character of mCLL[43]. Beyond a small number of specific differences, the inferred gene regulatory networks were highly similar between mCLL and uCLL, consistent with the low number of differentially expressed genes that were previously observed between CLL subtypes[16,44,45].

From a technological perspective, our study describes broadly applicable methods for dissecting chromatin profiles in large cohorts of primary patient samples. The differential chromatin analysis outlined in Fig. 3 starts from clinical and/or diagnostic data and uses supervised learning techniques to identify and cross-validate discriminatory chromatin signatures. We focused specifically on *IGHV* mutation status, but the method can be applied to any type of patient grouping, for example, based on disease progression or therapy response. Moreover, the described method for ATAC-seq-based inference of gene regulatory networks (Fig. 5) establishes a data-driven approach for dissecting regulatory cell states—including their differences between disease subtypes—that is highly complementary to previous work aimed at inferring regulatory networks from transcriptome data[46–48]. Finally, the 'chromatin accessibility corridor' (Fig. 2) adapts a related concept[49] to provide intuitive browser-based visualization of chromatin data across large cohorts, while accounting for regulatory heterogeneity.

Relevant limitations of our study include the following: (i) lack of a clearly defined and experimentally accessible cell-of-origin for uCLL and mCLL, making it difficult to distinguish with certainty between chromatin patterns that are CLL specific and those that are derived from the disease's cell-of-origin; (ii) clonal heterogeneity of CLL within patients, which would be experimentally addressable only with single-cell sequencing technologies[50,51] that are currently limited in their genome-wide coverage; (iii) lack of scalable methods for distinguishing between functional and non-functional transcription factor binding; and (iv) ambiguities in the assignment of transcription factor binding sites to the genes that they regulate. In the light of these limitations, the inferred gene regulatory networks constitute an initial model that will require future refinement as additional data and validations become available.

In summary, our study establishes a chromatin accessibility landscape of CLL, which identifies shared gene regulatory networks as well as widespread heterogeneity between individual patients and between disease subtypes. It also provides a resource that can act as a starting point for deeper dissection of chromatin regulation in CLL, identification of therapeutically relevant mechanisms and eventual translation of relevant discoveries into clinical practice. Given that the chromatin profiling assays used here (ATAC-seq and ChIPmentation) are sufficiently fast and straightforward for use in a clinical sequencing laboratory, chromatin deregulation is becoming increasingly tractable as a promising source of biomarkers for stratified cancer therapy.

Methods

Sample acquisition and clinical data. All patients were diagnosed and treated at the Royal Bournemouth Hospital (UK) according to the revised guidelines of the International Workshop Chronic Lymphocytic Leukemia/National Cancer Institute. Patients were selected to reflect the clinical and biological heterogeneity of the disease. Sequential samples were included for a total of 24 patients. All samples contained more than 80% leukaemic cells. Established chromosomal rearrangements were diagnosed by fluorescence *in situ* hybridization (Abbott Diagnostics; DakoCytomation) or multiple ligation-dependent probe amplification using the MLPA P037 CLL-1 probemix (MRC Holland SALSA) according to the manufacturers' instructions. Chromosome analysis was performed and reported according to the International System for Human Cytogenetic Nomenclature. *IGHV* was sequenced as previously described[4], and a threshold of >98% germline homology was taken to define the unmutated subset[4]. The study was approved by the ethics committees of the contributing institutions (Royal Bournemouth Hospital and Medical University of Vienna). Informed consent was obtained from all participants.

ATAC sequencing. Accessible chromatin mapping was performed using the ATAC-seq method as previously described[25], with minor adaptations. In each experiment, 10^5 cells were washed once in 50 µl PBS, resuspended in 50 µl ATAC-seq lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$ and 0.1% IGEPAL CA-630) and centrifuged for 10 min at 4 °C. On centrifugation, the pellet was washed briefly in 50 µl $MgCl_2$ buffer (10 mM Tris pH 8.0 and 5 mM $MgCl_2$) before incubating in the transposase reaction mix (12.5 µl 2 × TD buffer, 2 µl transposase (Illumina) and 10.5 µl nuclease-free water) for 30 min at 37 °C. After DNA purification with the MinElute kit, 1 µl of the eluted DNA was used in a quantitative PCR (qPCR) reaction to estimate the optimum number of amplification cycles. Library amplification was followed by SPRI size selection to exclude fragments larger than 1,200 bp. DNA concentration was measured with a Qubit fluorometer (Life Technologies). Library amplification was performed using custom Nextera primers[25]. The libraries were sequenced by the Biomedical Sequencing Facility at CeMM using the Illumina HiSeq3000/4000 platform and the 25-bp paired-end configuration.

RNA sequencing. Total RNA was isolated using the AllPrep DNA/RNA Mini Kit (Qiagen). RNA amount was measured using Qubit 2.0 Fluorometric Quantitation (Life Technologies), and the RNA integrity number was determined using Experion Automated Electrophoresis System (Bio-Rad). RNA-seq libraries were prepared using a Sciclone NGS Workstation (PerkinElmer) and a Zepyhr NGS Workstation (PerkinElmer) with the TruSeq Stranded mRNA LT sample preparation kit (Illumina). Library amount and quality were determined using Qubit 2.0 Fluorometric Quantitation (Life Technologies) and Experion Automated Electrophoresis System (Bio-Rad). The libraries were sequenced by the Biomedical Sequencing Facility at CeMM using the Illumina HiSeq 3000/4000 platform and the 50-bp single-read configuration.

ChIPmentation. ChIPmentation was carried out as previously described[26], with minor adaptions. Briefly, cells were washed once with PBS and fixed with 1% paraformaldehyde in up to 1 ml PBS for 10 min at room temperature. Glycine was added to stop the reaction. Cells were collected at 500 *g* for 10 min at 4 °C (subsequent work was performed on ice and used cool buffers and solutions unless otherwise specified) and washed twice with up to 0.5 ml ice-cold PBS supplemented with 1 µM phenylmethyl sulfonyl fluoride (PMSF). The pellet was lysed in sonication buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 0.25% SDS, 1 × protease inhibitors (Sigma) and 1 µM PMSF) and sonicated with a Covaris S220 sonicator for 20–30 min in a milliTUBE or microTUBE until the size of most fragments was in the range of 200–700 bp. Lysates were centrifuged at full speed for 5 min at 4 °C, and the supernatant containing the sonicated chromatin was transferred to a new tube. The lysate was then brought to RIPA buffer conditions (final concentration: 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 140 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate, 1 × protease inhibitors (Sigma) and 1 µM PMSF) to a volume of 200 µl per immunoprecipitation. For each immunoprecipitation, 10 µl magnetic Protein A (Life Technologies) were washed twice and resuspended in PBS supplemented with 0.1% BSA. The antibody was added and bound to the beads by rotating 2 h at 4 °C. Used antibodies were H3K4me1 (0.5 µg per immunoprecipitation, Diagenode pAb-194-050), H3K27ac (1 µg per immunoprecipitation, Diagenode pAB-196-050) and H3K27me3 (1 µg per immunoprecipitation, Millipore 07-499). For control libraries, an immunoprecipitation with 2.5 µg of a nonspecific IgG rabbit antibody was used. Blocked antibody-conjugated beads were then placed on a magnet, supernatant was removed and the sonicated lysate was added to the beads followed by incubation

for 3–4 h at 4 °C on a rotator. Beads were washed subsequently with RIPA (twice), RIPA-500 (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 500 mM NaCl, 1% Triton X-100, 0.1% SDS and 0.1% DOC) (twice) and RIPA-LiCl (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 250 mM LiCl, 1% Triton X-100, 0.5% DOC and 0.5% NP40) (twice).

Beads were washed once with cold Tris-Cl pH 8.0, to remove detergent, salts and EDTA. Beads were washed once more with cold Tris-Cl pH 8.0 but the reaction was not placed on a magnet to discard supernatant immediately. Instead, the whole reaction including beads was transferred to a new tube and then placed on a magnet to remove supernatant to decrease background. Beads were then carefully resuspended in 25 µl of the tagmentation reaction mix (10 mM Tris pH 8.0, 5 mM $MgCl_2$, 10% v/v dimethylformamide) containing 1 µl Tagment DNA Enzyme from the Nextera DNA Sample Prep Kit (Illumina) and incubated at 37 °C for 1–3 min in a thermocycler. The beads were washed with RIPA (twice) and once with cold Tris-Cl pH 8. Beads were washed once more with cold Tris-Cl pH 8.0 but the reaction was not placed on a magnet to discard supernatant immediately. Instead, the whole reaction including beads was again transferred to a new tube and then placed on a magnet to remove supernatant. Beads were then incubated with 70 µl elution buffer (0.5% SDS, 300 mM NaCl, 5 mM EDTA and 10 mM Tris-HCl pH 8.0) containing 2 µl of Proteinase K (NEB) for 1 h at 55 °C and 8 h at 65 °C, to revert formaldehyde cross-linking, and supernatant was transferred to a new tube. Finally, DNA was purified with SPRI AMPure XP beads (sample-to-beads ratio 1:2) or Qiagen MinElute columns.

One microlitre of each library was amplified in a 10-µl qPCR reaction containing 0.15 µM primers, 1 × SYBR Green and 5 µl Kapa HiFi HotStart ReadyMix (Kapa Biosystems), to estimate the optimum number of enrichment cycles with the following programme: 72 °C for 5 min, 98 °C for 30 s, 24 cycles of 98 °C for 10 s, 63 °C for 30 s and 72 °C for 30 s, and a final elongation at 72 °C for 1 min. Kapa HiFi HotStart ReadyMix was incubated at 98 °C for 45 s before preparation of all PCR reactions (qPCR and final enrichment PCR), to activate the hot-start enzyme for successful nick translation at 72 °C in the first PCR step. Final enrichment of the libraries was performed in a 50-µl reaction using 0.75 µM primers and 25 µl Kapa HiFi HotStart ReadyMix. Libraries were amplified for $N+1$ cycles, where N is equal to the rounded-up Cq value determined in the qPCR reaction. Enriched libraries were purified using SPRI AMPure XP beads at a beads-to-sample ratio of 1:1, followed by a size selection using AMPure XP beads to recover libraries with a fragment length of 200–400 bp. Library preparation was performed using custom Nextera primers as described for ATAC-seq[25]. The libraries were sequenced by the Biomedical Sequencing Facility at CeMM using the Illumina HiSeq3000/4000 platform and the 25-bp paired-end configuration.

Preprocessing of the ATAC-seq data. Reads were trimmed using Skewer[52]. Trimmed reads were aligned to the GRCh37/hg19 assembly of the human genome using Bowtie2 (ref. 53) with the '-very-sensitive' parameter. Duplicate reads were removed using sambamba *markdup*[54], and only properly paired reads with mapping quality >30 and alignment to the nuclear genome were kept. All downstream analyses were performed on the filtered reads. Genome browser tracks were created with the *genomeCoverageBed* command in BEDTools[55] and normalized such that each value represents the read count per base pair per million mapped and filtered reads. Finally, the UCSC Genome Browser's *bedGraphToBigWig* tool was used to produce a bigWig file. Combined tracks with percentile signal across the cohort were created by quantifying ATAC-seq read coverage at every reference genome position using BEDTools *coverage* and normalizing it between samples. Normalization was done by dividing each value by the total number of filtered reads and multiplying it with ten million, to obtain numbers that are comparable and easy to visualize. Next, the mean as well as the 5th, 25th, 75th and 95th percentiles of signal across the whole cohort were calculated with Numpy, converted into bedgraph files and subsequently to bigwig format using *bedGraphToBigWig*. Peak calling was performed with MACS2 (ref. 56) using the '-nomodel' and '-extsize 147' parameters, and peaks overlapping blacklisted features as defined by the ENCODE project[57] were discarded.

Preprocessing of the RNA-seq data. Reads were trimmed with Trimmomatic[58] and aligned to the GRCh37/hg19 assembly of the human genome using Bowtie1 (ref. 59) with the following parameters: -q -p 6 -a -m 100—minins 0—maxins 5000—fr—sam—chunkmbs 200. Duplicate reads were removed with Picard's *MarkDuplicates* utility with standard parameters before transcript quantification with BitSeq[60] using the Markov chain Monte Carlo method and standard parameters. To obtain gene-level quantifications, we assigned the expression values of its highest expressed transcript to each gene. Differential gene-level expression between the three *IGHV* mutation status groups was performed using DESeq2 (ref. 61) from the raw count data with a significance threshold of 0.05. To produce genome browser tracks, we mapped the reads to the genomic sequence of the GRCh37/hg19 assembly of the human genome using Bowtie2 (ref. 53) with the '-very-sensitive' parameter, removed duplicates using sambamba *markdup*[54] and used the *genomeCoverageBed* command in BEDTools[55] to produce a bedgraph file. This file was normalized such that each value represents the read count per base pair per million filtered reads, and the UCSC Genome Browser's *bedGraphToBigWig* tool was used to convert it into a bigWig file.

Preprocessing of the ChIPmentation data. Reads were trimmed using Skewer[52]. Trimmed reads were aligned to the GRCh37/hg19 assembly of the human genome using Bowtie2 (ref. 53) with the '-very-sensitive' parameter. Duplicate reads were removed using sambamba *markdup*[54], and only properly paired reads with mapping quality >30 and alignment to the nuclear genome were kept. All downstream analyses were performed on the filtered reads. Genome browser tracks were created with the *genomeCoverageBed* command in BEDTools[55] and normalized such that each value represents the read count per base pair per million filtered reads. Finally, the UCSC Genome Browser's *bedGraphToBigWig* tool was used to produce a bigWig file.

Bioinformatic analysis of chromatin accessibility. The CLL consensus map was created by merging the ATAC-seq peaks from all samples using the BEDTools[55] *merge* command. To produce Fig. 1b, we counted the number of unique chromatin-accessible regions after merging peaks for each sample in an iterative manner, randomizing the sample order 1,000 times and computing 95% confidence intervals across all iterations. The chromatin accessibility of each region in each sample was quantified using Pysam, counting the number of reads from the filtered BAM file that overlapped each region. To normalize read counts across samples, we performed quantile normalization using the *normalize.quantiles* function from the preprocessCore package in R. For each genomic region we calculated the support as the percentage of samples with a called peak in the region, and we calculated four measures of ATAC-seq signal variation across the cohort: mean signal, s.d., variance-to-mean ratio and the squared coefficient of variation (the square of the s.d. over the mean). In addition, we used BEDTools *intersect* to annotate each region with the identity of and distance to the nearest transcription start site and the overlap with Ensembl gene annotations (promoters were defined as the 2,500-bp region upstream of the transcription start site). Annotation with chromatin states was based on the 15-state genome segmentation for CD19 + B cells from the Roadmap Epigenomics Project[62] (identifier: E032).

To summarize the chromatin accessibility signals into one value per gene (Fig. 2b and Supplementary Fig. 5), we used the accessibility values of the closest region (but no further than 1,000 bp from the transcription start site) to represent the promoter and the mean values of all distal regions (located more than 2,500 bp from the transcription start site) of each gene to represent distal regulatory elements. To test for overrepresentation of CpG islands in the promoters of genes with a known role in B-cell biology and/or CLL pathogenesis, we downloaded the position of CpG islands in the GRCh37/hg19 assembly from the UCSC Genome Browser[63], counted the number of promoters (as defined above) that overlapped by at least 1 bp with CpG islands in the gene set of interest and in all other genes with accessible elements in CLL, and used Fisher's exact test to assess the significance of the association. Unsupervised principal component analysis was performed with the scikit-learn[64] library (*sklearn.decomposition.PCA)* applied to the chromatin accessibility values of all chromatin-accessible regions across the CLL cohort.

To investigate variability within the mCLL and uCLL sample groups, we divided the samples in two groups based on their *IGHV* mutation status (samples below a 98% homology threshold were considered mutated, and samples with missing values for the *IGHV* mutation status were excluded from the analysis) and we used the F test from the *var.test* function in R on the chromatin accessibility values of all CLL cohort regions. Significantly variable regions were defined as having a Bonferroni-corrected P-value below 0.05 and mean accessibility above 1. Region set-enrichment analysis was performed on the significantly variable regions of each group using LOLA[31] with its core databases: transcription factor binding sites from ENCODE[57], tissue clustered DNase hypersensitive sites[65], the CODEX database[66], UCSC Genome Browser annotation tracks[63], the Cistrome database[67] and data from the BLUEPRINT project[68]. Motif enrichment analysis was performed with the AME tool from the MEME suite[69] using 250 bp sequences centred on the chromatin-accessible regions and randomly generated sequences of the same length and set size from a distribution of zeroth- and first-order Markov order (single nucleotides and dinucleotide) frequencies as background.

Machine learning analysis of disease subtypes. Random forest classifiers from the scikit-learn[64] Python library (*sklearn.ensemble.RandomForestClassifier*) were trained with the samples' *IGHV* mutation status as class label and the chromatin accessibility values for each sample at each of the 112,298 consensus regions as input features (prediction attributes). All samples with known *IGHV* mutation status were used for class prediction, the performance was evaluated by leave-one-out cross-validation, and the results were plotted as ROC curves using scikit-learn. Given that several patients contributed more than one sample to the cohort, in each iteration of the cross-validation we removed any samples from the training set that belonged to the same patient as the sample in the test set, to eliminate a potential risk of overtraining. Furthermore, we repeated the cross-validation 1,000 times based on randomly shuffled class labels to confirm that no overtraining occurred in our analysis. The most predictive regions for *IGHV* mutation status were selected by averaging the feature importance of the random forest classifiers over all iterations of the cross-validation and selecting those features with Gini importance higher than 10^{-4}. Region set enrichment was performed using LOLA[31] as described above. Pathway enrichment analysis was performed using seq2pathway[70]. The sample clustering in Fig. 4a was based on the pairwise correlation of ATAC-seq signal in the predictive regions between samples, and the

dendrogram was plotted using Scipy's hierarchical clustering function. With the same values of chromatin accessibility from above, we performed principal component analysis on the CLL samples using R's implementation in the *prcomp* function. To provide further validation of the machine learning analysis, we also identified differential ATAC-seq peaks between *IGHV*-mutated and *IGHV*-unmutated samples using the DESeq2 R package[61]. This statistical analysis was based on read counts for all CLL-accessible regions in each patient, testing for differential chromatin accessibility using a model based on the negative binomial distribution. Regions with Benjamini–Hochberg adjusted *P*-values below 0.01 and an absolute log2 fold change above 1 were used for comparison with those signature regions identified by the machine-learning analysis.

Gene regulatory network inference. Transcription factor binding maps as the basis for inferring gene regulatory networks were derived by footprinting analysis using the PIQ software[71] and a set of 366 human transcription factor motifs from the JASPAR database[38]. We retained only those transcription factors with at least 500 high-purity (>0.7) binding sites overlapping with an ATAC-seq peak, as previously described[72]. Binding sites located in the gene body or in the 2,500-bp region upstream of its transcription start site were assigned to the overlapping gene(s), and intergenic binding sites were assigned to the gene whose transcription start site was closest to the peak. This assignment was based on the Ensembl gene annotation version 75, and we treated non-protein-coding genes in the same way as protein-coding genes. To infer gene regulatory networks, an interaction score was calculated in a similar way as previously described[72]: the interaction score between a transcription factor t and a gene g (S_{tg}) was defined as the sum over all n transcription factor binding sites of t that can be assigned to g:

$$S_{tg} = \sum_{i=1}^{n} 2*(P_i - 0.5)*10^{-\left(\frac{d_{ig}}{100000}\right)}$$

In this formula P_i is the PIQ purity score, and d_{ig} is the distance of a particular transcription factor binding site i to gene g. This score establishes a unidirectional (transcription factors to genes) and weighted (based on the interaction score) relationship, providing the edges of the gene regulatory network. We inferred gene regulatory networks for all samples combined and also separately for the two disease subtypes (mCLL and uCLL) based on *IGHV* mutation status. We considered only transcription-factor-to-gene interactions with scores above 1, and in Fig. 5b and Supplementary Figs 17 and 19 we plotted only nodes with more than 200 connections. For the CD19+ B-cell gene regulatory network we used DNase-seq data from the Roadmap Epigenomics Project[62] (identifier: E032). Both the processing of the raw data and the network inference were performed in the same manner as for ATAC-seq. The comparison of composition and structural characteristics of the gene regulatory networks inferred from ATAC-seq data for the CLL cohort and from DNase-seq data for CD19+ B cells was done using functions from the *networkx*[73] library in Python. The inferred networks were visualized using the Gephi software, applying the Force Atlas 2 graph layout with LinLog and hub dissuasion. To compare the inferred mCLL and uCLL networks, we divided the degree of each node by the total number of edges in each network, which compensates for differences in the absolute number of detected interactions, and quantified differences by subtracting and log2-transforming this value between networks for each node.

Data availability. All data are available as genome browser tracks for interactive browsing and download from the Supplementary Website (http://cll-chromatin.computational-epigenetics.org/). The processed data are also openly available from NCBI GEO under the accession number GSE81274, whereas the raw sequencing data are available from EBI EGA under the accession number EGAS00001001821, under a controlled access regimen to protect the privacy of the patients who have donated the samples.

References

1. Byrd, J. C., Stilgenbauer, S. & Flinn, I. W. Chronic lymphocytic leukemia. *Hematology Am. Soc. Hematol. Educ. Program* **2004,** 163–183 (2004).
2. Zenz, T., Mertens, D., Kuppers, R., Dohner, H. & Stilgenbauer, S. From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nat. Rev. Cancer* **10,** 37–50 (2010).
3. Damle, R. N. *et al.* Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* **94,** 1840–1847 (1999).
4. Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G. & Stevenson, F. K. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* **94,** 1848–1854 (1999).
5. Tobin, G. *et al.* Somatically mutated Ig V(H)3-21 genes characterize a new subset of chronic lymphocytic leukemia. *Blood* **99,** 2262–2264 (2002).
6. Agathangelidis, A. *et al.* Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. *Blood* **119,** 4467–4475 (2012).
7. Rossi, D. *et al.* Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to Richter syndrome. *Clin. Cancer Res.* **15,** 4415–4422 (2009).
8. Di Giovanni, S., Valentini, G., Carducci, P. & Giallonardo, P. Beta-2-microglobulin is a reliable tumor marker in chronic lymphocytic leukemia. *Acta Haematol.* **81,** 181–185 (1989).
9. Hallek, M. *et al.* Elevated serum thymidine kinase levels identify a subgroup at high risk of disease progression in early, nonsmoldering chronic lymphocytic leukemia. *Blood* **93,** 1732–1737 (1999).
10. Dohner, H. *et al.* Genomic aberrations and survival in chronic lymphocytic leukemia. *N. Engl. J. Med.* **343,** 1910–1916 (2000).
11. Rossi, D. *et al.* Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood* **121,** 1403–1412 (2013).
12. Baliakas, P. *et al.* Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia* **29,** 329–336 (2015).
13. Oscier, D. G. *et al.* The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. *Blood* **121,** 468–475 (2013).
14. Stilgenbauer, S. *et al.* Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. *Blood* **123,** 3247–3254 (2014).
15. Crespo, M. *et al.* ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *N. Engl. J. Med.* **348,** 1764–1775 (2003).
16. Ferreira, P. G. *et al.* Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Res.* **24,** 212–226 (2014).
17. Landau, D. A. *et al.* Mutations driving CLL and their evolution in progression and relapse. *Nature* **526,** 525–530 (2015).
18. Puente, X. S. *et al.* Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* **526,** 519–524 (2015).
19. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
20. Kulis, M. *et al.* Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.* **44,** 1236–1242 (2012).
21. Landau, D. A. *et al.* Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* **26,** 813–825 (2014).
22. Oakes, C. C. *et al.* Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia. *Cancer Discov.* **4,** 348–361 (2014).
23. Queiros, A. C. *et al.* A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia* **29,** 598–605 (2015).
24. Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome—biological and translational implications. *Nat. Rev. Cancer* **11,** 726–734 (2011).
25. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10,** 1213–1218 (2013).
26. Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* **12,** 963–965 (2015).
27. Risca, V. I. & Greenleaf, W. J. Unraveling the 3D genome: genomics tools for multiscale exploration. *Trends Genet.* **31,** 357–372 (2015).
28. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).
29. Stevenson, F. K., Krysov, S., Davies, A. J., Steele, A. J. & Packham, G. B-cell receptor signaling in chronic lymphocytic leukemia. *Blood* **118,** 4313–4320 (2011).
30. Ecker, S., Pancaldi, V., Rico, D. & Valencia, A. Higher gene expression variability in the more aggressive subtype of chronic lymphocytic leukemia. *Genome Med.* **7,** 8 (2015).
31. Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32,** 587–589 (2016).
32. Morabito, F. *et al.* Surrogate molecular markers for IGHV mutational status in chronic lymphocytic leukemia for predicting time to first treatment. *Leuk. Res.* **39,** 840–845 (2015).
33. Nikitin, E. A. *et al.* Expression level of lipoprotein lipase and dystrophin genes predict survival in B-cell chronic lymphocytic leukemia. *Leuk. Lymphoma* **48,** 912–922 (2007).
34. Oppezzo, P. *et al.* The LPL/ADAM29 expression ratio is a novel prognosis indicator in chronic lymphocytic leukemia. *Blood* **106,** 650–657 (2005).
35. Hock, B. D. *et al.* Release and clinical significance of soluble CD83 in chronic lymphocytic leukemia. *Leuk. Res.* **33,** 1089–1095 (2009).

36. Malavasi, F. *et al.* CD38 and chronic lymphocytic leukemia: a decade later. *Blood* **118,** 3470–3478 (2011).
37. Forconi, F. *et al.* The normal IGHV1-69-derived B-cell repertoire contains stereotypic patterns characteristic of unmutated CLL. *Blood* **115,** 71–77 (2010).
38. Mathelier, A. *et al.* JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44,** D110–D115 (2016).
39. Alizadeh, A. A. *et al.* Toward understanding and exploiting tumor heterogeneity. *Nat. Med.* **21,** 846–853 (2015).
40. Sung, M. H., Baek, S. & Hager, G. L. Genome-wide footprinting: ready for prime time? *Nat. Methods* **13,** 222–228 (2016).
41. Rosati, E. *et al.* Constitutively activated Notch signaling is involved in survival and apoptosis resistance of B-CLL cells. *Blood* **113,** 856–865 (2009).
42. Tomic, J., Lichty, B. & Spaner, D. E. Aberrant interferon-signaling is associated with aggressive chronic lymphocytic leukemia. *Blood* **117,** 2668–2680 (2011).
43. Mittal, A. K. *et al.* Role of CTLA4 in the proliferation and survival of chronic lymphocytic leukemia. *PLoS ONE* **8,** e70352 (2013).
44. Klein, U. *et al.* Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J. Exp. Med.* **194,** 1625–1638 (2001).
45. Rosenwald, A. *et al.* Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J. Exp. Med.* **194,** 1639–1647 (2001).
46. Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37,** 382–390 (2005).
47. Lefebvre, C. *et al.* A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* **6,** 377 (2010).
48. Yepes, S., Torres, M. M. & Lopez-Kleine, L. Regulatory network reconstruction reveals genes with prognostic value for chronic lymphocytic leukemia. *BMC Genomics* **16,** 1002 (2015).
49. Bock, C. *et al.* Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144,** 439–452 (2011).
50. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523,** 486–490 (2015).
51. Jin, W. *et al.* Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* **528,** 142–146 (2015).
52. Jiang, H., Lei, R., Ding, S. W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15,** 182 (2014).
53. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
54. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31,** 2032–2034 (2015).
55. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
56. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
57. Hoffman, M. M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res.* **41,** 827–841 (2013).
58. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
59. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
60. Glaus, P., Honkela, A. & Rattray, M. Identifying differentially expressed transcripts from RNA-seq data with biological variation. *Bioinformatics* **28,** 1721–1728 (2012).
61. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
62. Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.* **33,** 364–376 (2015).
63. Rosenbloom, K. R. *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43,** D670–D681 (2015).
64. Pedregosa, F. *et al.* Scikit-learn: machine learning in Python. *J. Machine Learn. Res.* **12,** 2825–2830 (2011).
65. Sheffield, N. C. *et al.* Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res.* **23,** 777–788 (2013).
66. Sanchez-Castillo, M. *et al.* CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic Acids Res.* **43,** D1117–D1123 (2015).
67. Liu, T. *et al.* Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.* **12,** R83 (2011).
68. Adams, D. *et al.* BLUEPRINT to decode the epigenetic signature written in blood. *Nat. Biotechnol.* **30,** 224–226 (2012).
69. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37,** W202–W208 (2009).
70. Wang, B., Cunningham, J. M. & Yang, X. H. Seq2pathway: an R/Bioconductor package for pathway analysis of next-generation sequencing data. *Bioinformatics* **31,** 3043–3045 (2015).
71. Sherwood, R. I. *et al.* Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32,** 171–178 (2014).
72. Qu, K. *et al.* Individuality and variation of personal regulomes in primary human T cells. *Cell Syst.* **1,** 51–61 (2015).
73. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference pp.* 11–15 (2008).

Acknowledgements

We thank all patients who have donated their samples for this study. We also thank Amelie Kuchler and Thomas Penz for expert technical assistance, Johanna Klughammer and Nathan Sheffield for their contributions to the data analysis pipeline, the Biomedical Sequencing Facility at CeMM for assistance with next-generation sequencing and all members of the Bock lab for their help and advice. Moreover, we thank Ulrich Jäger, Medhat Shehata, Philipp Staber and Jörg Menche for their comments and for critically reading the manuscript. This work was performed in the context of the BLUEPRINT project (European Union's Seventh Framework Programme grant agreement number 282510) and the ERA-NET projects EpiMark (FWF grant agreement number I 1575-B19) and CINOCA (FWF grant agreement number I 1626-B22). C.S. was supported by a Feodor Lynen Fellowship of the Alexander von Humboldt Foundation. J.C.S. was supported by Bloodwise (11052 and 12036), the Kay Kendall Leukaemia Fund (873), Cancer Research UK (C34999/A18087 and ECMC C24563/A15581), Wessex Medical Research and the Bournemouth Leukaemia Fund. C.B. was supported by a New Frontiers Group award of the Austrian Academy of Sciences.

Author contributions

A.F.R., C.S., J.C.S., D.O. and C.B. planned the study. J.C.S., R.W., Z.D. and D.O. provided samples and clinical data. C.S. and M.F. performed the experiments. A.F.R. analysed the data with contributions from C.S., J.C.S., D.O. and C.B. C.B. supervised the research. All authors contributed to the writing of the manuscript.

Additional information

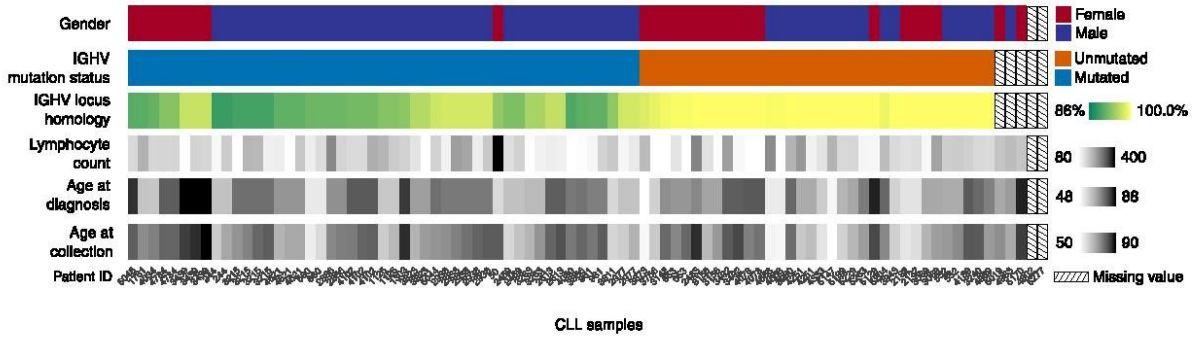
Accession codes: The processed high-throughput sequencing data are openly available from NCBI GEO under the accession number GSE81274, whereas the raw sequencing data are available from EBI EGA under the accession number EGAS00001001821, under a controlled access regimen to protect the privacy of the patients who have donated the samples.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

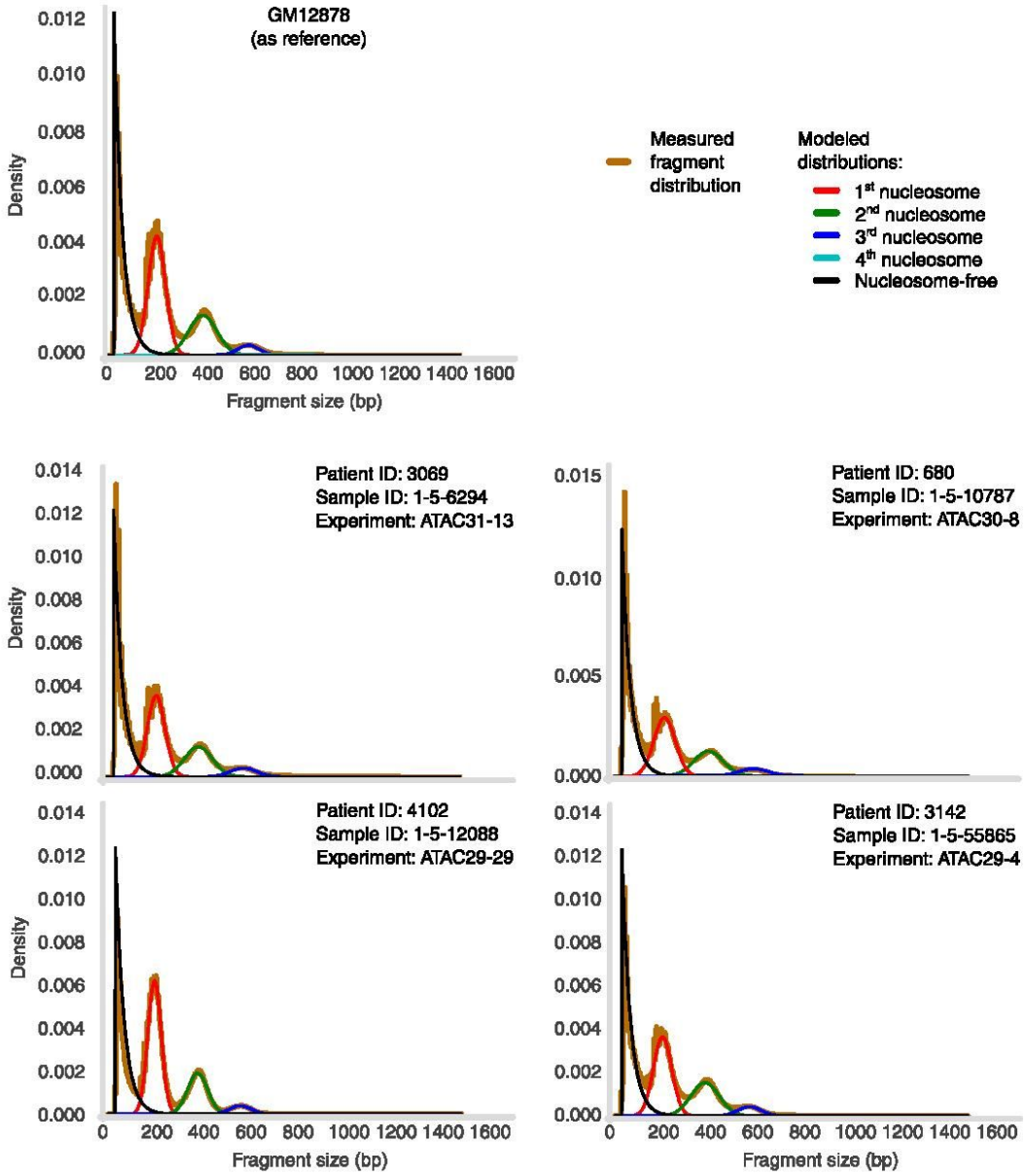
How to cite this article: Rendeiro, A. F. *et al.* Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. *Nat. Commun.* 7:11938 doi: 10.1038/ncomms11938 (2016).



Supplementary Figure 1

The cohort reflects the spectrum of CLL phenotypes commonly encountered in clinical care.

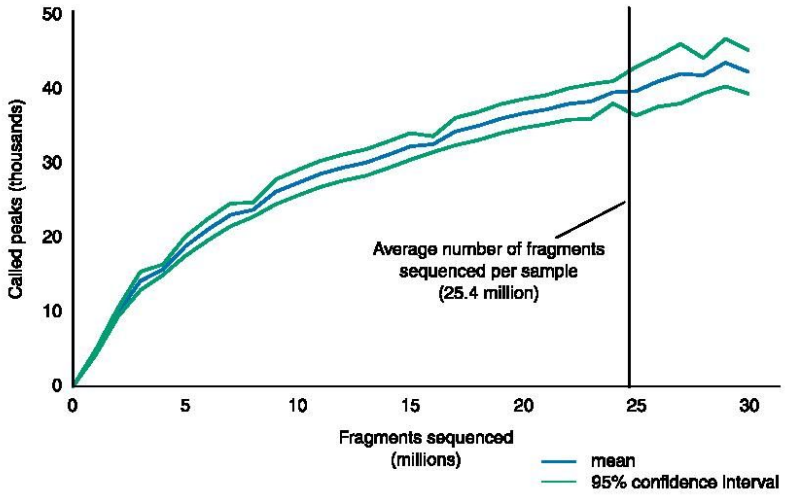
Visualization of clinical annotations for the patient samples included in this study.



Supplementary Figure 2

Observed ATAC-seq fragment length distributions indicate high data quality.

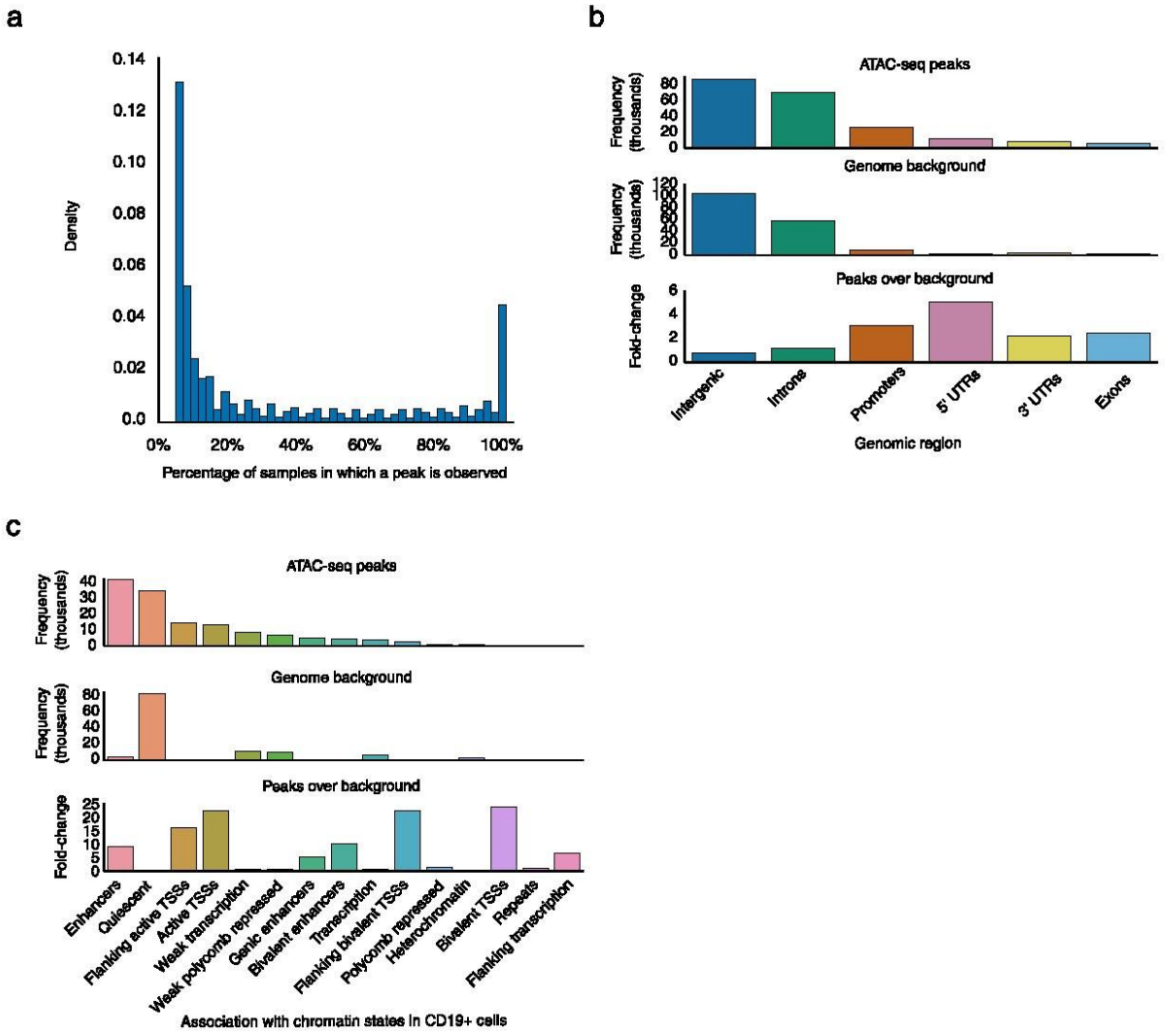
Distribution of ATAC-seq fragment lengths for published GM12878 data (Buenrostro et al. 2013 Nature Methods) and for four randomly selected CLL samples from this study. Fragment lengths were inferred based on paired-end sequencing data. The characteristic patterns of nucleosome-associated fragment length are observed in all samples.



Supplementary Figure 3

The chosen sequencing depth recovers the majority of ATAC peaks per sample.

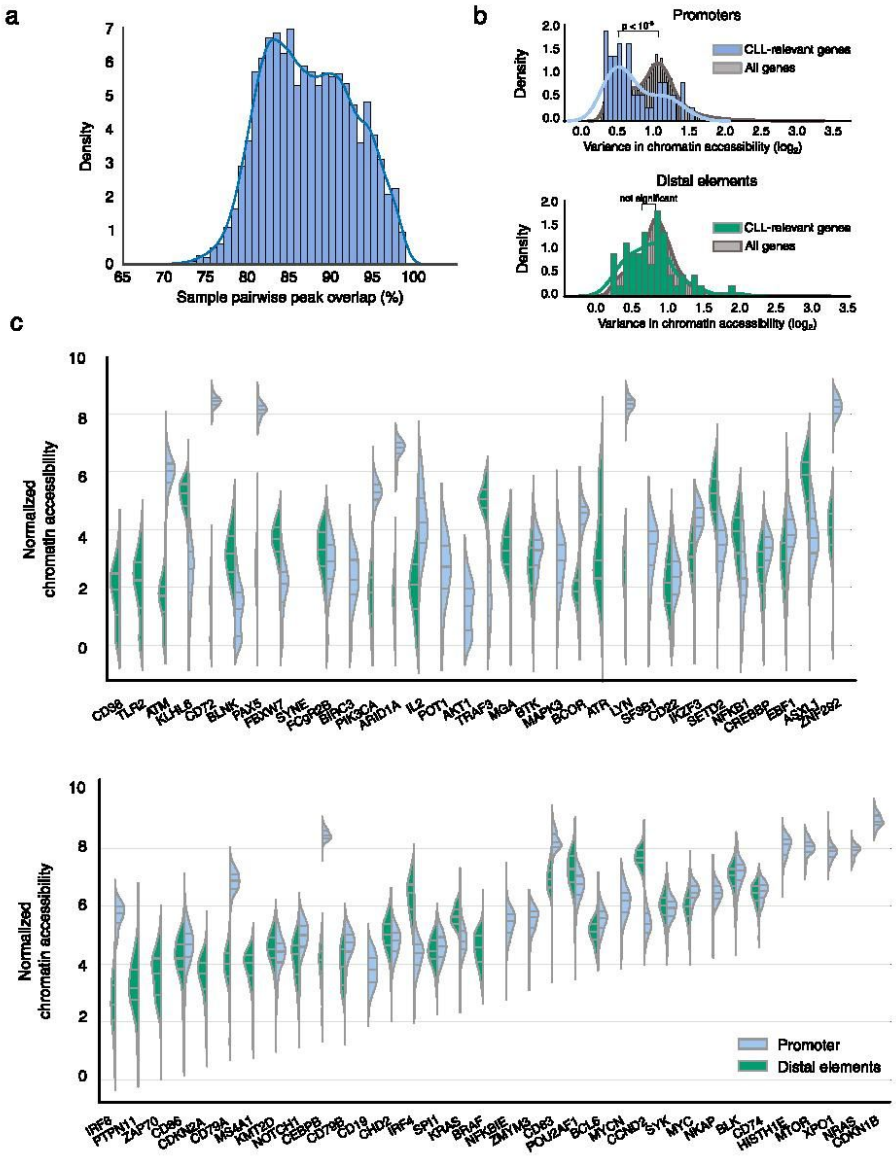
Relationship of the number of sequenced reads (x-axis) and the number of detected chromatin-accessible regions (y-axis), showing the average pattern across all 88 samples (blue line). The corridor indicated in green corresponds to a 95% confidence interval for random subsampling across samples.



Supplementary Figure 4

Chromatin-accessible regions in CLL are enriched for promoters and enhancers.

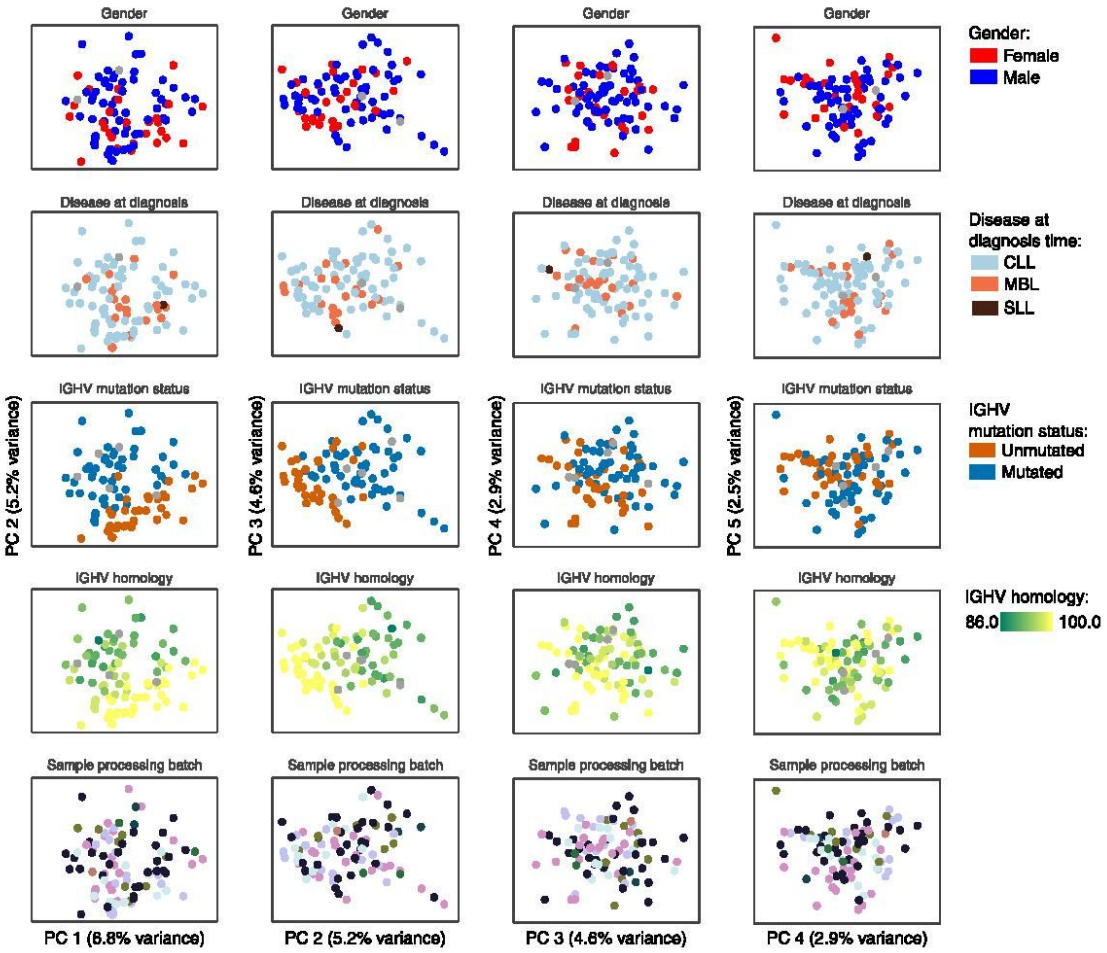
a) Histogram showing the number of samples in which a given chromatin-accessible region from the CLL consensus map was detected as a significant ATAC-seq peak. b) Frequency of overlap and enrichment of Ensembl gene annotation for regions in the CLL consensus map, compared to region sets of identical size and lengths that were randomized 1,000 times across the genome. c) Frequency of overlap and enrichment of chromatin state segmentations for CD19+ B cells (data from the Roadmap Epigenomics project), compared to region sets of identical size and lengths that were randomized 1,000 times across the genome.



Supplementary Figure 5

Heterogeneity in chromatin accessibility affects genes related to B cells and CLL.

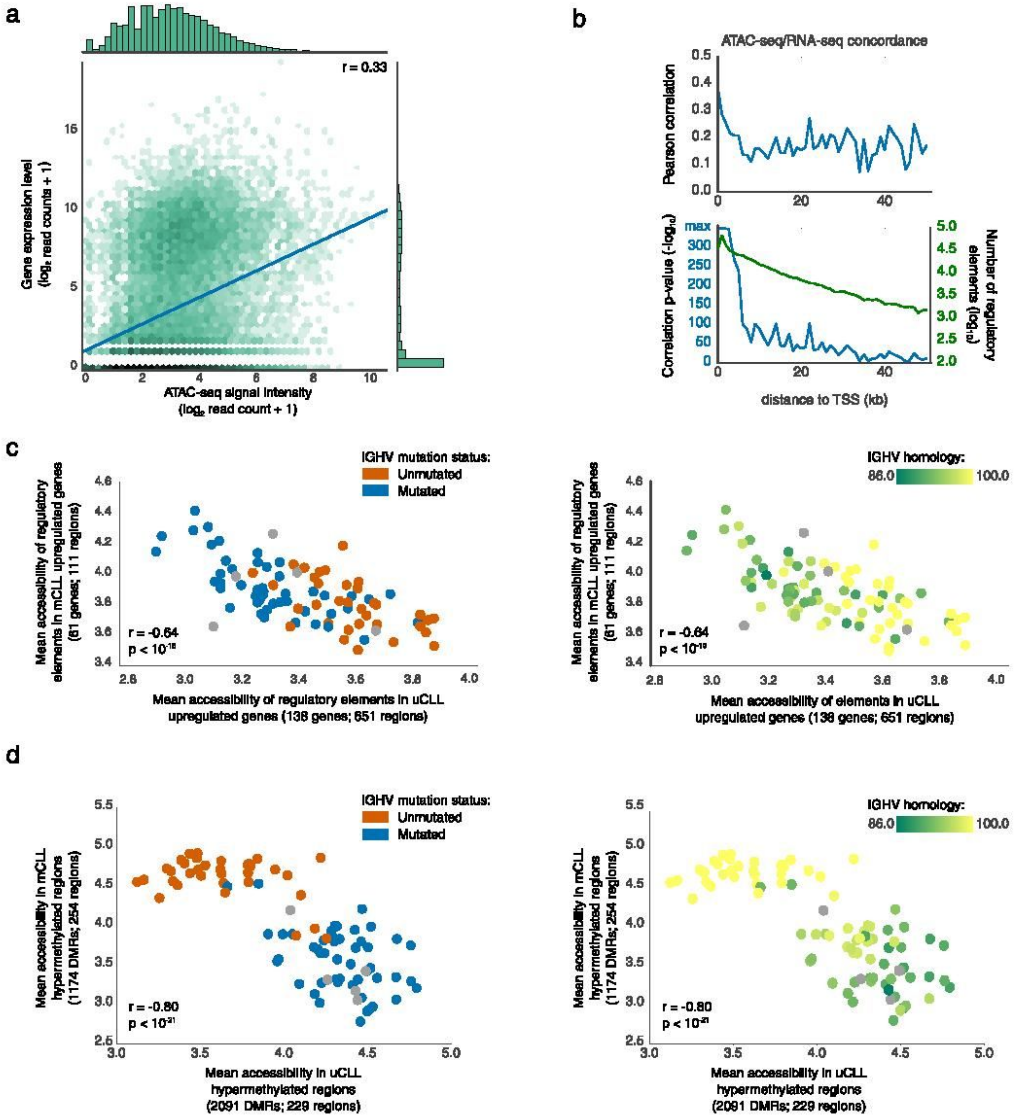
a) Histogram showing the percentage of chromatin-accessible regions that are shared between any two CLL samples. b) Distribution of variance in chromatin accessibility for promoter regions and putative distal regulatory regions across all genes (grey) and for a set of genes with a known role in B cell biology and/or CLL pathogenesis (blue/green). Chromatin accessibility scores were averaged across all regulatory regions assigned to a given gene. c) Violin plots of normalized chromatin accessibility values for gene promoters (regions located within 2,500 basepairs of the transcription start site) and distal regulatory elements (regions located at least 2,500 basepairs away from the nearest transcription start site) for the same genes as in panel b.



Supplementary Figure 6

Unsupervised analysis identifies IGHV mutation status as a key source of variation.

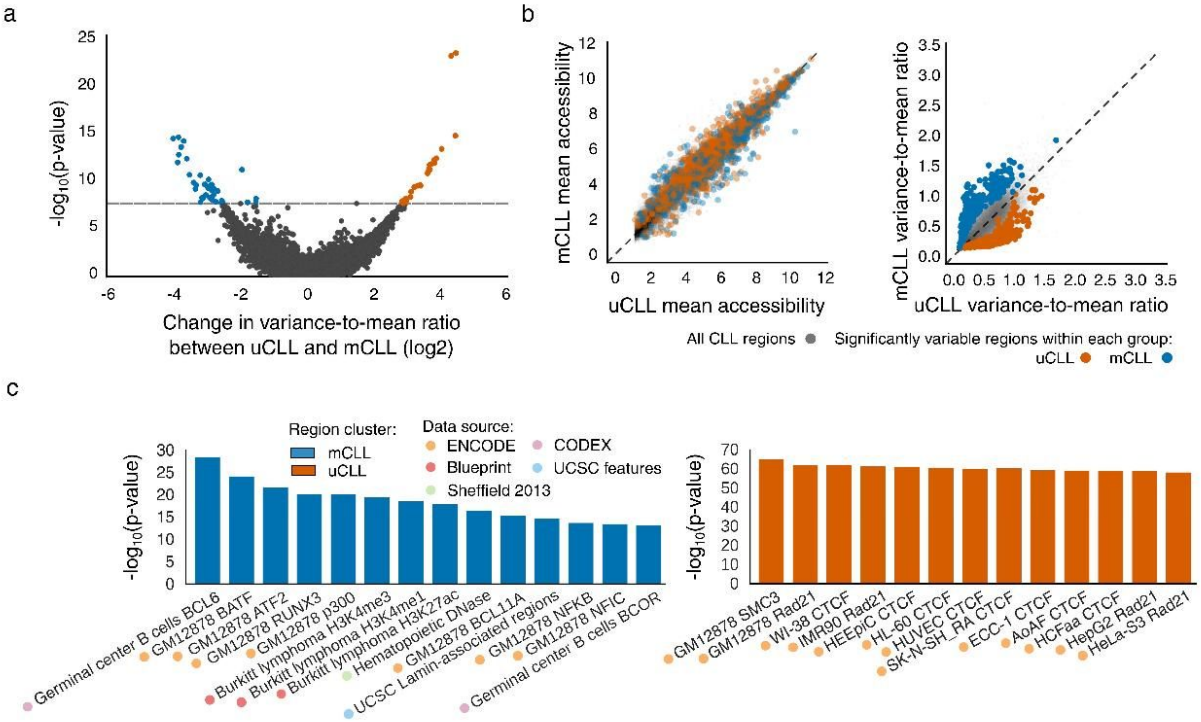
Principal component analysis based on the chromatin accessibility for all 88 samples at each of the 112,298 chromatin-accessible regions in the CLL cohort. The first five principal components are plotted, and samples are colored according to clinical annotations and molecular diagnostics data (top four rows) as well as the sample processing batch for the ATAC-seq experiments (bottom row).



Supplementary Figure 7

Chromatin accessibility is linked to differences in gene expression and DNA methylation.

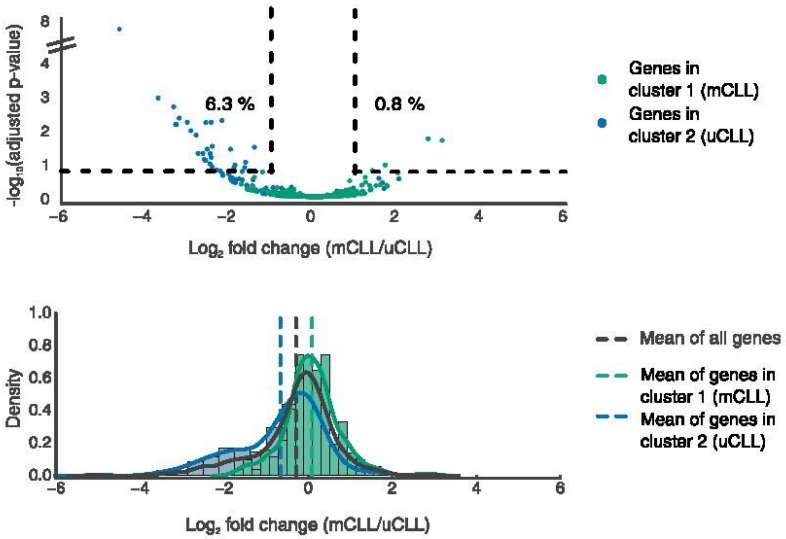
a) Hexbin scatterplot visualizing the weak correlation (Pearson's $r = 0.33$) between gene expression levels and the chromatin accessibility at associated regulatory regions. Shown are averages across ten samples with matched ATAC-seq and RNA-seq data. The color gradient is on a logarithmic scale. b) Pearson correlation (top) and significance of the association (bottom) between gene expression levels and chromatin accessibility values at associated regulatory regions, plotted over the distance of the accessible region to the gene's transcription start site. c) Mean chromatin accessibility across CLL-accessible regions associated with genes that were upregulated in *IGHV*-mutated or in *IGHV*-unmutated CLL. d) Mean chromatin accessibility in CLL-accessible regions that overlap with regions described as hypermethylated in *IGHV*-mutated or in *IGHV*-unmutated CLL (Kulis et al. 2012 Nature Genetics).



Supplementary Figure 8

Subtype-specific variable regions show characteristic enrichment patterns.

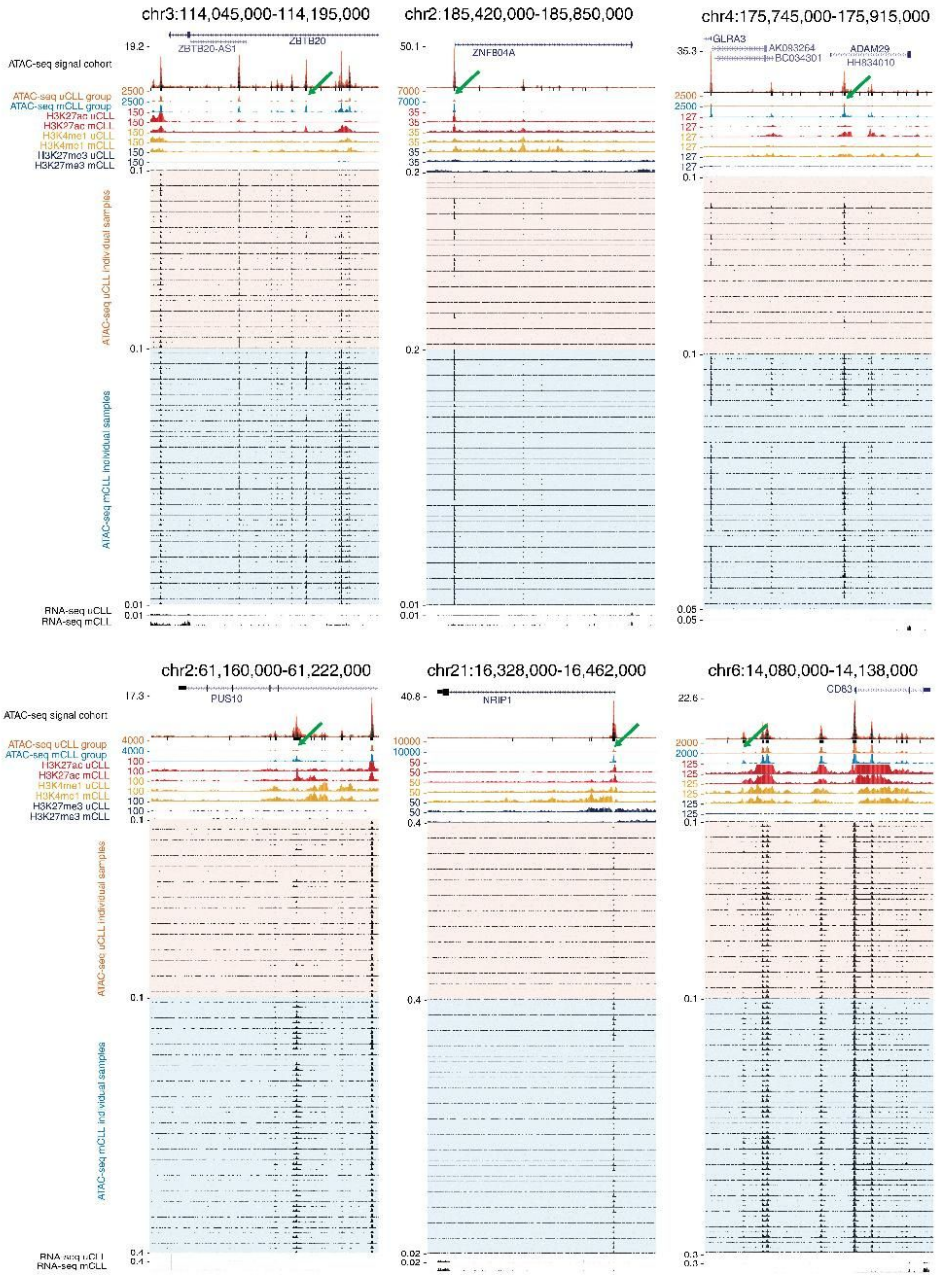
a) Differential variability between the mCLL and uCLL sample groups illustrated by each region's change in variance-to-mean ratio between groups (x-axis) and the *p*-value for variability within each group (y-axis). Blue and orange dots indicate significantly variable regions. b) Scatterplots of mean accessibility (left) and variance-to-mean ratio within each sample group (right). The plot on the left illustrates how significantly variable regions are dispersed across the accessibility range, rather than being strongly associated with differences in mean accessibility between the groups. The color coding is the same as in panel a. c) Most highly enriched region sets that significantly overlap with the differentially variable regions for mCLL (blue) and for uCLL (orange), based on LOLA analysis.



Supplementary Figure 10

Subtype-specific signature regions are weakly associated with differentially expressed genes.

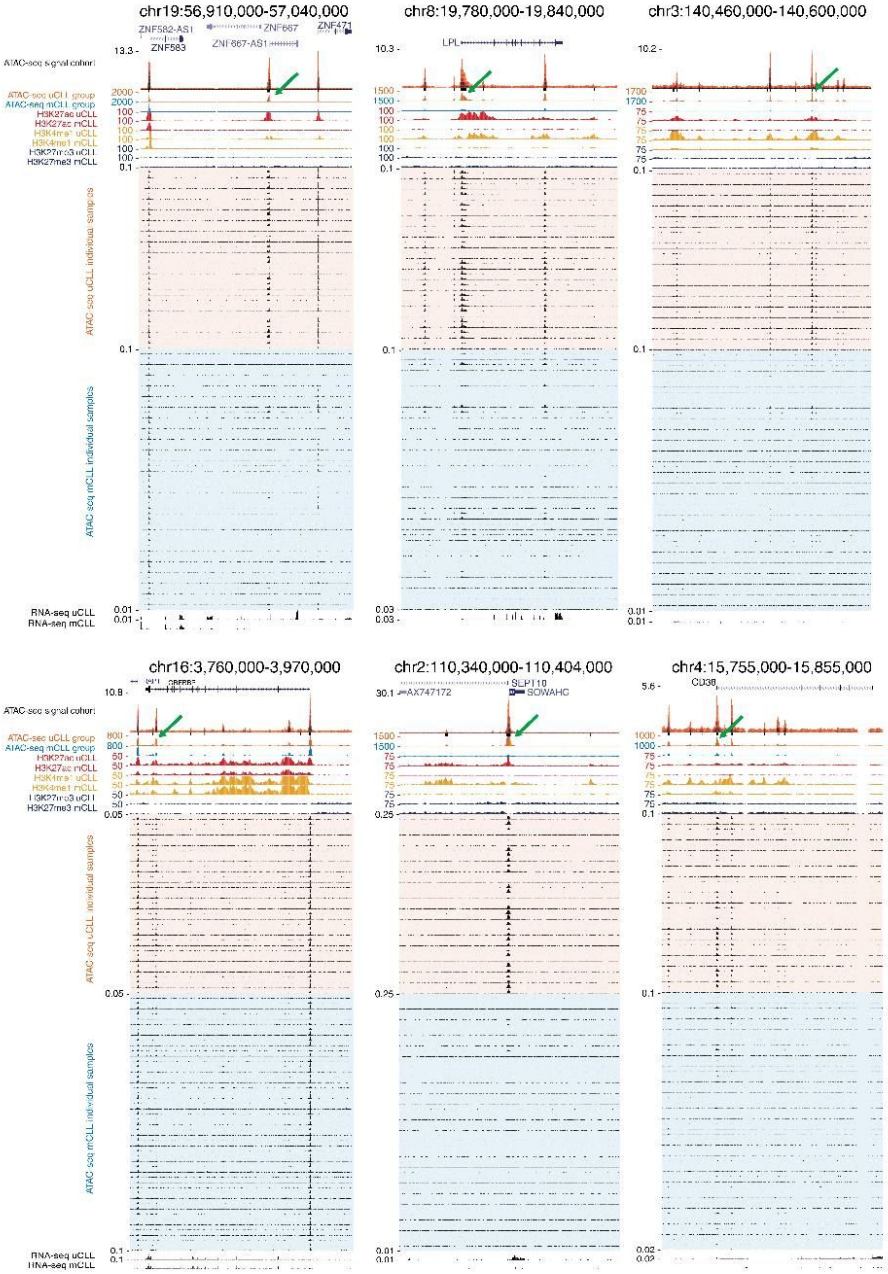
Volcano plot (top) and histogram (bottom) showing gene expression differences between mCLL and uCLL samples for genes that are co-localized with subtype-specific signature regions. Percentage values are based on the number of genes that were significantly differentially expressed in the RNA-seq analysis.



Supplementary Figure 11

Signature regions specific to mCLL show strong differences between disease subtypes but also heterogeneity within each subtype.

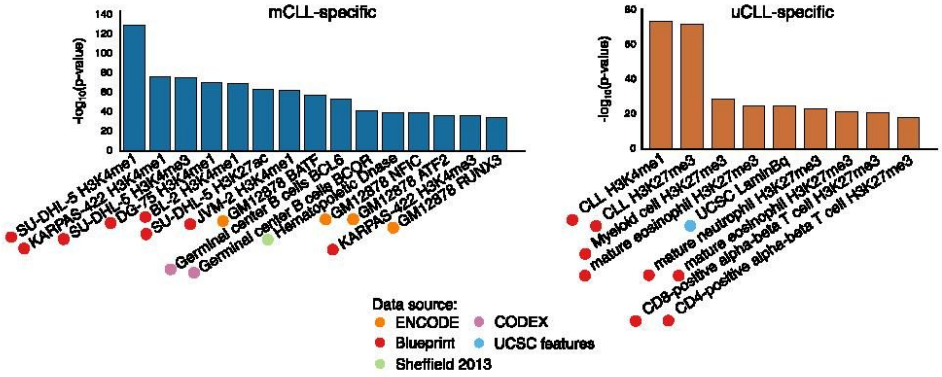
Genome browser plots for six gene loci that contain mCLL-specific signature regions (indicated by the green arrows). All ATAC-seq tracks were normalized by read depth to improve comparability between samples.



Supplementary Figure 12

Signature regions specific to uCLL show strong differences between disease subtypes but also heterogeneity within each subtype.

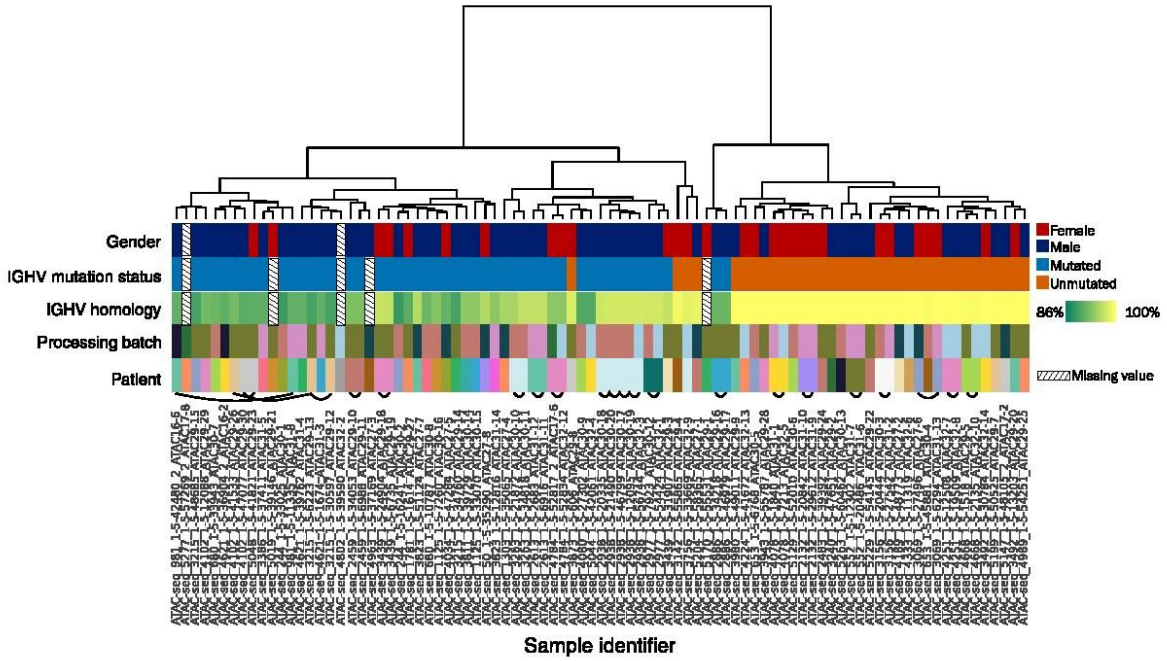
Genome browser plots for six gene loci that contain uCLL-specific signature regions (indicated by the green arrows). All ATAC-seq tracks were normalized by read depth to improve comparability between samples.



Supplementary Figure 13

Enrichment analysis for differential ATAC-seq peaks yields similar results as for the subtype-specific signature regions.

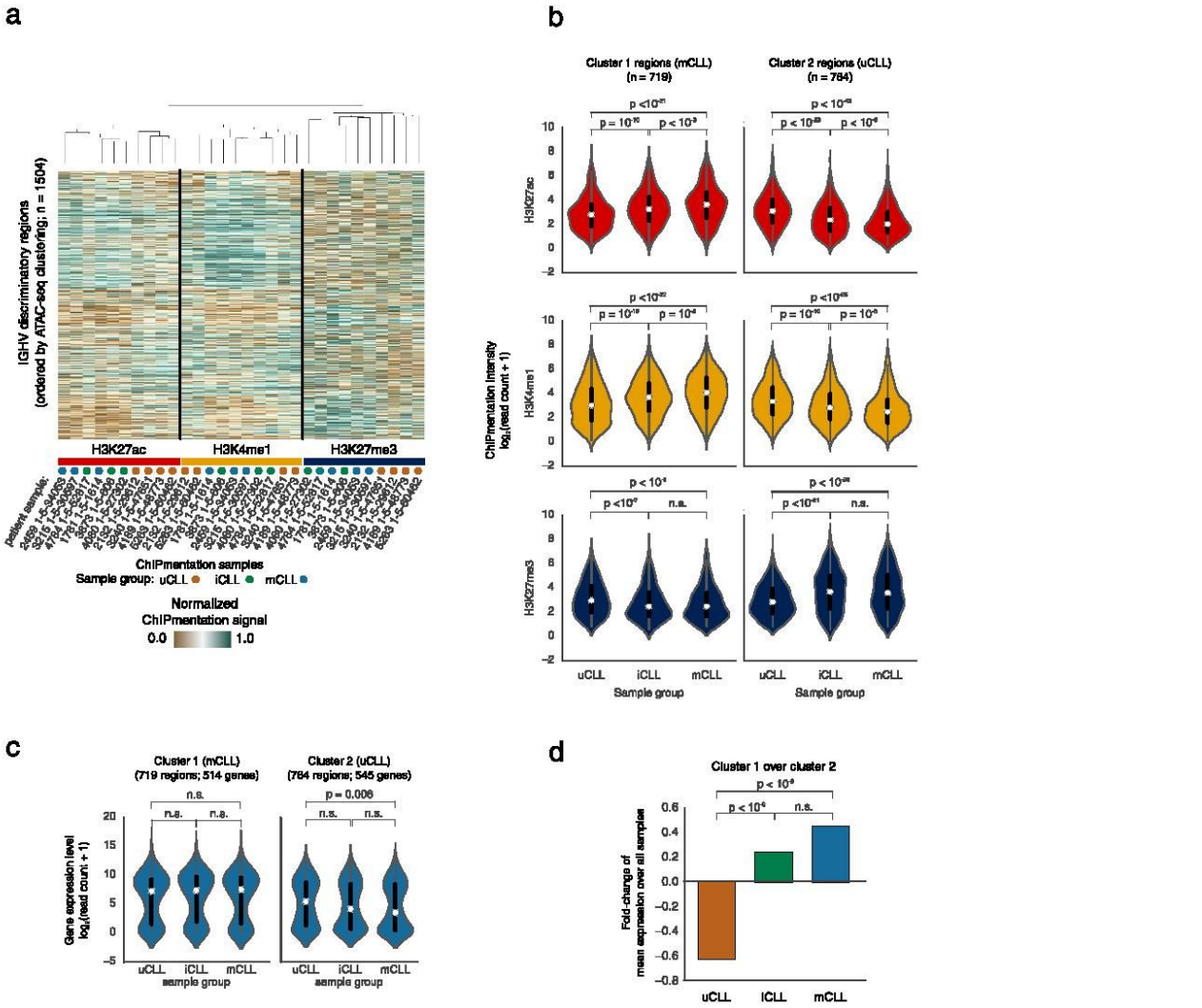
Complementing and validating the enrichment analysis shown in Figure 3g, this diagram lists the most highly enriched LOLA region sets for mCLL-specific (blue) and uCLL-specific (orange) differential peaks identified using DESeq2.



Supplementary Figure 14

Clustering based on CLL subtype-specific signature regions reflects IGHV mutation status.

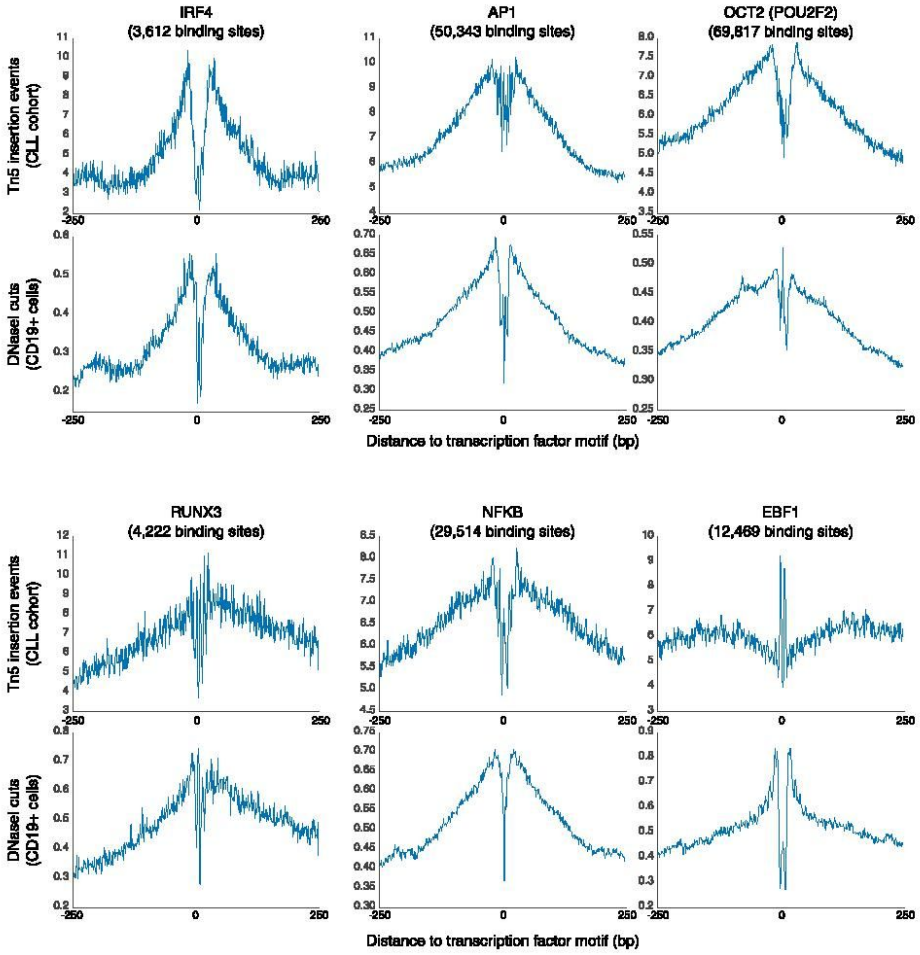
Hierarchical clustering of all CLL samples based on sample-wise correlation of chromatin accessibility for the most discriminatory regions that were identified between the *IGHV*-mutated and the *IGHV*-unmutated disease subtype. The clustering tree is annotated with clinical data, and samples from the same patient are connected by curved black lines.



Supplementary Figure 15

Histone marks and gene expression confirm the intermediate character of the iCLL sample cluster.

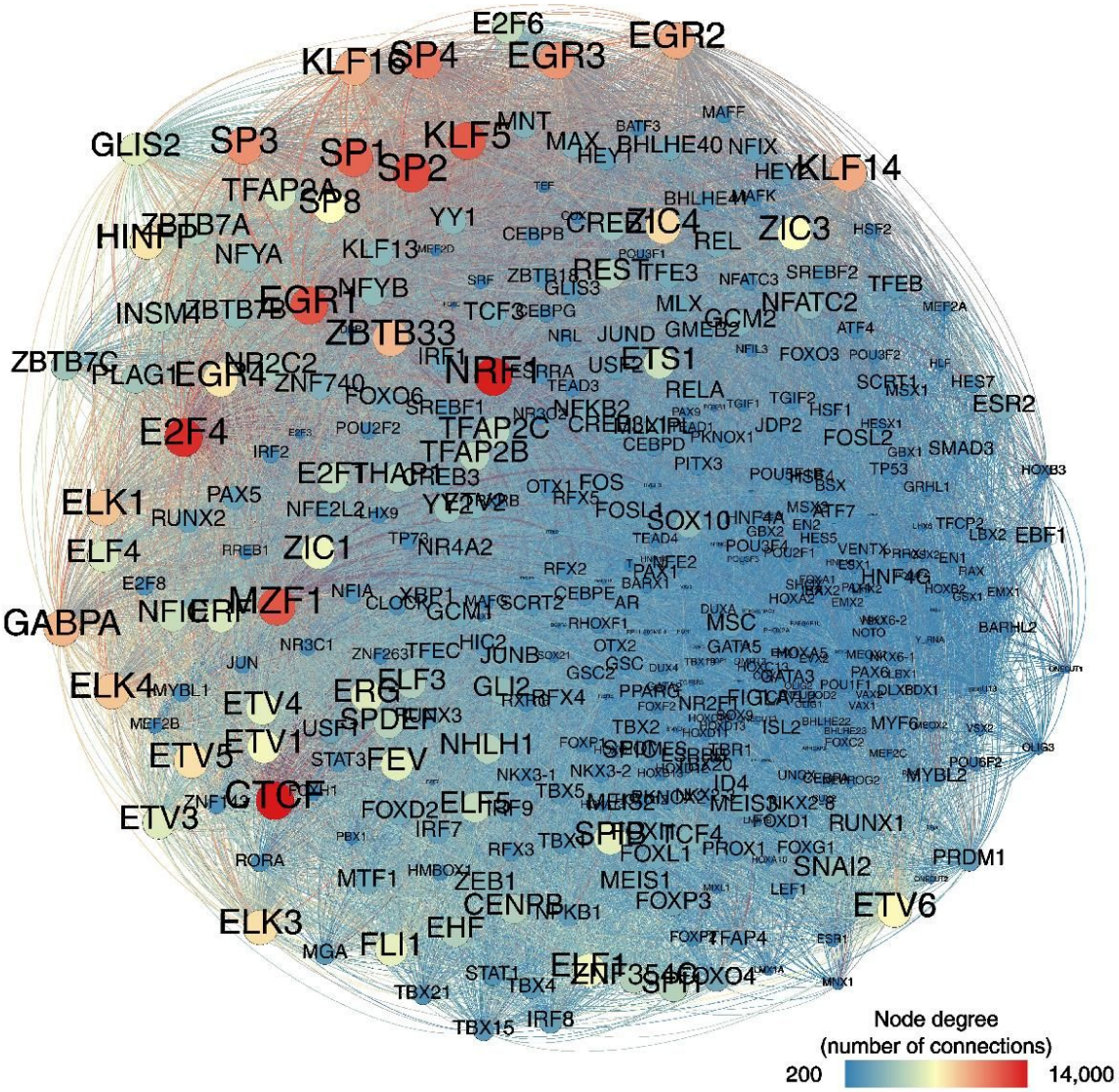
a) Hierarchical clustering and heatmap visualizing the ChIPmentation signal for three histone marks (H3K4me1, H3K27ac, H3K27me3) in ten CLL samples comprising three disease subtypes (mCLL, iCLL, uCLL). Regulatory regions were selected and sorted in the same way as in Figure 3c. b) Violin plots showing the distribution of ChIPmentation levels for each histone mark in the same regulatory regions as in panel a, grouped by disease subtype. In all panels, significance was assessed using the Mann-Whitney *U* test, and comparisons with *p*-values above 0.05 were labeled as not significant (n.s.). c) Mean gene expression values for genes associated with the regulatory regions from panel a, grouped by disease subtype. d) Barplot showing the mean fold change of genes associated with regulatory elements in cluster 1 (mCLL regions) over genes associated with cluster 2 (uCLL regions) across all genes, grouped by disease subtypes. Significance was assessed using the Mann-Whitney *U* test, and comparisons with *p*-values above 0.05 were labeled as not significant (n.s.).



Supplementary Figure 16

Transcription factor footprints for ATAC-seq and DNase-seq are similar.

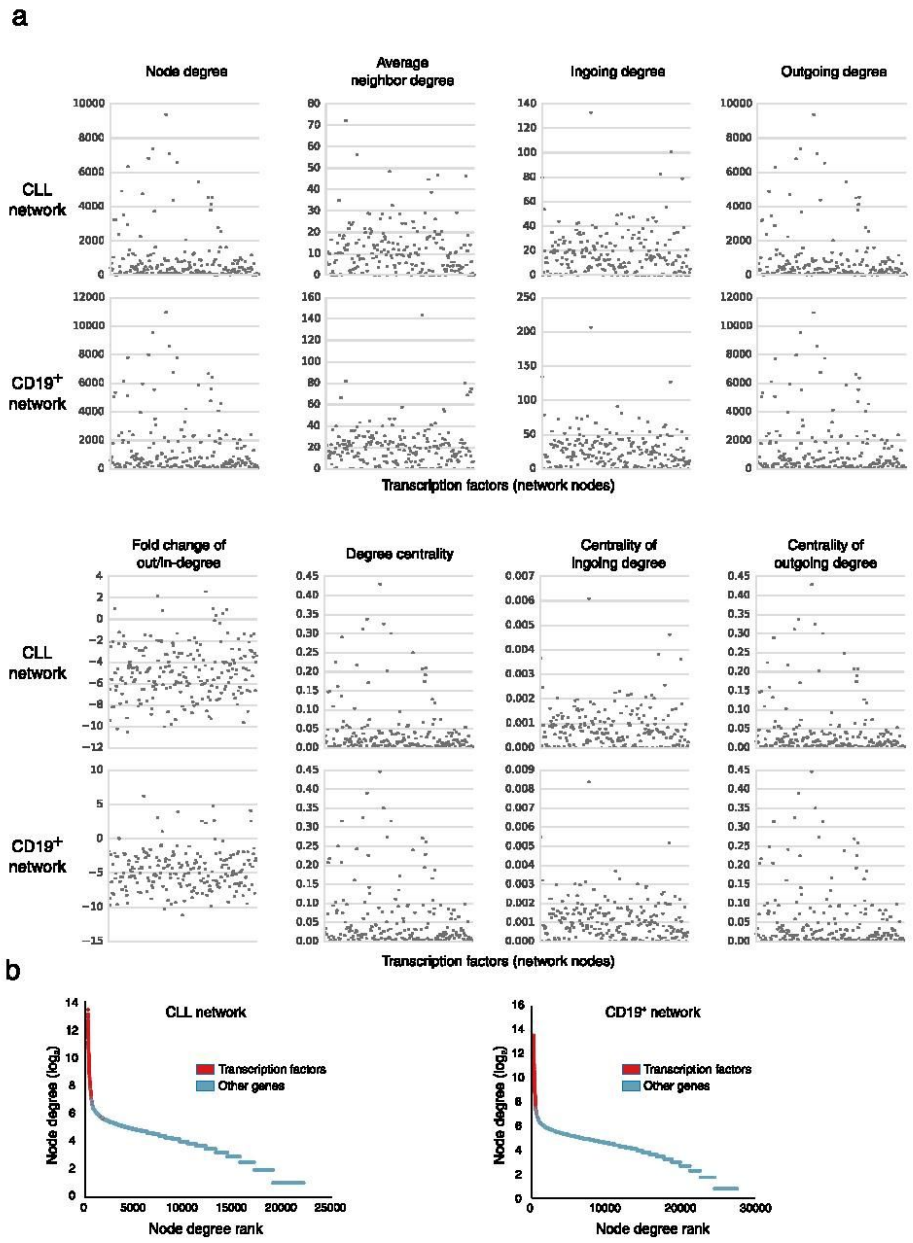
Footprinting diagrams showing the frequency of Tn5 transposase insertion events (for ATAC-seq) and DNase I cutting sites (for DNase-seq, based on data for CD19+ B cells from the Roadmap Epigenomics project) across a 500 basepair window around DNA binding motifs of transcription factors involved in B cell development.



Supplementary Figure 17

Cohort-level gene regulatory network identifies transcription factors relevant to CLL.

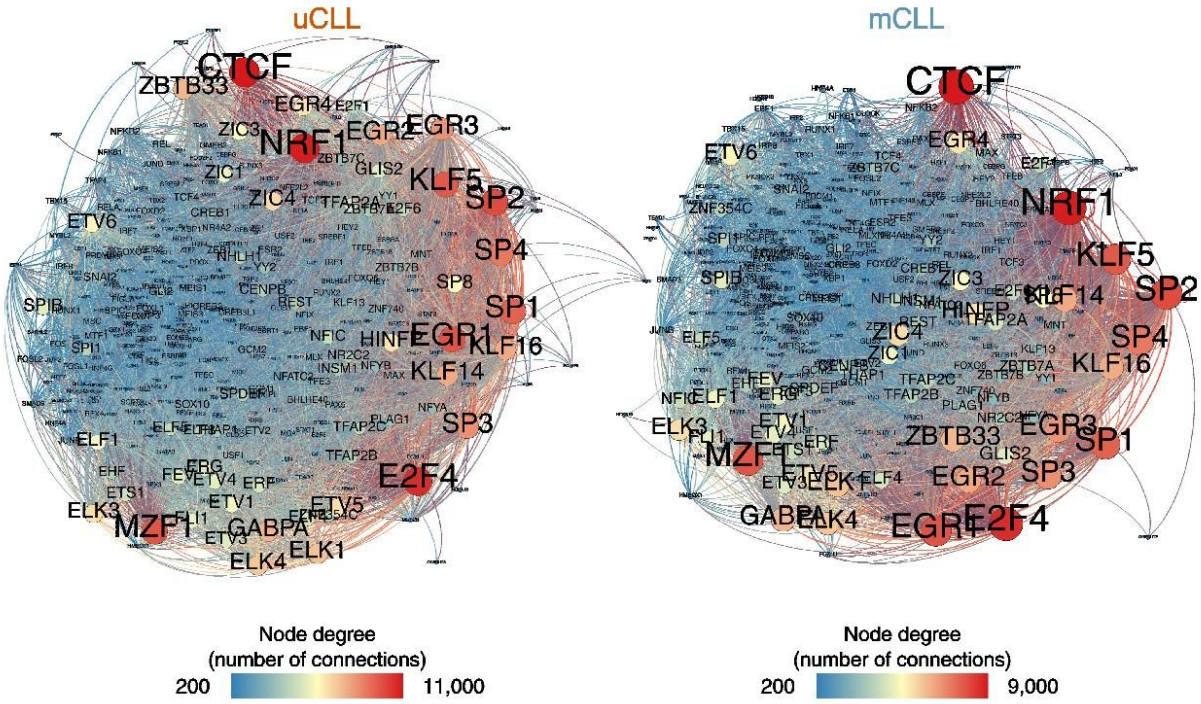
Gene regulatory network of CLL inferred from footprint predictions of transcription factor binding, based on the ATAC-seq data of all CLL samples. Only nodes with more than 200 connections are shown.



Supplementary Figure 18

Footprinting-based gene regulatory networks for ATAC-seq in CLL and DNase-seq in B cells show similar properties.

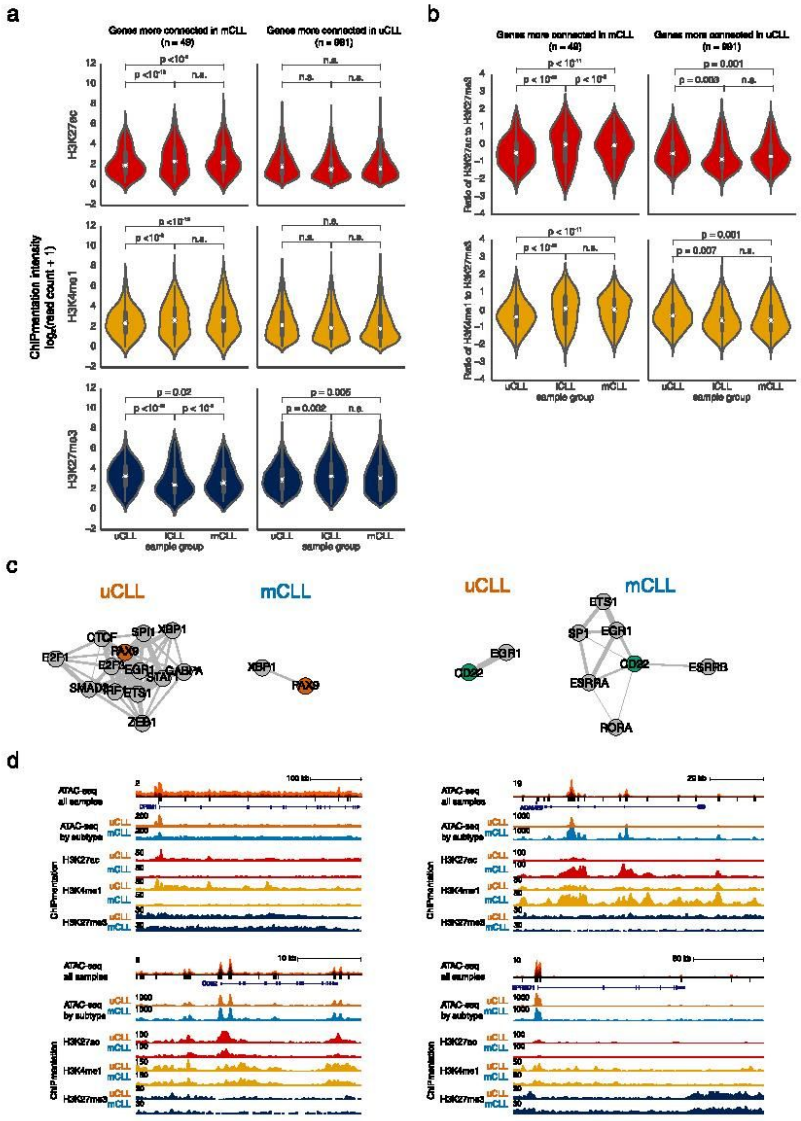
a) Structural properties of gene regulatory networks inferred from ATAC-seq data for the CLL cohort and from DNase-seq data for CD19+ B cells. b) Number of connections for all genes in the two gene regulatory networks (transcription factors are shown in red).



Supplementary Figure 19

Gene regulatory networks for mCLL and uCLL samples are globally similar.

Gene regulatory networks inferred based on the *IGHV*-unmutated samples (uCLL, left) and based on the *IGHV*-mutated samples (mCLL, right). Only nodes with more than 200 connections are shown.



Supplementary Figure 20

Disease subtype-specific networks detect differentially regulated genes and genomic regions.

a) Violin plots showing the distribution of ChIPmentation levels for each histone mark in regulatory regions associated with genes that are differentially connected between the subtype-specific networks. b) Violin plots showing the ratio between the ChIPmentation signal for histone marks associated with active (H3K4me1, H3K27ac) over repressed (H3K27me3) chromatin. c) Subnetworks with the neighbors of PAX9 and CD22, shown separately for the mCLL and uCLL networks. Edge width indicates the strength of the connection as measured by the calculated interaction score. d) ATAC-seq and ChIPmentation signal for three histone marks at representative differentially connected genes between the mCLL and uCLL networks. In panel a and b, significance was assessed using the Mann-Whitney *U* test, and comparisons with *p*-values above 0.05 were labeled as not significant (n.s.).

Manuscript #2

The following section contains the manuscript entitled “Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib drug response in chronic lymphocytic leukemia” which is available as a preprint in bioRxiv:

André F. Rendeiro, Thomas Krausgruber, Nikolaus Fortelny, Fangwen Zhao, Thomas Penz, Matthias Farlik, Linda C. Schuster, Amelie Nemc, Szabolcs Tasnády, Marienn Réti, Zoltán Mátrai, Donat Alpar, Csaba Bödör, Christian Schmidl, Christoph Bock. *Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib drug response in chronic lymphocytic leukemia*. **Nature Communications** (2020). doi:10.1038/s41467-019-14081-6

Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib drug response in chronic lymphocytic leukemia

André F. Rendeiro[1*], Thomas Krausgruber[1*], Nikolaus Fortelny[1], Fangwen Zhao[2], Thomas Penz[1], Matthias Farlik[1], Linda C. Schuster[1], Amelie Nemc[1], Szabolcs Tasnády[3], Marienn Réti[3], Zoltán Mátrai[3], Donat Alpar[1,4†], Csaba Bödör[4†], Christian Schmidl[1,7†], Christoph Bock[1,2,5,6†]

[1] CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria
[2] Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, Vienna, Austria
[3] Department of Haematology and Stem Cell Transplantation, Central Hospital of Southern Pest, National Institute of Hematology and Infectious Diseases, Budapest, Hungary
[4] MTA-SE Lendület Molecular Oncohematology Research Group, 1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary
[5] Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria
[6] Max Planck Institute for Informatics, Saarland Informatics Campus, Saarbrücken, Germany
[7] Current address: Regensburg Centre for Interventional Immunology (RCI) and University Medical Center of Regensburg, Regensburg, Germany

[*] These authors contributed equally to this work
[†] Co-last author / These authors jointly directed this work
Correspondence: Christoph Bock (cbock@cemm.oeaw.ac.at)

Keywords: Chronic lymphocytic leukemia, drug response profiling, ibrutinib therapy, chromatin mapping, single-cell RNA sequencing, time series analysis, machine learning, translational bioinformatics

Abstract

Chronic lymphocytic leukemia (CLL) is a genetically, epigenetically, and clinically heterogeneous disease. Despite this heterogeneity, the Bruton tyrosine kinase (BTK) inhibitor ibrutinib provides effective treatment for the vast majority of CLL patients. To define the underlining regulatory program, we analyzed high-resolution time courses of ibrutinib treatment in closely monitored patients, combining cellular phenotyping (flow cytometry), single-cell transcriptome profiling (scRNA-seq), and chromatin mapping (ATAC-seq). We identified a consistent regulatory program shared across all patients, which was further validated by an independent CLL cohort. In CLL cells, this program starts with a sharp decrease of NF-κB binding, followed by reduced regulatory activity of lineage-defining transcription factors (including PAX5 and IRF4) and erosion of CLL cell identity, finally leading to the acquisition of a quiescence-like gene signature which was shared across several immune cell types. Nevertheless, we observed patient-to-patient variation in the speed of its execution, which we exploited to predict patient-specific dynamics in the response to ibrutinib based on pre-treatment samples. In aggregate, our study describes the cellular, molecular, and regulatory effects of therapeutic B cell receptor inhibition in CLL at high temporal resolution, and it establishes a broadly applicable method for epigenome/transcriptome-based treatment monitoring.

Introduction

Chronic lymphocytic leukemia (CLL) is among the most frequent blood cancers[1]. It is characterized by clonal proliferation and accumulation of malignant B lymphocytes in the blood, bone marrow, spleen, and lymph nodes. On a cellular level, this process is driven by constitutively activated B cell receptor (BCR) signaling, which can be caused by erroneous (auto)antigen recognition and/or cell-autonomous mechanisms[2]. CLL shows remarkable clinical heterogeneity, with some patients pursuing an indolent course, while others progress rapidly and require early treatment. Extensive heterogeneity exists also at the genetic, epigenetic, and transcriptional level and has led to the identification of genetically defined CLL subtypes[3-6] and patient-specific transcriptional programs[7-9]. Moreover, characteristic DNA methylation patterns appear to reflect differences in the CLL's cell-of-origin[10-12], and chromatin profiles predict the BCR immunoglobulin heavy-chain variable (IGHV) gene mutation status[13].

Despite widespread clinical and molecular heterogeneity, therapeutic inhibition of BCR signaling has shown remarkable efficacy for CLL therapy in essentially all patients, with low rates of primary and secondary resistance. Most notably, treatment with the Bruton tyrosine kinase (BTK) inhibitor ibrutinib[14] achieves high clinical response rates even in patients carrying genetic markers predictive of fast disease progression such as TP53 aberrations[15,16]. As the result, ibrutinib is becoming the standard of care for a large percentage of patients with high-risk CLL.

The mechanism of action of ibrutinib is rooted in its inhibition of BTK, which leads to downregulation of BCR signaling. Previous studies have investigated specific aspects of the molecular response to ibrutinib, for example investigating immunosuppressive mechanisms[17] and identifying decreased NF-κB signaling as a cause of reduced cellular proliferation[18-20]; but they did not map the genome-scale, time-resolved regulatory response to ibrutinib in primary patient samples. A detailed understanding of these temporal dynamics is particularly relevant given that successful ibrutinib therapy often induces an initial increase (rather than decrease) of CLL cells in peripheral blood, which can take months to resolve[21,22]. This observation has been explained by the drug's effect on cell-cell contacts[23,24], which triggers relocation of CLL cells from a protective microenvironment to the peripheral blood. The fact that ibrutinib induces lymphocytosis also contributes to the low correlation between the CLL cell count in the blood and the clinical response to ibrutinib therapy[22], and there is an unmet need for early molecular markers of response to ibrutinib therapy.

To dissect the precise cellular and molecular changes induced by ibrutinib therapy, and to identify candidate molecular markers of therapy response, we followed individual CLL patients (n = 7) at high temporal resolution (eight time points) over a standardized 240-day time course of ibrutinib treatment. Peripheral blood samples were analyzed for cell composition by flow cytometry, for epigenetic/regulatory cell state by ATAC-seq[25] on six different FACS-purified immune cell populations (158 ATAC-seq profiles in total), and for cell type specific transcriptional changes by single-cell RNA-seq[26] applied to a subset of time points (>43,000 single-cell transcriptomes in total).

Integrative bioinformatic analysis of the resulting dataset identified a consistent regulatory program of ibrutinib-induced changes that was shared across all patients: Within the first days after the start of ibrutinib treatment, CLL cells displayed reduced NF-κB binding, followed by reduced activity of lineage-defining transcription factors, and erosion of CLL cell identity. Finally, after an extended period of ibrutinib treatment, a quiescence-like gene signature was acquired by CLL cells – and unexpectedly also by $CD8^+$ T cells and other immune cell populations. This drug-induced regulatory program was present in all patients, and we were able to validate it in an independent CLL cohort. However, we observed substantial patient-to-patient variation in the speed with which these events unfold. Taking advantage of our time series data, we identified patient-specific predictors of the time to acquire an ibrutinib-induced molecular response, and we found predictive regulatory patterns already in pre-treatment samples.

In aggregate, our study provides a comprehensive, time-resolved analysis of the molecular and cellular dynamics upon ibrutinib treatment in CLL. It constitutes one of the first high-resolution, multi-omics time series of the molecular response to targeted therapy in cancer patients. The study also establishes a broadly applicable approach

for analyzing drug-induced regulatory programs and identifying molecular response markers for targeted therapy. Importantly, the study's high temporal resolution and its use of three complementary assays provided robust and informative results based on a small number of samples. The presented approach may be particularly relevant for obtaining maximum insight from early-stage clinical trials and off-label drug use involving few individual patients.

Results

Ibrutinib therapy induces global changes in immune cell composition and single-cell transcription profiles

To investigate the cellular dynamics and regulatory program induced by the inhibition of BCR signaling in CLL patients, we followed seven individuals from the start of ibrutinib therapy over a standardized time course of 240 days (**Figure 1a**). All patients received the same treatment regimen with daily doses of ibrutinib and underwent extensive clinical monitoring. The patients covered a range of different demographic, clinical, and genetic parameters, representative of the spectrum of refractory CLL encountered in clinical practice (**Supplementary Table 1**).

For all patients and up to eight time points (0, 1, 2, 3, 8, 30, 120/150, 240 days after the start of ibrutinib therapy), we performed immunophenotyping by flow cytometric analysis of peripheral blood mononuclear cells (PBMCs), systematically quantifying changes in cell composition in response to ibrutinib therapy (**Supplementary Figure 1a and Supplementary Table 2**). A gradual decrease in the percentage of CLL cells was observed over time (**Figure 1b**), but with extensive temporal heterogeneity across patients (**Supplementary Figure 1b,c**). The progressive reduction in the percentage of CLL cells coincided with an increase in the percentage of non-malignant natural killer (NK) and T cell populations, consistent with a recent report[23]. This trend was most visible for $CD8^+$ T cells (**Figure 1b,c and Supplementary Table 2**), while $CD4^+$ T cells remained largely unaffected. Although these differences were not statistically significant due to small cohort size, they were consistent with published data and thus provided validation and cellular characterization of our cohort and time course of ibrutinib therapy.

Based on flow cytometry, we also observed a statistically significant loss of CLL-associated surface receptors (CD5, CD38), which was specific to CLL cells (**Figure 1d, Supplementary Figure 2, and Supplementary Table 3**). To investigate the ibrutinib-induced changes in gene expression more systematically – and simultaneously in CLL cells as well as in matched non-malignant immune cells, we performed droplet-based single-cell RNA-seq[26] on the total PBMC population for a subset of patients and time points (**Supplementary Table 4**). Overall, ~43,000 single-cell transcriptomes passed quality control (**Supplementary Figure 3a,b**) and were integrated into a two-dimensional map using the UMAP method for unsupervised dimensionality reduction **(Figure 1e)**.

Cell type specific marker genes (e.g., CD79A, CD3D, CD14, and NKG7) were readily detectable in the single-cell RNA-seq data and were largely unaffected by ibrutinib treatment **(Supplementary Figure 3c)**, thus allowing for robust marker-based assignment of cell types. Cell counts inferred from single-cell RNA-seq were almost perfectly correlated with those obtained by flow cytometry (Spearman's $\rho = 0.95$, **Supplementary Figure 3d**), which provided independent validation of our single-cell RNA-seq dataset. We were also able to infer patient-specific copy number aberrations from the single-cell RNA-seq data (**Figure 1f**), which identified characteristic CLL-specific chromosomal aberrations including the deletion of chromosome 11q and 17p, and trisomy of chromosome 12.

Comparing the single-cell transcriptomes for each sample and cell type to the patient's corresponding pre-treatment (day 0) sample (**Supplementary Figure 3e-j and Supplementary Table 5**), we found cell type specific trends in the molecular response to ibrutinib therapy (**Figure 1g-h and Supplementary Figure 4**). In CLL cells, we observed reduced expression of the ibrutinib target BTK, of CD52 (a CLL disease activity marker[27]), and of CD27 (a regulator of B cell activation[28]). Among the non-malignant immune cell types, $CD8^+$ T cells were most strongly affected, which included downregulation of genes important for immune cell activation such as CD28, JUN, and

ZAP70. This pattern was shared to a lesser extent by $CD4^+$ T cells, while $CD14^+$ cells were characterized by strong upregulation of the NF-KB regulator NFKBIA.

Looking beyond individual genes, we further characterized the molecular response to ibrutinib by quantifying the transcriptome dynamics of predefined gene sets and transcriptional modules relevant to CLL and immunity (**Figure 1i and Supplementary Figure 5, 6**). We observed robust downregulation of B cell specific genes in CLL cells, including target gene sets of NF-κB subunits RELA and NF-κB1, as well as target gene sets of the well-established NF-κB associated transcription factors ATF2 and SPI1/PU.1. Genes involved in oxidative phosphorylation were also downregulated, consistent with widespread dampening of cellular activities in CLL cells under ibrutinib therapy. Among the non-malignant immune cell types, $CD8^+$ T cells showed broad downregulation that was less pronounced but similar to the response observed in CLL cells, and $CD14^+$ monocytes/macrophages showed specific upregulation of inflammatory response signatures including interferon gamma, TNF, and NF-κB signaling.

In summary, immunophenotyping and single-cell RNA sequencing over a dense time course of ibrutinib therapy uncovered widespread changes not only in CLL cells, but also in non-malignant immune cells. Most notably, we observed downregulation of NF-κB signaling and loss of B-cell surface markers in CLL, suggesting these are key contributors to the progressive reduction of the CLL cell fraction over time, and we observed a surprising degree of transcriptional change in non-CLL immune cells concomitant with an increase in the $CD8^+$ T cell fraction.

Chromatin mapping in CLL cells defines an ibrutinib-induced regulatory program leading to loss of B cell identity

To dissect the regulatory basis of the ibrutinib-induced changes in the CLL cell transcriptomes and immunophenotypes, we performed ATAC-seq on the FACS-purified $CD19^+CD5^+$ cell compartment over the ibrutinib time course (**Figure 2a, Supplementary Figure 7, and Supplementary Table 6**). We modeled the temporal progression as Gaussian processes (a statistical method for handling time series data[29]) and identified 6,797 genomic regions that underwent significant changes in chromatin accessibility in response to ibrutinib treatment (**Supplementary Table 7**). Four major clusters were detected among these genomic regions (**Figure 2b**): (i) regions that gradually lost chromatin accessibility (n = 3,412); (ii) regions that gradually gained chromatin accessibility (n = 2,199); (iii) regions that followed a bimodal, oscillating pattern (n = 369); and (iv) regions characterized by a peak in chromatin accessibility around 30 days after the start of ibrutinib treatment (n = 354).

We inferred the putative regulatory roles of these four clusters by region set enrichment analysis using the LOLA software[30]. LOLA identified those region sets out of a large reference dataset that showed significant overlap with the regions of the respective cluster (**Figure 2c**). Cluster 1 (decrease in chromatin accessibility) was strongly enriched for binding sites of transcription factors with a role in lymphoid differentiation and gene regulation, and also for enhancers specific to CLL cells and/or B cells. Cluster 2 (increase in chromatin accessibility) was enriched for B cell as well as T cell specific enhancers. Cluster 3 (bimodal, oscillating chromatin accessibility) was enriched for NF-κB binding sites. Lastly, Cluster 4 (peak in chromatin accessibility around day 30) was enriched for intragenic, transcribed regions marked by histone H3K36me3 in a range of hematopoietic cell types.

To identify potential regulators of the ibrutinib-induced modulation of CLL cell state, we focused on the enriched transcription factors (from **Figure 2c**) and estimated their dynamic changes in global binding activity over the ibrutinib time course, aggregating the ATAC-seq signal across each factor's putative binding sites (as defined by publicly available ChIP-seq data). As expected, several key transcription factors involved in B cell development (including NF-κB and PAX5) and B cell proliferation (including MEF2C and FOXM1) showed marked reduction of chromatin accessibility at their binding sites (**Figure 2d and Supplementary Figure 8a**). This effect was shared by CLL cells and non-malignant B cells but not observed in other immune cell types.

Integrative analysis of chromatin accessibility and cell type specific transcription further refined this picture. When we performed parallel enrichment analysis for transcription factors and their putative binding sites (**Figure 2e**), we observed concerted changes for key regulators of B cell development such as BCL11A, EBF1, IKZF1, IRF4, MEF2A, NFATC1, PAX5, and POU2F2, indicating that BTK inhibition may trigger loss of B cell identity in CLL cells. In support of this interpretation, we found global B cell specific gene expression signatures consistently downregulated upon ibrutinib treatment in CLL cells (**Figure 2f and Supplementary Figure 8b**).

Taken together, these results define a characteristic temporal order in which the ibrutinib-induced changes in CLL cells unfold. Already after one day of ibrutinib treatment, CLL cells showed a reduction in chromatin accessibility at NF-κB binding sites. This was followed by a gradual decrease in chromatin accessibility at binding sites of transcription factors that are regulated by NF-κB (PU.1[31], IRF4[32,33]) or interact with NF-κB (ATF2[34]). Moreover, we observed widespread reduction in B cell specific regulatory activity including decreased chromatin accessibility at B cell specific elements and at the binding sites of B cell transcription factors such BCL11A, NFATC1, and RUNX3, which was also reflected at the transcriptional level. These results highlight NF-κB mediated loss of B cell identity as the central cell-intrinsic change in CLL cells from patients under ibrutinib therapy.

Analysis of five immune cell types identifies ibrutinib-induced acquisition of a shared quiescence gene signature

To characterize the effect of ibrutinib therapy on gene regulation in non-malignant immune cells, we performed ATAC-seq on FACS-purified $CD19^+CD5^-$ B cells, $CD3^+CD4^+$ T helper cells, $CD3^+CD8^+$ cytotoxic T cells, $CD56^+$ NK cells, and $CD14^+$ monocytes/macrophages from the same patients and time points (**Supplementary Table 8**). We identified a total of 12,574 temporally dynamic regulatory regions in these five cell types (**Figure 3a,b; Supplementary Figure 9; Supplementary Table 9**). Unsupervised clustering detected shared temporal dynamics across these cell types, with sets of regions showing gradually decreased or increased chromatin accessibility over the time course, and a bimodal, wave-like cluster that was characterized by an initial decrease followed by a subsequent increase in chromatin accessibility (**Figure 3c and Supplementary Figure 9a-c**). Despite the shared temporal dynamics, the affected regions were highly cell type specific (**Supplementary Figure 9d**), suggesting that the different immune cell types react in characteristic ways to the direct and indirect effects of ibrutinib treatment.

Of the five non-malignant immune cell types, $CD19^+CD5^-$ B cells were most strongly affected by ibrutinib therapy, consistent with the major role of BCR signaling and the ibrutinib target BTK that non-malignant B cells share with $CD19^+CD5^+$ CLL cells. Regions with decreasing chromatin accessibility in the non-malignant B cells were enriched for the same set of transcription factor binding sites as the CLL cells, and also, to a lesser extent, for NF-κB binding sites (**Figure 3d**). In contrast, we detected fewer regions with increasing chromatin accessibility upon ibrutinib treatment, and those regions lacked distinctive patterns of functional enrichment, suggesting that they are indirect effects downstream of the cells' direct response to ibrutinib treatment (**Supplementary Table 9**).

Biologically interesting changes were not restricted to B cells. For example, regions with decreasing chromatin accessibility in $CD4^+$ T cells upon ibrutinib treatment were enriched for binding sites of CTCF and RAD2, which are involved in three-dimensional chromatin organization; and regions with decreasing chromatin accessibility in $CD8^+$ T cells were enriched for histone marks associated with repressed chromatin in other cell types (**Figure 3d**). Conversely, regions with increased chromatin accessibility in $CD4^+$ T cells over the ibrutinib time course were enriched for interferon signaling and open, promoter-associated chromatin in T cells, while the enrichment observed for $CD8^+$ T cells included CpG islands and H3K4me1-marked regulatory regions (**Figure 3d**).

Despite the cell type specific effects of ibrutinib therapy on CLL cells and non-malignant immune cells, we also identified a characteristic set of genes that underwent consistent changes across all of the investigated cell types (**Figure 3e and Supplementary Figure 9e**). This shared ibrutinib response signature was enriched for genes in-

volved in ribosomal functions, mRNA processing, oxidative phosphorylation/metabolism, translation factors, senescence, and autophagy (**Figure 3f**). For example, the shared ibrutinib response signature included CD44, a pan-lymphocyte cell adhesion molecule; CD99, a regulator of leukocyte migration, T cell adhesion, and cell death; CD37, which mediates the interaction of B and T cells; various surface proteins involved in cell adhesion (CD52, CD164, ICAM3, ITGB7); the protein tyrosine kinase FGR, which is a negative regulator of cell migration; TPT1 (a regulator of cellular growth and proliferation); and several factors involved in protein translation (EEF2, EID1, EIF1, EIF3E) as well as ribosomal proteins (**Supplementary Figure 10a**).

Interestingly, we also found genes involved in senescence and/or quiescence, namely CXCR4, a chemokine receptor required for hematopoietic stem cell quiescence[35,36]; ZFP36L2, an RNA binding protein that promotes quiescence in developing B cells[37]; and HMGB2, a chromatin protein involved in the regulation of gene expression in senescent cells[38] (**Supplementary Figure 10**). Together, these results suggest that CLL cells and non-malignant immune cells respond to ibrutinib therapy with shared transcriptional changes that include downregulation of genes involved in leukocyte function and cell-cell interactions, as well as upregulation of genes involved in quiescence and cellular senescence.

To assess the reproducibility of this shared ibrutinib response signature in an independent validation cohort, we utilized recently published bulk RNA-seq data for PBMCs from CLL patients ($n = 19$) that underwent single-agent ibrutinib treatment at a different medical center[20]. We indeed observed consistent changes in the expression of our gene signature for the vast majority of patients from the validation cohort (**Figure 3g**). The difference was statistically significant at both time points compared to day 0 (month 1: $p = 7.6e\text{-}6$; month 6: $p = 1.0e\text{-}7$; paired *t*-test), and an accurate distinction was possible between patient samples collected before and during ibrutinib therapy (receiver operating characteristic area under curve values of 0.89 and 0.79, respectively) (**Figure 3h**).

In summary, our data show that ibrutinib therapy induces time-dependent regulatory changes not only in CLL cells but also in other immune cell types. Changes in non-malignant B cells mirrored those in CLL cells (albeit with a weaker NF-κB signature), while $CD4^+$ T cells, $CD8^+$ T cells, NK cells, and myeloid cells responded in cell type specific ways. We further identified and validated a gene expression signature that captures broad ibrutinib-induced downregulation of immune cell functions and acquisition a quiescent-like state in response to ibrutinib therapy.

Prediction of patient-to-patient variability over the time course provides a molecular marker of ibrutinib response

Our data and analyses strongly support the existence of a consistent, ibrutinib-induced regulatory program that is shared across all patients. Nevertheless, we also observed substantial patient-to-patient variability at the genetic (**Figure 1f**), transcriptional (**Figure 3g**), chromatin-regulatory (**Figure 2d**), and cellular level (**Supplementary Figure 1b**). This heterogeneity in the presence of a shared regulatory program could be explained by patient-to-patient differences in the speed of progression along the regulatory program. Analyzing the molecular progression may therefore provide us with an opportunity to monitor or even predict, based on molecular profiles, which patients pursue a faster or slower time course toward a sustained cellular response under ibrutinib therapy.

Along these lines, we first investigated whether there were changes in the subclonal composition of CLL cells under ibrutinib treatment. To that end, we analyzed the copy number profiles inferred from the single-cell RNA-seq data and indeed observed subclonal genetic differences within patients over the time course (**Supplementary Figure 10a-d**). We also inferred the molecular response to ibrutinib therapy for each individual CLL cell, based on the expression intensity of our validated ibrutinib response signature (**Figure 3e**) in the single-cell transcriptomes. When we correlated this "ibrutinib molecular response score" with the single-cell copy number profiles, we did not observe a clear association between individual copy number aberrations and the strength of the molecular response to ibrutinib in single cells (**Supplementary Figure 10e-h**). However, we did observe an increase of subclonal genetic diversity of the time course of ibrutinib response, based on a quantitative measure that we validated

on simulated data and on the changing ratio of CLL cells versus non-malignant cells in our time course (**Figure 4a and Supplementary Figure 10i-l**). This change of subclonal genetic heterogeneity within patients was indeed positively associated with a strong cellular response to ibrutinib treatment as measured the flow cytometry data (**Figure 4b**).

Second, we investigated the association of chromatin accessibility in CLL cells at day 0 with a range of patient-specific characteristics. To that end, we performed principal component analysis on the chromatin profiles for all patients and cell types, and we tested for statistical associations with the clinical annotation data (**Supplementary Figure 11a**). We observed a strong association between the second principal component of the chromatin profiles in CLL cells at day 0 and the cellular response to ibrutinib treatment at day 120, suggesting that this chromatin signature provides an epigenomic marker for the subsequent cellular response to ibrutinib treatment (**Figure 4c,d**). This chromatin signature separated patients into fast versus slow responders to ibrutinib treatment independently of other clinical annotations (**Supplementary Figure 11b**). Genomic regions associated with a slow response to ibrutinib treatment showed essentially the same enrichment as those that were downregulated in CLL cells (**Figure 2c**), including preferential overlap with a broadly active state of cellular activity (**Supplementary Figure 11c)**.

Third, we employed our single-cell RNA-seq dataset to derive and evaluate gene expression signatures that capture the molecular response to ibrutinib treatment in individual cells. Using machine learning, we predicted the time of sample collection (day 0, 30, or 120/150) for each of the ~19,000 single-cell transcriptomes for CLL cells from four donors. Both support vector machines and elastic net classifiers achieved excellent prediction performance with cross-validated test-set ROC area under curve (AUC) values in the range of 0.975 to 0.999, and these results were robust to differences in the number of detected genes among the single-cell transcriptome profiles (**Supplementary Figure 12a**). Our observations indicate that the transcriptome profiles of single CLL cells undergo changes that precisely reflect the duration of ibrutinib therapy – which we can exploit for molecular staging of the patient-specific ibrutinib response. Using a classifier that was trained and evaluated by patient-stratified cross-validation, we observed that cells from specific patients were consistently predicted to have progressed faster (CLL5) or slower (CLL6) along the trajectory of the transcriptional response to ibrutinib treatment (**Figure 4e**), indicating that individual patients indeed follow their own timelines in the molecular response to ibrutinib therapy.

Finally, for a more quantitative assessment of these temporal dynamics, we trained and evaluated regression models that predict the precise time (i.e., number of days) after the start of ibrutinib therapy for each individual CLL cell transcriptome. We observed excellent test set prediction performance for three patients (CLL1, CLL6, and CLL8), with r^2 values (i.e., percent variance explained) of 92.3%, 84.2%, and 78.1%, respectively **(Supplementary Figure 12b-c**). Lower performance was observed for CLL5 ($r^2 = 36.6\%$), where the day-0 time point already showed a signature reminiscent of ibrutinib treatment (**Figure 4f)**. Consistent with the results of the classification analysis (**Figure 4e**), the regression models predicted individual patients progressing faster (CLL5) or slower (CLL6) along the trajectory of transcriptional response, while the two remaining samples (CLL1, CLL8) followed similar timelines (**Figure 4f**). When we compared predictions based on CLL cell transcriptomes at day 0 across patients, we found that the observed molecular signature prior to the start of ibrutinib treatment indeed anticipated the subsequent cellular response (i.e., reduction of CLL cells on day 120/150 compared to day 0) (**Figure 4g**).

These results indicate that genetic, epigenetic, and transcriptional variation between patients capture inter-individual differences in the response to ibrutinib treatment and may provide molecular markers that predict the time to a strong cellular response for individual patients.

Discussion

Multi-omics analysis of clinical time courses provides an effective approach for dissecting the molecular response to targeted therapy, allowing us to define the temporal order of events and to unravel relevant regulatory programs. Here, we applied flow cytometry, single-cell RNA-seq, and chromatin mapping in six FACS-purified cell types to a dense time course of CLL patients starting ibrutinib therapy. These three assays provide comprehensive and complementary information comprising the cellular response (flow cytometry), transcriptional changes across all major immune cell populations (single-cell RNA-seq), and the underlying chromatin dynamics that may explain and predict the observed changes in transcription regulation and epigenetic cell state (ATAC-seq).

Integrative bioinformatic analysis identified a characteristic regulatory program that was shared across all patients. Among the earliest changes following the start of ibrutinib therapy, we observed a decrease of NF-κB binding in CLL cells, which was followed by a rapid reduction in the regulatory activity of transcription factors involved in B cell development and function (such as EBF1, FOXM1, IRF4, PAX5, and PU.1). This decrease was accompanied by (and it likely caused) downregulation of CLL-specific gene signatures and a decrease in surface marker levels (including CD5 and CD19), which together indicate a broad erosion of CLL cell identity.

Ibrutinib-induced changes were not exclusive to CLL cells but shared by several other immune cell types. Non-malignant B cells largely mirrored the changes observed in CLL cells – which was expected given the role of the ibrutinib target BTK in BCR signaling. We also observed a dampening effect of ibrutinib on immune pathway regulation in $CD8^+$ T cells, while there was an increase of inflammatory gene signatures in monocytes/macrophages. The changes in immune cell types that do not express BTK could be due to a combination of direct effects via ibrutinib's promiscuous inhibition of kinases other than BTK (including BLK, BMX, ITK, TEC, TXK, and EGFR[39]) and indirect effects arising from the ibrutinib-induced relocation of CLL cells from the protective microenvironment into the peripheral blood.

Interestingly, for both CLL cells and for non-malignant immune cell populations, sustained ibrutinib therapy eventually resulted in the acquisition of a shared, quiescence-like gene signature. We validated this gene signature in an independent clinical cohort and confirmed its reproducibility. Closer inspection of the contributing genes may help explain certain cellular and clinical phenotypes observed in CLL patients under ibrutinib therapy, including changes in the immune microenvironment[40] and increased susceptibility to infections[41-43]. For example, CD99 downregulation indicates that Fas-mediated T-cell death may be impaired[44,45], which has been proposed as a cause of $CD8^+$ T cell accumulation in peripheral blood[23]. Moreover, two genes in the signature (CXCR4 and ZNF36L2) have established biological functions in senescence and quiescence of hematopoietic stem cells and lymphocytes. Ibrutinib is known to inhibit CXCR4-mediated expression of CD20 in CLL cells[46], which could have implications for ongoing trials combining ibrutinib and anti-CD20 antibodies in CLL (e.g., NCT02007044).

Our comprehensive, time-resolved, multi-omics analysis of ibrutinib therapy thus provides integration and context for previous studies that have focused on specific aspects of the response to ibrutinib, including reduced proliferation[19], decreased cell-cell contacts[23,24], and downregulation of NF-κB[18-20]. Moreover, the identification of a regulatory program that was shared across all patients allowed us to explore patient-to-patient heterogeneity in the speed with which this program is executed, suggesting that it may be feasible to define predictive molecular markers for the cellular response to ibrutinib therapy. Most notably, our chromatin analysis identified a patient-specific signature present prior to treatment that correlated with the speed of CLL clearance, and our single-cell RNA-seq data predicted the cellular response measured 120/150 days after the start of ibrutinib therapy. While these results remain exploratory due to the small number of patients in our study, they raise the future perspective of molecular response monitoring and prediction for a growing class of targeted cancer therapies that are not primarily cytotoxic and for which simple cell-based biomarkers (such as leukemic cell count or minimal residual disease) are poor predictors of the eventual clinical response.

While we consider our approach broadly applicable in the context of targeted therapies and precision oncology, the following limitations apply: First, such comprehensive profiling (up to 8 time points, dozens of genome-wide ATAC-seq profiles, and thousands of single-cell transcriptomes per patient) is currently feasible only for a relatively small number of patients. This precludes systematic interaction analysis with genetic risk markers for CLL. (However, recent studies have shown that established prognostic markers have lost much of their predictive power with ibrutinib therapy[47,48], and the same may apply to other emerging treatments such as CAR T cell therapy[49].) Second, while we found evidence of subclonal heterogeneity in our single-cell transcriptome data, the current throughput of single-cell RNA-seq does not (yet) enable deep characterization of the subclonal architecture. Third, most patients that start ibrutinib treatment have previously been treated with other drugs or drug combinations (our cohort: 1 to 5 prior treatments), which may explain some of the differences in the speed of the molecular response to ibrutinib in the individual patients. Fourth, time series data supports only a weak form of causal inference (Granger causality[50,51]), where earlier events may cause later events but not vice versa (e.g., the observed decrease in NF-κB binding was followed by a downregulation of B cell transcription factors and an erosion of B cell identity among the CLL cells). The results should therefore be considered causal in a strict biological sense only after mechanistic experimental validation in suitable disease models. With these limitations taken into account, we expect that the presented approach will readily generalize to other targeted therapies, defining shared regulatory programs and identifying molecular markers of the temporal dynamics and response to targeted therapy.

In conclusion, our study demonstrates the power of high-throughput assays combined with integrative bioinformatic analysis for dissecting the regulatory impact of targeted therapies. A strength of this approach is the level of detail and biological insight that can be obtained from a small number of patients, which makes it well-suited for applications in personalized medicine, where each patient may behave differently. Moreover, the approach appears promising for early-stage clinical trials of new targeted therapies, where it is critical to obtain a robust assessment of the induced molecular and cellular dynamics, in order to inform dose finding and to provide biomarker candidates for molecular response monitoring.

Methods

Sample acquisition and clinical data

All patients were treated at the Department of Haematology and Stem Cell Transplantation, Central Hospital of Southern Pest, Budapest, Hungary, according to the revised guidelines of the International Workshop Chronic Lymphocytic Leukemia/National Cancer Institute[52]. The study was approved by the ethical committees of the contributing institutions (Dél-Pesti Centrumkórház, Semmelweis University, and Medical University of Vienna). Informed consent was obtained from all participants.

Flow cytometry and fluorescence activated cell sorting (FACS)

Patient PBMCs were thawed and washed twice with PBS containing 0.1% BSA and 5 mM EDTA (PBS + BSA + EDTA). Cells were then incubated with anti-CD16/CD32 (clone 93, Biolegend) to prevent nonspecific binding. Single-cell suspensions were stained with combinations of antibodies against CD3 (FITC, clone UCHT1), CD4 (PE-TxRed, clone OKT4), CD5 (PE-Cy7, clone UCHT2), CD8 (APC-Cy7, clone SK1), CD14 (PerCp-Cy5.5, clone M5E2), CD19 (APC, clone HIB19), CD25 (PE-Cy7, clone BC96), CD38 (PE, clone HB-7), CD45RA (PerCp-Cy5.5, clone HI100), CD45RO (AF700, clone 304218), CD56 (AF700, clone NCAM16.2), CD127 (APC, clone A019D5), CD197 (CCR7, PE, clone G043H7), and DAPI viability dye (all from Biolegend) for 30 min at 4 °C followed by two washes with PBS + BSA + EDTA. For flow cytometry, cells were acquired with an LSRFortessa Cell Analyzer (BD). For FACS, cells were sort-purified with a MoFlo Astrois (Beckman Coulter) using the gating

strategy depicted in **Supplementary Figure 1a**. Data analysis was performed with the FlowJo (Tree Star) software. In Figure 1d, control cells in CD5-PE-Cy7 channel are CD14+ myeloid cells, and control cells in CD38-PE channel are CD3+CD4-CD8- cells; these are cell populations which are known to not express the respective markers and based on which the background levels can be estimated.

Droplet-based single-cell RNA-seq

Single-cell libraries were generated using the Chromium Controller and Single Cell 3' Library & Gel Bead Kit v2 (10x Genomics) according to the manufacturer's protocol. Briefly, an aliquot of patient PBMCs was stained with DAPI for discrimination between live and dead cells, and a maximum of 100,000 live, doublet-excluded cells were sorted into 1.5 ml tubes. Cells were pelleted by centrifuging for 5 min at 4 °C at 300 x g and resuspended in PBS with 0.04% BSA. Up to 17,000 cells suspended in reverse transcription reagents, along with gel beads, were segregated into aqueous nanoliter-scale gel bead-in-emulsions (GEMs). The GEMs were then reverse-transcribed in a C1000 Thermal Cycler (Bio-Rad) programmed at 53 °C for 45 min, 85 °C for 5 min, and hold at 4 °C. After reverse transcription, single-cell droplets were broken, and the single-strand cDNA was isolated and cleaned with Cleanup Mix containing Dynabeads MyOne SILANE (Thermo Fisher Scientific). cDNA was then amplified with a C1000 Thermal Cycler programed at 98 °C for 3 min, 10 cycles of (98 °C for 15 s, 67 °C for 20 s, 72 °C for 1 min), 72 °C for 1 min, and hold at 4 °C. Subsequently, the amplified cDNA was fragmented, end-repaired, A-tailed, and index adaptor ligated, with cleanup in-between steps using SPRIselect Reagent Kit (Beckman Coulter). Post-ligation product was amplified with a T1000 Thermal Cycler programed at 98 °C for 45 s, 10 cycles of (98 °C for 20 s, 54 °C for 30 s, 72 °C for 20 s), 72 °C for 1 min, and hold at 4 °C. The sequencing-ready library was cleaned up with SPRIselect and sequenced by the Biomedical Sequencing Facility at CeMM using the Illumina HiSeq 3000/4000 platform and the 75 bp paired-end configuration.

Assay for transposable-accessible chromatin with sequencing (ATAC-seq)

Chromatin accessibility mapping was performed using the ATAC-seq method as previously described[25,53], with minor adaptations. In each experiment, a maximum of 50,000 sorted cells were pelleted by centrifuging for 5 min at 4 °C at 300 x g. After centrifugation, the pellet was carefully resuspended in the transposase reaction mix (12.5 µl 2xTD buffer, 2 µl TDE1 (Illumina), and 10.25 µl nuclease-free water, 0.25 µl 5% Digitonin (Sigma)) for 30 min at 37 °C. Following DNA purification with the MinElute kit eluting in 11 µl, 1 µl of the eluted DNA was used in a quantitative PCR reaction to estimate the optimum number of amplification cycles. Library amplification was followed by SPRI size selection to exclude fragments larger than 1,200 bp. DNA concentration was measured with a Qubit fluorometer (Life Technologies). Library amplification was performed using custom Nextera primers[25]. The libraries were sequenced by the Biomedical Sequencing Facility at CeMM using the Illumina HiSeq 3000/4000 platform and the 50 bp single-read configuration.

Preprocessing and analysis of single-cell RNA-seq data

Preprocessing of the single-cell RNA-seq data was performed using Cell Ranger version 2.0.0 (10x Genomics). Raw sequencing files were demultiplexed using the Cell Ranger command 'mkfastq'. Each sample was aligned to the human reference genome assembly 'refdata-cellranger-GRCh38-1.2.0' using the Cell Ranger command 'count', and all samples were aggregated using the Cell Ranger command 'aggr' without depth normalization. Raw expression data were then loaded into R version 3.4.0 and analyzed using the Seurat package version 2.0.1 with the parameters suggested by the developers[54]. Specifically, single-cell profiles with less than 200 detected genes (indicative of no cell in the droplet), more than 3,000 detected genes (indicative of cell duplicates), or more than

15% of UMIs stemming from mitochondrial genes were discarded. Read counts were normalized dividing by the total UMI count in each cell, multiplied by a factor of 10,000, and log transformed. The number of UMIs per cell and the percent of mitochondrial reads per cell were then regressed out using Seurat's standard analysis pipeline.

Dimensionality reduction and supervised analysis of gene expression

Principal component analysis, t-SNE analysis, hierarchical clustering, and differential expression analyses were carried out in R, using the respective functions of the Seurat package. t-SNE and cluster analyses were based on the first ten principal components. A negative binomial distribution test was used for differential analysis on genes expressed in at least 10% of cells in one group. Results were aggregated across patients by taking the mean for log fold changes and by Fisher's method for p-values. Enrichment analyses were done using Enrichr API[55] against the following databases: Transcription Factor PPIs, ENCODE, ChEA Consensus TFs from ChIP-X, NCI-Nature 2016, WikiPathways 2016, Human Gene Atlas, and Chromosome Location. Aggregate gene expression values for gene sets (signatures) were quantified as follows: Log-normalized transcript per 10^4 UMI counts were scaled between 0 and 1. The values for all genes of a given set were then summed to obtain a raw value for each gene set and cell. To remove cell specific effects such as differences in UMI distributions due to sequencing depth, raw values were transformed to Z-scores using a distribution of raw values of 500 randomly picked gene sets of the same size. Differences in signatures between time points were assessed using 't.test' in R. Results were aggregated across patients by taking the mean for log fold changes and by Fisher's method for p-values. Multiple testing correction of differentially expressed genes, enriched terms, and differences in signatures was carried out using the Benjamini-Hochberg procedure as implemented by the 'p.adjust' function in R. The selected gene sets included 50 'hallmark signatures' from MSigDB[56], as well as ATF2, BATF, NFIC, NFKB1, RELA, RUNX3, and SPI target genes, and B cell signatures from Human Gene Atlas, NCI Nature 2016, and WikiPathways 2016, all obtained from Enrichr[55]. For data representation, we denoised the dataset with the deep count autoencoder (DCA) in Python[57], using raw UMI counts as input and the 'Zero-Inflated Negative Binomial' model (which explained the relationship between mean expression and observed dropout rates significantly better than the 'Negative Binomial' model). The DCA-denoised data were then normalized per cell, log-transformed, and scaled. Dimensional reduction was performed by principal component analysis, and the resulting dimensions were used for neighbor graph construction followed by Uniform Manifold Approximation and Projection (UMAP) with Scanpy's default parameters[58].

Preprocessing and analysis of ATAC-seq data

ATAC-seq reads were trimmed using Skewer[59] and aligned to the GRCh37/hg19 assembly of the human genome using Bowtie2[60] with the '-very-sensitive' parameter. Duplicate reads were removed using the sambamba[61] 'mark-dup' command, and reads with mapping quality >30 and alignment to the nuclear genome were kept. All down-stream analyses were performed on these filtered reads. Peak calling was performed with MACS2[62] using the '-nomodel' and '-extsize 147' parameters, and peaks overlapping blacklisted features as defined by the ENCODE project[63] were discarded. We created a consensus region set by merging the called peaks from all samples across patients and cell types, and we quantified the accessibility of each region in each sample by counting the number of reads from the filtered BAM file that overlapped each region. To normalize the chromatin accessibility signal across samples, we first performed quantile normalization using the R implementation in the preprocessCore package ('normalize.quantiles' function). We then performed principal component analysis (scikit-learn, sklearn.decomposition.PCA implementation) on the normalized chromatin accessibility values of all chromatin-accessible regions across all samples. Upon inspection of the sample distribution along principal components, we noticed an association of several (but not all) samples from one processing batch with a specific principal component, while we did not observe any association of these samples with any known biological factor. To remove the effect of this

latent variable while retaining variation from other (biological) sources, we performed principal component analysis on the matrix of raw counts on a per cell type basis (except myeloid cells, which contained no such samples) and removed the latent variable (first principal component) by subtracting the outer product of the transformed values of each sample in this component and the loadings of each regulatory element in the same component from the original matrix. We then again performed normalization of the corrected count matrix and component analysis jointly for all cell types as before.

Time series modelling of chromatin accessibility dynamics

We modeled the temporal effect of ibrutinib in each cell type as a function of time by a latent process. To that end, we used the Python library GPy to fit Gaussian process regression models (GPy.models.GPRegression) on the log2 transformed sampling time on ibrutinib therapy (independent variable) and the normalized chromatin accessibility values for each regulatory element (dependent variable) for each cell type separately. We fitted a variable radial basis function (RBF) kernel as well as a constant kernel (both with an added noise kernel), and we compared the log-likelihood and standard deviation of the posterior probability of the two as previously described[64-66]. Dynamic regulatory elements were defined as those for which the survival function of the chi-square of the D statistic (twice the difference between the log-likelihood of the variable fit minus the log-likelihood of the constant fit) was smaller than 0.05 and the standard deviation of the posterior was higher than 0.05 (as described previously[66]). We then used the 'mixture of hierarchical Gaussian process' (MOHGP) method to cluster regulatory elements according to their temporal pattern. The MOHGP class from the GPclust library (GPclust.MOHGP) was fitted with the same data as before, this time with a Matern52 kernel (GPy.kern.Matern52) and an initial guess of four region clusters. Regions with posterior probability higher to 0.8 were selected as dynamic and included in the downstream analysis.

Region set enrichment analysis

We performed region set enrichment analysis on the clusters of dynamic genomic regions using LOLA[30] and its core database, which comprises transcription factor binding sites from ENCODE[63], tissue-specific DNase hypersensitive sites[67], the CODEX database[68], UCSC Genome Browser annotation tracks[69], the Cistrome database[70], and data from the BLUEPRINT project[71]. Motif enrichment analysis was performed with the HOMER[72] 'findMotifsGenome' command using the following parameters: '-mask -size 150 -length 8,10,12,14,16 -S 12'. Enrichment of genes associated with regulatory elements (annotated with the nearest transcription start site from Ensembl) was performed through the Enrichr API[55] for the following databases of gene sets: BioCarta 2016, ChEA 2016, Drug Perturbations from GEO down, Drug Perturbations from GEO up, ENCODE and ChEA Consensus TFs from ChIP-X, ENCODE TF ChIP-seq 2015, ESCAPE, GO Biological Process 2017b, GO Molecular Function 2017b, KEGG 2016, NCI-Nature 2016, Reactome 2016, Single Gene Perturbations from GEO down, Single Gene Perturbations from GEO up, and WikiPathways 2016.

Inference of global transcription factor activity

Global transcription factor accessibility was assessed by aggregating the normalized chromatin accessibility values of regulatory elements that overlap a consensus of regions (union of all sites) from ENCODE ChIP-seq peaks of the same factor across all cell types profiled. The mean accessibility of each sample in the sites overlapping binding sites of each factor was computed and subtracted by the mean accessibility of each sample across all measured regulatory elements. For visualization, we aggregated samples by cell type and sampling time point, displaying either the mean or a Z-score of chromatin accessibility. For all gene-level measures of chromatin accessibility, we

used the mean of all regulatory elements associated with a gene, defined as the gene with the closest transcription start site as annotated by the RefSeq gene models for the hg19 genome assembly.

Integrative analysis of ATAC-seq and single-cell RNA-seq data

To assess the agreement between the two analyses at the enrichment level, we performed enrichment analysis with Enrichr for genes differentially expressed across patients in the same cell type, and we compared the significance of terms for transcription factors in the 'ENCODE TF ChIP-seq 2015' gene set library with the significance of transcription factors enriched in the LOLA analysis for each ATAC-seq cluster. To identify a common transcriptional signature associated with ibrutinib treatment across cell types, we selected all genes that were differentially expressed with the same direction in at least 10 combinations of cell type and time point. These genes were split according to the direction of change with time and used for enrichment analysis with Enrichr as described above. The same genes were used to derive a score calculated as the mean expression of the upregulated genes over the mean expression of downregulated genes. An independent cohort of RNA-seq on bulk PBMCs from CLL patients[20] was used to assess the reproducibility of the signature by observing the significance of the difference between scores upon ibrutinib treatment with a paired-samples t-test. To assess the performance of the score as a classifier, we generated a ROC curve by counting true positive and negative rates with a sliding score threshold, and we calculated the area under the curve with scikit-learn's function 'sklearn.metrics.auc'.

Inference of DNA copy number variation from single-cell RNA-seq data

To infer DNA copy number profiles at the single-cell level, we started with DCA-denoised, normalized, and scaled single RNA-seq data of all cells. We removed per-cell differences by subtracting the median expression of each cell from all genes and per-gene differences by subtracting the median and dividing by the standard deviation. We then calculated a rolling mean of expression across genes ordered by their chromosomal position for each chromosome individually. To improve the representation of DNA copy number profiles, we centered the resulting matrix by subtracting the mean of all values in the matrix and applied smoothing by cubing the matrix values (which shrinks small changes relative to all cells) and multiplying them by 3 (which scales values back to usual copy number variation bounds). To discover clusters of genetically distinct cells within patients, we performed dimension reduction using principal component analysis on the smoothed matrix, computed a neighbor graph between cells, and fitted a UMAP manifold for the CLL cells of each sample (i.e. per patient and time point). This was overlaid with the response to ibrutinib of each single cell based on ibrutinib response signature described above. To assess global changes in genetic diversity within cells of a patient over time, we developed a global metric of genetic diversity based on inferred copy number profiles from single-cell RNA-seq data. We calculated pairwise Pearson correlation coefficients between all cells and used the square of the mean of this distribution as a measure of genetic diversity. To benchmark this approach, we first established simulated copy number profiles with the same dimensions are the inferred one but for varying total numbers of cells. We created two populations where we simulated gain or loss of chromosome 12 (log change: -1 or 1) whereas the remainder of the genome was Gaussian noise of mean zero and standard deviation 0.1. We assessed performance by mixing the two populations together in different ratios and computing Pearson correlation between the population fraction (ground truth) and the predicted global diversity. An additional benchmark was performed by taking advantage of natural, known mixtures of cell types in the data. For these data, the inferred change in genetic diversity is simply the difference between global diversity measures between time points of ibrutinib treatment for each patient.

Prediction of sample collection and patient-specific response time from single-cell RNA-seq data

The time point of sample collection (day 0, 30, or 120/150) for each CLL single-cell transcriptome was predicted using the glmnet package in R with a multinomial response variable (for classification) and the 'alpha' parameter (lasso penalty) set to 1. Prediction performance was assessed by 3-fold cross-validation for each patient, where optimal 'lambda' parameters were obtained separately for each (outer) fold in a 5-fold inner cross-validation using the function cv.glmnet. Parameter 'lambda' for the final prediction across patients were obtained by 5-fold cross-validation on all data for each patient using cv.glmnet. Predictions were aggregated for each patient by taking the mean of dummy variables (1: early, 2: mid, 3: late) across the three other patients. Classification performance for support vector machines was assessed using the LiblineaR package in R. Classifiers were trained with 'type' parameter 0 and 'cost' parameters estimated by the heuristicC method on the training data, where cells were split ten times into 70% for training and 30% for testing. Quantitative prediction of the precise time (number of days) after the start of ibrutinib therapy was performed using the glmnet package in R with a Gaussian response variable (for regression) and the 'alpha' parameter (lasso penalty) set to 1. Prediction performance was assessed using the 'lambda' parameter that provided the highest R^2 in the training data of each fold. The regularization parameter 'lambda' for the final prediction was obtained based on the mean squared error in a 3-fold cross-validation repeated five times on all data from each patient. Predictions were aggregated by taking the mean across three patients. For all Python analysis, we set the pseudo-random number generation seed state to 1142101101 in both the standard library 'random' and in 'numpy'.

Data availability

All data are available through the Supplementary Website (http://cll-timecourse.computational-epigenetics.org/). Single-cell RNA-seq and ATAC-seq data (sequencing reads, intensity values) have been deposited at NCBI GEO and are publicly available under accession number GSE111015.

Code availability

The analysis source code underlying the final version of the paper will be provided on the Supplementary Website (http://cll-timecourse.computational-epigenetics.org/).

Acknowledgements

We would like to thank all patients who have donated their samples for this study. We also thank the Biomedical Sequencing Facility at CeMM for assistance with next generation sequencing and all members of the Bock lab for their help and advice. C.S. was supported by a Feodor Lynen Fellowship of the Alexander von Humboldt Foundation. N.F. is supported by a fellowship from the European Molecular Biology Organization (EMBO ALTF 241-2017). T.K. is supported by a Lise-Meitner fellowship from the Austrian Science Fund (FWF M2403). D.A. and Cs.B. are supported by the K119950, KH17-126718, NVKP_16-1-2016-0004, and NVKP_16-1-2016-0005 grants of the Hungarian National Research, Development and Innovation Office, the Janos Bolyai research scholarship, and the LP95021 grant of the Hungarian Academy of Sciences. C.B. is supported by a New Frontiers Group award of the Austrian Academy of Sciences and by an ERC Starting Grant (European Union's Horizon 2020 research and innovation programme, grant agreement n° 679146).

Author contributions

A.F.R., T.K., D.A., C.S., and Ch.B. designed the study; S.T., M.R., Z.M., and Cs.B. provided samples and clinical data; T.K., F.Z., T.P., and C.S. performed the experiments with contributions from M.F., L.C.S., A.N., and D.A.; A.F.R., and N.F. analyzed the data with contributions from T.K. and Ch.B.; Ch.B. supervised the research. A.F.R., T.K., N.F., Z.M., Cs.B., D.A., C.S., and Ch.B. wrote the manuscript with contributions from all authors.

Competing financial interests

The authors declare no competing financial interests.

Figures

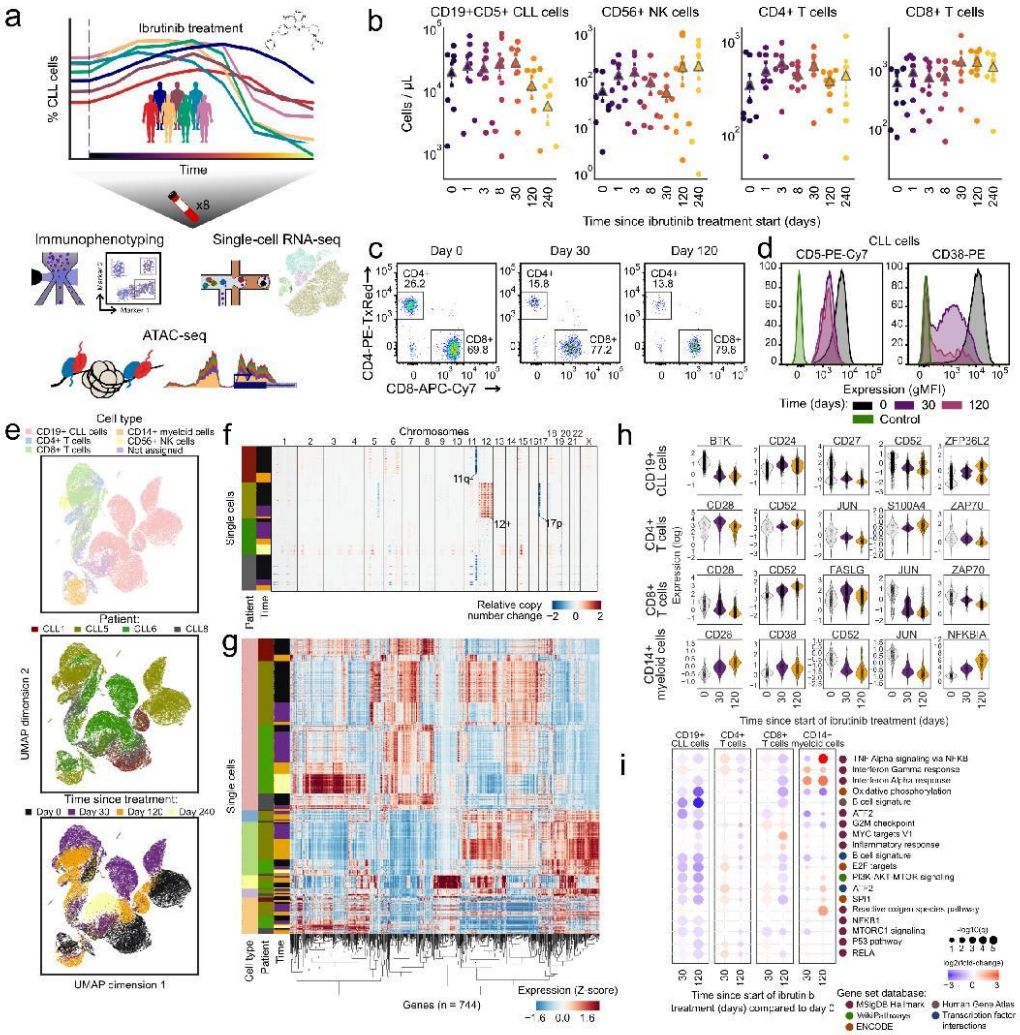


Figure 1: Time series analysis of the cellular and transcriptional response to ibrutinib in CLL patients identifies widespread changes in several immune cell types

a) Schematic representation of the study design. Peripheral blood from CLL patients undergoing single-agent ibrutinib therapy was collected at defined time points and assayed by flow cytometry (immunophenotype), single-cell RNA-seq (gene expression), and ATAC-seq (chromatin regulation). **b)** Cell type abundance over the ibrutinib time course, measured by flow cytometry. Triangles represent the mean for each time point. **c)** Flow cytometry scatterplots showing the abundance of T cell subsets for one representative patient at three time points (day 0: before the initiation of ibrutinib therapy, day 30 (120): 30 (120) days after the initiation of ibrutinib therapy). Cells positive for CD3 or CD8 were gated as indicated by the black rectangles and quantified as percentages of live

PBMCs. **d)** Flow cytometry histograms showing CD5 and CD38 expression on CLL cells (pre-gated for live, single $CD19^{+}CD5^{+}$ cells) for a representative patient and three time points. **e)** Two-dimensional similarity map (UMAP projection) showing all single-cell transcriptome profiles that passed quality control. Cells are color-coded according to their assigned cell types based on the expression of known marker genes. **f)** DNA copy number profiles for CLL cells inferred from single-cell RNA-seq data, which detect three genetic aberrations common in CLL (annotated in the pane). For illustration, 2,500 randomly selected CLL cells are shown for each patient. **g)** Clustered single-cell transcriptome heatmap for the most differentially expressed genes between time points. For illustration, 20,000 cells are shown. **h)** Violin plots showing the distribution of gene expression levels for selected differentially expressed genes over the time course. **i)** Differential gene expression signatures in four cell types, comparing each sample to the matched pre-treatment sample and averaging across patients. Patient-individual data are shown in Supplementary Figure 6.

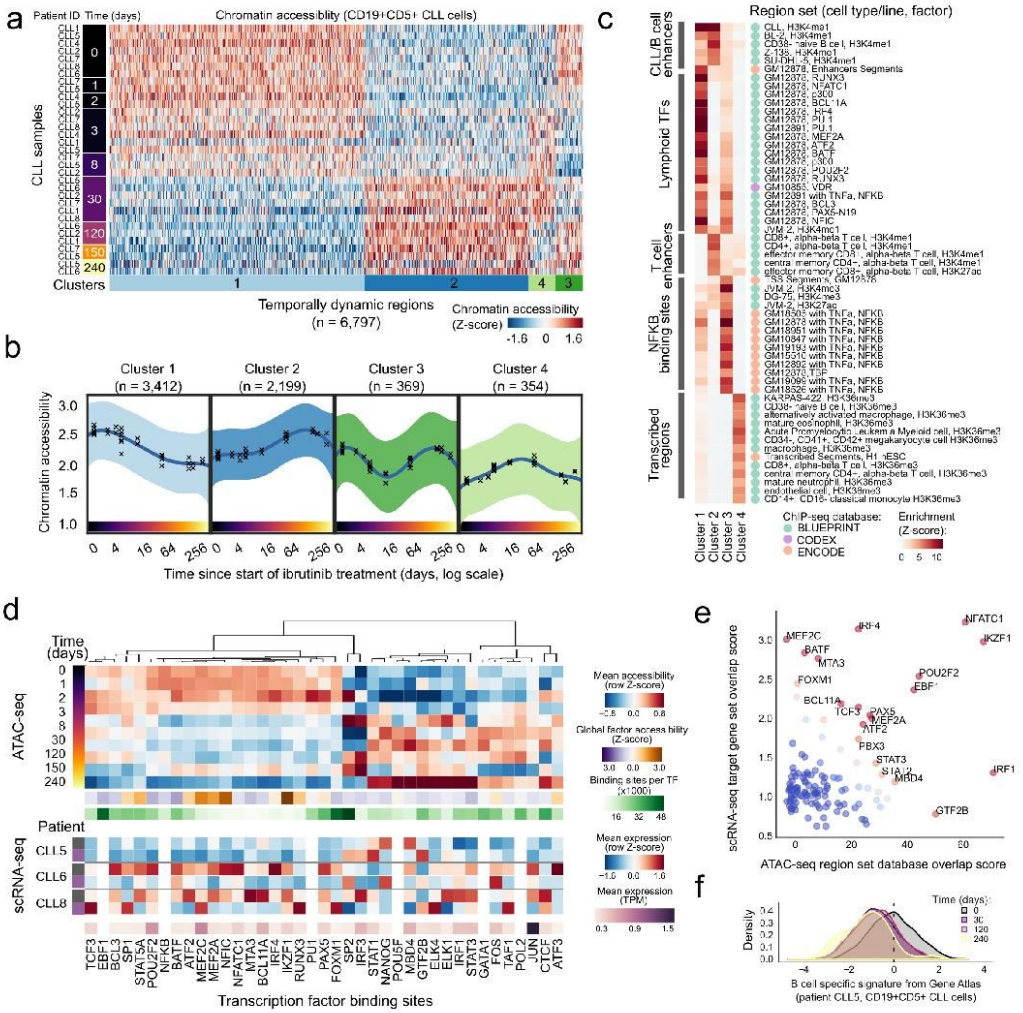


Figure 2: Integrated analysis of chromatin accessibility and gene expression in CLL cells uncovers a consistent regulatory program induced by ibrutinib therapy

a) Heatmap showing changes in chromatin accessibility for CLL cells over the time course of ibrutinib treatment. **b)** Mean chromatin accessibility across patients plotted over the ibrutinib time course in dynamically changing regulatory regions, highlighting the non-linear aspect of the ibrutinib effect on the chromatin. Crosses represent samples from a single patient at a specific time point, and 95% confidence intervals are shown as colored shapes. **c)** Region set enrichments for the clusters of dynamic regions, calculated using the LOLA software. Enrichment p-values were Z-score transformed per column. **d)** Heatmaps showing mean chromatin accessibility of regulatory regions overlapping with putative binding sites, expression of the corresponding transcription factor, and total number of its binding sites. Clustering was performed on the mean chromatin accessibility values. **e)** Scatterplot showing differential regulation of transcription factors upon ibrutinib treatment. The x-axis displays the enrichment

of transcription factors enriched in the LOLA analysis, and the y-axis displays the enrichment of their target genes among the differentially expressed genes. **f)** Gene expression histogram across CLL cells in one patient, demonstrating the decline of a B cell-specific expression signature over the time course of ibrutinib treatment. For illustration, the patient with most time points in the single-cell RNA-seq analysis (CLL5) is displayed.

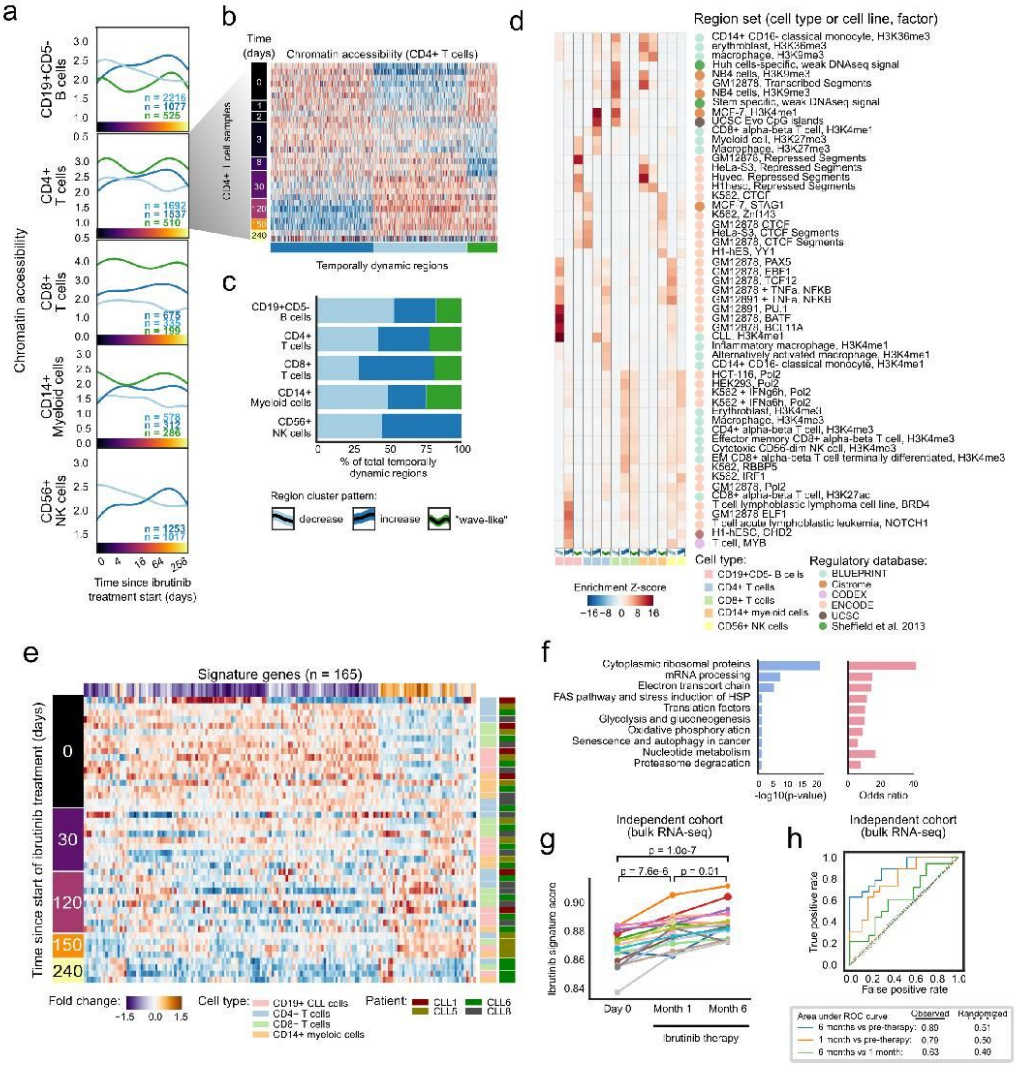


Figure 3: Integrated analysis of chromatin accessibility and gene expression for five immune cell types identifies regulatory changes in response to ibrutinib that converge on a shared quiescence-like gene signature

a) Mean chromatin accessibility across patients plotted over the ibrutinib time course for clusters of dynamically changing regulatory regions in five immune cell types. **b)** Heatmap of chromatin accessibility of CD4+ cells, illustrating dynamic regulation over the ibrutinib time course. **c)** Stacked bar plots indicating the percentage of dynamically changing regions in each cluster. **d)** Region set enrichments for the clusters of dynamically changing regions, calculated using the LOLA software and publicly available region sets as reference (mainly based on ChIP-seq data). Enrichment p-values were Z-score transformed per column. **e)** Heatmap showing mean expression levels for genes that were differentially expressed over the ibrutinib time course when combining the data for CLL cells and

for the five non-malignant immune cell types. Values represent column Z-scores of gene expression. **f)** Gene set enrichments for genes downregulated across cell types, using WikiPathways as reference (Fisher's exact test, left: FDR-corrected p-value, right: odds ratio as a measure of effect size). **g)** Expression score for the gene signature (as shown in panel e) in an independent cohort, calculated from bulk RNA-seq data for PBMCs collected before the start of ibrutinib therapy and at two subsequent time points. Significance was assessed using a paired t-test. **h)** ROC curve illustrating the prediction performance of the gene signature (from panel e) for classifying samples in the independent validation cohort into those collected before ibrutinib treatment and those collected during ibrutinib treatment. As negative controls, the prediction was repeated 100 times with permuted class labels for each combination of time points, and the mean ROC curves across iterations are shown as dotted lines.

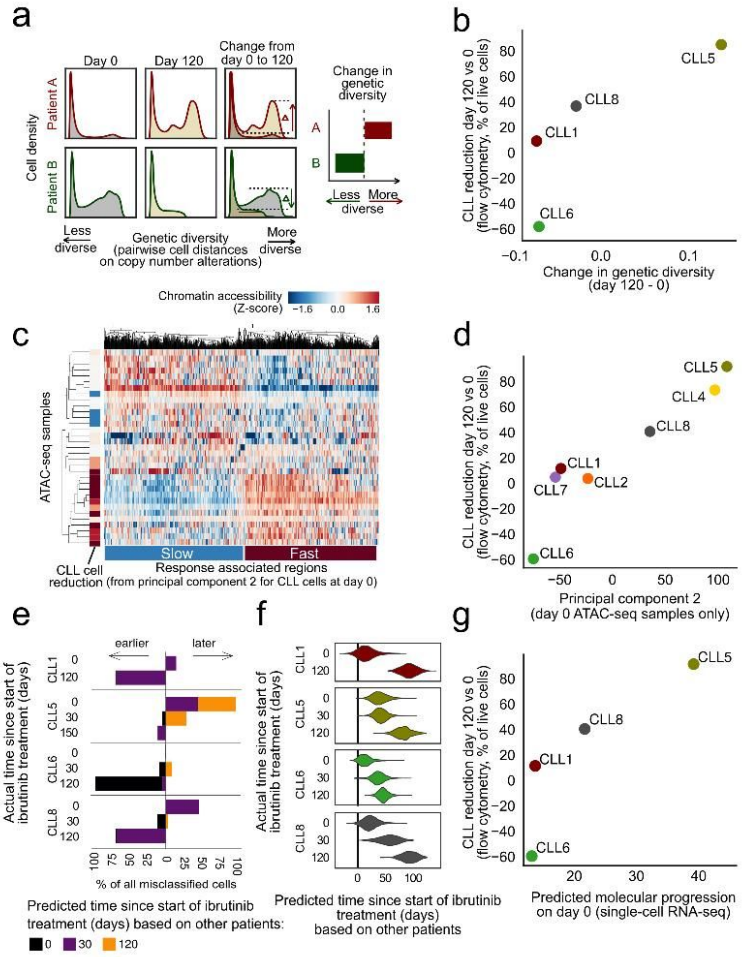
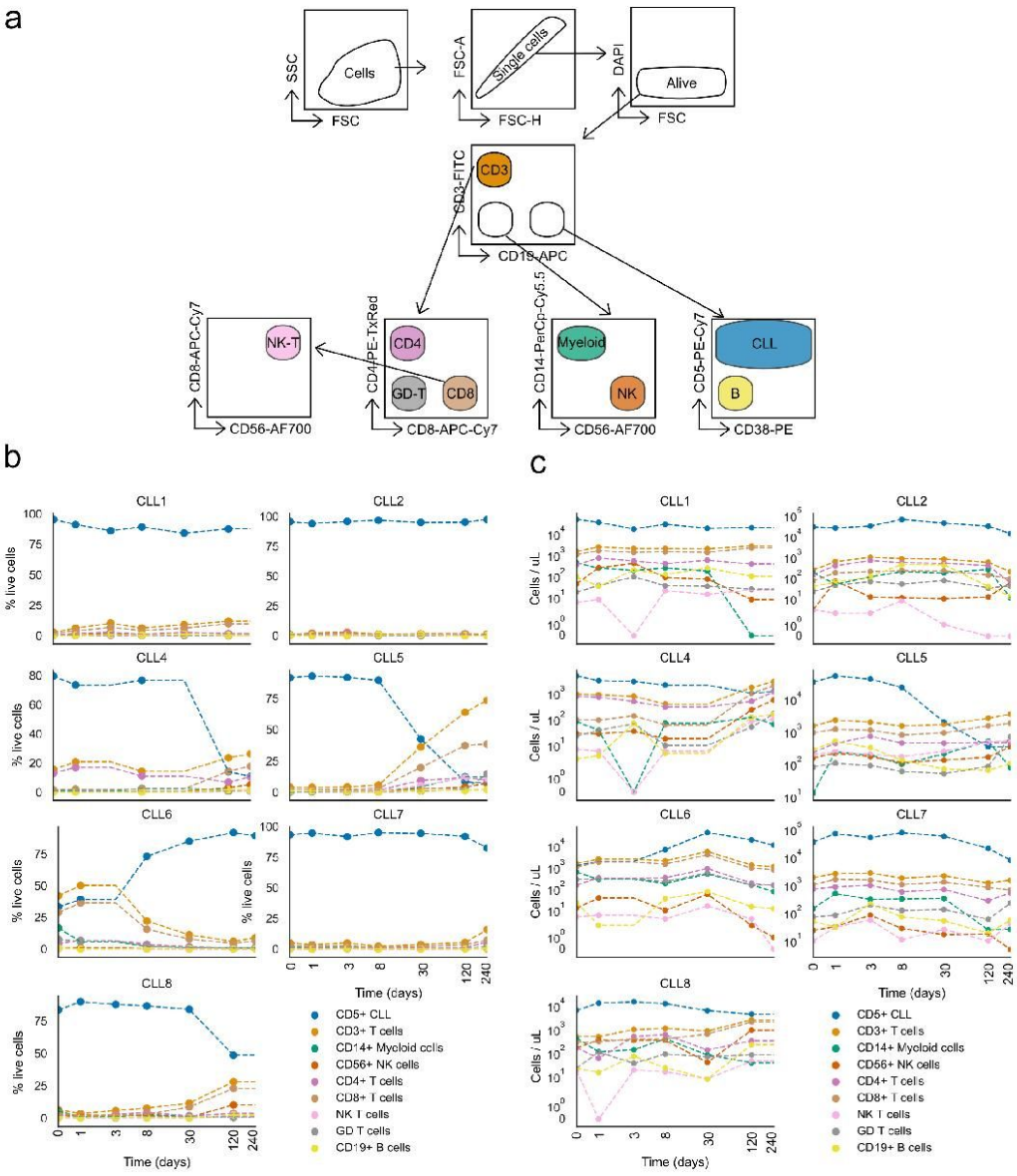


Figure 4: Analysis of copy number profiles, chromatin accessibility, and single-cell transcriptomes identified patient-specific associations with the speed of the cellular response to ibrutinib therapy

a) Computational approach to quantify changes in genetic diversity based on copy number profiles inferred from the single-cell RNA-seq data. Shifts in the distribution of pairwise distance similarities between time points indicate changes in the genetic diversity of the cell population. **b)** Scatterplot comparing across patients the change in genetic diversity between day 0 and 120/150 of ibrutinib treatment (x-axis) with the change in the CLL cell percentage on day 120 /150 of ibrutinib treatment compared to day 0 as measured by flow cytometry (y-axis). **c)** Clustered heatmap showing chromatin accessibility profiles for CLL cells at day 0 for the top 1000 genomic regions associated with the second principal component for these profiles (from **Supplementary Figure 11a**), annotated on the left with the change in CLL cell fraction (as in panel b). **d)** Scatterplot comparing across patients the average chromatin accessibility across regions linked to the second principal component (as in panel c, x-axis) with the change in CLL cell fraction (as in panel b, y-axis). **e)** Stacked bar charts showing the number and direction of

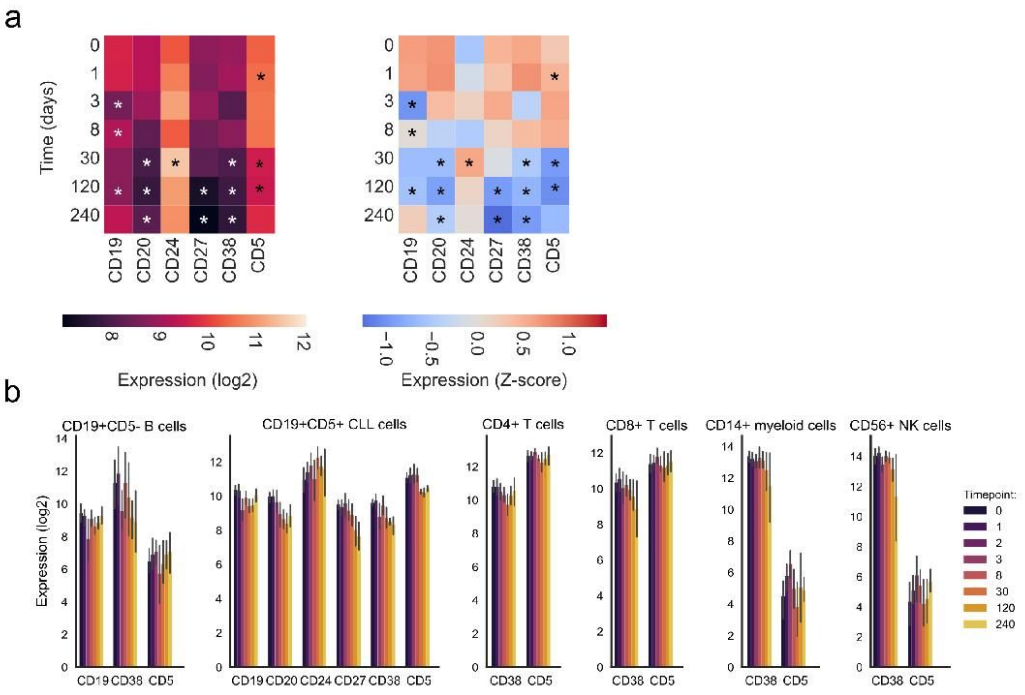
deviations from the actual collection time point when predicting time points in each patient after training the classifier in all other patients. **f**) Violin plots showing the predicted (x-axis) and actual (y-axis) number of days under ibrutinib therapy in each patient. Predictions are derived from regression models trained on all other patients. **g)** Scatterplot comparing the predicted time under ibrutinib therapy (from panel f, x-axis) with the change in CLL cell fraction (as in panel b, y-axis).

Supplementary figures



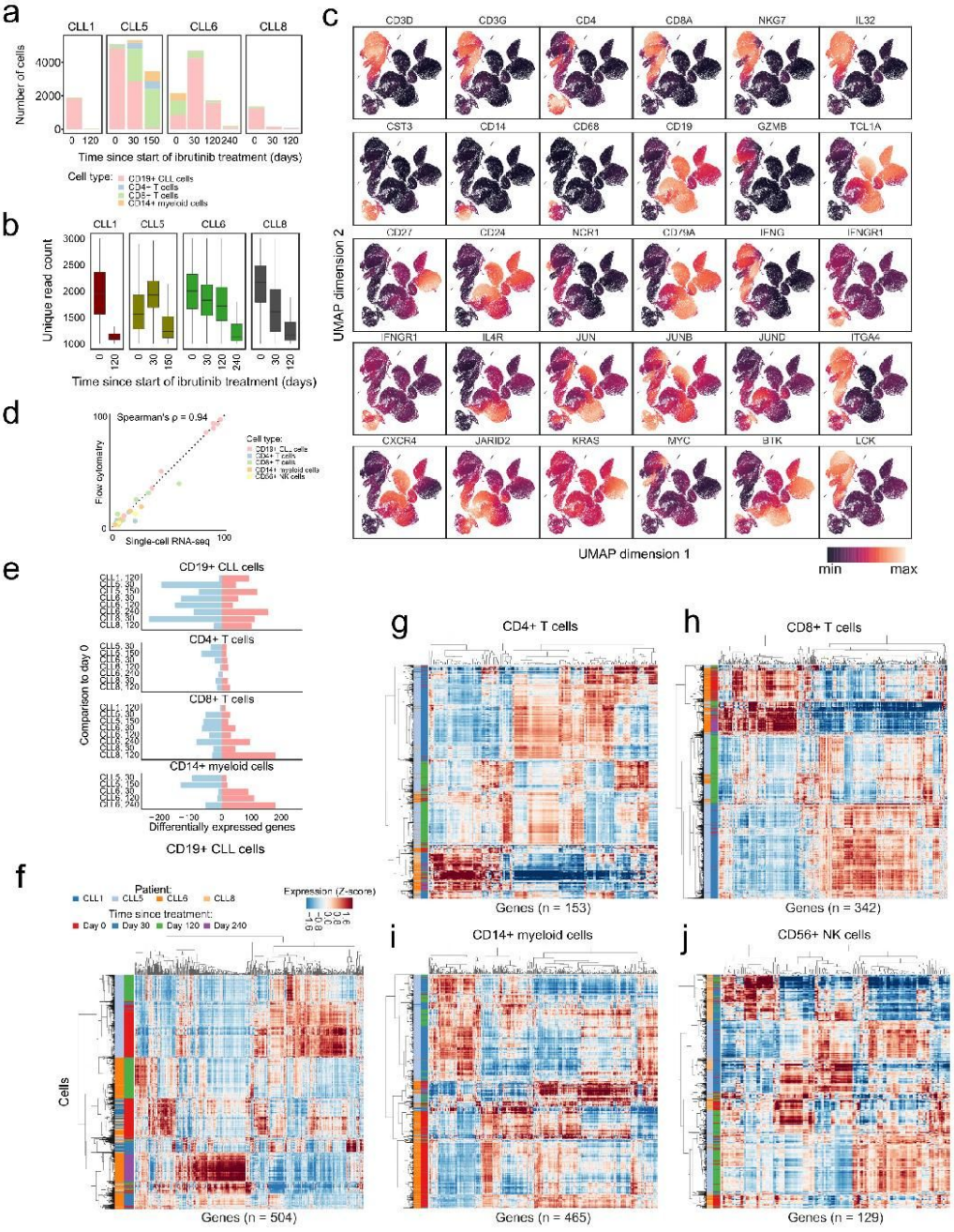
Supplementary Figure 1: Temporal dynamics of cell composition in CLL patients upon ibrutinib treatment

a) Schematic representation of the FACS approach for purifying CLL cells and five non-malignant immune cell types from PBMCs of CLL patients. **b-c)** Flow cytometry based quantification of the relative (b) or absolute (c) abundance of CLL cells and several non-malignant immune cell types in patients undergoing ibrutinib therapy.



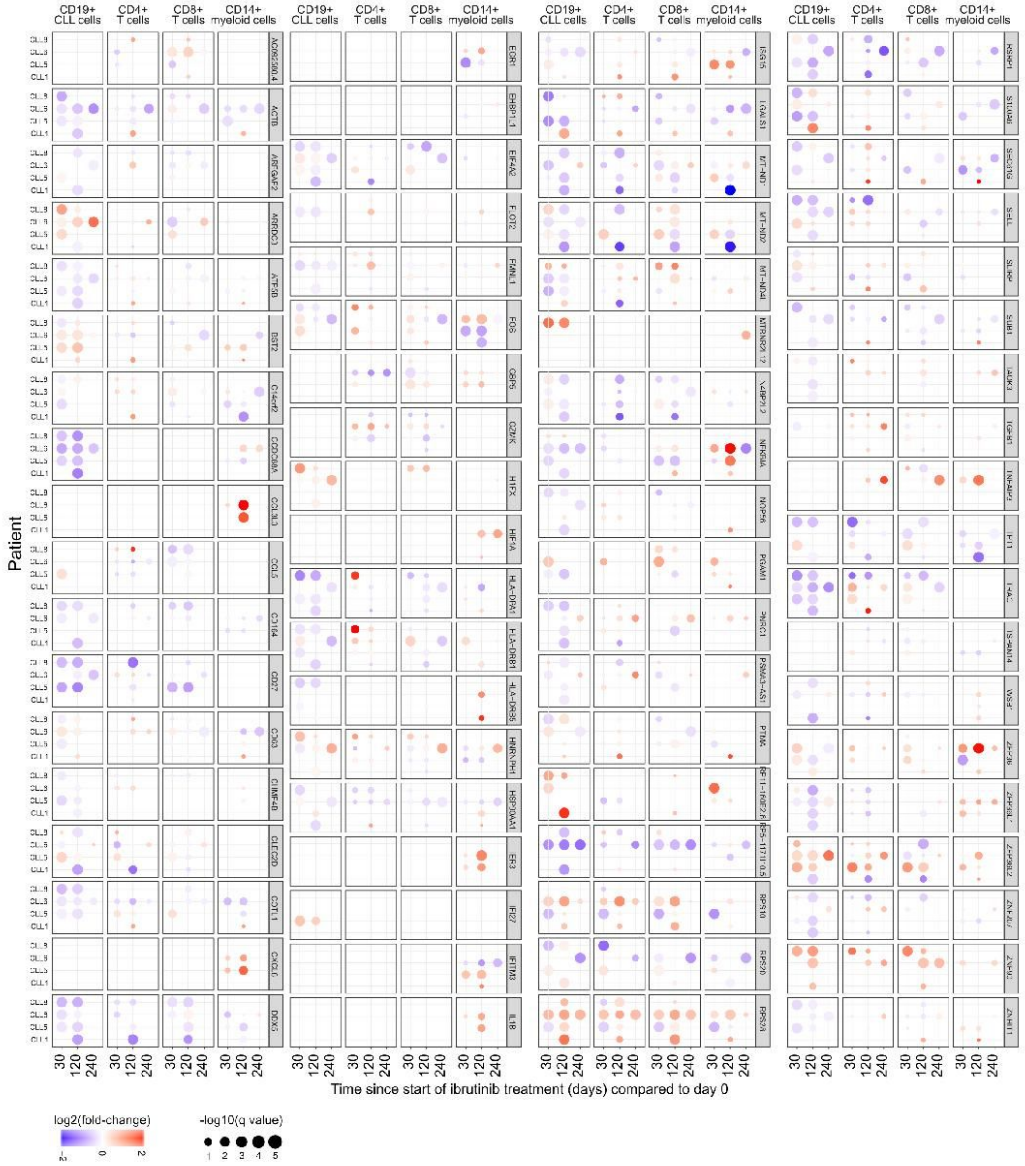
Supplementary Figure 2: Temporal dynamics of T cell subsets and surface marker expression

a) Mean expression of surface marker proteins in $CD19^{+}CD5^{+}$ CLL cells during ibrutinib treatment. The left panel displays absolute (log scale) expression, while the right panel displays column-wise Z-transformed values. Stars mark significant changes compared to time 0 (paired t-test, $p < 0.05$). **b)** Expression of surface marker proteins in immune cell subsets of CLL patients as measured by flow cytometry.



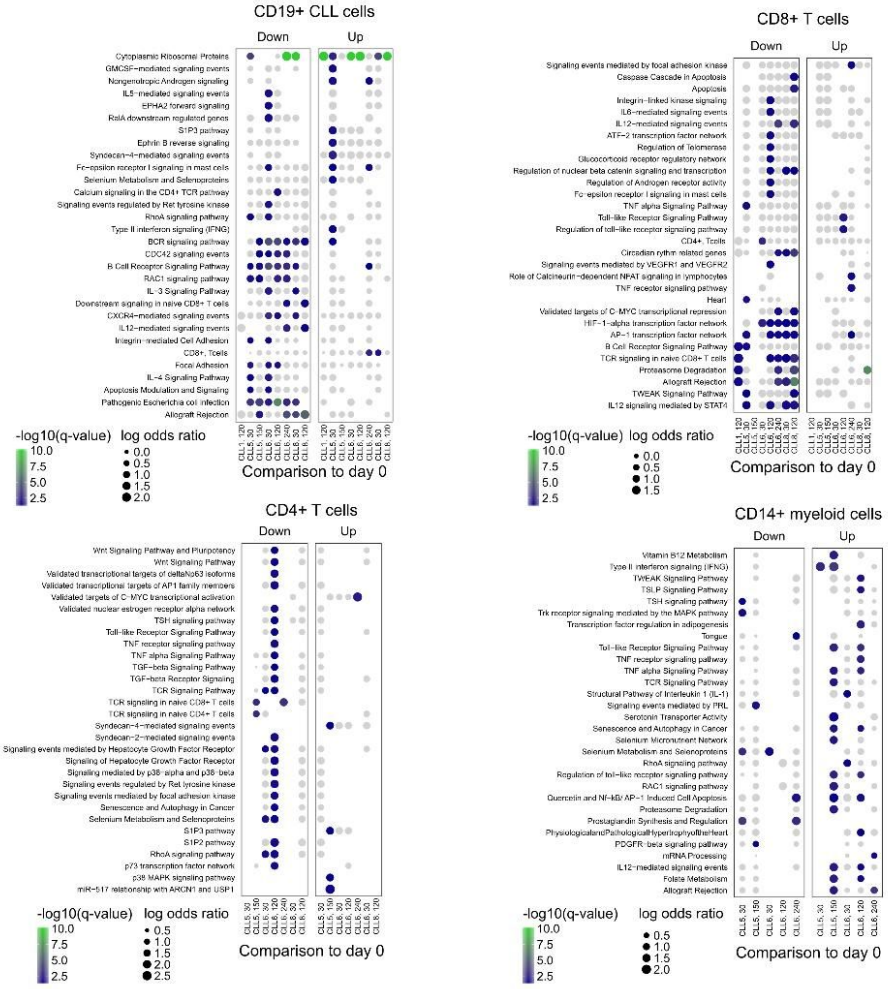
Supplementary Figure 3: Single-cell RNA-seq profiling over the ibrutinib time course

a) Bar plot displaying the number of single-cell transcriptome profiles that passed quality control, shown separately for each patient, cell type, and time point. **b)** Box plot displaying the number of unique molecular identifiers (UMIs) detected per single cell, shown separately for each patient, cell type, and time point. **c)** Heatmap showing mean expression levels of the marker genes that were used to assign the single-cell transcriptomes to defined cell types. Values represent expression levels (normalized UMI counts) scaled from minimum to maximum in each row. **d)** Scatterplot comparing the fraction of cells of each type based on single-cell RNA-seq versus flow cytometry across all patients and time points. **e)** Number of differentially expressed genes for each cell type, patient, and time point. **f-j)** Heatmaps of differentially expressed genes over the ibrutinib time course separately for each cell type.



Supplementary Figure 4: Transcriptional changes of differentially expressed genes upon ibrutinib treatment

Differential gene expression for those genes that were significantly differentially expressed (absolute log fold change >1 in more than two patients) over the course of ibrutinib therapy, shown separately for each cell type and comparing each sample to the pre-treatment (day 0) sample from the same patient.

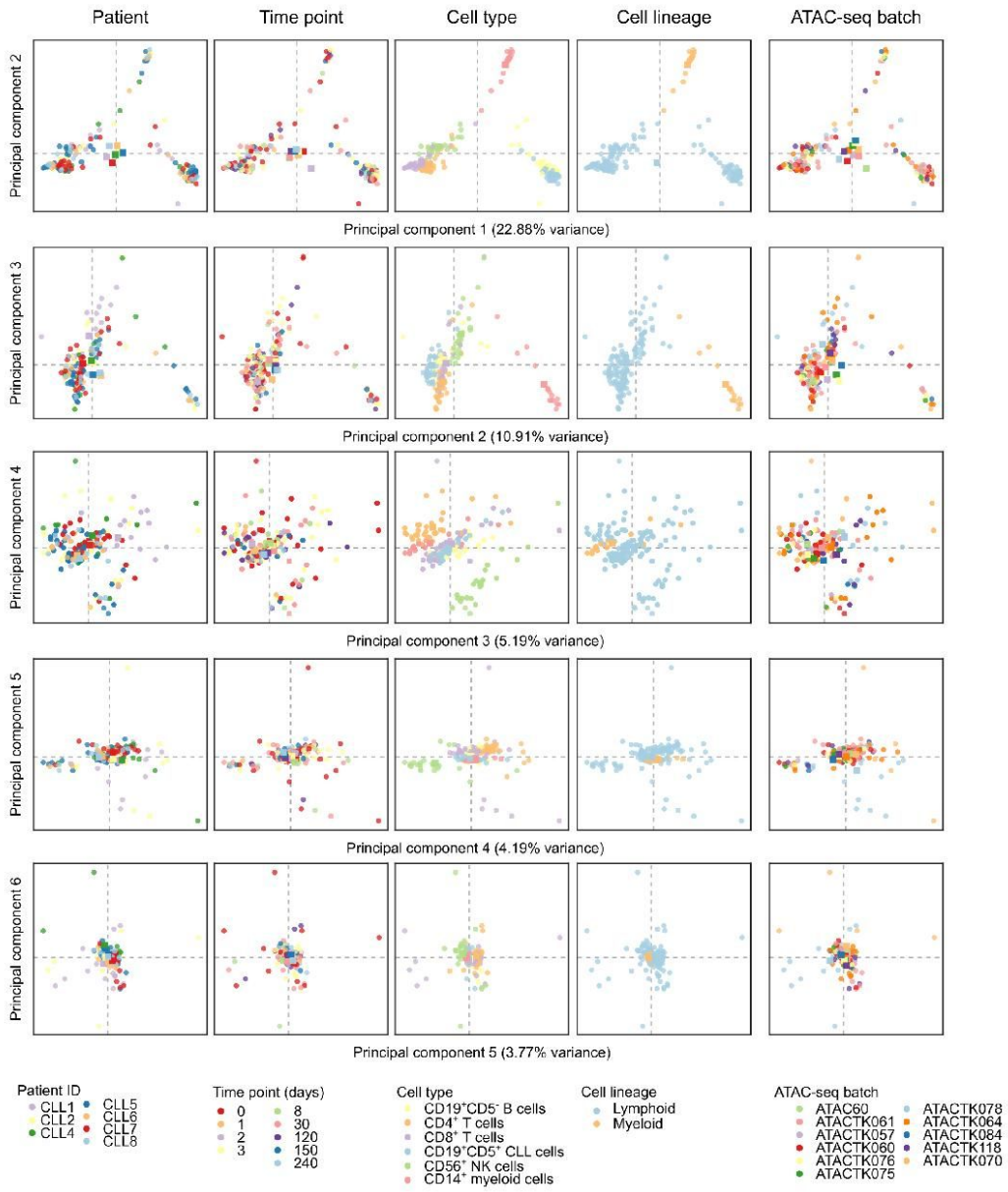


Supplementary Figure 5: Gene set enrichments for single-cell RNA-seq over the ibrutinib time course

Enrichment of differentially expressed genes over the course of ibrutinib therapy for gene sets and biological processes involved in immune regulation, calculated separately for each cell type.

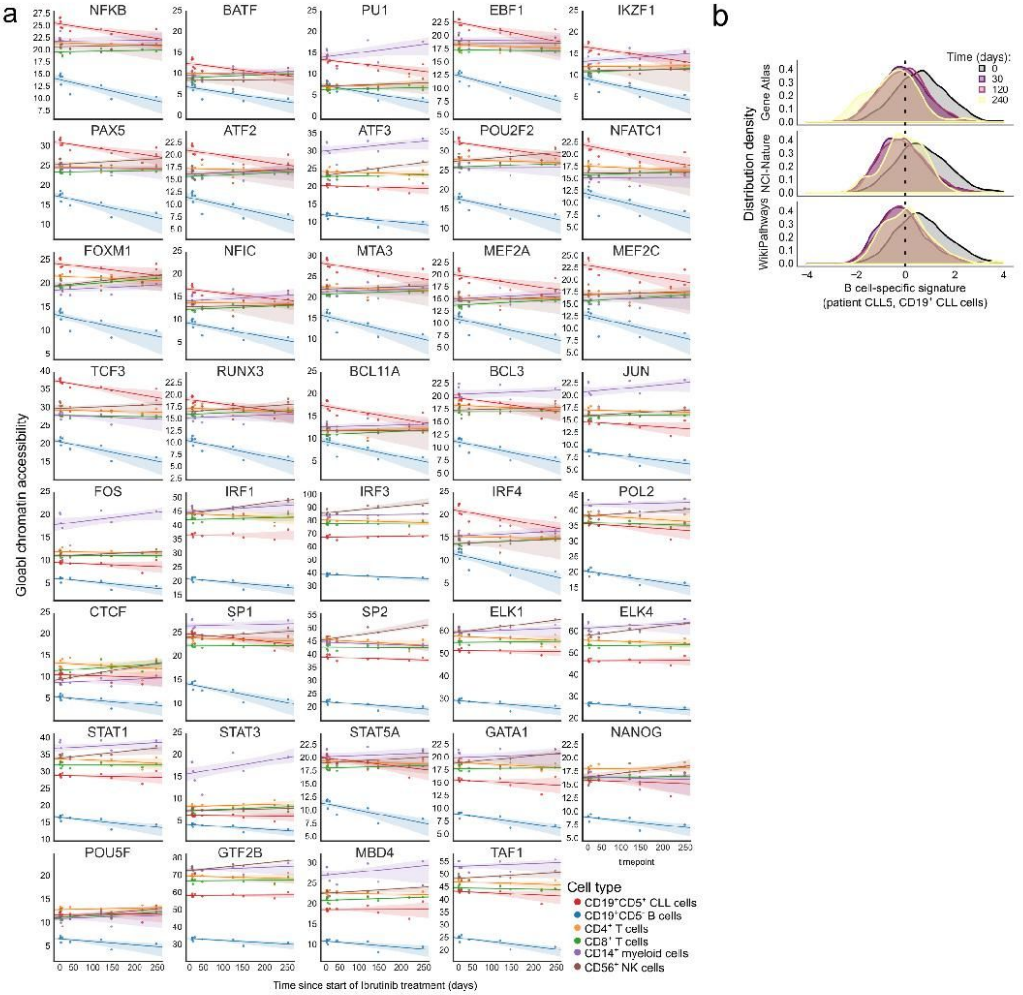
Supplementary Figure 6: Transcriptional changes of selected gene signatures upon ibrutinib treatment

Aggregate gene expression of selected gene signatures plotted over the ibrutinib time course. The list of gene signatures includes the hallmark signatures from MSigDB[56] (indicated by HM prefix) as well as sets of target genes for selected transcription factors (obtained from various sources).



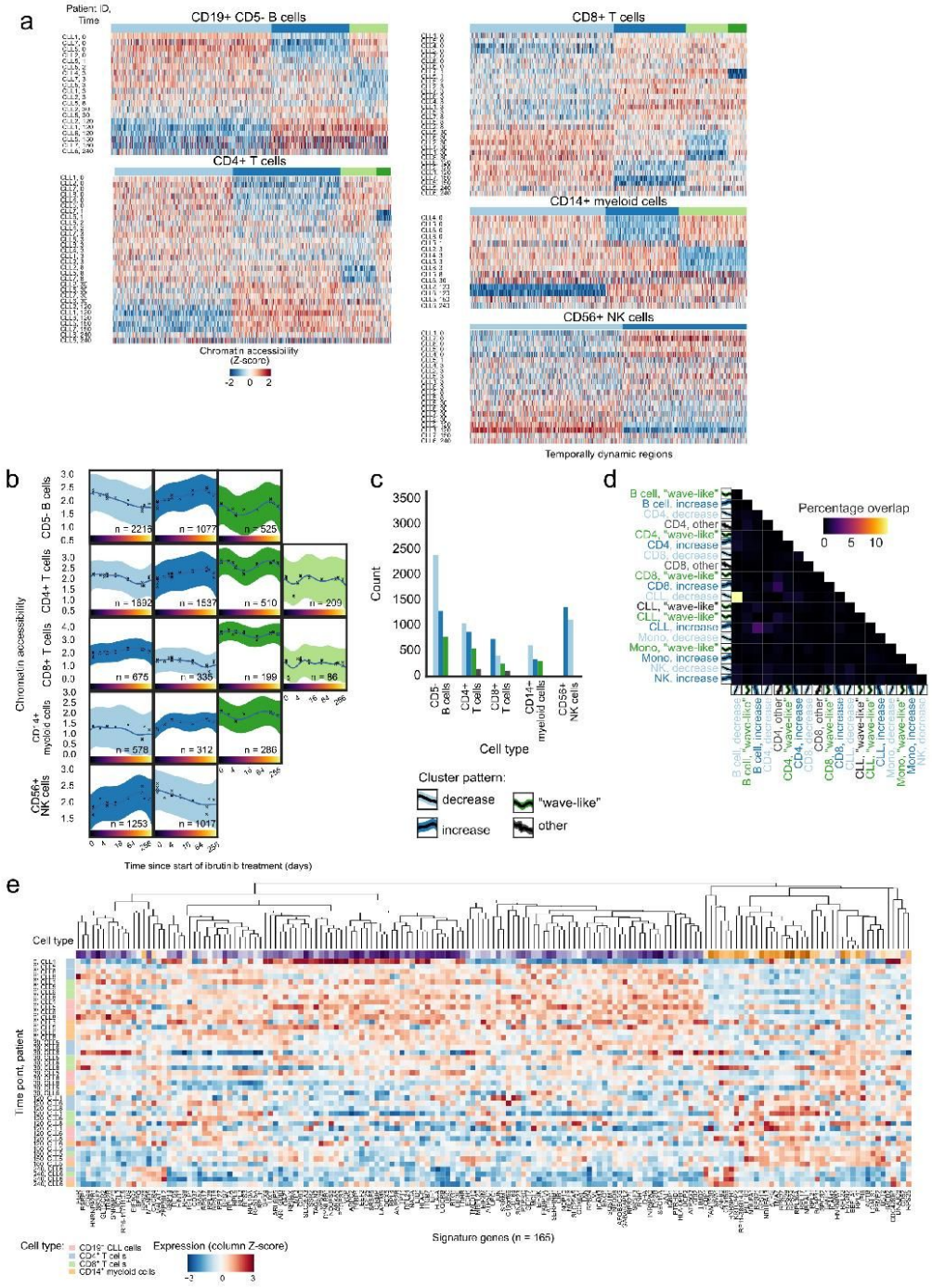
Supplementary Figure 7: Unsupervised analysis of chromatin accessibility over the ibrutinib time course

Principal component analysis of all chromatin accessibility profiles, highlighting biological and technical annotations of potential relevance. Samples are shown as circles, color-coded according to the shown annotations, and the centroid for each annotation is shown as a color-coded square.



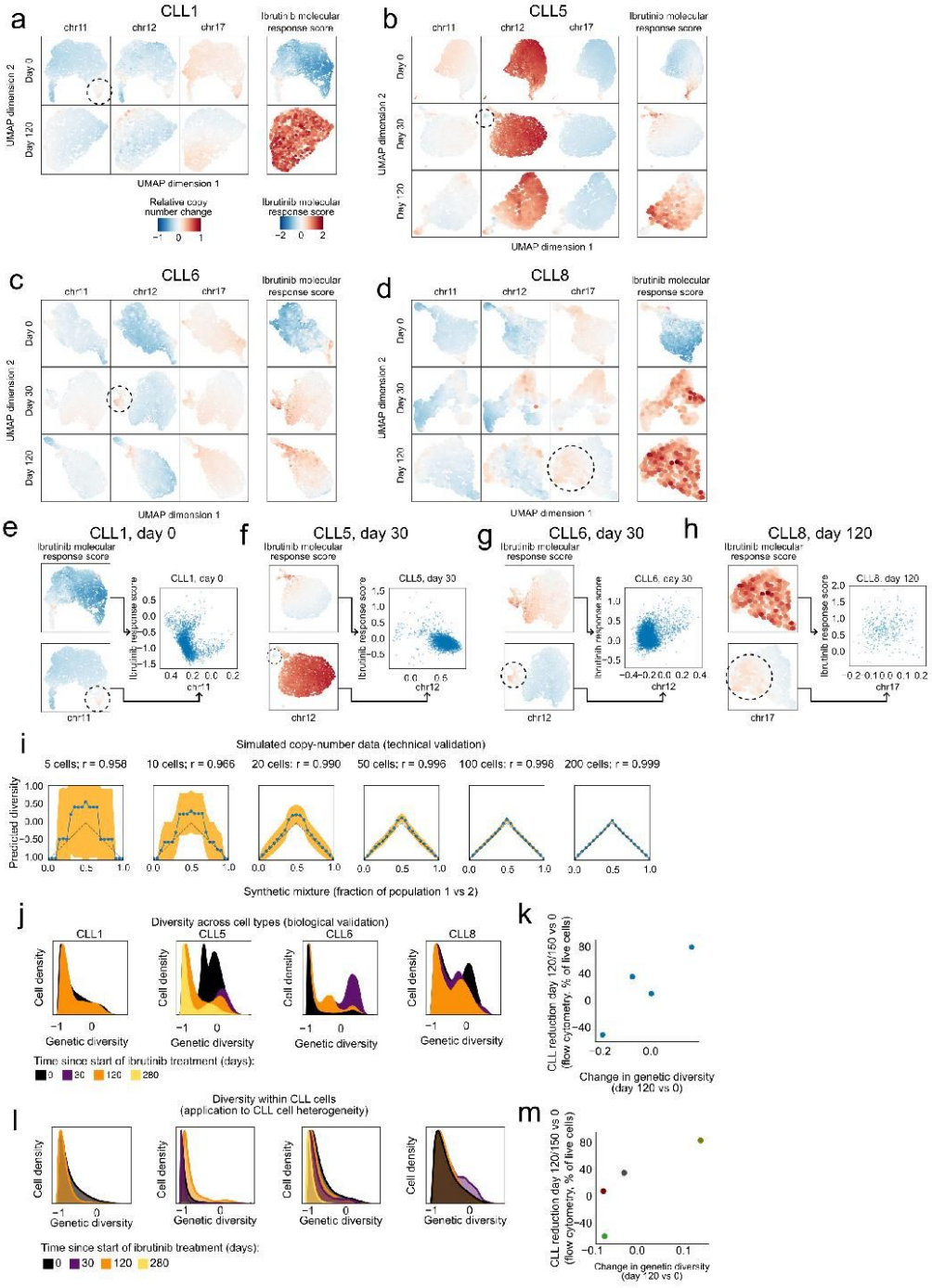
Supplementary Figure 8: Changes in transcription regulation and cell state over the ibrutinib time course

a) Line plots showing mean chromatin accessibility of regulatory regions overlapping putative binding sites of the respective transcription factors (based on publicly available ChIP-seq data) for each cell type and time point. Colored areas indicate 95 percent confidence intervals calculated over 1,000 bootstrap runs. **b)** Gene expression histogram across CLL cells in one patient, demonstrating the decline of B cell-specific expression signature (three alternative signatures are shown) over the time course of ibrutinib treatment. For illustration, the patient with most time points in the single-cell RNA-seq analysis (CLL5) is displayed.



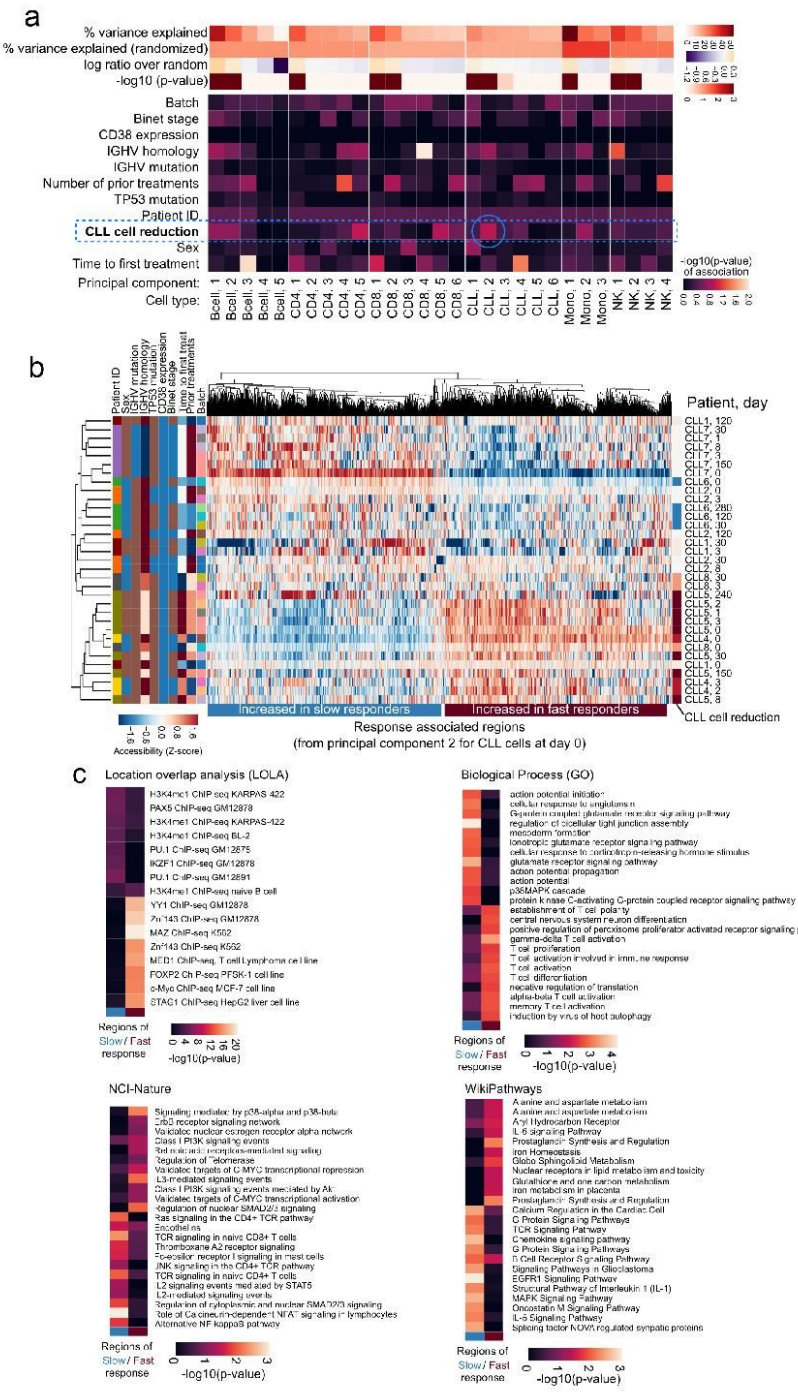
Supplementary Figure 9: Cluster analysis of regulatory regions in non-malignant immune cell types

a) Heatmaps showing chromatin accessibility at dynamically changing regulatory regions for five FACS-purified non-malignant immune cell types collected over the ibrutinib time course. Values represent column Z-scores of normalized ATAC-seq signal strength. **b)** Mean chromatin accessibility across patients plotted over the ibrutinib time course for each cluster of dynamically changing regulatory regions in each cell type. Each cross represents a single sample from a single patient at a specific time point, and 95% confidence intervals are shown as colored shapes. **c)** Absolute number of dynamic regulatory regions for each cell type and cluster. **d)** Pairwise overlap of dynamic regulatory regions between cell types and clusters. **e)** Clustered heatmap showing patient-specific gene expression levels for the quiescence-like gene expression signature (**Figure 3e**), based on the single-cell RNA-seq data over the ibrutinib time course. Values represent column Z-scores of gene expression.



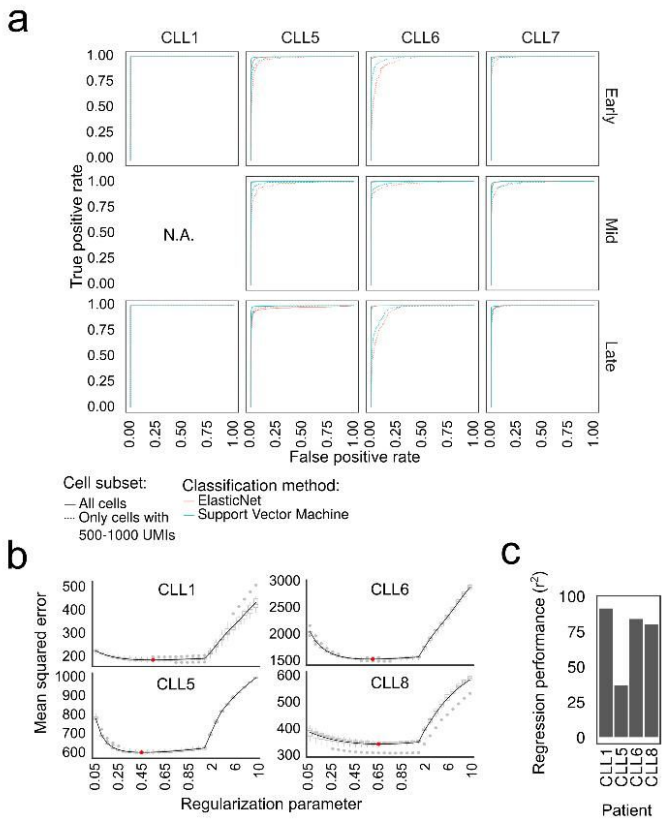
Supplementary Figure 10: Analysis of copy number and genetic diversity over the ibrutinib time course

a-d) Two-dimensional similarity map (UMAP projection) based on DNA copy number profiles for single cells inferred from the single-cell RNA-seq data. These maps were calculated separately for each patient and time point. Color-coding indicates the relative copy number change for three chromosomal aberrations common in CLL (left) and for the ibrutinib molecular response score (right), i.e., the change in the CLL cell percentage on day 120/150 of ibrutinib treatment compared to day 0 as measured by flow cytometry. Genetically distinct subclones are highlighted by dashed circles. **e-h)** Scatterplots comparing selected subclonal copy number aberrations (x-axis and UMAP plots on the bottom left) with the ibrutinib molecular response score (y-axis and UMAP plots on the top left) across single cells in individual patients and time points. **i)** Accuracy of the computational approach for quantifying genetic diversity benchmarked on simulated copy number profiles that were combined at defined percentages (x-axis). Dashed lines indicate expected values (based on the simulation's known ground truth), blue lines indicate inferred values, and yellow areas represents 95th confidence intervals for the inferred values. Correlation coefficients quantify the overall agreement between expected and inferred values. **j)** Histograms showing the change in genetic diversity across all cells (i.e., CLL cells and immune cells) in the single-cell RNA-seq dataset. **k**) Scatterplot showing the correlation between the change in genetic diversity across all cells (x-axis) between time points and the cellular response to ibrutinib treatment (y-axis). **l**) Histograms showing the change in genetic diversity specifically for CLL cells. **m**) Scatterplot showing the correlation between changes in genetic diversity specifically in CLL cells (x-axis) and the cellular response to ibrutinib treatment (y-axis). Panel m is a reproduction of Figure 4b for consistency with panels j-k.



Supplementary Figure 11: Analysis of chromatin profiles and their association with the response to ibrutinib

a) Heatmaps showing the association of various clinical annotations with the principal components of the cell type specific chromatin accessibility profiles of different cell types prior to the start of ibrutinib treatment. The blue circle highlights the association between the second principal component for CLL cells and the change in the CLL cell percentage on day 120/150 of ibrutinib treatment compared to day 0, as measured by flow cytometry (y-axis). **b)** Clustered heatmap showing patient-specific chromatin profiles for genomic regions associated with the second principal component (from panel a). **c)** Enrichment analysis for genomic regions associated with the second principal component, separately for regions associated with a slow versus a fast response to ibrutinib treatment.



Supplementary Figure 12: Prediction of the time point of sample collection from single-cell transcriptomes

a) ROC curves showing the cross-validated test set performance of classifiers predicting the time point of sample collection based on single-cell transcriptome profiles, using two different machine learning methods (logistic regression with elastic net regularization and support vector machines) and two different thresholds for single-cell RNA-seq data quality (all cells vs. only cells with 500 to 1,000 UMIs). **b)** Optimization of the regularization parameter (lambda) for predicting the time since the start of ibrutinib treatment using elastic net regularized linear regression. Red dots indicate the chosen parameter for each patient. **c)** Cross-validated test set performance of the regression models (coefficient of determination) for predicting the time since the start of ibrutinib treatment for each patient.

Supplementary table legends

Supplementary Table 1: Clinical annotation of the CLL patients included in the time course analysis

Supplementary Table 2: Cell type composition over the time course as measured by flow cytometry

Supplementary Table 3: Expression of cell surface marker proteins as measured by flow cytometry

Supplementary Table 4: Gene expression of individual cell types (single-cell RNA-seq) over the time course

Supplementary Table 5: Differentially expressed genes for individual cell types over the time course

Supplementary Table 6: Summary statistics for ATAC-seq chromatin mapping in CLL cells

Supplementary Table 7: Dynamic chromatin regions over the time course in CLL cells

Supplementary Table 8: Summary statistics for ATAC-seq chromatin mapping in immune cell types

Supplementary Table 9: Dynamic chromatin regions over the time course in immune cell type

References

1 Byrd, J. C., Stilgenbauer, S. & Flinn, I. W. Chronic lymphocytic leukemia. *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program*, 163--183, doi:10.1182/asheducation-2004.1.163 (2004).

2 Stevenson, F. K., Krysov, S., Davies, A. J., Steele, A. J. & Packham, G. B-cell receptor signaling in chronic lymphocytic leukemia. *Blood* **118**, 4313-4320, doi:10.1182/blood-2011-06-338855 (2011).

3 Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101--105, doi:10.1038/nature10113 (2011).

4 Quesada, V. *et al.* Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature Genetics* **44**, 47--52, doi:10.1038/ng.1032 (2011).

5 Puente, X. S. & Be. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*, doi:10.1038/nature14666 (2015).

6 Quesada, V. *et al.* The genomic landscape of chronic lymphocytic leukemia: clinical implications. *BMC Med* **11**, 124, doi:10.1186/1741-7015-11-124 (2013).

7 Klein, U. *et al.* Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *The Journal of experimental medicine* **194**, 1625--1638, doi:10.1084/jem.194.11.1625 (2001).

8 Rosenwald, A. *et al.* Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *The Journal of experimental medicine* **194**, 1639--1647, doi:10.1084/jem.194.11.1639 (2001).

9 Ferreira, P. G. *et al.* Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Research* **24**, 212--226, doi:10.1101/gr.152132.112 (2014).

10 Kulis, M. *et al.* Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nature Genetics* **44**, 1236--1242, doi:10.1038/ng.2443 (2012).

11 Oakes, C. C. *et al.* DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nature genetics*, doi:10.1038/ng.3488 (2016).

12 Oakes, C. C. *et al.* Evolution of DNA Methylation Is Linked to Genetic Aberrations in Chronic Lymphocytic Leukemia. *Cancer Discovery* **4**, 348--361, doi:10.1158/2159-8290.CD-13-0349 (2014).

13 Rendeiro, A. F. *et al.* Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. *Nature communications* **7**, 11938, doi:10.1038/ncomms11938 (2016).

14 Byrd, J. C., O'Brien, S. & James, D. F. Ibrutinib in relapsed chronic lymphocytic leukemia. *N Engl J Med* **369**, 1278-1279, doi:10.1056/NEJMc1309710 (2013).

15 O'Brien, S. *et al.* Single-Agent Ibrutinib in Treatment-Naive and Relapsed/Refractory Chronic Lymphocytic Leukemia: A 5-Year Experience. *Blood*, doi:10.1182/blood-2017-10-810044 (2018).

16 O'Brien, S. *et al.* Ibrutinib as initial therapy for elderly patients with chronic lymphocytic leukaemia or small lymphocytic lymphoma: an open-label, multicentre, phase 1b/2 trial. *Lancet Oncol* **15**, 48-58, doi:10.1016/S1470-2045(13)70513-8 (2014).

17 Kondo, K. *et al.* Ibrutinib modulates the immunosuppressive CLL microenvironment through STAT3-mediated suppression of regulatory B-cell function and inhibition of the PD-1/PD-L1 pathway. *Leukemia* **32**, 960-970, doi:10.1038/leu.2017.304 (2018).

18 Burger, J. A. *et al.* Safety and activity of ibrutinib plus rituximab for patients with high-risk chronic lymphocytic leukaemia: a single-arm, phase 2 study. *Lancet Oncol* **15**, 1090-1099, doi:10.1016/S1470-2045(14)70335-3 (2014).

19 Herman, S. E. *et al.* Ibrutinib inhibits BCR and NF-kappaB signaling and reduces tumor proliferation in tissue-resident cells of patients with CLL. *Blood* **123**, 3286-3295, doi:10.1182/blood-2014-02-548610 (2014).

20 Landau, D. A. *et al.* The evolutionary landscape of chronic lymphocytic leukemia treated with ibrutinib targeted therapy. *Nature Communications* **8**, doi:10.1038/s41467-017-02329-y (2017).

21 Ponader, S. *et al.* The Bruton tyrosine kinase inhibitor PCI-32765 thwarts chronic lymphocytic leukemia cell survival and tissue homing in vitro and in vivo. *Blood* **119**, 1182-1189, doi:10.1182/blood-2011-10-386417 (2012).

22 Woyach, J. A. *et al.* Prolonged lymphocytosis during ibrutinib therapy is associated with distinct molecular characteristics and does not indicate a suboptimal response to therapy. *Blood* **123**, 1810-1817, doi:10.1182/blood-2013-09-527853 (2014).

23 Long, M. *et al.* Ibrutinib treatment improves T cell number and function in CLL patients. *J Clin Invest* **127**, 3052-3064, doi:10.1172/JCI89756 (2017).

24 Sagiv-Barfi, I. *et al.* Therapeutic antitumor immunity by checkpoint blockade is enhanced by ibrutinib, an inhibitor of both BTK and ITK. *Proc Natl Acad Sci U S A* **112**, E966-972, doi:10.1073/pnas.1500712112 (2015).

25 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* **10**, 1213--1218, doi:10.1038/nmeth.2688 (2013).

26 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049, doi:10.1038/ncomms14049 (2017).

27 Vojdeman, F. J. *et al.* Soluble CD52 is an indicator of disease activity in chronic lymphocytic leukemia. *Leuk Lymphoma* **58**, 2356-2362, doi:10.1080/10428194.2017.1285027 (2017).

28 Borst, J., Hendriks, J. & Xiao, Y. CD27 and CD70 in T cell and B cell activation. *Curr Opin Immunol* **17**, 275-281, doi:10.1016/j.coi.2005.04.004 (2005).

29 Rasmussen, C. E. & Williams, C. K. I. *Gaussian processes for machine learning*. (MIT Press, 2006).

30 Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587-589, doi:10.1093/bioinformatics/btv612 (2016).

31 Bonadies, N. *et al.* PU.1 is regulated by NF-kappaB through a novel binding site in a 17 kb upstream enhancer element. *Oncogene* **29**, 1062-1072, doi:10.1038/onc.2009.371 (2010).

32 Grumont, R. J. & Gerondakis, S. Rel induces interferon regulatory factor 4 (IRF-4) expression in lymphocytes: modulation of interferon-regulated gene expression by rel/nuclear factor kappaB. *J Exp Med* **191**, 1281-1292 (2000).

33 Saito, M. *et al.* A signaling pathway mediating downregulation of BCL6 in germinal center B cells is blocked by BCL6 gene alterations in B cell lymphoma (vol 12, pg 280, 2007). *Cancer Cell* **12**, 403-403, doi:DOI 10.1016/j.ccr.2007.09.025 (2007).

34 Kaszubska, W. *et al.* Cyclic AMP-independent ATF family members interact with NF-kappa B and function in the activation of the E-selectin promoter in response to cytokines. *Mol Cell Biol* **13**, 7180-7190 (1993).

35 Nie, Y., Han, Y. C. & Zou, Y. R. CXCR4 is required for the quiescence of primitive hematopoietic cells. *J Exp Med* **205**, 777-783, doi:10.1084/jem.20072513 (2008).

36 Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* **25**, 977-988, doi:10.1016/j.immuni.2006.10.016 (2006).

37 Galloway, A. *et al.* RNA-binding proteins ZFP36L1 and ZFP36L2 promote cell quiescence. *Science* **352**, 453-459, doi:10.1126/science.aad5978 (2016).

38 Aird, K. M. *et al.* HMGB2 orchestrates the chromatin landscape of senescence-associated secretory phenotype gene loci. *J Cell Biol* **215**, 325-334, doi:10.1083/jcb.201608026 (2016).

39 Byrd, J. C. *et al.* Acalabrutinib (ACP-196) in Relapsed Chronic Lymphocytic Leukemia. *N Engl J Med* **374**, 323-332, doi:10.1056/NEJMoa1509981 (2016).

40 ten Hacken, E. & Burger, J. A. Microenvironment dependency in Chronic Lymphocytic Leukemia: The basis for new targeted therapies. *Pharmacol Ther* **144**, 338-348, doi:10.1016/j.pharmthera.2014.07.003 (2014).

41 Ghez, D. *et al.* Early-onset invasive aspergillosis and other fungal infections in patients treated with ibrutinib. *Blood* **131**, 1955-1959, doi:10.1182/blood-2017-11-818286 (2018).

42 Tillman, B. F., Pauff, J. M., Satyanarayana, G., Talbott, M. & Warner, J. L. Systematic review of infectious events with the Bruton tyrosine kinase inhibitor ibrutinib in the treatment of hematologic malignancies. *Eur J Haematol* **100**, 325-334, doi:10.1111/ejh.13020 (2018).

43 Varughese, T. *et al.* Serious Infections in Patients Receiving Ibrutinib for Treatment of Lymphoid Cancer. *Clin Infect Dis* **67**, 687-692, doi:10.1093/cid/ciy175 (2018).

44 Pettersen, R. D., Bernard, G., Olafsen, M. K., Pourtein, M. & Lie, S. O. CD99 signals caspase-independent T cell death. *J Immunol* **166**, 4931-4942 (2001).

45 Jung, K. C., Kim, N. H., Park, W. S., Park, S. H. & Bae, Y. The CD99 signal enhances Fas-mediated apoptosis in the human leukemic cell line, Jurkat. *FEBS Lett* **554**, 478-484 (2003).

46 Chen, S. S. *et al.* BTK inhibition results in impaired CXCR4 chemokine receptor surface expression, signaling and function in chronic lymphocytic leukemia. *Leukemia* **30**, 833-843, doi:10.1038/leu.2015.316 (2016).

47 Kipps, T. J. *et al.* Integrated Analysis: Outcomes of ibrutinib-related patients with Chronic Lymphocytic Leukemia/Small Lymphocytic Leukemia (CLL/SLL) with high-risk prognostic factors. *Hematological Oncology* **35**, 109-111, doi:doi:10.1002/hon.2437_99 (2017).

48 O'Brien, S. *et al.* Single-agent ibrutinib in treatment-naive and relapsed/refractory chronic lymphocytic leukemia: a 5-year experience. *Blood* **131**, 1910-1919, doi:10.1182/blood-2017-10-810044 (2018).

49 Fraietta, J. A. *et al.* Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia. *Nat Med* **24**, 563-571, doi:10.1038/s41591-018-0010-1 (2018).

50 Granger, C. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control* **2**, 329-352 (1980).

51 Granger, C. W. J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **37**, 424-438, doi:10.2307/1912791 (1969).

52 Hallek, M. *et al.* Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood* **111**, 5446-5456, doi:10.1182/blood-2007-06-093906 (2008).

53 Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* **48**, 1193-1203, doi:10.1038/ng.3646 (2016).

54 Satija, R., Farrell, J. a., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology* **33**, doi:10.1038/nbt.3192 (2015).

55 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).

56 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).

57 Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S. & Theis, F. J. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun* **10**, 390, doi:10.1038/s41467-018-07931-2 (2019).

58 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15, doi:10.1186/s13059-017-1382-0 (2018).

59 Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC bioinformatics* **15**, 182, doi:10.1186/1471-2105-15-182 (2014).

60 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357--359, doi:10.1038/nmeth.1923 (2012).

61 Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032--2034, doi:10.1093/bioinformatics/btv098 (2015).

62 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

63 Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57--74, doi:10.1038/nature11247 (2012).

64 Kalaitzis, A. A. & Lawrence, N. D. A simple approach to ranking differentially expressed gene expression time courses through Gaussian process regression. *BMC Bioinformatics* **12**, 180, doi:10.1186/1471-2105-12-180 (2011).

65 Hensman, J., Lawrence, N. D. & Rattray, M. Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters. *BMC Bioinformatics* **14**, 252, doi:10.1186/1471-2105-14-252 (2013).

66 Macaulay, I. C. *et al.* Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep* **14**, 966-977, doi:10.1016/j.celrep.2015.12.082 (2016).

67 Sheffield, N. C. *et al.* Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Research* **23**, 777--788, doi:10.1101/gr.152140.112 (2013).

68 Snchez-Castillo, M. *et al.* CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic acids research* **43**, D1117--1123, doi:10.1093/nar/gku895 (2015).

69 Rosenbloom, K. R. *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Research* **43**, D670--D681, doi:10.1093/nar/gku1177 (2015).

70 Liu, T. *et al.* Cistrome: an integrative platform for transcriptional regulation studies. *Genome biology* **12**, R83, doi:10.1186/gb-2011-12-8-r83 (2011).

71 Adams, D. *et al.* BLUEPRINT to decode the epigenetic signature written in blood. *Nature Biotechnology* **30**, 224--226, doi:10.1038/nbt.2153 (2012).

72 Heinz, S. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576--589, doi:10.1016/j.molcel.2010.05.004.Simple (2011).

Discussion

General discussion of results

A landscape of chromatin accessibility for the stratification of CLL

The large map of chromatin accessibility produced allowed us to uncover a previously unknown dynamic layer of regulation in cancer cells between different patient samples. The observation that distal regulatory elements of genes related to B cell functions tend to have higher variance in chromatin accessibility than their promoters highlights the high complexity of the human genome and reveals a layer of genetic regulation that warrants further investigation. In that view, the inference of gene regulatory networks specific to disease subtypes differing on IGHV mutation status has brought to light an intricate layer of the regulatory landscape mediated by the binding of transcription factors to DNA. While we uncovered a tight core network active across CLL lymphocytes, the comparison of the subtypes of the network highlighted differential transcription factor activity and its downstream effects for many genes with functions in cellular proliferation, signaling and immune processes, related to specific differences between the two IGHV groups.

The observation that the IGHV mutation status dominated the landscape of variation of CLL comes in agreement with previous observations in other individual layers of gene regulation such as DNA methylation (Kulis *et al*, 2012, 2015; Oakes *et al*, 2016) and transcription (Klein *et al*, 2001; Rosenwald *et al*, 2001; Ferreira *et al*, 2014), and more recently across layers of the regulatory landscape such as chromosomal conformation (Beekman *et al*, 2018) or profiles of drug sensitivity (Argelaguet *et al*, 2018). The impact of IGHV mutation and the likely differentiation state of CLL cells has not only a strong impact from the regulatory perspective but also clinically, as it is an independent prognostic marker for CLL (Hallek *et al*, 2018) affecting progression-free and overall survival even in the context of modern targeted treatment (Farooqui *et al*, 2015).

The importance of IGHV for prognostication of CLL, the fact that it had such a clear effect on the chromatin of CLL lymphocytes, along with being a routinely collected clinical variable of tangible importance made IGHV status an ideal case for the application of machine learning. Our Random Forest classifier approach was successful in learning highly predictive feature weights that could both immediately reveal insights into the underlying biological wiring of the two CLL subtypes while also providing proof-of-principle of the use of chromatin accessibility in the classification of patient samples for differential diagnosis and prognostication through patient stratification.

Taking together, our analytical framework revealed itself a promising and generalizable approach to balance discovery and interpretation of high dimensional data in precision medicine. While the

unsupervised learning of a reduced chromatin space jointly for all patient samples seeks to identify all important factors governing variation between samples, the supervised learning of clinical variables lends itself to biological interpretation and generates a model capable of classifying unseen samples in the future.

Our study of the chromatin landscape in CLL therefore demonstrates the technical feasibility of profiling large numbers of primary patient samples previously cryopreserved and biobanked thereby opening the prospect of using chromatin profiling assays for clinical tasks such as diagnosis, patient stratification, treatment recommendation and disease monitoring.

Deriving markers and models for the monitoring of CLL treatment

The goal of our analysis of immunological, regulatory and transcriptional changes of six immune cell types of patients undergoing ibrutinib therapy, was to understand the effects of this targeted treatment on the immune system and to derive computational models to predict and monitor the response to therapy over time. For this reason, we employed unprecedented depth (provided by the multiple orthogonal assays, particularly the single cell transcriptome) and breadth (in the number of profiled cell types and timepoints) in the characterization of the dynamic response to therapy.

The transcriptional landscape of ibrutinib treated CLL was characterized by large changes in transcription of genes related with pro-proliferative signaling pathways. In addition to bioinformatic gene expression analysis for single cells, we used the dataset to infer copy number variants in single cells. While we discovered interesting changes in the abundance of cells carrying specific chromosomal aberrations within patients along time, these did not always display a transcriptional profile that showed evidence of adaptation to ibrutinib treatment. This is likely related to the fact that the diversity of cells under analysis is rather low compared with the number of cells in the patients and likely to suffer from grave undersampling, which highlights the need for assays that can capture even higher number of single cells per sampling point per patient.

The analysis of chromatin accessibility changes during ibrutinib treatment of CLL revealed early reduction of inferred NFKB binding to chromatin, consistent with its established role in activating gene expression following BCR stimulation. The dense temporal profiling along the course of treatment allowed us to uncover of changes over the course of treatment, where we observed massive downregulation of regulatory elements usually bound by transcription factors that have important functions in establishing and maintaining B cell identify. The fact that inhibition of BCR leads to erosion of cellular identity in primary human lymphocytes, speaks to the tight connection between cellular signaling and maintenance of cellular identity through genetic regulation,

something that we also found reinforced by analyzing the expression of B cell and CLL-specific surface protein markers.

Our comprehensive analysis of transcriptional and regulatory changes across multiple cell types showed specific changes as well as shared patterns. In particular, a quiescence-like signature comprising 165 genes was upregulated over time in CLL cells, but also in other cell types, with CD8 T cells being the second-most affected. While the cause of such general cross-cell type effect are likely due to off-target effect of ibrutinib on other TEK-family kinases, the biological and clinical implications remain to be explored. For example, ibrutinib may be contributing to the impairment of the cytotoxic action of CD8 T cells on the CLL cells, particularly at a point where the immunosupressive phenotype of CLL cells is lowered. While the discovered signature was validated in an external cohort of CLL patients undergoing ibrutinib treatment, further investigation on the functional consequences of potential ibrutinib off-target effect healthy immune cell compartments is needed.

The discovery of a signature capable of quantifying the patients progression through the spectrum of molecular response to treatment, encouraged us to develop a regression model powered by robust regularization techniques in machine learning that leverages on the high dimensionality of the single cell dataset. While the model largely captured the global trend of response to ibrutinib across patients, it was also capable of positioning each patient in light of the other's position in the response spectrum, thus quantifying the relative difference between responses at a given sampled time. This difference corresponded to observed values of CLL cells reduction at day 120 of treatment, in an internal validation of the model.

This study highlights the power of high-dimensional assays to characterize changes in cancer cells and the immune system at a regulatory and transcriptional level but also the importance of longitudinal profiling particularly in chronic disease. The development of adequate computational methods to monitor and predict the response of each patient to treatment is probably its most direct contribution towards the development of multivariate markers and models for the monitoring of CLL.

Conclusions and future prospects

Our work establishes the use of large-scale high-dimensional assays in primary human samples of leukemia patients in a clear path towards the genome-aware, data-driven, personalized treatment and understanding of chronic disease.

In the future, for CLL as well as other chronic diseases, the availability of dense high-dimensional longitudinal data for high numbers of patients will enable the positioning of the patient as the unit

for learning as opposed to the contemporary situation where a specific sample at a given time is the unit. In that situation, the goal shifts towards learning the full trajectory of the patients based on longitudinal sampling as early as the time of diagnosis, through the disease progression, and treatment. Examples of similar approaches that leverage on large amounts of dense longitudinal data are rare but a few examples are appearing, albeit in the context of clinical trials (Zhou *et al*, 2019; Schüssler-Fiorenza Rose *et al*, 2019), and not routine clinical sampling.

Equally important for the success of such data gathering endeavors will be the aggregation of clinical metadata and the combination of orthogonal assays that profile for example several modalities of gene regulation and expression that characterize the current cellular states, but jointly rich phenotypes such as sensitivity to an array of drugs (Dietrich *et al*, 2018; Schmidl *et al*, 2019).

Pairing this wealth of dense longitudinal and high-dimensional data with computational models that have high predictive power at the same time being readily interpretable, while not unprecedented (Argelaguet *et al*, 2018), is however a challenge. If done in a generative and probabilistic framework (unsupervised or explicitly modeled), such models could generate potential outcomes for a future of a new patient with the intent of helping physicians have a genome-aware decision process that can reveal the reasoning of its decision based on tangible biological and clinically interpretable factors.

While this vision may take time to materialize, I consider the work of this thesis a step towards genome-centric, data-driven, personalized treatment and understanding of chronic disease.

References

Abken H, Koehler P, Schmidt P, Hombach AA & Hallek M (2012) Engineered T cells for the adoptive therapy of b-cell chronic lymphocytic leukaemia. *Adv. Hematol.* **2012:**

Akira S, Okazaki K & Sakano H (1987) Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. *Science (80-.).* **238:** 1134–1138 Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.3120312

Amin S, Walsh M, Wilson C, Parker AE, Oscier D, Willmore E, Mann D & Mann J (2012) Cross-talk between DNA methylation and active histone modifications regulates aberrant expression of ZAP70 in CLL. *J. Cell. Mol. Med.* **16:** 2074–2084

Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, Buettner F, Huber W & Stegle O (2018) Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **14:** 1–13 Available at: https://onlinelibrary.wiley.com/doi/abs/10.15252/msb.20178124

Autio K, Turunen O, Penttilä O, Erämaa E, de la Chapelle A & Schröder J (1979) Human chronic lymphocytic leukemia: Karyotypes in different lymphocyte populations. *Cancer Genet. Cytogenet.* **1:** 147–155 Available at: https://linkinghub.elsevier.com/retrieve/pii/0165460879900207

Bagg A & June CH (2016) Chimeric Antigen Receptor–Modified T Cells in Chronic Lymphoid Leukemia. *N. Engl. J. Med.* **374:** 998–998 Available at: http://www.nejm.org/doi/10.1056/NEJMx160005

Baliakas P, Hadzidimitriou A, Sutton L, Rossi D, Minga E, Villamor N, Larrayoz M, Kminkova J, Agathangelidis A, Davis Z, Tausch E, Stalika E, Kantorova B, Mansouri L, Scarfò L, Cortese D, Navrkalova V, Rose-Zerilli MJJ, Smedby KE, Juliusson G, et al (2014) Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia***:** 1–8 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24943832

Bannish G, Fuentes-Pananá EM, Cambier JC, Pear WS & Monroe JG (2001) Ligand-independent Signaling Functions for the B Lymphocyte Antigen Receptor and Their Role in Positive Selection during B Lymphopoiesis. *J. Exp. Med.* **194:** 1583–1596 Available at: http://www.jem.org/lookup/doi/10.1084/jem.194.11.1583

Barrero MJ, Boué S & Izpisúa Belmonte JC (2010) Epigenetic Mechanisms that Regulate Cell Identity. *Cell Stem Cell* **7:** 565–570

Bassing CH, Swat W & Alt FW (2002) The mechanism and regulation of chromosomal V(D)J recombination. *Cell* **109:** 45–55

Beekman R, Chapaprieta V, Russiñol N, Vilarrasa-Blasi R, Verdaguer-Dot N, Martens JHA, Duran-Ferrer M, Kulis M, Serra F, Javierre BM, Wingett SW, Clot G, Queirós AC, Castellano G, Blanc J, Gut M, Merkel A, Heath S, Vlasova A, Ullrich S, et al (2018) The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. *Nat. Med.* **24:** 868–880

Berndt SI, Camp NJ, Skibola CF, Vijai J, Wang Z, Gu J, Nieters A, Kelly RS, Smedby KE, Monnereau A, Cozen W, Cox A, Wang SS, Lan Q, Teras LR, Machado M, Yeager M, Brooks-Wilson AR, Hartge P, Purdue MP, et al (2016) Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. *Nat. Commun.* **7:** 1–9

Binet JL, Leporrier M, Dighiero G, Charron D, Vaugier G, Beral HM, Natali JC, Raphael M, Nizet B & Follezou JY (1977) A clinical staging system for chronic lymphocytic leukemia.Prognostic significance. *Cancer* **40:** 855–864 Available at: http://doi.wiley.com/10.1002/1097-0142%28197708%2940%3A2%3C855%3A%3AAID-CNCR2820400239%3E3.0.CO%3B2-1

Bologna L, Gotti E, Manganini M, Rambaldi A, Intermesoli T, Introna M & Golay J (2011) Mechanism of Action of Type II, Glycoengineered, Anti-CD20 Monoclonal Antibody GA101 in B-Chronic Lymphocytic Leukemia Whole Blood Assays in Comparison with Rituximab and Alemtuzumab. *J. Immunol.* **186:** 3762–3769

Braggio E, Kay NE, VanWier S, Tschumper RC, Smoley S, Eckel-Passow JE, Sassoon T, Barrett M, Van Dyke DL, Byrd JC, Jelinek DF, Shanafelt TD & Fonseca R (2012) Longitudinal genome-wide analysis of patients with chronic lymphocytic leukemia reveals complex evolution of clonal architecture at disease progression and at the time of relapse. *Leukemia* **26:** 1698–1701

Brander DM, Rhodes J, Pagel JM, Nabhan C, Tam CS, Jacobs R, Hill BT, Lamanna N, Lansigan F, Shadman M, Ujjani CS, Skarbnik AP, Cheson BD, Pu JJ, Sehgal AR, Furman PRR & Mato AR (2016) An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncol.* **17:** 779–790 Available at: https://linkinghub.elsevier.com/retrieve/pii/S1470204516300298

Brentjens RJ, Rivière I, Park JH, Davila ML, Wang X, Stefanski J, Taylor C, Yeh R, Bartido S, Borquez-Ojeda O, Olszewska M, Bernal Y, Pegram H, Przybylowski M, Hollyman D, Usachenko Y, Pirraglia D, Hosey J, Santos E, Halton E, et al (2011) Safety and persistence of adoptively transferred autologous CD19-targeted T cells in patients with relapsed or chemotherapy refractory B-cell leukemias. *Blood* **118:** 4817–4828

Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, Majeti R, Chang HY & Greenleaf WJ (2018) Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173:** 1535-1548.e16 Available at: https://doi.org/10.1016/j.cell.2018.03.074

Burger JA & Chiorazzi N (2013) B cell receptor signaling in chronic lymphocytic leukemia. *Trends Immunol.* **34:** 592–601 Available at: http://dx.doi.org/10.1016/j.it.2013.07.002

Burger JA, Keating MJ, Wierda WG, Hartmann E, Hoellenriegel J, Rosin NY, de Weerdt I, Jeyakumar G, Ferrajoli A, Cardenas-Turanzas M, Lerner S, Jorgensen JL, Nogueras-González GM, Zacharian G, Huang X, Kantarjian H, Garg N, Rosenwald A & O'Brien S (2014) Safety and activity of ibrutinib plus rituximab for patients with high-risk chronic lymphocytic leukaemia: A single-arm, phase 2 study. *Lancet Oncol.* **15:** 1090–1099

Burger JA, Landau DA, Taylor-Weiner A, Bozic I, Zhang H, Sarosiek K, Wang L, Stewart C, Fan J, Hoellenriegel J, Sivina M, Dubuc AM, Fraser C, Han Y, Li S, Livak KJ, Zou L, Wan Y,

Konoplev S, Sougnez C, et al (2016) Clonal evolution in patients with chronic lymphocytic leukaemia developing resistance to BTK inhibition. *Nat. Commun.* **7:** 11589 Available at: http://www.nature.com/articles/ncomms11589

Burger JA & Montserrat E (2013) Coming full circle: 70 years of chronic lymphocytic leukemia cell redistribution, from glucocorticoids to inhibitors of B-cell receptor signaling. *Blood* **121:** 1501–1509

Burger JA, Sivina M, Jain N, Kim E, Kadia T, Estrov Z, Nogueras-Gonzalez GM, Huang X, Jorgensen J, Li J, Cheng M, Clow F, Ohanian M, Andreeff M, Mathew T, Thompson P, Kantarjian H, O'Brien S, Wierda WG, Ferrajoli A, et al (2019) Randomized trial of ibrutinib vs ibrutinib plus rituximab in patients with chronic lymphocytic leukemia. *Blood* **133:** 1011–1019

Byrd JC, Lozanski G, Heerema NA, Flinn IW, Smith L, Harbison J, Webb J, Moran M, Lucas M, Lin T, Hackbarth ML, Proffitt JH, Lucas D & Grever MR (2004) Alemtuzumab is an effective therapy for chronic lymphocytic leukemia with p53 mutations and deletions. *Blood* **103:** 3278–3281

Campo E, Cymbalista F, Ghia P, Jäger U, Pospisilova S, Rosenquist R, Schuh A & Stilgenbauer S (2018) TP53 aberrations in chronic lymphocytic leukemia: an overview of the clinical implications of improved diagnostics. *Haematologica* **103:** 1956–1968 Available at: http://www.haematologica.org/lookup/doi/10.3324/haematol.2018.187583

Campo E, Swerdlow SH, Harris NL, Pileri S, Stein H & Jaffe ES (2011) The 2008 WHO classification of lymphoid neoplasms and beyond: Evolving concepts and practical applications. *Blood* **117:** 5019–5032

Chantepie SP, Vaur D, Grunau C, Salaün V, Briand M, Parienti JJ, Heutte N, Cheze S, Roussel M, Gauduchon P, Leporrier M & Krieger S (2010) ZAP-70 intron1 DNA methylation status: Determination by pyrosequencing in B chronic lymphocytic leukemia. *Leuk. Res.* **34:** 800–808

Chigrinova E, Rinaldi A, Kwee I, Rossi D, Rancoita PMV, Strefford JC, Oscier D, Stamatopoulos K, Papadaki T, Berger F, Young KH, Murray F, Rosenquist R, Greiner TC, Chan WC, Orlandi EM, Lucioni M, Marasca R, Inghirami G, Ladetto M, et al (2013) Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome. *Blood* **122:** 2673–2682

Cimmino A, Calin GA, Fabbri M, Iorio M V, Ferracin M, Shimizu M, Wojcik SE, Aqeilan RI, Zupo S, Dono M, Rassenti L, Alder H, Volinia S, Liu C -g., Kipps TJ, Negrini M & Croce CM (2005) miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc. Natl. Acad. Sci.* **102:** 13944–13949 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0506654102

Claus R, Lucas DM, Ruppert AS, Williams KE, Weng D, Patterson K, Zucknick M, Oakes CC, Rassenti LZ, Greaves AW, Geyer S, Wierda WG, Brown JR, Gribben JG, Barrientos JC, Rai KR, Kay NE, Kipps TJ, Shields P, Zhao W, et al (2014) Validation of ZAP-70 methylation and its relative significance in predicting outcome in chronic lymphocytic leukemia. *Blood* **124:** 42–48

Coiffier B, Lepretre S, Pedersen LM, Gadeberg O, Fredriksen H, van Oers MHJ, Wooldridge J, Kloczko J, Holowiecki J, Hellmann A, Walewski J, Flensburg M, Petersen J & Robak T (2007) Safety and efficacy of ofatumumab, a fully human monoclonal anti-CD20 antibody, in patients with relapsed or refractory B-cell chronic lymphocytic leukemia: a phase 1-2 study. *Blood* **111:** 1094–1100 Available at: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2007-09-111781

Compagno M, Wang Q, Pighi C, Cheong TC, Meng FL, Poggio T, Yeap LS, Karaca E, Blasco RB, Langellotto F, Ambrogio C, Voena C, Wiestner A, Kasar SN, Brown JR, Sun J, Wu CJ, Gostissa M, Alt FW & Chiarle R (2017) Phosphatidylinositol 3-kinase δ blockade increases genomic instability in B cells. *Nature* **542:** 483–493 Available at: http://dx.doi.org/10.1038/nature21406

Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ, Majeti R & Chang HY (2016) Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48:** 1193–1203 Available at: http://www.nature.com/articles/ng.3646

Corcoran M, Parker A, Orchard J, Davis Z, Wirtz M, Schmitz OJ & Oscier D (2005) ZAP-70 methylation status is associated with ZAP-70 expression status in chronic lymphocytic leukemia. *Haematologica* **90:** 1078–88

Criel A, Verhoef G, Vlietinck R, Mecucci C, Billiet J, Michaux L, Meeus P, Louwagie A, Van Orshoven A, Van Hoof A, Boogaerts M, Van den Berghe H & De Wolf-Peeters C (1997) Further characterization of morphologically defined typical and atypical CLL: a clinical, immunophenotypic, cytogenetic and prognostic study on 390 cases. *Br. J. Haematol.* **97:** 383–391 Available at: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2141.1997.402686.x

Dancey JE & Chen HX (2006) Strategies for optimizing combinations of molecularly targeted anticancer agents. *Nat. Rev. Drug Discov.* **5:** 649–659

Darvin P, Toor SM, Sasidharan Nair V & Elkord E (2018) Immune checkpoint inhibitors: recent progress and potential biomarkers. *Exp. Mol. Med.* **50:** 1–11 Available at: http://dx.doi.org/10.1038/s12276-018-0191-1

Depoil D, Fleire S, Treanor BL, Weber M, Harwood NE, Marchbank KL, Tybulewicz VLJ & Batista FD (2008) CD19 is essential for B cell activation by promoting B cell receptor-antigen microcluster formation in response to membrane-bound ligand. *Nat. Immunol.* **9:** 63–72

Dietrich S, Oleś M, Lu J, Sellner L, Anders S, Velten B, Wu B, Hüllein J, Liberio M da S, Walther T, Wagner L, Rabe S, Ghidelli-Disse S, Bantscheff M, Oleś AK, Słabicki M, Mock A, Oakes CC, Wang S, Oppermann S, et al (2018) Drug-perturbation-based stratification of blood cancer. *J. Clin. Invest.* **128:** 427–445

Druker BJ & Lydon NB (2000) Lessons learned from the development of an Abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J. Clin. Invest.* **105:** 3–7 Available at: http://www.jci.org/articles/view/9083

Eichhorst B & Hallek M (2016) Prognostication of chronic lymphocytic Leukemia in the era of new agents. *Hematology* **2016:** 149–155

Eshhar Z, Waks T, Gross G & Schindler DG (1993) Specific activation and targeting of cytotoxic lymphocytes through chimeric single chains consisting of antibody-binding domains and the gamma or zeta subunits of the immunoglobulin and T-cell receptors. *Proc. Natl. Acad. Sci.* **90:** 720–724 Available at: https://www.pnas.org/content/90/2/720 [Accessed August 3, 2019]

Fan C-M & Maniatis T (1991) Generation of p50 subunit of NF-kB by processing of p105 through an ATP-dependent pathway. *Nature* **354:** 395–398 Available at: http://www.nature.com/articles/354395a0

Farlik M, Halbritter F, Müller F, Choudry FA, Ebert P, Klughammer J, Farrow S, Santoro A, Ciaurro V, Mathur A, Uppal R, Stunnenberg HG, Ouwehand WH, Laurenti E, Lengauer T, Frontini M & Bock C (2016) DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell* **19:** 808–822

Farooqui MZH, Valdez J, Martyr S, Aue G, Saba N, Niemann CU, Herman SEM, Tian X, Marti G, Soto S, Hughes TE, Jones J, Lipsky A, Pittaluga S, Stetler-Stevenson M, Yuan C, Lee YS, Pedersen LB, Geisler CH, Calvo KR, et al (2015) Ibrutinib for previously untreated and relapsed or refractory chronic lymphocytic leukaemia with TP53 aberrations: A phase 2, single-arm trial. *Lancet Oncol.* **16:** 169–176 Available at: http://dx.doi.org/10.1016/S1470-2045(14)71182-9

Fearon DT, Carroll MC & Carroll MC (2000) Regulation of B Lymphocyte Responses to Foreign and Self-Antigens by the CD19/CD21 Complex. *Annu. Rev. Immunol.* **18:** 393–422 Available at: http://www.annualreviews.org/doi/10.1146/annurev.immunol.18.1.393

Fenn JE & Udelsman R (2011) First use of intravenous chemotherapy cancer treatment: Rectifying the record. *J. Am. Coll. Surg.* **212:** 413–417 Available at: http://dx.doi.org/10.1016/j.jamcollsurg.2010.10.018

Ferreira PG, Jares P, Rico D, Gomez-Lopez G, Martinez-Trillos A, Villamor N, Ecker S, Gonzalez-Perez A, Knowles DG, Monlong J, Johnson R, Quesada V, Djebali S, Papasaikas P, Lopez-Guerra M, Colomer D, Royo C, Cazorla M, Pinyol M, Clot G, et al (2014) Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Res.* **24:** 212–226 Available at: http://genome.cshlp.org/cgi/doi/10.1101/gr.152132.112

Flinn IW, Hillmen P, Montillo M, Nagy Z, Illés Á, Etienne G, Delgado J, Kuss BJ, Tam CS, Gasztonyi Z, Offner F, Lunin S, Bosch F, Davids MS, Lamanna N, Jaeger U, Ghia P, Cymbalista F, Portell CA, Skarbnik AP, et al (2018) The phase 3 DUO trial: Duvelisib vs ofatumumab in relapsed and refractory CLL/SLL. *Blood* **132:** 2446–2455

Forconi F, Potter KN, Wheatley I, Darzentas N, Sozzi E, Stamatopoulos K, Mockridge CI, Packham G & Stevenson FK (2010) The normal IGHV1-69-derived B-cell repertoire contains stereotypic patterns characteristic of unmutated CLL. *Blood* **115:** 71–77 Available at: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2009-06-225813

Fraietta JA, Lacey SF, Orlando EJ, Pruteanu-Malinici I, Gohil M, Lundh S, Boesteanu AC, Wang Y, O'connor RS, Hwang WT, Pequignot E, Ambrose DE, Zhang C, Wilcox N, Bedoya F, Dorfmeier C, Chen F, Tian L, Parakandi H, Gupta M, et al (2018) Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia. *Nat. Med.* **24:** 563–571

Fu SM, Winchester RJ & Kunkel HG (1974) Occurrence of surface IgM, IgD, and free light chains on human lymphocytes. *J. Exp. Med.* **139:** 451–456 Available at: http://www.jem.org/cgi/doi/10.1084/jem.139.2.451

Furman RR, Sharman JP, Coutre SE, Cheson BD, Pagel JM, Hillmen P, Barrientos JC, Zelenetz AD, Kipps TJ, Flinn I, Ghia P, Eradat H, Ervin T, Lamanna N, Coiffier B, Pettitt AR, Ma S, Stilgenbauer S, Cramer P, Aiello M, et al (2014) Idelalisib and Rituximab in Relapsed Chronic Lymphocytic Leukemia. *N. Engl. J. Med.* **370:** 997–1007

Gahrton G, Juliusson G, Robèrt K-H & Friberg K (1987) Role of chromosomal abnormalities in chronic lymphocytic leukemia. *Blood Rev.* **1:** 183–192 Available at: https://linkinghub.elsevier.com/retrieve/pii/0268960X87900348

Gaiti F, Chaligne R, Gu H, Brand RM, Kothen-Hill S, Schulman RC, Grigorev K, Risso D, Kim K, Pastore A, Huang KY, Alonso A, Sheridan C, Omans ND, Biederstedt E, Clement K, Wang L, Felsenfeld JA, Bhavsar EB, Aryee MJ, et al (2019) Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature* **569:** 576–580 Available at: http://dx.doi.org/10.1038/s41586-019-1198-z

Gibson G (2018) Population genetics and GWAS: A primer. *PLoS Biol.* **16:** 1–6

Gruber M, Bozic I, Leshchiner I, Livitz D, Stevenson K, Rassenti L, Rosebrock D, Taylor-Weiner A, Olive O, Goyetche R, Fernandes SM, Sun J, Stewart C, Wong A, Cibulskis C, Zhang W, Reiter JG, Gerold JM, Gribben JG, Rai KR, et al (2019) Growth dynamics in naturally progressing chronic lymphocytic leukaemia. *Nature* **570:** 474–479 Available at: http://dx.doi.org/10.1038/s41586-019-1252-x

Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, Hillmen P, Keating M, Montserrat E, Chiorazzi N, Stilgenbauer S, Rai KR, Byrd JC, Eichhorst B, O'Brien S, Robak T, Seymour JF & Kipps TJ (2018) iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood* **131:** 2745–2760 Available at: http://www.ncbi.nlm.nih.gov/pubmed/29540348

Hao T, Li-Talley M, Buck A & Chen W (2019) An emerging trend of rapid increase of leukemia but not all cancers in the aging population in the United States. *Sci. Rep.* **9:** 12070 Available at: http://www.nature.com/articles/s41598-019-48445-1

Herishanu Y, Perez-Galan P, Liu D, Biancotto A, Pittaluga S, Vire B, Gibellini F, Njuguna N, Lee E, Stennett L, Raghavachari N, Liu P, McCoy JP, Raffeld M, Stetler-Stevenson M, Yuan C, Sherry R, Arthur DC, Maric I, White T, et al (2011) The lymph node microenvironment promotes B-cell receptor signaling, NF- B activation, and tumor proliferation in chronic lymphocytic leukemia. *Blood* **117:** 563–574 Available at: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2010-05-284984

Herman SEM, Mustafa RZ, Gyamfi J a, Pittaluga S, Chang S, Chang B, Farooqui M & Wiestner A (2014) Ibrutinib inhibits B-cell receptor and NF-κB signaling and reduces tumor proliferation in tissue-resident cells of patients with chronic lymphocytic leukemia. *Blood* **123:** 3286–3295 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24659631

Jaglowski SM, Alinari L, Lapalombella R, Muthusamy N & Byrd JC (2010) The clinical application of monoclonal antibodies in chronic lymphocytic leukemia. *Blood* **116:** 3705–3714 Available at: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2010-04-001230

Jain N & O'Brien S (2015) Initial treatment of CLL: Integrating biology and functional status. *Blood* **126:** 463–470

Jondal M (1974) Surface Markers on Human B and T Lymphocytes. *Scand. J. Immunol.* **3:** 269–276 Available at: http://doi.wiley.com/10.1111/j.1365-3083.1974.tb01257.x

Juliusson G, Oscier DG, Fitchett M, Ross FM, Stockdill G, Mackie MJ, Parker AC, Castoldi GL, Cuneo A, Knuutila S, Elonen E & Gahrton G (1990) Prognostic Subgroups in B-Cell Chronic Lymphocytic Leukemia Defined by Specific Chromosomal Abnormalities. *N. Engl. J. Med.* **323:** 720–724 Available at: http://www.nejm.org/doi/abs/10.1056/NEJM199009133231105

Kaisho T, Takeda K, Tsujimura T, Kawai T, Nomura F, Terada N & Akira S (2001) IκB kinase α is essential for mature B cell development and function. *J. Exp. Med.* **193:** 417–426

Kambayashi T & Laufer TM (2014) Atypical MHC class II-expressing antigen-presenting cells: can anything replace a dendritic cell? *Nat. Rev. Immunol.* **14:** 719–730 Available at: http://www.nature.com/articles/nri3754

Kipps TJ, Stevenson FK, Wu CJ, Croce CM, Packham G, Wierda WG, O'Brien S, Gribben J & Rai K (2017) Chronic lymphocytic leukaemia. *Nat. Rev. Dis. Prim.* **3:**

Klein U, Tu Y, Stolovitzky G a, Mattioli M, Cattoretti G, Husson H, Freedman a, Inghirami G, Cro L, Baldini L, Neri a, Califano a & Dalla-Favera R (2001) Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J. Exp. Med.* **194:** 1625–1638

Koyasu S (2003) The role of PI3K in immune cells. *Nat. Immunol.* **4:** 313–319 Available at: http://www.nature.com/articles/ni0403-313

Kulis M, Heath S, Bibikova M, Queirós AC, Navarro A, Clot G, Martínez-Trillos A, Castellano G, Brun-Heath I, Pinyol M, Barberán-Soler S, Papasaikas P, Jares P, Beà S, Rico D, Ecker S, Rubio M, Royo R, Ho V, Klotzle B, et al (2012) Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.* **44:** 1236–1242 Available at: http://dx.doi.org/10.1038/ng.2443

Kulis M, Merkel A, Heath S, Queirós AC, Schuyler RP, Castellano G, Beekman R, Raineri E, Esteve A, Clot G, Verdaguer-Dot N, Duran-Ferrer M, Russiñol N, Vilarrasa-Blasi R, Ecker S, Pancaldi V, Rico D, Agueda L, Blanc J, Richardson D, et al (2015) Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat. Genet.* **47:** 746–756 Available at: http://www.nature.com/doifinder/10.1038/ng.3291

Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, Sougnez C, Stewart C, Sivachenko A, Wang L, Wan Y, Zhang W, Shukla SA, Vartanov A, Fernandes SM, Saksena G, Cibulskis K, Tesar B, Gabriel S, Hacohen N, et al (2013) Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia. *Cell* **152:** 714–726 Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867413000718

Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, Stevenson K, Sougnez C, Wang L, Li S, Kotliar D, Zhang W, Ghandi M, Garraway L, Fernandes SM, Livak KJ, Gabriel S, Gnirke A, Lander ES, Brown JR, et al (2014) Locally Disordered Methylation Forms the Basis of Intratumor Methylome Variation in Chronic Lymphocytic Leukemia. *Cancer Cell* **26:** 813–825 Available at: http://linkinghub.elsevier.com/retrieve/pii/S1535610814004164

Landau DA, Sun C, Rosebrock D, Herman SEM, Fein J, Sivina M, Underbayev C, Liu D, Hoellenriegel J, Ravichandran S, Farooqui MZH, Zhang W, Cibulskis C, Zviran A, Neuberg DS, Livitz D, Bozic I, Leshchiner I, Getz G, Burger JA, et al (2017) The evolutionary landscape of chronic lymphocytic leukemia treated with ibrutinib targeted therapy. *Nat. Commun.* **8:** 2185 Available at: http://dx.doi.org/10.1038/s41467-017-02329-y

Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, Kluth S, Bozic I, Lawrence M, Böttcher S, Carter SL, Cibulskis K, Mertens D, Sougnez CL, Rosenberg M, Hess JM, Edelmann J, Kless S, Kneba M, Ritgen M, et al (2015) Mutations driving CLL and their evolution in progression and relapse. *Nature* **526:** 525–530 Available at: http://www.nature.com/doifinder/10.1038/nature15395

Lenartova A, Johannesen TB & Tjønnfjord GE (2016) National trends in incidence and survival of chronic lymphocytic leukemia in Norway for 1953-2012: a systematic analysis of population-based data. *Cancer Med.* **5:** 3588–3595 Available at: http://doi.wiley.com/10.1002/cam4.849

Li Z (2004) The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev.* **18:** 1–11 Available at: http://www.genesdev.org/cgi/doi/10.1101/gad.1161904

Linet MS, Schubauer-Berigan MK, Weisenburger DD, Richardson DB, Landgren O, Blair A, Silver S, Field RW, Caldwell G, Hatch M & Dores GM (2007) Chronic lymphocytic leukaemia: An overview of aetiology in light of recent developments in classification and pathogenesis. *Br. J. Haematol.* **139:** 672–686

Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee A V., Omberg L, Wolf DM, Shriver CD, Thorsson V, Hu H, Caesar-Johnson SJ, Demchok JA, Felau I, Kasapi M, Ferguson ML, et al (2018) An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173:** 400-416.e11 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0092867418302290

Mavromatis B & Cheson BD (2003) Monoclonal Antibody Therapy of Chronic Lymphocytic Leukemia. *J. Clin. Oncol.* **21:** 1874–1881 Available at: http://ascopubs.org/doi/10.1200/JCO.2003.09.113

McGranahan N & Swanton C (2017) Cancer Evolution Constrained by the Immune Microenvironment. *Cell* **170:** 825–827 Available at: http://dx.doi.org/10.1016/j.cell.2017.08.012

Montserrat E, Gomis F, Vallespi T, Rios A, Romero A, Soler J, Alcala A, Morey M, Ferran C, Diaz-Mediavilla J, Flores A, Woessner S, Batlle J, Gonzalez-Aza C, Rovira M, Reverter JC & Rozman C (1991) Presenting features and prognosis of chronic lymphocytic leukemia in younger adults. *Blood* **78:** 1545

Moreno C, Greil R, Demirkan F, Tedeschi A, Anz B, Larratt L, Simkovic M, Samoilova O, Novak J, Ben-Yehuda D, Strugov V, Gill D, Gribben JG, Hsu E, Lih CJ, Zhou C, Clow F, James DF, Styles L & Flinn IW (2019) Ibrutinib plus obinutuzumab versus chlorambucil plus obinutuzumab in first-line treatment of chronic lymphocytic leukaemia (iLLUMINATE): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol.* **20:** 43–56 Available at: http://dx.doi.org/10.1016/S1470-2045(18)30788-5

Mukhtar E, Adhami VM & Mukhtar H (2014) Targeting microtubules by natural agents for cancer therapy. *Mol. Cancer Ther.* **13:** 275–284

Nel AE, Landreth GE, Goldschmidt-Clermont PJ, Tung HE & Galbraith RM (1984) Enhanced tyrosine phosphorylation in B lymphocytes upon complexing of membrane immunoglobulin. *Biochem. Biophys. Res. Commun.* **125:** 859–866 Available at: https://linkinghub.elsevier.com/retrieve/pii/0006291X84913627

Nemazee D (2017) Mechanisms of central tolerance for B cells. *Nat. Rev. Immunol.* **17:** 281–294 Available at: http://dx.doi.org/10.1038/nri.2017.19

Ninomiya S, Narala N, Huye L, Yagyu S, Savoldo B, Dotti G, Heslop HE, Brenner MK, Rooney CM & Ramos CA (2015) Tumor indoleamine 2,3-dioxygenase (IDO) inhibits CD19-CAR T cells and is downregulated by lymphodepleting drugs. *Blood* **125:** 3905–3916 Available at: http://www.bloodjournal.org/lookup/doi/10.1182/blood-2015-01-621474

O'Brien S, Furman RR, Coutre S, Flinn IW, Burger JA, Blum K, Sharman J, Wierda W, Jones J, Zhao W, Heerema NA, Johnson AJ, Luan Y, James DF, Chu AD & Byrd JC (2018) Single-agent ibrutinib in treatment-naïve and relapsed/refractory chronic lymphocytic leukemia: a 5-year experience. *Blood* **131:** 1910–1919 Available at: http://www.bloodjournal.org/content/bloodjournal/early/2018/02/01/blood-2017-10-810044.full.pdf?sso-checked=true

O'Brien SM, Lamanna N, Kipps TJ, Flinn I, Zelenetz AD, Burger JA, Keating M, Mitra S, Holes L, Yu AS, Johnson DM, Miller LL, Kim Y, Dansey RD, Dubowy RL & Coutre SE (2015) A phase 2 study of idelalisib plus rituximab in treatment-naïve older patients with chronic lymphocytic leukemia. *Blood* **126:** 2686–2694

Oakes CC, Claus R, Gu L, Assenov Y, Hullein J, Zucknick M, Bieg M, Brocks D, Bogatyrova O, Schmidt CR, Rassenti L, Kipps TJ, Mertens D, Lichter P, Dohner H, Stilgenbauer S, Byrd JC, Zenz T & Plass C (2014) Evolution of DNA Methylation Is Linked to Genetic Aberrations in Chronic Lymphocytic Leukemia. *Cancer Discov.* **4:** 348–361 Available at: http://cancerdiscovery.aacrjournals.org/cgi/doi/10.1158/2159-8290.CD-13-0349

Oakes CC, Seifert M, Assenov Y, Gu L, Przekopowitz M, Ruppert AS, Wang Q, Imbusch CD, Serva A, Brocks D, Koser SD, Lipka DB, Bogatyrova O, Weichenhan D, Brors B, Rassenti L, Kipps TJ, Mertens D, Zapatka M, Lichter P, et al (2016) DNA methylation dynamics during B cell

maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat. Genet.*

Ouillette P, Saiya-Cork K, Seymour E, Li C, Shedden K & Malek SN (2013) Clonal Evolution, Genomic Drivers, and Effects of Therapy in Chronic Lymphocytic Leukemia. *Clin. Cancer Res.* **19:** 2893–2904 Available at: http://clincancerres.aacrjournals.org/cgi/doi/10.1158/1078-0432.CCR-13-0138

Packard TA & Cambier JC (2013) B lymphocyte antigen receptor signaling: Initiation, amplification, and regulation. *F1000Prime Rep.* **5:** 5–7

Pardoll DM (2012) The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* **12:** 252–264 Available at: http://dx.doi.org/10.1038/nrc3239

Parikh SA (2018) Chronic lymphocytic leukemia treatment algorithm 2018. *Blood Cancer J.* **8:** 93 Available at: http://dx.doi.org/10.1038/s41408-018-0131-2

Park JH, Geyer MB & Brentjens RJ (2016) CD19-targeted CAR T-cell therapeutics for hematologic malignancies: Interpreting clinical outcomes to date. *Blood* **127:** 3312–3320

Park YJ, Kuen DS & Chung Y (2018) Future prospects of immune checkpoint blockade in cancer: from response prediction to overcoming resistance. *Exp. Mol. Med.* **50:** 1–13 Available at: http://dx.doi.org/10.1038/s12276-018-0130-1

Parker DC (1993) T Cell-Dependent B Cell Activation. *Annu. Rev. Immunol.* **11:** 331–360 Available at: http://www.annualreviews.org/doi/10.1146/annurev.iy.11.040193.001555

Parker WB (2009) Enzymology of purine and pyrimidine antimetabolites used in the treatment of cancer. *Chem. Rev.* **109:** 2880–2893

Pastore A, Gaiti F, Lu SX, Brand RM, Kulm S, Chaligne R, Gu H, Huang KY, Stamenova EK, Béguelin W, Jiang Y, Schulman RC, Kim K, Alonso A, Allan JN, Furman RR, Gnirke A, Wu CJ, Melnick AM, Meissner A, et al (2019) Corrupted coordination of epigenetic modifications leads to diverging chromatin states and transcriptional heterogeneity in CLL. *Nat. Commun.* **10:** 1874 Available at: http://dx.doi.org/10.1038/s41467-019-09645-5

Pogue SL, Kurosaki T, Bolen J & Herbst R (2000) B Cell Antigen Receptor-Induced Activation of Akt Promotes B Cell Survival and Is Dependent on Syk Kinase. *J. Immunol.* **165:** 1300–1306

Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI, Munar M, Rubio-Pérez C, Jares P, Aymerich M, Baumann T, Beekman R, Belver L, Carrio A, Castellano G, Clot G, Colado E, Colomer D, Costa D, Delgado J, et al (2015) Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* Available at: http://www.nature.com/doifinder/10.1038/nature14666

Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, Escaramis G, Jares P, Beà S, González-Díaz M, Bassaganyas L, Baumann T, Juan M, López-Guerra M, Colomer D, Tubío JMC, López C, Navarro A, Tornador C, Aymerich M, et al (2011) Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475:** 101–105

Pufall MA (2015) Glucocorticoids and Cancer. **872:** 315–333 Available at: http://link.springer.com/10.1007/978-1-4939-2895-8

Quail DF & Joyce JA (2013) Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* **19:** 1423–1437

Quesada V, Conde L, Villamor N, Ordóñez GR, Jares P, Bassaganyas L, Ramsay AJ, Beà S, Pinyol M, Martínez-Trillos A, López-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Giné E, Hernández JM, González-Díaz M, et al (2012) Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* **44:** 47–52 Available at: http://dx.doi.org/10.1038/ng.1032

Radaev S, Zou Z, Tolar P, Nguyen K, Nguyen A, Krueger PD, Stutzman N, Pierce S & Sun PD (2010) Structural and Functional Studies of Igαβ and Its Assembly with the B Cell Antigen Receptor. *Structure* **18:** 934–943 Available at: https://linkinghub.elsevier.com/retrieve/pii/S0969212610002315

Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN & Pasternack BS (1975) Clinical staging of chronic lymphocytic leukemia. *Blood* **46:** 219–34 Available at: http://www.ncbi.nlm.nih.gov/pubmed/1139039

Ramsay AJ, Quesada V, Foronda M, Conde L, Martínez-Trillos A, Villamor N, Rodríguez D, Kwarciak A, Garabaya C, Gallardo M, López-Guerra M, López-Guillermo A, Puente XS, Blasco MA, Campo E & López-Otín C (2013) POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat. Genet.* **45:** 526–530

Rawstron AC, Kreuzer KA, Soosapilla A, Spacek M, Stehlikova O, Gambell P, McIver-Brown N, Villamor N, Psarra K, Arroz M, Milani R, de la Serna J, Cedena MT, Jaksic O, Nomdedeu J, Moreno C, Rigolin GM, Cuneo A, Johansen P, Johnsen HE, et al (2018) Reproducible diagnosis of chronic lymphocytic leukemia by flow cytometry: An European Research Initiative on CLL (ERIC) & European Society for Clinical Cell Analysis (ESCCA) Harmonisation project. *Cytom. Part B - Clin. Cytom.* **94:** 121–128

Reff ME, Carner K, Chambers KS, Chinn PC, Leonard JE, Raab R, Newman RA, Hanna N & Anderson DR (1994) Depletion of B cells in vivo by a chimeric mouse human monoclonal antibody to CD20. *Blood* **83:** 435–45 Available at: http://www.ncbi.nlm.nih.gov/pubmed/7506951

Roberts AW, Davids MS, Pagel JM, Kahl BS, Puvvada SD, Gerecitano JF, Kipps TJ, Anderson MA, Brown JR, Gressick L, Wong S, Dunbar M, Zhu M, Desai MB, Cerri E, Heitner Enschede S, Humerickhouse RA, Wierda WG & Seymour JF (2016) Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia. *N. Engl. J. Med.* **374:** 311–322 Available at: http://www.nejm.org/doi/10.1056/NEJMoa1513257

Rosenthal A (2017) Small Molecule Inhibitors in Chronic Lymphocytic Lymphoma and B Cell Non-Hodgkin Lymphoma. *Curr. Hematol. Malig. Rep.* **12:** 207–216

Rosenwald A, Alizadeh AA, Widhopf G, Simon R, Davis RE, Yu X, Yang L, Pickeral OK, Rassenti LZ, Powell J, Botstein D, Byrd JC, Grever MR, Cheson BD, Chiorazzi N, Wilson WH, Kipps

TJ, Brown PO & Staudt LM (2001) Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J. Exp. Med.* **194:** 1639–47

Ross FM & Stockdill G (1987) Clonal chromosome abnormalities in chronic lymphocytic leukemia patients revealed by TPA stimulation of whole blood cultures. *Cancer Genet. Cytogenet.* **25:** 109–121 Available at: https://linkinghub.elsevier.com/retrieve/pii/016546088790166X

Rozman C & Montserrat E (1995) Current concepts: Chronic Lymphocytic Leukemia. *N. Engl. J. Med.* **333:** 1052–1057 Available at: http://www.nejm.org/doi/abs/10.1056/NEJM199510193331606

Schmidl C, Vladimer GI, Rendeiro AF, Schnabl S, Krausgruber T, Taubert C, Krall N, Pemovska T, Araghi M, Snijder B, Hubmann R, Ringler A, Runggatscher K, Demirtas D, de la Fuente OL, Hilgarth M, Skrabs C, Porpaczy E, Gruber M, Hoermann G, et al (2019) Combined chemosensitivity and chromatin profiling prioritizes drug combinations in CLL. *Nat. Chem. Biol.* **15:** 232–240 Available at: http://www.nature.com/articles/s41589-018-0205-2

Schroeder HW & Cavacini L (2010) Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **125:** S41–S52 Available at: http://dx.doi.org/10.1016/j.jaci.2009.09.046

Schüssler-Fiorenza Rose SM, Contrepois K, Moneghetti KJ, Zhou W, Mishra T, Mataraso S, Dagan-Rosenfeld O, Ganz AB, Dunn J, Hornburg D, Rego S, Perelman D, Ahadi S, Sailani MR, Zhou Y, Leopold SR, Chen J, Ashland M, Christle JW, Avina M, et al (2019) A longitudinal big data approach for precision health. *Nat. Med.* **25:** 792–804 Available at: http://dx.doi.org/10.1038/s41591-019-0414-6

Siegel RL, Miller KD & Jemal A (2018) Cancer statistics, 2018. *CA. Cancer J. Clin.* **68:** 7–30 Available at: http://doi.wiley.com/10.3322/caac.21442

Spina V & Rossi D (2019) Overview of non-coding mutations in chronic lymphocytic leukemia. *Mol. Oncol.* **13:** 99–106 Available at: http://doi.wiley.com/10.1002/1878-0261.12416

Swift LH & Golsteyn RM (2014) Genotoxic anti-cancer agents and their relationship to DNA damage, mitosis, and checkpoint adaptation in proliferating cancer cells. *Int. J. Mol. Sci.* **15:** 3403–3431

Tam CS, O'Brien S, Wierda W, Kantarjian H, Wen S, Do K-A, Thomas DA, Cortes J, Lerner S & Keating MJ (2008) Long-term results of the fludarabine, cyclophosphamide, and rituximab regimen as initial therapy of chronic lymphocytic leukemia. *Blood* **112:** 975–980 Available at: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2008-02-140582

The International CLL-IPI working Group (2016) An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncol.* **17:** 779–790 Available at: http://dx.doi.org/10.1016/S1470-2045(16)30029-8

Turtle CJ, Riddell SR & Maloney DG (2016) CD19-Targeted chimeric antigen receptor-modified T-cell immunotherapy for B-cell malignancies. *Clin. Pharmacol. Ther.* **100:** 252–258

Tzifi F, Economopoulou C, Gourgiotis D, Ardavanis A, Papageorgiou S & Scorilas A (2012) The role of BCL2 family of apoptosis regulator proteins in acute and chronic leukemias. *Adv. Hematol.* **2012:**

Wang J, Lunyak V V. & Jordan IK (2011) Genome-wide prediction and analysis of human chromatin boundary elements. *Nucleic Acids Res.* **40:** 511–529 Available at: http://nar.oxfordjournals.org/cgi/content/abstract/40/2/511 [Accessed September 20, 2011]

Wang L, Shalek AK, Lawrence M, Ding R, Gaublomme JT, Pochet N, Stojanov P, Sougnez C, Shukla SA, Stevenson KE, Zhang W, Wong J, Sievers QL, MacDonald BT, Vartanov AR, Goldstein NR, Neuberg D, He X, Lander E, Hacohen N, et al (2014) Somatic mutation as a mechanism of Wnt/ -catenin pathway activation in CLL. *Blood* **124:** 1089–1098 Available at: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2014-01-552067

Wang LD & Clark MR (2003) B-cell antigen-receptor signalling in lymphocyte development. *Immunology* **110:** 411–420

Woyach JA & Johnson AJ (2015) Targeted therapies in CLL: mechanisms of resistance and strategies for management. *Blood* **126:** 471–477 Available at: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2015-03-585075

Yam-Puc JC, Zhang L, Zhang Y & Toellner K-M (2018) Role of B-cell receptors for B-cell development and antigen-induced differentiation. *F1000Research* **7:** 429 Available at: https://f1000research.com/articles/7-429/v1

Zhang J, Yang PL & Gray NS (2009) Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9:** 28–39

Zhou W, Sailani MR, Contrepois K, Zhou Y, Ahadi S, Leopold SR, Zhang MJ, Rao V, Avina M, Mishra T, Johnson J, Lee-McMullen B, Chen S, Metwally AA, Tran TDB, Nguyen H, Zhou X, Albright B, Hong B, Petersen L, et al (2019) Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature* **569:** 663–671 Available at: http://www.nature.com/articles/s41586-019-1236-x

Curriculum Vitae

André Figueiredo Rendeiro
Born 21st January 1990, Murtosa, Portugal
andre.rendeiro@pm.me
andre-rendeiro.com
ORCID: 0000-0001-9362-5373

Current position

9/2014-1/2020 PhD student
CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences
Laboratory of Christoph Bock

Education

9/2012-6/2014 Masters in Molecular and Cell Biology, University of Aveiro, Portugal
9/2018-7/2012 Bachelor in Biology, University of Aveiro, Portugal

Publications

Pre-peer review publications are marked with (⚑); equal contributions are marked with (*).

First-author publications

1. ⚑ Paul Datlinger*, André F. Rendeiro*, Thorina Boenke, Thomas Krausgruber, Daniele Barreca, Christoph Bock. Ultra-high throughput single-cell RNA sequencing by combinatorial fluidic indexing. bioRxiv (2019). doi:10.1101/2019.12.17.879304
2. André F. Rendeiro*, Thomas Krausgruber*, Nikolaus Fortelny, Fangwen Zhao, Thomas Penz, Matthias Farlik, Linda C. Schuster, Amelie Nemc, Szabolcs Tasnády, Marienn Réti, Zoltán Mátrai, Donat Alpar, Csaba Bödör, Christian Schmidl, Christoph Bock. Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib drug response in chronic lymphocytic leukemia. Nature Communications (2020). doi:10.1038/s41467-019-14081-6
3. Christian Schmidl*, Gregory I Vladimer*, André F. Rendeiro*, Susanne Schnabl*, Thomas Krausgruber, Christina Taubert, Nikolaus Krall, Tea Pemovska, Mohammad Araghi, Berend Snijder, Rainer Hubmann, Anna Ringler, Kathrin Runggatscher, Dita Demirtas, Oscar Lopez de la Fuente, Martin Hilgarth, Cathrin Skrabs, Edit Porpaczy, Michaela Gruber, Gregor Hoermann, Stefan Kubicek, Philipp B Staber, Medhat Shehata, Giulio Superti-Furga, Ulrich Jäger, Christoph Bock. Combined chemosensitivity and chromatin profiling prioritizes drug combinations in CLL. Nature Chemical Biology (2019). doi:10.1038/s41589-018-0205-2
4. André F. Rendeiro*, Christian Schmidl*, Jonathan C. Strefford*, Renata Walewska, Zadie Davis, Matthias Farlik, David Oscier, Christoph Bock. Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. Nature Communications (2016). doi:10.1038/ncomms11938
5. Christian Schmidl*, André F. Rendeiro*, Nathan C Sheffield, Christoph Bock. 2015. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nature Methods (2015). doi:10.1038/nmeth.3542

Additional publications

1. Michael Delacher, Charles D Imbusch, Agnes Hotz-Wagenblatt, Jan-Philipp Mallm, Katharina Bauer, Malte Simon, Dania Riegel, André F Rendeiro, Sebastian Bittner, Lieke Sanderink, Asmita Pant, Lisa Schmidleithner, Kathrin L Braband, Bernd Echtenachter, Alexander Fischer, Valentina Giunchiglia, Petra Hoffmann, Matthias Edinger, Christoph

Bock, Michael Rehli, Benedikt Brors, Christian Schmidl, Markus Feuerer. Precursors for Nonlymphoid-Tissue Treg Cells Reside in Secondary Lymphoid Organs and Are Programmed by the Transcription Factor BATF. Immunity. (2020) doi:10.1016/j.immuni.2019.12.002

2. Christopher JM Piper, Elizabeth C Rosser, Kristine Oleinika, Kiran Nistala, Thomas Krausgruber, André F. Rendeiro, Aggelos Banos, Ignat Drozdov, Matteo Villa, Scott Thomson, Georgina Xanthou, Christoph Bock, Brigitta Stockinger, Claudia Mauri. Aryl Hydrocarbon Receptor Contributes to the Transcriptional Program of IL-10-Producing Regulatory B Cells. Cell Reports. (2019) doi:10.1016/j.celrep.2019.10.018
3. Florian Puhm, Taras Afonyushkin, Ulrike Resch, Georg Obermayer, Manfred Rohde, Thomas Penz, Michael Schuster, Gabriel Wagner, André F. Rendeiro, Imene Melki, Christoph Kaun, Johann Wojta, Christoph Bock, Bernd Jilma, Nigel Mackman, Eric Boilard, Christoph J Binder. Mitochondria are a subset of extracellular vesicles released by activated monocytes and induce type I IFN and TNF responses in endothelial cells. Circulation Research. (2019) doi:10.1161/CIRCRESAHA.118.314601
4. Sandra Schick, André F. Rendeiro, Kathrin Runggatscher, Anna Ringler, Bernd Boidol, Melanie Hinkel, Peter Májek, Loan Vulliard, Thomas Penz, Katja Parapatics, Christian Schmidl, Jörg Menche, Guido Boehmelt, Mark Petronczki, André C. Müller, Christoph Bock, Stefan Kubicek. Systematic characterization of BAF mutations provides insights into intracomplex synthetic lethalities in human cancers. Nature Genetics (2019). doi:10.1038/s41588-019-0477-9
5. Sara Sdelci, André F. Rendeiro, Philipp Rathert, Wanhui You, Jung-Ming G. Lin, Anna Ringler, Gerald Hofstätter, Herwig P. Moll, Bettina Gürtl, Matthias Farlik, Sandra Schick, Freya Klepsch, Matthew Oldach, Pisanu Buphamalai, Fiorella Schischlik, Peter Májek, Katja Parapatics, Christian Schmidl, Michael Schuster, Thomas Penz, Dennis L. Buckley, Otto Hudecz, Richard Imre, Shuang-Yan Wang, Hans Michael Maric, Robert Kralovics, Keiryn L. Bennett, Andre C. Müller, Karl Mechtler, Jörg Menche, James E. Bradner, Georg E. Winter, Kristaps Klavins, Emilio Casanova, Christoph Bock, Johannes Zuber & Stefan Kubicek. MTHFD1 interaction with BRD4 links folate metabolism to transcriptional regulation. Nature Genetics (2019). doi:10.1038/s41588-019-0413-z
6. Alexander Swoboda, Robert Soukup, Katharina Kinslechner, Bettina Wingelhofer, David Schoerghofer, Christina Sternberg, Ha Pham, Maria Vallianou, Jaqueline Horvath, Dagmar Stoiber, Lukas Kenner, Lionel Larue, Valeria Poli, Friedrich Beer-mann, Takashi Yokota, Stefan Kubicek, Thomas Krausgruber, André F. Rendeiro, Christoph Bock, Rainer Zenz, Boris Kovacic, Fritz Aberger, Markus Hengstschlaeger, Peter Petzelbauer, Mario Mikula, Richard Moriggl. STAT3 promotes melanomametastasis by CEBP-induced repression of the MITF pigmentation path-way. bioRxiv (2018). doi:10.1101/422832
7. Tahsin Stefan Barakat*, Florian Halbritter*, Man Zhang, André F. Rendeiro, Christoph Bock, Ian Chambers. Functional dissection of the enhancer repertoire in human embryonic stem cells. Cell Stem Cell. (2018) doi:10.1016/j.stem.2018.06.014
8. Paul Datlinger, André F. Rendeiro*, Christian Schmidl*, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda C Schuster, Amelie Kuchler, Donat Alpar, Christoph Bock. Pooled CRISPR screening with single-cell transcriptome readout. Nature Methods. (2017) doi:10.1038/nmeth.4177
9. Roman A Romanov, Amit Zeisel, Joanne Bakker, Fatima Girach, Arash Hellysaz, Raju Tomer, Alán Alpár, Jan Mulder, Frédéric Clotman, Erik Keimpema, Brian Hsueh, Ailey K Crow, Henrik Martens, Christian Schwindling, Daniela Calvigioni, Jaideep S Bains, Zoltán Máté, Gábor Szabó, Yuchio Yanagawa, Ming-Dong Zhang, André F. Rendeiro, Matthias Farlik, Mathias Uhlén, Peer Wulff, Christoph Bock, Christian Broberger, Karl Deisseroth, Tomas Hökfelt, Sten Linnarsson, Tamas L Horvath & Tibor Harkany. Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. Nature Neuroscience (2017). doi:10.1038/nn.4462

10. Clara Jana-Lui Busch, Tim Hendrikx, David Weismann, Sven Jäckel, Sofie M. A. Walenbergh, André F. Rendeiro, Juliane Weißer, Florian Puhm, Anastasiya Hladik, Laura Göderle, Nikolina Papac-Milicevic, Gerald Haas, Vincent Millischer, Saravanan Subramaniam, Sylvia Knapp, Keiryn L. Bennett, Christoph Bock, Christoph Reinhardt, Ronit Shiri-Sverdlov, Christoph J. Binder. Malondialdehyde epitopes are sterile mediators of hepatic inflammation in hypercholesterolemic mice. Hepatology (2017). doi:10.1002/hep.28970
11. Michaela Schwaiger, Anna Schönauer, André F. Rendeiro, Carina Pribitzer, Alexandra Schauer, Anna F Gilles, Johannes B Schinko, Eduard Renfer, David Fredman, Ulrich Technau. Evolutionary conservation of the eumetazoan gene regulatory landscape. Genome Research (2014). doi:10.1101/gr.162529.113
12. André F. Rendeiro, Pavla Navratilova, Eric Thompson. Chromatin preparation for ChIP-seq in Oikopleura dioica. Figshare (2014). doi:10.6084/m9.figshare.884562

Communications

Conference talks

1. André F. Rendeiro. Chromatin mapping and single-cell immune profiling define the temporal dynamics of Ibrutinib response in CLL. Young Scientist Association of the Medical University of Vienna PhD Symposia, June 2019, Vienna, Austria.
2. André F. Rendeiro. Chromatin mapping and single-cell immune profiling define the temporal dynamics of Ibrutinib response in CLL. Frontiers in Single Cell Genomics Meeting - Cold Spring Harbour Asia, November 2018, Suzhou, China.
3. André F. Rendeiro. CROP-seq: updates on the single cell CRISPR screening method. 10X User Group Meeting 2018, April 2018, EMBL, Heidelberg, Germany.
4. André F. Rendeiro. Pooled CRISPR screening with single-cell transcriptome read-out. SLAS 2018, February 2018, San Diego, USA.
5. André F. Rendeiro. Large-scale ATAC-seq profiling to identify disease subtypes, regulatory networks and monitoring treatment in CLL. Illumina User Group Meeting 2017, February 2018, Switzerland.
6. Paul Datlinger, André F. Rendeiro*, Christian Schmidl*, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda C Schuster, Amelie Kuchler, Donat Alpar, Christoph Bock. Pooled CRISPR screening with single-cell transcriptome readout. Ascona Work- shop 2017, May 2017, Ascona, Switzerland.
7. André F. Rendeiro. Large-scale chromatin profiling uncovers heterogeneity of molecular phenotypes and gene regulatory networks of chronic lymphocytic leukemia. Illumina User Meeting, February 2017, Cologne, Germany.
8. Michaela Schwaiger, Anna Schönauer, André F. Rendeiro, Carina Pribitzer, Alexandra Schauer, Anna Gilles, Johannes Schinko, David Fredman, and Ulrich Technau. Evolutionary conservation of the eumetazoan gene regulatory landscape. XVIII Portuguese Genetics Society Meeting, June 2013. Porto, Portugal.

Conference posters

1. André F. Rendeiro*, Thomas Krausgruber*, Nikolaus Fortelny, Fangwen Zhao, Thomas Penz, Matthias Farlik, Linda C. Schuster, Amelie Nemc, Szabolcs Tasnády, Marienn Réti, Zoltán Mátrai, Donat Alpar, Csaba Bödör, Christian Schmidl, Christoph Bock. Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib drug response in chronic lymphocytic leukemia. SCOG Workshop Computational Single Cell Genomics, May 2019. Munich, Germany. doi:10.6084/m9.figshare.7892663.v1

2. Christian Schmidl*, André F. Rendeiro*, Gregory I Vladimer*, Thomas Krausgruber, Tea Pemovska, Nikolaus Krall, Berend Snijder, Oscar Lopez de la Fuente, Anna Ringler, Stefan Kubicek, Philipp B. Staber, Medhat Shehata, Giulio Superti-Furga, Ulrich Jäger, Christoph Bock. Combined chromatin accessibility and chemosensitivity profiling identifies targetable pathways and rational drug combinations in Ibrutinib-treated chronic lymphocytic leukemia. Young Scientist Association of the Medical University of Vienna PhD Symposia, June 2017. Vienna, Austria
3. André F. Rendeiro*, Christian Schmidl*, Jonathan C. Strefford*, Renata Walewska, Zadie Davis, Matthias Farlik, David Oscier, Christoph Bock. Large-scale chromatin profiling uncovers heterogeneity of molecular phenotypes and gene regulatory networks of chronic lymphocytic leukemia. Young Scientist Association of the Medical University of Vienna PhD Symposia, June 2016. Vienna, Austria. doi:10.6084/m9.figshare.3479528.v1 Best poster award in "Malignant Diseases" category.
4. André F. Rendeiro*, Christian Schmidl*, Jonathan C. Strefford*, Renata Walewska, Zadie Davis, Matthias Farlik, David Oscier, Christoph Bock. Large-scale chromatin profiling uncovers heterogeneity of molecular phenotypes and gene regulatory networks of chronic lymphocytic leukemia. Keystone Symposia on Chromatin and Epigenetics, March 2016. Whistler, Vancouver, Canada. doi:10.6084/m9.figshare.3479528.v1
5. Anna Schönauer, André F. Rendeiro, Michaela Schwaiger, Ulrich Technau. Identification of cis-regulatory elements in the sea anemone Nematostella vectensis. Evonet Symposium, September 2012. Vienna, Austria. doi:10.6084/m9.figshare.107026

Skills

Data science	Pipeline development for NGS data preprocessing; ATAC-/ChIP-/RNA-seq data analysis; single cell RNA-/ATAC-seq analysis Data-driven and statistical analysis; Machine learning; Bayesian modeling with probabilistic programming;
Programming	Experienced in Python and R programming; Beginner level Rust, V programming; Software development: version control, testing, versioning, continuous integration.
Mol. Biology	Chromatin imunoprecipitation (ChIP), NGS library preparation, Western blotting, PCR, molecular cloning

Additional experience

Scientific Activity

2013-2014	The role of E2F regulation and H3K79 methylation in *Oikopleura dioica*'s cell cycle Sars International Centre for Marine Molecular Biology, Bergen, Norway Laboratory of Eric Thompson
2011-2012	Identification of cis-regulatory elements in *Nematostella vectensis* using ChIP-seq Dept. of Molecular Evolution and Development, University of Vienna, Austria Laboratory of Uli Technau
2010-2011	Tol2-mediated zebrafish transgenesis for studies in protein mistranslation Biology Department, University of Aveiro, Portugal Laboratory of Manuel Santos
2009-2010	Transciptome studies with microarrays in yeast Biology Department, University of Aveiro, Portugal Laboratory of Manuel Santos

Associative/Administrative

2010-2012 Member of the Biology department counsel, University of Aveiro, Portugal
2009-2011 Member of the undergraduate Biology program committee
University of Aveiro, Portugal

Advanced courses

2015/09 Summer School on Machine Learning for Personalized Medicine
Marie Curie Initial Training Network, Manchester, UK
2012/09 Scientific writing course,
Biology Department, University of Aveiro, Portugal

Awards/Scholarships

2016/06 Best poster award - "Malignant diseases" category, YSA Symposium
Young Scientist Association of the Medical University of Vienna, Austria
2016/06 Best artwork award - "Illustrations and digital simulations" category
“ScienceArt” Competition of the YSA Symposium
Young Scientist Association of the Medical University of Vienna, Austria
2013-2014 Erasmus studies mobility program scholarship
European Commission
2011-2012 Erasmus internship mobility program scholarship
European Commission
2009-2010 "Integration into Research" Grant
Science and Technology Foundation – Portugal

Languages

Portuguese	Native speaker	German	Basic knowledge
English	Professional competence	French	Basic knowledge
Spanish	Conversational		