

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: Feb 20, 2021

Drug - microenvironment - gene interplay in Chronic Lymphocytic Leukemia

Referees: Dr. Judith Zaugg
Prof. Dr. Michael Boutros

To Mum, Dad and Milly x

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 Bećirović, 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 “omic”-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	v
Abstract	vii
List of Abbreviations	xiv
List of Figures	xxiii
List of Tables	xxv
1 Introduction	1
1.1 CLL	1
1.1.1 Disease characteristics	1
1.1.2 Cell of Origin	2
1.1.3 IGHV status	3
1.1.4 The central role of BCR signalling	4
1.1.5 Recurrent genetic features in CLL	5
1.2 Therapies in CLL	11
1.2.1 Chemotherapy and chemoimmunotherapy	11
1.2.2 Targeted therapies	11
1.2.3 Management algorithm	12
1.3 The tumour microenvironment in CLL	14
1.3.1 The role of the tumour microenvironment in CLL	14
1.3.2 Components of the tumour microenvironment	14
1.3.3 Microenvironmental pathways	16

1.3.4	The influence of the microenvironment on drug response	18
1.3.5	Modelling the tumour microenvironment	19
1.4	Background to the approaches used in this thesis	21
1.4.1	<i>Ex-vivo</i> drug perturbation screens	21
1.4.2	Multiomics datasets to study CLL	22
1.4.3	Mathematical modelling	23
1.4.4	CLL as a model for studying tumour biology	25
2	Methods	27
2.1	Experimental methods:	27
2.1.1	Drug - Stimulus Combinatorial Perturbation Assay	27
2.1.2	Preparation of screening plates	27
2.2	Drug-Stimulation assay	28
2.3	WES, RNA Sequencing, Targeted Sequencing and DNA Copy number variants	29
2.3.1	Follow - up investigations	29
2.3.2	Investigation of Ibrutinib + IL4 + IBET-762	30
2.4	Data availability	32
2.5	Statistical Analysis	32
2.5.1	Data processing	32
2.5.2	Analysis of Screening Data	33
2.5.3	Ibrutinib - IBET-762 - IL-4 analysis	41
3	Data	43
3.1	Aims of the project	43
3.2	Experimental overview	44
3.2.1	High-throughput combinatorial perturbation assay	44
3.2.2	Additional external datasets	44
3.3	Processing of raw values obtained from cell viability assay	45
3.3.1	Raw values and Quality control	45
3.3.2	Data normalisation and log transformation	46
3.4	Characteristics of drugs used in the screen	46
3.4.1	Drug pathways	46
3.4.2	Drug - Drug Correlations	47
3.4.3	Drug responses	47
3.5	Characteristics of stimuli used in the screen	47
3.5.1	The panel of stimuli	47

3.5.2	Stimulus pathways	48
3.5.3	Heterogeneity of response is not caused by differences in receptor expression	48
3.6	Characteristics of patient samples used in the screen	48
3.6.1	Genetic Data available for each patient	49
3.6.2	Things form other script	49
3.7	Generation of public resource	49
3.7.1	Shiny app	49
3.7.2	Package	50
4	Ex-vivo sensitivity to microenvironmental stimulation in primary CLL cells	55
4.1	Proliferating responses to the panel of stimuli	56
4.1.1	<i>ex vivo</i> assay demonstrates 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	60
4.2.1	C1 - C4 show distinct response profiles with the panel of stimuli	60
4.2.2	The clusters show differences in disease dynamics	61
4.2.3	The clusters show differential responses to drugs <i>in vitro</i>	62
4.2.4	The clusters are enriched for different genetic features	63
4.2.5	GSEA of DE genes between subgroups	65
4.3	Summary	68
5	Genetic modulators of responses to microenvironmental stimulation	71
5.1	Systematic analysis of the effect of genetic features on responses to stimuli	72
5.1.1	Univariate analysis identifies IGHV status and trisomy 12 as key modulators of microenvironmental response	72
5.1.2	Multivariate analysis of gene - stimulus associations confirms IGHV status and trisomy 12 as key modulators of stimulus response	72
5.1.3	Response to TLR stimulation is dependent on IGHV status, trisomy 12 and mutations in DNA Damage Response genes	74
5.2	Investigating Trisomy 12 as a modulator of microenvironmental response	77
5.2.1	Trisomy 12 is a modulator of microenvironmental response	77
5.2.2	STAT6, IRAK4 and SMAD3 are more highly expressed in trisomy 12 CLL	78

5.2.3	Classification analysis identifies trisomy 12 phenocopies that show increased expression of IRAK4 and SMAD3	80
5.2.4	Spi-B and PU.1 TFs show higher activity in trisomy 12 CLL	82
5.2.5	Spi-B and PU.1 targets are enriched for immune signalling pathways	84
5.2.6	Double knockdown of Spi-B and PU.1 reduces proliferation of trisomy 12+ cell lines	85
5.3	Summary	86
5.4	Discussion	87
5.5	Contributions statement	88
6	Molecular and microenvironmental modulators of drug response in CLL	91
6.1	Mapping drug - stimulus interactions	92
6.1.1	Linear modelling quantifies interactions between drugs and stimuli	92
6.1.2	INF γ induces resistance to ibrutinib <i>in vitro</i>	95
6.2	Genetic modulators of drug response	97
6.2.1	Univariate analysis identifies genetic modulators of drug response	97
6.3	The modulatory effect of mutations on drug - stimuli interactions	99
6.3.1	Patient-specific linear modelling identifies drug - stimulus - gene interactions	99
6.3.2	Patient - specific drug - stimulus interactions may have clinical significance	101
6.4	Summary	102
6.5	Discussion	103
6.6	Contributions Statement	105
7	Results	111
7.0.1	IL4 induces resistance to BCR inhibitors and chemotherapeutics .	111
7.1	Review of Results	113
8	Discussion	115
References		117
Appendix		149
Figures		149
Tables		149

List of Abbreviations

AML	Acute myeloid leukaemia		
AKT	Protein kinase B	Ig	Immuno
ATACseq	Assay for Transposase-Accessible Chromatin using sequencing	IGHV	Immuno
ATP	Adenosine triphosphate	IL	Interleu
AUC	Area under curve	JAK	Janus
BAFF	B cell-activating factor	M-CLL	CLL w
BCR	B-cell receptor	U-CLL	CLL w
BH	Benjamini-Hochberg	MAPK	Mitoge
BMSC	Mesenchymal bone marrow stromal cell	MRD	Minima
BTK	Brutons tyrosine kinase	MSCs	Meseri
CD40L	CD40 Ligand	NF κ B	Nuclea
CDF	Cumulative Distribution Function	NLC	nursel
ChIPseq	Chromatin immunoprecipitation sequencing	NOTCH	Neuro
CLL	Chronic lymphocytic leukemia	OS	Overal
CNV	Copy number variation	PBMC	Periph
CpG ODN	CpG oligodeoxynucleotides	PCA	Princip
DMSO	Dimethyl sulfoxide	PKC	Protei
DNA	Deoxyribonucleic acid	PLC- γ 2	phosp
DNAseq	DNA sequencing	PI3K	Phosp
ERK	Ras-dependent extracellular signal-regulated kinase	RNA	Ribon
FBS	Fetal bovine serum	RNAseq	RNA s
FDC	Follicular dendritic cell	SDF1 α	Stroma
FDR	False discovery rate	SF3B1	Splicin
GSEA	Gene set enrichment analysis	STAT	Signall
IFN γ	Interferon γ		
SYK	Spleen tyrosine kinase		
TF	Transcription factor		
TGF β	Transforming growth factor β		
Th cell	T-helper cell	xiv	
TLR	Toll-like receptor		
TSS	Transcription start site		
TTFT	Time to first treatment		
TTT	Time to next treatment		
U-CLL	CLL without somatic hypermutations in IGHV loci		
VST	Variance Stabilising Transformation		
WES	Whole exome sequencing		

List of Figures

1.1	Wright–Giemsa-stained blood smear of a CLL malignant B cell, depicting typical morphology. <i>Figure from Thomas J. Kipps et al. (2017)</i>	2
1.2	Graphical depiction of the BCR pathway. <i>Figure originally published in Thomas J. Kipps et al. (2017)</i>	5
1.3	Graphical depiction of commonly mutated genes in CLL, grouped into cellular pathways (blue boxes). Minus sign indicates negative regulation. <i>_Figure originally published in (Kipps2017_?)</i>	8
1.4	Management algorithm for patients with CLL. CIT (chemoimmunotherapy), LDT (lymphocyte doubling time). <i>Figure originally published in Thomas J. Kipps et al. (2017)</i>	13
1.5	Haematoxylin and eosin stain of CLL-infiltrated lymph node tissue section, showing pale-staining proliferation centres (circled). <i>_Figure originally published in (Kipps2017_?)</i>	15
1.6	Graphic summarising soluble factors and cell-cell contacts involved in cross-talk between CLL cells and non-neoplastic cells of the tumour microenvironment <i>Figure originally published in Thomas J. Kipps et al. (2017)</i>	17
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. <i>Figure and caption from Bruch & Giles et al. 2021</i>	45

3.2	Overview of the approach for the validation experiments. Four CLL PBMC samples were incubated with ibrutinib, IL4 and IBET-762, both individually and in all combinations. Samples were taken for ATAC sequencing and RNA sequencing after 6 hours and for mass spectrometry after 24 hours. Sequencing data was processed for downstream analysis, outline in Chapter 7.	46
3.3	Boxplots of raw viability count data prior to normalisation and log transformation. For each of the 192 patient samples, there were 50 DMSO-treated wells. Dashed horizontal lines indicate potential cut-offs for samples that showed low viability with DMSO control treatment.	47
3.4	Bar plot of the drugs used in screen.	47
3.5	Raw viabilities	49
3.6	Log transformed viability values for all drugs that were included in the screen after quality control. p values from student's t test.	50
3.7	(ref:microenvironmentOverview)	51
3.8	Home page of the shiny app accompanying this project. The app was published along side Bruch et al. (2021)	52
3.9	Image of one of the tabs, in which the user can explore log-transformed viability values for different drug and stimulus treatments, stratified by genetic features.	52
3.10	(ref:package1)	53
4.1	Heatmap of Pearson correlation coefficients of each pair of stimuli, based on log transformed viability values. See Methods section 2.5.2. <i>Figure adapted from Bruch et al. (2021)</i>	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. Stimuli are grouped by corresponding pathway, and responses are stratified by IGHV status. <i>Figure adapted from Bruch et al. (2021)</i>	58

4.4	The heatmap matrix shows viability measurements for 192 samples (columns) and 17 stimuli (rows). The data are shown z-scores of log-transformed, control-normalised viability values. 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. See Methods section 2.5.2. <i>Figure from Bruch et al. (2021)</i>	59
4.5	(left) CDFs of the consensus matrices for $k = 2 - 7$, as indicated in the legend, estimated using 100 bin histogram. (right) 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 . See Methods section 2.5.2.	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. See Methods section 2.5.2.	61
4.7	Log-transformed normalised viability values stratified by cluster, for each stimulus. Stimuli activating the same pathway are grouped together. P-values from Student's t-test. <i>Figure adapted from Bruch et al. (2021)</i> .	62
4.8	(A) 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. See Methods sections ?? and 2.5.2. <i>Figure from Bruch et al. (2021)</i>	63
4.9	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.	65
4.10	Distribution of selected genetic features (rows) within each cluster for all patient samples (columns). White indicates patient sample is not annotated for that feature.	66

4.11 Enrichment and depletion of genetic features within each cluster, quantified using lasso-penalised regression. x axis shows genetic features, 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 plus and minus standard deviation. See Methods section 2.5.2. <i>Figure from Bruch & Giles et al 2021.</i>	67
4.12 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 -log10(BH-adjusted p value). Darker grey points are labelled where adjusted p < 0.01. See Methods section 2.5.2. <i>Figure from Bruch & Giles et al. 2021.</i>	68
4.13 Bar plot showing gene set enrichment to compare gene expression of samples in C3 and C4. Normalised enrichment scores (NES) are shown for top ten pathways with smallest adjusted p-value. Gene ranking performed based on Wald statistic, calculated using the DESeq2 package. See Methods section 2.5.2. <i>Figure from Bruch & Giles et al. 2021.</i>	69
4.14 Enrichment plot 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. See Methods section 2.5.2. <i>Figure and caption from Bruch & Giles et al. 2021.</i>	70
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 or is it 54??). 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. <i>Figure from Bruch & Giles et al. 2021. See Methods section 2.5.2.</i>	73

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>	74
5.3 Control-normalised log transformed viability values after treatment with Resiquimod (TLR 7/8), stratified by named genetic features.	75
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>	76
5.5 Control-normalised log transformed viability values after treatment with IL4, sCD40L + IL4 and TGF β , stratified by trisomy 12.	78
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: 2.5.2.	79
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 2.5.2.	80
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.	81
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.	82

5.10 diffTF workflow from Berest et al. (2019). diffTF accepts a list of TFs along with the genomic locations of their bindings 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.	83
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).	85
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>	86
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>	87

6.1	(A) Graphical line plots representing typical response patterns for each of the four drug - stimulus interaction categories (types I - IV). x-axis shows treatment, y-axis shows log transformed viability values, facet labels indicate interaction type. Blue and black horizontal lines demonstrate effect of interaction on viability with combinatorial treatment. Blue line shows predicted viability based on additive effects alone and black line shows viability accounting for additive effects and interaction. (B) Histogram showing number of drug-stimulus interactions within each category, where p value for β_{int} is <0.05. See Methods section 2.5.2. <i>Figure and caption from (Giles2021._?)</i>	94
6.2	Heatmap of all β_{int} values for all drug - stimulus combinations where p for $\beta_{int} < 0.05$, annotated with interaction type. Scale indicates size and sign of β_{int} . Rows and columns clustered according to hierarchical clustering. See Methods section 2.5.2. <i>Figure and caption from (Giles2021._?)</i>	95
6.3	Line plots showing examples of drug-stimulus interactions. x axis indicates treatment, y axis shows log transformed viability values, with matching samples linked across treatments. Black horizontal lines represent predicted viability for each treatment, using coefficients from linear model fit. For combinatorial treatment, blue line indicates predicted viability based on additive effect alone, and black line indicates predicted viability accounting for additive effects plus interaction. See Methods section 2.5.2. <i>Figure and caption from (Bruch et al. 2021)</i>	97
6.4	Plot showing BH-adjusted p values from Student's t-tests (two-sided, with equal variance), for all tested drug-gene associations. Tests performed for IGHV status and somatic mutations and copy number aberrations (n=54). Each coloured circle represents a gene-stimulus association meeting 10% FDR cut off. Method 2.5.2 <i>Figure from (Bruch et al. 2021)</i>	98
6.5	Beeswarm-boxplots showing control-normalised log transformed viability values after treatment with (A) Ralimetinib, (B) Nutlin-3a and (C) IBET-762, stratified by trisomy 12.	99

6.6 Heatmap summarising genetic predictors of drug - stimulus interactions. Each row depicts a single drug - stimulus combination, and each tile indicates that β_{int} for given drug and stimulus combination is modulated by corresponding genetic feature. Thus each row represents the output of a single model fit (as in Figure 6.7). Colour of tile indicates size and sign of coefficient assigned to genetic feature, where a positive coefficient corresponds to a more positive β_{int} if the feature is present. Only top 8 most commonly selected genetic features are shown. Drug - stimulus combinations with no genetic predictors of β_{int} amongst top 8 shown are omitted for clarity. See Methods section 2.5.2. <i>Figure from (Bruch et al. 2021)</i>	106
6.7 Genetic features that modulate the size of β_{int} between fludarabine and CpG ODN. Scatter plot, heatmap and bar plots indicate inputs and outputs of equation (6.2). Bar plots indicate size and sign of coefficients assigned to genetic features named on right. Scatter plot depicts β_{int} values for each patient sample i.e.response matrix. Heatmap tiles indicate mutation status for the selected genetic features (i.e. feature matrix) corresponding to same sample in scatter plot below, to show how size of β_{int} varies with each feature. See Methods section 2.5.2. <i>Figure and caption from (Giles2021._?)</i>	107
6.8 Beeswarm-boxplots of log(viability) values for fludarabine and CpG ODN single and combinatorial treatments, faceted by IGHV status and trisomy 12 status. P-values from paired Student's t-tests. <i>Figure and caption from (Giles2021._?)</i>	107
6.9 Predictor profile depicting genetic features that modulate the size of β_{int} between Nutlin-3a (MDM2) and CpG ODN. Scatter plot (bottom) and heatmap (right) and bar plots (left) indicate inputs and outputs of fitting model in Equation (6.2). Bar plots indicate size and sign of coefficients assigned to genetic features named on the right. Scatter plot depicts ranked patient-specific β_{int} values i.e. the response matrix. Heatmap above indicates patient mutation status for the selected genetic features i.e. feature matrix. Heatmap tiles correspond to points in scatter plot (i.e. patient data is aligned), to indicate how the size of β_{int} varies with selected genetic feature. See Methods section 2.5.2. <i>Figure and caption from (Giles2021._?)</i>	108

6.10 Beeswarm-boxplot showing control-normalised log transformed viability values, after treatment with CpG ODN + Nutlin-3a, stratified by del(11q) and IGHV status. p-values from Student's t-tests. <i>Figure from (Bruch et al. 2021)</i>	108
6.11 Predictor profile depicting genetic features that modulate the size of β_{int} between ibrutinib (BTK) and IL4. Plot generated as in Figure 6.7. See also Methods section 2.5.2. <i>Figure from (Giles2021._?)</i>	109
6.12 Beeswarm boxplots of log(viability) values for ibrutinib and IL4 single and combinatorial treatments, faceted by IGHV status and trisomy 12 status. P-values from paired Student's t-tests. <i>Figure and caption from (Giles2021._?)</i>	109
7.1 Examples of IL4-induced resistance. Plots show log transformed viability values with each treatment, for all samples. Matching samples are linked across treatments. Black and blue horizontal lines indicate the predicted viability from the linear model for each single and combinatorial treatment. In combinatorial treatment, both the expected viability based on the additive effect of drug and stimulus (blue), and the viability with interaction (black) are shown, to indicate the impact of the interaction. Plots show IL4 + (A) Idelalisib (Pi3K), (B) PRT062607 (SYK) (C) Everolimus (mTOR), and (D) Selumetinib (p38 MAPK). See Methods section 2.5.2. <i>Figure from Bruch & Giles et al. 2021.</i>	111
7.2 Estimated residuals from model XXX4.	112
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.	149
8.2 (ref:drugcytGeneIntAll)	150
8.3 ADD NUTLIN-3A Predictor profile	150

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).	152

Chapter 1

Introduction

1.1 CLL

1.1.1 Disease characteristics

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

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

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

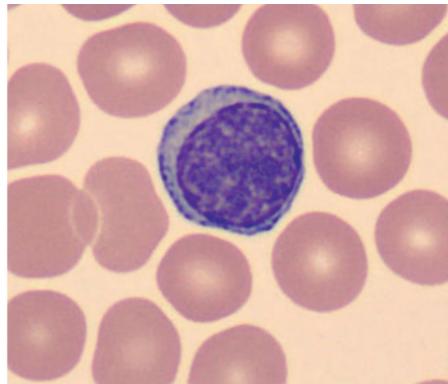


Figure 1.1: Wright–Giemsa-stained blood smear of a CLL malignant B cell, depicting typical morphology. *Figure from Thomas J. Kipps et al. (2017) .*

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

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

1.1.2 Cell of Origin

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

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

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

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

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

1.1.3 IGHV status

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

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

microenvironment, including stimulation of the BCR.

1.1.4 The central role of BCR signalling

The BCR pathway is central to the process of selection, development, proliferation and survival of CLL clones (Chiorazzi and Ferrarini 2003; Stevenson and Caligaris-Cappio 2004; Agathangelidis et al. 2012; Jan A. Burger and Chiorazzi 2013; Iacovelli et al. 2015; Dubois et al. 2020). Evidence of the importance of this pathway comes from the observation that BCRs of CLL cells show a biased selection of IGHV and IGLV κ/λ genes. This generates BCRs that are remarkably similar across patients (T. J. Kips et al. 1989; Fais et al. 1998; Widhopf et al. 2004; B. T. Messmer et al. 2004; K. Stamatopoulos et al. 2007), suggesting that BCR binding to certain antigens may drive selection and proliferation of CLL clones. Other compelling studies have shown that BCR signalling genes are upregulated in CLL cells taken from lymph nodes (**Heris-hanu2011?**) and that cells from IGHV-U CLL patients (with poorer outcomes), also show activation of BCR-related genes (Rosenwald et al. 2001). The success of BCR inhibitors in the clinic (John C. Byrd et al. 2013), also underlines the critical importance of this pathway in CLL.

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

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

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

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

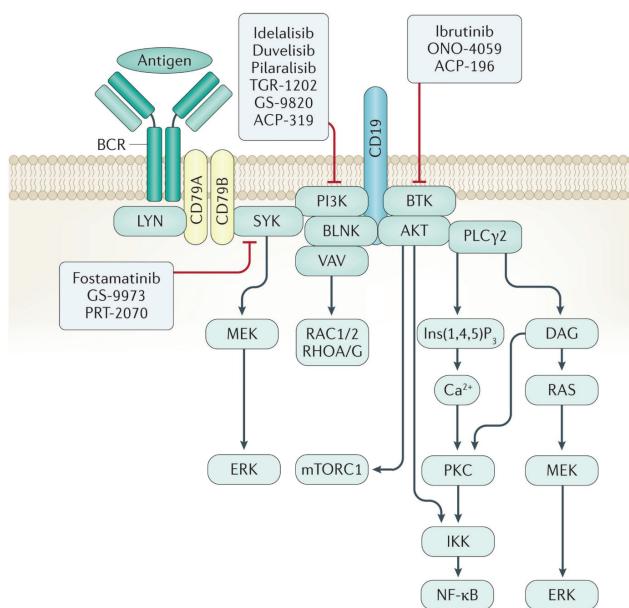


Figure 1.2: Graphical depiction of the BCR pathway. *Figure originally published in Thomas J. Kipps et al. (2017)*

1.1.5 Recurrent genetic features in CLL

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

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

Somatic Mutations

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

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

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

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

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

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

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

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

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

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

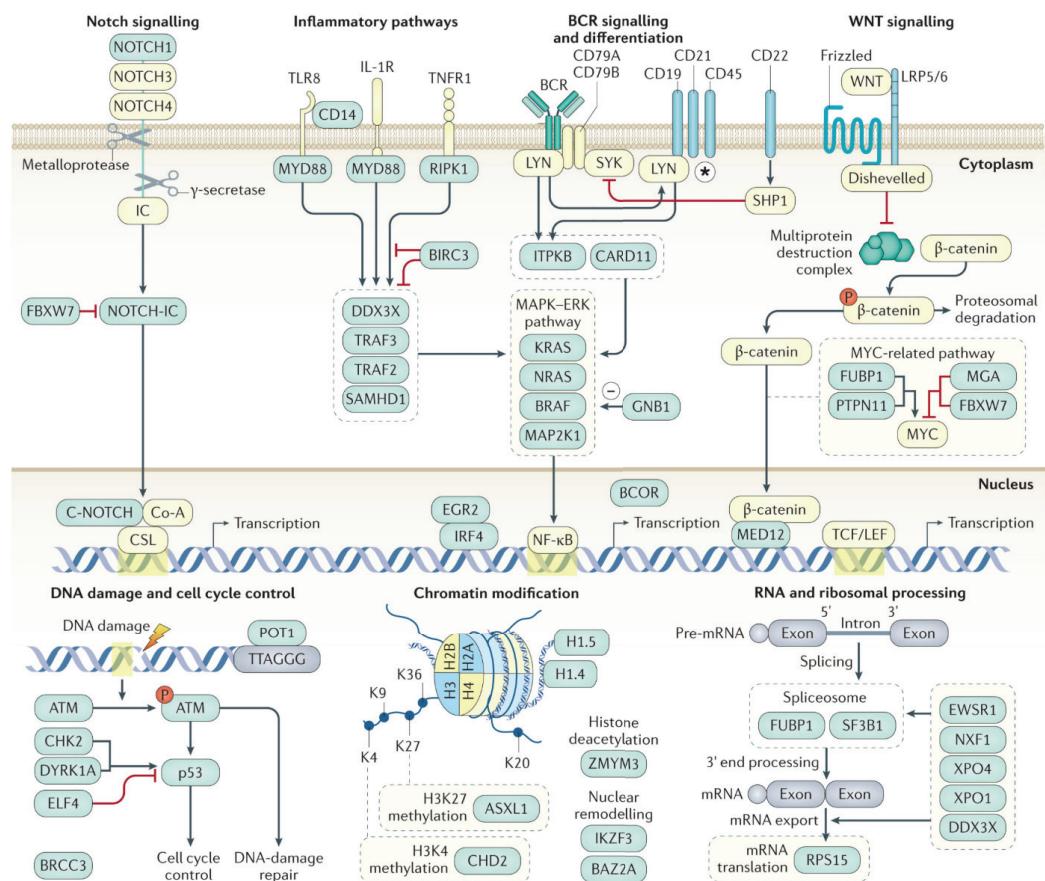


Figure 1.3: Graphical depiction of commonly mutated genes in CLL, grouped into cellular pathways (blue boxes). Minus sign indicates negative regulation. _Figure originally published in (Kipps2017_?)

Structural Aberrations

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

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

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

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

The incompletely understood role of trisomy12

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

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

Epigenetic alterations

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

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

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

1.2 Therapies in CLL

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

1.2.1 Chemotherapy and chemoimmunotherapy

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

1.2.2 Targeted therapies

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

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

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

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

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

1.2.3 Management algorithm

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

et al. 2017)).

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

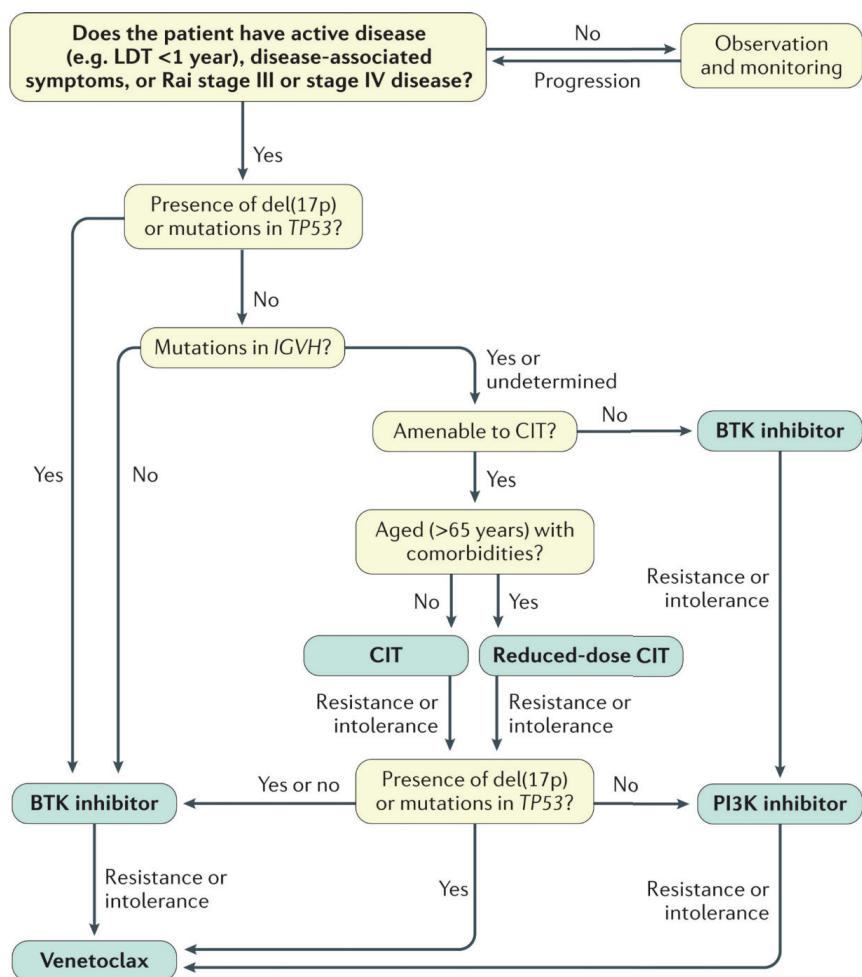


Figure 1.4: Management algorithm for patients with CLL. CIT (chemoimmunotherapy), LDT (lymphocyte doubling time). *Figure originally published in Thomas J. Kipps et al. (2017).*

1.3 The tumour microenvironment in CLL

1.3.1 The role of the tumour microenvironment in CLL

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

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

1.3.2 Components of the tumour microenvironment

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

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

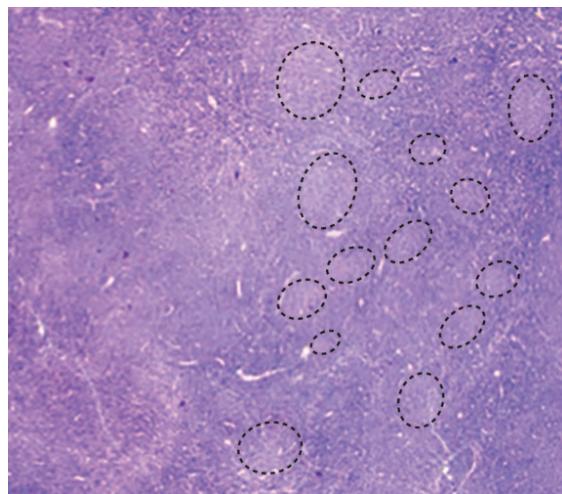


Figure 1.5: Haematoxylin and eosin stain of CLL-infiltrated lymph node tissue section, showing pale-staining proliferation centres (circled). _Figure originally published in (Kipps2017_?).

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

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

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

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

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

1.3.3 Microenvironmental pathways

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

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

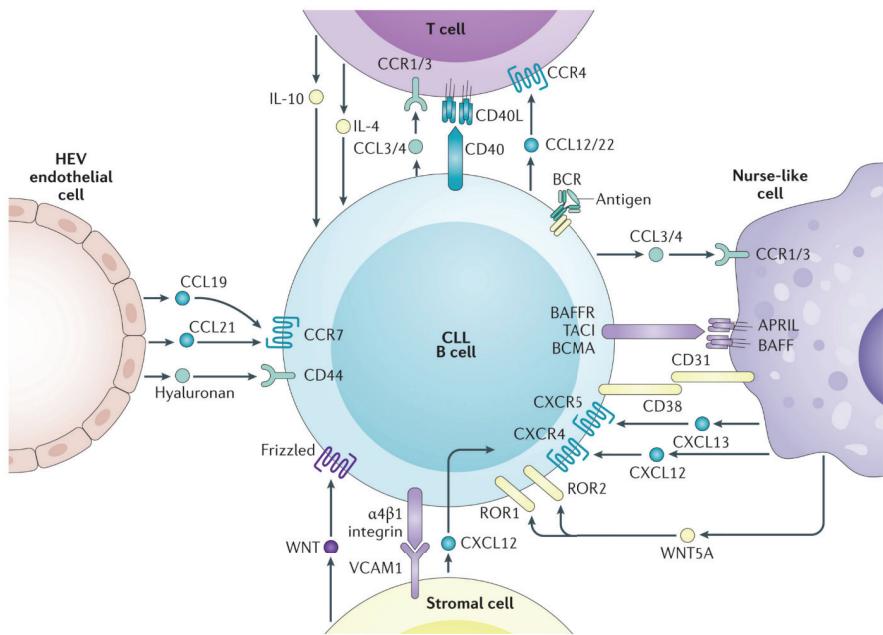


Figure 1.6: Graphic summarising soluble factors and cell-cell contacts involved in cross-talk between CLL cells and non-neoplastic cells of the tumour microenvironment *Figure originally published in Thomas J. Kipps et al. (2017)* .

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

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

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

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

1.3.4 The influence of the microenvironment on drug response

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

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

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

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

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

1.3.5 Modelling the tumour microenvironment

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

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

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

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

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

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

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

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

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

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

1.4 Background to the approaches used in this thesis

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

1.4.1 *Ex-vivo* drug perturbation screens

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

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

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

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

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

1.4.2 Multiomics datasets to study CLL

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

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

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

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

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

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

1.4.3 Mathematical modelling

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

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

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (1.1)$$

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

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

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

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

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \quad (1.2)$$

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

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon \quad (1.3)$$

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

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

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

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

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

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

1.4.4 CLL as a model for studying tumour biology

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

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

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

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

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 andneaten / 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, Med-ChemExpress 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. 5µL of drug-stimulation dilution were added into each well of the assay plates, 20 µL 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 µg/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 En-Vision 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 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.”

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), (Wilkerson and Waltman 2021), (G. Yu 2021b), (G. 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 Section 4.1.1. Pearson correlation coefficients were calculated for each drug - drug and stimulus - stimulus pair, using the `cor` functions of the (R Core Team 2021) package with log transformed viability values which were normalised to untreated controls.

Characterising stimulus responses across all samples

myself original complete (just check) Section 4.1.2. 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 (Wilkerson and Waltman 2021) package. The rows were ordered using the dendrogram order produced by `hclust` with default branch arrangement.

ADD detail on clustering

Univariate analysis of gene - stimulus and gene - drug response associations

myself written for thesis complete Section ?? & 6.2.1. 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 = 63/54). A 10% FDR cut off was used to determine significance.

Penalised multivariate regression of gene - stimulus and gene - drug associations

myself written for thesis (but some overlap with JCI) Section 5.1.2 & ??.

ADAPT FOR DRUGS ALSO 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

myself original Section 6.1.1 Linear models were fitted for each drug - stimulus combination, to extract a β_{int} term and associated p value for each combinatorial treatment. Linear model was fitted using equation (6.1), using the `lm` function of the R package stats (R Core Team 2021):

$$\log(V) = \beta_{drug}X_{drug} + \beta_{stimulus}X_{stimulus} + \beta_{int}X_{drug}X_{stimulus} + \epsilon \quad (2.1)$$

where V is the predicted viability with a given treatment, β_{drug} , $\beta_{stimulus}$ and β_{int} are coefficients for the drug, stimulus and combinatorial terms and X_{drug} and $X_{stimulus}$ are indicator variables (0 or 1) for the presence or absence of a drug/stimulus. ϵ is the vector of model residuals. Equation from Bruch & Giles et al. 2021.

To fit model, the matrix of log-transformed viability values, for control, single and combinatorial treatments was used. Interactions were defined as significant if p value for $\beta_{int} < 0.05$. The matrix of significant β_{int} was then plotted as a heatmap, with the package `pheatmap`(Kolde 2019) where the rows (stimuli) and columns (drugs) were ordered according to the dendrogram order produced by `hclust` (R Core Team 2021) using default branch arrangement.

To define the four interaction categories, drug - stimulus combinations were first divided with respect to the sign of β_{int} , whereby a positive β_{int} indicates that the viability with combinatorial treatment is higher than would be expected based on additive effects alone and vice versa. The groups were further divided into synergies and antagonisms according to the values of the model coefficients (β_{drug} , $\beta_{stimulus}$ and β_{int}). Synergisms were assigned when coefficients for single treatments (β_{drug} and $\beta_{stimulus}$) were both greater than, or both less than, the observed coefficient for the combinatorial treatment (i.e. $\beta_{drug} + \beta_{stimulus} + \beta_{int}$). For positive antagonisms, $\beta_{drug} + \beta_{stimulus} + \beta_{int}$ was less than either β_{drug} or $\beta_{stimulus}$. For negative antagonisms, $\beta_{drug} + \beta_{stimulus} + \beta_{int}$ was greater than either β_{drug} or $\beta_{stimulus}$. All drug - stimulus interactions, for which p value for $\beta_{int} < 0.05$ fit into one of these groups.

Modelling of drug - stimulus - gene interactions

myself original Section 6.3. 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.

Equation (6.2) quantifies how the viability with any combination can be predicted:

$$\log(V) = \beta_{drug}X_{drug} + \beta_{stimulus}X_{stimulus} + \beta_{patient}X_{patient} + \beta_{drug-stimulus}X_{drug}X_{stimulus} + \beta_{drug-patient}X_{drug} \\ (2.2)$$

where V is the predicted viability of a patient sample with a given treatment, β_{drug} , $\beta_{stimulus}$, β_{int} , $\beta_{drug-stimulus}$, $\beta_{drug-patient}$, $\beta_{drug-stimulus}$ and $\beta_{int}X_{drug}$ are regression coefficients for the drug, stimulus, patient sample and combinatorial terms and X_{drug} , $X_{stimulus}$ and $X_{patient}$ are indicator variables (0 or 1) for the presence or absence of a drug/stimulus/patient sample. ϵ is the vector of model residuals. See also Methods section 2.5.2. Equation from Bruch & Giles et al. 2021.

This generates a higher order interaction term $\beta_{int}X_{drug}X_{stimulus}X_{patient}$. This term represents a patient sample-specific β_{int} for each drug - stimulus combination, quantifying the size of an interaction between a drug and stimulus in each patient genetic background.

With these patient sample-specific β_{int} terms, it was possible to search for associations between the size of β_{int} and genetic features. The aim was to screen for molecular features that increased or decreased the size of a drug - stimulus interactions, using multivariate regression with L1 (lasso) regularisation.

As input to the model, the response matrix was composed of the sample - specific β_{int} values for each drug-stimulus combination. To generate the feature matrix (137 samples versus 40 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 ($p=39$) and IGHV status (coded as 0-1). 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 90% of 30 bootstrapped repeats.

^elaborate? ### Follow up ##### Lymphocyte doubling times {#LDT-method} Herbst et al., 2020c - update Section 4.2.2. “Patients which had lymphocyte counts available for less than 4 time points 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 Section 4.2.2. 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 collection 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 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) Section ???. 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 Section 4.2.5. 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` (G. 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 based 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 tirsomy 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

myeslf 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 ChIPpeak. We filtered peaks within \leq 1kb 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.

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

2.5.3 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

Plan: -headings [x] -add all figures <- -all text

proof text proof figures proof code proof numbers

Done - send

3.1 Aims of the project

```
#add label column to df
df <- df %>%
  dplyr::mutate(Cytokine = as.character(Cytokine)) %>%
  dplyr::mutate(Cytlabel = ifelse(Cytokine == "IL-2", "IL2",
                                   ifelse(Cytokine == "IL-4", "IL4",
                                         ifelse(Cytokine == "IL-6", "IL6",
                                               ifelse(Cytokine == "IL-10", "IL10",
                                                     ifelse(Cytokine == "IL-21", "IL21",
                                                       ifelse(Cytokine == "sCD40L", "sCD40L",
                                                         ifelse(Cytokine == "IL-23", "IL23",
                                                               ifelse(Cytokine == "IL-17", "IL17")))))))))
```

```

dplyr::mutate(Cytokine = factor(Cytokine))

#generate table of Cytokines and assosciated labels
Cytokine_labels <- df %>%
  filter(PatientID == "Pat_001", Drug == "DMSO") %>%
  select( Cytokine, Cytlabel)

colnames(Cytokine_labels) <- c("name", "Cytlabel")

```

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.

3.2 Experimental overview

3.2.1 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 controls.”

See whether to include this here, or just with other stuff ### Validation experiments In addition to the screen, we acquired the following validation data (Figure ??).

Ibrutinib + IL4 - IBET figure

Lymph nodes

Cell line experiments

3.2.2 Additional external datasets

Additional datasets (just describe) - CLL ATACseq data³² from the EGA (EGAD00001002110)
 ChIPseq data for Spi-B binding³⁵ from the NCBI GEO database³⁶ (GSE56857, GSM1370276)
 Proteomics data from Sophie Clinical data: LDT TTT TFS OS

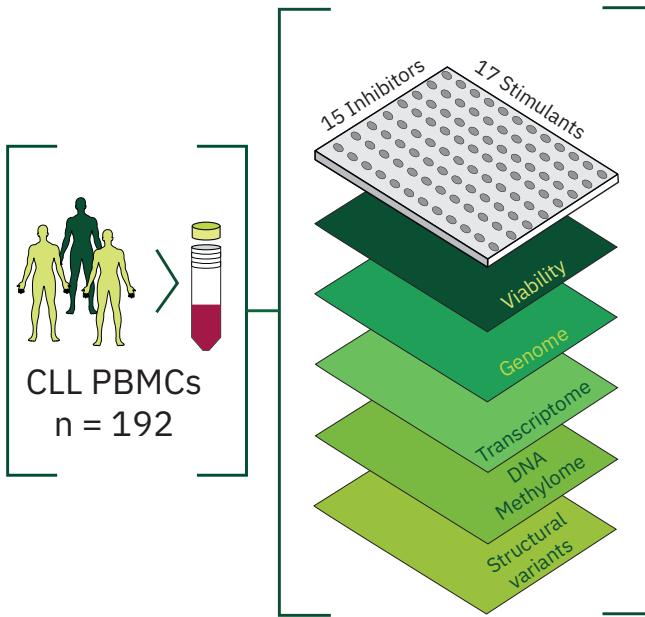


Figure 3.1: Schematic of experimental protocol. By combining 12 drugs and 17 stimuli, we systematically queried the effects of simultaneous stimulation and inhibition of critical pathways in CLL ($n = 192$). Integrating functional drug-stimulus response profiling with four additional omics layers, we identified pro-survival pathways, underlying molecular modulators of drug and microenvironment responses, and drug-stimulus interactions in CLL. *Figure and caption from Bruch & Giles et al. 2021.*

3.3 Processing of raw values obtained from cell viability assay

3.3.1 Raw values and Quality control

raw viabilities

(Figure 3.3).

Rest might not be necessary? **any QC plots**

(Figure ??).

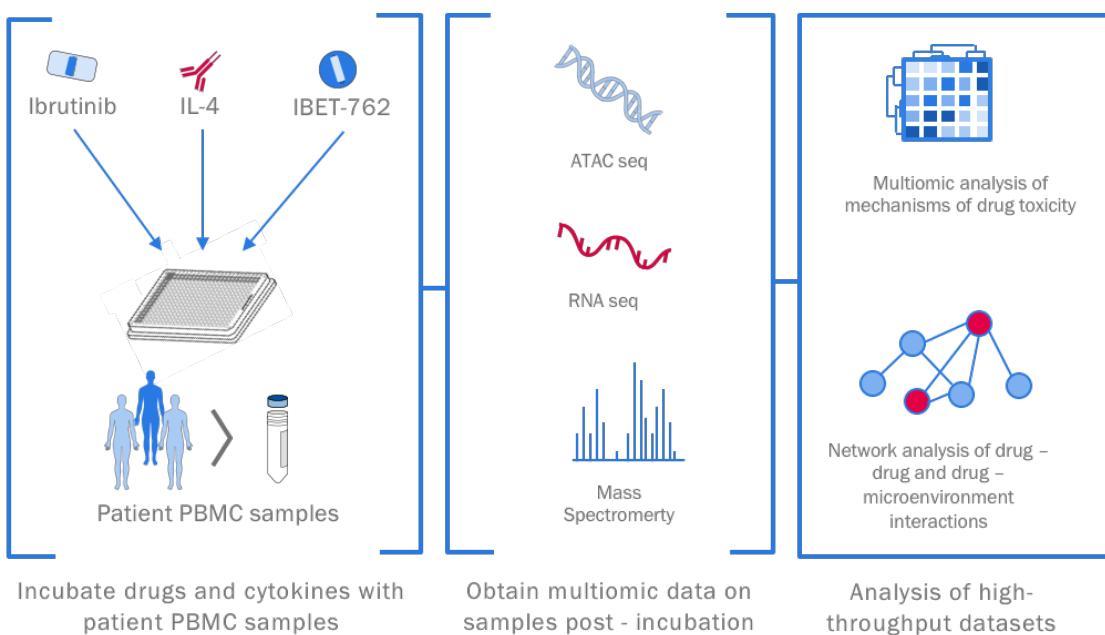


Figure 3.2: Overview of the approach for the validation experiments. Four CLL PBMC samples were incubated with ibrutinib, IL4 and IBET-762, both individually and in all combinations. Samples were taken for ATAC sequencing and RNA sequencing after 6 hours and for mass spectrometry after 24 hours. Sequencing data was processed for downstream analysis, outline in Chapter 7.

3.3.2 Data normalisation and log transformation

demo plot

(Figure ??).

3.4 Characteristics of drugs used in the screen

3.4.1 Drug pathways

We screened 17 different drugs (Figure 3.4). **Add table with details of drugs**

(Table ??).

add licencing status , otherwise this is good on the drugs

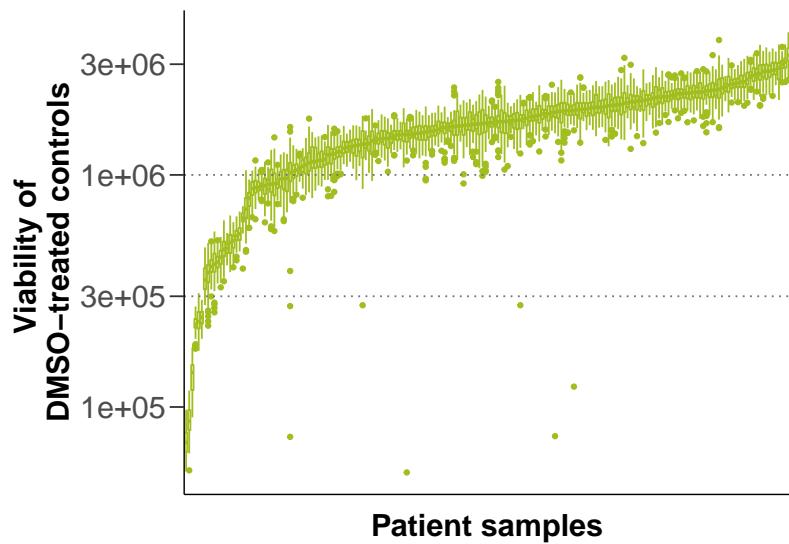


Figure 3.3: Boxplots of raw viability count data prior to normalisation and log transformation. For each of the 192 patient samples, there were 50 DMSO-treated wells. Dashed horizontal lines indicate potential cut-offs for samples that showed low viability with DMSO control treatment.

Figure 3.4: Bar plot of the drugs used in screen.

3.4.2 Drug - Drug Correlations

(sense check) **corr plot**

(Figure 3.5).

3.4.3 Drug responses

The drug responses were as follows (Figure 3.6).

3.5 Characteristics of stimuli used in the screen

3.5.1 The panel of stimuli

table of stimuli -details of how were selected

(Table ??).

	Drug	Main targets	Target category	Drug Group	
D_001	Ibrutinib	BTK	BCR	FDA approved	E
D_002	Idelalisib	PI3K delta	BCR	FDA approved	E
D_003	Fludarabine	Purine analogue	DDR	FDA approved	C
D_004	Nutlin-3a	MDM2	DDR	Clinical development/ tool compound	C
D_005	Selumetinib	MEK1/2	MAPK	Clinical development/ tool compound	C
D_006	BAY-11-7085	NFkB	NFkB	Clinical development/ tool compound	C
D_007	Everolimus	mTOR	mTOR	FDA approved	C
D_008	PRT062607	SYK	BCR	Clinical development/ tool compound	E
D_009	Pyridone-6	JAK1/2/3	JAK/STAT	Clinical development/ tool compound	C
D_010	Ralimetinib	p38 MAPK	MAPK	Clinical development/ tool compound	C
D_011	Luminespib	HSP90	HSP90	Clinical development/ tool compound	C
D_012	I-BET 762	BRD2/3/4	Epigenome	Clinical development/ tool compound	C

more on pathways? or is the diagram good enough? more on how they were chosen?
not really my PhD?

3.5.2 Stimulus pathways

details of which stimuli activates with pathways

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 Characteristics of patient samples used in the screen

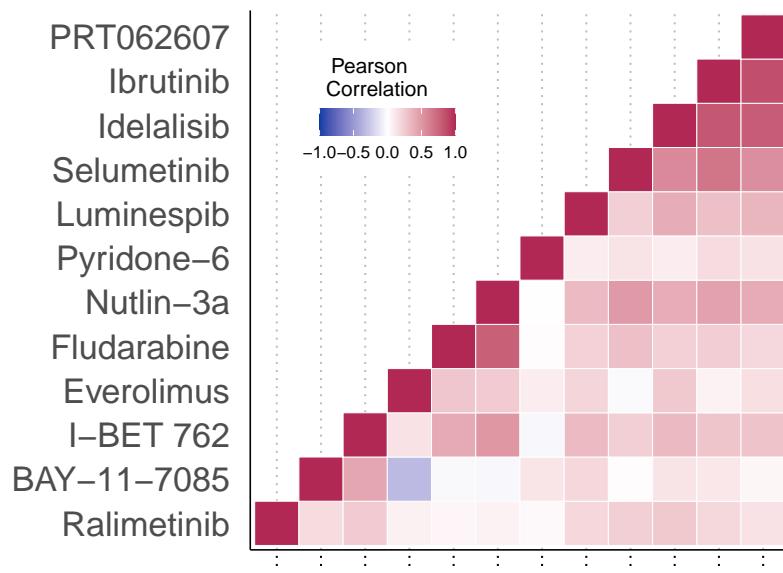


Figure 3.5: Raw viabilities

3.6.1 Genetic Data available for each patient

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

(Figure ??).

3.6.2 Things form other script

any otehrs?

(Figure ??).

3.7 Generation of public resource

3.7.1 Shiny app

Shiny app

(Figure 3.8).

(Figure 3.9).

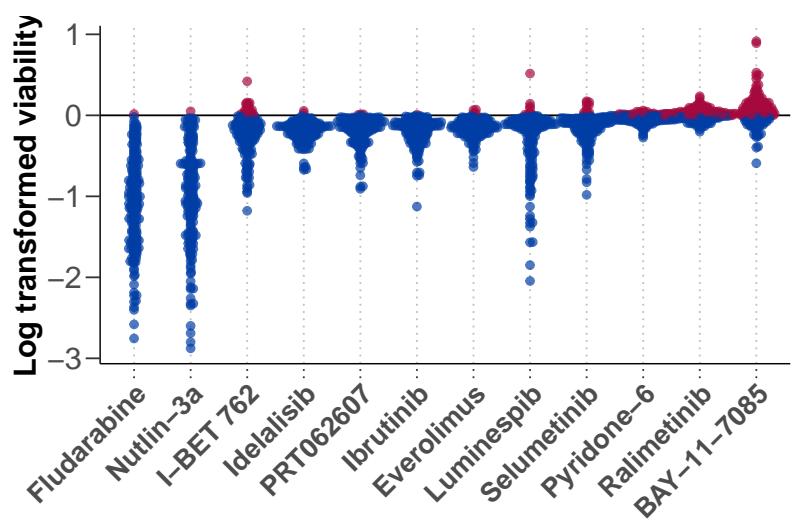


Figure 3.6: Log transformed viability values for all drugs that were included in the screen after quality control. p values from student's t test.

3.7.2 Package

package

(Figure 3.10).

(ref:package1)

Stimulus	Name	Supplier	Concentration
IL4 human recombinant animal component free	IL4	Sigma-Aldrich	10 ng/ml
IL10 human Animal component free	IL10	Sigma-Aldrich	10 ng/ml
IL2 human recombinant animal component free	IL2	Sigma-Aldrich	10 ng/ml
R-848	Resiquimod	Enzo Life Sciences	1000 ng/ml
Human IL-21	IL21	Peprotech	10 ng/ml
Human BAFF	BAFF	Peprotech	250 ng/ml
Human IL-1 beta	IL1	Peprotech	10 ng/ml
Human sCD40 Ligand	sCD40L	Peprotech	1000 ng/ml
Goat F(AB')2 Fragment to human IgM	soluble anti-IgM	MP Biomedicals	20000 ng/ml
Human TGF1	TGF1	Peprotech	10 ng/ml
Human IL15	IL15	Peprotech	10 ng/ml
Human IL6	IL6	Peprotech	10 ng/ml
ODN 2006 (ODN 7909)	CpG ODN	Invivogen	1000 ng/ml
Human SDF1 alpha (CXCL12)	SDF-1	Peprotech	200 ng/ml
Human Interferon gamma	Interferon	Peprotech	5 ng/ml
HS-5 conditioned medium	HS-5 CM	self produced	20 %

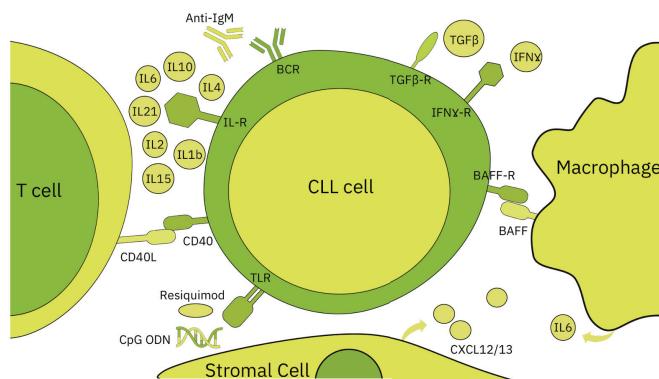


Figure 3.7: (ref:microenvironmentOverview)

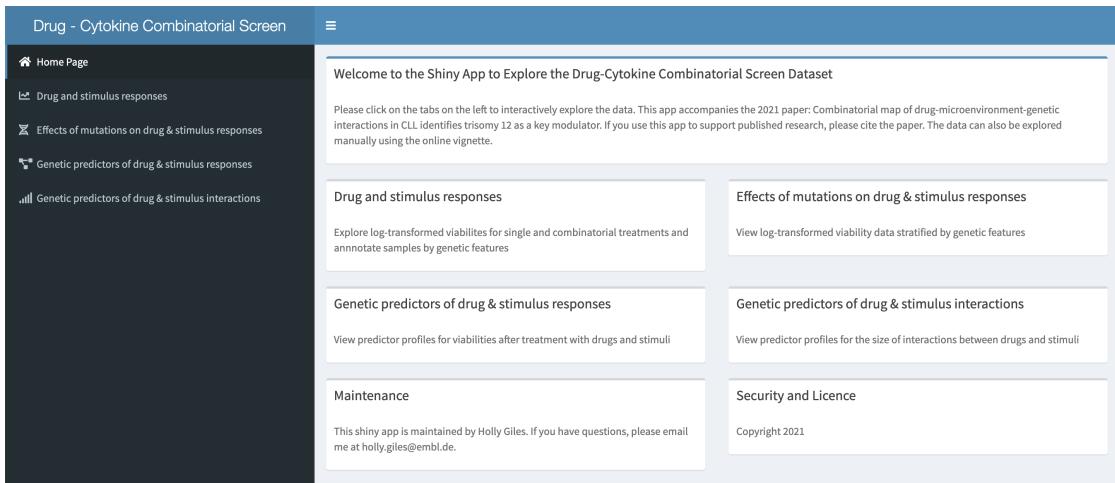


Figure 3.8: Home page of the shiny app accompanying this project. The app was published along side Bruch et al. (2021) .

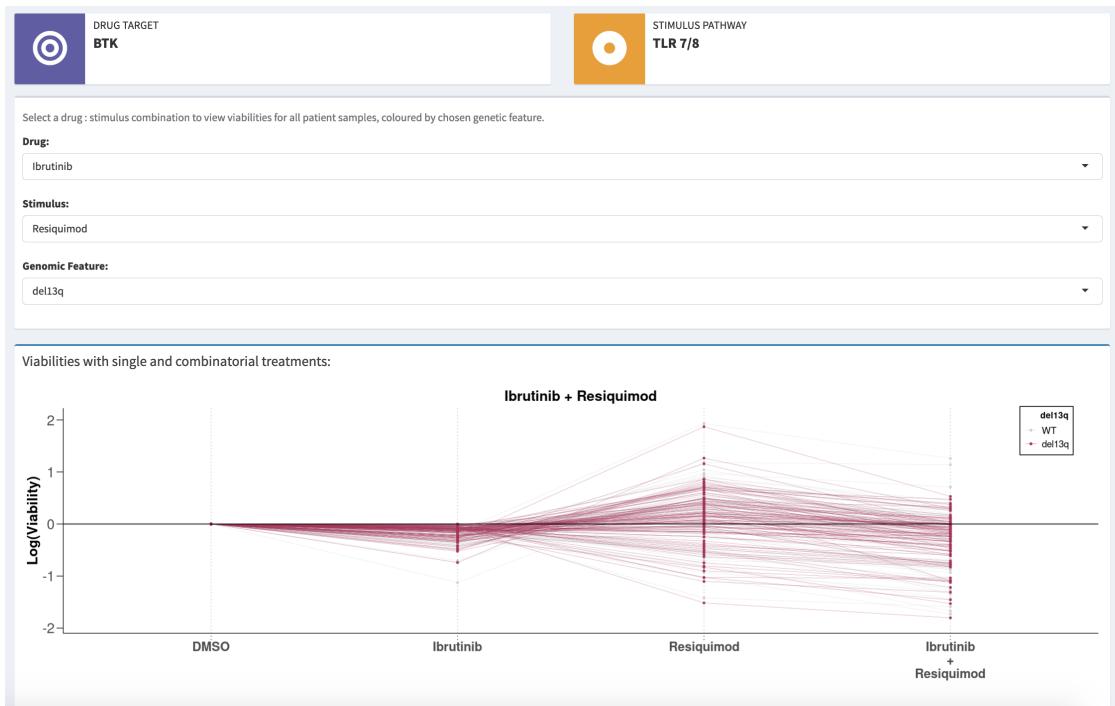


Figure 3.9: Image of one of the tabs, in which the user can explore log-transformed viability values for different drug and stimulus treatments, stratified by genetic features.

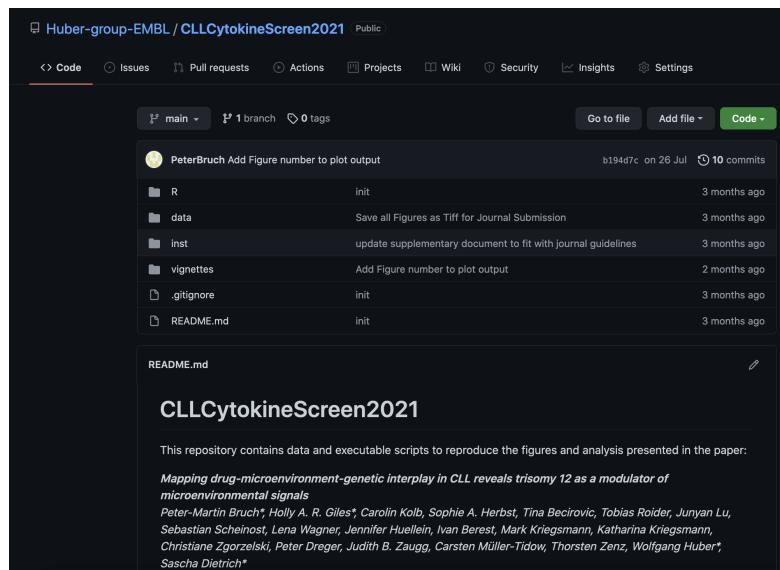


Figure 3.10: (ref:package1)

Chapter 4

Ex-vivo sensitivity to microenvironmental stimulation in primary CLL cells

Our assay quantified the effects of 17 cytokines and microenvironmental stimuli on cell viability in 192 primary CLL samples. The 17 stimuli were selected based on evidence in the literature that each stimulus had been shown to impact on CLL viability *in vitro*, aiming to minimise redundancy among the compounds (see also Section 3.5.1) (Bruch et al. 2021). Many studies have applied various methods to model the impact of microenvironmental signalling on CLL, and each method has its strengths (see also Section 1.3.5). Our assay represents a reductionist model of microenvironmental signalling, making it possible to dissect the effect of individual soluble factors within the protective niche on baseline viability.

This chapter details my analysis of the phenotypic effects of stimuli on CLL viability, leading to the identification of four patient subgroups that differ in their stimulus response profiles. This chapter also covers the clinical and molecular characterisation of these four subgroups.

The results presented in this chapter centre on viability values of CLL PMBCs samples treated with our panel of stimuli. CLL cells do not proliferate *in vitro*, but rather undergo spontaneous apoptosis (Collins et al. 1989). Treatment with various stimuli, or co-culture with NLCs or BMSCs can extend survival of CLL PBMCs *ex vivo* (Collins et al. 1989; Jan A. Burger et al. 2000; Kurtova et al. 2009; Deaglio and Malavasi 2009; Purroy

et al. 2015). To measure the individual phenotypes generated by each of our stimuli, we thus quantified viability by comparing ATP counts in treated samples, with those in DMSO wells, after 48 hours. A positive viability value indicates that the sample viability was increased relative to control. Values shown have additionally been log-transformed for the purposes of statistical analysis.

4.1 Proliferating responses to the panel of stimuli

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

To begin the analysis, I started by investigating heterogeneity amongst responses to the stimuli. I calculated Pearson correlation coefficients for each stimulus pair, using the log-transformed normalised viabilities (Bruch et al. 2021). The resulting coefficients were ordered using hierarchical clustering and visualised in a symmetrical heatmap (Figure 4.1).

In the resulting heatmap, several clusters of stimuli could be identified, including a larger group corresponding to agonists of TLR and Nf κ B pathways and a smaller group encompassing IL4 and TLR stimuli.

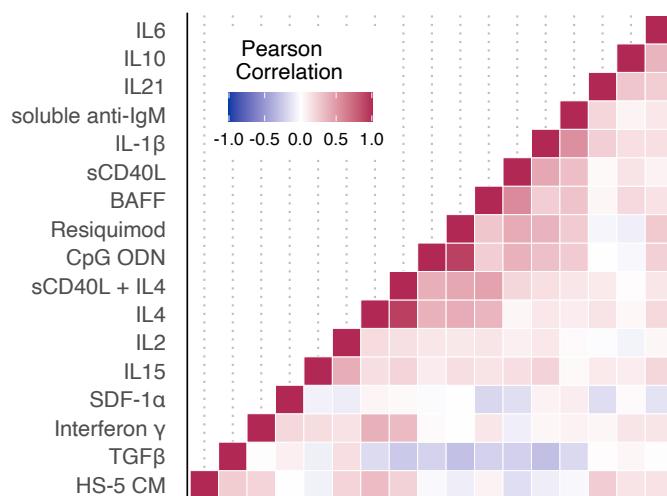


Figure 4.1: Heatmap of Pearson correlation coefficients of each pair of stimuli, based on log transformed viability values. See Methods section 2.5.2. *Figure adapted from Bruch et al. (2021)*.

However, whilst certain stimuli clustered into groups, very few stimulus pairs showed

any significant correlation. Almost all stimulus pairs showed little correlation ($R < 0.6$), including those that targeted similar downstream pathways, indicating a high degree of functional diversity amongst soluble factors in the CLL microenvironment. 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.

In contrast, correlation of 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).

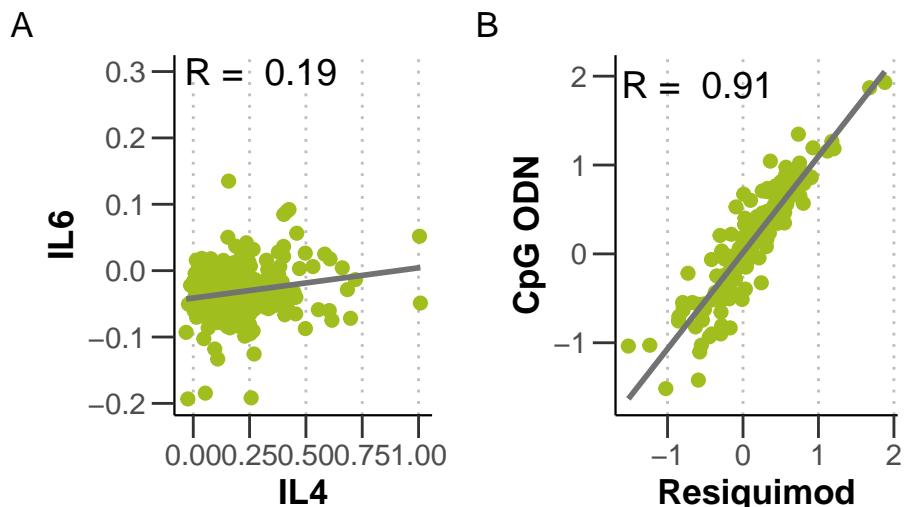


Figure 4.2: Scatter plot of log-transformed viability values, normalised to DMSO controls, for (A) treatment with JAK-STAT agonists IL4 and IL6 and (B) treatment with TLR agonists CpG ODN and Resiquimod.

Having observed that microenvironmental stimulation induced diverse phenotypes between patient samples, I next aimed to visualise a global overview of these phenotypes. I plotted log-transformed viability values normalised to DMSO controls for all patient samples and all stimuli (Figure 4.3, (Bruch et al. 2021)).

Figure 4.3 shows that the majority of the stimuli increased viability, underlining the supportive nature of the microenvironment in CLL. However, four out of 17 reduced CLL viability relative to control, namely IL6, TGF β and TLR 7/8/9 agonists CpG ODN and

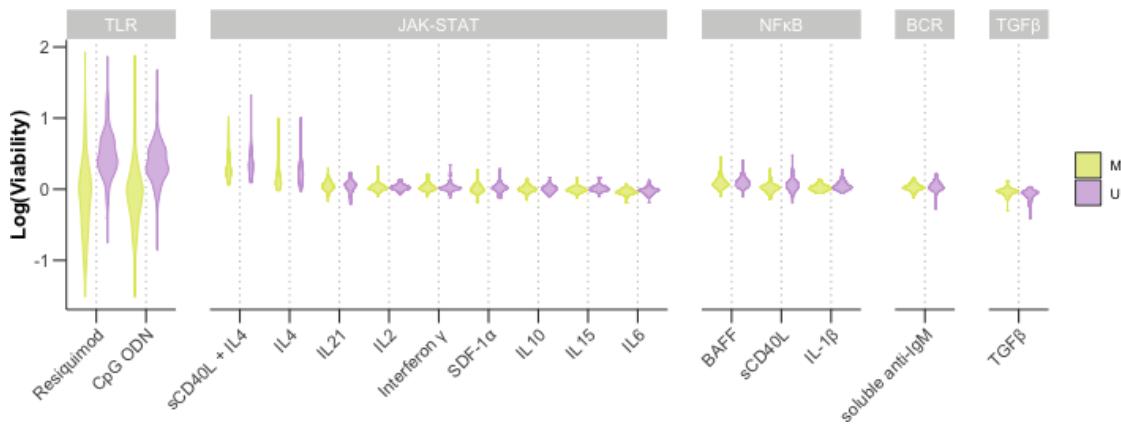


Figure 4.3: Log transformed viabilities after treatment with each stimulus. Stimuli are grouped by corresponding pathway, and responses are stratified by IGHV status. *Figure adapted from Bruch et al. (2021)*.

Resiquimod 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. Our assay identified IL4 and TLR7/8/9 as key players in CLL-microenvironment cross-talk, and thus remain central 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 next generated a heatmap of all stimuli responses across all samples, using z-scores for optimal visualisation. We performed consensus clustering on the resulting heatmap to group patients according to their response profiles (Figure 4.4).

The consensus clustering method generates robust hierarchical clustering by subsampling from the matrix of values (in this case, the viability z-scores). Using subsampling, it is possible to calculate a “consensus matrix” which indicates the proportion of times each pair of values occupy the same cluster when subsampled together. This is repeated for different numbers of clusters, denoted by k, allowing the user to select the optimal number of clusters for a given dataset.

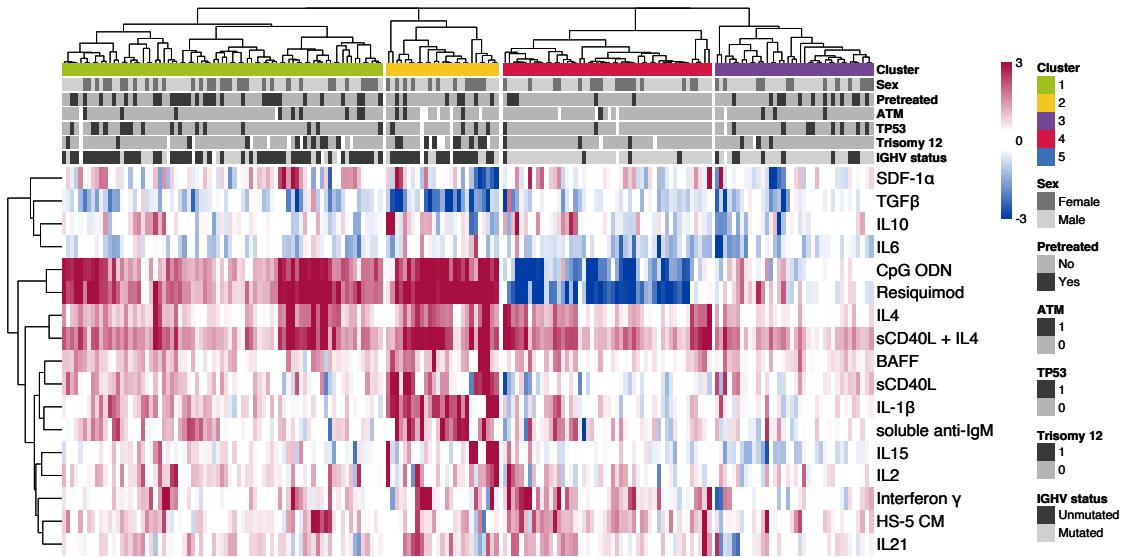


Figure 4.4: The heatmap matrix shows viability measurements for 192 samples (columns) and 17 stimuli (rows). The data are shown z-scores of log-transformed, control-normalised viability values. The colour bars to the right show sample annotations. Consensus clustering was used to define column tree layout, using hierarchical clustering with the Euclidean metric. See Methods section 2.5.2. *Figure from Bruch et al. (2021)*

I ran consensus clustering on the matrix of z-scores for different values of k , and arranged the corresponding heatmaps accordingly. Figure @ref(fig:stimuliHeatmap shows the arrangement for $k = 4$. Comparing the heatmaps, I 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.

To validate the choice of four clusters, I additionally visualised summaries of the consensus matrices for each value of k , to quantify the degree of confidence in the clusters for each k (Figure 4.5).

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).

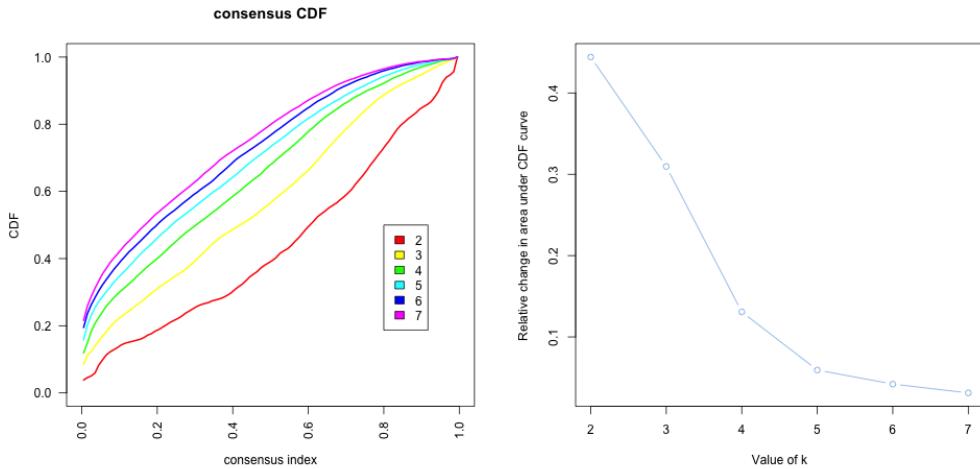


Figure 4.5: (left) CDFs of the consensus matrices for $k = 2 - 7$, as indicated in the legend, estimated using 100 bin histogram. (right) Relative change in area under the CDF curve, for $k = 2 - 7$, to compare k with $k - 1$. In the case of $k = 2$, there is no $k - 1$, so the total area is plotted. Line shows relative increase in consensus between each value of k . See Methods section 2.5.2.

4.2 Functional characterisation of patient clusters

4.2.1 C1 - C4 show distinct response profiles with the panel of stimuli

Consensus clustering identified four patient subgroups, which each responded differently to the panel of stimuli. This raised the possibility that responses to microenvironmental stimulation are linked to cell-intrinsic features, and could be prognostic. Thus, I next aimed to characterise the phenotypic and molecular differences between these four clusters.

I began by investigating the differences between the response profiles of each cluster. I refer to the four clusters as C1 to C4: C1 and C2 were enriched for IGHV-U whilst the samples in C3 and C4 were mostly IGHV-M. Amongst the IGHV-U enriched C1 and C2, both showed strong, positive responses to IL4 and TLR7/8/9 stimulation. C2 could be distinguished by stronger responses to the stimuli overall, in particular to NF κ B agonists IL1 β , anti-IgM, BAFF and sCD40L. 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 4.7 summarises

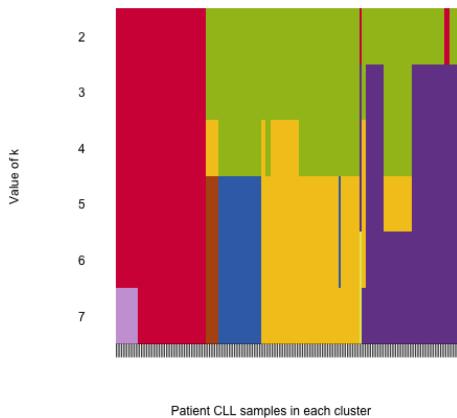


Figure 4.6: Assignment of patient samples (columns), to each cluster, for $k = 1 - 7$ (rows) to demonstrate stability of cluster membership. Cluster colour for $k = 4$ match those in heatmap in 4.4. See Methods section 2.5.2.

these findings in more detail, showing responses stratified by cluster for a subset of the stimuli.

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)¹ and time to next treatment (TTT)² 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 4.8A). 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).

¹ LDT indicates the rate at which malignant B cells accumulate, and is an independent biomarker that correlates with overall survival (OS) (Baumann et al. 2021).

² TTT reflects the duration of time between the initiation of one therapy and the start of the next treatment (Campbell et al. 2020; Delgado and Guddati 2021). This measurement accounts for the period of treatment, along with the period in which the disease and/or symptoms are controlled.

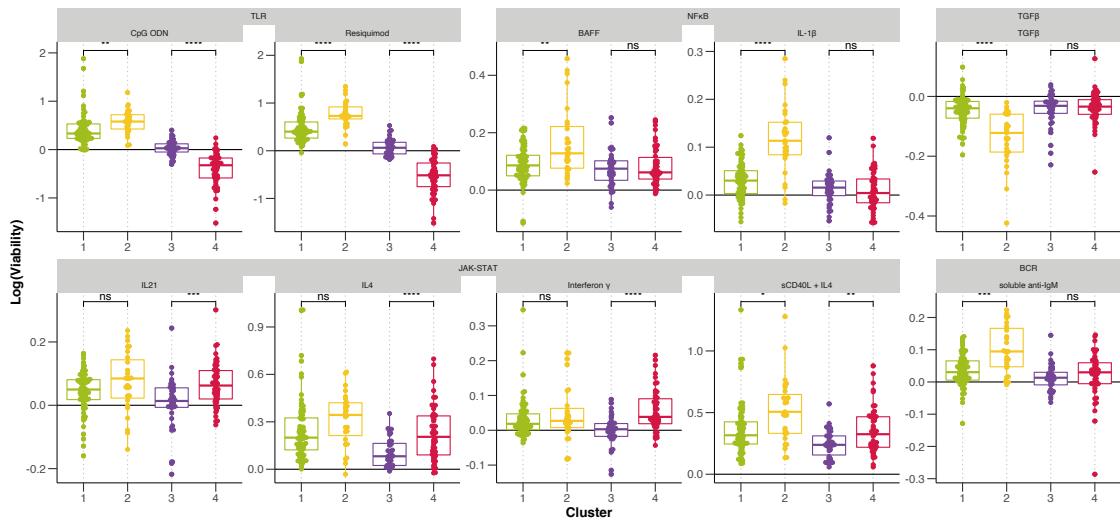


Figure 4.7: Log-transformed normalised viability values stratified by cluster, for each stimulus. Stimuli activating the same pathway are grouped together. P-values from Student's t-test. *Figure adapted from Bruch et al. (2021).*

To further validate this, we observed that TTT in the IGHV-M enriched C3 was significantly shorter than C4 (p -value = 0.005) and comparable to the progression dynamics of IGHV-U enriched C1 and 2 (Figure 4.8B).

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).

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 show 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 4.9).

As expected, the IGHV-U enriched clusters C1 and 2 were more sensitive to BCR

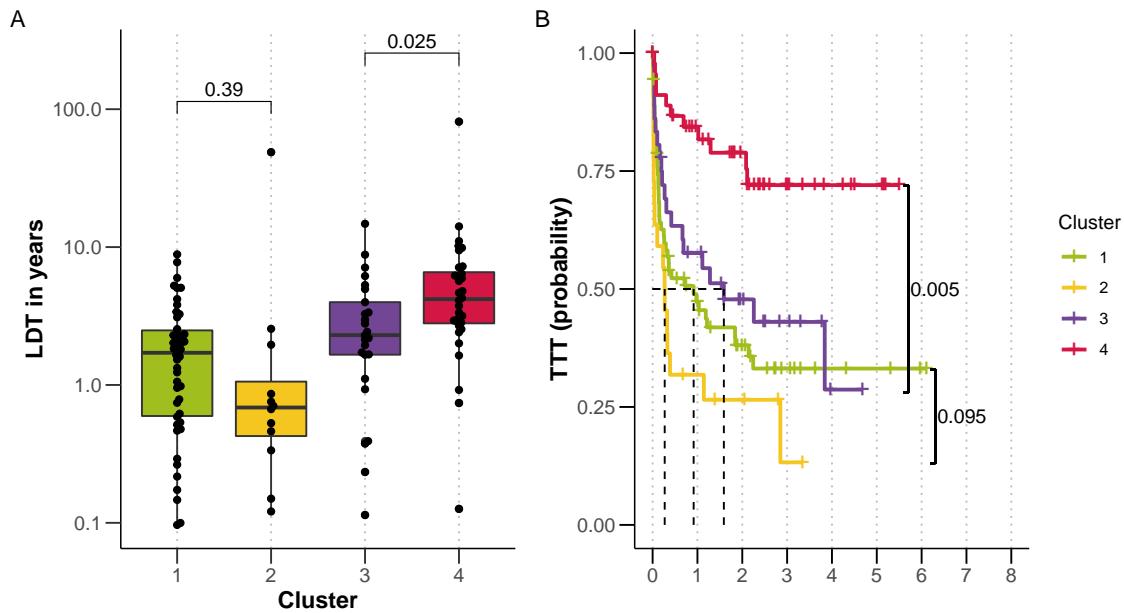


Figure 4.8: (A) LDT stratified by cluster, p-values from Student's t-test. (B) Kaplan-Meier curves to show TTT for each cluster. p-values from univariate Cox proportional hazard models comparing IGHV-U enriched C1 with C2, and IGHV-M enriched C3 with C4. See Methods sections ?? and 2.5.2. *Figure from Bruch et al. (2021)*.

inhibition by ibrutinib, idelalisib and PRT062607, than C3 and 4. Between C1 and C2, C2 was more sensitive to a number of the drugs, including idelalisib (SYK) ($p = 0.012$), everolimus (mTOR) ($p = 0.02$) and the chemotherapeutics fludarabine ($p = 0.031$) and nutlin-3a ($p = 0.042$). Amongst C3 and C4, C3 showed lower sensitivity to everolimus ($p = 0.051$) and to fludarabine ($p < 0.001$) and nutlin-3a ($p = 0.01$). This aligns with the observation that patients in C3 have a poorer prognosis, despite the fact most of these samples were 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.

4.2.4 The clusters are enriched for different genetic features

Next I 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 4.10).

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

To quantify whether each cluster was enriched or depleted for certain genetic features, I used a multinomial modelling approach, with lasso penalisation (for more background on generalised linear models, see section 1.4.3 (Bruch et al. 2021)). I ran the multinomial model to assign coefficients to genetic features that were associated with each cluster, C1 to C4. Running the regression with lasso regularisation meant that only genetic features that were enriched or depleted in a cluster were selected by the model, and coefficients for all other genetic features were shrunk towards 0.

To run the model, I generated a feature matrix, consisting of the genetic data for patient samples and a discrete response matrix, which included the cluster assigned to each patient. Genetic features with >20% missing values were excluded ($n = 39$), and only patients with complete annotation were included in the feature matrix ($n = 137$). I used the function `cv.glmnet` from the R package `glmnet` (Friedman et al. 2021) to generate the model, using 3-fold cross validation, and selecting the optimal model using λ_{min} .

Each cluster was thus assigned a set of genetic coefficients. To ensure that these were robust, I applied certain cut-offs. I ran 50 bootstrapped repeats, and removed coefficients if they were not selected in at least 60% of cases. Additionally, I applied a minimum value of 0.35.

Figure 4.11 shows the resulting mean coefficients. A positive coefficient indicated that this feature was enriched in the cluster, and a negative coefficient indicated it was depleted. The associated standard deviations are also indicated.

As expected, IGHV status was the main feature that predicted cluster membership. Beyond IGHV status, trisomy 12 and *SF3B1* mutations were enriched in C2, which

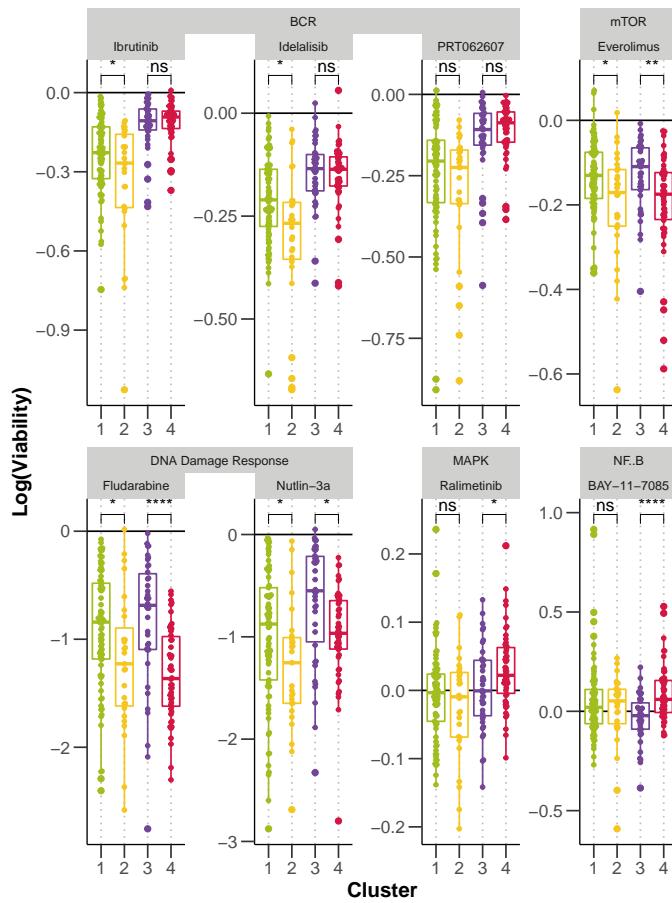


Figure 4.9: Log-transformed normalised viability values, stratified by cluster, for each drug. Drugs targeting the same pathway are grouped together. P-values from Student's t-test.

showed enhanced responses to many stimuli. C4, which was associated with slow in-vivo progression, showed depletion of *TP53*, *ATM*, mutations in *RAS* and *RAF* and gain(8q).

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 C3 and C4, for which 21 RNAseq samples were available (Bruch & Giles et al. 2021).

To quantify differential gene expression, I began by filtering out immunoglobulin genes,

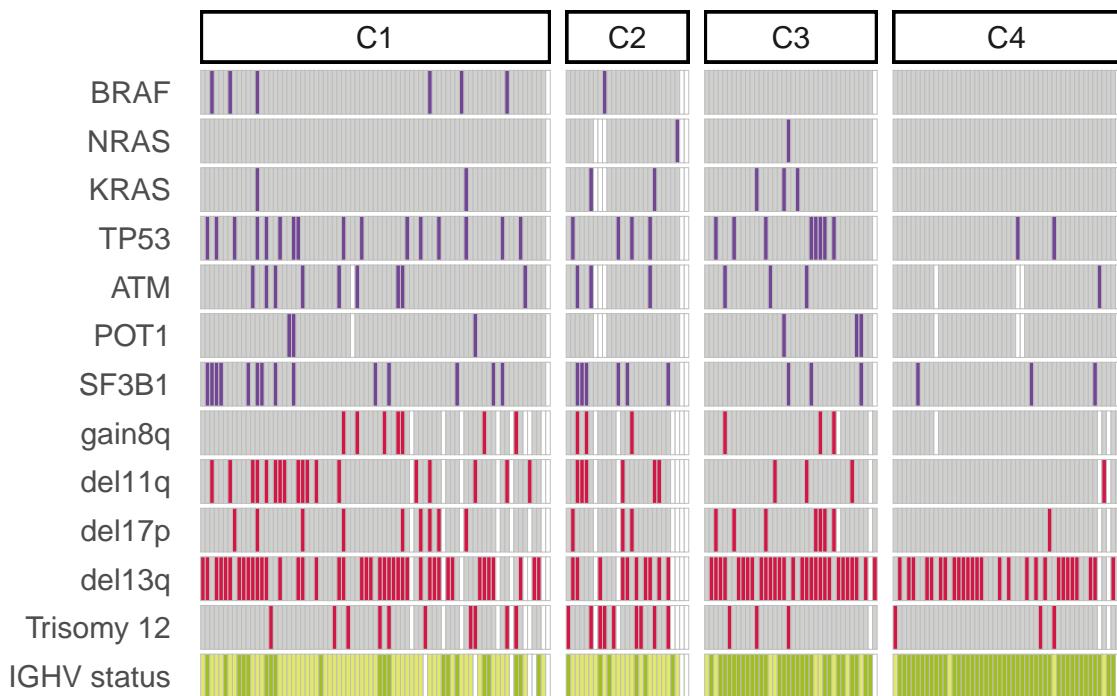


Figure 4.10: Distribution of selected genetic features (rows) within each cluster for all patient samples (columns). White indicates patient sample is not annotated for that feature.

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 affected by this biomarker.

I followed the DESeq2 protocol (Love, Anders, and Huber 2021) 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 4.12).

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 using the fgsea algorithm (Figure 4.13) (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

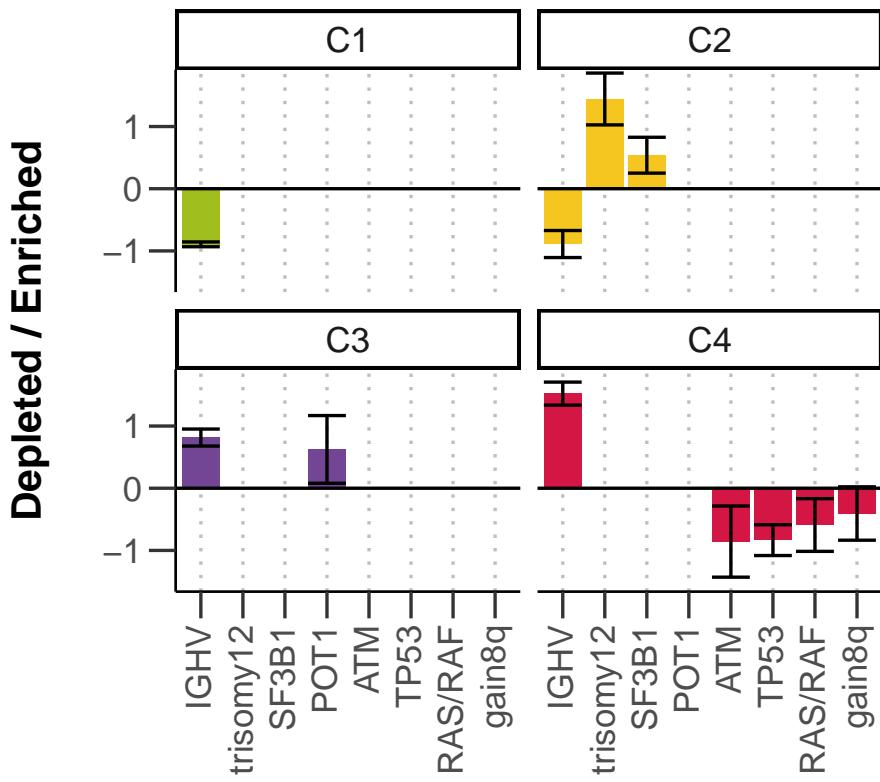


Figure 4.11: Enrichment and depletion of genetic features within each cluster, quantified using lasso-penalised regression. x axis shows genetic features, y axis indicates value and sign of coefficient assigned to feature, for each cluster (positive coefficients are enriched in the cluster, negative coefficients are depleted). Coefficients shown are mean coefficients from 50 bootstrapped repeats and error bars represent the mean plus and minus standard deviation. See Methods section 2.5.2. *Figure from Bruch & Giles et al 2021.*

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 4.14).

In addition, C3 showed upregulation of microenvironmental signalling pathways relative to C4, including TNFa Signalling via NFKB and Interferon Gamma Response (Figure 4.14) (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.

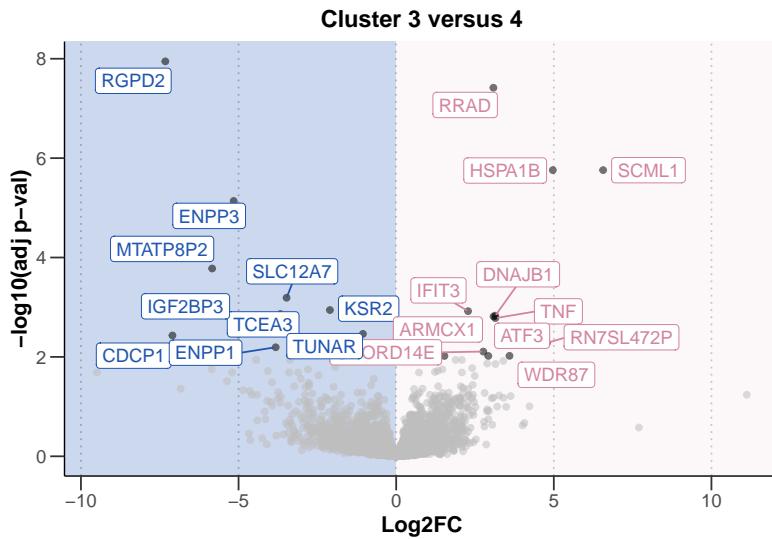


Figure 4.12: Volcano plot of differentially expressed genes between C3 and C4. x axis indicates log₂ fold change values, calculated using the DESeq2 package (**DESeq2?**), y axis gives corresponding -log₁₀(BH-adjusted p value). Darker grey points are labelled where adjusted p < 0.01. See Methods section 2.5.2. *Figure from Bruch & Giles et al. 2021.*

4.3 Summary

The screen enabled a systematic study of the impact of individual microenvironmental pathways on CLL viability. The assay highlighted potent pro-survival signals that are active across heterogeneous genetic backgrounds, such as IL4, and others 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 screen demonstrates the ability of microenvironmental response profiling, to distinguish disease subgroups. The assay revealed four subgroups 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.

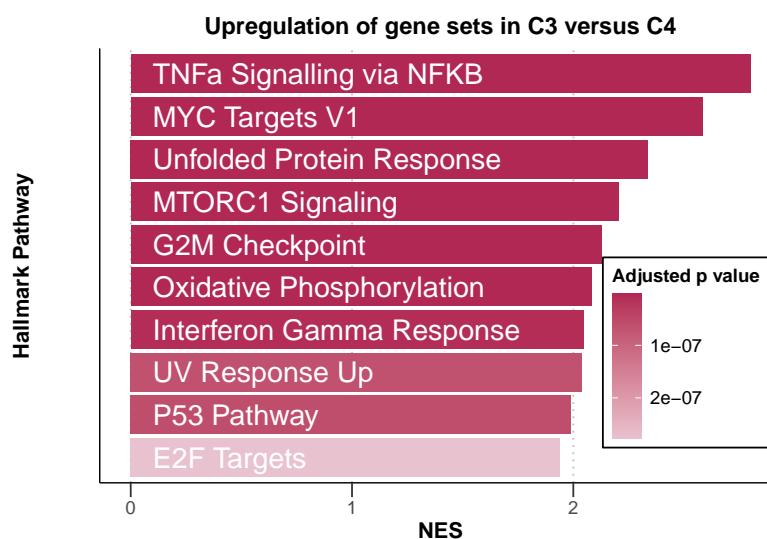


Figure 4.13: Bar plot showing gene set enrichment to compare gene expression of samples in C3 and C4. Normalised enrichment scores (NES) are shown for top ten pathways with smallest adjusted p-value. Gene ranking performed based on Wald statistic, calculated using the DESeq2 package. See Methods section 2.5.2.
Figure from Bruch & Giles et al. 2021.

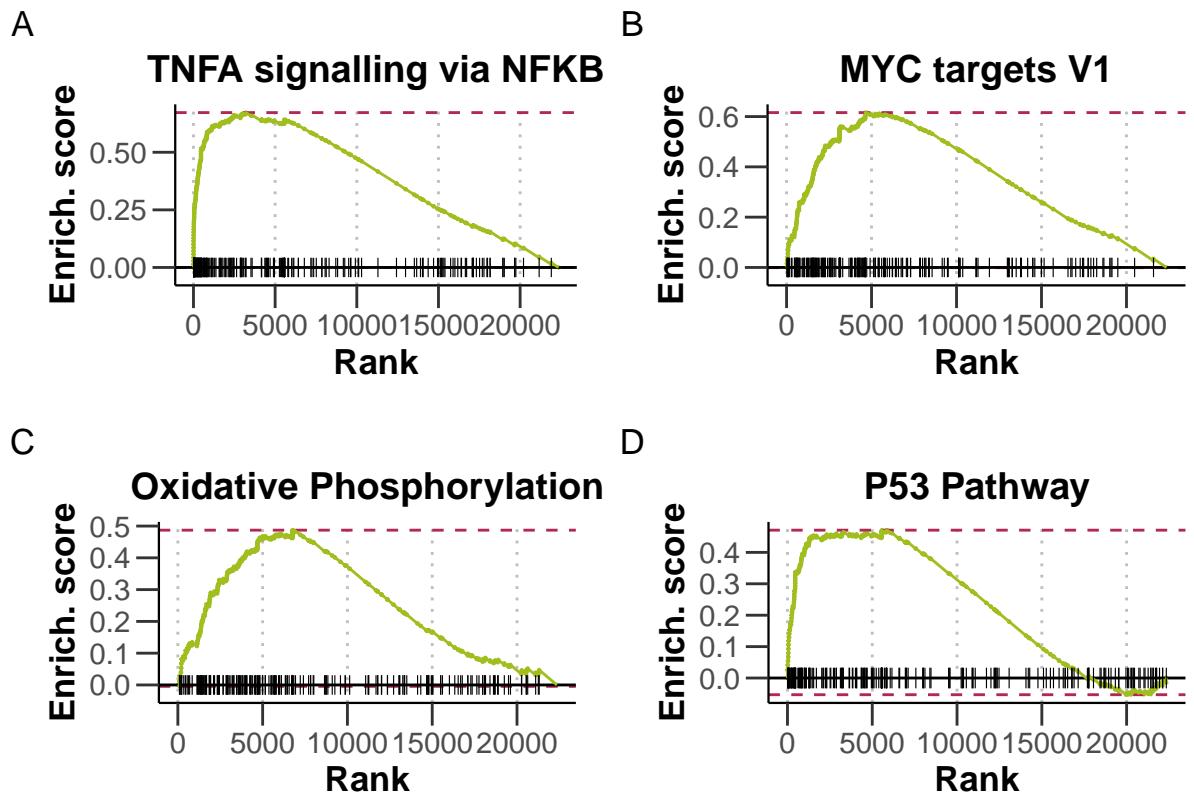


Figure 4.14: Enrichment plot of selected pathways. Gene set enrichment analysis (GSEA) was performed with the Hallmark gene sets from the GSEA Molecular Signatures Database. Wald statistic was used to rank the genes. The green curve corresponds to the Enrichment Score curve, which is the running sum of the weighted enrichment score obtained from GSEA software. See Methods section 2.5.2. *Figure and caption 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 (H. Döhner et al. 2000), had no impact on the responses to the panel of stimuli.

5.1.2 Multivariate analysis of gene - stimulus associations 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 1.4.3.

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 41 features), I excluded genetic features for which >20% of the values were missing, and

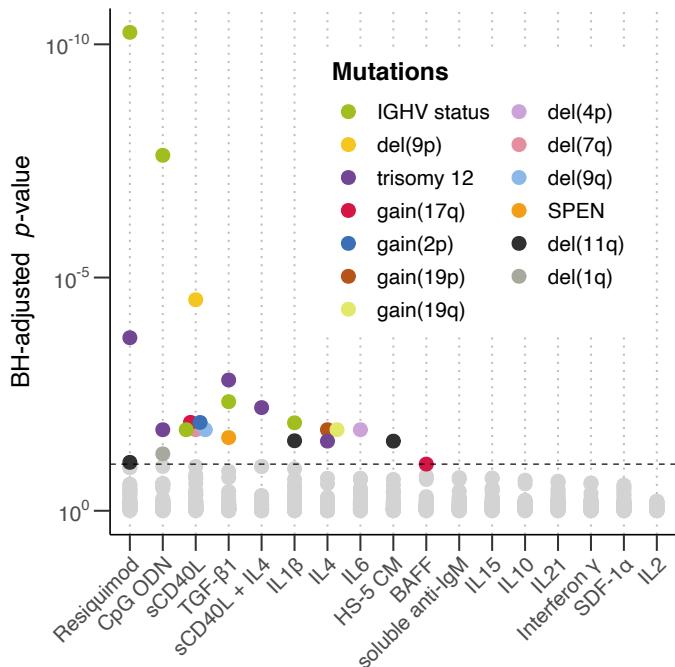


Figure 5.1: Plot showing BH-adjusted p values from Student's t-tests (two-sided, with equal variance), for all tested gene-stimulus associations. Tests performed for IGHV status and somatic mutations and copy number aberrations with 3 patient samples in each group ($n = 63$ or is it 54??). Each circle 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. See Methods section 2.5.2.*

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

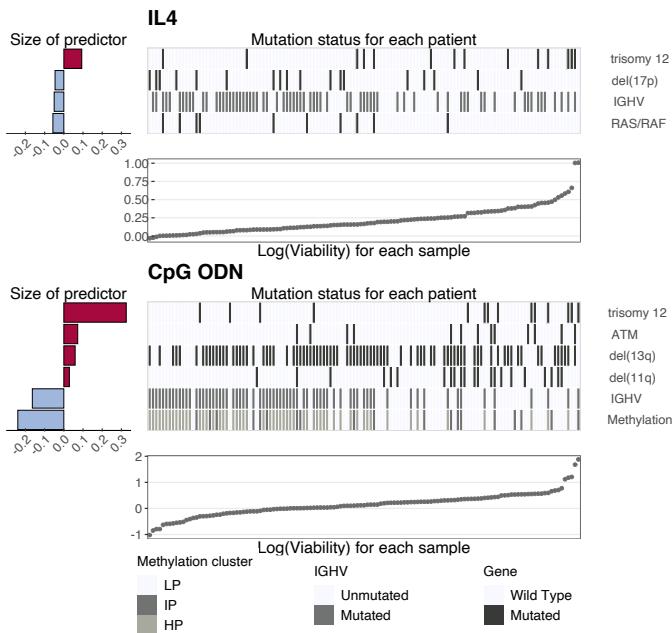


Figure 5.2: Predictor profiles for IL4 and CpG ODN depicting gene - stimulus associations. Bar plots (left) show size and sign of assigned coefficients from Gaussian linear modelling. A positive coefficient indicates that stimulated increase in viability is larger when feature is present. Scatter plots (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.*

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.

Discussion that Inhibition of ERK (p38) also increased IL4 protective effect

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 Resiquimod 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 cohort, 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).

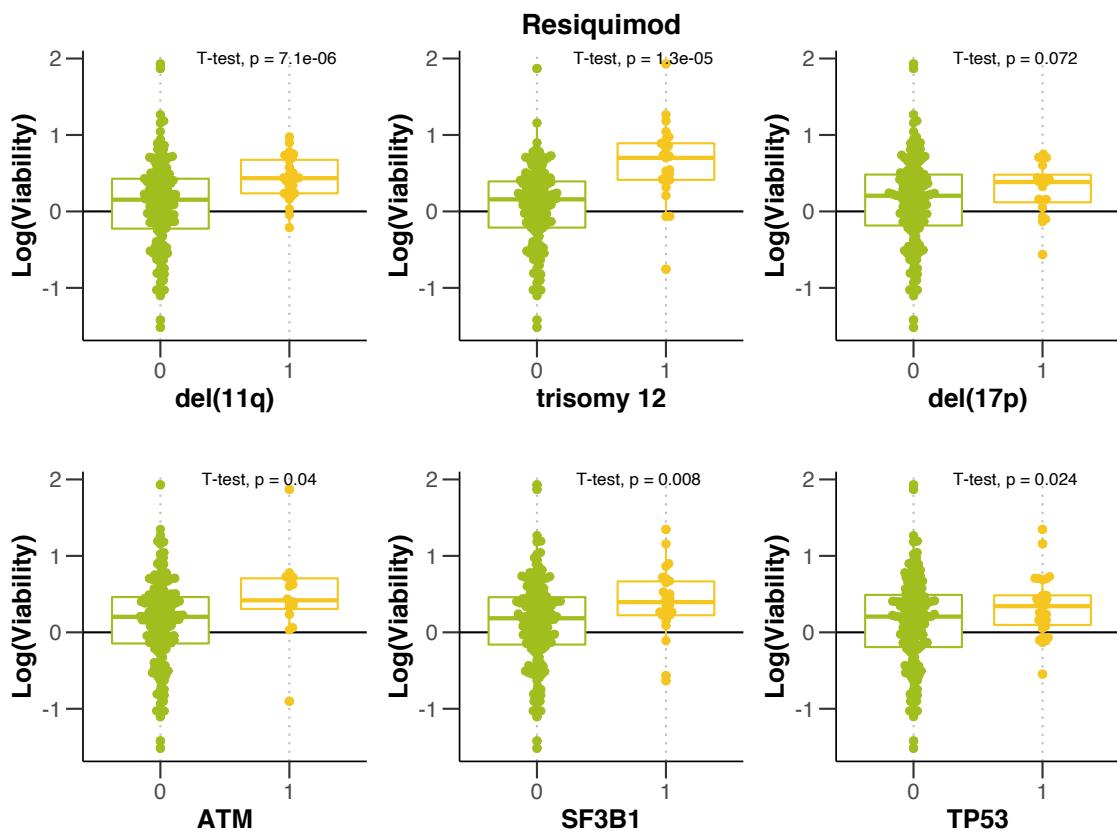


Figure 5.3: Control-normalised log transformed viability values after treatment with Resiquimod (TLR 7/8), stratified by named genetic features.

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 ??). 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.

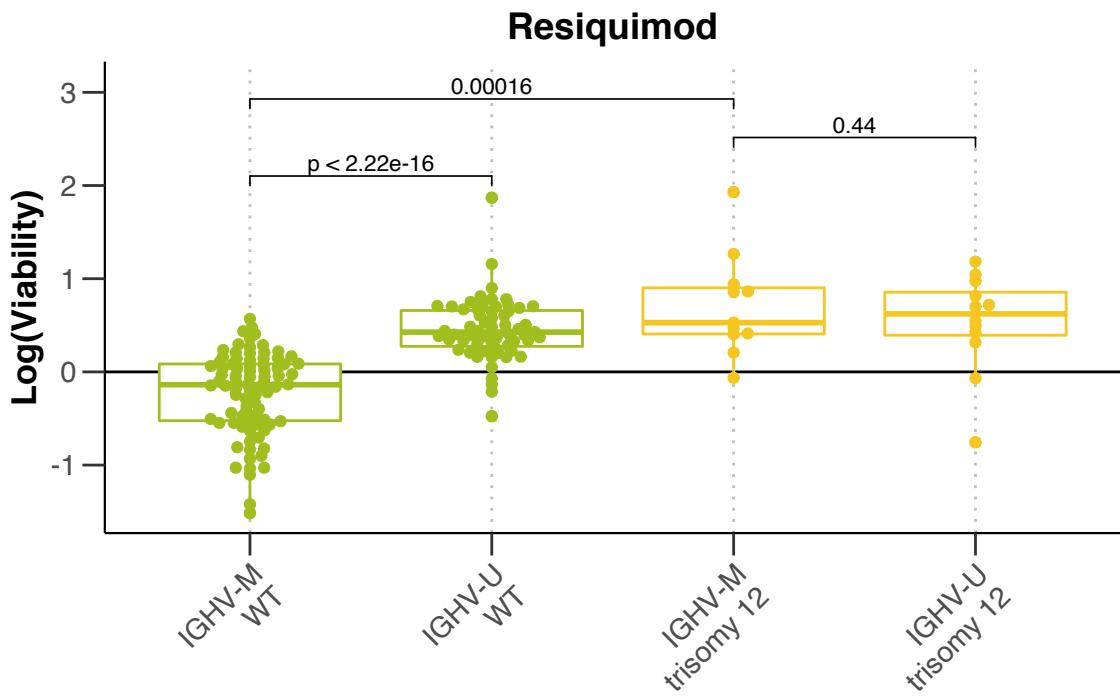


Figure 5.4: Beeswarm-boxplot showing control-normalised log transformed viability values, after treatment with Resiquimod, stratified by trisomy 12 and IGHV status. p-values from Student's t-tests. *Figure from (Bruch & Giles et al. 2021)*

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 (Thomas J. 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)(H. 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 1.1.5).

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

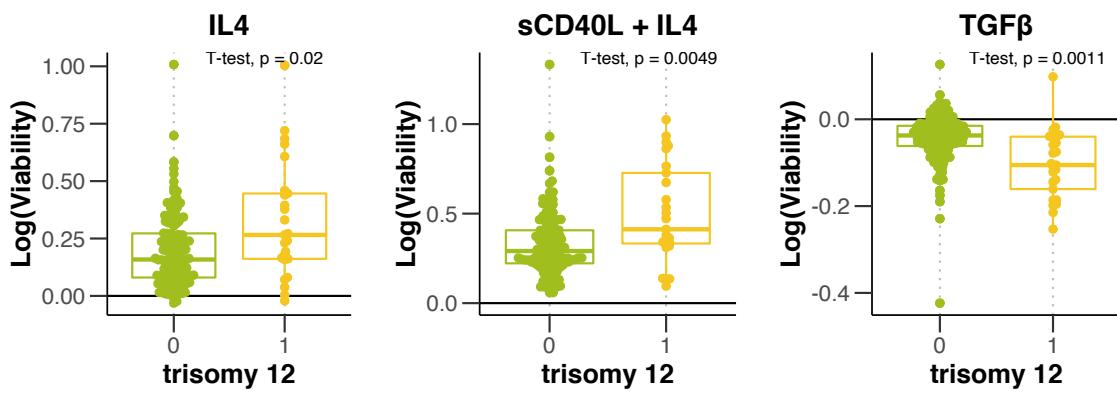


Figure 5.5: Control-normalised log transformed viability values after treatment with IL4, sCD40L + IL4 and TGF β , stratified by trisomy 12.

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; Meier-Abt et al. 2021). 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 sig-

nalling 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.

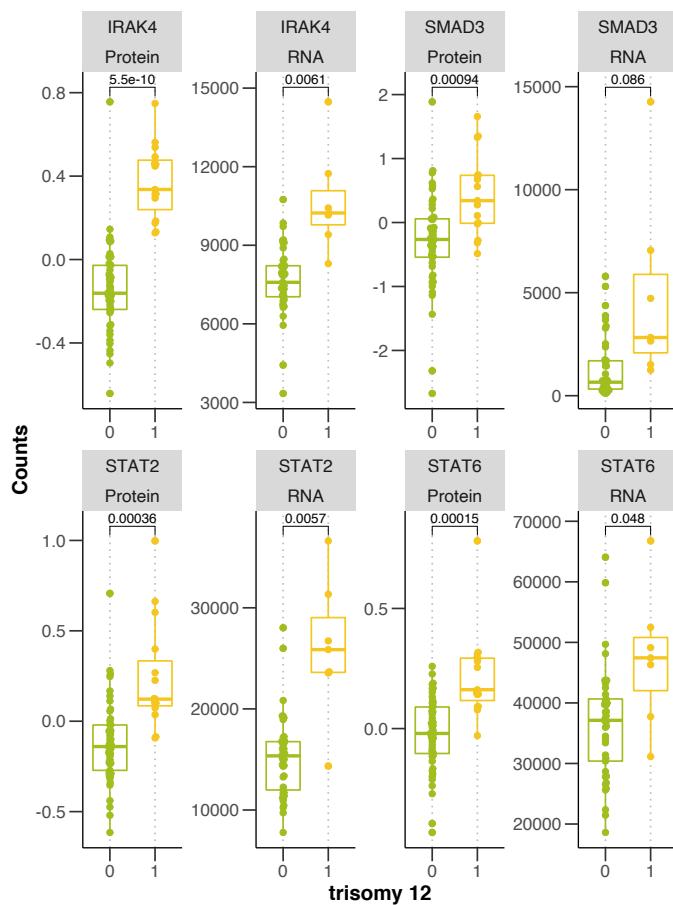


Figure 5.6: Beeswarm-boxplots showing RNA counts and protein abundances stratified by trisomy 12 status. P values from Student's t-test. _Proteomics dataset from ([HerbstThesis_?](#)) Method: 2.5.2.

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).

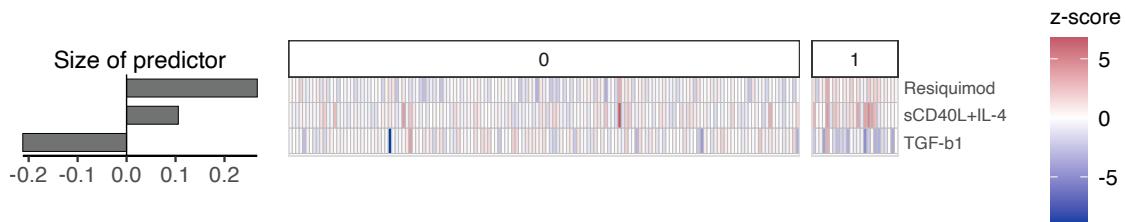


Figure 5.7: Predictor profile showing coefficients selected as predictors of trisomy 12 status. Here, binomial modelling with L1-penalty was used to identify associations between responses to stimuli and trisomy 12 status. Bar plots (left) show size and sign of assigned coefficients named on the right. 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 2.5.2.

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 $\text{TGF}\beta$ (Figure 5.8, and Patient sample B showed the highest viability with IL4, and the second highest with Resiquimod.

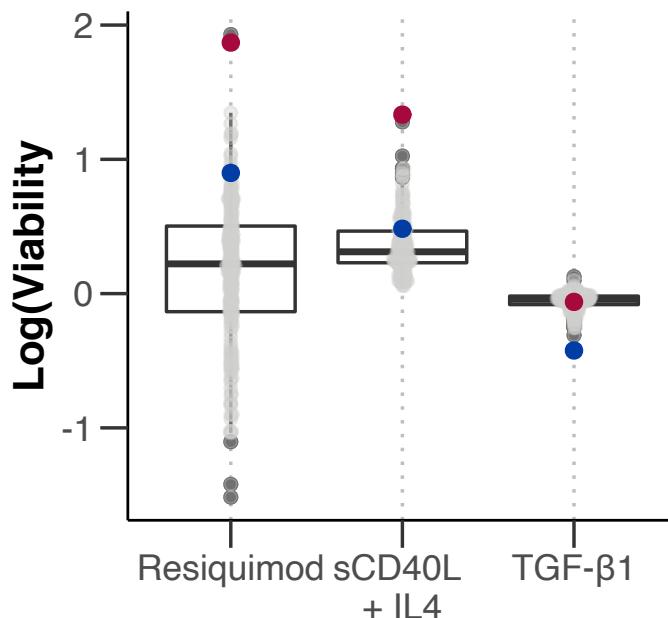


Figure 5.8: Control-normalised log-transformed viability values, for all samples after treatment with Resiquimod, sCD40L + IL4 and $\text{TGF}\beta$. Patient sample A and B are indicated in blue and red, respectively.

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-trisomy 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.

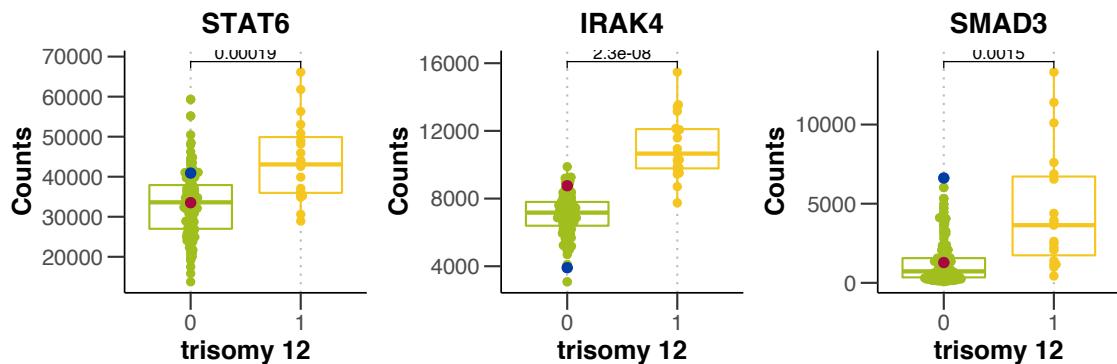


Figure 5.9: Raw RNA counts for SMAD3, IRAK4 and STAT6, stratified by trisomy 12 status. Patient sample A and B are indicated in blue and red, respectively. P values from Student's t-test.

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 [??](#). We provided a list of 636 TFs, from the HOCOMOCO v10 database (Kulakovskiy et al. 2016) to measure their change in activity between trisomy 12 and WT samples (Bruch & Giles et al. 2021).

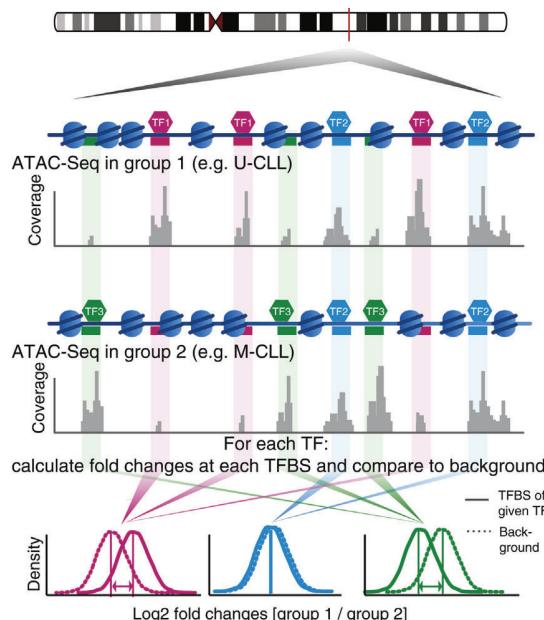


Figure 5.10: `diffTF` workflow from Berest et al. (2019). `diffTF` accepts a list of TFs along with the genomic locations of their binding sites. For each TF, the software computes the distribution of fold changes between the trisomy 12 and WT samples, using the ATACseq peaks at each TF binding site 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.

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

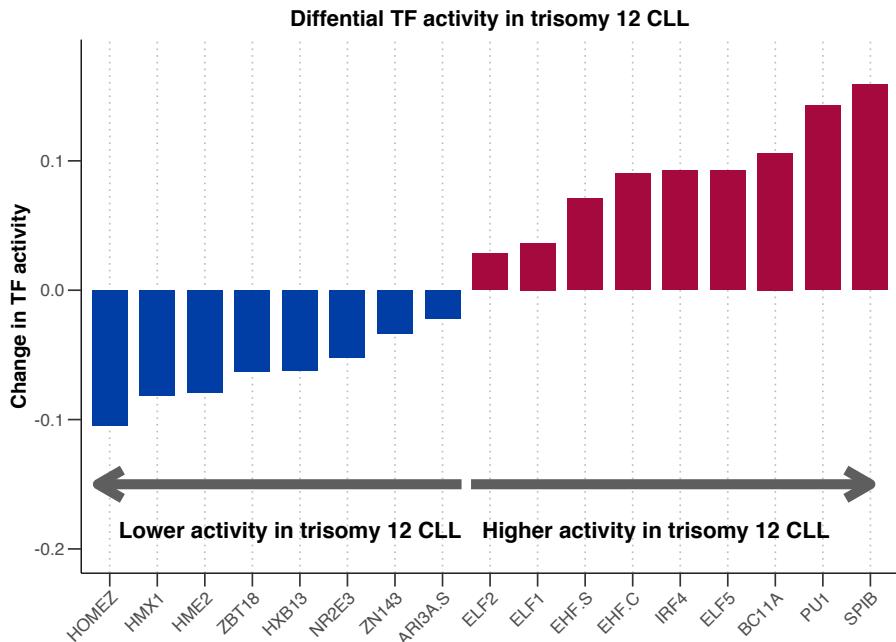


Figure 5.11: Bar plot showing the results of the diffTF analysis for the Rendeiro et al. (2016) dataset. y axis shows change in TF activity (weighted mean difference) between trisomy 12 ($n = 9$) and non-trisomy 12 samples ($n = 43$), x axis indicates TF names. 17 / 636 TFs, with BH adjusted $p < 0.05$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

selected KEGG(**KEGG?**) and Reactome (**Reactome?**) genesets, using the ‘R’ package `clusterProfiler` (G. Yu 2021b). This method corresponded to a one-sided version of Fisher’s exact test. This analysis showed TLR, BCR and $TGF\beta$ 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

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.*

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. Secondly, 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.

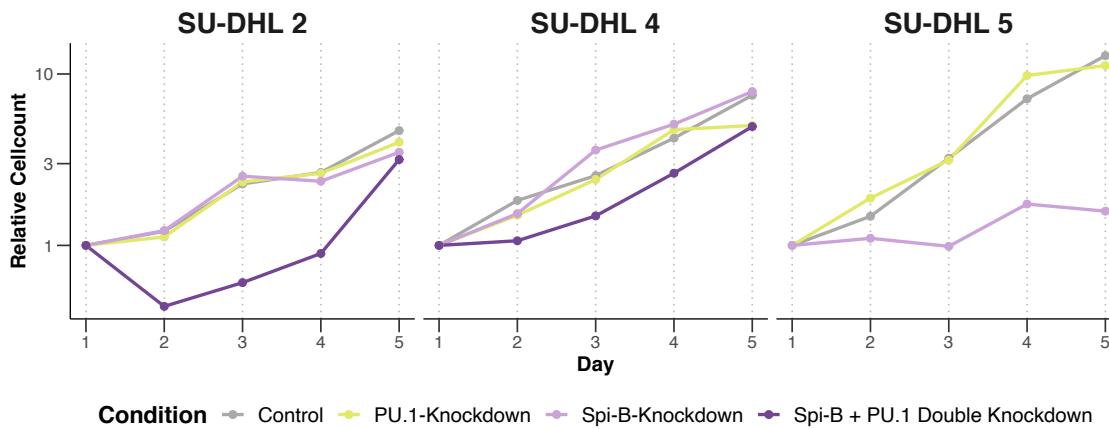


Figure 5.13: Knockdown of Spi-B and PU.1 led to growth restriction and cell death.

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.*

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 (H. 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 and PU1 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 SPIB binding (which is on chromosome 12 and is higher) <https://pubmed.ncbi.nlm.nih.gov/7566969/> PU1 and SPIB bind in mouse lymphoma cell line <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 mechanism by which these mutations contribute to CLL drive. It also points to TLR response as a useful prognostic marker.

Power of multivariate modelling; 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

BCR signalling - which is important with IgHV and trisomy 12, is elevated in the LN according to gene signatures: <https://ashpublications.org/blood/article/117/2/563/28102/The-lymph-node-microenvironment-promotes-B-cell>

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 Bericovic 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 facts - main analysis nad key finding, this hsould be routine Then outline in detail at the bottom

Chapter 6

Molecular and microenvironmental modulators of drug response in CLL

Chapters 4 and 5 provide a systematic exploration of the effect of microenvironmental stimulation on *ex vivo* CLL biology, and how these pathways can be modulated by genetic features. Collectively, these results underline the fact that cancer biology is determined by an integrated network of cell-intrinsic molecular aberrations and cell-intrinsic signals generated by cell-cell contacts and soluble factors. The comprehensive nature of this dataset enables the study of these multiple axis.

In this chapter, we examine the third axis of the dataset: drug response. As well as providing protective signals that increase CLL viability and support tumour growth, the microenvironment is also believed to interfere with drug action. Several studies have demonstrated the impact of individual microenvironmental signals on drug response *ex vivo*, and is also thought to occur within the protective niche (Ahn et al. 2018). Guided by this observation, we next wanted to investigate the impact of each of the stimuli on responses to the panel of drugs, and how this can be further modulated by genetic features.

^sharpen this once more

In section 6.1, the impact of the stimuli on drug response is quantified by identifying drug - stimulus interactions. In section 6.2, the impact of genetic features on drug response

is explored. Finally, in section 6.3, multivariate modelling is applied to integrate all three axes of the dataset, to investigate how drug - microenvironment interactions can be further modulated by genetic features. These analyses benefit from linear regression and generalised linear modelling approaches, described in more detail in section 1.4.3.

6.1 Mapping drug - stimulus interactions

6.1.1 Linear modelling quantifies interactions between drugs and stimuli

To begin, I evaluated computational methods for quantifying the impact of the stimuli on drug action. The aim was to screen for cases where the stimulus pathway specifically interacted with the target pathway of the drug.

When a stimulus is individually applied to a CLL sample, the stimulus activates signalling cascades that modulate CLL viability and impact the rate of spontaneous apoptosis. When a stimulus is co-applied with a drug, the stimulus will continue to impact upon baseline viability, at the same time as the drug inhibits baseline viability. Assuming the drug and stimulus do not interact, the viability of the cells with the combinatorial treatment will equate to the additive impact of both these compounds. In the case that there is some interaction between the stimulus pathway and drug target pathway, the resulting viability will be more complex. The difference between the additive, or expected viability, and the true measured viability, can be quantified by an interaction factor.

I fitted a linear model (Equation (6.1)) to quantify how each stimulus modulates drug efficacy beyond its individual effect (Bruch et al. 2021). I used the `lm` function implemented in the R package stats(R Core Team 2021). The model accounts for the individual effects of the drug and stimulus treatment on spontaneous apoptosis and baseline viability, in addition to an interaction term, named β_{int} . β_{int} , the interaction factor, quantifies how the combined treatment effect differs from the sum of the individual treatments.

Equation (6.1) quantifies how the viability with any combination can be predicted:

$$\log(V) = \beta_{drug}X_{drug} + \beta_{stimulus}X_{stimulus} + \beta_{int}X_{drug}X_{stimulus} + \epsilon \quad (6.1)$$

where V is the predicted viability with a given treatment, β_{drug} , $\beta_{stimulus}$ and β_{int} are coefficients for the drug, stimulus and combinatorial terms and X_{drug} and $X_{stimulus}$

are indicator variables (0 or 1) for the presence or absence of a drug/stimulus. ϵ is the vector of model residuals. See also Methods section 2.5.2. Equation from ([Giles2021._?](#))

This approach identified 45 drug-stimulus combinations (out of 204), where β_{int} had $p < 0.05$, i.e. the stimulus had a specific effect on the efficacy of the drug, highlighting the extent to which drug action can be modulated by cell-external signals *ex vivo*.

These interactions are of potential clinical interest, to support our understanding of *in vivo* drug response, and the development of strategies to target drug resistance pathways. They are also relevant to laboratory studies of drug action, which are very often performed in the absence of such signals.

To establish which interactions may be important in these contexts, we began by classifying the drug - stimulus interactions into four categories (Bruch et al. 2021). The categories were firstly defined based on whether the interaction was antagonistic or synergistic, i.e. the stimulus acted to oppose or reinforce the activity of the drug. These were further divided based on the sign of β_{int} i.e. whether the combinatorial viability was higher (positive) or lower (negative) than would be expected based on additive effects (Figure 6.1A).

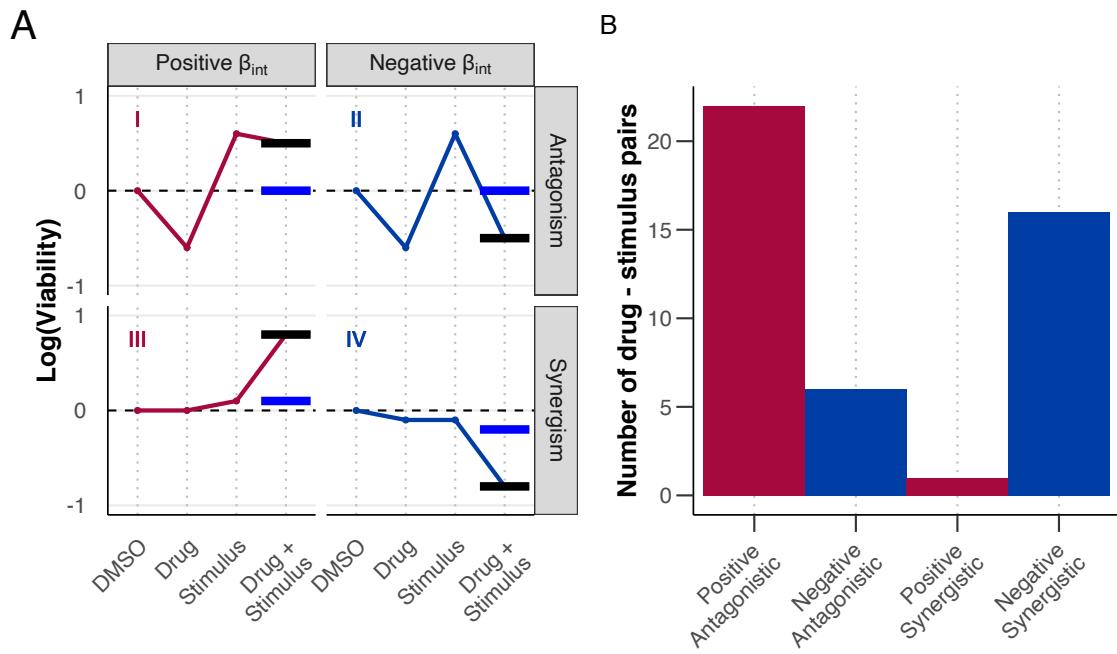


Figure 6.1: (A) Graphical line plots representing typical response patterns for each of the four drug - stimulus interaction categories (types I - IV). x-axis shows treatment, y-axis shows log transformed viability values, facet labels indicate interaction type. Blue and black horizontal lines demonstrate effect of interaction on viability with combinatorial treatment. Blue line shows predicted viability based on additive effects alone and black line shows viability accounting for additive effects and interaction. (B) Histogram showing number of drug-stimulus interactions within each category, where p value for β_{int} is <0.05 . See Methods section 2.5.2. _Figure and caption from (Giles2021._?)

These categories encompassed the following interaction types. Type I defined positive antagonisms in which stimuli reversed drug action, leading to decreased drug efficacy and increased CLL viability. This group highlights drug - stimulus interactions that could be relevant to treatment resistance pathways *in vivo*. Type II negative antagonisms highlighted cases where the drug inhibited the stimulus, reducing the effect of the stimulus on baseline viability, and decreasing CLL viability. These interactions include drugs that provide potential therapies to target treatment - resistance pathways *in vivo*. Type III and type IV interactions cover synergies, in which the drug and stimulus increase the effect of the other. Type III positive synergies lead to increased CLL viability with the combinatorial treatment, whilst type IV negative synergies lead to increased drug efficacy and lower viability. In the latter case, these may highlight cases where drug ef-

ficacy is dependent on external pathways, and may be useful for understanding *ex vivo* drug action. Figure 6.1B quantifies the number of interactions in each category.

6.1.2 INF γ induces resistance to ibrutinib *in vitro*

Using the linear modelling approach in combination with our classification system, we attempted to generate the first comprehensive map of drug - stimulus interactions in CLL (Bruch et al. 2021). This map provides a resource to explore the impact of the microenvironment on the majority of drugs that are clinically-licensed in CLL (Figure 6.2). We use this resource to investigate a number of drug - stimulus interactions here, and in section 7.

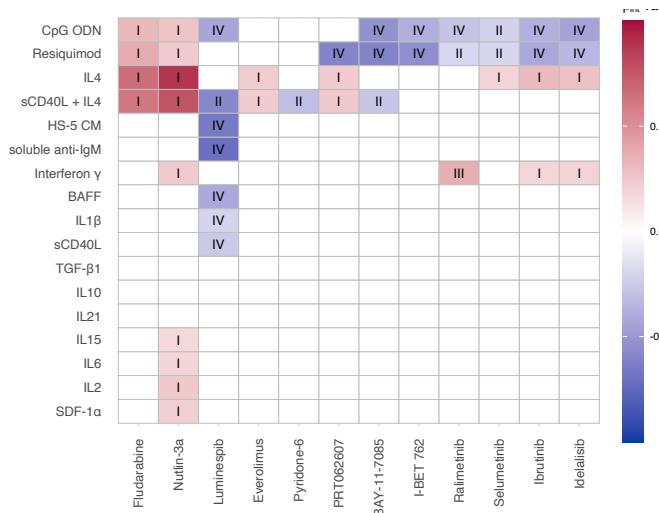


Figure 6.2: Heatmap of all β_{int} values for all drug - stimulus combinations where p for $\beta_{int} < 0.05$, annotated with interaction type. Scale indicates size and sign of β_{int} . Rows and columns clustered according to hierarchical clustering. See Methods section 2.5.2. _Figure and caption from (Giles2021._?)

To complement this broader approach, I next examined each interaction individually, to select the most clinically interesting examples (Figure 6.3).

Positive antagonistic interactions were the most common, in which stimuli reversed drug action. This group included resistance mechanisms such as the inactivation of ibrutinib by IL4 stimulation (Figure 6.3A)(Aguilar-Hernandez2016?), indicating that the modelling and classification approach successfully recapitulated known drug - stimulus interactions. Our approach demonstrated that IL4 induced resistance to a range of drugs, beyond ibrutinib, and showed the highest number of positive antagonistic interactions

of all of the stimuli.

Amongst the positive antagonistic interactions, we noted reduced efficacy of ibrutinib in the context of IFN γ stimulation (Figure 6.3B, (Bruch et al. 2021)). [ADD FOLLOW UP / STAT1 experiment].

There were six cases of negative antagonisms, whereby drug action inhibited the pro-survival effect of the stimulus. As IL4 appeared to be the most potent conferrer of drug resistance, it was of greatest interest to identify drugs that may inhibit this pathway. For example, the Pan-JAK inhibitor pyridone-6 inhibited the increase in viability with sCD40L + IL4 stimulation (Figure 6.3C, (Bruch et al. 2021)).

The model identified a single positive synergism, in which IFN γ treatment in combination with ralimetinib, a p38 MAPK inhibitor, stimulated a large increase in viability that was not observed in each of the single treatments (Figure 6.3D, (Bruch et al. 2021)). This finding suggests a potential inhibitory effect of p38 MAPK activity on signalling via IFN γ . [ADD TO THIS NOS THEORY]

16 drug - stimulus combinations demonstrated negative synergistic interactions. This implied that the drug efficacy was somehow dependent upon, or increased by, activation of the stimulus pathway. For example, luminespib, a HSP90 inhibitor, showed higher efficacy with a number of stimuli, including soluble anti-IgM, indicating that *ex vivo* measurements of luminespib action are likely to be affected by the presence or absence of microenvironmental signals (Figure 6.3F, (Bruch et al. 2021)).

Collectively, these results represent a comprehensive and systematic study of the influence of the microenvironment on drug efficacy in CLL. This work highlights key resistance pathways, strategies for targeting microenvironmental resistance, and underlines important pathways in *ex vivo* studies of drug efficacy. In this thesis, I have selected some of the more interesting examples, such as resistance to ibrutinib induced by IL4 and IFN γ , to investigate in more detail here and in Chapter 7. This work also aims to serve as a resource for others, to provide a basis for further follow-up studies beyond the ones in this thesis. It also aims to provide a proof-of-principle highlighting the value of combinatorial screening, and linear modelling approaches, to study the impact of the microenvironment in cancers in general. All drug - stimulus combinations and interactions can be explored on the online shiny app, and the whole analysis can be replicated, or adapted, from the online code repository, published alongside (Bruch et al. 2021).

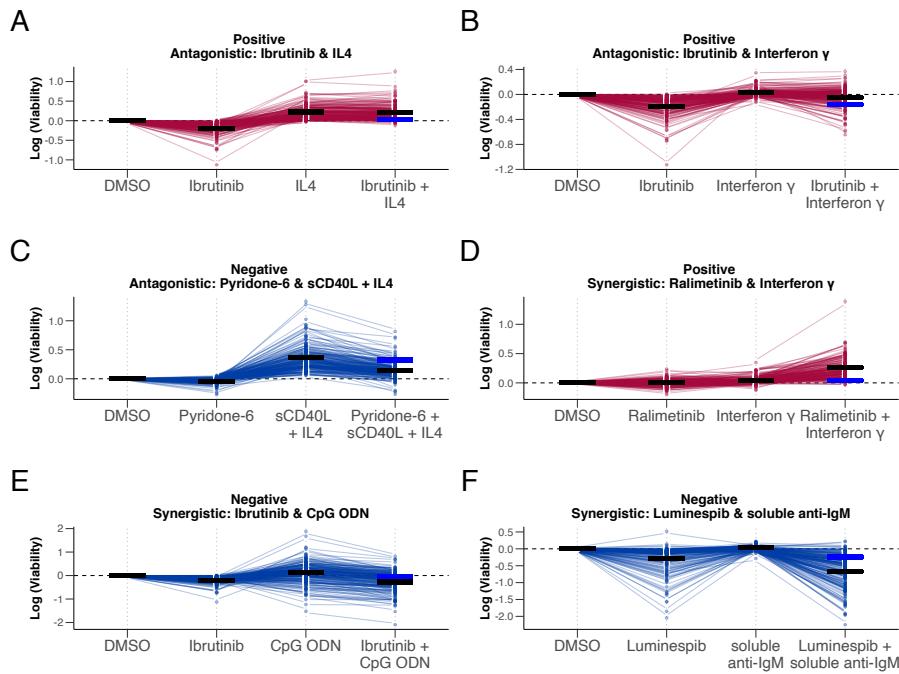


Figure 6.3: Line plots showing examples of drug-stimulus interactions. x axis indicates treatment, y axis shows log transformed viability values, with matching samples linked across treatments. Black horizontal lines represent predicted viability for each treatment, using coefficients from linear model fit. For combinatorial treatment, blue line indicates predicted viability based on additive effect alone, and black line indicates predicted viability accounting for additive effects plus interaction. See Methods section 2.5.2. *Figure and caption from (Bruch et al. 2021).*

6.2 Genetic modulators of drug response

6.2.1 Univariate analysis identifies genetic modulators of drug response

Most mutations in cancer have not been linked to drug response (Dietrich et al. 2017). CLL has provided an powerful model system in which to study the link between recurrent genetic features and drug response: liquid biopsies are easy to obtain and sequence, and screen *ex vivo* and it is much easier than in many cancers to take samples over a time course. Previous work in our lab (Dietrich et al. 2017) has taken advantage of this, generating the largest survey of molecular determinants of response in CLL to date.

Before investigating the impact of mutations on drug - stimulus interactions, I performed

a similar analysis of genetic determinants of drug response, to first establish the effects of mutations independently of stimuli on drug response. I used the genetic data, including all mutations, copy number variants and IGHV status for which there were greater than 3 cases in the cohort ($n = 54$), to perform t-tests to identify molecular features that impact on drug response (Figure 6.4).

This analysis recapitulated the findings of previous work (Dietrich et al. 2017), including the impact of IGHV status on responses to BCR inhibitors ibrutinib (BTK) and idelalisib (Pi3K), and the impact of TP53 and del(17p) mutations on Nutlin-3a (MDM2) and Fludarabine (Purine analogue).

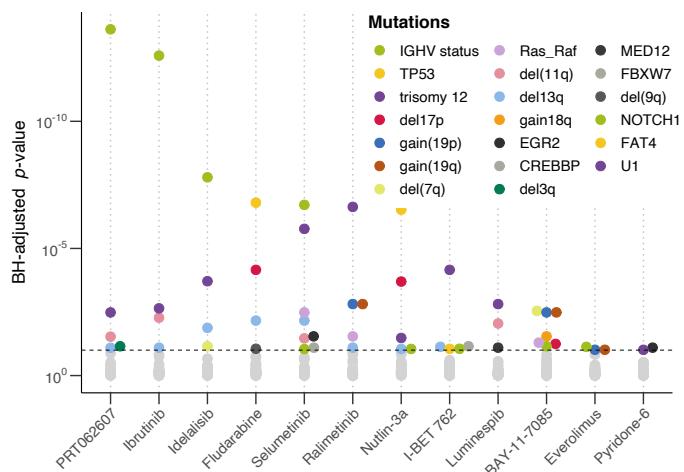


Figure 6.4: Plot showing BH-adjusted p values from Student's t-tests (two-sided, with equal variance), for all tested drug-gene associations. Tests performed for IGHV status and somatic mutations and copy number aberrations ($n=54$). Each coloured circle represents a gene-stimulus association meeting 10% FDR cut off. Method 2.5.2 Figure from (Bruch et al. 2021).

In addition, it also highlighted the broad impact of trisomy 12 status on drug response *ex vivo* (Figure 6.5). While trisomy 12 is known to accelerate disease progression, tumours with trisomy 12 are more treatable due to higher sensitivity to chemotherapeutics and ibrutinib. In my results, trisomy 12 increased sensitivity to all drugs except —fludarabine, BAY-11-7085 and Everolimus. These included increased sensitivity to p38 MAPK inhibition by ralimetinib, agreeing with previous work demonstrating higher trisomy 12 sensitivity with a range of MEK and ERK inhibitors, suggesting an essential role for MEK/ERK signalling in trisomy 12 CLL (Dietrich et al. 2017). Trisomy 12 samples also showed higher sensitivity to Nutlin-3a, an MDM2 inhibitor. Interestingly, the MDM2 gene is located on chromosome 12, and is more abundant in trisomy 12

samples [ADD SUPP FIG], which may explain this. I also noted increased trisomy 12 sensitivity to IBET-762, a BET inhibitor that has been shown to suppress transcriptional responses to cytokine signalling via JAK-STAT (Chan et al. 2015). If trisomy 12 samples are more dependent on cytokine signals, as the results in section 5.2.1 suggest, this makes sense.

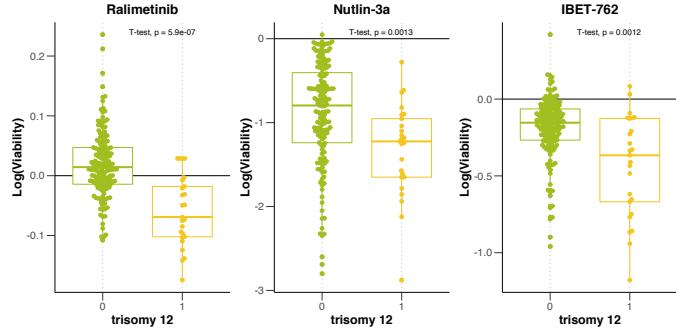


Figure 6.5: Beeswarm-boxplots showing control-normalised log transformed viability values after treatment with (A) Ralimetinib, (B) Nutlin-3a and (C) IBET-762, stratified by trisomy 12.

6.3 The modulatory effect of mutations on drug - stimuli interactions

6.3.1 Patient-specific linear modelling identifies drug - stimulus - gene interactions

Sections 6.1 & 6.2, explore the effects of microenvironmental signals and molecular features on drug response independently. I next aimed to investigate their collective effect on *in vitro* drug efficacy in CLL, by quantifying the extent to which genetic driver mutations modulated the interactions between stimuli and drugs (Bruch et al. 2021).

I began by conceptualising methods to quantify the collective effect of the drugs, stimuli and genetic features on CLL viability and biology. I decided to adapt the linear model in Equation (6.1), by fitting this model in a patient sample - specific manner (Bruch et al. 2021):

Equation (6.2) quantifies how the viability with any combination can be predicted:

$$\log(V) = \beta_{drug} X_{drug} + \beta_{stimulus} X_{stimulus} + \beta_{patient} X_{patient} + \beta_{drug-stimulus} X_{drug} X_{stimulus} + \beta_{drug-patient} X_{drug} X_{patient} \quad (6.2)$$

where V is the predicted viability of a patient sample with a given treatment, β_{drug} , $\beta_{stimulus}$, β_{int} , $\beta_{drug-stimulus}$, $\beta_{drug-patient}$, $\beta_{drug-stimulus}$ and $\beta_{int}X_{drug}$ are regression coefficients for the drug, stimulus, patient sample and combinatorial terms and X_{drug} , $X_{stimulus}$ and $X_{patient}$ are indicator variables (0 or 1) for the presence or absence of a drug/stimulus/patient sample. ϵ is the vector of model residuals. See also Methods section 2.5.2. Equation from (Bruch et al. 2021).

This generates a higher order interaction term $\beta_{int}X_{drug}X_{stimulus}X_{patient}$. This term represents a patient sample-specific β_{int} for each drug - stimulus combination, quantifying the size of an interaction between a drug and stimulus in each patient genetic background.

With these patient sample-specific β_{int} terms, it was possible to search for associations between the size of β_{int} and genetic features. The aim was to screen for molecular features that increased or decreased the size of a drug - stimulus interactions, using multivariate regression with L1 (lasso) regularisation.

As input to the model, the response matrix was composed of the sample - specific β_{int} values for each drug-stimulus combination. To generate the feature matrix (137 samples versus 40 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 ($p=39$) and IGHV status (coded as 0-1). 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 90% of 30 bootstrapped repeats.

This analysis revealed that 60/204 drug - stimulus interactions were modulated by at least one genetic feature (Figure 6.6, Appendix Figure 8.2). A positive coefficient here indicates that the presence of the genetic feature is associated with more positive β_{int} , in other words, the viability with the drug and stimulus is higher than expected in the presence of the genetic feature.

Have to make sure this figure takes a whole page but doesn't float to the bottom

Applying this broad scale screening approach established some wider trends. Firstly, Trisomy 12 and IGHV status impacted the largest number of drug - stimulus interactions, indicating that their impact on stimulus and drug response individually also extends to drug-stimulus interactions. Secondly, out of all the stimuli, IL4 and sCD40L + IL4-based interactions were modulated by the largest number of features. IL4 alone in-

creased viability uniformly across genetic backgrounds; its interaction with drugs may be more influenced by genetic features. Notably, the presence of trisomy 12 acted to increase IL4-induced drug resistance, in almost all drugs for which $\beta_{int}>0.05$. Interestingly, the presence of TP53 and del(17q) reduced the degree to which IL4 induced resistance to the chemotherapeutics Nutlin-3a and Fludarabine, most likely because both these mutations already induce resistance alone.

6.3.2 Patient - specific drug - stimulus interactions may have clinical significance

This broad scale screening approach was not perfect in itself. Modelling can only go so far, and it is likely that some cases identified here are artefacts of the modelling approach. Therefore, I decided to investigate each individual case one - by - one to establish which drug - stimulus - gene interactions are the most biologically interesting. This analysis highlighted several drug - stimulus - gene interactions that may have clinical importance and warrant further investigation (Bruch et al. 2021).

The first of these concerned the interaction between fludarabine and CpG ODN. The value of β_{int} for fludarabine and CpG ODN was modulated by six genetic factors; most strongly by IGHV status, del(11q) and trisomy 12 (Figure 6.7, (Bruch et al. 2021)).

To look in more detail at what this means, I visualised sample responses to CpG ODN and fludarabine, stratified by IGHV status and trisomy 12 (Figure 6.8, (Bruch et al. 2021)). In IGHV-M non-trisomy 12 samples, TLR stimulation increased fludarabine efficacy. In samples that were either IGHV-U or trisomy 12+, TLR stimulation induced resistance to fludarabine. TLR stimulation *ex vivo* can increase or decrease viability, depending on the genetic background and this also impacts on drug response. It would be useful to establish whether these dynamics are replicated within the *in vivo* niche.

TLR-induced drug resistance in certain genetic backgrounds extended to other chemotherapeutic drugs. Nutlin-3a is an MDM2 inhibitor, which in our dataset shows higher efficacy in certain backgrounds including IGHV-U, trisomy12, and del(11q) (Figure 6.4, Figure 6.5, Appendix Figure 8.3). Despite this, TLR stimulation reduces Nutlin-3a toxicity in these same genetic backgrounds, such that treatment with Nutlin-3a and CpG ODN is more toxic to IGHV-M, non-trisomy 12 and non-del(11q) samples i.e. the opposite(Figure 6.9, Figure 6.10). This observation underlines the need to observe drug activity *in vitro* in the context of microenvironmental signals, in CLL and in other can-

cers.

The modelling approach also identified interactions, where the differential size of β_{int} was driven by the single treatment effect, rather than the combination. For example, if a drug is more efficacious in a certain genetic background, but a stimulus nonetheless eradicates drug toxicity, the size of the interaction will be larger in this genetic background. This is because the increase in expected viability in these samples (quantified by β_{int}) would be even higher than in other genetic backgrounds. This is the case with Ibrutinib + IL4 for instance (Bruch et al. 2021). Ibrutinib showed higher efficacy in trisomy 12 and IGHV-U samples and IL4 induced complete resistance to ibrutinib independently of genetic background. Thus in trisomy 12 and IGHV-U samples treated with ibrutinib, the increase in viability in the context of IL4 stimulation is larger than in non-trisomy 12 and IGHV-M samples, and thus these features are assigned positive coefficients. This highlights the breadth of IL4-induced resistance to ibrutinib in CLL across genetic backgrounds, even where genetic features would indicate that ibrutinib efficacy should be high, and underlines the importance of understanding the activity of this resistance pathway *in vivo* (Figure 6.11, Figure 6.12, (Bruch et al. 2021)).

this is a resource, all models , and drug - stimuli responses stratified by genetic features can be explored:

6.4 Summary

This work represents the first large - scale attempt to integrate the effects of mutations and the microenvironment in drug response in lymphoma, and to the best of my knowledge, in other cancers. The results provide a proof-of-principle of using combinatorial screening, molecular profiling and statistical modelling can generate more complex biological insights to guide further clinical studies.

More specifically, these results identify a number of key drug-resistance pathways, including the impact of IFN γ on ibrutinib. I highlight the importance of the NOS pathway in IL4 and IFN γ inhibitor of apoptosis. I also show how some of these resistance pathways are context-dependent. In particular, the ability of TLR stimulation to induce resistance to chemotherapeutics is modulated by many genetic features. THis warrants further follow-up.

Overall these findings illustrate how the presence of known genetic alterations determines how drugs and external stimuli interact with each other.

6.5 Discussion

IL4and ibrutinib- what's known implications of ifn gamma and also p38 and ralitmetinib and ifn gamma

strategies for targeting these - pyridone 6, but not licensed? BET inhibition - paper showing that bet specifically inhibits ifn and il4 targets - so maybe something more broad spectrum is useful

implciationsfor studying drug action ex vivo In addition, we aimed to demonstrate that combinatorial screening in conjunction with linear modelling is a powerful tool to study the impact of the microenvironment on drug response in cancer, which is often under-appreciated relative to the impact of molecular features.

In addition, we aimed to demonstrate that combinatorial screening in combination with linear modelling is a powerful approach to study the impact of the microenvironment on drug response in cancer, which is often under-appreciated relative to the impact of molecular features.

How each category might be useful

the importanceof the resource here

OUR approach distinguished between baseline and interaction s- this is a powerful proof of principle cancer biology is integratively influenced by cell-intrinsic and cell-extrinsic features <- what attempts have been, persuade why this needs to continue, how can this be helpful in the clinic

MUST READ: BET inhibitors and suppression of cytokine signalling : BET Bromodomain Inhibition Suppresses Transcriptional Responses to Cytokine-Jak-STAT Signaling in a Gene-specific Manner in Human Monocytes

Nutline and IBET nutlin-3ain trisomy 12 - MDM2 is higher

also IBET - these also correlate - its possible that they work via a similar mechanism - ibet inhibits brd4 which is inhibiting p53 targets, so p53 is more active in these cells

<https://www.nature.com/articles/s41467-020-20378-8>

NOS pathway

TLR - what's going on in vivo, is it really having this bipolar effect in vivo Fludarabine response - does this help to explain - more work needed

In Vitro Sensitivity of CLL Cells to Fludarabine May Be Modulated by the Stimulation of Toll-like Receptors Eleonora Fonte

IL4-induced resistance- this is clearly important across a range of drugs and genetic background, more needed in vivo, possible combinatorial therapies

in vivo IL4 evidence: from T-cells in chronic lymphocytic leukemia: Guardians or drivers of disease? Philipp M. Roessner & Martina Seiffert

Leukemia volume 34, pages 2012–2024 (2020) In 2005, Görgün and colleagues investigated T-cell function in CLL. Gene expression profiling (GEP) of sorted T-cells from peripheral blood mononuclear cells (PBMCs) of CLL patients suggested that CD4+ T-cells are skewed toward a Th2 phenotype due to decreased expression of c-Jun NH₂-terminal kinase (JNK) and p38 mitogen-activated protein kinase (MAPK) pathway-related genes [12]. This is supported by an increased number of interleukin (IL)-4 secreting CD4+ T-cells in PBMCs of CLL patients in comparison to cells of healthy controls (HC). Studies using the immunocompetent Et-TCL1 mouse model of CLL and after adoptive transfer of TCL1 leukemia (TCL1 AT) into BL/6 wild-type (WT) mice revealed more IL-4 secretion of CD4+ T-cells in comparison to control mice [13, 14]. Similarly, conditioned medium of CLL cells was able to induce secretion of IL-4 in allogeneic mixed lymphocyte reactions (MLR) of healthy individuals, suggesting that soluble factors secreted by CLL cells trigger Th2 T-cell polarization in vitro [15].

in other cancers we should integrate drugs - stimuli and genetics

fludarabine and CpG, different outcomes in IGHV M and U patients

the importance of IL4 in all patient backgrounds

TLR Signaling Is Activated in Lymph Node–Resident CLL Cells and Is Only Partially Inhibited by Ibrutinib Eman L. Dadashian <https://cancerres.aacrjournals.org/content/79/2/360.short>

There are models to investigate ibrutinib resistance in vivo, but they are lacking in form
ME: Development and characterization of prototypes for in vitro and in vivo mouse models of ibrutinib-resistant CLL Burcu Aslan

also check out discussion from :<https://www.sciencedirect.com/science/article/pii/S0006497119402206>

6.6 Contributions Statement

The author of this thesis performed the statistical inference and regression modelling outlined in this chapter. In essence, she conceptualised and performed the linear modelling and multivariate modelling in Equations In addition, she performed univariate analysis of genetic determinants of drug response and data visualisation and investigation of hits.

Peter-Martin Bruch generated the viability screening data used in the model. The genetic annotation data was taken from (Dietrich et al. 2017). The categorisation of drug - stimulus interactions, and investigation of examples of drug - stimulus interactions was performed in collaboration with Peter-Martin Bruch for the paper Bruch & Giles et al 2021. The author also received support from Frederich Zwiebell and Junyan Lu, who provided advice with regards to modelling approaches.

- check pdf and tidy referencing linear model figures tidy extra text
- check of text = 6 done!

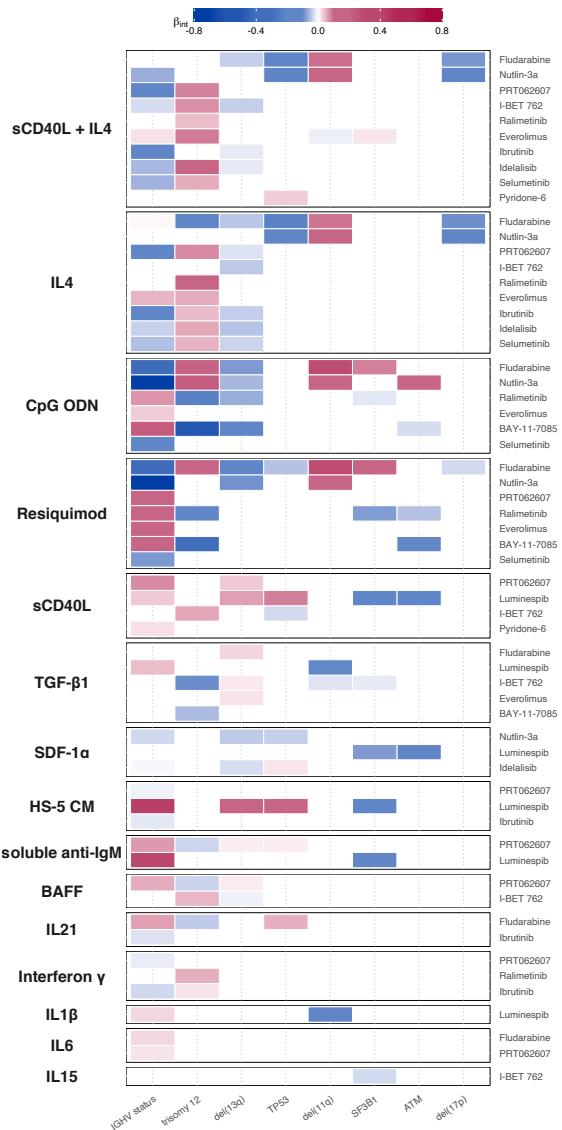


Figure 6.6: Heatmap summarising genetic predictors of drug - stimulus interactions. Each row depicts a single drug - stimulus combination, and each tile indicates that β_{int} for given drug and stimulus combination is modulated by corresponding genetic feature. Thus each row represents the output of a single model fit (as in Figure 6.7). Colour of tile indicates size and sign of coefficient assigned to genetic feature, where a positive coefficient corresponds to a more positive β_{int} if the feature is present. Only top 8 most commonly selected genetic features are shown. Drug - stimulus combinations with no genetic predictors of β_{int} amongst top 8 shown are omitted for clarity. See Methods section 2.5.2. *Figure from (Bruch et al. 2021).*

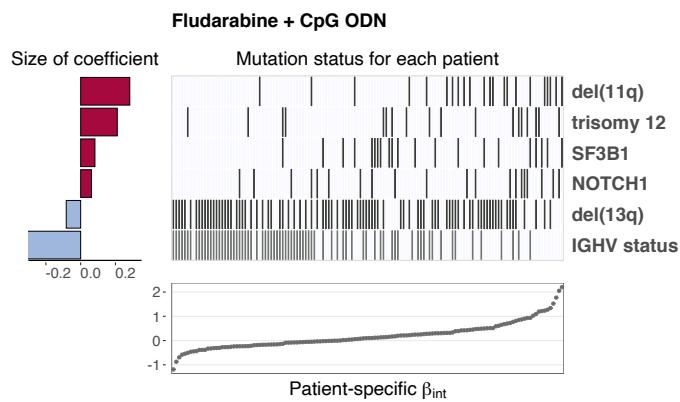


Figure 6.7: Genetic features that modulate the size of β_{int} between fludarabine and CpG ODN. Scatter plot, heatmap and bar plots indicate inputs and outputs of equation (6.2). Bar plots indicate size and sign of coefficients assigned to genetic features named on right. Scatter plot depicts β_{int} values for each patient sample i.e. response matrix. Heatmap tiles indicate mutation status for the selected genetic features (i.e. feature matrix) corresponding to same sample in scatter plot below, to show how size of β_{int} varies with each feature. See Methods section 2.5.2.
 Figure and caption from (Giles2021.?)

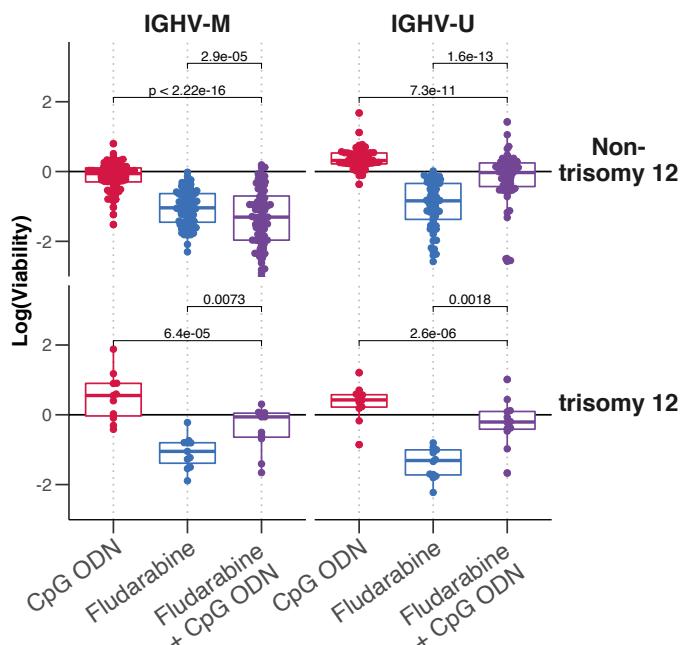


Figure 6.8: Beeswarm-boxplots of log(viability) values for fludarabine and CpG ODN single and combinatorial treatments, faceted by IGHV status and trisomy 12 status. P-values from paired Student's t-tests.
 Figure and caption from (Giles2021.?)

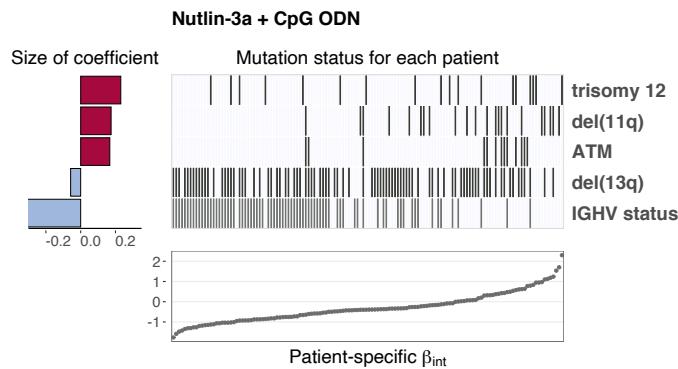


Figure 6.9: Predictor profile depicting genetic features that modulate the size of β_{int} between Nutlin-3a (MDM2) and CpG ODN. Scatter plot (bottom) and heatmap (right) and bar plots (left) indicate inputs and outputs of fitting model in Equation (6.2). Bar plots indicate size and sign of coefficients assigned to genetic features named on the right. Scatter plot depicts ranked patient-specific β_{int} values i.e. the response matrix. Heatmap above indicates patient mutation status for the selected genetic features i.e. feature matrix. Heatmap tiles correspond to points in scatter plot (i.e. patient data is aligned), to indicate how the size of β_{int} varies with selected genetic feature. See Methods section 2.5.2. Figure and caption from (Giles2021.?)

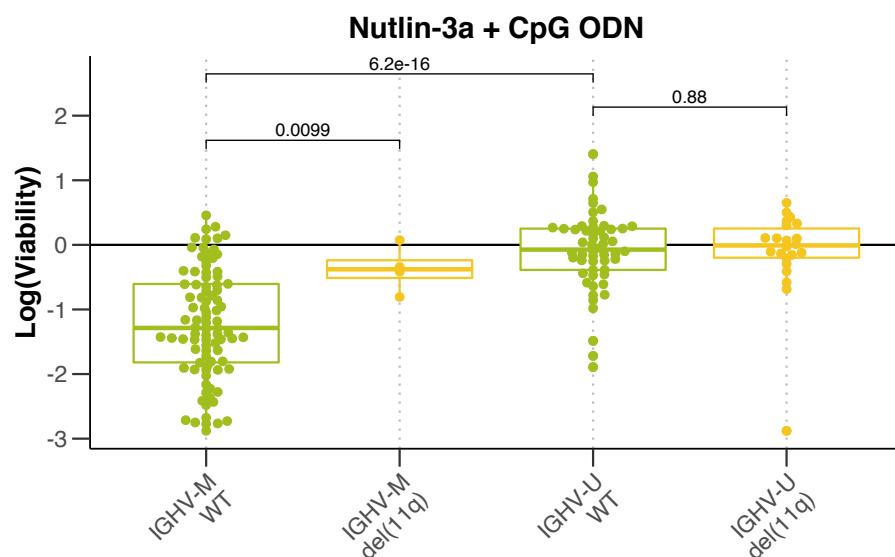


Figure 6.10: Beeswarm-boxplot showing control-normalised log transformed viability values, after treatment with CpG ODN + Nutlin-3a, stratified by del(11q) and IGHV status. p-values from Student's t-tests. *Figure from (Bruch et al. 2021)*

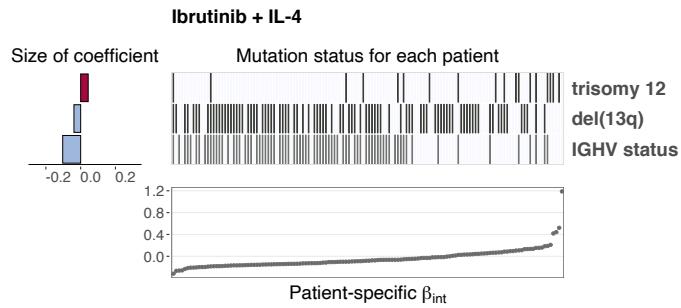


Figure 6.11: Predictor profile depicting genetic features that modulate the size of β_{int} between ibrutinib (BTK) and IL4. Plot generated as in Figure 6.7. See also Methods section 2.5.2. _Figure from (Giles2021._?)

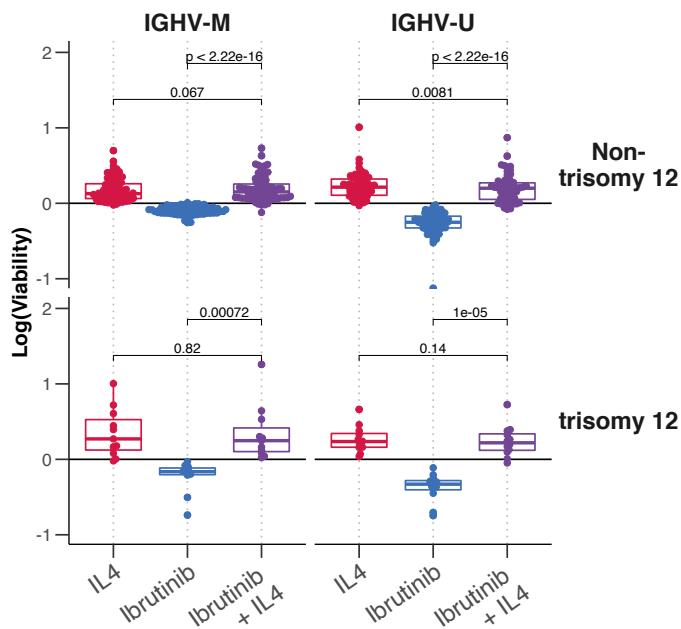


Figure 6.12: Beeswarm boxplots of log(viability) values for ibrutinib and IL4 single and combinatorial treatments, faceted by IGHV status and trisomy 12 status. P-values from paired Student's t-tests. _Figure and caption from (Giles2021._?)

Chapter 7

Results

7.0.1 IL4 induces resistance to BCR inhibitors and chemotherapeutics (Figure 7.1).

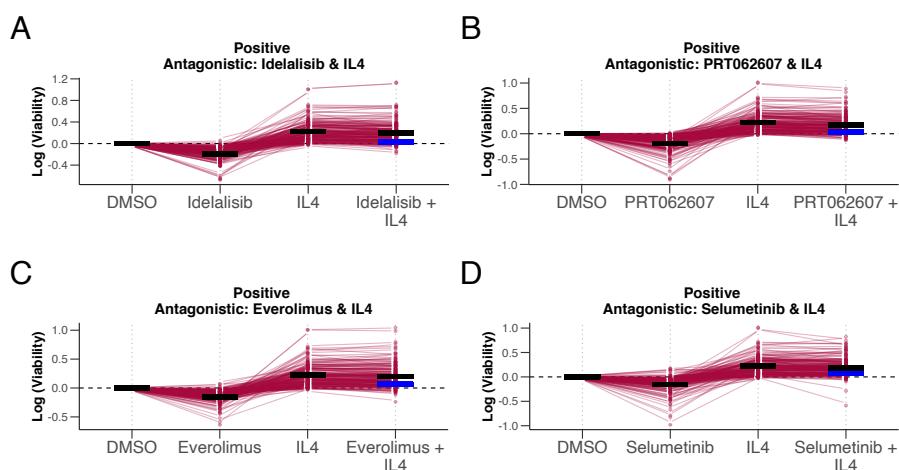


Figure 7.1: Examples of IL4-induced resistance. Plots show log transformed viability values with each treatment, for all samples. Matching samples are linked across treatments. Black and blue horizontal lines indicate the predicted viability from the linear model for each single and combinatorial treatment. In combinatorial treatment, both the expected viability based on the additive effect of drug and stimulus (blue), and the viability with interaction (black) are shown, to indicate the impact of the interaction. Plots show IL4 + (A) Idelalisib (Pi3K), (B) PRT062607 (SYK) (C) Everolimus (mTOR), and (D) Selumetinib (p38 MAPK). See Methods section 2.5.2. *Figure from Bruch & Giles et al. 2021.*

- 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.

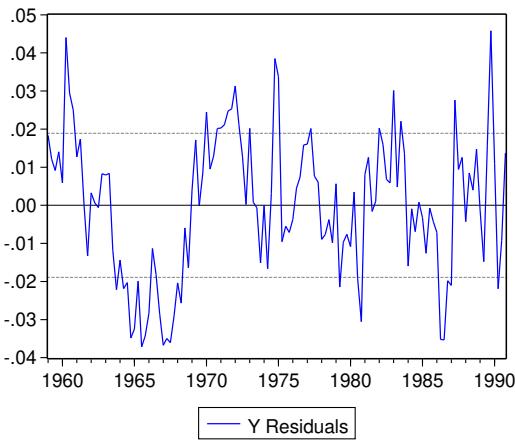


Figure 7.2: Estimated residuals from model XXX4. ...

- 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*!
 - *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>.
- Agathangelidis, Andreas, Nikos Darzentas, Anastasia Hadzidimitriou, Xavier Brochet, Fiona Murray, Xiao Jie Yan, Zadie Davis, et al. 2012. “Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: A molecular classification with implications for targeted therapies.” *Blood* 119 (19): 4467–75. <https://doi.org/10.1182/blood-2011-11-393694>.
- Aguilar-Hernandez, Maria M, Matthew D Blunt, Rachel Dobson, Alison Yeomans, Stephen Thirdborough, Marta Larrayoz, Lindsay D Smith, et al. n.d. “IL-4 enhances expression and function of surface IgM in CLL cells.” <https://doi.org/10.1182/blood-2015-11>.
- Ahn, Inhye E, Mohammed Z H Farooqui, Xin Tian, Janet Valdez, Clare Sun, Susan Soto, Jennifer Lotter, et al. 2018. “Depth and durability of response to ibrutinib in CLL: 5-year follow-up of a phase 2 study.” *Blood* 131 (21): 2357–66. <https://doi.org/10.1182/blood-2017-12-820910>.
- 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/>.
- Alexandrov, Ludmil B., Serena Nik-Zainal, David C. Wedge, Samuel A. J. R. Aparicio, Sam Behjati, Andrew V. Biankin, Graham R. Bignell, et al. 2013. “Signatures of

mutational processes in human cancer." *Nature* 500 (7463): 415–21. <https://doi.org/10.1038/nature12477>.

Allen, Christopher D. C., and Jason G. Cyster. 2008. "Follicular dendritic cell networks of primary follicles and germinal centers: Phenotype and function." *Seminars in Immunology* 20 (1): 14–25. <https://doi.org/10.1016/J.SMIM.2007.12.001>.

Amigo-Jiménez, Irene, Elvira Bailón, Noemí Aguilera-Montilla, María José Terol, José A García-Marco, and Angeles García-Pardo. 2015. "Bone marrow stroma-induced resistance of chronic lymphocytic leukemia cells to arsenic trioxide involves Mcl-1 up-regulation and is overcome by inhibiting the PI3K δ or PKC β signaling pathways." *Oncotarget* 6 (42): 44832–48. <https://doi.org/10.18632/oncotarget.6265>.

Anda-Jáuregui, Guillermo de, and Enrique Hernández-Lemus. 2020. "Computational Oncology in the Multi-Omics Era: State of the Art." *Frontiers*. <https://doi.org/10.3389/fonc.2020.00423>.

Argelaguet, Ricard, Britta Velten, Damien Arnol, Sascha Dietrich, Thorsten Zenz, John C Marioni, Florian Buettner, Wolfgang Huber, and Oliver Stegle. 2018. "MultiOmics Factor Analysis—a framework for unsupervised integration of multiomics data sets." *Molecular Systems Biology* 14 (6): e8124. <https://doi.org/10.15252/msb.20178124>.

Asslaber, Daniela, Eva M Grössinger, Tamara Girbl, Sebastian W Hofbauer, Alexander Egle, Lukas Weiss, Richard Greil, and Tanja N Hartmann. 2013. "Mimicking the microenvironment in chronic lymphocytic leukaemia - where does the journey go?" *Br J Haematol.* <https://doi.org/10.1111/bjh.12151>.

Austen, Belinda, Judith E Powell, Azra Alvi, Ian Edwards, Laura Hooper, Jane Starczynski, A Malcolm R Taylor, Christopher Fegan, Paul Moss, and Tatjana Stankovic. 2005. "Mutations in the ATM gene lead to impaired overall and treatment-free survival that is independent of IGVH mutation status in patients with B-CLL." *Blood* 106 (9): 3175–82. <https://doi.org/10.1182/blood-2004-11-4516>.

Axelrod, M, Z Ou, L K Brett, L Zhang, E R Lopez, A T Tamayo, V Gordon, et al. 2014. "Combinatorial drug screening identifies synergistic co-targeting of Bruton's tyrosine kinase and the proteasome in mantle cell lymphoma." Nature Publishing Group. <https://doi.org/10.1038/leu.2013.249>.

Bagnara, Davide, Matthew S. Kaufman, Carlo Calissano, Sonia Marsilio, Piers E. M. Patten, Rita Simone, Philip Chum, et al. 2011. "A novel adoptive transfer model of

chronic lymphocytic leukemia suggests a key role for T lymphocytes in the disease.” *Blood* 117 (20): 5463–72. <https://doi.org/10.1182/blood-2010-12-324210>.

Balatti, Veronica, Arianna Bottini, Alexey Palamarchuk, Hansjuerg Alder, Laura Z. Rassenti, Thomas J. Kipps, Yuri Pekarsky, and Carlo M. Croce. 2012. “NOTCH1 mutations in CLL associated with trisomy 12.” *Blood* 119 (2): 329–31. <https://doi.org/10.1182/blood-2011-10-386144>.

Barretina, Jordi, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A. Margolin, Sungjoon Kim, Christopher J. Wilson, et al. 2012. “The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.” *Nature* 483 (7391): 603–7. <https://doi.org/10.1038/nature11003>.

Basu, Amrita, Nicole E Bodycombe, Jaime H Cheah, Edmund V Price, Ke Liu, Giannina I Schaefer, Richard Y Ebright, et al. 2013. “XAn interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules.” *Cell* 154 (5): 1151–61. <https://doi.org/10.1016/j.cell.2013.08.003>.

Baumann, Tycho, Riccardo Moia, Gianluca Gaidano, Julio Delgado, Adalgisa Condoluci, Neus Villamor, Anil Babu Payedimarri, et al. 2021. “Lymphocyte doubling time in chronic lymphocytic leukemia modern era: a real-life study in 848 unselected patients.” *Leukemia* 35 (8): 2325–31. <https://doi.org/10.1038/s41375-021-01149-w>.

Beekman, Renée, Vicente Chapaprieta, Núria Russiñol, Roser Vilarrasa-Blasi, Núria Verdaguer-Dot, Joost H. A. Martens, Martí Duran-Ferrer, et al. 2018. “The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia.” *Nature Medicine* 24 (6): 868–80. <https://doi.org/10.1038/s41591-018-0028-4>.

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>.

Bhattacharya, Nupur, Michaela Reichenzeller, Maiwen Caudron-Herger, Sarah Haebe, Nathan Brady, Susanne Diener, Maria Nothing, et al. 2015. “Loss of cooperativity of secreted CD40L and increased dose-response to IL4 on CLL cell viability correlates with enhanced activation of NF- κ B and STAT6.” *International Journal of Cancer* 136 (1): 65–73. <https://doi.org/10.1002/ijc.28974>.

- Bhoi, Sujata, Viktor Ljungström, Panagiotis Baliakas, Mattias Mattsson, Karin E Smedby, Gunnar Juliusson, Richard Rosenquist, and Larry Mansouri. 2016. *Epigenetics* 11 (6): 449–55. <https://doi.org/10.1080/15592294.2016.1178432>.
- Binder, Mascha, Barbara Léchenne, Ramesh Ummanni, Christian Scharf, Stefan Balabanov, Maria Trusch, Hartmut Schlüter, Ingke Braren, Edzard Spillner, and Martin Trepel. 2010. “Stereotypical Chronic Lymphocytic Leukemia B-Cell Receptors Recognize Survival Promoting Antigens on Stromal Cells.” *PLOS ONE* 5 (12): e15992. <https://doi.org/10.1371/JOURNAL.PONE.0015992>.
- Boissard, F, J. J. Fournié, A Quillet-Mary, L Ysebaert, and M Poupot. 2015. “Nurse-like cells mediate ibrutinib resistance in chronic lymphocytic leukemia patients.” *Blood Cancer Journal* 5 (10): e355. <https://doi.org/10.1038/bcj.2015.74>.
- Borche, L, A Lim, JL Binet, and G Dighiero. 1990. “Evidence that chronic lymphocytic leukemia B lymphocytes are frequently committed to production of natural autoantibodies.” *Blood* 76 (3): 562–69. <https://doi.org/10.1182/blood.v76.3.562.562>.
- 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>.
- Bröker, Barbara M., Adek Klajman, Pierre Youinou, Jean Jouquan, Colin P. Worman, John Murphy, Lorna Mackenzie, et al. 1988. “Chronic lymphocytic leukemic (CLL) cells secrete multispecific autoantibodies.” *Journal of Autoimmunity* 1 (5): 469–81. [https://doi.org/10.1016/0896-8411\(88\)90068-6](https://doi.org/10.1016/0896-8411(88)90068-6).
- Brown, Jennifer R, John C Byrd, Steven E Coutre, Don M Benson, Ian W Flinn, Nina D Wagner-Johnston, Stephen E Spurgeon, et al. 2014. “Idelalisib, an inhibitor of phosphatidylinositol 3-kinase p110 δ , for relapsed/refractory chronic lymphocytic leukemia.” *Blood* 123 (22): 3390–97. <https://doi.org/10.1182/blood-2013-11-535047>.
- Bruch, Peter-Martin, Holly A. R. Giles, Carolin Kolb, Sophie A. Herbst, Tina Becirovic, Tobias Roider, Junyan Lu, et al. 2021. “Mapping drug-microenvironment-genetic interplay in CLL reveals trisomy 12 as a modulator of microenvironmental signals.” *bioRxiv Cancer Biology*, July, 2021.07.23.453514. <https://doi.org/10.1101/2021.07.23.453514>.
- Burger, J A, N J Zvaifler, N Tsukada, G S Firestein, and T J Kipps. 2001. “Fibroblast-like

synoviocytes support B-cell pseudoemperipolesis via a stromal cell-derived factor-1- and CD106 (VCAM-1)-dependent mechanism.” *The Journal of Clinical Investigation* 107 (3): 305–15. <https://doi.org/10.1172/JCI11092>.

Burger, Jan A. 2020. “Treatment of Chronic Lymphocytic Leukemia.” Edited by Dan L. Longo. *New England Journal of Medicine* 383 (5): 460–73. <https://doi.org/10.1056/NEJMra1908213>.

Burger, Jan A., Meike Burger, and Thomas J. Kipps. 1999. “Chronic Lymphocytic Leukemia B Cells Express Functional Cxcr4 Chemokine Receptors That Mediate Spontaneous Migration Beneath Bone Marrow Stromal Cells.” *Blood* 94 (11): 3658–67. <https://doi.org/10.1182/BLOOD.V94.11.3658>.

Burger, Jan A., and Nicholas Chiorazzi. 2013. “B cell receptor signaling in chronic lymphocytic leukemia.” *Trends Immunol.* <https://doi.org/10.1016/j.it.2013.07.002>.

Burger, Jan A., and John G Gribben. 2014. “The microenvironment in chronic lymphocytic leukemia (CLL) and other B cell malignancies: insight into disease biology and new targeted therapies.” *Seminars in Cancer Biology* 24 (February): 71–81. <https://doi.org/10.1016/j.semcaner.2013.08.011>.

Burger, Jan A., Nobuhiro Tsukada, Meike Burger, Nathan J. Zvaifler, Marie Dell’Aquila, and Thomas J. Kipps. 2000. “Blood-derived nurse-like cells protect chronic lymphocytic leukemia B cells from spontaneous apoptosis through stromal cell-derived factor-1.” *Blood* 96 (8): 2655–63. <https://doi.org/10.1182/BLOOD.V96.8.2655>.

Burger, Meike, Tanja Hartmann, Myriam Krome, Justyna Rawluk, Hirokazu Tamamura, Nobutaka Fujii, Thomas J. Kipps, and Jan A. Burger. 2005. “Small peptide inhibitors of the CXCR4 chemokine receptor (CD184) antagonize the activation, migration, and antiapoptotic responses of CXCL12 in chronic lymphocytic leukemia B cells.” *Blood* 106 (5): 1824–30. <https://doi.org/10.1182/blood-2004-12-4918>.

Bürkle, Andrea, Matthias Niedermeier, Annette Schmitt-Gräff, William G. Wierda, Michael J. Keating, and Jan A. Burger. 2007. “Overexpression of the CXCR5 chemokine receptor, and its ligand, CXCL13 in B-cell chronic lymphocytic leukemia.” *Blood* 110 (9): 3316–25. <https://doi.org/10.1182/blood-2007-05-089409>.

Byrd, John C., Jennifer R. Brown, Susan O’Brien, Jacqueline C. Barrientos, Neil E. Kay, Nishitha M. Reddy, Steven Coutre, et al. 2014. “Ibrutinib versus Ofatumumab in

Previously Treated Chronic Lymphoid Leukemia.” *New England Journal of Medicine* 371 (3): 213–23. <https://doi.org/10.1056/nejmoa1400376>.

Byrd, John C, Richard R Furman, Steven E Coutre, Ian W Flinn, Jan A Burger, Kristie A Blum, Barbara Grant, et al. 2013. “Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia.” *The New England Journal of Medicine* 369 (1): 32–42. <https://doi.org/10.1056/NEJMoa1215637>.

Cahill, N, A. C. Bergh, M Kanduri, H Göransson-Kultima, L Mansouri, A Isaksson, F Ryan, et al. 2013. “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* 27 (1): 150–58. <https://doi.org/10.1038/leu.2012.245>.

Calin, George Adrian, Calin Dan Dumitru, Masayoshi Shimizu, Roberta Bichi, Simona Zupo, Evan Noch, Hansjuerg Aldler, et al. 2002. “Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (24): 15524–29. <https://doi.org/10.1073/pnas.242606799>.

Campbell, Belinda A, Julia J Scarisbrick, Youn H Kim, Ryan A Wilcox, Christopher McCormack, and H Miles Prince. 2020. “Time to next treatment as a meaningful endpoint for trials of primary cutaneous lymphoma.” *Cancers* 12 (8): 1–11. <https://doi.org/10.3390/cancers12082311>.

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 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>.

Carey, Alyssa, David K Edwards, Christopher A Eide, Laura Newell, Elie Traer, Bruno C Medeiros, Daniel A Polleyea, et al. 2017. “Identification of Interleukin-1 by Functional Screening as a Key Mediator of Cellular Expansion and Disease Progression in Acute Myeloid Leukemia.” *Cell Reports* 18 (13): 3204–18. <https://doi.org/10.1016/j.celrep.2017.03.018>.

Chan, Chun Hin, Celestia Fang, Anna Yarilina, Rab K Prinjha, Yu Qiao, and Lionel B Ivashkiv. 2015. “BET bromodomain inhibition suppresses transcriptional responses to cytokine-Jak-STAT signaling in a gene-specific manner in human monocytes.” *Eu-*

ropean Journal of Immunology 45 (1): 287–97. <https://doi.org/10.1002/eji.201444862>.

Chang, Julie Elizabeth, and Brad Steven Kahl. 2012. “Bendamustine for treatment of chronic lymphocytic leukemia.” *Expert Opinion on Pharmacotherapy* 13 (10): 1495–505. <https://doi.org/10.1517/14656566.2012.693163>.

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>.

Cheng, S, J Ma, A Guo, P Lu, J P Leonard, M Coleman, M Liu, J J Buggy, R R Furman, and Y L Wang. 2014. “BTK inhibition targets in vivo CLL proliferation through its effects on B-cell receptor signaling activity.” *Leukemia* 28 (3): 649–57. <https://doi.org/10.1038/leu.2013.358>.

Chiorazzi, Nicholas, and Manlio Ferrarini. 2003. “B cell chronic lymphocytic leukemia: Lessons learned from studies of the B cell antigen receptor.” Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA. <https://doi.org/10.1146/annurev.immunol.21.120601.141018>.

Chiorazzi, Nicholas, Kanti R Rai, and Manlio Ferrarini. 2005. “Chronic lymphocytic leukemia.” *The New England Journal of Medicine* 352 (8): 804–15. <https://doi.org/10.1056/NEJMra041720>.

Chu, Charles C., Rosa Catera, Katerina Hatzi, Xiao Jie Yan, Lu Zhang, Xiao Bo Wang, Henry M. Fales, et al. 2008. “Chronic lymphocytic leukemia antibodies with a common stereotypic rearrangement recognize nonmuscle myosin heavy chain IIA.” *Blood* 112 (13): 5122–29. <https://doi.org/10.1182/blood-2008-06-162024>.

Chung, James B., Michael Silverman, and John G. Monroe. 2003. “Transitional B cells: step by step towards immune competence.” *Trends in Immunology* 24 (6): 342–48. [https://doi.org/10.1016/S1471-4906\(03\)00119-4](https://doi.org/10.1016/S1471-4906(03)00119-4).

Cimmino, Amelia, George Adrian Calin, Muller Fabbri, Marilena V Iorio, Manuela Fer racin, Masayoshi Shimizu, Sylwia E Wojcik, et al. 2005. “miR-15 and miR-16 induce apoptosis by targeting BCL2.” *Proceedings of the National Academy of Sciences*

of the United States of America 102 (39): 13944–49. <https://doi.org/10.1073/pnas.0506654102>.

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 Bchronic 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>.

Crassini, Kyle, Yandong Shen, Stephen Mulligan, and O. Giles Best. 2017. “Modeling the chronic lymphocytic leukemia microenvironment in vitro.” *Leukemia and Lymphoma* 58 (2): 266–79. <https://doi.org/10.1080/10428194.2016.1204654>.

Cromptot, Emerence, Michael Van Damme, Karljen Pieters, Marjorie Vermeersch, David Perez-Morga, Philippe Mineur, Marie Maerevoet, et al. 2017. “Extracellular vesicles of bone marrow stromal cells rescue chronic lymphocytic leukemia B cells from apoptosis, enhance their migration and induce gene expression modifications.” *Haematologica* 102 (9): 1594–604. <https://doi.org/10.3324/haematol.2016.163337>.

Damle, Rajendra N., Tarun Wasil, Franco Fais, Fabio Ghiotto, Angelo Valetto, Steven L. Allen, Aby Buchbinder, et al. 1999. “Ig V Gene Mutation Status and CD38 Expression As Novel Prognostic Indicators in Chronic Lymphocytic Leukemia.” *Blood* 94 (6): 1840–47. <https://doi.org/10.1182/BLOOD.V94.6.1840>.

Dancescu, M, M Rubio-Trujillo, G Biron, D Bron, G Delespesse, and M Sarfati. 1992. “Interleukin 4 protects chronic lymphocytic leukemic B cells from death by apoptosis and upregulates Bcl-2 expression.” *The Journal of Experimental Medicine* 176 (5): 1319–26. <https://doi.org/10.1084/jem.176.5.1319>.

De Rooij, Martin F. M., Annemieke Kuil, Christian R Geest, Eric Eldering, Betty Y Chang, Joseph J Buggy, Steven T Pals, and Marcel Spaargaren. 2012. “The clinically active BTK inhibitor PCI-32765 targets B-cell receptor- and chemokine-controlled adhesion and migration in chronic lymphocytic leukemia.” *Blood* 119 (11): 2590–94. <https://doi.org/10.1182/blood-2011-11-390989>.

Deaglio, Silvia, and Fabio Malavasi. 2009. “Chronic lymphocytic leukemia microenvironment: Shifting the balance from apoptosis to proliferation.” <https://doi.org/10.3324/haematol.2009.006676>.

Decker, Thomas, Christian Bogner, Madlen Oelsner, Christian Peschel, and Ingo Ring-

shausen. 2010. "Antiapoptotic effect of interleukin-2 (IL-2) in B-CLL cells with low and high affinity IL-2 receptors." *Annals of Hematology* 89 (11): 1125–32. <https://doi.org/10.1007/s00277-010-0994-1>.

Defoiche, Julien, Christophe Debacq, Becca Asquith, Yan Zhang, Arsène Burny, Dominique Bron, Laurence Lagneaux, Derek Macallan, and Luc Willems. 2008. "Reduction of B cell turnover in chronic lymphocytic leukaemia." *British Journal of Haematology* 143 (2): 240–47. <https://doi.org/10.1111/j.1365-2141.2008.07348.x>.

Del Giudice, Ilaria, Davide Rossi, Sabina Chiaretti, Marilisa Marinelli, Simona Tavolaro, Sara Gabrielli, Luca Laurenti, et al. 2012. "NOTCH1 mutations in +12 chronic lymphocytic leukemia (CLL) confer an unfavorable prognosis, induce a distinctive transcriptional profiling and refine the intermediate prognosis of +12 CLL." *Haematologica* 97 (3): 437–41. <https://doi.org/10.3324/haematol.2011.060129>.

Delgado, Amanda, and Achuta Kumar Guddati. 2021. "Clinical endpoints in oncology - a primer." *American Journal of Cancer Research* 11 (4): 1121–31. <http://www.ncbi.nlm.nih.gov/pubmed/33948349%20http://www.ncbi.nlm.nih.gov/pmc/articles/PMC8085844/>

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>.

Ding, Wei, Grzegorz S. Nowakowski, Traci R. Knox, Justin C. Boysen, Mary L. Maas, Susan M. Schwager, Wenting Wu, et al. 2009. "Bi-directional activation between mesenchymal stem cells and CLL B-cells: implication for CLL disease progression." *British Journal of Haematology* 147 (4): 471–83. <https://doi.org/10.1111/j.1365-2141.2009.07868.x>.

Döhner, Hartmut, Konstanze Fischer, Martin Bentz, Katrin Hansen, Axel Benner, Georges Cabot, Daniela Diehl, et al. 1995. "p53 Gene Deletion Predicts for Poor Survival and Non-Response to Therapy With Purine Analogs in Chronic B-Cell Leukemias." *Blood* 85 (6): 1580–89. <https://doi.org/10.1182/BLOOD.V85.6.1580.BLOODJOURNAL8561580>.

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/>.
- Du, W, and O Elemento. 2015. “Cancer systems biology: Embracing complexity to develop better anticancer therapeutic strategies.” *Oncogene* 34 (25): 3215–25. <https://doi.org/10.1038/onc.2014.291>.
- Dubois, Nathan, Emerence Cromptot, Nathalie Meuleman, Dominique Bron, Laurence Lagneaux, and Basile Stamatopoulos. 2020. “Importance of Crosstalk Between Chronic Lymphocytic Leukemia Cells and the Stromal Microenvironment: Direct Contact, Soluble Factors, and Extracellular Vesicles.” Frontiers Media SA. <https://doi.org/10.3389/fonc.2020.01422>.
- Elston, Lauren, Chris Fegan, Robert Hills, Shaikh S Hashimdeen, Elisabeth Walsby, Peter Henley, Chris Pepper, and Stephen Man. 2020. “Increased frequency of CD4+ PD-1+ HLA-DR+ T cells is associated with disease progression in CLL.” *British Journal of Haematology* 188 (6): 872–80. <https://doi.org/10.1111/bjh.16260>.
- Enzler, Thomas, Arnon P Kater, Weizhou Zhang, George F Widhopf, Han Yu Chuang, Jason Lee, Esther Avery, Carlo M Croce, Michael Karin, and Thomas J Kipps. 2009. “Chronic lymphocytic leukemia of E μ -TCL1 transgenic mice undergoes rapid cell turnover that can be offset by extrinsic CD257 to accelerate disease progression.” *Blood* 114 (20): 4469–76. <https://doi.org/10.1182/blood-2009-06-230169>.
- Fabbri, Giulia, and Riccardo Dalla-Favera. 2016. “The molecular pathogenesis of chronic lymphocytic leukaemia.” *Nat Rev Cancer*. <https://doi.org/10.1038/nrc.2016.8>.
- Fabbri, Giulia, Silvia Rasi, Davide Rossi, Vladimir Trifonov, Hossein Khiabanian, Jing Ma, Adina Grunn, et al. 2011. “Analysis of the chronic lymphocytic leukemia coding genome: Role of NOTCH1 mutational activation.” *Journal of Experimental Medicine* 208 (7): 1389–1401. <https://doi.org/10.1084/jem.20110921>.
- Fais, Franco, Fabio Ghiotto, Shiori Hashimoto, Brian Sellars, Angelo Valetto, Steven L Allen, Philip Schulman, et al. 1998. “Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors.” *Journal of Clinical Investigation* 102 (8): 1515–25. <https://doi.org/10.1172/JCI3009>.
- Farinello, Diego, Monika Woziska, Elisa Lenti, Luca Genovese, Silvia Bianchessi, Edoardo

- Migliori, Nicolò Sacchetti, et al. 2018. “A retinoic acid-dependent stroma-leukemia crosstalk promotes chronic lymphocytic leukemia progression.” *Nature Communications* 9 (1): 1787. <https://doi.org/10.1038/s41467-018-04150-7>.
- Fedorchenko, Oleg, Marius Stiefelhagen, Abdul A. Peer-Zada, Romy Barthel, Petra Mayer, Laura Eckei, Alexandra Breuer, et al. 2013. “CD44 regulates the apoptotic response and promotes disease development in chronic lymphocytic leukemia.” *Blood* 121 (20): 4126–36. <https://doi.org/10.1182/blood-2012-11-466250>.
- Ferreira, Pedro G, Pedro Jares, Daniel Rico, Gonzalo Gómez-López, Alejandra Martínez-Trillo, Neus Villamor, Simone Ecker, et al. 2014. “Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia.” *Genome Research* 24 (2): 212–26. <https://doi.org/10.1101/gr.152132.112>.
- Fischer, Ute, Jun J. Yang, Tomokatsu Ikawa, Daniel Hein, Carolina Vicente-Dueñas, Arndt Borkhardt, and Isidro Sánchez-García. 2020. “Cell Fate Decisions: The Role of Transcription Factors in Early B-cell Development and Leukemia.” *Blood Cancer Discovery* 1 (3): 224–33. <https://doi.org/10.1158/2643-3230.bcd-20-0011>.
- Friedberg, Jonathan W, Jeff Sharman, John Sweetenham, Patrick B Johnston, Julie M Vose, Ann LaCasce, Julia Schaefer-Cutillo, et al. 2010. “Inhibition of Syk with fostamatinib disodium has significant clinical activity in non-Hodgkin lymphoma and chronic lymphocytic leukemia.” *Blood* 115 (13): 2578–85. <https://doi.org/10.1182/blood-2009-08-236471>.
- 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>.
- Furman, Richard R, Jeff P Sharman, Steven E Coutre, Bruce D Cheson, John M Pagel, Peter Hillmen, Jacqueline C Barrientos, et al. 2014. “Idelalisib and Rituximab in Relapsed Chronic Lymphocytic Leukemia.” *New England Journal of Medicine* 370 (11): 997–1007. <https://doi.org/10.1056/nejmoa1315226>.
- Gaiti, Federico, Ronan Chaligne, Hongcang Gu, Ryan M Brand, Steven Kothen-Hill, Rafael C Schulman, Kirill Grigorev, et al. 2019. “Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia.” *Nature* 569 (7757): 576–80. <https://doi.org/10.1038/s41586-019-1198-z>.

- Garnett, Mathew J., Elena J. Edelman, Sonja J. Heidorn, Chris D. Greenman, Anahita Dastur, King Wai Lau, Patricia Greninger, et al. 2012. "Systematic identification of genomic markers of drug sensitivity in cancer cells." *Nature* 483 (7391): 570–75. <https://doi.org/10.1038/nature11005>.
- Ghia, Paolo, Andrzej Pluta, Małgorzata Wach, Daniel Lysak, Tomas Kozak, Martin Simkovic, Polina Kaplan, et al. 2020. "Ascend: Phase III, randomized trial of acalabrutinib versus idelalisib plus rituximab or bendamustine plus rituximab in relapsed or refractory chronic lymphocytic leukemia." In *Journal of Clinical Oncology*, 38:2849–61. 25. J Clin Oncol. <https://doi.org/10.1200/JCO.19.03355>.
- Goede, Valentin, Kirsten Fischer, Raymonde Busch, Anja Engelke, Barbara Eichhorst, Clemens M. Wendtner, Tatiana Chagorova, et al. 2014. "Obinutuzumab plus Chlorambucil in Patients with CLL and Coexisting Conditions." *New England Journal of Medicine* 370 (12): 1101–10. <https://doi.org/10.1056/nejmoa1313984>.
- Gonzalez, David, Pilar Martinez, Rachel Wade, Sarah Hockley, David Oscier, Estella Matutes, Claire E Dearden, Sue M Richards, Daniel Catovsky, and Gareth J Morgan. 2011. "Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: Results from the LRF CLL4 trial." *Journal of Clinical Oncology* 29 (16): 2223–29. <https://doi.org/10.1200/JCO.2010.32.0838>.
- Goodspeed, Andrew, Laura M Heiser, Joe W Gray, and James C Costello. 2016. "Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics." *Molecular Cancer Research : MCR* 14 (1): 3–13. <https://doi.org/10.1158/1541-7786.MCR-15-0189>.
- Granziero, Luisa, Paolo Ghia, Paola Circosta, Daniela Gottardi, Giuliana Strola, Massimo Geuna, Licia Montagna, Paola Piccoli, Marco Chilosi, and Federico Caligaris-Cappio. 2001. "Survivin is expressed on CD40 stimulation and interfaces proliferation and apoptosis in B-cell chronic lymphocytic leukemia." *Blood* 97 (9): 2777–83. <https://doi.org/10.1182/blood.V97.9.2777>.
- Grioni, Matteo, Arianna Brevi, Elena Cattaneo, Alessandra Rovida, Jessica Bordini, Maria Teresa Sabrina Bertilaccio, Maurilio Ponzoni, et al. 2021. "CD4+ T cells sustain aggressive chronic lymphocytic leukemia in E_{μ} -TCL1 mice through a CD40L-independent mechanism." *Blood Advances* 5 (14): 2817–28. <https://doi.org/10.1182/bloodadvances.2020003795>.

- Guarini, Anna, Sabina Chiaretti, Simona Tavolaro, Roberta Maggio, Nadia Peragine, Franca Citarella, Maria Rosaria Ricciardi, et al. 2008. "BCR ligation induced by IgM stimulation results in gene expression and functional changes only in IgVH unmutated chronic lymphocytic leukemia (CLL) cells." *Blood* 112 (3): 782–92. <https://doi.org/10.1182/blood-2007-12-127688>.
- Guo, Ailin, Pin Lu, Greg Coffey, Pamela Conley, Anjali Pandey, and Y Lynn Wang. 2017. "Dual SYK/JAK inhibition overcomes ibrutinib resistance in chronic lymphocytic leukemia: Cerdulatinib, but not ibrutinib, induces apoptosis of tumor cells protected by the microenvironment." *Oncotarget* 8 (8): 12953–67. <https://doi.org/10.18632/oncotarget.14588>.
- Guruharsha, K. G., Mark W. Kankel, and Spyros Artavanis-Tsakonas. 2012. "The Notch signalling system: Recent insights into the complexity of a conserved pathway." Nature Publishing Group. <https://doi.org/10.1038/nrg3272>.
- Hallek, M, K Fischer, G Fingerle-Rowson, A M Fink, R Busch, J Mayer, M Hensel, et al. 2010. "Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: A randomised, open-label, phase 3 trial." *The Lancet* 376 (9747): 1164–74. [https://doi.org/10.1016/S0140-6736\(10\)61381-5](https://doi.org/10.1016/S0140-6736(10)61381-5).
- Hallek, Michael, Bruce D Cheson, Daniel Catovsky, Federico Caligaris-Cappio, Guillaume Dighiero, Hartmut Döhner, Peter Hillmen, et al. 2008. "Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines." *Blood* 111 (12): 5446–56. <https://doi.org/10.1182/blood-2007-06-093906>.
- Hamblin, Terry J., Zadie Davis, Anne Gardiner, David G. Oscier, and Freda K. Stevenson. 1999. "Unmutated Ig VH Genes Are Associated With a More Aggressive Form of Chronic Lymphocytic Leukemia." *Blood* 94 (6): 1848–54. <https://doi.org/10.1182/BLOOD.V94.6.1848>.
- Hamilton, Emma, Laurence Pearce, Liam Morgan, Sophie Robinson, Vicki Ware, Paul Brennan, N Shaun B Thomas, et al. 2012. "Mimicking the tumour microenvironment: Three different co-culture systems induce a similar phenotype but distinct proliferative signals in primary chronic lymphocytic leukaemia cells." *British Journal of Haematology* 158 (5): 589–99. <https://doi.org/10.1111/j.1365-2141.2012.09191.x>.

- Hayden, Rachel E, Guy Pratt, Claudia Roberts, Mark T Drayson, and Chris M Bunce. 2012. "Treatment of chronic lymphocytic leukemia requires targeting of the protective lymph node environment with novel therapeutic approaches." *Leuk Lymphoma*. <https://doi.org/10.3109/10428194.2011.610014>.
- 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>.
- Herman, Sarah E. M., Amber L. Gordon, Erin Hertlein, Asha Ramanunni, Xiaoli Zhang, Samantha Jaglowski, Joseph Flynn, et al. 2011. "Bruton tyrosine kinase represents a promising therapeutic target for treatment of chronic lymphocytic leukemia and is effectively targeted by PCI-32765." *Blood* 117 (23): 6287–96. <https://doi.org/10.1182/blood-2011-01-328484>.
- Hernández-Lemus, Enrique. 2014. "Systems Biology and Integrative Omics in Breast Cancer," 333–52. https://doi.org/10.1007/978-81-322-0843-3_17.
- Herndon, T M, S-S Chen, N S Saba, J Valdez, C Emson, M Gatmaitan, X Tian, et al. 2017. "Direct in vivo evidence for increased proliferation of CLL cells in lymph nodes compared to bone marrow and peripheral blood." *Leukemia* 31 (6): 1340–47. <https://doi.org/10.1038/leu.2017.11>.
- Hervé, Maxime, Kai Xu, Yen Shing Ng, Hedda Wardemann, Emilia Albesiano, Bradley T Messmer, Nicholas Chiorazzi, and Eric Meffre. 2005. "Unmutated and mutated chronic lymphocytic leukemias derive from self-reactive B cell precursors despite expressing different antibody reactivity." *Journal of Clinical Investigation* 115 (6): 1636–43. <https://doi.org/10.1172/JCI24387>.
- Hillmen, Peter, Tadeusz Robak, Ann Janssens, K Govind Babu, Janusz Kloczko, Sebastian Grosicki, Michael Doubek, et al. 2015. "Chlorambucil plus ofatumumab versus chlorambucil alone in previously untreated patients with chronic lymphocytic leukaemia (COMPLEMENT 1): a randomised, multicentre, open-label phase 3 trial." *Lancet (London, England)* 385 (9980): 1873–83. [https://doi.org/10.1016/S0140-6736\(15\)60027-7](https://doi.org/10.1016/S0140-6736(15)60027-7).
- Hoellenriegel, Julia, Sarah A Meadows, Mariela Sivina, William G Wierda, Hagop Kantarjian, Michael J Keating, Neill Giese, et al. 2011. "The phosphoinositide 3-kinase delta inhibitor, CAL-101, inhibits B-cell receptor signaling and chemokine networks

in chronic lymphocytic leukemia.” *Blood* 118 (13): 3603–12. <https://doi.org/10.1182/blood-2011-05-352492>.

Huber, Wolfgang, and Susan Holmes. 2019. “Modern Statistics for Modern Biology.” <https://www.huber.embl.de/msmb/index.html>.

Iacovelli, Stefano, Eva Hug, Sara Bennardo, Marcus Duehren-Von Minden, Stefania Gobessi, Andrea Rinaldi, Mirza Suljagic, et al. 2015. “Two types of BCR interactions are positively selected during leukemia development in the E μ -TCL1 transgenic mouse model of CLL.” *Blood* 125 (10): 1578–88. <https://doi.org/10.1182/blood-2014-07-587790>.

Iorio, Francesco, Theo A Knijnenburg, Daniel J Vis, Graham R Bignell, Michael P Menden, Michael Schubert, Nanne Aben, et al. 2016. “A Landscape of Pharmacogenomic Interactions in Cancer.” *Cell* 166 (3): 740–54. <https://doi.org/10.1016/j.cell.2016.06.017>.

Jayappa, Kallesh D., Konrad J. Cios, Vicki L. Gordon, Puja C. Arora, Timothy P. Bender, Michael E. Williams, Craig A. Portell, and Michael J. Weber. 2018. “A Multi-Drug Tolerant Phenotype Is Induced By Diverse Microenvironmental Agonists in Chronic Lymphocytic Leukemia and Mantle Cell Lymphoma.” *Blood* 132 (Supplement 1): 4120–20. <https://doi.org/10.1182/blood-2018-99-119521>.

Jayappa, Kallesh D, Craig A Portell, Vicki L Gordon, Brian J Capaldo, Stefan Bekiranov, Mark J Axelrod, L Kyle Brett, et al. 2017. “Microenvironmental agonists generate de novo phenotypic resistance to combined ibrutinib plus venetoclax in CLL and MCL.” *Blood Advances* 1 (14): 933–46. <https://doi.org/10.1182/bloodadvances.2016004176>.

Jensen, Caleb, and Yong Teng. 2020. “Is It Time to Start Transitioning From 2D to 3D Cell Culture?” *Frontiers*. <https://doi.org/10.3389/fmolb.2020.00033>.

Kalachikov, S, A Migliazza, E Cayanis, N S Fracchiolla, M F Bonaldo, L Lawton, P Jelenc, et al. 1997. “Cloning and gene mapping of the chromosome 13q14 region deleted in chronic lymphocytic leukemia.” *Genomics* 42 (3): 369–77. <https://doi.org/10.1006/geno.1997.4747>.

Kassambara, Alboukadel, Marcin Kosinski, and Przemyslaw Biecek. 2021. *Survminer: Drawing Survival Curves Using Ggplot2*. <https://rpkgs.datanovia.com/survminer/index.html>.

Kay, Neil E, Tait D Shanafelt, Ann K Strelge, Yean K Lee, Nancy D Bone, and Azra Raza.

2007. "Bone biopsy derived marrow stromal elements rescue chronic lymphocytic leukemia B-cells from spontaneous and drug induced cell death and facilitates an "angiogenic switch"." *Leukemia Research* 31 (7): 899–906. <https://doi.org/10.1016/j.leukres.2006.11.024>.

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, T. J., E. Tomhave, L. F. Pratt, S. Duffy, P. P. Chen, and D. A. Carson. 1989. "Developmentally restricted immunoglobulin heavy chain variable region gene expressed at high frequency in chronic lymphocytic leukemia." *Proceedings of the National Academy of Sciences of the United States of America* 86 (15): 5913–17. <https://doi.org/10.1073/pnas.86.15.5913>.

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>.

Kitada, S, J M Zapata, M Andreeff, and J C Reed. 1999. "Bryostatin and CD40-ligand enhance apoptosis resistance and induce expression of cell survival genes in B-cell chronic lymphocytic leukaemia." *British Journal of Haematology* 106 (4): 995–1004. <https://doi.org/10.1046/j.1365-2141.1999.01642.x>.

Klein, Ulf, Yuhai Tu, Gustavo A Stolovitzky, Michela Mattioli, Giorgio Cattoretti, Hervé Husson, Arnold Freedman, et al. 2001. "Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells." *Journal of Experimental Medicine* 194 (11): 1625–38. <https://doi.org/10.1084/jem.194.11.1625>.

Kolde, Raivo. 2019. *Pheatmap: Pretty Heatmaps*. <https://CRAN.R-project.org/package=pheatmap>.

Kondo, Motonari. 2010. "Lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors." *Immunological Reviews* 238 (1): 37–46. <https://doi.org/10.1111/j.1600-065X.2010.00963.x>.

- Kostareli, E., M. Gounari, A. Janus, F. Murray, X. Brochet, V. Giudicelli, S. Pospisilova, et al. 2012. "Antigen receptor stereotypy across B-cell lymphoproliferations: The case of IGHV4-59/IGKV3-20 receptors with rheumatoid factor activity." *Leukemia*. <https://doi.org/10.1038/leu.2011.311>.
- Kraus, Manfred, Marat B. Alimzhanov, Nikolaus Rajewsky, and Klaus Rajewsky. 2004. "Survival of resting mature B lymphocytes depends on BCR signaling via the Ig α/β heterodimer." *Cell* 117 (6): 787–800. <https://doi.org/10.1016/j.cell.2004.05.014>.
- Krysov, Sergey, Kathleen N. Potter, C. Ian Mockridge, Vania Coelho, Isla Wheatley, Graham Packham, and Freda K. Stevenson. 2010. "Surface IgM of CLL cells displays unusual glycans indicative of engagement of antigen in vivo." *Blood* 115 (21): 4198–4205. <https://doi.org/10.1182/blood-2009-12-254847>.
- 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>.
- Kulis, Marta, Simon Heath, Marina Bibikova, Ana C Queirós, Alba Navarro, Guillem Clot, Alejandra Martínez-Trillo, et al. 2012. "Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia." *Nature Genetics* 44 (11): 1236–42. <https://doi.org/10.1038/ng.2443>.
- Kumar, Dinesh. n.d. "A Complete understanding of LASSO Regression." Accessed October 3, 2021. <https://www.mygreatlearning.com/blog/understanding-of-lasso-regression/#introduction>.
- Kurtova, Antonina V., Kumudha Balakrishnan, Rong Chen, Wei Ding, Susanne Schnabl, Maite P. Quiroga, Mariela Sivina, et al. 2009. "Diverse marrow stromal cells protect CLL cells from spontaneous and drug-induced apoptosis: development of a reliable and reproducible system to assess stromal cell adhesion-mediated drug resistance." *Blood* 114 (20): 4441–50. <https://doi.org/10.1182/blood-2009-07-233718>.
- Lagneaux, L., A. Delforge, D. Bron, C. De Bruyn, and P. Stryckmans. 1998. "Chronic Lymphocytic Leukemic B Cells But Not Normal B Cells Are Rescued From Apoptosis by Contact With Normal Bone Marrow Stromal Cells." *Blood* 91 (7): 2387–96. <https://doi.org/10.1182/BLOOD.V91.7.2387>.

- Landau, Dan A, Scott L Carter, Petar Stojanov, Aaron McKenna, Kristen Stevenson, Michael S Lawrence, Carrie Sougnez, et al. 2013. "Evolution and impact of subclonal mutations in chronic lymphocytic leukemia." *Cell* 152 (4): 714–26. <https://doi.org/10.1016/j.cell.2013.01.019>.
- Landau, Dan A, Kendall Clement, Michael J Ziller, Patrick Boyle, Jean Fan, Hongcang Gu, Kristen Stevenson, et al. 2014. "Locally Disordered Methylation Forms the Basis of Intratumor Methylome Variation in Chronic Lymphocytic Leukemia." *Cancer Cell* 26 (6): 813–25. <https://doi.org/10.1016/j.ccr.2014.10.012>.
- Landau, Dan A, Eugen Tausch, Amaro N Taylor-Weiner, Chip Stewart, Johannes G Reiter, Jasmin Bahlo, Sandra Kluth, et al. 2015. "Mutations driving CLL and their evolution in progression and relapse." *Nature* 526 (7574): 525–30. <https://doi.org/10.1038/nature15395>.
- Lebien, Tucker W., and Thomas F Tedder. 2008. "B lymphocytes: How they develop and function." *Blood* 112 (5): 1570–80. <https://doi.org/10.1182/blood-2008-02-078071>.
- Letai, Anthony. 2017. "Functional precision cancer medicine-moving beyond pure genomics." *Nature Medicine* 23 (9): 1028–35. <https://doi.org/10.1038/nm.4389>.
- Li, Yang, Yu Wang, Zhuoyue Wang, Danhui Yi, and Shuangge Ma. 2015. "Racial differences in three major NHL subtypes: Descriptive epidemiology." *Cancer Epidemiology* 39 (1): 8–13. <https://doi.org/10.1016/j.canep.2014.12.001>.
- Lipsky, Andrew, Danny Luan, Shirley Chen, Ronan Chaligne, Neville Dusaj, Erica B Bhavsar, Chelston Ang, et al. 2020. "Single-Cell Multi-Omics Reveals Distinct Paths to Survival of Admixed BTKC481 Mutant Vs. Wild-Type Cells in Clinically Progressing Chronic Lymphocytic Leukemia." *Blood* 136 (Supplement 1): 40–42. <https://doi.org/10.1182/blood-2020-138374>.
- Lotz, M, E Ranheim, and T J Kipps. 1994. "Transforming growth factor beta as endogenous growth inhibitor of chronic lymphocytic leukemia B cells." *Journal of Experimental Medicine* 179 (3): 999–1004. <https://doi.org/10.1084/jem.179.3.999>.
- Love, Michael, Simon Anders, and Wolfgang Huber. 2021. *DESeq2: Differential Gene Expression Analysis Based on the Negative Binomial Distribution*. <https://github.com/mikelove/DESeq2>.
- Lu, Desheng, Yandong Zhao, Rommel Tawatao, Howard B Cottam, Malini Sen, Lorenzo

- M Leoni, Thomas J Kipps, Maripat Corr, and Dennis A Carson. 2004. "Activation of the Wnt signaling pathway in chronic lymphocytic leukemia." *Proceedings of the National Academy of Sciences of the United States of America* 101 (9): 3118–23. <https://doi.org/10.1073/pnas.0308648100>.
- Lu, Junyan, Ester Cannizzaro, Fabienne Meier-Abt, Sebastian Scheinost, Peter-Martin Bruch, Holly Ar Giles, Almut Lütge, et al. 2021. "Multi-omics reveals clinically relevant proliferative drive associated with mTOR-MYC-OXPHOS activity in chronic lymphocytic leukemia." *Nature Cancer* 2 (8): 853–64. <https://doi.org/10.1038/s43018-021-00216-6>.
- Lukas, Marina, Britta Velten, Leopold Sellner, Katarzyna Tomska, Jennifer Hüellein, Tatjana Walther, Lena Wagner, et al. 2020. "Survey of ex vivo drug combination effects in chronic lymphocytic leukemia reveals synergistic drug effects and genetic dependencies." *Leukemia* 34 (11): 2934–50. <https://doi.org/10.1038/s41375-020-0846-5>.
- Lukenbill, Joshua, and Matt Kalaycio. 2013. "Fludarabine: A review of the clear benefits and potential harms." *Leuk Res.* <https://doi.org/10.1016/j.leukres.2013.05.004>.
- Mallm, Jan-Philipp, Murat Iskar, Naveed Ishaque, Lara C Klett, Sabrina J Kugler, Jose M Muino, Vladimir B Teif, et al. 2019. "Linking aberrant chromatin features in chronic lymphocytic leukemia to transcription factor networks." *Molecular Systems Biology* 15 (5): e8339. <https://doi.org/10.15252/msb.20188339>.
- Mangolini, Maurizio, Frederik Götte, Andrew Moore, Tim Ammon, Madlen Oelsner, Gloria Lutzny-Geier, Ludger Klein-Hitpass, et al. 2018. "Notch2 controls non-autonomous Wnt-signalling in chronic lymphocytic leukaemia." *Nature Communications* 9 (1): 3839. <https://doi.org/10.1038/s41467-018-06069-5>.
- Mansouri, Larry, Nikos Papakonstantinou, Stavroula Ntoufa, Kostas Stamatopoulos, and Richard Rosenquist. 2016. "NF- κ B activation in chronic lymphocytic leukemia: A point of convergence of external triggers and intrinsic lesions." *Semin Cancer Biol.* <https://doi.org/10.1016/j.semcaner.2016.07.005>.
- Mårtensson, Inga Lill, Nina Almqvist, Ola Grimsholm, and Angelina I. Bernardi. 2010. "The pre-B cell receptor checkpoint." *FEBS Lett.* <https://doi.org/10.1016/j.febslet.2010.04.057>.

- Martínez-Trillos, Alejandra, Alba Navarro, Marta Aymerich, Julio Delgado, Armando López-Guillermo, Elias Campo, and Neus Villamor. 2016. "Clinical impact of MYD88 mutations in chronic lymphocytic leukemia." *Blood*. <https://doi.org/10.1182/blood-2015-10-678490>.
- Matutes, E., K. Owusu-Ankomah, R. Morilla, J. G. García Marco, A. Houlihan, T. Que, and D. Catovsky. 1994. "The immunological profile of B-cell disorders and proposal of a scoring system for the diagnosis of CLL." *Undefined*. <https://www.semanticscholar.org/paper/The-immunological-profile-of-B-cell-disorders-and-a-Matutes-41ab989ce5c271aa59bd588e88919f4bea23dce6>.
- McWilliams, Emily M, Christopher R Lucas, Timothy Chen, Bonnie K Harrington, Ronni Wasmuth, Amanda Campbell, Kerry A Rogers, et al. 2019. "Anti-BAFF-R antibody VAY-736 demonstrates promising preclinical activity in CLL and enhances effectiveness of ibrutinib." *Blood Advances* 3 (3): 447–60. <https://doi.org/10.1182/bloodadvances.2018025684>.
- Meier-Abt, Fabienne, Junyan Lu, Ester Cannizzaro, Marcel F Pohly, Sandra Kummer, Sibylle Pfammatter, Laura Kunz, et al. 2021. "The Protein Landscape of Chronic Lymphocytic Leukemia (CLL)." *Blood*, June. <https://doi.org/10.1182/blood.2020009741>.
- Menyhárt, Otília, and Balázs Gyrfy. 2021. "Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis." *Comput Struct Biotechnol J*. <https://doi.org/10.1016/j.csbj.2021.01.009>.
- Messmer, Bradley T, Emilia Albesiano, Dimitar G Efremov, Fabio Ghiotto, Steven L Allen, Jonathan Kolitz, Robin Foa, et al. 2004. "Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia." *The Journal of Experimental Medicine* 200 (4): 519–25. <https://doi.org/10.1084/jem.20040544>.
- Messmer, Bradley T, Davorka Messmer, Steven L Allen, Jonathan E Kolitz, Prasad Kudalkar, Denise Cesar, Elizabeth J Murphy, et al. 2005. "In vivo measurements document the dynamic cellular kinetics of chronic lymphocytic leukemia B cells." *The Journal of Clinical Investigation* 115 (3): 755–64. <https://doi.org/10.1172/JCI23409>.
- Messmer, Davorka, Jessie-F Fecteau, Morgan O'heyre, Ila S Bharati, Tracy M Handel, and Thomas J Kipps. 2011. "Chronic lymphocytic leukemia cells receive RAF-dependent survival signals in response to CXCL12 that are sensitive to inhibition by

sorafenib.” <https://doi.org/10.1182/blood-2010>.

Migliazza, Anna, Francesc Bosch, Hirokazu Komatsu, Eftihia Cayanis, Stefano Martinotti, Elena Toniato, Ernesto Guccione, et al. 2001. “Nucleotide sequence, transcription map, and mutation analysis of the 13q14 chromosomal region deleted in B-cell chronic lymphocytic leukemia.” *Blood* 97 (7): 2098–2104. <https://doi.org/10.1182/blood.V97.7.2098>.

Miller, Kimberly D., Leticia Nogueira, Angela B. Mariotto, Julia H. Rowland, K. Robin Yabroff, Catherine M. Alfano, Ahmedin Jemal, Joan L. Kramer, and Rebecca L. Siegel. 2019. “Cancer treatment and survivorship statistics, 2019.” *CA: A Cancer Journal for Clinicians* 69 (5): 363–85. <https://doi.org/10.3322/caac.21565>.

Mittal, Amit K., Nagendra K. Chaturvedi, Karan J. Rai, Christine E. Gilling-Cutucache, Tara M. Nordgren, Margaret Moragues, Runqing Lu, et al. 2014. “Chronic Lymphocytic Leukemia Cells in a Lymph Node Microenvironment Depict Molecular Signature Associated with an Aggressive Disease.” *Molecular Medicine* 20 (1): 290–301. <https://doi.org/10.2119/molmed.2012.00303>.

Moore, Victoria Del Gaizo, Jennifer R Brown, Michael Certo, Tara M Love, Carl D Novina, and Anthony Letai. 2007. “Chronic lymphocytic leukemia requires BCL2 to sequester prodeath BIM, explaining sensitivity to BCL2 antagonist ABT-737.” *Journal of Clinical Investigation* 117 (1): 112–21. <https://doi.org/10.1172/JCI28281>.

Moreau, Elisabeth J, Estella Matutes, Roger P A'Hern, Alison M Morilla, Ricardo M Morilla, Kwasi A Owusu-Ankomah, Ben K Seon, and Daniel Catovsky. 1997. “Improvement of the chronic lymphocytic leukemia scoring system with the monoclonal antibody SN8 (CD79b).” *American Journal of Clinical Pathology* 108 (4): 378–82. <https://doi.org/10.1093/ajcp/108.4.378>.

Moreton, Paul, Ben Kennedy, Guy Lucas, Michael Leach, Saad M. B. Rassam, Andrew Haynes, Jane Tighe, et al. 2005. “Eradication of minimal residual disease in B-cell chronic lymphocytic leukemia after alemtuzumab therapy is associated with prolonged survival.” *Journal of Clinical Oncology* 23 (13): 2971–79. <https://doi.org/10.1200/JCO.2005.04.021>.

Mraz, Marek, Clive S Zent, Amy K Church, Diane F Jelinek, Xiaosheng Wu, Sarka Pospisilova, Stephen M Ansell, et al. 2011. “Bone marrow stromal cells protect lymphoma B-cells from rituximab-induced apoptosis and targeting integrin α -4- β -1 (VLA-4) with natalizumab can overcome this resistance.” *British Journal of Haematology*

tology 155 (1): 53–64. <https://doi.org/10.1111/j.1365-2141.2011.08794.x>.

Muzio, Marta, Cristina Scielzo, Maria T. S. Bertilaccio, Michela Frenquelli, Paolo Ghia, and Federico Caligaris-Cappio. 2009. “Expression and function of toll like receptors in chronic lymphocytic leukaemia cells.” *British Journal of Haematology* 144 (4): 507–16. <https://doi.org/10.1111/j.1365-2141.2008.07475.x>.

Myhrinder, Anna Lanemo, Eva Hellqvist, Ekaterina Sidorova, Anita Söderberg, Helen Baxendale, Charlotte Dahle, Kerstin Willander, et al. 2008. “A new perspective: Molecular motifs on oxidized LDL, apoptotic cells, and bacteria are targets for chronic lymphocytic leukemia antibodies.” *Blood* 111 (7): 3838–48. <https://doi.org/10.1182/blood-2007-11-125450>.

Nabhan, Chadi, Briseis Aschebrook-Kilfoy, Brian C-H Chiu, Sonali M Smith, Tait D Shanafelt, Andrew M Evens, and Neil E Kay. 2014. “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.” *Leukemia and Lymphoma* 55 (12): 2778–84. <https://doi.org/10.3109/10428194.2014.898758>.

Nadeu, Ferran, Julio Delgado, Cristina Royo, Tycho Baumann, Tatjana Stankovic, Magda Pinyol, Pedro Jares, et al. 2016. “Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia.” *Blood* 127 (17): 2122–30. <https://doi.org/10.1182/blood-2015-07-659144>.

Nelder, J. A., and R. W. M. Wedderburn. 1972. “Generalized Linear Models.” *Journal of the Royal Statistical Society. Series A (General)* 135 (3): 370. <https://doi.org/10.2307/2344614>.

Nishio, Mitsufumi, Tomoyuki Endo, Nobuhiro Tsukada, Junko Ohata, Shinichi Kitada, John C Reed, Nathan J Zvaifler, and Thomas J Kipps. 2005. “Nurselike cells express BAFF and APRIL, which can promote survival of chronic lymphocytic leukemia cells via a paracrine pathway distinct from that of SDF-1alpha.” *Blood* 106 (3): 1012–20. <https://doi.org/10.1182/blood-2004-03-0889>.

Nutt, Stephen L., Philip D. Hodgkin, David M. Tarlinton, and Lynn M. Corcoran. 2015. “The generation of antibody-secreting plasma cells.” *Nature Reviews Immunology* 2015 15:3 15 (3): 160–71. <https://doi.org/10.1038/nri3795>.

O’Brien, Susan, and Neil E Kay. 2011. “Maintenance therapy for B-Chronic lymphocytic

- leukemia.” *Clinical Advances in Hematology and Oncology* 9 (1): 22–31. http://www.hematologyandoncology.net/files/2013/07/ho0111_obrien1.pdf.
- O’Hayre, Morgan, Catherina L. Salanga, Thomas J. Kipps, Davorka Messmer, Pieter C. Dorrestein, and Tracy M. Handel. 2010. “Elucidating the CXCL12/CXCR4 signaling network in chronic lymphocytic leukemia through phosphoproteomics analysis.” Edited by Syed A. Aziz. *PLoS ONE* 5 (7): e11716. <https://doi.org/10.1371/journal.pone.0011716>.
- Oakes, Christopher C, Marc Seifert, Yassen Assenov, Lei Gu, Martina Przekopowitz, Amy S Ruppert, Qi Wang, et al. 2016. “DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia.” *Nature Genetics* 48 (3): 253–64. <https://doi.org/10.1038/ng.3488>.
- 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.*
- Oppezzo, P, and G Dighiero. 2013. “Role of the B-cell receptor and the microenvironment in chronic lymphocytic leukemia.” *Blood Cancer Journal* 3 (9): e149–49. <https://doi.org/10.1038/bcj.2013.45>.
- Os, Audun, Simone Bürgler, Anna Parente Ribes, Ane Funderud, Dong Wang, Keith M. Thompson, Geir E. Tjønnfjord, Bjarne Bogen, and Ludvig A. Munthe. 2013. “Chronic Lymphocytic Leukemia Cells Are Activated and Proliferate in Response to Specific T Helper Cells.” *Cell Reports* 4 (3): 566–77. <https://doi.org/10.1016/J.CELREP.2013.07.011>.
- Palamarchuk, Alexey, Alexey Efanov, Natalya Nazaryan, Urmila Santanam, Hansjuerg Alder, Laura Rassenti, Thomas Kipps, Carlo M Croce, and Yuri Pekarsky. 2010. “13q14 deletions in CLL involve cooperating tumor suppressors.” *Blood* 115 (19): 3916–22. <https://doi.org/10.1182/blood-2009-10-249367>.
- Palma, Marzia, Giusy Gentilcore, Kia Heimersson, Fariba Mozaffari, Barbro Näsman-Glaser, Emma Young, Richard Rosenquist, Lotta Hansson, Anders Österborg, and Håkan Mellstedt. 2017. “T cells in chronic lymphocytic leukemia display dysregulated expression of immune checkpoints and activation markers.” *Haematologica* 102 (3): 562–72. <https://doi.org/10.3324/haematol.2016.151100>.

- Panayiotidis, P., D. Jones, K. Ganeshaguru, L. Foroni, and A. V. Hoffbrand. 1996. "Human bone marrow stromal cells prevent apoptosis and support the survival of chronic lymphocytic leukaemia cells in vitro." *British Journal of Haematology* 92 (1): 97–103. <https://doi.org/10.1046/j.1365-2141.1996.00305.x>.
- Pascutti, Maria Fernanda, Margot Jak, Jacqueline M. Tromp, Ingrid A. M. Derkx, Esther B. M. Remmerswaal, Rachel Thijssen, Martijn H. A. van Attekum, et al. 2013. "IL-21 and CD40L signals from autologous T cells can induce antigen-independent proliferation of CLL cells." *Blood* 122 (17): 3010–19. <https://doi.org/10.1182/blood-2012-11-467670>.
- Pedersen, Irene M., Shinichi Kitada, Lorenzo M. Leoni, Juan M. Zapata, James G. Karras, Nobuhiro Tsukada, Thomas J. Kipps, Yong Sung Choi, Frank Bennett, and John C. Reed. 2002. "Protection of CLL B cells by a follicular dendritic cell line is dependent on induction of Mcl-1." *Blood* 100 (5): 1795–1801. https://doi.org/10.1182/BLOOD.V100.5.1795.H81702001795_1795_1801.
- Pelanda, Roberta, and Raul M. Torres. 2012. "Central B-Cell tolerance: Where selection begins." *Cold Spring Harbor Perspectives in Biology* 4 (4). <https://doi.org/10.1101/cshperspect.a007146>.
- Pemovska, Tea, Mika Kontro, Bhagwan Yadav, Henrik Edgren, Samuli Eldfors, Agnieszka Szwajda, Henrikki Almusa, et al. 2013. "Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia." *Cancer Discovery* 3 (12): 1416–29. <https://doi.org/10.1158/2159-8290.CD-13-0350>.
- Puente, Xose S, Silvia Beà, Rafael Valdés-Mas, Neus Villamor, Jesús Gutiérrez-Abril, José I Martín-Subero, Marta Munar, et al. 2015. "Non-coding recurrent mutations in chronic lymphocytic leukaemia." *Nature* 526 (7574): 519–24. <https://doi.org/10.1038/nature14666>.
- Puente, Xose S., Magda Pinyol, Victor Quesada, Laura Conde, Gonzalo R. Ordóñez, Neus Villamor, Georgia Escaramis, et al. 2011. "Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia." *Nature* 475 (7354): 101–5. <https://doi.org/10.1038/nature10113>.
- Pulte, Dianne, Maria Theresa Redaniel, Jenny Bird, and Mona Jeffreys. 2015. "Survival for patients with chronic leukemias in the US and Britain: Age-related disparities and changes in the early 21st century." *European Journal of Haematology* 94 (6): 540–45. <https://doi.org/10.1111/ejh.12468>.

- Purroy, Noelia, Pau Abrisqueta, Júlia Carabia, Cecilia Carpio, Carles Palacio, Francesc Bosch, and Marta Crespo. 2015. "Co-culture of primary CLL cells with bone marrow mesenchymal cells, CD40 ligand and CpG ODN promotes proliferation of chemoresistant CLL cells phenotypically comparable to those proliferating in vivo." *Oncotarget* 6 (10): 7632–43. <https://doi.org/10.18632/oncotarget.2939>.
- Quesada, Víctor, Laura Conde, Neus Villamor, Gonzalo R. Ordóñez, Pedro Jares, Laia Bassaganyas, Andrew J. Ramsay, et al. 2012. "Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia." *Nature Genetics* 44 (1): 47–52. <https://doi.org/10.1038/ng.1032>.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- 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>.
- Robak, Tadeusz. 2005. "Therapy of chronic lymphocytic leukaemia with purine nucleoside analogues: Facts and controversies." *Drugs Aging*. <https://doi.org/10.2165/00002512-200522120-00002>.
- Robak, Tadeusz, Anna Dmoszynska, Philippe Solal-Célyny, Krzysztof Warzocha, Javier Loscertales, John Catalano, Boris V Afanasiev, et al. 2010. "Rituximab plus fludarabine and cyclophosphamide prolongs progression-free survival compared with fludarabine and cyclophosphamide alone in previously treated chronic lymphocytic leukemia." *Journal of Clinical Oncology* 28 (10): 1756–65. <https://doi.org/10.1200/JCO.2009.26.4556>.
- Roberts, Andrew W., Matthew S. Davids, John M. Pagel, Brad S. Kahl, Soham D. Puvvada, John F. Gerecitano, Thomas J. Kipps, et al. 2016. "Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia." *New England Journal of Medicine* 374 (4): 311–22. <https://doi.org/10.1056/nejmoa1513257>.
- Roessner, Philipp M, and Martina Seiffert. 2020. "T-cells in chronic lymphocytic leukemia: Guardians or drivers of disease?" *Leukemia* 34 (8): 2012–24. <https://doi.org/10.1038/s41375-020-0873-2>.

- Rosenwald, Andreas, Ash A Alizadeh, George Widhopf, Richard Simon, R Eric Davis, Xin Yu, Liming Yang, et al. 2001. "Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia." *Journal of Experimental Medicine* 194 (11): 1639–47. <https://doi.org/10.1084/jem.194.11.1639>.
- Rossi, Davide, Michaela Cerri, Clara Deambrogi, Elisa Sozzi, Stefania Cresta, Silvia Rasi, Lorenzo De Paoli, et al. 2009. "The prognostic value of TP53 mutations in chronic lymphocytic leukemia is independent of Del17p13: Implications for overall survival and chemorefractoriness." *Clinical Cancer Research* 15 (3): 995–1004. <https://doi.org/10.1158/1078-0432.CCR-08-1630>.
- Rossi, Davide, Silvia Rasi, Giulia Fabbri, Valeria Spina, Marco Fangazio, Francesco Forconi, Roberto Marasca, et al. 2012. "Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia." *Blood* 119 (2): 521–29. <https://doi.org/10.1182/blood-2011-09-379966>.
- Rossi, Davide, Silvia Rasi, Valeria Spina, Alessio Bruscaggin, Sara Monti, Carmela Ciardullo, Clara Deambrogi, et al. 2013. "Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia." *Blood* 121 (8): 1403–12. <https://doi.org/10.1182/blood-2012-09-458265>.
- Scheffold, Annika, and Stephan Stilgenbauer. 2020. "Revolution of Chronic Lymphocytic Leukemia Therapy: the Chemo-Free Treatment Paradigm." Springer. <https://doi.org/10.1007/s11912-020-0881-4>.
- Schulz, A., G. Toedt, T. Zenz, S. Stilgenbauer, P. Lichter, and M. Seiffert. 2011. "Inflammatory cytokines and signaling pathways are associated with survival of primary chronic lymphocytic leukemia cells in vitro: a dominant role of CCL2." *Haematologica* 96 (3): 408–16. <https://doi.org/10.3324/haematol.2010.031377>.
- Scielzo, Cristina, and Paolo Ghia. 2020. "Modeling the Leukemia Microenvironment In Vitro." Frontiers Media SA. <https://doi.org/10.3389/fonc.2020.607608>.
- Scrivener, S., R. V. Goddard, E. R. Kaminski, and A. G. Prentice. 2003. "Abnormal T-cell Function in B-cell Chronic Lymphocytic Leukaemia." *Leukemia & Lymphoma* 44 (3): 383–89. <https://doi.org/10.1080/1042819021000029993>.
- Shiloh, Yosef, and Yael Ziv. 2013. "The ATM protein kinase: Regulating the cellular response to genotoxic stress, and more." *Nat Rev Mol Cell Biol*. <https://doi.org/>

10.1038/nrm3546.

- Shin, Chanseok, and James L Manley. 2004. "Cell signalling and the control of pre-mRNA splicing." *Nat Rev Mol Cell Biol*. <https://doi.org/10.1038/nrm1467>.
- Shlomchik, Mark J., and Florian Weisel. 2012. "Germinal center selection and the development of memory B and plasma cells." *Immunological Reviews* 247 (1): 52–63. <https://doi.org/10.1111/J.1600-065X.2012.01124.X>.
- Siegel, Rebecca, Carol DeSantis, Katherine Virgo, Kevin Stein, Angela Mariotto, Tenbroeck Smith, Dexter Cooper, et al. 2012. "Cancer treatment and survivorship statistics, 2012." *CA: A Cancer Journal for Clinicians* 62 (4): 220–41. <https://doi.org/10.3322/caac.21149>.
- Simonetti, Giorgia, Maria Teresa Sabrina Bertilaccio, Paolo Ghia, and Ulf Klein. 2014. "Mouse models in the study of chronic lymphocytic leukemia pathogenesis and therapy." American Society of Hematology. <https://doi.org/10.1182/blood-2014-05-577122>.
- Skowronska, Anna, Anton Parker, Gulshanara Ahmed, Ceri Oldreive, Zadie Davis, Sue Richards, Martin Dyer, et al. 2012. "Biallelic ATM inactivation significantly reduces survival in patients treated on the United Kingdom leukemia research fund chronic lymphocytic leukemia 4 trial." *Journal of Clinical Oncology* 30 (36): 4524–32. <https://doi.org/10.1200/JCO.2011.41.0852>.
- Snijder, Berend, Gregory I Vladimer, Nikolaus Krall, Katsuhiro Miura, Ann Sofie Schmolke, Christoph Kornauth, Oscar Lopez de la Fuente, et al. 2017. "Image-based ex-vivo drug screening for patients with aggressive haematological malignancies: interim results from a single-arm, open-label, pilot study." *The Lancet Haematology* 4 (12): e595–606. [https://doi.org/10.1016/S2352-3026\(17\)30208-9](https://doi.org/10.1016/S2352-3026(17)30208-9).
- Srinivasan, Vishrut K, Shano Naseem, Neelam Varma, Deepesh P Lad, and Pankaj Malhotra. 2020. "Genomic alterations in chronic lymphocytic leukemia and their correlation with clinico-hematological parameters and disease progression." *Blood Research* 55 (3): 131–38. <https://doi.org/10.5045/br.2020.2020080>.
- Stamatopoulos, Basile, Nathalie Meuleman, Cécile De Bruyn, Karlien Pieters, Philippe Mineur, Christine Le Roy, Stéphane Saint-Georges, et al. 2012. "AMD3100 disrupts the cross-talk between chronic lymphocytic leukemia cells and a mesenchymal stromal or nurse-like cell-based microenvironment: pre-clinical evidence for its association with chronic lymphocytic leukemia treatments." *Haematologica* 97 (4): 608–15.

<https://doi.org/10.3324/haematol.2011.052779>.

Stamatopoulos, Basile, Michaël Van Damme, Emerence Cromptot, Barbara Dessars, Hakim El Housni, Philippe Mineur, Nathalie Meuleman, Dominique Bron, and Laurence Lagneaux. 2015. "Opposite Prognostic Significance of Cellular and Serum Circulating MicroRNA-150 in Patients with Chronic Lymphocytic Leukemia." *Molecular Medicine* 21 (1): 123–33. <https://doi.org/10.2119/molmed.2014.00214>.

Stamatopoulos, B., N. Meuleman, C. De Bruyn, A. Delforge, D. Bron, and L. Lagneaux. 2010. "The histone deacetylase inhibitor suberoylanilide hydroxamic acid induces apoptosis, down-regulates the CXCR4 chemokine receptor and impairs migration of chronic lymphocytic leukemia cells." *Haematologica* 95 (7): 1136–43. <https://doi.org/10.3324/haematol.2009.013847>.

Stamatopoulos, Kostas, Chrysoula Belessi, Carol Moreno, Myriam Boudjogra, Giuseppe Guida, Tatjana Smilevska, Lynda Belhoul, et al. 2007. *Blood* 109 (1): 259–70. <https://doi.org/10.1182/blood-2006-03-012948>.

Stankovic, Tatjana, and Anna Skowronska. 2014. "The role of ATM mutations and 11q deletions in disease progression in chronic lymphocytic leukemia." *Leuk Lymphoma*. <https://doi.org/10.3109/10428194.2013.829919>.

Stevenson, Freda K, and Federico Caligaris-Cappio. 2004. "Chronic lymphocytic leukemia: Revelations from the B-cell receptor." *Blood*. <https://doi.org/10.1182/blood-2003-12-4312>.

Sthoeger, Z. M., M. Wakai, D. B. Tse, V. P. Vinciguerra, S. L. Allen, D. R. Budman, S. M. Lichtman, P. Schulman, L. R. Weiselberg, and N. Chiorazzi. 1989. "Production of autoantibodies by CD5-expressing B lymphocytes from patients with chronic lymphocytic leukemia." *Journal of Experimental Medicine* 169 (1): 255–68. <https://doi.org/10.1084/jem.169.1.255>.

Stilgenbauer, Stephan, Barbara Eichhorst, Johannes Schetelig, Steven Coutre, John F Seymour, Talha Munir, Soham D Puvvada, et al. 2016. "Venetoclax in relapsed or refractory chronic lymphocytic leukaemia with 17p deletion: a multicentre, open-label, phase 2 study." *The Lancet Oncology* 17 (6): 768–78. [https://doi.org/10.1016/S1470-2045\(16\)30019-5](https://doi.org/10.1016/S1470-2045(16)30019-5).

Stilgenbauer, Stephan, Andrea Schnaiter, Peter Paschka, Thorsten Zenz, Marianna Rossi, Konstanze Döhner, Andreas Bühl, et al. 2014. "Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial." *Blood*

123 (21): 3247–54. <https://doi.org/10.1182/blood-2014-01-546150>.

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

Tibshirani, Robert. 1996. “Regression Shrinkage and Selection via the Lasso.” 1. Vol. 58. <https://www.jstor.org/stable/pdf/2346178.pdf>.

Timp, Winston, and Andrew P Feinberg. 2013. “Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host.” *Nat Rev Cancer*. <https://doi.org/10.1038/nrc3486>.

Totero, Daniela de, Raffaella Meazza, Simona Zupo, Giovanna Cutrona, Serena Matis, Monica Colombo, Enrico Balleari, et al. 2006. “Interleukin-21 receptor (IL-21R) is up-regulated by CD40 triggering and mediates proapoptotic signals in chronic lymphocytic leukemia B cells.” *Blood* 107 (9): 3708–15. <https://doi.org/10.1182/blood-2005-09-3535>.

Trbusek, Martin, Jana Smardova, Jitka Malcikova, Ludmila Sebejova, Petr Dobes, Miluse Svitakova, Vladimira Vranova, et al. 2011. “Missense mutations located in structural p53 DNA-binding motifs are associated with extremely poor survival in chronic lymphocytic leukemia.” *Journal of Clinical Oncology* 29 (19): 2703–8. <https://doi.org/10.1200/JCO.2011.34.7872>.

Trentin, Livio, Andrea Cerutti, Renato Zambello, Rosaria Sancetta, Cristina Tassinari, Monica Facco, Fausto Adami, Francesco Rodeghiero, Carlo Agostini, and Gianpietro Semenzato. 1996. “Interleukin-15 promotes the growth of leukemic cells of patients with B-cell chronic lymphoproliferative disorders.” *Blood* 87 (8): 3327–35. <https://doi.org/10.1182/blood.v87.8.3327.bloodjournal18783327>.

Trimarco, Valentina, Elisa Ave, Monica Facco, Giorgia Chiodin, Federica Frezzato, Veronica Martini, Cristina Gattazzo, et al. 2015. “Cross-talk between chronic lymphocytic leukemia (CLL) tumor b cells and mesenchymal stromal cells (MSCs): Implications for neoplastic cell survival.” *Oncotarget* 6 (39): 42130–49. <https://doi.org/10.18632/oncotarget.6239>.

Tsukada, Nobuhiro, Jan A. Burger, Nathan J. Zviaifler, and Thomas J. Kipps. 2002. “Distinctive features of “nurselike” cells that differentiate in the context of chronic lymphocytic leukemia.” *Blood* 99 (3): 1030–37. <https://doi.org/10.1182/blood.V99.3.1030>.

- 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>.
- Tyner, Jeffrey W, Wayne F Yang, Armand Bankhead, Guang Fan, Luke B Fletcher, Jade Bryant, Jason M Glover, et al. 2013. "Kinase pathway dependence in primary human leukemias determined by rapid inhibitor screening." *Cancer Research* 73 (1): 285–96. <https://doi.org/10.1158/0008-5472.CAN-12-1906>.
- Vangapandu, Hima V., Mary L. Ayres, Christopher A. Bristow, William G. Wierda, Michael J. Keating, Kumudha Balakrishnan, Christine M. Stellrecht, and Varsha Gandhi. 2017. "The Stromal Microenvironment Modulates Mitochondrial Oxidative Phosphorylation in Chronic Lymphocytic Leukemia Cells." *Neoplasia (United States)* 19 (10): 762–71. <https://doi.org/10.1016/j.neo.2017.07.004>.
- Vogelstein, Bert, Nickolas Papadopoulos, Victor E Velculescu, Shabin Zhou, Luis A Diaz, and Kenneth W Kinzler. 2013. "Cancer genome landscapes." *Science*. <https://doi.org/10.1126/science.1235122>.
- Wahlfors, J., H. Hiltunen, K. Heinonen, E Hamalainen, L. Alhonen, and J. Janne. 1992. "Genomic hypomethylation in human chronic lymphocytic leukemia." *Blood* 80 (8): 2074–80. <https://doi.org/10.1182/blood.v80.8.2074.2074>.
- Walker, Jeffrey A. 2018. "Elementary Statistical Modeling for Applied Biostatistics." https://www.middleprofessor.com/files/applied-biostatistics_bookdown/_book/index.html#what-is-unusual-about-this-book.
- Walsby, Elisabeth, Andrea Buggins, Stephen Devereux, Ceri Jones, Guy Pratt, Paul Brennan, Chris Fegan, and Chris Pepper. 2014. "Development and characterization of a physiologically relevant model of lymphocyte migration in chronic lymphocytic leukemia." *Blood* 123 (23): 3607–17. <https://doi.org/10.1182/blood-2013-12-544569>.
- Wang, Lili, Michael S. Lawrence, Youzhong Wan, Petar Stojanov, Carrie Sougnez, Kristen Stevenson, Lillian Werner, et al. 2011. "SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia." *New England Journal of Medicine* 365 (26): 2497–2506. <https://doi.org/10.1056/nejmoa1109016>.
- Widhopf, George F., Laura Z. Rassenti, Traci L. Toy, John G. Gribben, William G. Wierda, and Thomas J. Kipps. 2004. *Blood* 104 (8): 2499–2504. <https://doi.org/10.1182/blood-2004-03-0818>.

- Wilkerson, Matt, and Peter Waltman. 2021. *ConsensusClusterPlus: ConsensusClusterPlus*.
- 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>.
- Woyach, Jennifer A., Richard R. Furman, Ta-Ming Liu, Hatice Gulcin Ozer, Marc Zapatka, Amy S. Ruppert, Ling Xue, et al. 2014. "Resistance Mechanisms for the Bruton's Tyrosine Kinase Inhibitor Ibrutinib." *New England Journal of Medicine* 370 (24): 2286–94. <https://doi.org/10.1056/nejmoa1400029>.
- Woyach, Jennifer A., and Amy J. Johnson. 2015. "Targeted therapies in CLL: Mechanisms of resistance and strategies for management." *Blood* 126 (4): 471–77. <https://doi.org/10.1182/blood-2015-03-585075>.
- Xu, Zhenshu, Donglian Xiong, Jushun Zhang, Jingyan Zhang, Xiuli Chen, Zhihe Chen, and Rong Zhan. 2018. "Bone marrow stromal cells enhance the survival of chronic lymphocytic leukemia cells by regulating HES-1 gene expression and H3K27me3 demethylation." *Oncology Letters* 15 (2): 1937–42. <https://doi.org/10.3892/ol.2017.7450>.
- Yosifov, Deyan Y, Christine Wolf, Stephan Stilgenbauer, and Daniel Mertens. 2019. "From biology to therapy: The CLL success story." Wolters Kluwer Health. <https://doi.org/10.1097/HS9.0000000000000175>.
- 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/>.
- Yu, Lijian, Haesook T Kim, Siddha N. Kasar, Parul Benien, Wei Du, Kevin Hoang, Andrew Aw, et al. 2017. "Survival of Del17p CLL depends on genomic complexity and somatic mutation." *Clinical Cancer Research* 23 (3): 735–45. <https://doi.org/10.1158/1078-0432.CCR-16-0594>.
- Zenz, Thorsten, Barbara Eichhorst, Raymonde Busch, Tina Denzel, Sonja Häbe, Dirk Winkler, Andreas Bühler, et al. 2010. "TP53 mutation and survival in chronic lym-

phocytic leukemia.” *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 28 (29): 4473–79. <https://doi.org/10.1200/JCO.2009.27.8762>.

Zenz, Thorsten, Alexander Kröber, Katrin Scherer, Sonja Häbe, Andreas Bühler, Axel Benner, Tina Denzel, et al. 2008. “Monoallelic TP53 inactivation is associated with poor prognosis in chronic lymphocytic leukemia: Results from a detailed genetic characterization with long-term follow-up.” *Blood* 112 (8): 3322–29. <https://doi.org/10.1182/blood-2008-04-154070>.

Zenz, Thorsten, Almut Luetge, Junyan Lu, Huellein Jennifer, Sascha Dietrich, Leopold Sellner, and Wolfgang Huber. 2019. “Transcriptional Profiling Reveals Strong Impact of Major Molecular Disease Subgroups and Mixed Epistasis in Chronic Lymphocytic Leukemia.” *Blood* 134 (Supplement_1): 1742–42. <https://doi.org/10.1182/blood-2019-126388>.

Zhang, Wan, Dunyaporn Trachootham, Jinyun Liu, Gang Chen, Helene Pelicano, Celia Garcia-Prieto, Weiqin Lu, et al. 2012. “Stromal control of cystine metabolism promotes cancer cell survival in chronic lymphocytic leukaemia.” *Nature Cell Biology* 14 (3): 276–86. <https://doi.org/10.1038/ncb2432>.

Ziller, Michael J., Hongcang Gu, Fabian Müller, Julie Donaghey, Linus T.-Y. Tsai, Oliver Kohlbacher, Philip L. De Jager, et al. 2013. “Charting a dynamic DNA methylation landscape of the human genome.” *Nature* 500 (7463): 477–81. <https://doi.org/10.1038/nature12433>.

Appendix

Figures

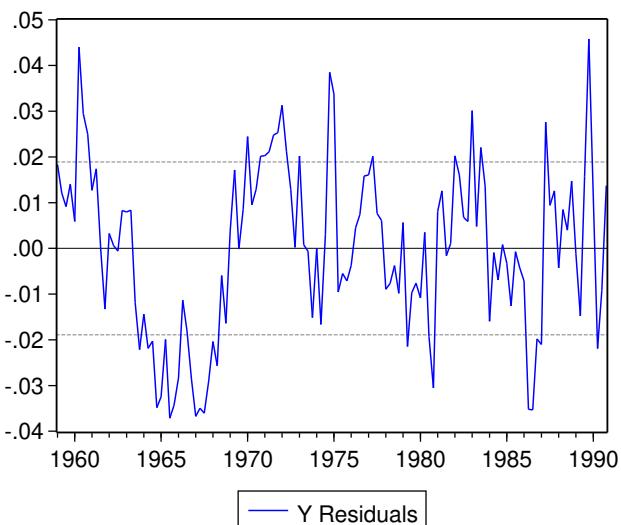


Figure 8.1: Analysis of ATACseq dataset of two trisomy 12 and two non-trisomy 12 untreated CLL PMBC samples. The volcano plot depicts change in TF activity (x axis) versus BH-adjusted p-values (y axis) for 636 TFs, comparing trisomy 12 and non-trisomy 12 samples. The (**diffTF?**)[REFERENCE] package 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.

(ref:drugcytGeneIntAll)

Tables

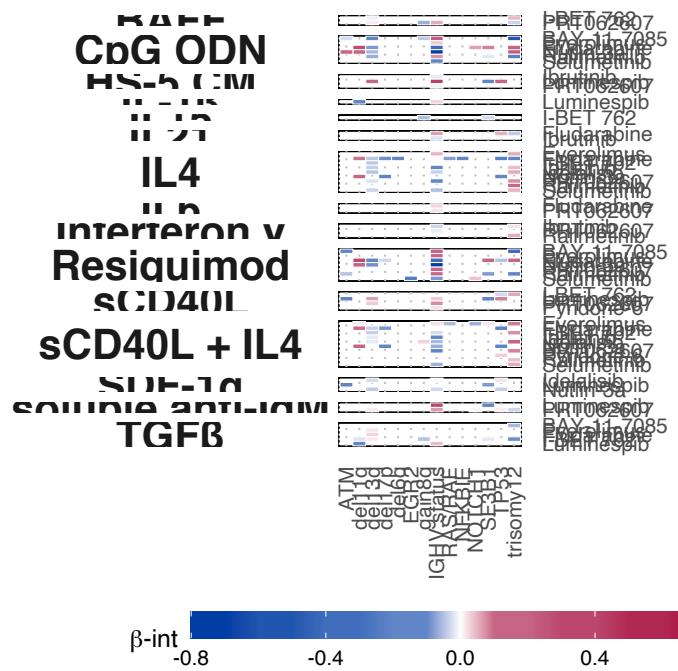


Figure 8.2: (ref:drugcytGeneIntAll)

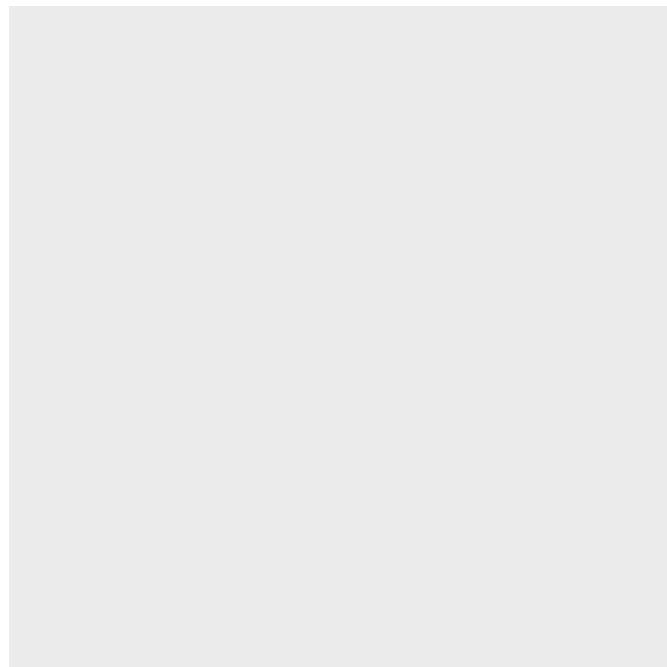


Figure 8.3: ADD NUTLIN-3A Predictor profile

Cytokine	Receptor	ENSEMBLID	Kegg
SDF-1a	CXCR 4	ENSG00000121966	7852
SDF-1a	CXCR 7	ENSG00000144476	57007
IL-6	IL-6 R	ENSG00000160712	3570
IL-6	IL-6 ST	ENSG00000134352	3572
IL-15	IL-15 R alpha	ENSG00000134470	3601
IL-15	IL-2 R beta	ENSG00000100385	3560
IL-15	IL-2 R gamma	ENSG00000147168	3561
IL-21	IL-21 R	ENSG00000103522	50615
IL-21	IL-21 R gamma	ENSG00000147168	3561
IL-2	IL-2 R alpha	ENSG00000134460	3559
IL-2	IL-2 R beta	ENSG00000100385	3560
IL-2	IL-2 R gamma	ENSG00000147168	3561
IL-10	IL-10 R alpha	ENSG00000110324	3587
IL-10	IL-10 R beta	ENSG00000243646	3588
IL-4	IL-4 R	ENSG00000077238	3566
IL-4	IL-2 R gamma	ENSG00000147168	3561
IL-4	IL-13 R A1	ENSG00000131724	3597
soluble anti-IgM	B cell receptor	ENSG00000275063	102723407
soluble anti-IgM	CD79A	ENSG00000105369	973
soluble anti-IgM	CD79B	ENSG00000007312	974
Resiquimod	TLR 7	ENSG00000196664	51284ă
Resiquimod	TLR 8	ENSG00000101916	51311
CpG ODN	TLR9	ENSG00000239732	54106
BAFF	BAFF R	ENSG00000159958	115650
BAFF	TACI	ENSG00000240505	23495
BAFF	BCMA	ENSG00000048462	608
IL-1b	IL-1 R1	ENSG00000115594	3554
IL-1b	IL-1 R2	ENSG00000115590	7850
sCD40L	CD40	ENSG00000101017	958
Interferon gamma	IFN-gamma R1	ENSG00000027697	3459
Interferon gamma	IFN-gamma R2	ENSG00000159128	3460
TGF-b1	TGF-beta R1	ENSG00000106799	7046
TGF-b1	TGF-beta R2	ENSG00000163513	7048

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