

Dissertation
submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
M.A. Holly Amelia Rebecca Giles
born in Portsmouth
Oral examination: Dec 15, 2021?

Drug - microenvironment - gene interplay in chronic lymphocytic
leukemia

Referees: Dr. Judith Zaugg

Prof. Dr. Michael Boutros

To . . .

Acknowledgements

I want to thank a few people.

List of publications

Thesis related

- Peter-Martin Bruch*, Holly A. R. Giles*, Carolin Kolb, Sophie Herbst, Tina Becirovic, Tobias Roider, Junyan Lu, Sebastian Scheinost, Lena Wagner, Jennifer Huellein, Ivan Berest, Mark Kriegsmann, Katharina Kriegsmann, Christiane Zgorzelski, Peter Dreger, Judith Zaugg, Carsten Mueller-Tidow, Thorsten Zenz, Wolfgang Huber, Sascha Dietrich et al. *in preparation.* "Mapping drug-microenvironment-genetic interplay in CLL reveals trisomy 12 as a modulator of microenvironment." *bioRxiv*. [doi: 10.1101/2021.07.23.453514](<https://doi.org/10.1101/2021.07.23.453514>)

Other contributions

The author of this thesis also contributed to a number of other projects throughout the PhD. The following have been published:

- Berest, Ivan*, Christian Arnold*, Armando Reyes-Palomares, Giovanni Palla, Kasper Dindler Rasmussen, Holly Giles, and Peter-Martin Bruch et al. 2019. "Quantification Of Differential Transcription Factor Activity And Multiomics-Based Classification Into Activators And Repressors: DiffTF". *Cell Reports* 29 (10): 3147-3159.e12. [doi: 10.1016/j.celrep.2019.10.106](<https://doi.org/10.1016/j.celrep.2019.10.106>)
- Lu, Junyan*, Ester Cannizzaro*, Fabienne Meier-Abt, Sebastian Scheinost, Peter-Martin Bruch, Holly A. R. Giles, and Almut Lütge et al. 2021. "Multi-Omics

Reveals Clinically Relevant Proliferative Drive Associated With Mtor-MYC-OXPHOS Activity In Chronic Lymphocytic Leukemia". *Nature Cancer*. [doi: 10.1038/s43018-021-00216-6](<https://doi.org/10.1038/s43018-021-00216-6>)

Zusammenfassung

Abstract in German Aktuelle Entwicklungen im Bereich der “omik”-Technologien tragen wesentlich zur Beschleunigung des Fortschritts in der Krebsforschung bei etc

Abstract

The tumour microenvironment and genetic alterations collectively influence disease biology and drug resistance in Chronic Lymphocytic Leukaemia (CLL). To establish an integrative understanding of these factors in CLL biology, we performed a combinatorial assay using 12 drugs individually co-applied with each of 17 microenvironmental stimuli on 192 samples of CLL peripheral blood mononucleated cells. We examined microenvironment response across a heterogeneous patient cohort and identified four distinct CLL subgroups that differed in their response landscapes and in patient outcomes. By combining our data with whole-exome sequencing, DNA-methylation, RNA-sequencing and copy number variant data of the same tumours, we systematically searched for molecular determinants of stimulus response and found trisomy 12 as a key modulator. Our data suggest that the amplifying effect of trisomy 12 on the response to environmental signals is mediated by the transcription factors Spi-B and PU.1. We generated a comprehensive map of drug-microenvironment interactions in CLL, and profiled the modulating impact of genetic features on these antagonistic and synergistic effects. Interleukin (IL) 4 and Toll-Like Receptor (TLR) 7/8/9 stimuli showed the most interactions. Both pathways were more active in CLL-infiltrated lymph nodes than in healthy samples ($p<0.001$), and high IL4 activity in lymph nodes correlated with shorter survival ($p=0.038$). We provide a multi-layered resource to investigate microenvironmental and drug interplay in CLL (Repository & Shiny). Our results highlight the importance of cell-extrinsic influences on drug response and disease progression, and how these further depend on molecular features.

Table of Contents

Acknowledgements	i
List of publications	iii
Zusammenfassung	vii
Abstract	ix
List of Abbreviations	xiv
List of Figures	xv
List of Tables	xvii
1 Introduction	1
1.0.1 Disease Characteristics	1
1.0.2 IGHV status and the role of BCR signalling	3
1.0.3 Recurrent genetic features in CLL	9
1.0.4 The incompletely understood role of trisomy12	17
1.0.5 Stratification of CLL patients	18
1.0.6 Prognosis and where the problems lie	21
1.1 The tumour microenvironment in CLL	24
1.1.1 The concept of the protective niche	24
1.1.2 Microenvironmental pathways and cross-talk	25

1.2	Therapies in CLL	27
1.2.1	Clinical regimes	27
1.2.2	New Targeted therapies	29
1.2.3	Drug resistances in CLL	30
1.2.4	BET inhibitors, and their potential for the treatment of leukemia .	30
1.3	Background to the approach used in this thesis	30
1.3.1	Ex-vivo drug sensitivity screens	30
1.3.2	Studies of the microenvironment, and attempts to replicate the impact of the microenvironment	31
1.3.3	ATAC sequencing as a proxy for TF activity justification	31
1.3.4	Measurement of outcomes?	31
1.3.5	Multiomics datasets to study CLL	32
1.3.6	Mathematical modelling	32
1.3.7	CLL as a model for studying tumour biology	32
1.4	Aims of this thesis	33
1.5	Overview of the approach	33
1.6	Note on contributions	33
2	Methods	35
2.1	Experimental methods:	35
2.1.1	Drug - Stimulus Combinatorial Perturbation Assay	35
2.1.2	Preparation of screening plates	36
2.2	Drug-Stimulation assay	37
2.3	WES, RNA Sequencing, Targeted Sequencing and DNA Copy number variants	37
2.3.1	Follow - up investigations	37
2.3.2	Investigation of Ibrutinib + IL4 + IBET-762	39
2.4	Data availability	41
2.5	Statistical Analysis	41

2.5.1	Data processing	41
2.5.2	Analysis of Screening Data	42
2.5.3	Follow up	44
2.5.4	Ibrutinib - IBET-762 - IL-4 analysis	49
3	Data	51
3.1	Drug screens and experiments	51
3.1.1	High-throughput combinatorial perturbation assay	51
3.1.2	Validation experiments	51
3.2	Characteristics of drugs used in the screen	52
3.2.1	Drug pathways	52
3.2.2	Drug responses	52
3.2.3	Genetic predictors of drug responses	52
3.2.4	Drug - Drug Correlations	52
3.3	Characteristics of stimuli used in the screen	52
3.3.1	Stimulus responses	52
3.3.2	Stimulus - Stimulus Correlations	52
3.4	Characteristics of patient samples used in the screen	53
3.4.1	Genetic Data available for each patient	53
3.5	Processing of raw values obtained from cell viability assay	53
3.5.1	Data normalization and quality control	53
3.5.2	How viability is calculated, why we use log values etc	53
3.5.3	Heterogeneity of response is not caused by differences in receptor expression	53
3.6	Generation of public resource	53
3.6.1	Shiny app	53
3.6.2	Package	54
4	Ex-vivo sensitivity to microenvironmental stimulation in primary CLL cells	55

4.1	Prolifing responses to the panel of stimuli	56
4.1.1	<i>ex vivo</i> assay demonstrated functional diversity of cytokines and microenvironmental stimuli	56
4.1.2	Microenvironmental response profiling identifies discrete patient subgroups	58
4.2	Functional characterisation of patient clusters	61
4.2.1	C1 - C4 showed distinct response profiles with the panel of stimuli	61
4.2.2	The clusters show differences in disease dynamics	62
4.2.3	The clusters showed differential responses to drugs <i>in vitro</i>	63
4.2.4	The clusters are enriched for different genetic features	64
4.2.5	GSEA of DE genes between subgroups	65
4.3	Additional Analysis	66
4.4	Summary	66
4.5	Discussion	67
5	Genetic modulators of responses to microenvironmental stimulation	75
5.1	Systematic analysis of the effect of genetic features on responses to stimuli	76
5.1.1	Univariate analysis identifies IGHV status and trisomy 12 as key modulators of microenvironmental response	76
5.1.2	Multivariate analysis of gene - stimulus assosciations confirms IGHV status and trisomy 12 as key modulators of stimulus response	77
5.1.3	Response to TLR stimulation is dependent on IGHV status, trisomy 12 and mutations in DNA Damage Response genes	78
5.2	Investigating Trisomy 12 as a modulator of microenvironmental response .	80
5.2.1	Trisomy 12 is a modulator of microenvironmental response	80
5.2.2	STAT6, IRAK4 and SMAD3 are more highly expressed in trisomy 12 CLL	81
5.2.3	Classification analysis identifies trisomy 12 phenocopies that show increased expression of IRAK4 and SMAD3	82

5.2.4	Spi-B and PU.1 TFs show higher activity in trisomy 12 CLL	84
5.2.5	Spi-B and PU.1 targets are enriched for immune signalling pathways	85
5.2.6	Double knockdown of Spi-B and PU.1 reduces proliferation of tri- somy 12+ cell lines	86
5.3	Summary	86
5.4	Discussion	87
5.5	Contributions statement	88
6	Results	101
6.1	Review of Results	102
7	Results	103
7.1	Review of Results	104
8	Discussion	105
References		107
Appendix		113
Figures	113
Tables	114

List of Abbreviations

AKT	Protein kinase B
AUC	Area under curve
BCR	B-cell receptor
BH	Benjamini-Hochberg
BTK	Brutons tyrosine kinase
CDF	Cumulative Distribution Function
CLL	Chronic lymphocytic leukemia
CNV	Copy number variation
CpG ODN	CpG oligodeoxynucleotides
ERK/MAPK signalling	Mitogen-activated protein kinase signalling
FBS	Fetal bovine serum
FDR	False discovery rate
GSEA	Gene set enrichment analysis
Ig	Immunoglobulin
IGHV	Immunoglobulin heavy chain variable region
IL	Interleukin
JAK	Janus kinase
M-CLL	CLL with somatic hypermutations in IGHV loci
NF κ B	Nuclear factor kappa-light-chain-enhancer of activated B cells
NOTCH	Neurogenic locus notch
OS	Overall survival
PBMC	Peripheral blood mononuclear cells
PCA	Principal component analysis
PI3K	Phosphoinositide 3-kinase
U-CLL	CLL without somatic hypermutations in IGHV loci
TGF β	Transforming growth factor β
TF	Transcription factor
TLR	Toll-like receptor
TSS	Transcription start site

List of Figures

1.1	(ref:CLLmorphology)	2
1.2	(ref:proliferationCentres)	3
1.3	(ref:BCRsignalling)	5
1.4	(ref:IGHVstatus)	6
1.5	(ref:CLLmutations)	13
1.6	(ref:CLLstratification)	20
1.7	(ref:microenvironmentTissues)	25
1.8	(ref:microenvironmentOverview)	26
1.9	(ref:microenvironmentSignalling)	27
3.1	Schematic of experimental protocol. By combining 12 drugs and 17 stimuli, we systematically queried the effects of simultaneous stimulation and inhibition of critical pathways in CLL (n=192). Integrating functional drug-stimulus response profiling with four additional omics layers, we identified pro-survival pathways, underlying molecular modulators of drug and microenvironment responses, and drug-stimulus interactions in CLL. . .	52
3.2	Schematic of validationary data.	52
3.3	Bar plot of the drugs used in screen.	52
3.4	Log transformed viability values for all drugs that were included in the screen after quality control. p values from student's t test.	53

4.1	Heatmap of Pearson correlation coefficients. Coefficients for each pair of stimuli were calculated using log transformed viability values normalised to untreated control, and ordered according to hierarchical clustering. Figure adapted from Bruch & Giles et al. 2021.	56
4.2	Scatter plot of log-transformed viability values, normalised to DMSO controls, for (A) treatment with JAK-STAT agonists IL4 and IL6 and (B) treatment with TLR agonists CpG ODN and Resiquimod.	57
4.3	Log transformed viabilities after treatment with each stimulus. Where stimuli decreased viability relative to control, points are shown in blue, whilst increased viability is shown in red. Figure adapted from Bruch & Giles et al. 2021.	58
4.4	The heatmap matrix shows the viability measurements for 192 samples (columns) and 17 stimuli (rows). Viability was measured via ATP-based assay after 48h stimulation. The data are shown normalised to DMSO-treated controls, and scaled within each row according to the Median Absolute Deviation (MAD). Limits were applied to scaling factor for optimal visualisation. The colour bars to the right show sample annotations. Consensus Clustering was used to define column tree layout, using hierarchical clustering with the Euclidean metric. Figure from Bruch & Giles et al. 2021.	59
4.5	Summaries of the CDFs of the consensus matrices. Consensus CDF graphic showing the CDFs of the consensus matrix for $k = 2 - 7$, as indicated in the legend, estimated using 100 bin histogram (left). Relative change in area under the CDF curve, for $k = 2 - 7$, to compare k with $k - 1$. In the case of $k = 2$, there is no $k - 1$, so the total area is plotted. Line shows relative increase in consensus between each value of k (right).	60
4.6	Assignment of patient samples (columns), to each cluster, for $k = 1 - 7$ (rows) to demonstrate stability of cluster membership. Cluster colour for $k = 4$ match those in heatmap in 4.4.	61

4.7 (A) Lymphocyte doubling time (LDT) stratified by cluster, p-values from Student's t-test. (B) Kaplan-Meier curves to show TTT for each cluster. p-values from univariate Cox proportional hazard models comparing IGHV-U enriched C1 with C2, and IGHV-M enriched C3 with C4. Figure from Bruch & Giles et al. 2021.	63
4.8 Log-transformed normalised viability values, stratified by cluster, for each drug. Drugs targeting the same pathway are grouped together. P-values from Student's t-test.	68
4.9 Distribution of selected genetic features (rows) within each cluster for all patient samples (columns). Where a patient sample is not annotated for a feature, this is marked in white.	69
4.10 Multinomial regression with lasso penalisation to identify enrichment or depletion of genetic features within each cluster. Matrix of genetic features ($p=39$), and IGHV status (encoded as M = 1 and U = 0) were used to identify multivariate predictors of cluster assignment. x axis shows genetic predictors, y axis indicates value and sign of coefficient assigned to feature, for each cluster (positive coefficients are enriched in the cluster, negative coefficients are depleted). Coefficients shown are mean coefficients from 50 bootstrapped repeats and error bars represent the mean \pm standard deviation. Genetic features with $>20\%$ missing values were excluded, and only patients with complete annotation were included in the model ($n=137$). Figure from Bruch & Giles et al 2021.	70
4.11 Volcano plot of differentially expressed genes between C3 and C4. X axis indicates log ₂ fold change values, calculated using the DESeq2 package (DESeq2?), y axis gives corresponding -log ₁₀ (adjusted p value). P values adjusted using BH method. Genes are labeled where adjusted p < 0.05. Figure from Bruch & Giles et al. 2021.	71

5.2 Predictor profiles for IL4 and CpG ODN depicting gene - stimulus associations. Bar plots (left) show size and sign of assigned coefficients from Gaussian linear modelling. A positive coefficient indicates that stimulated increase in viability is larger when feature is present. Scatter plots (bottom) and corresponding heatmaps above show how presence of selected genetic feature relates to sample viabilities. Scatter plots show ranked log(viability) values for each sample and heatmaps show mutation status for each predictor, for corresponding sample in scatter plot. See method 2.5.2. <i>Figure adapted from Bruch & Giles et al. 2021.</i>	90
5.3 Control-normalised log transformed viability values after treatment with Resiquimod (TLR 7/8), stratified by named genetic features.	91
5.4 Beeswarm-boxplot showing control-normalised log transformed viability values, after treatment with Resiquimod, stratified by trisomy 12 and IGHV status. p-values from Student's t-tests. <i>Figure from (Bruch & Giles et al. 2021)</i>	92
5.5 Control-normalised log transformed viability values after treatment with IL4, sCD40L + IL4 and TGF β , stratified by trisomy 12.	92
5.6 Beeswarm-boxplots showing RNA counts and protein abundances stratified by trisomy 12 status. P values from Student's t-test. <i>_Proteomics dataset from (HerbstThesis_?) Method: 2.5.3.</i>	93
5.7 Predictor profile showing coefficients selected as predictors of trisomy 12 status. Here, binomial modelling with L1-penalty was used to identify associations between responses to stimuli and trisomy 12 status. Bar plots (left) show size and sign of assigned coefficients named on the right. A positive coefficient indicates that higher viability upon treatment named stimulus is associated with trisomy 12. Facet labels (top) and corresponding heatmap below show how the viability with each named stimulus relates to trisomy 12 status. The model shown was selected from 50 bootstrapped repeats, based on maximal AUC. Method 2.5.3.	94

5.8 Control-normalised log-transformed viability values, for all samples after treatment with Resiquimod, sCD40L + IL4 and TGF β . Patient sample A and B are indicated in blue and red, respectively.	94
5.9 Raw RNA counts for SMAD3, IRAK4 and STAT6, stratified by trisomy 12 status. Patient sample A and B are indicated in blue and red, respectively. P values from Student's t-test.	95
5.10 diffTF workflow from Berest et al. (2019). diffTF accepts a list of TFs along with the genomic locations of their binding sites. For each TF, the software computes the distribution of fold changes between the trisomy 12 and WT samples, using the ATACseq peaks at each TF binding site in each condition. The software compares this distribution to a background set of fold changes generated using GC content-matched loci that do not contain the same TF binding site motif. Each TF is thus assigned assigned a weighted mean difference value, which quantifies the change in activity, and a p value.	96
5.11 Bar plot showing the results of the diffTF analysis for the Rendeiro et al. (2016) dataset. y axis shows change in TF activity (weighted mean difference) between trisomy 12 (n = 9) and non-trisomy 12 samples (n = 43),x axis indicates TF names. 17 / 636 TFs, with BH adjusted p <0.05 are shown. p values generated by diffTF in permutation mode. TF binding sites defined in HOCOMOCO v10 (Kulakovskiy et al. 2016).	97

5.12	Table shows results from over-representation tests of selected KEGG and Reactome pathways amongst Spi-B and PU.1 targets. Columns show geneset pathways, corresponding database, the number of genes within geneset, the number of Spi-B and PU.1 target genes within geneset (total number of target genes defined also shown), and p-value from over-representation test. TF targets defined as closest gene to each significant ChIP peak (q value<0.05), and within \$±\$1kb of TSS. ChIPseq data from Care et al. (2014). Over-representation tests run using (clusterProfiler?) package, method corresponds to one-sided version of Fisher's exact test. <i>Figure and caption from Bruch & Giles et al 2021.</i>	98
5.13	Knockdown of Spi-B and PU.1 led to growth restriction and cell death. Data shown is the mean of three technical replicates. While SU-DHL 5 was already growth impaired after Spi-B KD, SU-DHL 2 and SU-DHL 4 were only impaired after double knock-down. Double knockdown lead to rapid cell death in SU-DHL 5 (data not shown). Cell count normalised to seeded cell number of diffuse large b cell lymphoma cell lines after knockdown of Spi-B or double knockdown of Spi-B and PU.1 or shRNA as control. <i>Figure and caption from Bruch & Giles et al. 2021.</i>	99
6.1	Estimated residuals from model XXX3.	102
7.1	Estimated residuals from model XXX4.	104
8.1	Analysis of ATACseq dataset of two trisomy 12 and two non-trisomy 12 untreated CLL PMBC samples. The volcano plot depicts change in TF activity (x axis) versus BH-adjusted p-values (y axis) for 636 TFs, comparing trisomy 12 and non-trisomy 12 samples. The (diffTF?)[REFERENCE] package was ran in analytical mode to calculate TF activity, measured as weighted mean difference. TFs are labeled if adjusted p-value < 0.01 and absolute weighted mean difference > 0.15.	113

List of Tables

4.1	Table depicting results of Multivariate Cox Proportional Hazard Model to test prognostic value of key genetic features and clusters using Time to Next Treatment and C3 as reference.	64
8.1	Detailed descriptive statistics of location and dispersion for 2100 observed swap rates for the period from February 15, 1999 to March 2, 2007. Swap rates measured as 3.12 (instead of 0.0312).	114

Chapter 1

Introduction

Next tasks: Give a bit of time each day for next few weeks, then full tie when chapter 5 and 6 finished Continue to go through add the next papers Then go through each section, do a lit review, and read what i need for each, add the text and references and keep going Should then be easy to tie this together

(theoretical background) ## CLL Disease biology

1.0.1 Disease Characteristics

CD5, CD19, CD23+, low levels of CD20

Reading:

From Kipps 2017: Incidence (get for 2021) Siegel R, et al. Cancer treatment and survivorship statistics, 2012. CA Cancer J. Clin. 2012; 62:220–241. [PubMed: 22700443]

1.1 Taken from Kipps et al 2017: “Wright–Giemsa-stained blood smears showing the typical chronic lymphocytic leukaemia (CLL) B lymphocyte (part a). Figure from (Kipps 2017.?)” *Figure taken from Kipps et al 2017*

1.2 Taken from Kipps et al 2017: “a | Tissue sections of a lymph node stained with haematoxylin and eosin showing numerous pale-staining pseudofollicles, which are circled

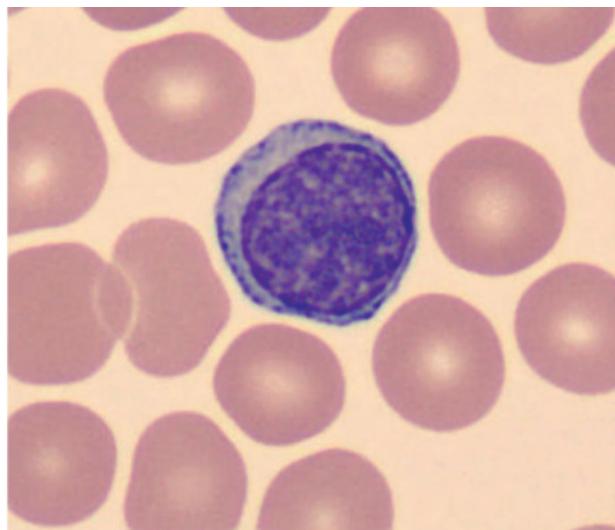


Figure 1.1: (ref:CLLmorphology)

(original magnification $\times 20$).” —Figure from (**Kipps2017**) *Figure taken from Kipps et al 2017*

(Kipps et al. 2017) 10.1038/nrdp.2016.96.Chronic “Chronic lymphocytic leukaemia (CLL) is a malignancy of CD5+ B cells that is characterized by the accumulation of small, mature-appearing lymphocytes in the blood, marrow and lymphoid tissues. Signalling via surface immunoglobulin, which constitutes the major part of the B cell receptor, and several genetic alterations play a part in CLL pathogenesis, in addition to interactions between CLL cells and other cell types, such as stromal cells, T cells and nurse-like cells in the lymph nodes. The clinical progression of CLL is heterogeneous and ranges from patients who require treatment soon after diagnosis to others who do not require therapy for many years, if at all. Several factors, including the immunoglobulin heavy-chain variable region gene (IGHV) mutational status, genomic changes, patient age and the presence of comorbidities, should be considered when defining the optimal management strategies, which include chemotherapy, chemoimmunotherapy and/or drugs targeting B cell receptor signalling or inhibitors of apoptosis, such as BCL-2. Research on the biology of CLL has profoundly enhanced our ability to identify patients who are at higher risk for disease progression and our capacity to treat patients with drugs

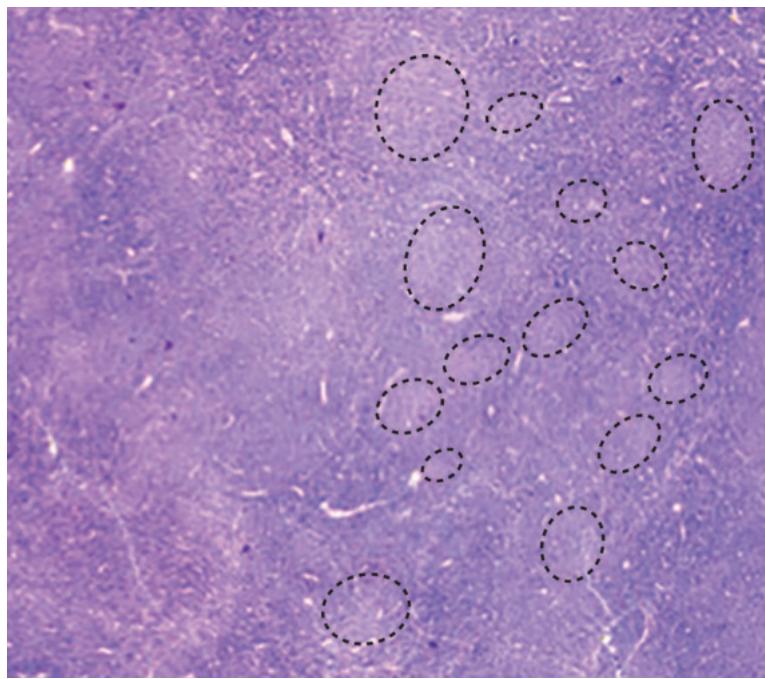


Figure 1.2: (ref:proliferationCentres)

that selectively target distinctive phenotypic or physiological features of CLL.”

Get for 2021: CLL is estimated to account for ~19,000 of all newly detected cancers in the United States in 2016 (REF. 11). The average incidence of CLL varies between individuals in different geographical regions and ranges from <0.01% of individuals in eastern Asia to ~0.06% of individuals in Europe and the United States. The risk of developing CLL is about two-times higher for men than for women and increases with age; the median age at diagnosis ranges from 70 to 72 years^{11–14}. The

1.0.2 IGHV status and the role of BCR signalling

Reading: From Kipps 2017: IGHV status: Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*. 1999; 94:1848– 1854. [PubMed: 10477713]

Damle RN, et al. Ig V gene mutation status and CD38 expression as novel prognostic

indicators in chronic lymphocytic leukemia. *Blood.* 1999; 94:1840–1847. [PubMed: 10477712] References 1 and 2 are landmark papers that describe two main subsets of patients with different disease progression tendencies based on IGHV mutation status of the immunoglobulins that are expressed by CLL cells.

Restricted repertoire: Kipps TJ, et al. Developmentally restricted immunoglobulin heavy chain variable region gene expressed at high frequency in chronic lymphocytic leukemia. *Proc. Natl Acad. Sci. USA.* 1989; 86:5913–5917. [PubMed: 2503826] This paper describes the discovery that the immunoglobulin repertoire of CLL cells may be highly restricted, suggesting that the antibodies expressed by CLL cells are most likely selected based on their capacity to bind to some common self-antigens.

Fais F, et al. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J. Clin. Invest.* 1998; 102:1515–1525. [PubMed: 9788964]

Widhopf G, F 2nd, et al. Chronic lymphocytic leukemia B cells of more than 1% of patients express virtually identical immunoglobulins. *Blood.* 2004; 104:2499–2504. [PubMed: 15217828]

1.3 Adapted from: Kipps et al 2017: “B cell receptor (BCR) signalling is initiated by SRC-family kinase-dependent phosphorylation (mainly LYN) of CD79A and CD79B that creates a docking site for the binding and activation of spleen tyrosine kinase (SYK). SYK then triggers the formation of a multi-component ‘signalosome,’ comprising Bruton tyrosine kinase (BTK), AKT, phosphoinositide 3-kinase (PI3K), phospholipase C2 (PLC2) and B cell-linker protein (BLNK), among others. CD19 is a co-receptor for BCR and is important for PI3K activation, which recruits and activates PLC2, BTK and AKT. PLC2 generates diacylglycerol (DAG) and inositol-1,4,5-trisphosphate (Ins(1,4,5)P3), which triggers Ca²⁺ release from the endoplasmic reticulum, leading to the activation of the MEK–extracellular signal-regulated kinase (ERK) and nuclear factor-B (NF-B) signalling pathways. Other effects of BCR signalling include activation of mechanistic target of rapamycin complex 1 (mTORC1) and of Rho-family GTPases, RAC1 and

RHOA, which can affect the cytoskeleton. Inhibitors of SYK, PI3K and BTK are shown. Note that this figure describes the main molecules and interactions that are involved in positive BCR signalling, but is not an exhaustive description of all signalling pathways or molecules activated. IKK, IB kinase; PKC, protein kinase C.” 1.4 Taken from Kipps et

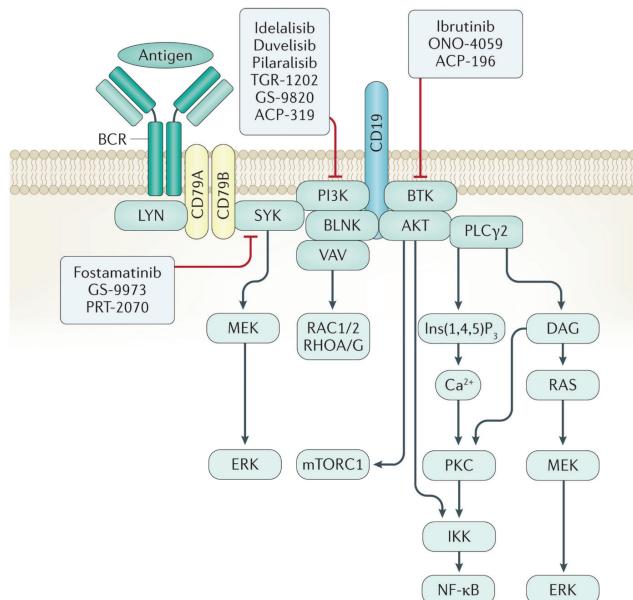


Figure 1.3: (ref:BCRsignalling)

al 2017: “Normal naive B cells that have undergone successful V(D)J recombination and express functional B cell receptors that are capable of binding to antigen interact with CD4+ T cells and accessory cells, which aggregate to form follicles that become germinal centres. Germinal cells each have a dark zone, comprising rapidly dividing B cells, and a light zone, comprising B cells mixed with follicular dendritic cells (FDCs), macrophages and helper T cells (TH cells). The B cells enter the dark zone of the germinal centre where they experience rapid proliferation and somatic hypermutation (SHM) in the genes encoding the immunoglobulin variable regions of the heavy chain (IGHV) and the light chain (IGVL). As they pass through to the light zone, the B cells that express the fittest B cell receptors for binding antigen are selected and may undergo immunoglobulin class-switch recombination. Chronic lymphocytic leukaemia (CLL) cells that use

unmutated IGHV apparently originate from CD5+ B cells prior to experiencing SHM, whereas CLL cells that use mutated IGHV most likely originate from CD5+ B cells that have passed through and differentiated in the germinal centre. Some CLL cells might be derived from B cells that also have undergone immunoglobulin class-switch recombination and express immunoglobulin isotypes other than IgM and IgD, for example, IgG or IgA. Another subset is one with CLL cells that express immunoglobulin with only modest somatic mutations, such as CLL cells that use IGHV3-21 with ~97% homology to the inherited IGHV3-21 gene and an immunoglobulin light chain encoded by an unmutated IGLV3-21; these cells might derive from a B cell that has had constrained SHM, possibly owing to a limited need for immunoglobulin somatic diversification and selection. Dashed arrows indicate speculated pathways.” _Figure from (Kipps2017_?). (Kipps

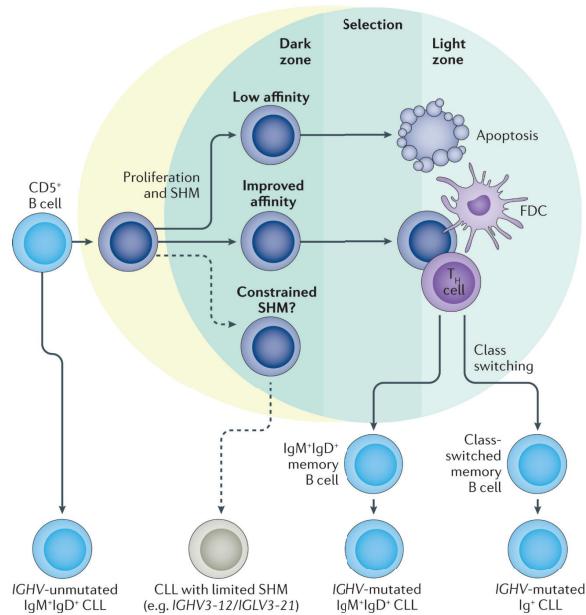


Figure 1.4: (ref:IGHVstatus)

et al. 2017) The BCR is composed of a ligand-binding trans-membrane immunoglobulin molecule (either IgA, IgD, IgE, IgG or IgM) and the signalling Ig (also known as CD79A)–Ig (also known as CD79B) heterodimer. CLL cells typically co-express IgD and IgM, although at low levels compared with normal B cells; less than a few per

cent of CLL cases express class-switched isotypes, most commonly IgG. The CD79A and CD79B molecules contain immunoreceptor tyrosine-based activation motifs, which can be phosphorylated following the crosslinking of surface immunoglobulin, thereby triggering BCR signalling. A functional BCR is required for the survival of mature B cells⁷⁶ and is maintained in most mature B cell malignancies, including CLL. In CLL, evidence suggests that the surface immunoglobulin of CLL B cells is engaged by autoantigen, which leads to constitutive BCR signalling *in vivo*^{77–79}. The importance of this interaction is underscored by the clinical success of kinase inhibitors that block BCR signalling (see Management), although effects on other receptors might also have a role⁸⁰. Like most cancers, CLL is heterogeneous and the outcome of BCR signalling ranges from enhanced B cell activation to B cell anergy^{81,82}. The main pathways that lead to cell survival and proliferation downstream of the BCR are shown in FIG. 3, along with drugs targeted against key signalling intermediates. BCR signalling that leads to anergy is less well defined, but seems to involve biased activation of inhibitory molecules with only partial activation of the pathways that are typically associated with B cell activation⁸¹. One important molecule that may be involved is the inositol lipid phosphatase SHIP1. SHIP1 is activated by the tyrosine-protein kinase LYN and may limit B cell activation by counteracting phosphoinositide 3-kinase (PI3K) activity at both chronically engaged receptors and distant non-ligated BCRs, rendering them insensitive to stimulation^{82,83}. Enhanced B cell activation is more commonly observed in CLL that expresses unmutated IGHV, whereas anergy predominates in most cases of CLL that express mutated IGHV⁸⁴. Anergy is a state of cellular lethargy induced by chronic engagement of the surface antigen receptors in the absence of adequate T cell help. Although capable of reversing their phenotype, anergic cells are less likely to proliferate in response to BCR signalling than more activated cells, which might, in part, account for the observation that patients with CLL cells that express mutated IGHV generally have more indolent disease than patients with CLL cells with unmutated IGHV⁸⁵. The fate of the cell (activation versus anergy) might be influenced by the CLL cell of origin (FIG. 1), as the cell types that can form CLL differ in their patterns of DNA methylation.

lation⁷³, and are likely to respond differently to autoantigens. An unresolved question is whether anergy can be reversed *in vivo*, mirroring what occurs *in vitro*⁷⁸. The BCR also coordinates the activity of other cell surface receptors, including integrins, such as 41 integrin. BCR stimulation can result in increased adhesion of CLL cells to 41 integrin substrates, for example, fibronectin and vascular cell adhesion protein 1 (REF. 86). By contrast, CXC-chemokine receptor 4 (CXCR4) is downmodulated by BCR engagement and both can trigger ‘inside-out’ signalling, resulting in the activation of 41 integrin^{87,88}. Thus, recognized antigen encountered in lymphoid tissue is likely to affect adhesion and migration of CLL cells. Modulation of these pathways, coupled with the role of BTK and PI3K in chemokine receptor signalling⁸⁹, contribute to the increased lymphocytosis observed in patients upon initiation of treatment with inhibitors of BTK or PI3K (see Management).

(Kipps et al. 2017) 10.1038/nrdp.2016.96.Chronic “CLL can be divided into two main subsets, which differ in their clinical behaviour. These subsets are distinguished by whether CLL cells express an unmutated or mutated immunoglobulin heavy-chain variable region gene (IGHV), reflecting the stage of normal B cell differentiation from which they originate^{1,2}. CLL cells that express an unmutated IGHV originate from a B cell that has not undergone differentiation in germinal centres, which are the sites in the lymph nodes where B cells experience somatic hypermutation in their immunoglobulin variable region genes and selection during an immune response. Patients with CLL cells that express an unmutated IGHV typically have more-aggressive disease than patients with CLL cells that express a mutated IGHV. CLL cells with mutated IGHV arise from a post-germinal centre B cell that expresses immunoglobulin that has undergone somatic hypermutation and, in some cases, also immunoglobulin isotype switching (FIG. 1), similar to what occurs in normal B cells during an immune response to antigen. It should be emphasized that the high level of somatic mutations that arise in IGHV in the germinal centre are a natural part of affinity maturation of antibodies and, unlike mutations in other genes, are not pathological. The tumours are simply reflecting the stage of maturation of the parental B cell. In addition, some CLL cells have been

described that are similar to unmutated IGHV CLL, but originate from B cells with limited somatic mutation, such as CLL with immunoglobulin heavy chains encoded by mutated IGHV3–21 and immunoglobulin light chains encoded by unmutated IGLV3–21 (REFS 3,4). The repertoire of immunoglobulin molecules produced by the CLL cells of all patients is considerably more limited than the repertoire of immunoglobulin molecules that can be made by the B cells of any one person^{5,6}, reflecting the biased use in CLL of certain IGHV genes that have restricted somatic mutation and limited junctional and heavy-light chain combinatorial diversity. In as many as one-third of patients, the CLL cells express immunoglobulin ‘stereotypes,’ which are stretches of primary structure in the variable region that can also be identified in the immunoglobulins produced by the CLL cells of other patients⁷. The restricted immunoglobulin repertoire in CLL is underscored by the finding that ~1 in 75 patients have CLL cells that express immunoglobulin molecules that are virtually identical⁸. The limited immunoglobulin diversity provides compelling evidence that CLL B cells are selected based on the binding activity of their expressed surface immunoglobulin, suggesting that B cell receptor (BCR) signalling plays a crucial part in CLL pathogenesis”

1.0.3 Recurrent genetic features in CLL

Genetic alterations in CLL can include chromosomal alterations, mutations, alterations in the expression of mi RNAs and epigenetic modifications.

(for purpose of TLR sotry) These included del(11q) and *ATM* , del(17p) and *TP53*, *SF3B1*.

REading: From Kipps et al: 38 Dohner H, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. N. Engl. J. Med. 2000; 343:1910–1916. [PubMed: 11136261]

Del(13q)mechanism: 39. Klein U, et al. The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. Cancer Cell. 2010; 17:28–40. [PubMed: 20060366]

Del(17p) mechanism 40. Zenz T, Mertens D, Kuppers R, Dohner H, Stilgenbauer S. From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nat. Rev. Cancer.* 2010; 10:37–50. [PubMed: 19956173]

del(11q)

Mutations and outcome: 38: Dohner H, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N. Engl. J. Med.* 2000; 343:1910–1916. [PubMed: 11136261]

JIC paper, explaon PACE Improved in recent years: 41: Van Dyke DL, et al. The Dohner fluorescence in situ hybridization prognostic classification of chronic lymphocytic leukaemia (CLL): the CLL Research Consortium experience. *Br. J. Haematol.* 2016; 173:105–113. [PubMed: 26848054]

CLL abounds a high degree of genetic diversity: Fabbri G, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J. Exp. Med.* 2011; 208:1389–1401. [PubMed: 21670202] 43. Pleasance ED, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature.* 2010; 463:191–196. [PubMed: 20016485] 44. Puente XS, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature.* 2011; 475:101–105. [PubMed: 21642962] 45. Wang L, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* 2011; 365:2497–2506. [PubMed: 22150006]

Functional role of mutations: WNT: 47 Wang L, et al. Somatic mutation as a mechanism of Wnt/beta-catenin pathway activation in CLL. *Blood.* 2014; 124:1089–1098. [PubMed: 24778153]

POT1: reference thereview

SF3B1: 45 Wang L, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* 2011; 365:2497–2506. [PubMed: 22150006] 48 Ferreira PG, et al. Transcriptome characterization by RNA sequencing identifies a major molec-

ular and clinical subdivision in chronic lymphocytic leukemia. *Genome Res.* 2014; 24:212–226. [PubMed: 24265505]

49. Quesada V, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* 2012; 44:47–52.
50. Te Raa GD, et al. The impact of SF3B1 mutations in CLL on the DNA-damage response. *Leukemia.* 2015; 29:1133–1142. [PubMed: 25371178]

Seqencing studies: two seminal studies: 51. Landau DA, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature.* 2015; 526:525–530. [PubMed: 26466571] 52. Puente XS, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature.* 2015; 526:519–524. [PubMed: 26200345] References 51 and 52 describe landmark studies on whole-exome sequencing of CLL cells obtained from a large cohort of patients, reporting new driver mutations in CLL.

Epigenetic changes: Global hypomethylation and local hypermethylation 66. Cahill N, et al. 450K–array analysis of chronic lymphocytic leukemia cells reveals global DNA methylation to be relatively stable over time and similar in resting and proliferative compartments. *Leukemia.* 2013; 27:150–158. [PubMed: 22922567] 67. Wahlfors J, et al. Genomic hypomethylation in human chronic lymphocytic leukemia. *Blood.* 1992; 80:2074–2080. [PubMed: 1382719] 68. Ziller MJ, et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature.* 2013; 500:477–481. [PubMed: 23925113]

Methylation for classification: 69. Kulis M, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.* 2012; 44:1236–1242. [PubMed: 23064414] This paper describes an analysis of DNA methylation in CLL cases, showing that the two molecular subtypes of CLL have differing DNA methylomes that seem to represent epigenetic imprints from distinct normal B cell subpopulations.

74. Bhoi S, et al. Prognostic impact of epigenetic classification in chronic lympho-

cytic leukemia: the case of subset #2. *Epigenetics*. 2016; 11:449–455. [PubMed: 27128508]

1.5 Taken from Kipps et al 2017: “Genes that are mutated in chronic lymphocytic leukaemia (CLL) are involved in several cellular pathways (blue boxes). As such, mutations in these genes can lead to a range of cellular consequences, such as aberrant DNA repair and B cell receptor (BCR) signalling, among others^{51,213}. The minus sign from GBN1 to the MAPK–ERK pathway indicates negative regulation. *For more detail of the BCR and its associated signalling, see FIG. 3. ASXL1, additional sex comb-like protein 1; ATM, ataxia telangiectasia mutated; BAZ2A, bromodomain adjacent to zinc-finger domain protein 2A; BCOR, BCL-6 co-repressor; BIRC3, baculoviral IAP repeat-containing protein 3; BRCC3, BRCA1/BRCA2-containing complex subunit 3; C-NOTCH, carboxy-terminal domain of NOTCH; CARD11, caspase recruitment domain-containing protein 11; CHD2, chromodomainhelicase-DNA-binding protein 2; CHK2, checkpoint kinase 2; Co-A, co-activator; CSL, CBF1–Suppressor of Hairless–LAG1 (also known as RBPJ); DDX3X, ATP-dependent RNA helicase DDX3X; DYRK1A, dual-specificity tyrosine-phosphorylation-regulated kinase 1A; EGR2, early growth response 2; ELF4, ETS-related transcription factor Elf-4; ERK, extracellular signal-regulated kinase; EWSR1, Ewing sarcoma breakpoint region 1 protein; FBXW7, F-box/WD repeat-containing protein 7; FUBP1, far upstream element-binding protein 1; GNB1, guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit 1; H3K4, histone H3 lysine 4; IC, intracellular domain; IKZF3, Ikaros family zinc-finger protein 3; IL-1R, IL-1 receptor; IRF4, interferon regulatory factor 4; ITPKB, inositol-trisphosphate 3-kinase B; LRP, low-density lipoprotein receptor-related protein; MAP2K1, dual-specificity mitogen-activated protein kinase kinase 1; MAPK, mitogen-activated protein kinase; MED12, Mediator of RNA polymerase II transcription subunit 12; MGA, MAX gene-associated protein; MYD88, myeloid differentiation primary response protein MyD88; NF-B, nuclear factor-B; NXF1, nuclear RNA export factor 1; P, phosphate; POT1, protection of telomeres protein 1; PTPN11, tyrosine-protein phosphatase non-receptor type 11; RIPK1, receptor-interacting serine/threonine-protein kinase 1; RPS15, 40S ribosomal protein

S15; SAMHD1, SAM domain and HD domain-containing protein 1; SF3B1, splicing factor 3B subunit 1; SHP1, Src homology region 2 domain-containing phosphatase 1 (also known as PTPN6); SYK, spleen tyrosine kinase; TCF/LEF, T cell factor/lymphoid enhancer factor; TLR8, Toll-like receptor 8; TNFR1, tumour necrosis factor receptor 1 (also known as TNFRSF1A); TRAF, TNFR-associated factor; XPO, exportin; ZMYM3, zinc-finger MYM-type protein 3. Adapted with permission from REF. 51, Macmillan Publishers Limited.” _Figure from (Kipps2017_?) mutations (Kipps et al. 2017)

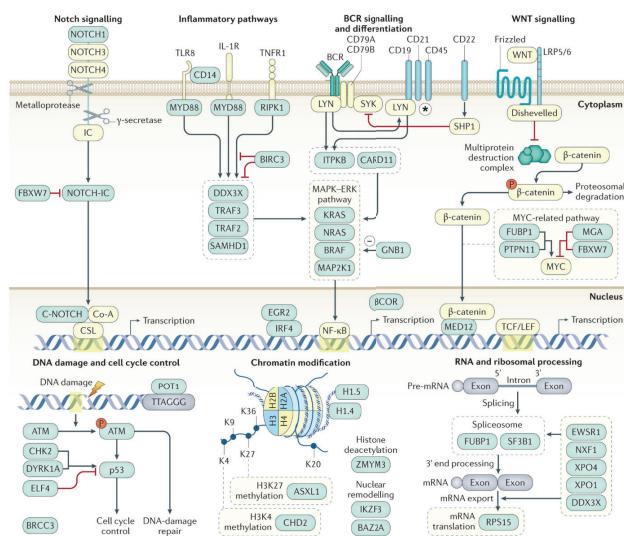


Figure 1.5: (ref:CLLmutations)

Several large genetic studies have revealed numerous genetic alterations in CLL, including single- nucleotide polymorphisms (SNPs), chromosomal alterations and alterations in non- coding RNA, such as microRNA (miRNA), some of which can be used to determine prognosis and to guide management strategies. Interactions between CLL cells and their microenvironment, including interactions with other cell types, such as T cells, nurse-like cells and stromal cells, can induce B cell proliferation and contribute to disease.

(Kipps et al. 2017) The altered expression of genes that are located in or near CLL-associated SNPs might contribute to disease development. For example, a SNP in IRF4 is associated with low expression of interferon regulatory factor 4; mice that are de-

ficient in this protein can develop CLL23, partly owing to hyperactivation of Notch signalling²⁴.

23: Shukla V, Ma S, Hardy RR, Joshi SS, Lu R. A role for IRF4 in the development of CLL. *Blood*. 2013; 122:2848–2855. [PubMed: 23926303] 24. Shukla V, Shukla A, Joshi SS, Lu R. Interferon regulatory factor 4 attenuates notch signaling to suppress the development of chronic lymphocytic leukemia. *Oncotarget*. 2016; 7:41081–41094. [PubMed: 27232759]

(Kipps et al. 2017) Somatic mutations—The advent of massively parallel sequencing and the application of whole-exome sequencing to CLL have transformed our understanding of the genetic heterogeneity of CLL and have established that CLL harbours a high degree of genetic variability^{42–45} (FIG. 2). From these studies, recurrent somatic mutations have been consistently observed in genes that have a role in DNA damage (for example, TP53 and ATM), mRNA processing (for example, SF3B1 and XPO1), chromatin modification (for example, HIST1H1E, CHD2 and ZMYM3), WNT signalling, Notch signalling (for example, NOTCH1) and inflammatory pathways (for example, MYD88). Other mutations, such as those found in EGR2 or BRAF, can affect B cell-related signalling and transcription⁴⁶ (FIG. 2). The functional role of several putative driver mutations has been confirmed; for example, silencing mutated WNT pathway genes in primary CLL cells resulted in decreased cell viability⁴⁷. Mutations in POT1, which has a role in the protection of telomeres, prevented the binding of protection of telomeres protein 1 to telomeric DNA, resulting in numerous chromosomal abnormalities, in addition to the development of abnormal telomeres. Mutations in SF3B1 have been found to be associated with aberrant RNA splicing^{45,48,49} and an altered DNA-damage response⁵⁰. SAMHD1 encodes a protein involved in the regulation of intracellular deoxy-nucleotide pools, which are recruited to the site of DNA damage and are probably involved in the response to DNA double-strand breaks⁵⁰. The detection of somatic mutations and their relative frequencies is variable, which possibly reflects differences in the composition of the cohorts studied worldwide. Two seminal studies have provided the largest sequenced

collections to date^{51,52}, in which >500 CLL samples were characterized by whole-exome sequencing or whole- genome sequencing. The clinical and/or biological features of the patients examined in each study were notably distinct; one study analysed matched pre-treatment samples from patients who required initial treatment and noted mutations in SF3B1 (21% of patients), ATM (15% of patients), TP53 (7% of patients), NOTCH1 (6% of patients) or BIRC3 (4% of patients). The other study assessed patients with early stage CLL and patients with monoclonal B cell lymphocytosis and identified NOTCH1 (12.6% of patients), ATM (11% of patients), BIRC3 (8.8% of patients) and SF3B1 (8.6% of patients) as the most frequently mutated genes. These large sample cohorts have provided the sensitivity to discover novel candidate cancer genes that are altered in CLL. Both studies also identified somatic mutations in MGA and PTPN11, which encode modulators of MYC, IKZF3, which encodes a key transcription factor, and RPS15, which encodes 40S ribosomal protein S15 and is recurrently mutated in ~20% of patients who relapse after combination chemotherapy⁵³. Other recurrent somatic mutations include those in the 3 untranslated region of NOTCH1 and a PAX5 enhancer, which increases the expression of these B cell-associated transcription factors^{54,55}. Patients with mutations in the 3 untranslated region of NOTCH1 have a shorter time from diagnosis to treatment and poorer overall survival, similar to that of patients with non- synonymous mutations, which alter the amino acid sequence of NOTCH1. Next-generation sequencing has revealed intra-tumoural heterogeneity in CLL. Some somatic mutations, such as those in MYD88, or chromosomal abnormalities, such trisomy 12 or del(13q), are most often found in all the CLL cells of any one patient, indicating that these genetic alterations occurred early in the evolution of the leukaemia. Other mutations, such as those found in SF3B1 or NOTCH1, or chromosomal alterations, such as del(17p), are typically found in only a fraction of the leukaemia cells and thus represent subclonal events, which occur after the development of CLL. Across studies, subclonal driver mutations are associated with more-aggressive disease, particularly when two or more are found concurrent in a leukaemia cell population^{51,56,57}. In addition, studies have demonstrated that large clonal shifts can occur following chemotherapy, owing to increases in the proportions of

CLL cells that have a TP53 mutation or del(17p), indicating that such genetic changes provide a strong fitness advantage in the setting of therapy⁵¹. By contrast, one study of CLL cells from patients treated with ibrutinib (an inhibitor of Bruton tyrosine kinase (BTK)), revealed mutations associated with drug resistance that were distinct from those observed in CLL cells of patients treated with standard chemotherapy⁵⁸. CLL was the first human disease that was found to be associated with alterations in miRNA.

(can compare size of large studies with ours)

copy number variants (Kipps et al. 2017) Approximately 80% of patients with CLL carry at least one of four common chromosomal alterations: a deletion in chromosome 13q14.3 (del(13q)), del(11q), del(17p) and trisomy 12 (REF. 38). Del(13q) is the most common chromosomal alteration, evident in >50% of patients, and is associated with favourable prognosis. Within this deleted region is the DLEU2-mir-15–16 cluster, which regulates the expression of proteins that can inhibit apoptosis or that are involved in cell cycle progression³⁹ (see Supplementary information S2 (table)). Del(17p) is found in 7% of patients and is associated with loss of the tumour suppressor gene TP53 (REF. 40), whereas del(11q) is found in 18% of patients and is often associated with alterations in ATM (which encodes a protein involved in DNA repair); each of these chromosomal alterations is associated with adverse clinical outcome³⁸, although this has improved in recent years⁴¹. Trisomy 12 is found in 16% of patients with CLL and is associated with an intermediate prognosis. Unlike the neoplastic B cells in mantle cell lymphoma, CLL cells do not have the translocation t(11;14) (q13;q32) or other genetic alterations that enhance the expression of CCND1, which encodes the cell cycle regulator cyclin D1

Epigenetic changes The CLL epigenome shows global hypomethylation combined with local hypermethylation, as has been observed in other cancers^{66–68}. Indeed, comprehensive methylation profiling has demonstrated substantial intra-tumoural methylation heterogeneity^{52,69–73}. Increasing methylation heterogeneity has consistently been associated with increased genetic complexity owing to the acquisition of subclonal mutations,

thus linking genomic and methylomic evolution in CLL^{52,70,71}. Indeed, locally disordered methylation in CLL might enhance the evolutionary adaptive capacity of CLL cells by increasing the background ‘noise’ of the genome, thereby providing increased opportunities for somatic mutations within the leukaemia clone. In support of this notion is the observed association between methylation evolution and adverse clinical outcome^{52,70,71}.

(Kipps et al. 2017) Methylation signatures can classify distinct clinical CLL subgroups^{69,74}. As these methylation patterns are a heritable trait, they have been used to ‘trace’ back to the type of normal B cell from which the CLL cells were derived⁷⁵. These studies revealed that the CLL cells of different patients derive from a continuum of B cell maturation states, which are not restricted to discrete maturation stages. Nevertheless, CLL cells that use unmutated IGHV versus mutated IGHV generally have distinctive methylation patterns, which respectively approximate to those of pre-germinal centre versus post-germinal centre memory B cells, as depicted in FIG. 1. The diversity in the likely cells of origin of CLL cells highlights the biological and phenotypic heterogeneity of this disease. These findings also suggest that epigenetic programming that is dependent of transcription factors has a potentially important role in the development of CLL.

1.0.4 The incompletely understood role of trisomy12

Bosch, F., Dalla-Favera, R. Chronic lymphocytic leukaemia: from genetics to treatment. Nat Rev Clin Oncol 16, 684–701 (2019). <https://doi.org/10.1038/s41571-019-0239-8>

Trisomy 12 chronic lymphocytic leukemia expresses a unique set of activated and targetable pathways DOI: <https://doi.org/10.3324/haematol.2018.190132> Issue: Vol. 103 No. 12 (2018): December, 2018

The Protein Landscape of Chronic Lymphocytic Leukemia (CLL) Fabienne Meier-Abt <https://doi.org/10.1182/blood.2020009741>

what’s understood about the overexpression of genes and proliferative drive; DOI: 10.1200/JCO.2005.02.568

Journal of Clinical Oncology 23, no. 16 (June 01, 2005) 3780-3792.

1.0.5 Stratification of CLL patients

Reading: (from Kipps)

12. Nabhan C, et al. The impact of race, ethnicity, age and sex on clinical outcome in chronic lymphocytic leukemia: a comprehensive Surveillance, Epidemiology, and End Results analysis in the modern era. *Leuk. Lymphoma.* 2014; 55:2778–2784. [PubMed: 24588735]
 13. Li Y, Wang Y, Wang Z, Yi D, Ma S. Racial differences in three major NHL subtypes: descriptive epidemiology. *Cancer Epidemiol.* 2015; 39:8–13. [PubMed: 25560974]
 14. Pulte D, Redaniel MT, Bird J, Jeffreys M. Survival for patients with chronic leukemias in the US and Britain: age-related disparities and changes in the early 21st century. *Eur. J. Haematol.* 2015; 94:540–545. [PubMed: 25315799]
- 1.6 Taken from Kipps et al 2017: “Indications for therapy of patients with chronic lymphocytic leukaemia (CLL) include late-stage disease, evidence for rapid disease progression or disease-related symptoms. Patients with del(17p) or mutated TP53 should be treated with therapy that does not require functional TP53, such as ibrutinib (a Bruton tyrosine kinase (BTK) inhibitor), given the relatively poor outcome for such patients with chemotherapy. For patients without del(17p) or known mutations in TP53, immunoglobulin heavy-chain variable region (IGHV) mutational status can help to define the treatment strategy; patients with unmutated IGHV could be considered for therapy with a BTK inhibitor (such as ibrutinib) and patients with mutated IGHV might be good candidates for chemoimmunotherapy (CIT), if amenable. Indeed, patients with mutated IGHV can have excellent outcomes with CIT regimens, such as fludarabine, cyclophosphamide and rituximab, with >50% of patients having a median progression-free survival of >10 years, including the potential for cure. If the patient is amenable to CIT, age, medical comorbidities and myeloid reserve should be taken into consid-

eration. Patients >65 years of age commonly have medical comorbidities and are less able to tolerate myelosuppressive regimens, such as fludarabine, cyclophosphamide and rituximab. Thus, considerations should be given to using reduced dose or less myelosuppressive chemotherapy regimens, such as chlorambucil or reduced-dose bendamustine and an anti-CD20 monoclonal antibody for patients with limited myeloid reserve. Patients who either do not respond, have a poor tolerance to CIT or relapse following CIT, should be re-evaluated for del(17p) or TP53 mutations. Patients who develop de novo del(17p) or TP53 mutations, or have known del(17p) and/or TP53 mutations, or who develop resistance or intolerance to ibrutinib, could be considered for therapy with idelalisib and rituximab or the BCL-2 inhibitor venetoclax. Patients treated with CIT who do not have del(17p) or TP53 mutations could be considered for repeat CIT if their progression-free survival after CIT is >2 years and the patient has sufficient myeloid reserve. Such patients also might be treated with a BTK inhibitor or a phosphoinositide 3-kinase (PI3K) inhibitor, which also could be considered for patients who develop intolerance or resistance to therapy with ibrutinib. Patients who develop resistance or intolerance to inhibitors of BTK, PI3K and/or BCL-2 should be considered for clinical trials or alternative agents. LDT, lymphocyte doubling time.” *Figure from Kipps et al. (2017)*. Disease staging (Kipps et al. 2017) Two clinical staging systems are widely used to divide patients with CLL into three broad prognostic groups^{114,115}. The Rai staging system (TABLE 1) is more commonly used in the United States, whereas the Binet classification (TABLE 2) is more commonly used in Europe. The staging systems each recognize the importance of marrow function and define late-stage, or high-risk, disease by the presence of pronounced anaemia or thrombocytopenia.

(Kipps et al. 2017) Prognostic factors that can help to identify patients who require therapy relatively soon after diagnosis include certain clinical features and genetic, molecular and biochemical characteristics of the CLL cell. Multivariable models, prognostic indexes^{116–118} and nomograms¹¹⁹ have been developed to consolidate such prognostic factors so that they can more robustly predict clinical outcome. Commonly used parameters that are associated with poorer outcome are male sex, 65 years of age, poor

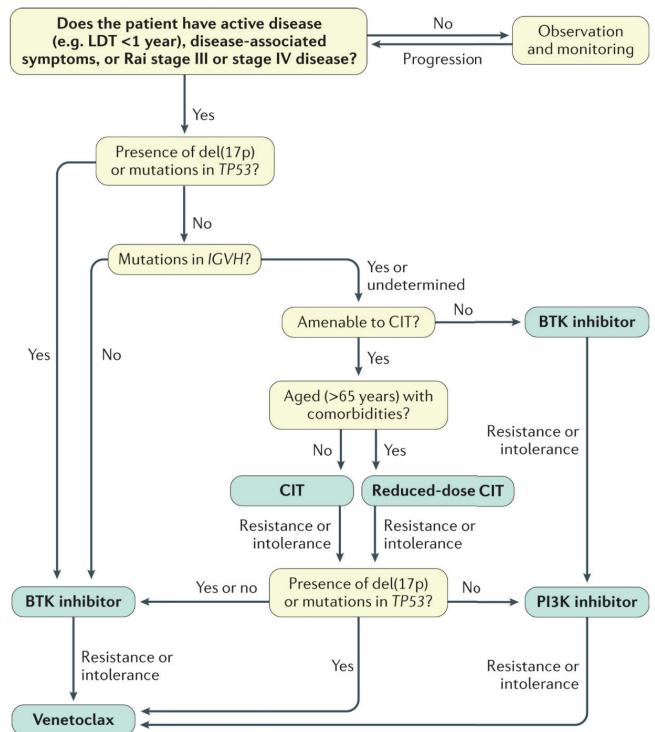


Figure 1.6: (ref:CLLstratification)

performance status due to medical comorbidities, certain CLL cell characteristics, such as the expression of unmutated IGHV1,2, ZAP70 (REFS 120,121), CD49d (also known as integrin 4)122 or CD38 (REF.2), the presence of del(17p)38 or del(11q)123, high serum levels of 2-microglobulin (>3.5 mg per l)124, complex karyotype (that is, the presence of three or more chromosomal aberrations observed on a karyotype test)125,126, or a high absolute lymphocyte count ($>50,000$ cells per μl) and/or late-stage disease at initial presentation. Del(17p) is often associated with inactivating mutations in TP53 and is a predictor of poor outcome to treatment with regimens that involve conventional chemotherapy¹²⁷.

(Kipps et al. 2017) Generally, indications to initiate therapy include pronounced disease-related anaemia or thrombocytopenia (patients with Rai stage III or stage IV disease, or Binet stage C disease), symptomatic lymphadenopathy and/or symptoms that are

associated with active disease, such as night sweats, fatigue, unintentional weight loss or fever without evidence of infection¹⁰⁷. However, when basing a treatment decision on constitutional symptoms alone, the physician should consider other medical conditions, such as hypothyroidism, hyperthyroidism, hypoglycaemia, chronic inflammation, uncommon opportunistic infections or sleep disorders, including sleep apnoea.

For patients who need treatment, the presence of del(17p) or mutated TP53 are the most important features that are currently directing the choice of therapy (FIG. 8). Next, advanced age of >65 years, the presence of medical comorbidities and the objectives of treatment have substantial bearing on the choice of therapy. Increasingly, IGHV mutational status is considered as a parameter when determining the type of therapy; for example, chemotherapy-based regimens are reserved for patients with CLL and mutated IGHV. Conversely, the specific Rai or Binet stage of the patient who requires treatment does not necessarily influence the choice of therapy. Systemic

1.0.6 Prognosis and where the problems lie

harsh treatments, especially when most are elderly not a proper cure - combo therapies may cover all bases heterogenous - combo therapies may cover all bases The clinical course of newly diagnosed CLL is extremely variable; some patients remain free of symptoms and are fully active for decades, whereas others rapidly become symptomatic or develop high-risk disease, which requires treatment soon after diagnosis and might result in death due to therapy-related and/or disease-related complications. However, most patients have a clinical course that is in between these two extremes.

(Kipps et al. 2017) Adverse effects of ibrutinib include fatigue, diarrhoea, bleeding, ecchymoses, rash, arthralgia, myalgia, increased blood pressure and atrial fibrillation. Clinical trials are currently evaluating second-generation BTK inhibitors (for example, acalabrutinib¹⁴⁸, ONO/ GS-4059 (REF. 149) or BGB-3111) to determine whether any one of these drugs has a superior therapeutic index than that of ibrutinib¹⁵⁰.

(Kipps et al. 2017) Toxicities of venetoclax include gastrointestinal disturbances, neu-

tropenia and tumour lysis syndrome¹⁵⁹, which is characterized by hyperkalaemia, hyperuricaemia and/or azotaemia. Tumour lysis syndrome results from the rapid destruction of cancer cells and the release of their cellular contents into the blood. Tumour lysis syndrome typically occurs when initiating venetoclax therapy or when dosing is increased. Thus, patients start venetoclax with a low daily dose, which is escalated each week over 5 weeks to mitigate the risk of developing tumour lysis syndrome. Even with this strategy, patients who are at high risk for tumour lysis syndrome because of bulky lymphadenopathy and/or lymphocytosis of >25,000 cells per μ l must be hydrated and closely monitored during therapy initiation and during dose escalation.

(Kipps et al. 2017) The outlook for patients with CLL has improved substantially over the past several years. Through research on the immune biology and genetics of CLL, patients can be stratified into subgroups with distinctive clinical features, which has improved our capacity to assess prognosis or govern therapy. However, an understanding of the mechanisms that contribute to immune dysfunction or how it contributes to autoimmune disease, such as autoimmune haemolytic anaemia, therapy resistance or therapy-related complications is unknown. Whether tyrosine kinase inhibitors can affect clonal evolution, induce and/or select for drug resistance, or can achieve complete responses if used earlier in the course of the disease is also unknown.

New therapies:

(Kipps et al. 2017) Several therapies are currently under preclinical and clinical investigation for the treatment of patients with CLL, including new drugs and treatment modalities that can modulate the immune system, and cell transplantation.

Car-T Therapies (Kipps et al. 2017) CD19-targeted CAR T cells have yielded long-term PFS and relapse-free survival durations in patients with CLL; in 14 patients with relapsed or refractory CLL, four patients achieved a complete response and four patients achieved a partial response¹⁹⁷. None of the complete responders had MRD and none relapsed (a median follow-up of 19 months). However, the efficacy of CAR T cell therapy in patients with CLL has been modest compared with that in patients with acute lymphoblastic

leukaemia; this might be owing to qualitative defects in the T cells of patients with CLL, who are generally older than patients with acute lymphoblastic leukaemia and already have immune dysfunction that reflects disease-associated anergy (see BCR and B cell signalling). Ibrutinib might partially correct some of these defects¹⁹⁸. Larger

Allogeneic stem cell transplantation (Kipps et al. 2017) Allogeneic stem cell transplantation is a potentially curative strategy for patients with relapsed or refractory CLL, including patients with high-risk features such as del(17p). Research efforts are ongoing to develop better-tolerated cell-based therapy with a similar curative potential that can be used without the immunosuppression and associated long-term morbidity and mortality of allogeneic stem cell transplantation. Donor availability, advanced patient age, associated toxicities of myelosuppression, graft- versus-host disease and impaired resistance to infections limit the application of allogeneic stem cell transplantation in patients with CLL. In addition, the advent of BCR signalling inhibitors and BCL-2 inhibitors provide multiple treatment options that afford well-tolerated, long-term disease control, making allogeneic stem cell transplantation the least desirable option for most patients.

Immune-modulatory drugs (Kipps et al. 2017) Immune-modulatory drugs, such as thalidomide and lenalidomide, are approved for the treatment of patients with multiple myeloma, mantle cell lymphoma or myelodysplastic disease. Although these drugs have clinical activity in patients with CLL, they have had limited application unless used in combination with an anti-CD20 monoclonal antibody^{166,186}. In CLL, lenalidomide can induce the expression of p21WAF1/CIP1, which inhibits cyclin-dependent kinase and CLL cell proliferation¹⁸⁷, and can improve immune synapse formation, potentially enhancing immune function.

Immune Checkpoint inhibitors (Kipps et al. 2017) The immune checkpoint receptor PD-1, and its ligands PD-L1 and PD-L2, is the most important cognate receptor involved in the suppression of cellular immune activation. CLL cells express high levels of PD-L1 and PD-L2 and can suppress the responses of PD-1-expressing effector T cells¹⁰⁶,

leading to an exhausted (that is, no longer functional) T cell phenotype. Preclinical studies in mouse models have demonstrated that checkpoint inhibitors can reactivate immune effector cells to have anti-leukaemia activity²⁰². However,

Combination targeted therapies (Kipps et al. 2017) We can now target the distinctive phenotypic or physiological features of CLL with targeted therapeutic agents, which have a higher therapeutic index than standard chemotherapy. Through the use of combination therapy, which targets different B cell survival signalling pathways and/or achieves better eradication of CLL cells, we might be able to define curative treatments for most patients with this disease. Research on leukaemia cell survival signalling pathways, such as those governed by interactions between leukaemia cells and cells or secreted factors within the microenvironment (FIG. 4), might identify pathways that are not affected by BCR inhibitors. For example, BCR inhibitors, such as ibrutinib, cannot block ROR1-dependent WNT5A signalling, which enhances CLL cell proliferation, migration and survival⁹³; as such, antibodies that block ROR1-dependent signalling could potentially have synergistic activity when used in combination with BCR inhibitors²⁰³. Furthermore, the interaction between CLL cells and accessory cells in the microenvironment might enhance CLL cell expression of anti-apoptotic proteins other than BCL-2, such as MCL1, thereby contributing to therapy resistance. As such, the therapeutic use of a selective BCL-2 antagonist, such as venetoclax, might be more effective when used in combination with BCL-2 inhibitors^{161,162}, which also interfere with the homing of CLL cells to the microenvironment. Conceivably, combination target therapy with agents that have synergistic activity will provide highly effective and potentially curative treatment of patients with CLL.

1.1 The tumour microenvironment in CLL

1.1.1 The concept of the protective niche

1.7 Adapted from Davids et al 2012 Lymph nodes etc

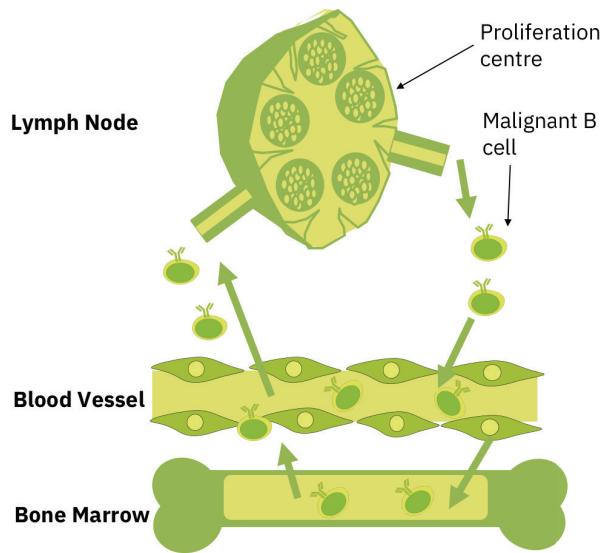


Figure 1.7: (ref:microenvironmentTissues)

(Kipps et al. 2017) CLL cells depend on survival signals that they receive in lymphoid tissues from neighbouring non-neoplastic cells within the so-called cancer microenvironment. CLL cells follow chemokine gradients into lymph nodes, where they form ‘proliferation centres’ (REF. 77), as opposed to normal germinal centres. In these proliferation centres, the CLL cells contact non- malignant stromal cells, nurselike cells (also known as lymphoma-associated macrophages), T cells and mesenchymal-derived stromal cells (FIG. 4). Engagement with autoantigen may occur during this transit, thereby stimulating CLL cell activation and proliferation if sufficient T cell help is available. Only a few per cent of the CLL cells undergo proliferation at any one time; the remainder of the cells are either unstimulated or driven into anergy⁸⁴. However, within such proliferation centres, all CLL cells are exposed to chemokines, integrins, cytokines and survival factors.

1.1.2 Microenvironmental pathways and cross-talk

pathways esp IL4 and TLR

The microenvironment in CLL constitutes cross talk via soluble factors and cell-cell

contacts, in the blood, bone marrow and lymph nodes.

?? Adapted from: Weistner et al 2015 1.9 Taken from Kipps et al 2017; “Migration of

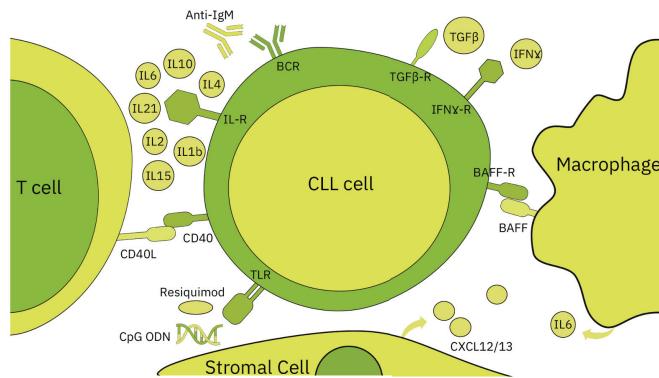


Figure 1.8: (ref:microenvironmentOverview)

chronic lymphocytic leukaemia (CLL) cells into the lymphoid tissue is primarily mediated through CXC-chemokine receptor 4 (CXCR4) in response to CXC-chemokine ligand 12 (CXCL12), which is secreted mainly by nurse-like cells (NLCs) and mesenchymal-derived stromal cells. Migration of CLL cells into lymph nodes also occurs via CC-chemokine receptor 7 (CCR7) in response to CC-chemokine ligand 19 (CCL19) and CCL21, which are produced by the endothelial cells of high endothelial venules (HEVs). HEV endothelial cells also express hyaluronan, which can interact with CD44, to facilitate B cell signalling and might enhance the production of active matrix metalloproteinase 9 (MMP9). Once in tissues, several chemokines promote B cell survival, including CXCL12, B cell-activating factor (BAFF; also known as TNFSF13B) and a proliferation-inducing ligand (APRIL; also known as TNFSF13). In addition, CLL cell survival can be promoted through cognate interactions between CD31 and CD38, and the production by stromal cells of WNT factors, which can interact with ROR1, ROR2 and/or various Frizzled receptors. CLL cell contact with mesenchymal stromal cells can also be established through vascular cell adhesion protein 1 (VCAM1)-4 1 integrin interactions that contribute to CLL cell survival. In turn, CLL cells can secrete chemokines, such as CCL3 and CCL4, which can recruit T cells and NLC-precursor cells (monocytes) to the CLL microen-

vironment. Activated T cells can provide CLL cells with proliferative signals through CD40 ligand (CD40L)-CD40 interactions and the secretion of several cytokines, such as IL-2, IL-4 and IL-10. In turn, activated CLL cells secrete CCL12 and CCL22, which attract more T cells into the CLL microenvironment. In tissues, CLL cells can be exposed to environmental and/or self-antigens that might trigger B cell activation through interactions with the surface immunoglobulin; this could amplify the responsiveness of CLL cells to the signals and factors that are provided by the CLL microenvironment. BAFFR, BAFF receptor (also known as TNFRSF13C); BCMA, B cell maturation protein (also known as TNFRSF17); BCR, B cell receptor; TACI, transmembrane activator and CAML interactor (also known as TNFRSF13B). *Figure from Kipps et al. (2017) .*

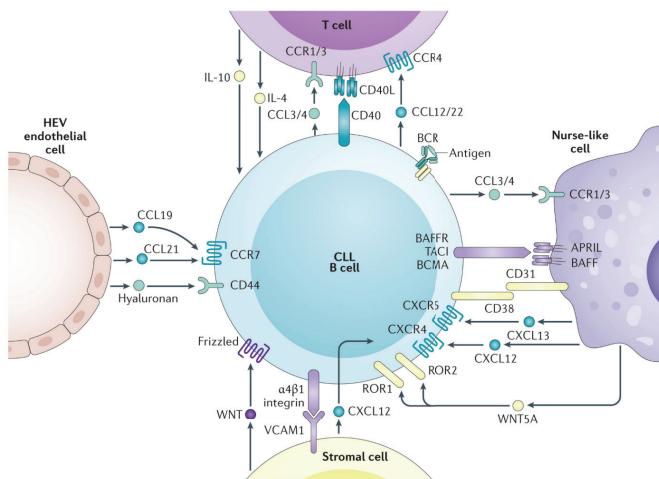


Figure 1.9: (ref:microenvironmentSignalling)

1.2 Therapies in CLL

1.2.1 Clinical regimes

(Kipps et al. 2017) The treatment of patients with CLL can include chemotherapy, a combination of chemotherapy and immunotherapy, or drugs that target the signalling pathways that promote the growth and/or survival of CLL cells (for example, BCR

signalling and BCL-2)128,129.

Chemotherapy—Chemotherapy has been the mainstay of therapy for the past 50 years. Purine analogues (most commonly fludarabine, but also pentostatin or cladribine) and alkylating agents (including chlorambucil, cyclophosphamide or bendamustine) are used in the treatment of CLL^{130–132}. Chemotherapy-based regimens can cause myelosuppression, an increased risk of infections and, in a small subset of patients, post-therapy myelodysplasia or secondary cancers, such as acute myeloid leukaemia (see Secondary cancers).

Chemoimmunotherapy—Phase III clinical trials have validated the benefit of anti-CD20 monoclonal antibodies, such as rituximab, obinutuzumab or ofatumumab, in combination with chemotherapy for the treatment of patients with CLL. In one trial (the CLL8 trial of the German CLL Study Group), patients who received fludarabine and cyclophosphamide with rituximab had higher response rates and a longer median progression-free survival (PFS) than patients who were treated with fludarabine and cyclophosphamide¹³³. In a separate study (the CLL11 trial), patients >65 years of age with medical comorbidities who were treated with chlorambucil and either obinutuzumab or rituximab had improved response rates and longer median PFS than patients who were treated with chlorambucil alone¹³⁴. However, the median PFS was significantly longer for patients who received obinutuzumab (26.7 months) than in those who received rituximab (11.1 months). In a third phase III trial, median PFS significantly improved from 13.1 months for patients treated with chlorambucil to only 22.4 months for patients treated with chlorambucil and ofatumumab¹³⁵. As a consequence of these three trials, the US FDA approved the use of rituximab, obinutuzumab or ofatumumab in combination with chemotherapy for the first-line treatment of patients with CLL. The FDA also approved the use of ofatumumab as a single agent for the treatment of patients with relapsed or refractory disease based on data from a phase II study¹³⁶.

1.2.2 New Targeted therapies

in particular Ibrutinib and idelalisib

(Kipps et al. 2017) Three main classes of drugs that each can inhibit BCR signalling have been evaluated in patients with CLL: BTK inhibitors, PI3K inhibitors and spleen tyrosine kinase (SYK) inhibitors^{86,143}. CLL cells with unmutated IGHV seem to be more sensitive to inhibitors of BCR signalling than CLL cells with mutated IGHV144, but whether inhibitors, such as ibrutinib, are more effective in patients with CLL and unmutated IGHV, remains to be validated in clinical trials. Ibrutinib has been approved in the United States and Europe for use as initial therapy, as well as in patients with relapsed disease, which followed results from a randomized trial that showed a significantly higher response rate to therapy with ibrutinib than with ofatumumab¹⁴⁵. In addition, with continuous therapy, patients treated with ibrutinib had a significantly longer median PFS and overall survival than patients treated for 8 months with ofatumumab. Approval of ibrutinib as initial therapy was based on the results of a randomized trial that showed a significant improvement in median PFS and overall survival in patients 65 years of age without del(17p) who were treated indefinitely with ibrutinib than in patients treated for up to 48 weeks with chlorambucil¹⁴⁶. Upon initiation of treatment with ibrutinib, lymphadenopathy is rapidly reduced, which is associated with a concomitant increase in absolute lymphocyte count¹⁴⁷. The rise in absolute lymphocyte count is related to the inhibition of chemokine receptor signalling, which inhibits the migration of CLL cells from the blood into the lymphoid tissues. This resulting lymphocytosis should not be considered a sign of progression; over time, the lymphocytosis subsides as the overall tumour burden decreases with continued therapy.

(Kipps et al. 2017) PI3K inhibitors include idelalisib, duvelisib (also known as IPI-145), TGR-1022 and ACP-319 (also known as AMG-319)¹⁵¹; the latter three drugs are being evaluated in clinical trials, whereas idelalisib was approved in the United States and Europe for the treatment of patients with relapsed CLL

(Kipps et al. 2017) Phase I/II clinical trials of fostamatinib, an oral SYK inhibitor,

caused reduction in lymphadenopathy with concomitant lymphocytosis, an improvement in disease-related cytopenias and relief of disease-related symptoms in most of the treated patients with CLL¹⁵⁶. However, dose-limiting toxicities of fostamatinib treatment include neutropenia, thrombocytopenia and diarrhoea. Other inhibitors of SYK, such as entospletinib, are being evaluated in preclinical and clinical studies.

(Kipps et al. 2017) BCL-2 inhibitors—Venetoclax is a small molecule that functions as a BH3 mimetic to inhibit BCL-2 (REF. 157). This drug is highly potent in inducing apoptosis in CLL cells, possibly by diminishing the capacity of BCL-2 to sequester the pro-apoptotic molecule BCL-2-interacting mediator of cell death (BIM; also known as BCL2L11)¹⁵⁸. As such, venetoclax is effective in patients with relapsed and/or refractory disease¹⁵⁹ or in patients with relapsed disease and del(17p)¹⁶⁰. Indeed, the overall response rate for patients with relapsed disease and del(17p) was 79%, with 8% achieving a complete response. In addition, the estimated 12-month PFS was 72% and overall survival was 87%. On the basis of these results, the FDA approved the use of venetoclax for patients with relapsed disease and del(17p). Ongoing studies have shown that venetoclax can be safely combined with rituximab or obinutuzumab. Moreover, studies are examining the use of venetoclax with or without an anti-CD20 monoclonal antibody, and with or without ibrutinib^{161,162}, which might provide higher response rates to therapy than that with venetoclax alone.

1.2.3 Drug resistances in CLL

incl ME -mediated and ibrutinib - IL4

1.2.4 BET inhibitors, and their potential for the treatment of leukemia

1.3 Background to the approach used in this thesis

1.3.1 Ex-vivo drug sensitivity screens

1.3.2 Studies of the microenvornment, and attempts to replicate the impact of the microenvironment

1.3.3 ATAC sequncing as a proxy for TF activity justification

1.3.4 Measurement of outcomes?

LDT, TTT, TFT, OS

(Kipps et al. 2017) Currently, the most reliable prognostic models are those developed for treatment-free survival, as evolving treatments have yet to change the indications for therapy. Predictive models to define overall survival with a given type of therapy are challenged by the chronicity of CLL and the fact that patients often receive serial treatments, each of which can affect outcome; moreover, death might be due to an indirect or unrelated cause. Furthermore, treatment options are changing, with newly identified, highly effective agents that are clearly prolonging survival and have activity among patients who would have been considered high risk when the only option was conventional chemotherapy.

(Kipps et al. 2017) Historically, a favourable response to therapy has been defined as a partial remission or complete remission. Partial remission requires a 50% reduction in tumour bulk (for example, lymphadenopathy and splenomegaly), a 50% reduction in lymphocytosis, and platelet counts of >100,000 cells per μl (or 50% improvement over baseline) or a haemoglobin level of >11 g per dl (or 50% improvement over baseline) without requiring transfusions or exogenous growth factors¹⁰⁷. Complete remission requires the normalization of blood counts, resolution in lymphadenopathy and splenomegaly, and normal marrow function. The use of CT to assess response in CLL is becoming more common, particularly in clinical trials. However, the benefit of using repeated CT scans to monitor disease is uncertain, and seems unlikely to change patient outcome. Because of the distinct pattern of response observed with BCR inhibitors, a new response category, namely, partial response with lymphocytosis, has been described. Partial response with lymphocytosis is defined as a >50% reduction in lymphadenopathy and

splenomegaly, with persistent lymphocytosis; often the blood lymphocyte counts are equal to or greater than those observed prior to therapy.

In clinical trials, it is becoming more common to evaluate for MRD with 0.01% of CLL cells among the total population of mononuclear cells in the blood or marrow. MRD can be measured by flow cytometry or PCR with next-generation sequencing of the clonal immunoglobulin variable region gene rearrangements¹⁶³. In most clinical trials for patients with CLL, particularly those conducted in Europe, evaluation of MRD has been performed by flow cytometry of mononuclear cells from the marrow aspirate (the preferred method) or from the peripheral blood. In the 6 months following anti-CD20 monoclonal antibody treatment, the assessment of MRD is more sensitive on the mononuclear cells of the marrow aspirate than on cells that are isolated from the blood, which will often lack detectable CLL even when they are readily found in the marrow. Beyond a complete response, the best predictor of long-term PFS and overall survival is the achievement of a complete remission without evidence for MRD.

1.3.5 Multiomics datasets to study CLL

The power of multi-omics approaches especially the use of ATAC sequencing and combining ATAC and RNA

1.3.6 Mathematical modelling

Mathematical background in particular, lasso-regularised regression, and the caveats associated with feature selection wrt cluster assignment

use method, explain how it works we need mor ltieracy about these things discouvery of mulitple layers of biology rather than just using t tests

how this can be used to build classifiers

1.3.7 CLL as a model for studying tumour biology

1.4 Aims of this thesis

1.5 Overview of the approach

Including diagrams

1.6 Note on contributions

Many of the figures, and their captions are taken directly from the manuscript. Where this is the case, it is indicated in the figure caption.

Chapter 2

Methods

Methods in quotation marks are taken from Bruch & Giles et al. 2021 and I have authored the original text, unless stated otherwise. Paragraphs without quotation marks were rewritten for the thesis.

Routine: all headings should be here at the end of each chapter, make sure all methods are included Copy things in and put in quotation marks if from Bruch & Giles et al. 2021, mark if they were not written by me, or if they were written jointly Add REFERENCE and TABLE for now Return to this later and neaten / expand on methods, and provide all referencing and tables -fine to use quotations (although check plagiarism email), suggested rewriting as i / he / pmb throughout

2.1 Experimental methods:

2.1.1 Drug - Stimulus Combinatorial Perturbation Assay

Patient Sample Preparation

with Peter-Martin Bruch Original Peripheral Blood was taken from 192 patients for the initial drug-stimulation assay. Blood was separated by Ficoll gradient (GE Healthcare) and mononuclear cells were cryopreserved.

2.1.2 Preparation of screening plates

with Peter-Martin Bruch Original “Sample preparation, cell-culture, drug-stimulation profiling and genomic annotation was performed on 192 CLL patient samples as previously described[REFERENCE] with the following adjustments. Stimuli and drugs were mixed and preplated in the culture plates directly before adding the cell suspensions. RPMI-1640 and supplements were acquired from Gibco by Life Technologies, human serum was acquired from PAN Biotech (Cat.No. P40-2701, Lot.No:P-020317). Luminescence was measured after 48h on a Perkin Elmer EnVision.”

with Peter-Martin Bruch Original Compounds were obtained from Selleckchem, MedChemExpress and Sigma-Aldrich, dissolved in DMSO and stored at -20°C. 12 drugs were used in two concentrations [TABLE]. Final DMSO concentration did not exceed 0.3% in all experiments. Carfilzomib, Panobinostat and Venetoclax were removed from the analysis as they showed inconsistent toxicity depending on used media. Insert table

Recombinant cytokines and stimulatory agents were dissolved according to manufacturer’s protocol. 17 stimuli were selected [TABLE]. HS-5 conditioned medium was produced by incubating HS-5 stroma cell line (gifted by Martina Seiffert, DKFZ, Heidelberg) for 4 days at 37°C and 5% CO₂, after which the supernatant was centrifuged and stored at -20°C, the final concentration of HS-5 CM was 20%. Bead immobilised anti-IgM was removed from the analysis due to storage instability. ”

Paper: “Compounds and stimulatory agents were dissolved, stored and diluted according to manufacturer’s protocol. HS-5 conditioned medium was produced by incubating HS-5 stromal cell line to >80% confluence and cell removal by centrifugation. For a detailed list of stimuli and drugs and associated concentrations see Supp. Tables 1 and 2. Final DMSO concentration did not exceed 0.3%.”

Original Compounds were preplated in 384-well polypropylene storage plates (Greiner Bio-One Cat.No.:781271), which were stored at -20°C. For each batch of samples tested on the same day a new storage plate was thawed and diluted in serum free RPMI, with or

without stimuli. 5 μ L of drug-stimulation dilution were added into each well of the assay plates, 20 μ L of cell suspension were added. The final cell concentration was 8*10⁵ cells/ml. Cells were thawed as previously described[REFERENCE].

2.2 Drug-Stimulation assay

with Peter-Martin Bruch Original Drug-stimulation assays were performed with RPMI-1640 (Gibco by Life Technologies) supplemented with Penicillin Streptomycin (Gibco) final concentration of 100 Units/ml and 100 μ g/ml respectively, L-Glutamine (Gibco) final concentration 2mM, and 10% pooled, heat-inactivated and sterile filtered human type AB male off the clot serum (PAN Biotech, Cat.No. P40-2701, Lot.No:P-020317).

Drugs and stimuli were added to the cells simultaneously. Culture was performed in 384-well plates (Greiner Bio-One Cat.No.: 781904) and cells were incubated at 37°C and 5% CO₂ for 48h.

Cell Viability was determined using the ATP-based CellTiter-Glo assay (Promega, Cat.No.:G7573). Luminescence was measured for the drug-stimulation assays using a Perkin Elmer EnVision and for the drug-drug interaction experiments using a Perkin Elmer EnSight, with a measurement time of 100ms per well.

2.3 WES, RNA Sequencing, Targeted Sequencing and DNA Copy number variants

Myself Original Sequencing data generated by Dietrich et al[REFERENCE], and all sequencing and data processing were performed as described there. For

2.3.1 Follow - up investigations

Lymphocyte doubling time

Survival data

ATACsequencing of 4 CLL samples

with Peter-Martin Bruch (mostly edited by me) Paper “Peripheral blood was taken from 4 CLL patients and separated by Ficoll gradient (GE Healthcare), mononuclear cells were cryopreserved on liquid nitrogen. Samples were later thawed from frozen as previously described[REFERENCE] and MACS sorted for CD19 positive cells (Milteny autoMACS). The cells were resuspended in RPMI (GIBCO, Cat.No. 21875-034), with the addition of 2mM glutamine (GIBCO, Cat.No. 25030-24), 1% Pen/Strep (GIBCO, Cat.No. 15140-122) and 10% pooled, heat-inactivated and sterile filtered human type AB male off the clot serum (PAN Biotech, Cat.No. P40-2701, Lot.No:P-020317). 5ml of cell suspension was cultured in 6-well plates (Greiner Bio-One Cat.No. 657160). After thawing, cells were incubated at 37°C and 5% CO₂ for 6 hours in 0.2% DMSO. The final cell concentration was 2x10⁶ cells/ml. Cell viability and purity was assessed using FACS. All samples had a viability over 90% and over 95% of CD19+/CD5+/CD3- cells.”

with Nayara Paper “ATAC-seq libraries were generated as described previously[REFERENCE]. Cell preparation and transposition was performed according to the protocol, starting with 5x10⁴ cells per sample. Purified DNA was stored at 20°C until library preparation was performed. To generate multiplexed libraries, the transposed DNA was initially amplified for 5x PCR cycles using 2.5 μL each of 25 M PCR Primer 1 and 2.5 μL of 25 μM Barcoded PCR Primer 2 (included in the Nextera index kit, Illumina, San Diego, CA, USA), 25 μL of NEBNext High-Fidelity 2x PCR Master Mix (New England Biolabs, Boston, Massachusetts) in a total volume of 50 μL. 5 μL of the amplified DNA was used to determine the appropriate number of additional PCR cycles using qPCR. Additional number of cycles was calculated through the plotting of the linear Rn versus cycle, and corresponds to one-third of the maximum fluorescent intensity. Finally, amplification was performed on the remaining 45 μL of the PCR reaction using the optimal number of cycles determined for each library by qPCR (max. 13 cycles in total). The amplified fragments were purified with two rounds of SPRI bead clean-up (1.4x). The size distribution of the libraries was assessed on Bioanalyzer with a DNA High Sensitiv-

ity kit (Agilent Technologies, Santa Clara, CA), concentration was measured with Qubit DNA High Sensitivity kit in Qubit 2.0 Flurometer (Life Technologies, Carlsbad, CA). Sequencing was performed on NextSeq 500 (Illumina, San Diego, CA, USA) using 75bp paired-end sequencing, generating 450 million paired-reads per run, with an average of 55 million reads per sample.”

Spi-B and PU.1 shRNA Knockdowns

Immunohistochemistry of patient lymph nodes

Peter-Martin Bruch Paper “LN biopsies of CLL-infiltrated and non-neoplastic lymph nodes were Paraffin fixed and arranged in Tissue Microarrays. Consecutively they were stained for PAX-5 (790-4420, Roche), CD3 (790-4341, Roche), pSTAT6 (ab28829, abcam), STAT6 (519-4290, Zytomed Systems) and pIRAK4 (ab216513, abcam). The slides were analysed using QuPath 50 and the recommended protocol. Cell based data on mean staining intensity was exported and further analysed using R.”

Ibrutinib - IL4 - STAT6i interaction assay

2.3.2 Investigation of Ibrutinib + IL4 + IBET-762

Ibrutinib - IL4 - IBET-762 interaction assay

Ibrutinib - IL4 - IBET -762 treatment

ATACsequencing

with Peter Martin Bruch get from diffTF Peripheral blood was taken from 4 CLL patients and separated by Ficoll gradient (GE Healthcare), mononuclear cells were cryopreserved on liquid nitrogen. Samples were later thawed from frozen as previously described[REFERENCE] and MACS sorted for CD19 positive cells (Milteny autoMACS). The cells were resuspended in RPMI (GIBCO, Cat.No. 21875-034), with the addition of 2mM glutamine (GIBCO, Cat.No. 25030-24), 1% Pen/Strep (GIBCO, Cat.No. 15140-122) and 10% pooled, heat-inactivated and sterile filtered human type AB male off the clot

serum (PAN Biotech, Cat.No. P40-2701, Lot.No:P-020317). 5ml of cell suspension was cultured in 6-well plates (Greiner Bio-One Cat.No. 657160). After thawing, cells were incubated at 37°C and 5% CO₂ for 6 hours in 0.2% DMSO. The final cell concentration was 2x10⁶ cells/ml. Cell viability and purity was assessed using FACS. All samples had a viability over 90% and over 95% of CD19+/CD5+/CD3- cells.

with Nayara get from diffTF ATAC-seq libraries were generated as described previously[REFERENCE]. Cell preparation and transposition was performed according to the protocol, starting with 5x10⁴ cells per sample. Purified DNA was stored at 20°C until library preparation was performed. To generate multiplexed libraries, the transposed DNA was initially amplified for 5x PCR cycles using 2.5 μL each of 25 μM PCR Primer 1 and 2.5 μL of 25 μM Barcoded PCR Primer 2 (included in the Nextera index kit, Illumina, San Diego, CA, USA), 25 μL of NEBNext High-Fidelity 2x PCR Master Mix (New England Biolabs, Boston, Massachusetts) in a total volume of 50μL. 5 μ L of the amplified DNA was used to determine the appropriate number of additional PCR cycles using qPCR. Additional number of cycles was calculated through the plotting of the linear Rn versus cycle, and corresponds to one-third of the maximum fluorescent intensity. Finally, amplification was performed on the remaining 45 μL of the PCR reaction using the optimal number of cycles determined for each library by qPCR (max. 13 cycles in total). The amplified fragments were purified with two rounds of SPRI bead clean-up (1.4x). The size distribution of the libraries was assessed on Bioanalyzer with a DNA High Sensitivity kit (Agilent Technologies, Santa Clara, CA), concentration was measured with Qubit DNA High Sensitivity kit in Qubit 2.0 Flurometer (Life Technologies, Carlsbad, CA). Sequencing was performed on NextSeq 500 (Illumina, San Diego, CA, USA) using 75bp paired-end sequencing, generating 450 million paired-reads per run, with an average of 55 million reads per sample.

RNAsequencing

Proteomics

2.4 Data availability

European Genome-Phenome Archive (EGA) accession The data for and computational analysis code used in this study are available from

Additionally, we obtained CLL ATACseq data[REFERENCE] from the European Genome-phenome Archive (EGA: EGAD00001002110) and from (Bruch & Giles et al. 2021) For the ChIPseq analysis of SPIB binding sites, we accessed the data 37 from the NCBI GEO database 51, accession GEO: GSE56857. We made use of the data for SPIB in the OCILY3 DLBCL cell line (GSM1370276).

2.5 Statistical Analysis

Mixutre of paper and original Integrative data analysis of screening data, DNA and RNA sequencing, CNV and methylation profiles and follow up experiments was performed using R version 4 (REFERENCE R Core Team, 2018) with the RStudio interface (REFERENCE RStudio Team, 2016) and using packages that included (Love, Anders, and Huber 2021), (Therneau 2021), (Friedman et al. 2021), (M. Wilkerson and Waltman 2021), (Yu 2021b), (Yu 2021a), (Akalin et al. 2021) and (Oles et al. 2021) to perform univariate association tests, multivariate regression with and without lasso penalization, Cox regression, generalised linear modelling and clustering.

2.5.1 Data processing

Processing of screening data

with Peter-Martin Bruch Paper To quantify the response of a patient sample to each treatment condition, we used viability relative to the control, i.e., the CellTiter Glo luminescence readout of the respective well divided by the median of luminescence readouts of the DMSO control wells on the same plate.

Quality Assessment and Control

RNA data

RNA data was taken from the PACE repository (**PACE?**). For detailed explanation of processing see (**PACE?**).

ATACsequencing

Data from PACE

2.5.2 Analysis of Screening Data

Drug-drug and stimulus - stimulus correlations

myself original complete Pearson correlation coefficients were calculated for each drug - drug and stimulus - stimulus pair, using the `cor` functions of the (**R-stats?**) package with log transformed viability values which were normalised to untreated controls.

Characterising stimulus responses across all samples

myself original complete (just check) For the heatmap in figure 4.4, the viability data is represented as z scores. Log transformed viability values, normalised to DMSO controls, are row-scaled according to the Median Absolute Deviance. Limits were then applied to this row scaling factor for optimal visualisation. The matrix of z scores was plotted using the (Kolde 2019) package. The ordering of the columns (patient samples) was obtained from the dendrogram that resulted from running `ConsensusClusterPlus`, from the (M. Wilkerson and Waltman 2021) package. The rows were ordered using the dendrogram order produced by `hclust` with default branch arrangement.

Univariate analysis of gene - stimulus response associations

myself written for thesis complete Section ???. Two-sided Student's t-tests, with equal variance were performed for IGHV status and somatic mutations and copy number aberrations with 3 patient samples in each group (n = 63). A 10% FDR cut off was used to determine significance.

Penalised multivariate regression of gene - stimulus associations

myself written for thesis (but some overlap with JCI) Section 5.1.2. A Gaussian linear model with L1-penalty implemented in the R package `glmnet` (Friedman et al. 2021), with mixing parameter alpha = 1, was used to identify gene-stimulus associations. Matrix of genetic features ($p=39$), IGHV status (encoded as M = 1 and U = 0), and Methylation Cluster (encoded as 0, 0.5, 1) were used to identify multivariate predictors. All features were thus encoded on a similar scale to ensure equal treatment by lasso constraint in model fitting. Genetic features with more than 20% missing values were excluded from the analysis, and samples without complete annotation for remaining features were removed, resulting in $n = 129$ samples. The matrix of control-normalised log-transformed viability values for these 129 samples was provided as the dependent variable. Using 3-fold cross-validation, the optimal penalty parameter λ was selected so as to minimise the cross-validated R². The reduction in cross-validated mean squared error compared to the null model was used as loss. The model was fitted for 30 bootstrapped repeats, and the resulting coefficients are the mean of those coefficients that were selected in >75% model fits.

Correlation of RNA-Receptor Expression with viability

myself original complete RNA count data for matched samples was available for 49 patients and was transformed using the variance stabilising transformation. Stimulus - receptor pairs were defined using the available literature, see table ???. For each stimulus, pearson correlation coefficient were calculated between the control-normalised log-transformed viability value of a sample and its expression of the corresponding stimulus receptor.

Linear modelling of drug - stimulus interactions

Modelling of drug - stimulus - gene interactions

myself original To identify drug - stimulus interactions that were dependent on genetic features, we first fitted linear models to the viability matrix, for each drug - stimulus combination, using the formula in Eqn2. We extracted drug - stimulus interaction coefficients for each patient to generate a response matrix for n = 137 patients. Next, we performed multivariate regression using a Gaussian linear model with L1-penalty (i.e., lasso regression) as implemented in the R package `glmnet`(Friedman et al. 2021). As the dependent variable, the matrix of drug - stimulus interaction coefficients for each patient sample was used. As input to the model, genetic features with more than 20% missing values were excluded, and only patients with complete annotation were included in the model (n= 137). As predictors, the genetic mutations and CNVs (p= 39), IGHV status (coded as 0-1) and Methylation CLuster (coded as 0, 0.5, 1) were used, using 3-fold cross-validation. Misclassification error was used as loss for cross- validation.

2.5.3 Follow up

Lymphocyte doubling times

Herbst et al., 2020c - update “Patients which had lymphocyte counts available for less than 4 timepoints between the sample collection date and the time of the next treatment and patients currently in treatment were ex-cluded. Thus, . . . patients with enough data remained. Lymphocyte growth rates were calculated by fitting a linear model to the log10 transformed lymphocyte counts of all timepoints between the sample collection date and the time of the next treatment versus the period of time.”

Survival analyses

myself original Survival analyses were performed using TTT, TFT and OS as metrics. Follow-up information to calculate OS was available for all 192 CLL patients. For 188 of 192 CLL patients treatment information after sample collection was available.

Herbst et al "Time to next treatment (TTT) was calculated from the date of sample col-

lection to subsequent treatment initiation. Patients without treatment initiation during the observation time and patients who died before treatment initiation were censored at the latest follow-up contact.”

myself original For the purpose of visualising Kaplan-Meier plots , optimal cut-points of staining levels were calculated using maximally selected rank statistics as computed by the R package **maxstat** (**R-maxstat?**). Based on these cut points, patients were split into two groups, and their survival data were plotted using the Kaplan-Meier method, using the R package **survminer** (Kassambara, Kosinski, and Biecek 2021).

For Cox proportional hazards regression, the **coxph** function of the R package **survival** (Therneau 2021) was used. Expand

Penalised multivariate regression genetic predictors of cluster membership

Myself original (edit so not similar to JCI) I used a multinomial linear model with L1-penalty, implemented in the **glmnet** (Friedman et al. 2021) package. As the dependent variable, the cluster assignment for each patient was used. As input to the model ,genetic features with more than 20% missing values were excluded, and only patients with complete annotation were included in the model (n= 137). As predictors, the genetic mutations and CNVs (p= 39) and IGHV status (coded as 0-1) were used, using 3-fold cross-validation. Misclassification error was used as loss for cross- validation. The resulting coefficients indicated associations between genetic features and each cluster.

Gene expression and gene set enrichment analysis between clusters

myself original For the n=49 patient samples for which viability data and RNA–Seq data for matching samples was available, the R package **DESeq2** (Love, Anders, and Huber 2021) was used to search for associations of these two data types.RNA-Seq read count data was regressed on to the patient clusters C3, C4 (design formula ~ IGHV.status + Cluster). Genes were ranked by their test statistics and Gene Set Enrichment Analysis (GSEA) (implementing the fgsea algorithm with the **clusterProfiler** (Yu 2021b)

package) was applied to the ranked lists with the KEGG pathway gene set selections from the MSigDBdatabase[REFERENCE].

Associations of ex-vivo stimulus responses with genomic features

We tested for associations between stimulus viability assay results and genomic features by Student's t-tests (two-sided, with equal variance). We tested somatic mutations (aggregated at the gene level), copy number aberrations and IGHV status. We restricted the analysis to features that were present in at least 3 patient samples (63 features). p-values were adjusted for multiple testing by applying the Benjamini-Hochberg procedure.

Analysis of differential gene dosage in trisomy 12 CLL

myself written for thesis Section 5.2.2. For all RNA samples available in PACE (i.e. not just those that matched the samples in the screen), differential expression was called using the `DESeq2` package (Love, Anders, and Huber 2021), with the design formula `~trisomy12`. Raw RNA counts were visualised if the gene had BH-adjusted $p < 0.1$ and belonged to TGF β , JAK-STAT or TLR pathways genesets, as defined in the KEGG database (**Kegg?**) downloaded using the `msigdbr` package (Dolgalev 2021). Proteomic abundance data was also plotted, provided by Sophie Herbst (Herbst 2020). Proteomics data partially overlaps with RNAseq data. Protein quantification was performed as described in Herbst (2020).

Generation of a Trisomy 12 classifier

edit and finish Section 5.2.3. To build the classifier, a binomial linear model with L1-penalty implemented in the R package `glmnet` (Friedman et al. 2021), with mixing parameter alpha = 1, was used. To identify coefficients that predict trisomy 12 status, the viability matrix for all stimuli, as described in section 2.5.2 was used, i.e. z scores of the control-normalised log-transformed viability values.

Matrix of genetic features (p=39), IGHV status (encoded as M = 1 and U = 0), and

Methylation Cluster (encoded as 0, 0.5, 1) were used to identify multivariate predictors. All features were thus encoded on a similar scale to ensure equal treatment by lasso constraint in model fitting. Genetic features with more than 20% missing values were excluded from the analysis, and samples without complete annotation for remaining features were removed, resulting in $n = 129$ samples. The matrix of control-normalised log-transformed viability values for these 129 samples was provided as the dependent variable. Using 3-fold cross-validation, the optimal penalty parameter λ was selected so as to minimise the cross-validated R². The reduction in cross-validated mean squared error compared to the null model was used as loss. The model was fitted for 30 bootstrapped repeats, and the resulting coefficients are the mean of those coefficients that were selected in >75% model fits.

The classifier was built using binomial regression, with lasso penalisation, as implemented in the R package `glmnet` (Friedman et al. 2021). The feature matrix consisted of z scores of the viability values after treatment with each stimulus, and was used to predict the response (trisomy 12 status). I ran the model for 50 bootstrapped repeats, using three-fold cross-validation and mean absolute error as loss. Resiquimod, sCD40L+IL4 and TGF β were selected as coefficients that predict trisomy 12 status (Figure ??, as would be expected bed on the observations in section 5.2.1).

ATACseq processing

with Ivan Berest original (update to paper) We downloaded CLL data (Rendeiro et al. 2016) from the European Genome-phenome Archive (EGA: EGAD00001002110). Original dataset had 88 ATAC-seq samples from 55 patients, however we used for the analysis only one sample per patient passing quality checks, resulting in 52 samples. We used an in-house constructed ATAC-seq processing pipeline (Berest et al. 2019) to obtain final bam files mapped to the hg19 annotation genome that were corrected for CG bias.

Annotation of trisomy 12 status

with Ivan Berest original (update to paper) As trisomy 12 status was not included in the original metadata, we used a mean amount of reads in the chromatin accessible peaks for each sample to distinguish trisomy 12 patients. All samples containing 1.4 times more reads in the peaks located on chromosome 12, compared to the peaks on all other chromosomes, were classified as trisomy 12 patients.

diffTF analysis of TF activity in trisomy 12 CLL

with Ivan Berest original (update to paper) For the larger ATACseq dataset, we ran analytical mode of diffTF (**R-diffTF?**) pipeline to investigate differential TF activity between trisomy 12 and non-trisomy 12 patients using the HOCOMOCO v10 database (Kulakovskiy et al. 2016) with the following design formula: “~ Batch + Gender + IGHV status + trisomy12 status.”

for trisomy 12 software uses the ATACseq counts at each of the binding sites across the genome to generate a distribution of fold changes between trisomy 12 and WT samples. This fold change distribution is compared to a background distribution of fold changes, calculated similarly using ATACseq counts for a GC-matched motif, that does not contain the TF binding motif. The TF is classified as differentially active between trisomy 12 and WT samples where these two distributions are significantly different. Each TF is assigned a weighted mean difference value, which quantifies the change in activity, and a p value.

For the smaller ATACseq dataset, ATAC-seq data generated from our CLL samples were processed similarly with the in-house ATAC-seq pipeline, as described previously (Berest et al. 2019), with the only exception that we didn’t use CG bias correction step. We used a similarly analytical mode of diffTF with HOCOMOCO v10 database (Kulakovskiy et al. 2016) using the following parameters: minOverlap = 1; design formula = “~ Patient + Trisomy 12 status.”

Functional enrichment analysis of SPIB ChIPseq data

myself original (update to paper) We downloaded SPIB ChIPseq data [REFERENCE] from the NCBI GEO database[REFERENCE], accession GEO:GSE56857. We made use of the data for SPIB in the OCILY3 DLBCL cell line (GSM1370276). SPIB ChIP peaks were filtered for significance (q value <0.05). We used the annotatePeaks function from the package clusterProfiler[REFERENCE] to annotate the nearest gene for each ChIP-peak. We filtered peaks within $\leq 1\text{kb}$ of a TSS of a gene and performed over-representation of KEGG and Reactome pathways (using the clusterProfiler package[REFERENCE]) amongst the resulting list of genes.

Linear modelling of drug - stimulus interactions

myself original To identify drug-stimulus interactions, we fitted linear models to the viability matrix, for each drug - stimulus combination, using the formula in Eqn1. Drug - stimulus interaction coefficients and associated p values were extracted and used to define significant interactions.

Drug-Drug interaction assay

An independent patient cohort of 16 patients was used in the drug-drug interaction experiments.

Survival analysis of immunohistology chemistry data

2.5.4 Ibrutinib - IBET-762 - IL-4 analysis

RNAseq processing

Deseq2 analysis

ATACseq processing

diffTF analysis

Generation of Gene Regulatory Network

Proteomics Processing

Proteomics Analysis

Chapter 3

Data

Characterisation of primary CLL samples by high-throughput combinatorial screening and multi-omic profiling. An introduction to the dataset that my PhD is based on.

3.1 Drug screens and experiments

3.1.1 High-throughput combinatorial perturbation assay

We measured the effects of 17 cytokines and microenvironmental stimuli on cell viability in 192 primary CLL samples and combined each one with 12 drugs to investigate the influence on spontaneous and drug-induced apoptosis (Figure 3.1). Viability was assessed by ATP measurement via CellTiterGlo after 48 h of culture and normalised to untreated controls⁸.

3.1.2 Validation experiments

In addition to the screen, we acquired the following validationary data (Figure 3.2).

Proper explamation of data srouces and how we acuried them, the format they are in

Figure 3.1: Schematic of experimental protocol. By combining 12 drugs and 17 stimuli, we systematically queried the effects of simultaneous stimulation and inhibition of critical pathways in CLL ($n=192$). Integrating functional drug-stimulus response profiling with four additional omics layers, we identified pro-survival pathways, underlying molecular modulators of drug and microenvironment responses, and drug-stimulus interactions in CLL.

Figure 3.2: Schematic of validationary data.

3.2 Characteristics of drugs used in the screen

3.2.1 Drug pathways

We screened 17 different drugs (Figure 3.3).

Figure 3.3: Bar plot of the drugs used in screen.

3.2.2 Drug responses

The drug responses were as follows (Figure 3.4).

3.2.3 Genetic predictors of drug responses

3.2.4 Drug - Drug Correlations

3.3 Characteristics of stimuli used in the screen

3.3.1 Stimulus responses

3.3.2 Stimulus - Stimulus Correlations

Figure 3.4: Log transformed viability values for all drugs that were included in the screen after quality control. p values from student's t test.

3.4 Characteristics of patient samples used in the screen

3.4.1 Genetic Data available for each patient

WES, CNVs, Methylation, Transcriptomic, ATACseq, LDT, survT, IHC PLus Ibrutinib + IBET + IL4 treated samples - ATACseq, RNaseq, proteomics

3.5 Processing of raw values obtained from cell viability assay

3.5.1 Data normalization and quality control

3.5.2 How viability is calculated, why we use log values etc

3.5.3 Heterogeneity of response is not caused by differences in receptor expression

Next, we tested whether the observed heterogeneity of response was caused by differences in receptor expression. We calculated Pearson correlation coefficients comparing control - normalised log viability values for each stimulus with transformed RNA counts of the corresponding stimulus receptor(s). No stimulus-receptor pair showed a Pearson coefficient greater than 0.4, confirming that the heterogeneity of response was not caused by differential receptor expression.

3.6 Generation of public resource

3.6.1 Shiny app

3.6.2 Package

- Present the underlying economic model/theory and give reasons why it is suitable to answer the given problem¹.

¹ Here is an example of a footnote.

Chapter 4

Ex-vivo sensitivity to microenvironmental stimulation in primary CLL cells

The screen included 17 cytokines and microenvironmental stimuli, which were selected based on evidence in the literature that each stimulus had been shown to impact on CLL viability *in vitro*. We attempted to put together the largest panel of stimuli for an assay of this kind in CLL, aiming to minimise redundancy among the compounds (Bruch & Giles et al. 2021).

CLL cells do not proliferate *in vitro*, but rather undergo spontaneous apoptosis in the absence of stimulation (Collins et al. 1989). The effect of each stimuli on CLL viability was thus quantified by comparing ATP counts in treated primary samples, compared with those in DMSO wells where a positive value indicates that the sample viability was increased relative to control.

The assay represents a reductionist model of microenvironmental signalling, making it possible to dissect the effect of individual stimuli on baseline viability. The large

patient cohort also enabled us to study the differential impact of these stimuli across heterogeneous samples, and identify distinct subgroups of patients whose samples show similar response profiles to the panel of stimuli.

4.1 Proliferating responses to the panel of stimuli

4.1.1 *ex vivo* assay demonstrated functional diversity of cytokines and microenvironmental stimuli

To investigate heterogeneity amongst responses to the stimuli, I calculated Pearson correlation coefficients for each stimulus pair, using the log-transformed normalised viabilities (Bruch & Giles et al. 2021). The resulting coefficients were ordered using hierarchical clustering and visualised in a symmetrical heatmap (Figure 4.1). The hierarchical clustering distinguished clusters of stimuli, including a larger group corresponding to agonists of TLR and Nf κ B pathways and a smaller group encompassing IL4 and TLR stimuli.

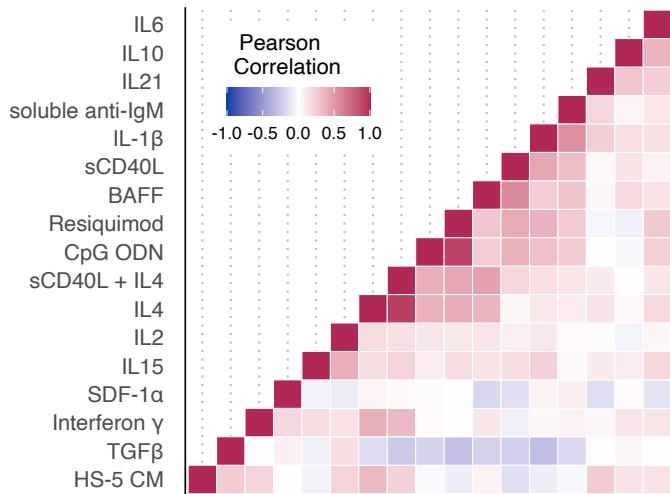


Figure 4.1: Heatmap of Pearson correlation coefficients. Coefficients for each pair of stimuli were calculated using log transformed viability values normalised to untreated control, and ordered according to hierarchical clustering. Figure adapted from Bruch & Giles et al. 2021.

98.5% of stimulus pairs showed little correlation ($R < 0.6$), including those that targeted similar downstream pathways, indicating a high degree of functional diversity amongst microenvironmental signals. For example, JAK-STAT agonists such as IL4 and IL6 showed little correlation (Figure 4.2A).

Only two stimulus pairs showed correlations where $R > 0.6$, and in both cases these targeted near identical receptors or downstream pathways. These included CpG ODN (TLR 9) and Resiquimod (TLR 7 and 8) (Figure 4.2B), and IL4 and IL4 + CD40L which primarily target JAK3 - STAT6.

Repeating the analysis to correlate drug - drug pairs demonstrated that drugs targeting components of the same pathway were highly correlated. For quality control purposes, this indicated that our data sensitively and specifically reflect inter-individual differences in pathway dependencies (Dietrich et al. 2017).

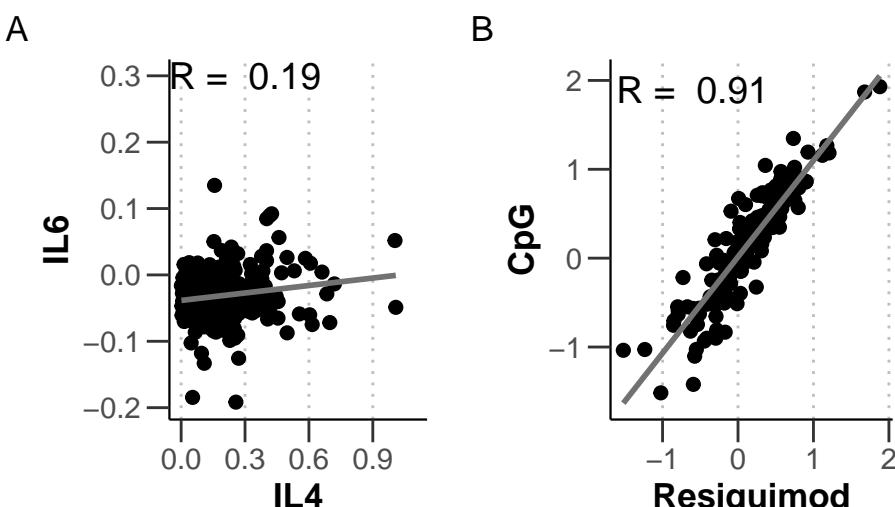


Figure 4.2: Scatter plot of log-transformed viability values, normalised to DMSO controls, for (A) treatment with JAK-STAT agonists IL4 and IL6 and (B) treatment with TLR agonists CpG ODN and Resiquimod.

Microenvironmental stimulation induced diverse phenotypes between patient samples, and across different stimuli. To gain a global overview of these phenotypes, I visualised log-transformed viability values normalised to DMSO controls for all patient samples

and all stimuli (Figure 4.3 (Bruch & Giles et al. 2021)).

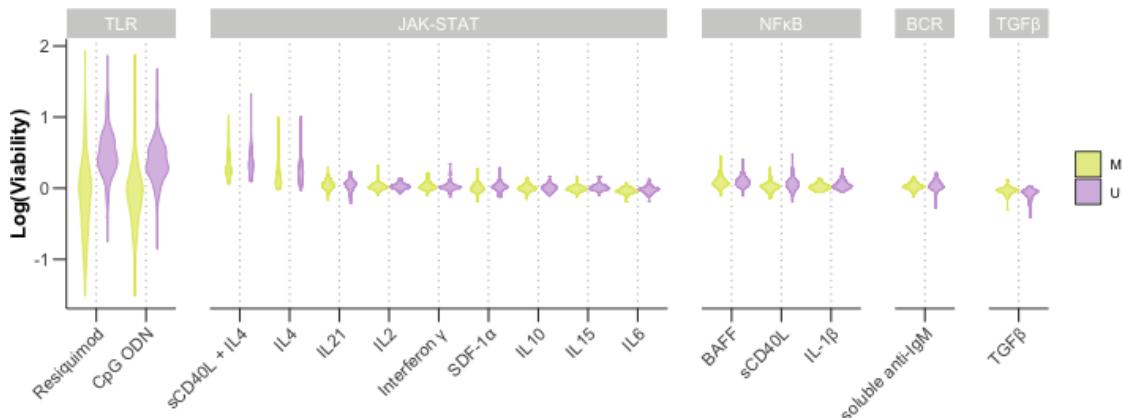


Figure 4.3: Log transformed viabilities after treatment with each stimulus.

Where stimuli decreased viability relative to control, points are shown in blue, whilst increased viability is shown in red. Figure adapted from Bruch & Giles et al. 2021.

The majority of the stimuli increased viability, underlining the supportive nature of the microenvironment in CLL. However, three out of 17 reduced CLL viability relative to control, namely IL6, TGF β and TLR 7/8/9 agonists in IGHV-mutated (IGHV-M) samples.

IL4 and TLR7/8/9 agonists Resiquimod and CpG ODN induced the strongest responses, an indication of their potency in modulating CLL cell survival. Notably, TLR agonists increased viability in certain samples, in most cases IGHV-U, and decreased viability in others, mostly IGHV-M. Both IL4 and TLR7/8/9 emerged as the key pathways in the screen, and play an important role throughout the results of this thesis.

4.1.2 Microenvironmental response profiling identifies discrete patient subgroups

To further investigate the variability in responses across the cohort, we visualised z scores of the log-transformed viabilities and performed consensus clustering on the resulting

matrix to generate a heatmap of all stimuli responses across all samples (Figure 4.4).

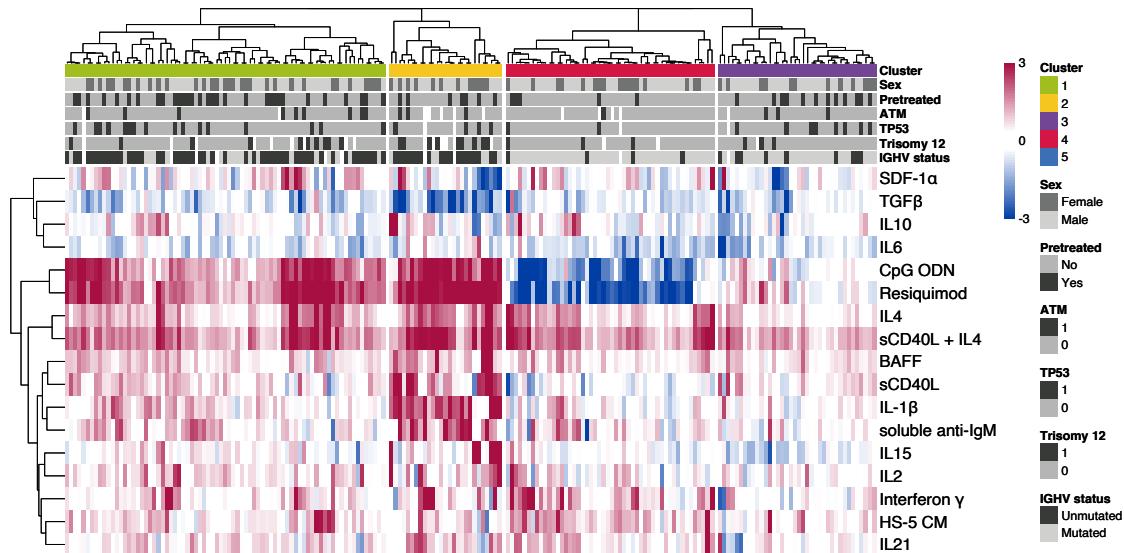


Figure 4.4: The heatmap matrix shows the viability measurements for 192 samples (columns) and 17 stimuli (rows). Viability was measured via ATP-based assay after 48h stimulation. The data are shown normalised to DMSO-treated controls, and scaled within each row according to the Median Absolute Deviation (MAD). Limits were applied to scaling factor for optimal visualisation. The colour bars to the right show sample annotations. Consensus Clustering was used to define column tree layout, using hierarchical clustering with the Euclidean metric. Figure from Bruch & Giles et al. 2021.

The heatmap reflects heterogeneity in responses across samples and stimuli, once again underlining the potency of IL4 in increasing sample viability across diverse genetic backgrounds and the diversity in responses to TLR stimulation by Resiquimod and CpG ODN.

To generate the column-wise clustering, consensus clustering was performed on the matrix of z scores. Consensus clustering allows the user to subsample from the matrix of values, to generate hierarchical clustering for a given number of clusters, k. From this, it is possible to calculate a consensus matrix for each value of k, indicating for each

pair of values the proportion of time they occupy the same cluster when subsampled together.

I visualised the clustered heatmap in Figure 4.4 for different values of k , and concluded on the existence of four robust clusters within the cohort (add these to appendix?). Each cluster shows a unique response profile to the panel of stimuli. We termed the clusters C1 to C4: C1 and C2 were enriched for IGHV-U whilst the samples in C3 and C4 were mostly IGHV-M.

To validate the choice of four clusters, I visualised the summaries of the consensus matrix, using the **ConsensusClusterPlus** package (M. D. Wilkerson and Hayes 2010), to quantify the degree of confidence in the clusters for different values of k .

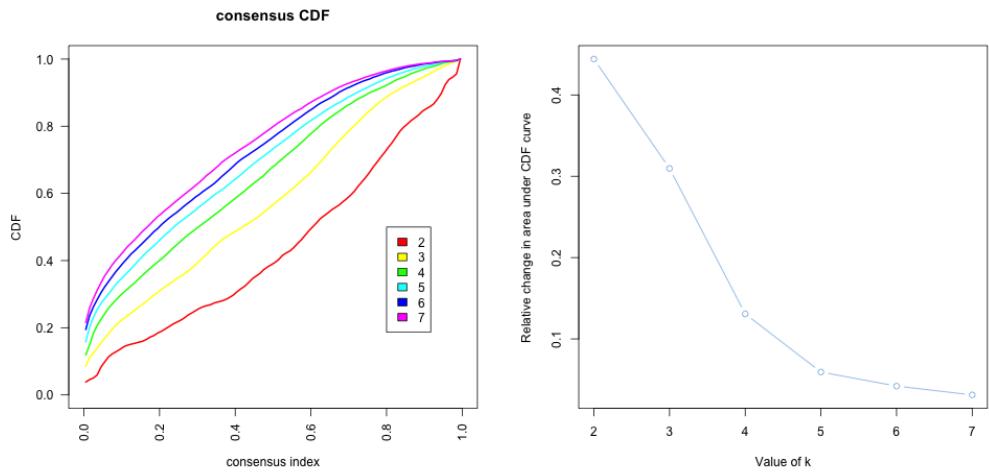
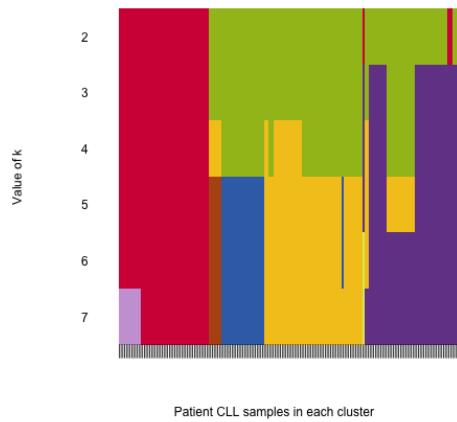


Figure 4.5: Summaries of the CDFs of the consensus matrices. Consensus CDF graphic showing the CDFs of the consensus matrix for $k = 2 - 7$, as indicated in the legend, estimated using 100 bin histogram (left). Relative change in area under the CDF curve, for $k = 2 - 7$, to compare k with $k - 1$. In the case of $k = 2$, there is no $k - 1$, so the total area is plotted. Line shows relative increase in consensus between each value of k (right).

The graph of the CDFs of the consensus matrix for each k indicated that the CDF reaches a maximum and cluster confidence is maximised at $k = 7$, though above $k = 4$

there is little appreciable increase. Figure 4.5. This is confirmed in the graph showing relative change in the area under the CDF curve, showing there is only a small increase in consensus between $k = 4$ and $k = 5$, supporting the choice of $k = 4$. The cluster tracking plot depicts how each patient sample is assigned for each value of k . For $k = 4$, the plot indicates that C3 and 4 in particular are highly stable (Figure 4.6).



Amongst the IGHV-M enriched clusters, C3 showed weaker responses to the majority of stimuli, and C4 was defined by a negative response to TLR7/8/9 stimulation (Bruch & Giles et al. 2021). Figure ?? summarises these findings in more detail, showing responses stratified by cluster for a subset of the stimuli.

cairo_pdf

2

4.2.2 The clusters show differences in disease dynamics

To validate the potential biological significance of these four clusters, we investigated whether the groups showed differential *in vivo* disease progression (Bruch & Giles et al. 2021). The study design was such that not all patients in the cohort were treatment - free, which confounded the analysis. For that reason, lymphocyte doubling time (LDT) and time to next treatment (TTT) were used to quantify CLL proliferative capacity, independently of treatment.

C1 and C2 showed a shorter LDT than C3 and C4, which is expected due to the differential proportions of IGHV-U and M patient samples in these groups (Figure 4.7A). Notably, within the IGHV-M enriched clusters C3 and C4, samples in C3 showed a significantly shorter LDT (Student's t-test, p-value = 0.025).

To further validate this, we observed that TTT in the IGHV-M enriched C3 was significantly shorter than C4 and comparable to the progression dynamics of IGHV-U enriched C1 and 2 (Figure 4.7B).

The difference in disease progression between the clusters indicated that microenvironmental response represents an additional biological layer, holding information relevant to disease dynamics. To validate that these clusters were not simply an indication of any underlying genetic features, we checked whether the observed differences in progression dynamics could be explained by other prognostic markers (Bruch & Giles et al. 2021).

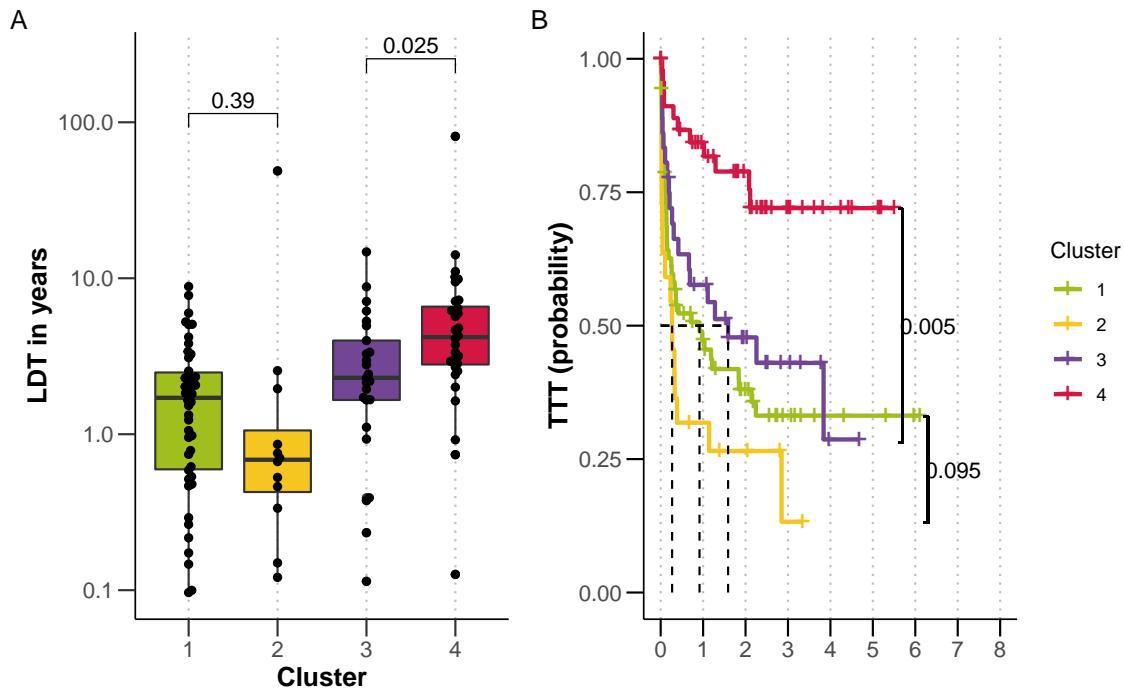


Figure 4.7: (A) Lymphocyte doubling time (LDT) stratified by cluster, p-values from Student's t-test. (B) Kaplan-Meier curves to show TTT for each cluster. p-values from univariate Cox proportional hazard models comparing IGHV-U enriched C1 with C2, and IGHV-M enriched C3 with C4. Figure from Bruch & Giles et al. 2021.

A multivariate Cox proportional hazard model accounting for IGHV status, trisomy 12 and TP53 in addition to the cluster assignment indicated an independent prognostic value for cluster assignment between C3 and C4 ($p= 0.039$, Table 4.1).

4.2.3 The clusters showed differential responses to drugs *in vitro*

The potential clinical relevance of the clusters was underlined by my observation that the samples within each group showed differential responses to drugs *in vitro* (Figure 4.8).

As expected, the IGHV-U enriched clusters C1 and 2 were more sensitive to BCR inhibition by ibrutinib, idelalisib and PRT062607, than C3 and 4. Between C1 and C2,

Table 4.1: Table depicting results of Multivariate Cox Proportional Hazard Model to test prognostic value of key genetic features and clusters using Time to Next Treatment and C3 as reference.

Factor	coef	exp(coef)	se(coef)	z	p.value
Cluster 3 vs Cluster 1	-0.03979	0.96099	0.29813	-0.13347	0.89382
Cluster 3 vs Cluster 2	0.51595	1.67522	0.37741	1.36708	0.1716
Cluster 3 vs Cluster 4	-0.82011	0.44038	0.39760	-2.06267	0.03914
IGHV.status	0.55192	1.73658	0.27253	2.02513	0.04285
trisomy 12	-0.13357	0.87496	0.35617	-0.37503	0.70764
TP53	1.38977	4.01395	0.26072	5.33058	<0.0001

C2 was more sensitive to a number of the drugs, including idelalisib (SYK) (p-value = 0.012), everolimus (mTOR) (p-value = 0.02) and the chemotherapeutics fludarabine (p-value = 0.031) and nutlin-3a (p-value = 0.042). Amongst C3 and C4, C3 showed lower sensitivity to everolimus (p-value = 0.051) and to fludarabine (p-value < 0.001) and nutlin-3a (p-value = 0.01). This aligns with the observation that patients in C3 have a poorer prognosis, despite most of these samples annotated as IGHV-M. C4 also showed a positive response to Nf κ B inhibition by BAY-11-7085, and p38 MAPK inhibition by Ralimetinib.

Such differential drug response patterns suggests that microenvironmental response may reflect disease-specific CLL biology, in the same way as molecular profiling, and thus may have the potential to guide therapy decisions in future.

4.2.4 The clusters are enriched for different genetic features

Next we assessed differences in the molecular profiles of samples within each cluster. Visually, it appeared that certain clusters were enriched or depleted for various genetic features recurrent in CLL (Figure 4.9).

To quantify this, I ran a multinomial model, with lasso regularisation, to predict cluster

membership (C1 -4) based on the matrix of genetic features for all the patient samples (Bruch & Giles et al 2021). The model assigned coefficients to genetic features, where a positive coefficient indicated that this feature was enriched in the cluster, and a negative coefficient indicated it was depleted.

The approach used 3-fold cross validation, selecting the optimal model using lambda min. To ensure that the resulting coefficients were robust, we selected coefficients that satisfied certain cut-offs. Coefficients were selected if they were assigned in > 60% of 50 bootstrapped repeats, and were larger than 0.35. Figure 4.10 shows the mean coefficients and associated standard deviation, for each genetic feature that met these criteria in each cluster.

As we expected, IGHV status was the main feature that predicted cluster membership. Beyond IGHV status, trisomy 12 and *SF3B1* mutations were enriched in C2, which showed enhanced responses to many stimuli. C4, which was associated with slow in-vivo progression, showed depletion of *TP53*, *ATM*, RAS/RAF mutations and gain8q.

4.2.5 GSEA of DE genes between subgroups

In addition to genetic features, I investigated differential expression of genes within each cluster. For n = 49 samples, RNAseq data was available for matched PBMC samples. I focused on the difference between clusters 3 and 4, for which 21 RNAseq samples were available (Bruch & Giles et al. 2021).

To quantify differential gene expression, I began by filtering out immunoglobulin genes, including genes at the heavy, light and kappa loci that encode the antigen receptor of B cells. The clusters each show differential enrichment of IGHV-M and U samples, and thus the differential expression analysis would otherwise be dominated by immunoglobulin genes that are well known to be affected by this biomarker.

I followed the Deseq2 protocol using a design formula to quantify the difference between clusters, and accounting for the confounding effect of IGHV status. 87 genes were differentially expressed (adjusted p < 0.05) between C3 and 4 (Figure 4.11).

To assign biological meaning to the differentially expressed genes, I quantified the enrichment of Hallmark pathways amongst the genes. I ranked the genes based on the Wald statistic, and then ran GSEA was using the fgsea algorithm (Figure 4.12) (Bruch & Giles et al. 2021).

Several pathways were upregulated amongst samples in C3, compared to C4, indicating that these pathways may relate in some way to the shorter TTT and LDT of patients within this cluster. Pathways associated with higher disease aggression were regulated in C3 including genesets relating to stress response (Unfolded Protein Response, UV Response Up, P53 Pathway), metabolism (Oxidative Phosphorylation) and proliferation (G2M Checkpoint, MYC Targets V1, MTORC1 Signaling, E2F Targets) (Figure 4.13).

In addition, C3 showed upregulation of microenvironmental signalling pathways relative to C4, including TNFa Signalling via NFKB and Interferon Gamma Response (Figure 4.13) (Bruch & Giles et al. 2021). This finding underlines our hypothesis that differential activity of microenvironmental signalling, both *in vivo* and *ex vivo* may be relevant to disease prognosis.

4.3 Additional Analysis

Can microenvironmental response aid prognostic models? Run a Multivariate model to predict survival/drug response with genetic matrix and stimuli response matrix?

4.4 Summary

The screen represents an attempt to comprehensively dissect the impact of individual microenvironmental pathways on CLL viability. Our assay has enabled us to highlight key, broad spectrum signals such as IL4. Our heterogeneous cohort also reveals pathways that operate in subsets of patients, such as TLR. IL4 and TLR represent the key pathways that we focus on throughout the rest of this thesis.

In addition, this approach enables us to use microenvironmental response profiles, as an alternative to molecular profiling, to look for the existence of subgroups.

We uncovered four such subgroup with distinct response profiles and molecular properties and clinical outcomes, suggesting that microenvironmental response holds biologically significant information that may be relevant to prognosis and treatment decision making.

4.5 Discussion

Gosia suggested discussing each results section individually?

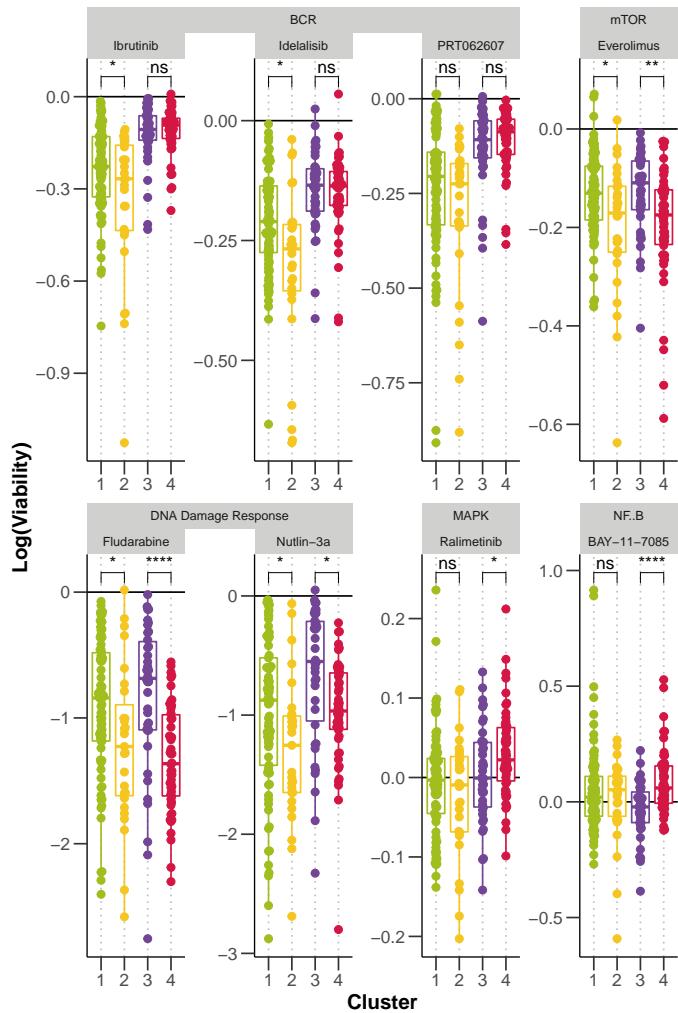


Figure 4.8: Log-transformed normalised viability values, stratified by cluster, for each drug. Drugs targeting the same pathway are grouped together. P-values from Student's t-test.

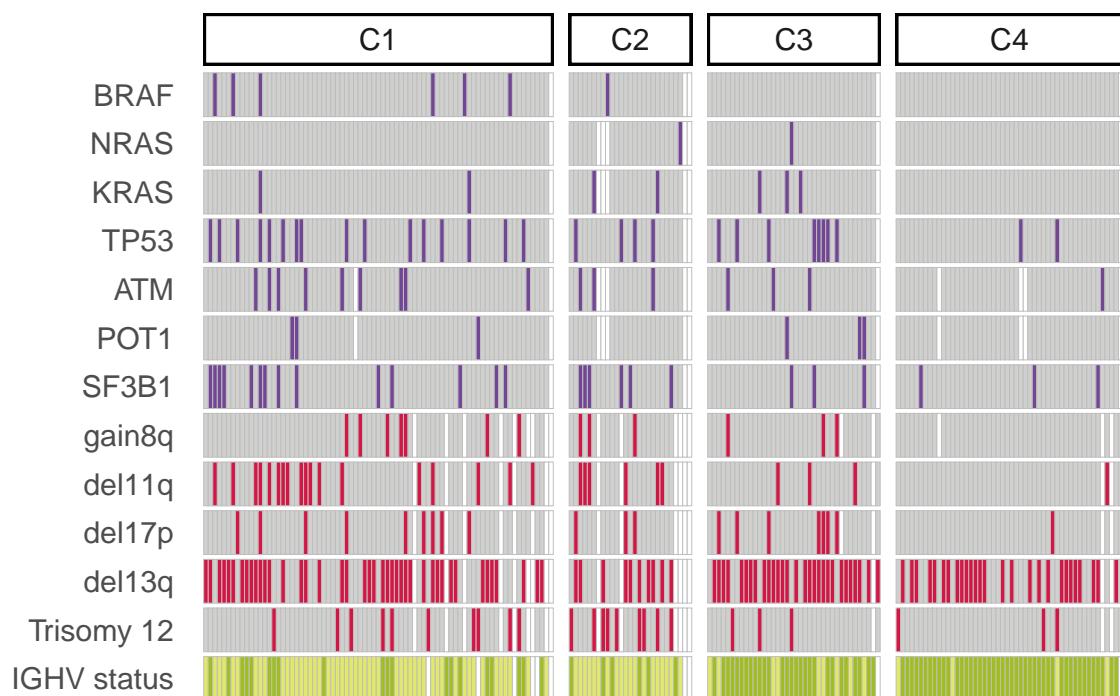


Figure 4.9: Distribution of selected genetic features (rows) within each cluster for all patient samples (columns). Where a patient sample is not annotated for a feature, this is marked in white.

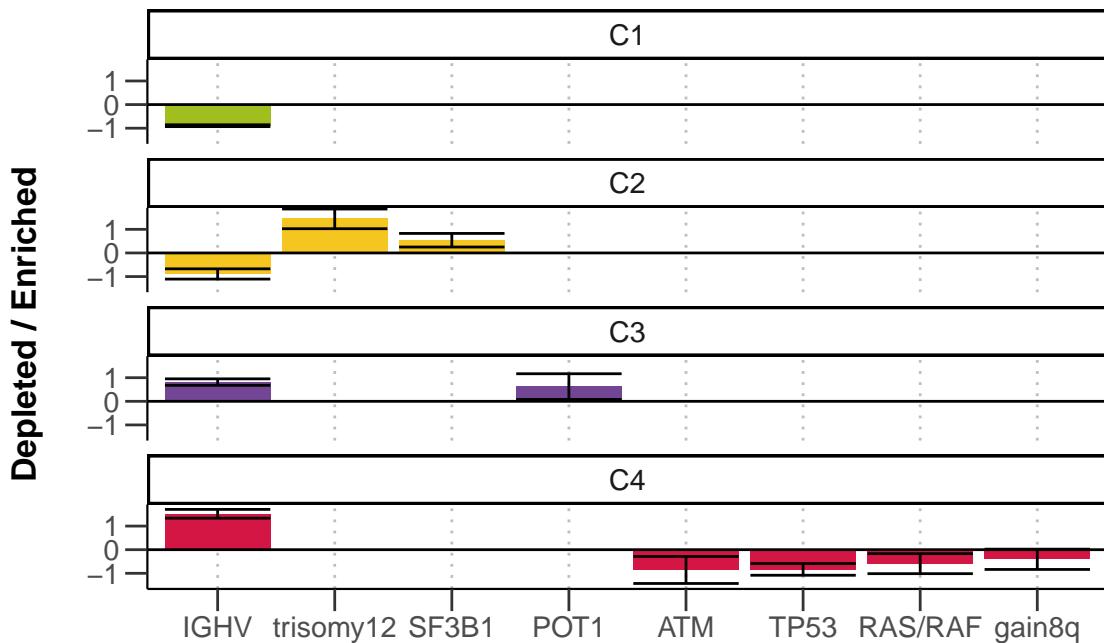


Figure 4.10: Multinomial regression with lasso penalisation to identify enrichment or depletion of genetic features within each cluster. Matrix of genetic features ($p=39$), and IGHV status (encoded as $M = 1$ and $U = 0$) were used to identify multivariate predictors of cluster assignment. x axis shows genetic predictors, y axis indicates value and sign of coefficient assigned to feature, for each cluster (positive coefficients are enriched in the cluster, negative coefficients are depleted). Coefficients shown are mean coefficients from 50 bootstrapped repeats and error bars represent the mean \pm standard deviation. Genetic features with $>20\%$ missing values were excluded, and only patients with complete annotation were included in the model ($n=137$). Figure from Bruch & Giles et al 2021.

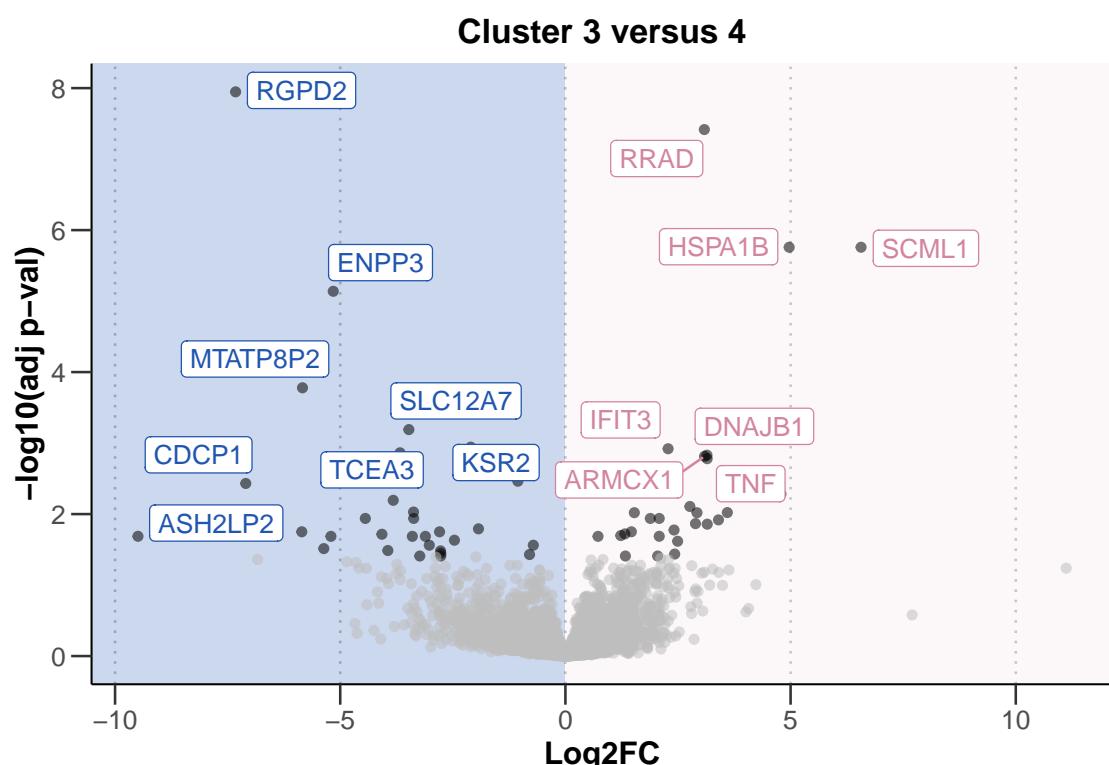


Figure 4.11: Volcano plot of differentially expressed genes between C3 and C4. X axis indicates log₂ fold change values, calculated using the DESeq2 package (**DESeq2?**), y axis gives corresponding $-\log_{10}(\text{adjusted p value})$. P values adjusted using BH method. Genes are labeled where adjusted p < 0.05. Figure from Bruch & Giles et al. 2021.

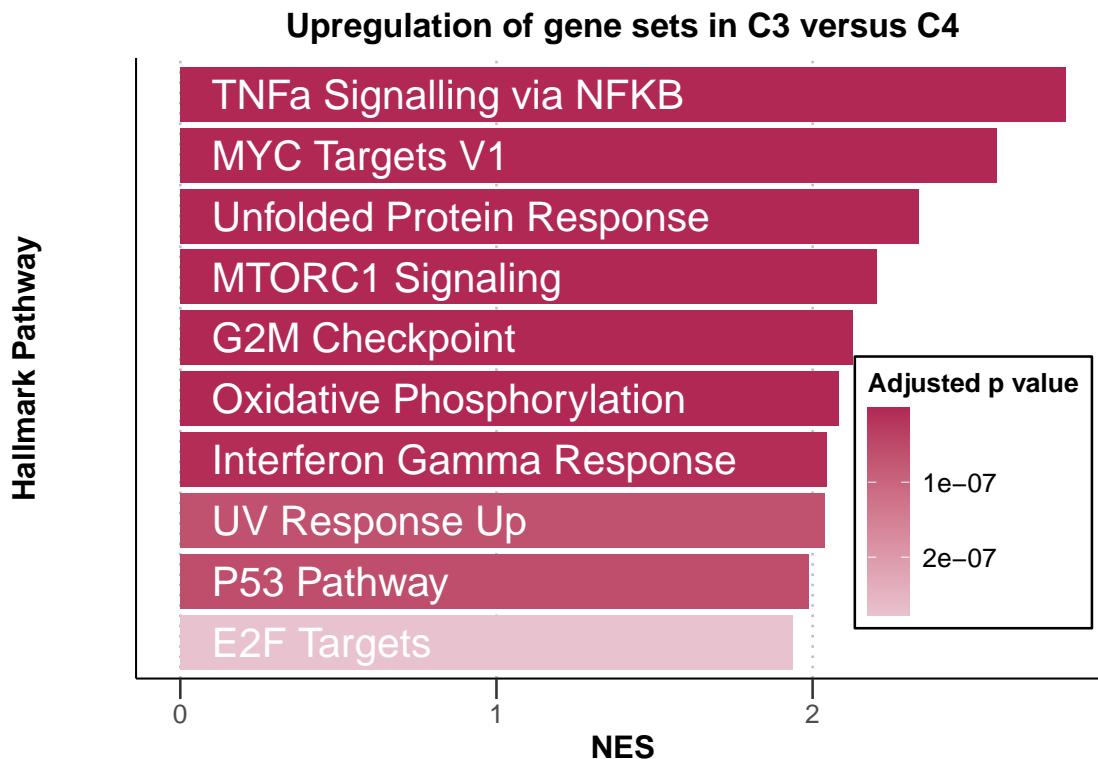


Figure 4.12: Gene set enrichment analysis (GSEA) to compare expression of genes in samples from C3 and C4 reveals upregulation of Hallmark pathways involved in microenvironmental signalling, stress response, metabolism and proliferation in C3. Normalised enrichment scores (NES) are shown for top 10 most significant pathways upregulated in C3 versus C4. Bars coloured according to adjusted p-value. Genes are ranked based on Wald statistic, calculated using the Deseq2 package and GSEA performed using the fgsea algorithm. Figure from Bruch & Giles et al. 2021.

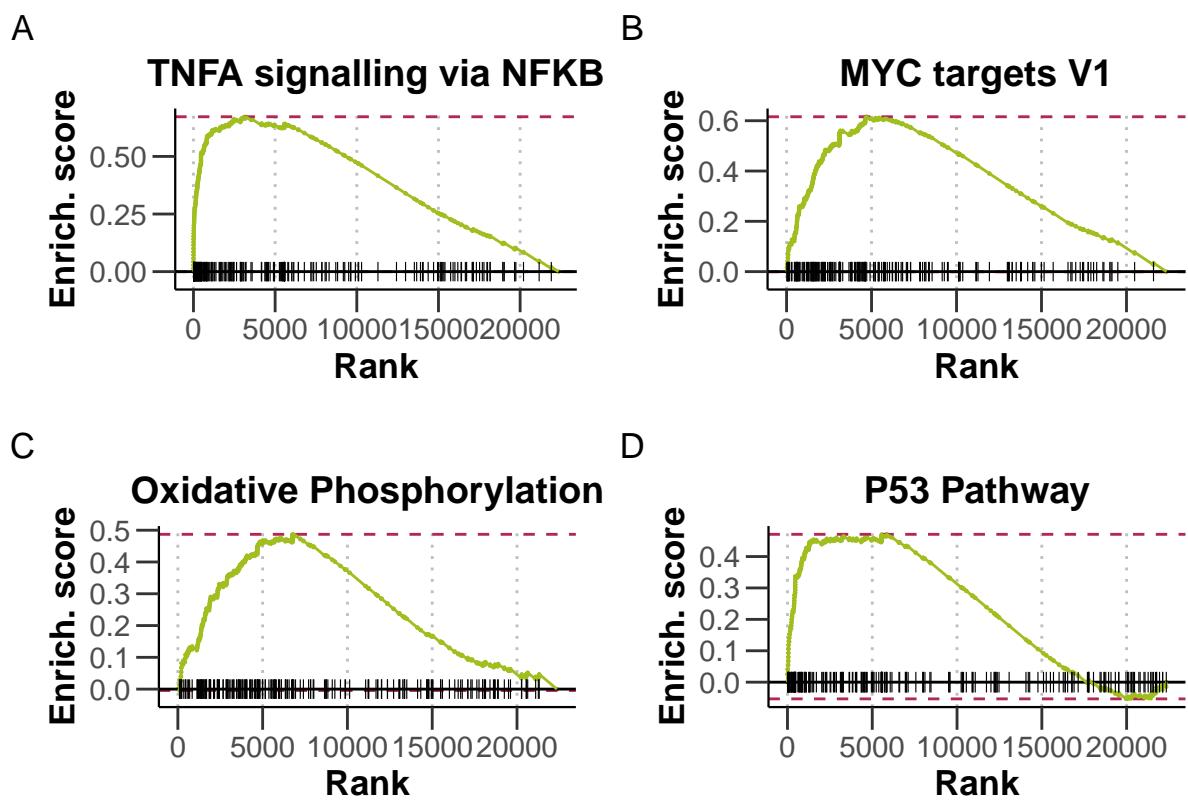


Figure 4.13: Enrichment plots of selected pathways. Gene set enrichment analysis (GSEA) was performed with the Hallmark gene sets from the GSEA Molecular Signatures Database. Wald statistic was used to rank the genes. The green curve corresponds to the Enrichment Score curve, which is the running sum of the weighted enrichment score obtained from GSEA software. Figure from Bruch & Giles et al. 2021

Chapter 5

Genetic modulators of responses to microenvironmental stimulation

Profiling the effects of the panel of stimuli revealed the heterogenous nature of responses to stimulation. The analysis revealed four patient sample groups, with distinct responses to the stimuli, unique molecular profiles, and differential disease progression. Two pathways emerged as the most potent modulators of CLL viability, namely IL4 and TLR. These too induced heterogeneous responses across the cohort, in particular stimulation of TLR 7/8/9 increased viability in some samples, whilst it reduced viability in others.

Next we asked to what extent this heterogeneity of response relates to the molecular profiles of the tumours. We sought to understand which underlying genetic features might modulate responses to external signals, and how these interactions may occur. We combined the screening dataset with multi-omics profiles of the patient samples taken from the PACE repository (Dietrich et al. 2017; Oles et al. 2021) and performed a systematic survey of molecular determinants of stimulus response, using whole-exome sequencing,

DNA-methylation, RNA-sequencing and copy number variant data. In addition, for a small subset of patients we also generated ATAC sequencing and Mass Spectrometry data.

Collectively, these data enabled us to probe genetic and epigenetic modulators of microenvironmental signalling, in a heterogeneous cohort that encompasses the clinical and molecular diversity of CLL. In the first part of this chapter, I apply a broad systematic approach to identify important genetic modulators, and in the second, I outline my follow-up investigations into the impact of trisomy 12 on stimulus response.

5.1 Systematic analysis of the effect of genetic features on responses to stimuli

5.1.1 Univariate analysis identifies IGHV status and trisomy 12 as key modulators of microenvironmental response

To begin, I ran a univariate analysis to compare viability values post-stimulation for patient samples with and without each genetic feature. In total 63 genetic features were surveyed, including IGHV status, somatic gene mutations and structural variants, where there were at least three patient samples in each group (Figure 5.1).

This analysis revealed the extent to which genetic features modulate microenvironmental response: for ten out of 17 stimuli, at least one genetic feature determined response and for six out of 17 stimuli, two or more genetic features significantly altered the response (Student's t-tests, FDR = 10%, Bruch & Giles et al. 2021). The most common features were IGHV status and trisomy 12. Del(11q) also affected response to several stimuli. Notably, del(13q) and del(17p), which like trisomy 12 and del(11q) are the most common aberrations in CLL and act as prognostic markers (Döhner et al. 2000), had no impact on the responses to the panel of stimuli.

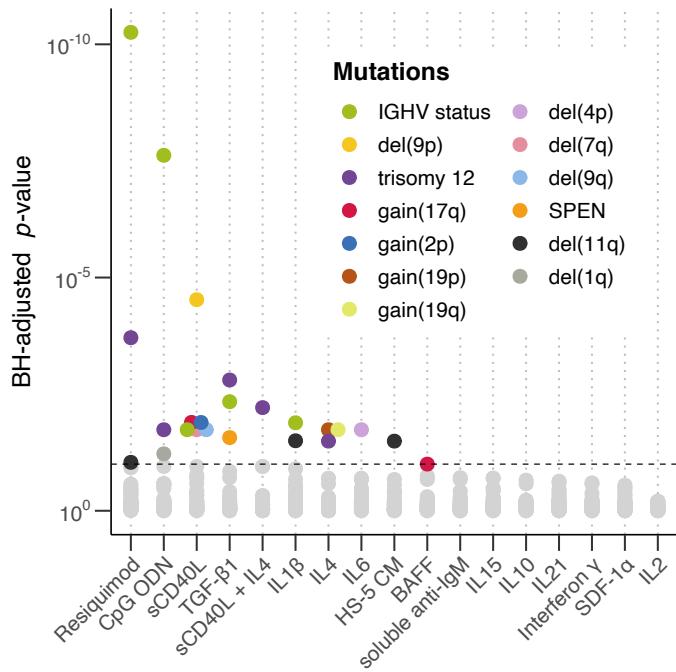


Figure 5.1: Plot showing BH-adjusted p values from Student's t-tests (two-sided, with equal variance), for all tested gene-stimulus associations. Tests performed for IGHV status and somatic mutations and copy number aberrations with 3 patient samples in each group ($n = 63$). Each circle represents a gene-stimulus association. Associations that meet 10% FDR cut off are indicated in colour, where the colour denotes the genetic feature. *Figure from Bruch & Giles et al. 2021.*

5.1.2 Multivariate analysis of gene - stimulus assosciations confirms IGHV status and trisomy 12 as key modulators of stimulus response

It was possible that there may be interplay between genetic factors in determining responses to external signals. To address this, I applied multivariate modelling to integrate the influence of genetic features, IGHV status and DNA methylation on the size of response (Bruch & Giles et al. 2021). I used a Gaussian linear model with L1-penalty (i.e., lasso regression), to derive a predictor for each stimulus, comprised of these covariates.

For background on this approach, see section 1.3.6.

As input to the model, the response matrix was composed of the log transformed viability values for each stimulus. To generate the feature matrix (137 samples versus 39 features), I excluded genetic features for which >20% of the values were missing, and patient samples with incomplete annotation. As predictors, I included genetic mutations and CNVs ($p= 39$), IGHV status (coded as 0-1) and Methylation Cluster (coded as 0, 0.5, 1). I ran lasso regression, as implemented in the R package `glmnet`(Friedman et al. 2021), using three-fold cross-validation with misclassification error as loss. The resulting predictors are the mean of those coefficients that were selected in at least 75% of 30 bootstrapped repeats.

Using the output of the regression, I generated predictor profiles for each stimulus. For 5 / 17 stimuli, there was at least one genetic predictor that met the cut-offs (a selection are shown in figure 5.2).

The multivariate analysis demonstrated that responses to IL4 and sCD40L + IL4, TLR and TGF β were all affected by multiple genetic features (Bruch & Giles et al. 2021). For example, higher viability in response to IL4 stimulation was associated with trisomy 12 and unmutated IGHV. In contrast, IL4 generated little or no increase in viability in samples with a mutation in KRAS, NRAS or BRAF. These tumours benefited less from the anti-apoptotic effects of IL4, indicating that the ability of IL4 to increase viability could be dependent on signalling via Ras-Raf-MEK-ERK, which is already active in mutant samples.

5.1.3 Response to TLR stimulation is dependent on IGHV status, trisomy 12 and mutations in DNA Damage Response genes

The multivariate analysis highlighted that TLR stimulation by CpG ODN and Re-siquimod both showed the largest number of predictors, reflecting the multiple layers of biology involved here. These included del(11q) and *ATM*, del(17p) and *TP53*, IGHV status, trisomy 12 and *SF3B1* (Bruch & Giles et al. 2021). Amongst the screening co-

hort, TLR stimulation increased viability in certain samples and decreased in others. The clustering groups identified in section ?? also showed highly divergent responses to TLR, and this related to disease progression, thus warranting a more detailed look at the underlying features modulating TLR response (Figure 5.3).

IGHV status (along with Methylation Cluster, which is related to IGHV) and trisomy 12 had the strongest effect on TLR response. Chatzouli et al. (2014) have previously shown that TLR responses is dependent on IGHV status (see also, section 1.1.2. We noted trisomy 12 as an additional, novel determinant of TLR response (Bruch & Giles et al. 2021). In samples that do not have trisomy 12, TLR stimulation increases the viability of IGHV-U samples, whilst it decreases viability in IGHV-M samples, as expected. In contrast, in samples with trisomy 12, TLR stimulation increases viability regardless of IGHV status (Student's t-test, $p<0.001$ and $p=0.018$, Figure 5.4).

Chatzouli et al. (2014)'s observations point to the existence of specific types of BCR/TLR collaboration in CLL, leading to activation of pro-survival pathways, or apoptosis depending on the IGHV status of the tumour. Our results suggest that synergy between BCR and TLR may also be dependent on trisomy 12 status.

In addition to trisomy 12, I noted that TLR stimulation also increased viability in samples with mutations in the DNA damage response pathway, namely del(11q), del(17p), and *ATM* and *TP53*. Del(17p) occurs in 7% of CLL cases and is associated with loss of TP53, a tumour suppressor gene involved in DNA damage and repair (Zenz et al. 2010). Del(11q) is more common (18% of cases) and is associated with loss of ATM, another protein involved in DNA repair(Kipps et al. 2017). This suggests that there may also be cross-talk between the TLR and DNA damage response pathway, and that this affects the outcome of TLR stimulation.

The effect of TLR stimulation on CLL viability is dependent on the molecular make-up of the tumour. The nature of TLR response may also be important in prognosis, underlined by our the observation in section ?? that a subgroup of patients (C3) shows slower disease progression, and is the only group in which TLR induces apoptosis. The role of TLR

signalling in pathogenesis and prognosis may thus so far be under appreciated.

Collectively, this work represents the first large-scale attempt to profile the integrative effects of cell-extrinsic signals and cell-intrinsic features in lymphoma on a large scale. This systematic approach highlighted trisomy 12 as the most common feature to modulate responses. The rest of this chapter outlines my work to investigate the role of trisomy 12 in microenvironmental response.

5.2 Investigating Trisomy 12 as a modulator of microenvironmental response

5.2.1 Trisomy 12 is a modulator of microenvironmental response

Trisomy 12 modulated responses to IL4, TGF β , soluble CD40L + IL4 and TLR stimuli (Bruch & Giles et al. 2021, Figure 5.5). For example, the increase in viability induced by IL4 was enhanced in trisomy 12 samples, as was the decrease in viability in response to TGF β stimulation.

Trisomy 12 is commonly mutated in CLL (15% of patients) (Döhner et al. 2000). Until recently, this genetic lesion was regarded as conferring intermediate risk, though novel therapies have improved outcomes for trisomy 12 patients (Bosch and Dalla-Favera 2019). However, the functional explanation for its recurrence is incompletely understood. Previous work has shown a role for gene dosage effects in the pathogenic mechanism: Kienle et al. (2005) show that overexpression of genes including CDK4 and E2F1 leads to increased cell cycling and higher proliferative capacity. Likewise, Herbst (2020) showed that BCR signalling proteins are also upregulated, which may also contribute to the increased proliferative capacity. Despite this, outcomes in trisomy 12 CLL are more favourable, and these cases show increased susceptibility to BCR inhibition (Dietrich et al. 2017) (see also, section 1.0.4).

In the following, I outline my work to investigate the incompletely understood role of trisomy 12 in CLL.

5.2.2 STAT6, IRAK4 and SMAD3 are more highly expressed in trisomy 12 CLL

Trisomy 12 samples contain a third copy of chromosome 12: transcriptomic and proteomic profiling of CLL samples with this lesion have demonstrated that this has a major impact on gene expression and protein abundances (Abruzzo et al. 2018; Herbst 2020; MeierAbt2021?). It has also been shown that BCR signalling proteins are amongst those with increased expression in trisomy 12 (Herbst 2020), and this is thought to contribute to increased susceptibility to BCR inhibition in trisomy 12 CLL (Dietrich et al. 2017).

Guided by this observation, I began by investigating RNA and protein expression levels amongst genes involved in the TGF β , IL4 and TLR pathways, to which trisomy 12 samples respond more strongly. The aim was to determine whether proteins in these pathways are more abundant in trisomy 12 CLL, thus contributing to the enhanced response.

I ran differential expression analysis to compare CLL samples from the screen with and without trisomy 12. Next, I filtered the differentially expressed genes (adjusted p < 0.1) for those belonging to the TGF β , JAK-STAT and TLR pathways genesets, from the KEGG database. I visualised the RNA counts and protein abundances (Herbst 2020) for those genes.

Only a small proportion of the differentially expressed genes belonged to the TGF β , JAK-STAT and TLR pathways genesets. However of those that were, several genes were key downstream mediators of these pathways (Figure 5.6. Amongst TGF β signalling genes, 7 / 95 were upregulated in trisomy 12, including SMAD3. 12 / 160 IL4 signalling genes were differentially expressed, including STAT2. Only 2 / 116 genes in the TLR geneset were differentially expressed, though this included IRAK4 on chromosome 12. Amongst these key mediators SMAD3, STAT2, and IRAK4, all showed higher protein abundance in addition to increased RNA expression. Notably, STAT6, the key downstream mediator of IL4 signalling, was not differentially expressed, but also showed

differential protein abundance. Interestingly, many of the over-expressed genes are not located on chromosome 12, indicating the extent to which the differential dosage of this chromosome has on the expression of the entire genome.

Next I looked for further evidence to directly link the higher gene dosage of cytokine signalling genes, to enhanced responses to cytokine signals.

5.2.3 Classification analysis identifies trisomy 12 phenocopies that show increased expression of IRAK4 and SMAD3

I next investigated whether any non-trisomy 12 samples might display a trisomy 12-like phenotype (referred to as a phenocopy), in that they respond in a similar way to the panel of stimuli. The aim was to identify trisomy 12 phenocopies and to isolate the feature of these samples that might explain the underlying cause of enhanced response to external signals in trisomy 12 CLL. In particular, I was interested to see if these phenocopies showed higher expression, or even gene amplification, of the signalling genes identified in section 5.2.2.

To identify trisomy 12 phenocopies, I began by generating a classifier to predict the trisomy 12 status of a sample from its stimulus response matrix. I aimed to find non-trisomy 12 samples that were consistently misclassified as trisomy 12.

The classifier was built using binomial regression, with lasso penalisation, as implemented in the R package `glmnet` (Friedman et al. 2021). The feature matrix consisted of z scores of the viability values after treatment with each stimulus, and was used to predict the response (trisomy 12 status). I ran the model for 50 bootstrapped repeats, using three-fold cross-validation and mean absolute error as loss. Resiquimod, sCD40L+IL4 and TGF β were selected as coefficients that predict trisomy 12 status (Figure ??, as would be expected based on the observations in section 5.2.1).

I then predicted trisomy 12 status with the same viability matrix, using each of the 50 bootstrapped. I compared the results of the classification with the true trisomy 12 status. Two patient samples were misclassified as trisomy 12 in more than 50% of repeats; I refer

to these as patient sample A and B in the below. Patient sample A showed the lowest viability with TGF β (Figure 5.8, and Patient sample B showed the highest viability with IL4, and the second highest with Resiquimod.

I next investigated whether either of these non-trisomy 12 samples contained any regional amplifications on chromosome 12, which may help to isolate the gene(s) causing the tumour to respond more strongly to these signals. The whole exome sequencing for both patient samples indicated that Patient sample A had several amplified regions at 12p13.31 (42 copies), 12q24.13 (10 copies) and 12q24.33 (21 copies). An examination of the genes in these regions indicated that none of the signalling genes identified in 5.2.2 could be found in these regions, and no clear candidate gene(s) emerged that may be involved in responses to external signals.

In the absence of gene amplification, I next checked patient samples A and B might show higher expression of the signalling genes identified in 5.2.2 by visualising the RNA expression levels for SMAD3, IRAK4 and STAT6 for these patients (Figure 5.9. Patient sample A, which responded most strongly to Resiquimod, showed high levels of IRAK4 compared to the other non-triosmy 12 samples. Patient sample B, which responded strongly to TGF β showed the highest level of SMAD3 expression amongst the non-trisomy 12 samples. Thus, whilst neither patient sample contained an amplicon of SMAD3 or IRAK4, its possible that increased expression of these proteins enables a stronger response to the corresponding pathway.

These results collectively suggest that increased gene dosage of key genes in the IL4, TLR and TGF β pathways may underlie the increased response of trisomy 12 samples to these pathways. This is also reflected in the patient samples A and B, which responded more strongly to the TGF β and TLR pathways, respectively, and correspondingly higher levels of SMAD3 and STAT6 proteins. However, computational techniques can only go so far in providing biological proof, and more work is needed here to confirm this finding. In addition to gene dosage effects, I also investigated differences in transcription factor activity in trisomy 12. The results of this warranted further follow up, and this forms

the focus of rest of this chapter.

5.2.4 Spi-B and PU.1 TFs show higher activity in trisomy 12 CLL

Trisomy 12 has been well-studied at both the transcriptomic and proteomic level, and yet the cause of its recurrence in CLL is not fully understood. Thus, I next decided to investigate the impact of trisomy 12 on the epigenetic landscape of CLL, which is less well studied. In particular, I wanted to investigate differential transcription factor activity in CLL, which would give an indicator of which pathways are differentially active.

We acquired two independent ATAC sequencing datasets. The first, generated and processed in our lab, consisted of two WT and two trisomy 12 samples (Bruch & Giles et al. 2021). The second, taken from ([Rendeiro2017?](#)) and processed in our lab, comprised 43 WT and nine trisomy 12 samples.

In the external dataset, trisomy 12 status was not annotated. To do this, we used the ATACseq reads to call trisomy 12 in samples that contained > 1.4 times more reads per peak (i.e. genomic region) on average in chromosome 12, compared to peaks on other chromosomes.

Next, we used the R package `diffTF` ([R-diffTF?](#)) to identify TFs that showed differential binding site accessibility between the WT and trisomy 12 samples (Figure 5.10. `diffTF` enables comparison of TF activity between two conditions, using chromatin accessibility data, see section 1.3.3. We provided a list of 636 TFs, from the HOCOMOCO v10 database (Kulakovskiy et al. 2016) to measure their change in actiitvy between trisomy 12 and WT samples (Bruch & Giles et al. 2021).

Running this analysis on both ATACseq datasets indicated that the binding sites of nine TFs were more accessible ($p<0.05$) in the trisomy 12 samples of the larger, external dataset (Figure 5.11. In the smaller in-house dataset, the binding sites of 92 TFs were likewise more accessible (Supp. Figure 8.1). RNA and proteomics data show different abundances of transcripts and proteins in trisomy 12 CLL: ATACseq data here shows that this also corresponds to a specific signalling signature in trisomy 12 CLL.

We confirmed that the results were not affected by the additional copy of chromosome 12. Rerunning diffTF analysis without the ATACseq reads from chromosome 12 had minor impact on the significant TFs.

In both datasets, the TFs with the largest increase in activity in trisomy 12 were Spi-B and/or PU.1 (Bruch & Giles et al. 2021). Both TFs share similar binding motifs and exhibit functional redundancy (**GarrettSinha2001?**), which make it difficult to distinguish from ATACseq data alone whether either or both are more active.

Spi-B and PU.1 are haematopoietic regulators that are known to be key regulators of healthy B-cell function (Turkistany and Dekoter 2011)[READ MORE], controlling B-cell responses to environmental cues including CD40L, TLR ligands and IL4 (Willis et al. 2017).

Spi-B and PU.1 appeared to be upregulated in trisomy 12 CLL, and evidence in the literature indicated that these may regulate environmental sensing genes, providing a link between trisomy 12, and enhanced responses to external signals.

To provide further evidence of this, I next aimed to profile the downstream effects of Spi-B and PU.1 in lymphoma.

5.2.5 Spi-B and PU.1 targets are enriched for immune signalling pathways

I hypothesised that Spi-B and PU.1 might coordinate transcriptional response to external signals, thus modulating CLL proliferation in response to the microenvironment. To identify Spi-B and PU.1 target genes specifically in lymphoma, I acquired a ChIPseq dataset (Care et al. 2014) containing data on Spi-B and PU.1 binding in lymphoma cell lines. I used this dataset to test for functional enrichment of immune signalling pathways amongst the TF targets.

To define TF targets, I took the closest gene to each significant ChIP peak (q value <0.05), and within $\pm kb$ of TSS. I then tested for over-representation of these TF targets amongst

selected KEGG (**KEGG?**) and Reactome (**Reactome?**) genesets, using the ‘R’ package **clusterProfiler** (Yu 2021b). This method corresponded to a one-sided version of Fisher’s exact test. This analysis showed TLR, BCR and TGF β signalling genes to be enriched ($p<0.01$) amongst Spi-B and PU.1 targets (Figure 5.12).

5.2.6 Double knockdown of Spi-B and PU.1 reduces proliferation of trisomy 12+ cell lines

To establish the functional impact of Spi-B and PU.1 inhibition in trisomy 12+ lymphoma, we tested the impact of inhibiting these TFs on proliferation of lymphoma cell lines (Figure 5.13). We generated single and double shRNA knockdowns in lymphoma cell lines, namely SU-DHL4 and SU-DHL5 (trisomy 12) and SU-DHL2 (no trisomy 12) and then measured cell counts at 24 hour intervals (Bruch & Giles et al. 2021).

The single knockdowns had a small impact on proliferation: Spi-B inhibition reduced proliferation in SU-DHL5, as did PU.1 to a lesser extent. Double knockdown of both TFs markedly reduced proliferation in SU-DHL2 and SU-DHL4, and was lethal in SU-DHL5. This result suggested that both these TFs play an important role in the proliferative capacity of the tumour cells, and that there is functional redundancy between Spi-B and PU.1 in this context.

Collectively, these results demonstrate that trisomy 12 modulates responses to microenvironmental signals. Trisomy 12 appears to increase Spi-B and PU.1 activity, which regulate genes relating to environmental sensing, and reduce proliferation of cell lines when inhibited.

5.3 Summary

In this chapter, we perform a systematic survey of genetic determinants of microenvironmental response, leading to two key findings. Firstly, TLR signalling has a range of effects on CLL viability, and this is determined by many genetic features, including IGHV status, trisomy 12 and mutation in the DNA Damage Response pathway. Sec-

ondly, we identify trisomy 12 as a modulator of microenvironmental response, and show that higher activity of Spi-B and PU.1 may mediate this effect.

5.4 Discussion

Some do affect, some dont at all Notably, del(13q) and del(17p), which like trisomy 12 and del(11q) are the most common aberrations in CLL and act as prognostic markers (Döhner et al. 2000), had no impact on the responses to the panel of stimuli. ie some are really affecting signalling, others are not

TLR and BCR Chatzouli et al. (2014) show that stimulation of TLR in IGHV-U samples leads to pronounced p-ERK induction, whereas in M-CLL, stimulation of TLR concomitantly with BCR stimulation is required to induce a smaller increase in p-ERK. TLR stimulation without BCR stimulation in IGHV-M samples leads to an induction of caspase-8 and apoptosis, much more so than in IGHV-U samples(Chatzouli et al. 2014).

TLR as a prognostic tool?<- does this integrate multiple risk markers? Relevance in lymph node??

Paper whosng TLR response and survival that i found ages ago

Read up on trisomy 12 Dietrich et al. (2017).

Whats known about pathogenetic mechanisms: DOI: 10.1200/JCO.2005.02.568 Journal of Clinical Oncology 23, no. 16 (June 01, 2005) 3780-3792.

whole chromosom - is it lots of things - IRAK4, STAT6, gene dosage, SPIB and PU1 activity, BCR signalling?

gene dosage effects are obviously quite important here

the power of bioinformatics techniques , as well as traditional techniques, in biological discovery, to save time and resource

Read up on SpiB and PU1 SPib an dpU 1 regulate BTK: <https://pubmed.ncbi.nlm.nih.gov/8934542/> SPIB and PU1 activity in B cells: <https://pubmed.ncbi.nlm.nih.gov/15936902/> (these data indicate that Spi-1 and Spi-B directly regulate the expression of Grap2 and that Grap2 functions to modulate BCR signaling, but that reduced Grap2 expression is unlikely to account for the BCR signaling defects observed in Spi1+/- SpiB-/- B cells.) PU1 and SPIb and higher in Bc cells- this paper indicates this is to do with SP1 binding (which is on chromosome 12 and is higher) <https://pubmed.ncbi.nlm.nih.gov/7566969/> PU1 and PSIb bdnign in mouse lymphoma cell ine <https://pubmed.ncbi.nlm.nih.gov/25765478/> Spib and PU1 in B and T cell development <https://pubmed.ncbi.nlm.nih.gov/8691135/>

TLR signalling: REad (Chatzouli et al. 2014) TLR stimulation can lead to both activation of MAPK and increase in viability or activation of apoptosis. If mutations can affect which pathway is active, this points to additional mechanisms by which these mutations contribute to CLL drive. It also points to TLR response as a useful prognostic marker.

POwer of multivariate modellings; we need more literacy about these things discovery of multiple layers of biology rather than just using t tests

IL4 and KRAS - IL4 and ERK pathway reading

5.5 Contributions statement

The author of this thesis performed statistical inference and regression modelling outlined in this chapter. In essence, she performed the univariate and multivariate survey of genetic determinants of stimulus response. She identified trisomy 12 as a modulator of microenvironmental response, and conceptualised and performed the computational follow-up work on this finding, with support from Peter-Martin Bruch, Ivan Berest and Tina Becirovic.

This chapter integrates a number of datasets from the following sources. The screening

data was generated by Peter-Martin Bruch. The associated patient genetic data, including WES, WGS, CNV profiles, DNA methylation data and RNA was generated and processed by Dietrich et al. (2017). The proteomics dataset was generated by Herbst (2020), and kindly shared for the purpose of this thesis. The larger ATACseq dataset was generated by (**Rendeiro2017?**), and processed by Ivan Berest. The smaller ATACseq dataset was planned by the author and Peter-Martin Bruch, and generated by Peter-Martin Bruch. The data was processed by the author, with support from Ivan Berest. The Spi-B and PU.1 knockdown data was generated by Tina Bericoviv and published in Bruch & Giles et al. 2021.

In addition, the author received support on the following analysis work. Junyan Lu provided advice and support with multivariate modelling approaches. Ivan Berest provided support with running diffTF.

used I versus we and credited datasets, and my paper for figures and factrs - main analysis nad key finding, this hsould be routine Then outline in detail at the bottom

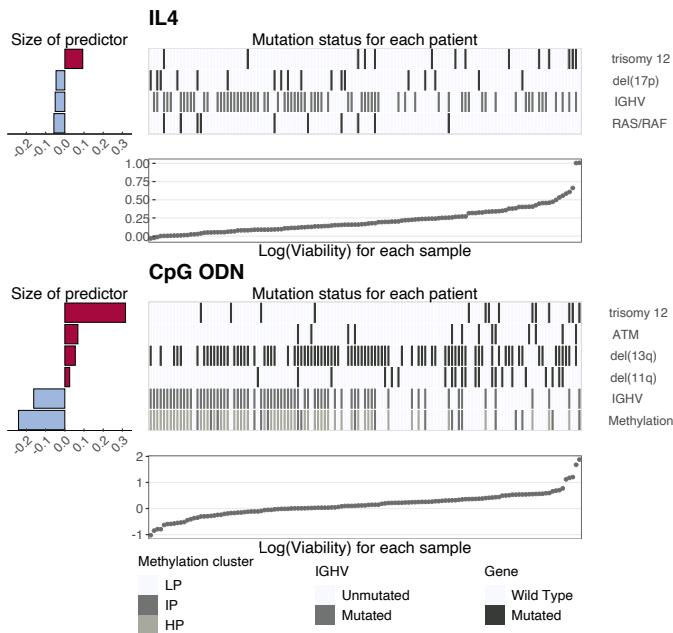


Figure 5.2: Predictor profiles for IL4 and CpG ODN depicting gene - stimulus associations. Bar plots (left) show size and sign of assigned coefficients from Gaussian linear modelling. A positive coefficient indicates that stimulated increase in viability is larger when feature is present. Scatter plots (bottom) and corresponding heatmaps above show how presence of selected genetic feature relates to sample viabilities. Scatter plots show ranked log(viability) values for each sample and heatmaps show mutation status for each predictor, for corresponding sample in scatter plot. See method 2.5.2. *Figure adapted from Bruch & Giles et al. 2021.*

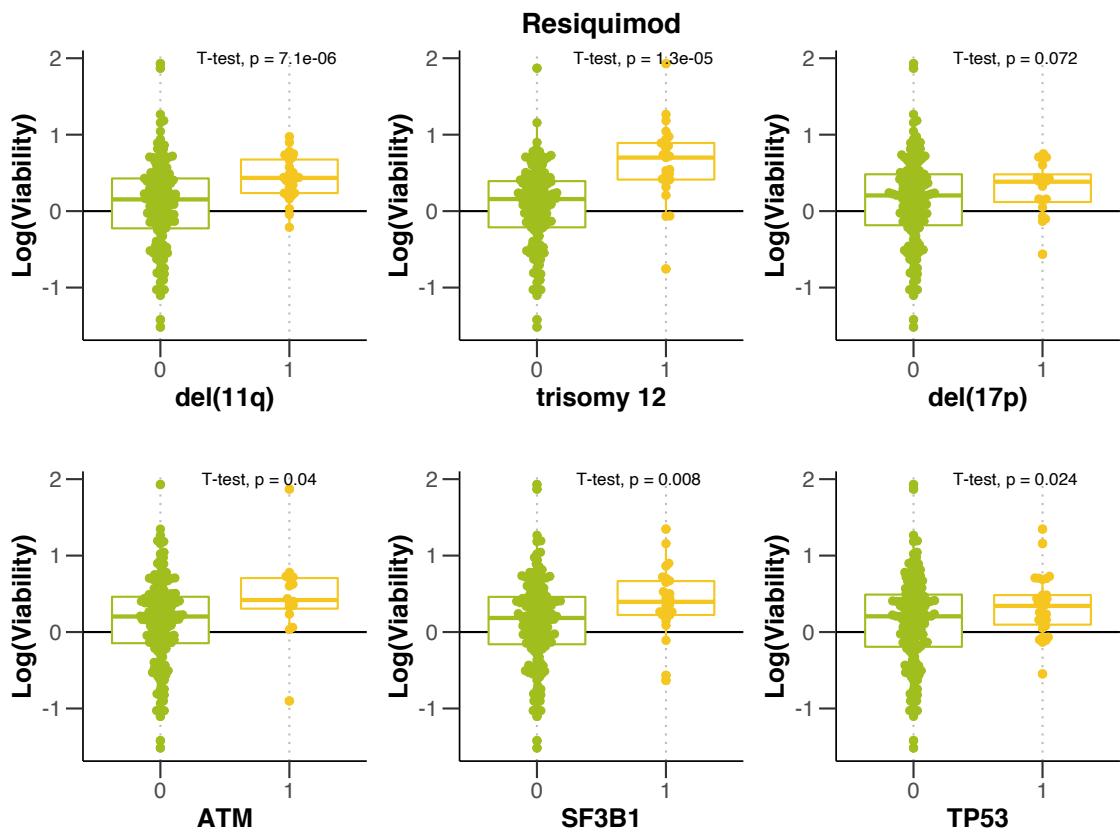


Figure 5.3: Control-normalised log transformed viability values after treatment with Resiquimod (TLR 7/8), stratified by named genetic features.

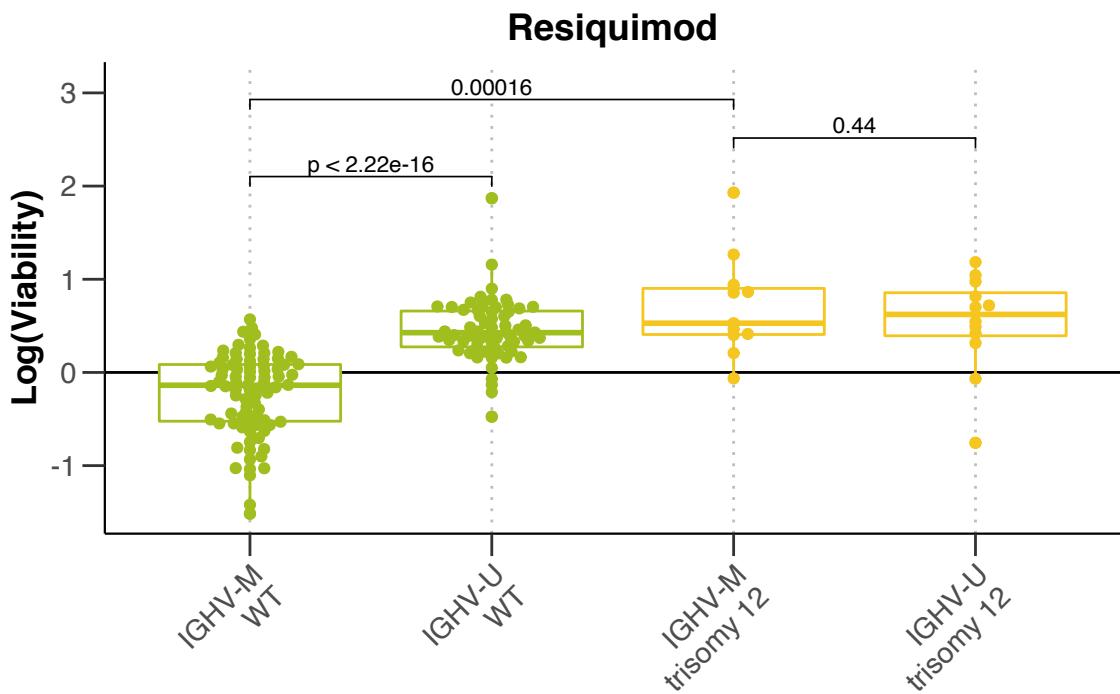


Figure 5.4: Beeswarm-boxplot showing control-normalised log transformed viability values, after treatment with Resiquimod, stratified by trisomy 12 and IGHV status. p-values from Student's t-tests. *Figure from (Bruch & Giles et al. 2021)*

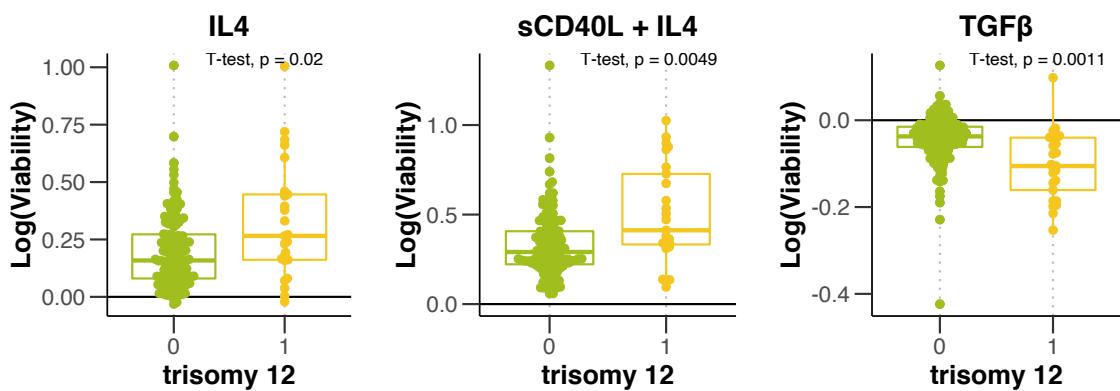


Figure 5.5: Control-normalised log transformed viability values after treatment with IL4, sCD40L + IL4 and TGF β , stratified by trisomy 12.

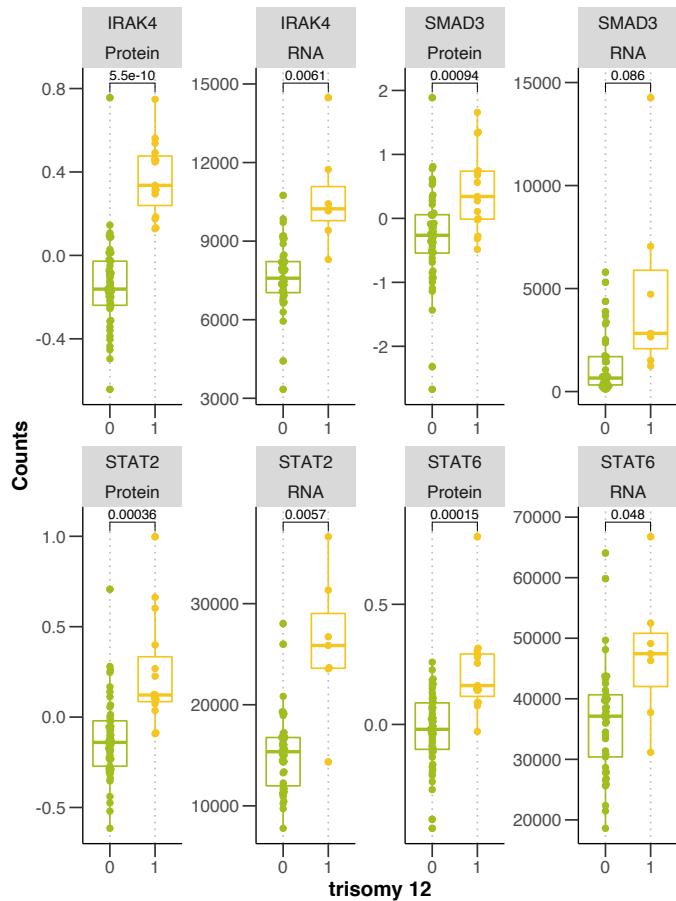


Figure 5.6: Beeswarm-boxplots showing RNA counts and protein abundances stratified by trisomy 12 status. P values from Student's t-test. _Proteomics dataset from ([HerbstThesis](#)_?) Method: 2.5.3.

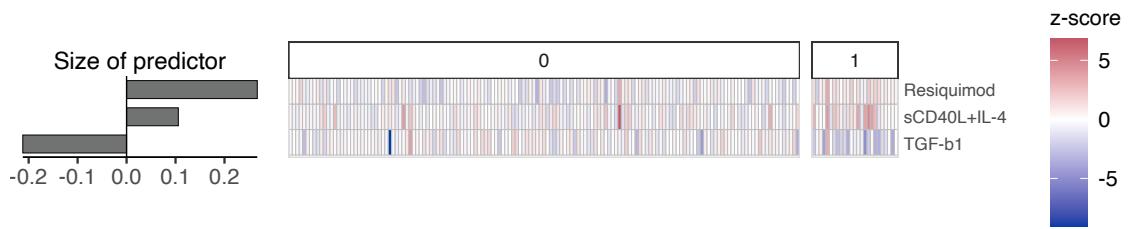


Figure 5.7: Predictor profile showing coefficients selected as predictors of trisomy 12 status. Here, binomial modelling with L1-penalty was used to identify associations between responses to stimuli and trisomy 12 status. Bar plots (left) show size and sign of assigned coefficients named on the right. A positive coefficient indicates that higher viability upon treatment named stimulus is associated with trisomy 12. Facet labels (top) and corresponding heatmap below show how the viability with each named stimulus relates to trisomy 12 status. The model shown was selected from 50 bootstrapped repeats, based on maximal AUC. Method 2.5.3.

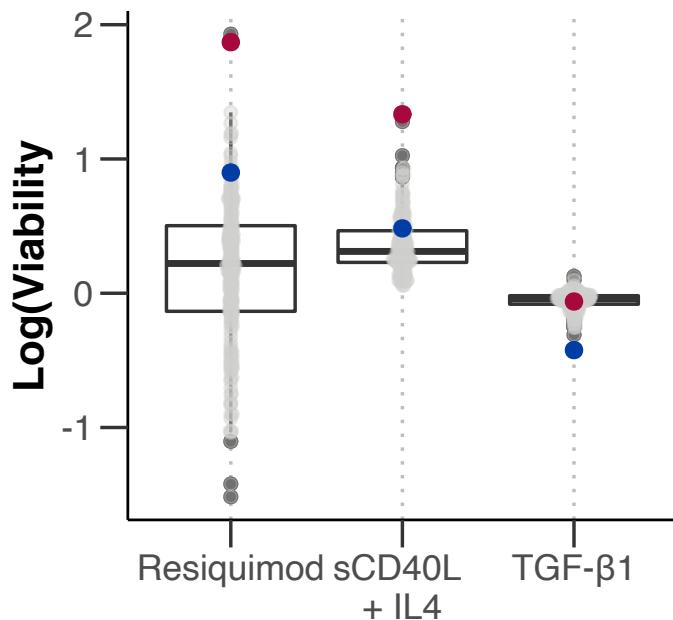


Figure 5.8: Control-normalised log-transformed viability values, for all samples after treatment with Resiquimod, sCD40L + IL4 and TGF β . Patient sample A and B are indicated in blue and red, respectively.

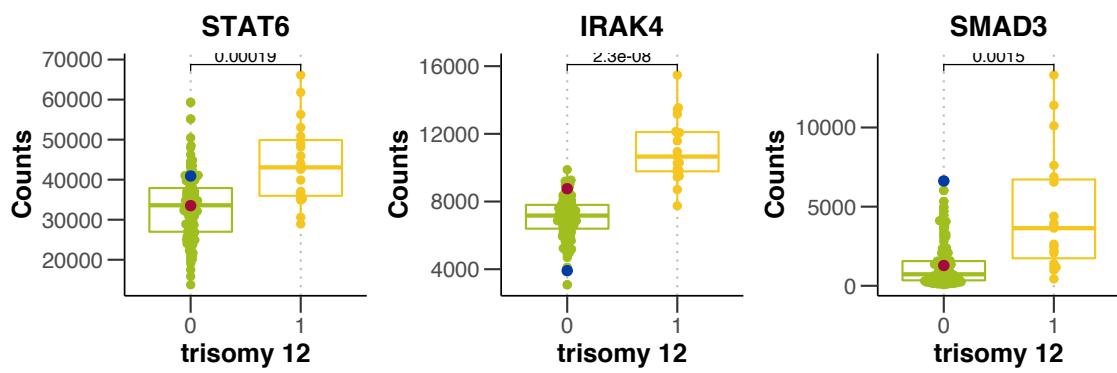


Figure 5.9: Raw RNA counts for SMAD3, IRAK4 and STAT6, stratified by trisomy 12 status. Patient sample A and B are indicated in blue and red, respectively. P values from Student's t-test.

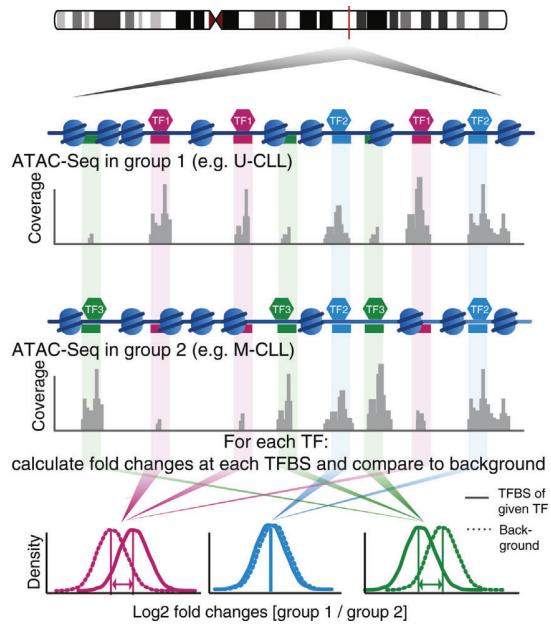


Figure 5.10: diffTF workflow from Berest et al. (2019). diffTF accepts a list of TFs along with the genomic locations of their binding sites. For each TF, the software computes the distribution of fold changes between the trisomy 12 and WT samples, using the ATACseq peaks at each TF binding site in each condition. The software compares this distribution to a background set of fold changes generated using GC content-matched loci that do not contain the same TF binding site motif. Each TF is thus assigned a weighted mean difference value, which quantifies the change in activity, and a p value.

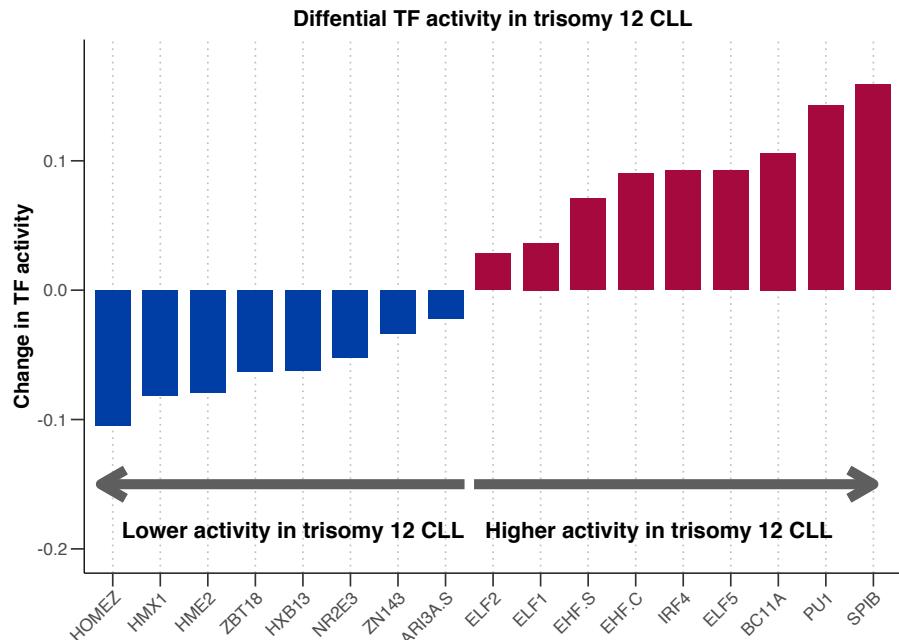


Figure 5.11: Bar plot showing the results of the diffTF analysis for the Rendeiro et al. (2016) dataset. y axis shows change in TF activity (weighted mean difference) between trisomy 12 ($n = 9$) and non-trisomy 12 samples ($n = 43$), x axis indicates TF names. 17 / 636 TFs, with BH adjusted $p < 0.05$ are shown. p values generated by diffTF in permutation mode. TF binding sites defined in HOCOMOCO v10 (Kulakovskiy et al. 2016).

Pathway	Geneset Database	Geneset Size	Spi-B targets (/3253)	SPIB p-value	PU.1 targets (/250)	PU1 p-value
BCR Signaling Pathway	KEGG	75	36	0.000	4	0.056
TLR Signaling Pathway	KEGG	102	40	0.000	5	0.046
Signaling by TGFbeta Receptor Complex	Reactome	73	26	0.007	0	1.000

Figure 5.12: Table shows results from over-representation tests of selected KEGG and Reactome pathways amongst Spi-B and PU.1 targets. Columns show geneset pathways, corresponding database, the number of genes within geneset, the number of Spi-B and PU.1 target genes within geneset (total number of target genes defined also shown), and p-value from over-representation test. TF targets defined as closest gene to each significant ChIP peak (q value < 0.05), and within $\pm \$1\text{kb}$ of TSS. ChIPseq data from Care et al. (2014). Over-representation tests run using (**clusterProfiler?**) package, method corresponds to one-sided version of Fisher's exact test. *Figure and caption from Bruch & Giles et al 2021.*

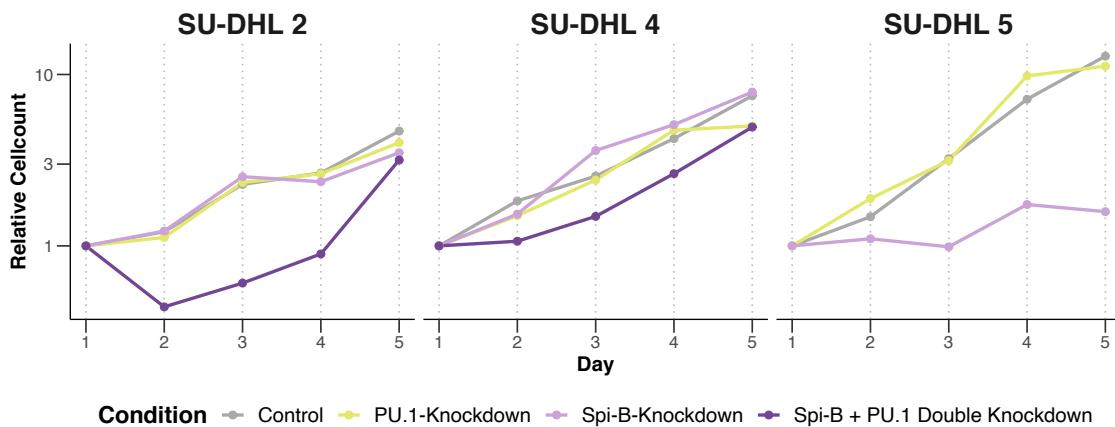


Figure 5.13: Knockdown of Spi-B and PU.1 led to growth restriction and cell death. Data shown is the mean of three technical replicates. While SU-DHL 5 was already growth impaired after Spi-B KD, SU-DHL 2 and SU-DHL 4 were only impaired after double knock-down. Double knockdown lead to rapid cell death in SU-DHL 5 (data not shown). Cell count normalised to seeded cell number of diffuse large b cell lymphoma cell lines after knockdown of Spi-B or double knockdown of Spi-B and PU.1 or shRNA as control.

Figure and caption from Bruch & Giles et al. 2021.

Chapter 6

Results

- Organize material and present results.
- Use tables, figures (but prefer visual presentation):
 - Tables and figures should supplement (and not duplicate) the text.
 - Tables and figures should be provided with legends.
 - *Figure ?? shows how to include and reference graphics. The graphic must be labelled before. Files must be in .eps format. You can do this really easily in R Markdown with knitr::include_graphics()!*
 - Figures can be referenced with `\@ref(fig:<name>)`, where `<name>` is the name of the code chunk.
- Tables and graphics may appear in the text or in the appendix, especially if there are many simulation results tabulated, but it also depends on the study and number of tables resp. figures. The key graphs and tables must appear in the text!
- R Markdown can also support math equations just like *LaTeX*!

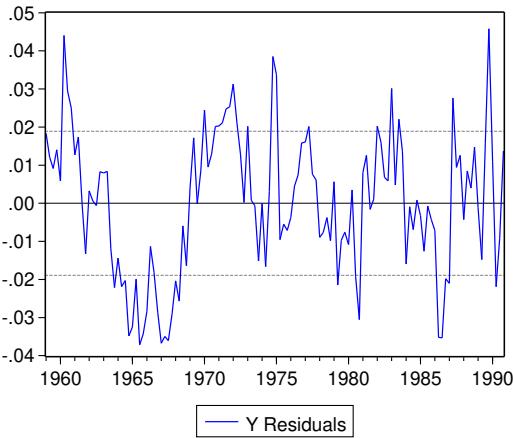


Figure 6.1: Estimated residuals from model XXX3. ...

- Equation (7.1) represents the ACs of a stationary stochastic process:

$$f_y(\lambda) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\lambda j} = (2\pi)^{-1} \left(\gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j \cos(\lambda j) \right) \quad (6.1)$$

where $i = \sqrt{-1}$ is the imaginary unit, $\lambda \in [-\pi, \pi]$ is the frequency and the γ_j are the autocovariances of y_t .

- Equations can be referenced with `\@ref(eq:<name>)`, where name is defined by adding `(\#eq:<name>)` in the line immediately before `\end{equation}`.

6.1 Review of Results

- Do the results support or do they contradict economic theory ?
- What does the reader learn from the results?
- Try to give an intuition for your results.
- Provide robustness checks.
- Compare to previous research.

Chapter 7

Results

- Organize material and present results.
- Use tables, figures (but prefer visual presentation):
 - Tables and figures should supplement (and not duplicate) the text.
 - Tables and figures should be provided with legends.
 - *Figure ?? shows how to include and reference graphics. The graphic must be labelled before. Files must be in .eps format. You can do this really easily in R Markdown with knitr::include_graphics()!*
 - Figures can be referenced with `\@ref(fig:<name>)`, where `<name>` is the name of the code chunk.
- Tables and graphics may appear in the text or in the appendix, especially if there are many simulation results tabulated, but it also depends on the study and number of tables resp. figures. The key graphs and tables must appear in the text!
- R Markdown can also support math equations just like *LaTeX*!

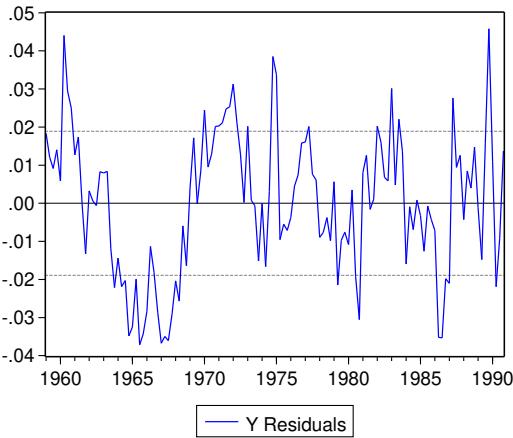


Figure 7.1: Estimated residuals from model XXX4. ...

- Equation (7.1) represents the ACs of a stationary stochastic process:

$$f_y(\lambda) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\lambda j} = (2\pi)^{-1} \left(\gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j \cos(\lambda j) \right) \quad (7.1)$$

where $i = \sqrt{-1}$ is the imaginary unit, $\lambda \in [-\pi, \pi]$ is the frequency and the γ_j are the autocovariances of y_t .

- Equations can be referenced with `\@ref(eq:<name>)`, where name is defined by adding `(\#eq:<name>)` in the line immediately before `\end{equation}`.

7.1 Review of Results

- Do the results support or do they contradict economic theory ?
- What does the reader learn from the results?
- Try to give an intuition for your results.
- Provide robustness checks.
- Compare to previous research.

Chapter 8

Discussion

References

- 10 Abruzzo, Lynne V., Carmen D. Herling, George A. Calin, Christopher Oakes, Lynn L. Barron, Haley E. Banks, Vikram Katju, Michael J. Keating, and Kevin R. Coombes. 2018. “Trisomy 12 chronic lymphocytic leukemia expresses a unique set of activated and targetable pathways.” *Haematologica* 103 (12): 2069–78. <https://doi.org/10.3324/haematol.2018.190132>.
- Akalin, Altuna, Vedran Franke, Katarzyna Wreczycka, Alexander Gosdschan, Liz Ing-Simmons, and Bozena Mika-Gospodorz. 2021. *Genomation: Summary, Annotation and Visualization of Genomic Data*. <http://bioinformatics.mdc-berlin.de/genomation/>.
- Berest, Ivan, Christian Arnold, Armando Reyes-Palomares, Giovanni Palla, Kasper Dindler Rasmussen, Holly Amelia Rebecca Giles, Peter-Martin Bruch, et al. 2019. “Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF.” *Cell Reports* 29 (10): 3147–3159.e12. <https://doi.org/10.1016/J.CELREP.2019.10.106>.
- Bosch, Francesc, and Riccardo Dalla-Favera. 2019. “Chronic lymphocytic leukaemia: from genetics to treatment.” *Nature Reviews Clinical Oncology* 16 (11): 684–701. <https://doi.org/10.1038/s41571-019-0239-8>.
- Care, Matthew A., Mario Cocco, Jon P. Laye, Nicholas Barnes, Yuanxue Huang, Ming Wang, Sharon Barrans, et al. 2014. “SPIB and BATF provide alternate determinants of T cell differentiation.” *Journal of Immunology* 193 (10): 5131–38. <https://doi.org/10.4244/JI14-0502>.

nants of IRF4 occupancy in diffuse large B-cell lymphoma linked to disease heterogeneity.” *Nucleic Acids Research* 42 (12): 7591–7610. <https://doi.org/10.1093/nar/gku451>.

Chatzouli, Maria, Stavroula Ntoufa, Nikos Papakonstantinou, Elisavet Chartomatsidou, Achilles Anagnostopoulos, Panagoula Kollia, Paolo Ghia, Marta Muzio, Kostas Stamatakopoulos, and Chrysoula Belessi. 2014. “Heterogeneous Functional Effects of Concomitant B Cell Receptor and TLR Stimulation in Chronic Lymphocytic Leukemia with Mutated versus Unmutated Ig Genes.” *The Journal of Immunology* 192 (10): 4518–24. <https://doi.org/10.4049/jimmunol.1302102>.

Collins, Russell J., Louise A. Verschuer, Brian V. Harmon, Roger L. Prentice, John H. Pope, and John F. R. Kerr. 1989. “Spontaneous programmed death (apoptosis) of B chronic lymphocytic leukaemia cells following their culture in vitro.” *British Journal of Haematology* 71 (3): 343–50. <https://doi.org/10.1111/j.1365-2141.1989.tb04290.x>.

Dietrich, Sascha, Magorzata Ole, Junyan Lu, Leopold Sellner, Simon Anders, Britta Velten, Bian Wu, et al. 2017. “Drug-perturbation-based stratification of blood cancer.” *Journal of Clinical Investigation* 128 (1): 427–45. <https://doi.org/10.1172/JCI93801>.

Döhner, H, Stephan Stilgenbauer, Axel Benner, Elke Leupolt, Alexander Kröber, Lars Bullinger, Konstanze Döhner, Martin Bentz, and Peter Lichter. 2000. “Genomic aberrations and survival in chronic lymphocytic leukemia.” *The New England Journal of Medicine* 343 (26): 1910–16. <https://doi.org/10.1056/NEJM200012283432602>.

Dolgalev, Igor. 2021. *Msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format*. <https://igordot.github.io/msigdbr/>.

Friedman, Jerome, Trevor Hastie, Rob Tibshirani, Balasubramanian Narasimhan, Kenneth Tay, and Noah Simon. 2021. *Glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models*. <https://CRAN.R-project.org/package=glmnet>.

- Herbst, Sophie. 2020. “Systematic analysis of cell-intrinsic and extrinsic factors in chronic lymphocytic leukemia to understand functional consequences for drug response and clinical outcome.” PhD diss., University of Heidelberg. <https://archiv.ub.uni-heidelberg.de/volltextserver/27438/1/Doktorarbeit.pdf>.
- Kassambara, Alboukadel, Marcin Kosinski, and Przemyslaw Biecek. 2021. *Survminer: Drawing Survival Curves Using Ggplot2*. <https://rpkgs.datanovia.com/survminer/index.html>.
- Kienle, Dirk L, Christian Korz, Beate Hosch, Axel Benner, Daniel Mertens, Annett Habermann, Alexander Kröber, et al. 2005. “Evidence for distinct pathomechanisms in genetic subgroups of chronic lymphocytic leukemia revealed by quantitative expression analysis of cell cycle, activation, and apoptosis-associated genes.” *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 23 (16): 3780–92. <https://doi.org/10.1200/JCO.2005.02.568>.
- Kipps, Thomas J, Freda K Stevenson, Catherine J Wu, Carlo M Croce, Graham Packham, William G Wierda, Susan O’Brien, John Gribben, and Kanti Rai. 2017. “Chronic lymphocytic leukaemia.” *Nature Reviews. Disease Primers* 3: 16096. <https://doi.org/10.1038/nrdp.2016.96>.
- Kolde, Raivo. 2019. *Pheatmap: Pretty Heatmaps*. <https://CRAN.R-project.org/package=pheatmap>.
- Kulakovskiy, Ivan V, Ilya E Vorontsov, Ivan S Yevshin, Anastasiia V Soboleva, Artem S Kasianov, Haitham Ashoor, Wail Ba-Alawi, et al. 2016. “HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models.” *Nucleic Acids Research* 44 (D1): D116–25. <https://doi.org/10.1093/nar/gkv1249>.
- Love, Michael, Simon Anders, and Wolfgang Huber. 2021. *DESeq2: Differential Gene Expression Analysis Based on the Negative Binomial Distribution*. <https://github.com/mikelove/DESeq2>.

Oles, Małgorzata, Sascha Dietrich, Junyan Lu, Britta Velten, Andreas Mock, Vladislav Kim, and Wolfgang Huber. 2021. *BloodCancerMultiOmics2017: "Drug-Perturbation-Based Stratification of Blood Cancer" by Dietrich s, Oles m, Lu j Et Al. - Experimental Data and Complete Analysis.*

Rendeiro, André F., Christian Schmidl, Jonathan C. Strefford, Renata Walewska, Zadie Davis, Matthias Farlik, David Oscier, and Christoph Bock. 2016. “Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks.” *Nature Communications* 7 (1): 11938. <https://doi.org/10.1038/ncomms11938>.

Therneau, Terry M. 2021. *Survival: Survival Analysis*. <https://github.com/therneau/survival>.

Turkistany, Shereen A, and Rodney P Dekoter. 2011. “The transcription factor PU.1 is a critical regulator of cellular communication in the immune system.” <https://doi.org/10.1007/s00005-011-0147-9>.

Wilkerson, Matthew D, and D Neil Hayes. 2010. “ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking.” *Bioinformatics (Oxford, England)* 26 (12): 1572–73. <https://doi.org/10.1093/bioinformatics/btq170>.

Wilkerson, Matt, and Peter Waltman. 2021. *ConsensusClusterPlus: ConsensusCluster-Plus*.

Willis, Simon N., Julie Tellier, Yang Liao, Stephanie Trezise, Amanda Light, Kristy O’Donnell, Lee Ann Garrett-Sinha, Wei Shi, David M. Tarlinton, and Stephen L. Nutt. 2017. “Environmental sensing by mature B cells is controlled by the transcription factors PU.1 and SpiB.” *Nature Communications* 8 (1): 1426. <https://doi.org/10.1038/s41467-017-01605-1>.

Yu, Guangchuang. 2021a. *ChIPseeker: ChIPseeker for ChIP Peak Annotation, Comparison, and Visualization*. <https://guangchuangyu.github.io/software/ChIPseeker>.

———. 2021b. *clusterProfiler: A Universal Enrichment Tool for Interpreting Omics Data*. <https://yulab-smu.top/biomedical-knowledge-mining-book/>.

Zenz, Thorsten, Daniel Mertens, Ralf Küppers, Hartmut Döhner, and Stephan Stilgenbauer. 2010. “From pathogenesis to treatment of chronic lymphocytic leukaemia.” *Nature Reviews Cancer* 10 (1): 37–50. <https://doi.org/10.1038/nrc2764>.

Appendix

Figures

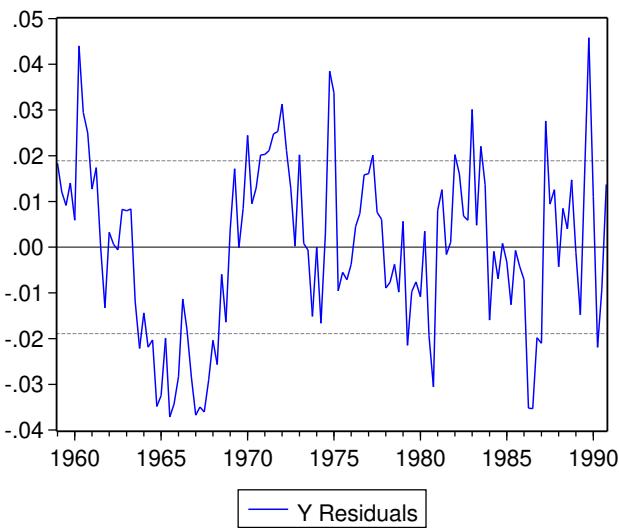


Figure 8.1: Analysis of ATACseq dataset of two trisomy 12 and two non-trisomy 12 untreated CLL PMBC samples. The volcano plot depicts change in TF activity (x axis) versus BH-adjusted p-values (y axis) for 636 TFs, comparing trisomy 12 and non-trisomy 12 samples. The **(diffTF?)**[REFERENCE] package was ran in analytical mode to calculate TF activity, measured as weighted mean difference. TFs are labeled if adjusted p-value < 0.01 and absolute weighted mean difference > 0.15.

Tables

Table 8.1: Detailed descriptive statistics of location and dispersion for 2100 observed swap rates for the period from February 15, 1999 to March 2, 2007. Swap rates measured as 3.12 (instead of 0.0312).

	3m	6m	1yr	2yr	3yr	5yr	7yr	10yr	12yr	15yr
Mean	3.138	3.191	3.307	3.544	3.756	4.093	4.354	4.621	4.741	4.878
Median	3.013	3.109	3.228	3.490	3.680	3.906	4.117	4.420	4.575	4.759
Min	1.984	1.950	1.956	2.010	2.240	2.615	2.850	3.120	3.250	3.395
Max	5.211	5.274	5.415	5.583	5.698	5.805	5.900	6.031	6.150	6.295
StD	0.915	0.919	0.935	0.910	0.876	0.825	0.803	0.776	0.768	0.762