Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M.A. Holly Amelia Rebecca Giles

born in Portsmouth

Oral examination: Dec 15, 2021?

Drug - microenvironment - gene interplay in chronic lymphocytic leukemia

Referees: Dr. Judith Zaugg

Prof. Dr. Michael Boutros

To ...

Acknowledgements

I want to thank a few people.

List of publications

Thesis related

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

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.
- 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.

Zusammenfassung

Abstract in German Aktuelle Entwicklungen im Bereich der “omik”-Technologien tragen wesentlich zurBeschleunigung 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

List of Abbreviations

AKT	Protein kinase B
BCR	B-cell receptor
BTK	Brutons tyrosine kinase
CDF	Cumulative Distribution Function
CLL	Chronic lymphocytic leukemia
CNV	Copy number variation
CpG ODN	CpG oligodeoxynucleotides
ERK/MAPK signalling	Mitogen-activated protein kinase signalling
FBS	Fetal bovine serum
FDR	False discovery rate
GSEA	Gene set enrichment analysis
Ig	Immunoglobulin
IGHV	Immunoglobulin heavy chain variable region
IL	Interleukin
JAK	Janus kinase
M-CLL	CLL with somatic hypermutations in IGHV loci
NFκB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NOTCH	Neurogenic locus notch
OS	Overall survival
PBMC	Peripheral blood mononuclear cells
PCA	Principal component analysis
PI3K	Phosphoinositide 3-kinase
U-CLL	CLL without somatic hypermutations in IGHV loci
TGFβ	Transforming growth factor β
TLR	Toll-like receptor
TTT	Time to next treatment
SF3B1	Splicing factor 3B subunit 1
STAT	Signal transducer and activator of transcription
SYK	Spleen tyrosine kinase

List of Figures

List of Tables

Chapter 1

Introduction

Brief: Referencing: * What do we already know about the subject (literature review)? Use citations: Dietrich et al. (2017) shows that... Alternative Forms of the Wald test are considered (Dietrich et al. 2017).

- Provide an overview of your results.
- Outline of the paper:
  The thesis is organized as follows. The next section describes the model under investigation. Section "Data" describes the data set and Section "Results" presents the results. Finally, Section "Conclusion" concludes.
- The introduction should be around 8 pages.

The microenvironment in CLL constitutes cross talk via soluble factors and cell-cell contacts, in the blood, bone marrow and lymph nodes. **??**

Chapter 2

Methods

- Describe the data and its quality.
- How was the data sample selected?
- Provide descriptive statistics such as:
  - time period,
  - item number of observations, data frequency,
  - item mean, median,
  - item min, max, standard deviation,
  - item skewness, kurtosis, Jarque–Bera statistic,
  - item time series plots, histogram.
- For example (you can decide whether you want to display the code below or not by setting `echo=TRUE` or `echo=FALSE`, respectively, in the code chunk header):

```
col_names <- c(
  "3m", "6m", "1yr", "2yr", "3yr", "5yr", "7yr", "10yr", "12yr", "15yr"
)
```

```r
means     <- c(
  3.138, 3.191, 3.307, 3.544, 3.756, 4.093, 4.354, 4.621, 4.741, 4.878
)
stddev    <- c(
  0.915, 0.919, 0.935, 0.910, 0.876, 0.825, 0.803, 0.776, 0.768, 0.762
)
row_names <- c("Mean", "StD")

df <- matrix(data = c(means, stddev), nrow = 2, byrow = T)
rownames(df) <- row_names
colnames(df) <- col_names
df <- data.frame(df)
knitr::kable(
  df,
  booktabs = TRUE,
  caption = "Detailed descriptive statistics of location and dispersion for 2100 observed
  col.names = col_names,
  escape = FALSE
) %>%
kable_styling(latex_options = c("HOLD_position")) %>%
row_spec(2, hline_after = T)
```

Table 2.1: 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
StD	0.915	0.919	0.935	0.910	0.876	0.825	0.803	0.776	0.768	0.762

- Allows the reader to judge whether the sample is biased or to evaluate possible impacts of outliers, for example.
- Here tables can be easily integrated using the `kable()` function in the `knitr` package (with perhaps some additional help from the `kableExtra` package). `kable()` will automatically generate a label for the table environment. That way you don't have to manually enter in the table in LaTex, you can embed tables from R code.
- Tables can be referenced using `\@ref(label)`, where `label` is `tab:<name>`, where `<name>` is the code chunk label.
- The appearance may look different to tables directly typed with LaTex, due to limitations in `kable()`. To compare:

Table 2.2: This table was handwritten with LaTeX.

	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
StD	0.915	0.919	0.935	0.910	0.876	0.825	0.803	0.776	0.768	0.762

Chapter 3

Data

```
-- Attaching packages --------------------------------------- tidyverse 1.3.1 --

v tibble  3.1.2     v purrr   0.3.4
v tidyr   1.1.3     v stringr 1.4.0
v readr   1.4.0     v forcats 0.5.1

-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter()          masks stats::filter()
x kableExtra::group_rows() masks dplyr::group_rows()
x dplyr::lag()             masks stats::lag()
```

> "C'est ne pas un pipe."
>
> Descartes

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

3.1 Drug screens and experiments

3.1.1 High-throughput combinatorial pertubation assay

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

.

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 reponses, and drug-stimulus interactions in CLL.

3.1.2 Validation experiments

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

.

Figure 3.2: Schematic of validatory data.

3.2 Characteristics of drugs used in the screen

3.2.1 Drug pathways

We screened 17 different drugs (Figure 3.3).

.

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

3.2.2 Drug responses

The drug responses were as follows (Figure 3.4).

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

3.2.3 Genetic predictors of drug responses

3.2.4 Drug - Drug Correlations

3.3 Characteristics of stimuli used in the screen

3.3.1 Stimulus responses

3.3.2 Stimulus - Stimulus Correlations

3.4 Characteristics of patient samples used in the screen

###Genetic Data available for each patient WES, CNVs, Methylation, Transcriptomic, ATACseq, LDT, survT, IHC PLus Ibrutinib + IBET + IL4 treated samples - ATACseq, RNAseq, proteomics

3.5 Processing of raw values obtained from cell viability assay

3.5.1 Data normalization and quality control

3.6 Generation of public resource

###Shiny app

###Package

- Present the underlying economic model/theory and give reasons why it is suitable to answer the given problem[1].

[1] Here is an example of a footnote.

Chapter 4

Ex-vivo sensitivity to microenvironmental stimulation in primary CLL cells

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

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

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

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

4.1 Prolifing responses to the panel of stimuli

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

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

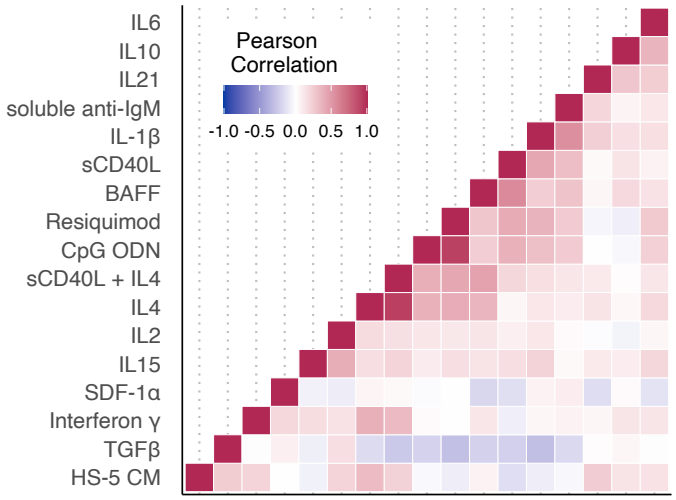

Figure 4.1: Heatmap of Pearson correlation coefficients. Coefficients for each pair of stimuli were calculated using log transformed viability values normalised to untreated control, and ordered according to hierarchical clustering. Figure adapted from Bruch & Giles et al. 2021.

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

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

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

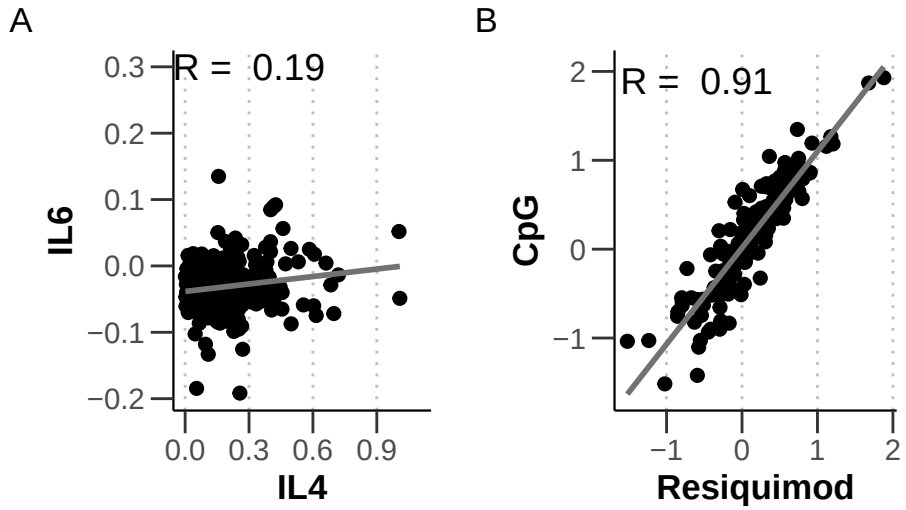

Figure 4.2: Scatter plot of log-transformed viability values, normalised to DMSO controls, for (A) treatment with JAK-STAT agonists IL4 and IL6 and (B) treatment with TLR agonists CpG ODN and Resiquimod.

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

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

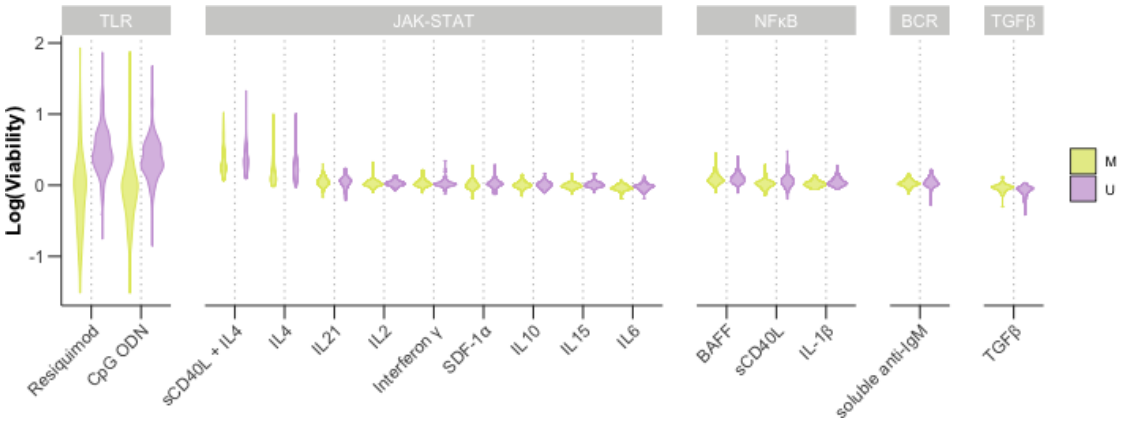

Figure 4.3: Log transformed viabilities after treatment with each stimulus. Where stimuli decreased viability relative to control, points are shown in blue, whilst increased viability is shown in red. Figure adapted from Bruch & Giles et al. 2021.

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

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

4.1.2 Microenvironmental response profiling identifies discrete patient subgroups

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

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

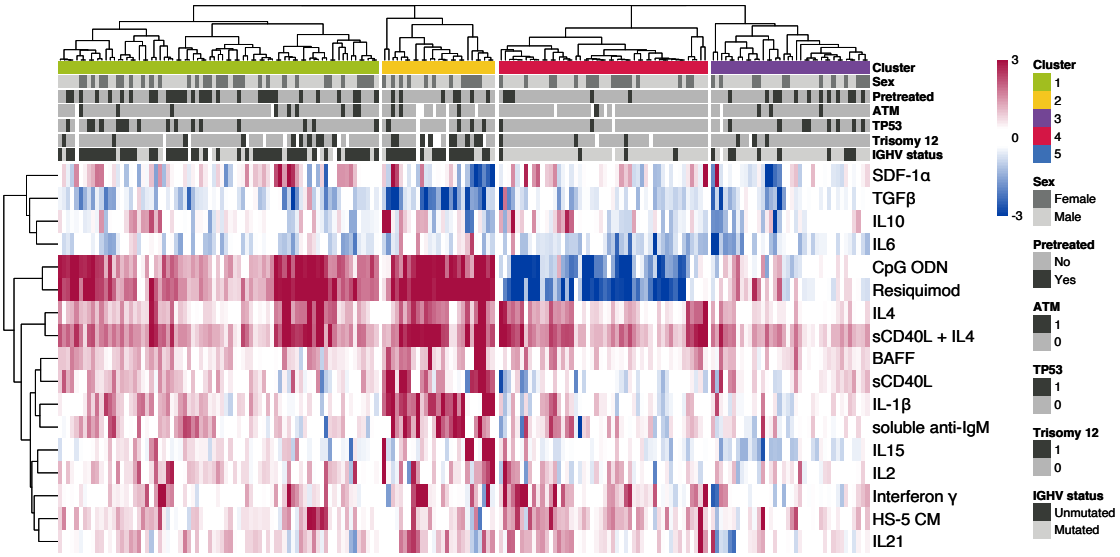

Figure 4.4: The heatmap matrix shows the viability measurements for 192 samples (columns) and 17 stimuli (rows). Viability was measured via ATP-based assay after 48h stimulation. The data are shown normalised to DMSO-treated controls, and scaled within each row according to the Median Absolute Deviation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

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

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

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

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

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

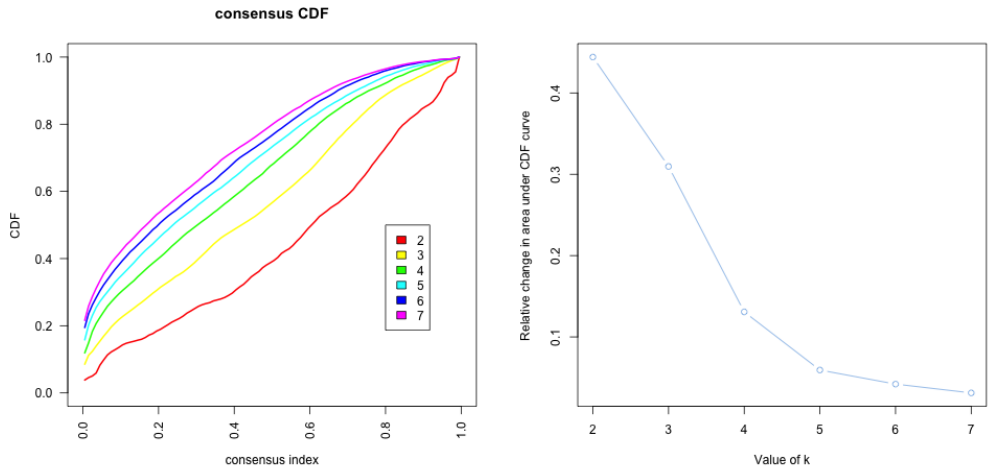

Figure 4.5: Summaries of the CDFs of the consensus matrices. Consensus CDF graphic showing the CDFs of the consensus matrix for k = 2 - 7, as indicated in the legend, estimated using 100 bin histogram (left). Relative change in area under the CDF curve, for k = 2 - 7, to compare k with k - 1. In the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

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

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

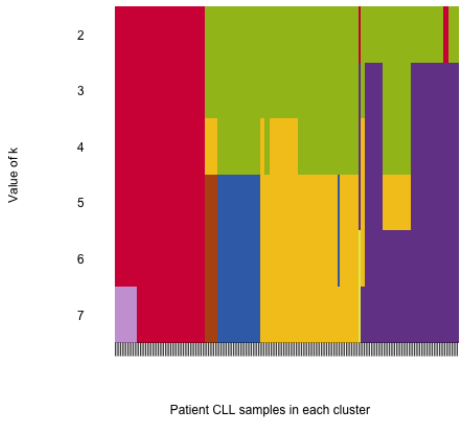

Figure 4.6: Assignment of patient samples (columns), to each cluster, for k = 1 - 7 (rows) to demonstrate stability of cluster membership. Cluster colour for k = 4 match those in heatmap in 4.4.

4.2 Functional characterisation of patient clusters

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

The heatmap in Figure 4.4 demonstrated that each cluster responded differently to the panel of stimuli.

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 NFκB agonists IL1β, 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 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) and time to next treatment (TTT) were used to quantify CLL proliferative capacity, independently of treatment.

C1 and C2 showed a shorter LDT than C3 and C4, which is expected due to the differential proportions of IGHV-U and M patient samples in these groups (Figure 4.8A). 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).

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

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).

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

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

The potential clinical relevance of the clusters was underlined by my observation that the samples within each group showed differential responses to drugs *in vitro* (Figure 4.9).

As expected, the IGHV-U enriched clusters C1 and 2 were more sensitive to BCR inhibition by ibrutinib, idelalisib and PRT062607, than C3 and 4. Between C1 and C2, C2 was more sensitive to a number of the drugs, including idelalisib (SYK) (p-value = 0.012), everolimus (mTOR) (p-value = 0.02) and the chemotherapeutics fludarabine (p-value = 0.031) and nutlin-3a (p-value = 0.042). Amongst C3 and C4, C3 showed lower sensitivity to everolimus (p-value = 0.051) and to fludarabine (p-value $<$ 0.001) and nutlin-3a (p-value = 0.01). This aligns with the observation that patients in C3 have a poorer prognosis, despite most of these samples annotated as IGHV-M. C4 also showed a positive repsonse to NfκB inhibition by BAY-11-7085, and p38 MAPK inhibition by Ralimetinib.

Such differential drug response patterns suggests that microenvironmental response may reflect disease-specific CLL biology, in the same way as molecular profiling, and thus may

have the potential to guide therapy decisions in future.

4.2.4 The clusters are enriched for different genetic features

Next we assessed differences in the molecular profiles of samples within each cluster. Visually, it appeared that certain clusters were enriched or depleted for various genetic features recurrent in CLL (Figure 4.10.

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

The approach used 3-fold cross validation, selecting the optimal model using lamda min. To ensure that the resulting coefficients were robust, we selected coefficients that satisfied certain cut-offs. Coefficients were selected if they were assigned in $> 60\%$ of 50 bootstrapped repeats, and were larger than 0.35. Figure 4.11 shows the mean coefficients and associated standard deviation, for each genetic feature that met these criteria in each cluster.

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

4.2.5 GSEA of DE genes between subgroups

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

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

I followed the Deseq2 protocol using a design formula to quantify the difference between clusters, and accounting for the confounding effect of IGHV status. 87 genes were differentially expressed (adjusted $p < 0.05$) between C3 and 4 (Figure 4.12).

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 was 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 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.

4.3 Additional Analysis

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

4.4 Summary

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

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

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

4.5 Discussion

Gosia suggested discussing each results section individually?

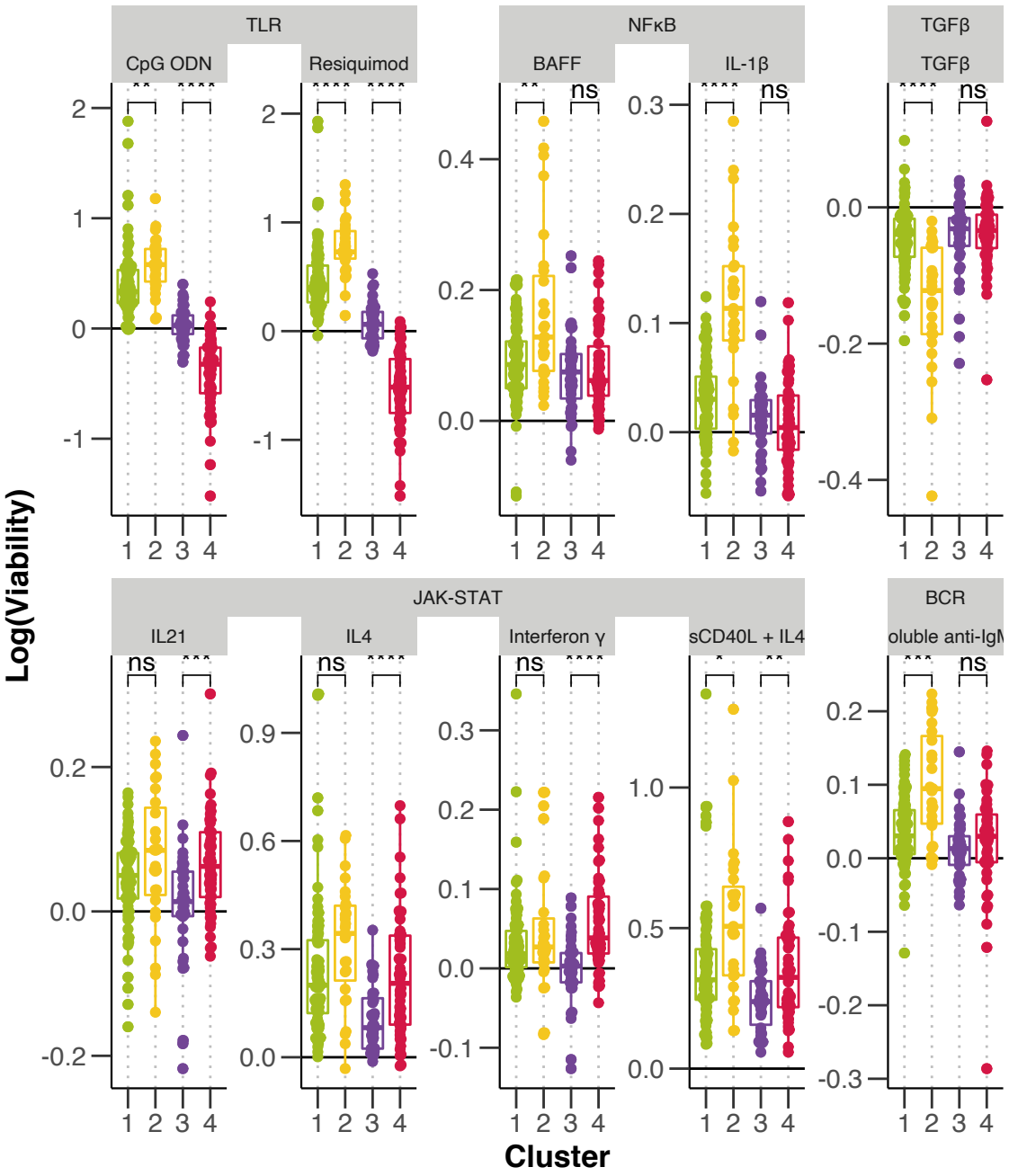

Figure 4.7: Log-transformed normalised viability values, stratified by cluster, for each stimulus. Stimuli activating the same pathway are grouped together. P-values from Student't t-test. Figure adapted from (Bruch & Giles et al. 2021).

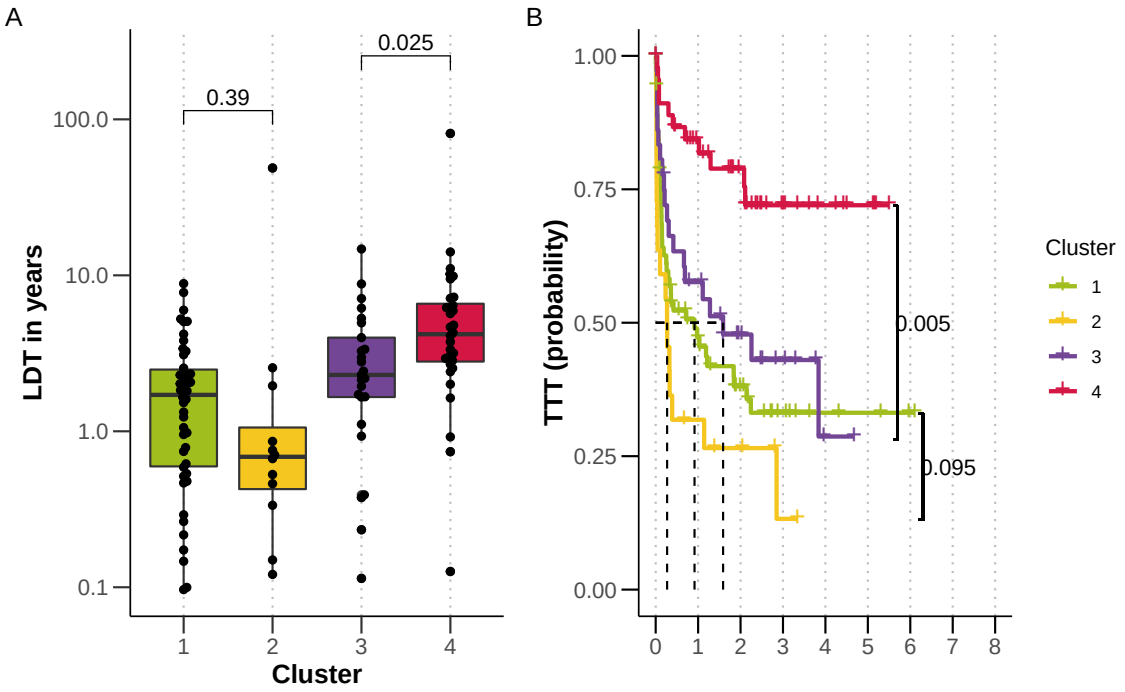

Figure 4.8: (A) Lymphocyte doubling time (LDT) stratified by cluster, p-values from Student's t-test. (B) Kaplan-Meier curves to show TTT for each cluster. p-values from univariate Cox proportional hazard models comparing IGHV-U enriched C1 with C2, and IGHV-M enriched C3 with C4. Figure from Bruch &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

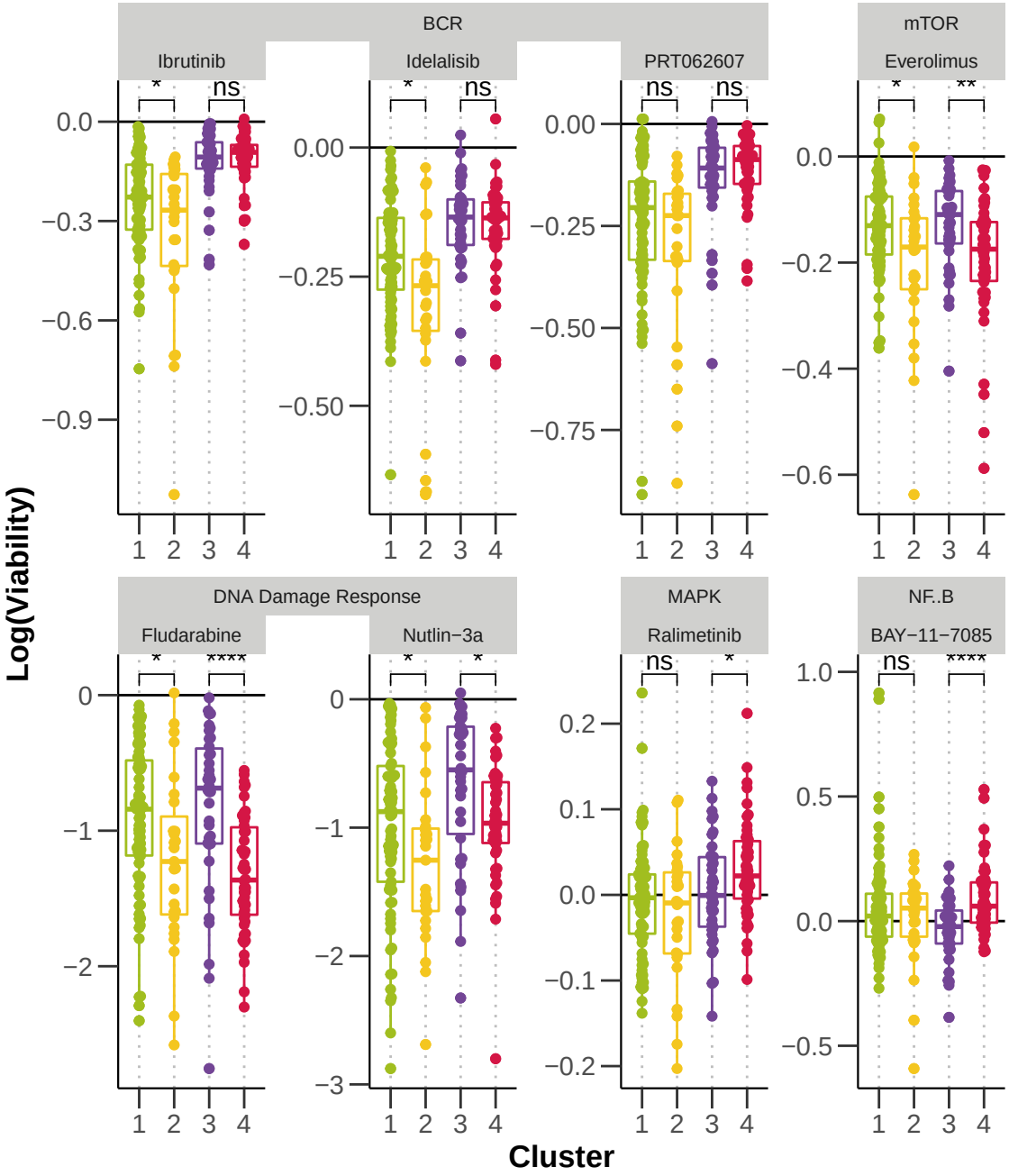

Figure 4.9: Log-transformed normalised viability values, stratified by cluster, for each drug. Drugs targeting the same pathway are grouped together. P-values from Student't t-test.

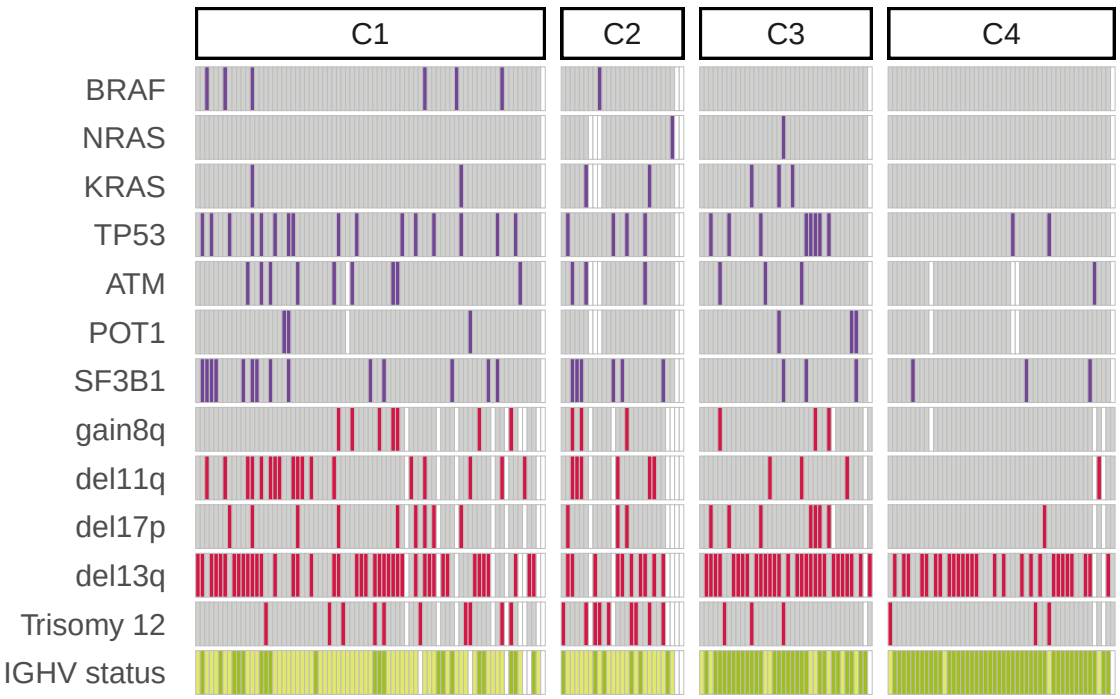

Figure 4.10: Distribution of selected genetic features (rows) within each cluster for all patient samples (columns). Where a patient sample is not annotated for a feature, this is marked in white.

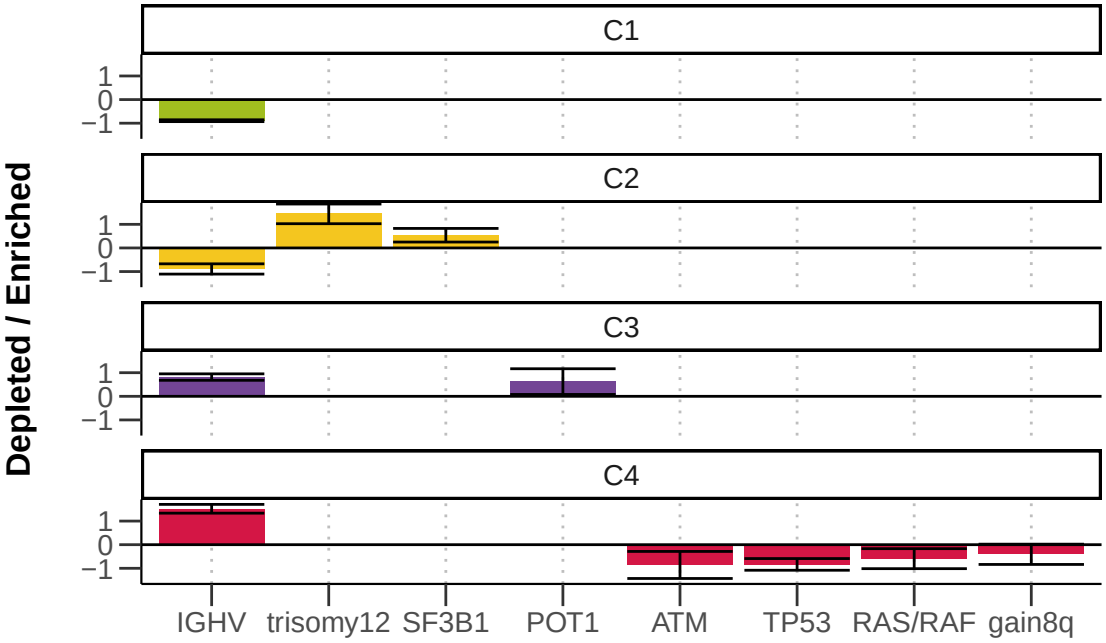

Figure 4.11: Multinomial regression with lasso penalisation to identify enrichment or depletion of genetic features within each cluster. Matrix of genetic features (p=39), and IGHV status (encoded as M = 1 and U = 0) were used to identify multivariate predictors of cluster assignment. x axis shows genetic predictors, y axis indicates value and sign of coefficient assigned to feature, for each cluster (positive coefficients are enriched in the cluster, negative coefficients are depleted). Coefficients shown are mean coefficients from 50 bootstrapped repeats and error bars represent the mean ś standard deviation. Genetic features with >20% missing values were excluded, and only patients with complete annotation were included in the model (n=137). Figure from Bruch & Giles et al 2021.

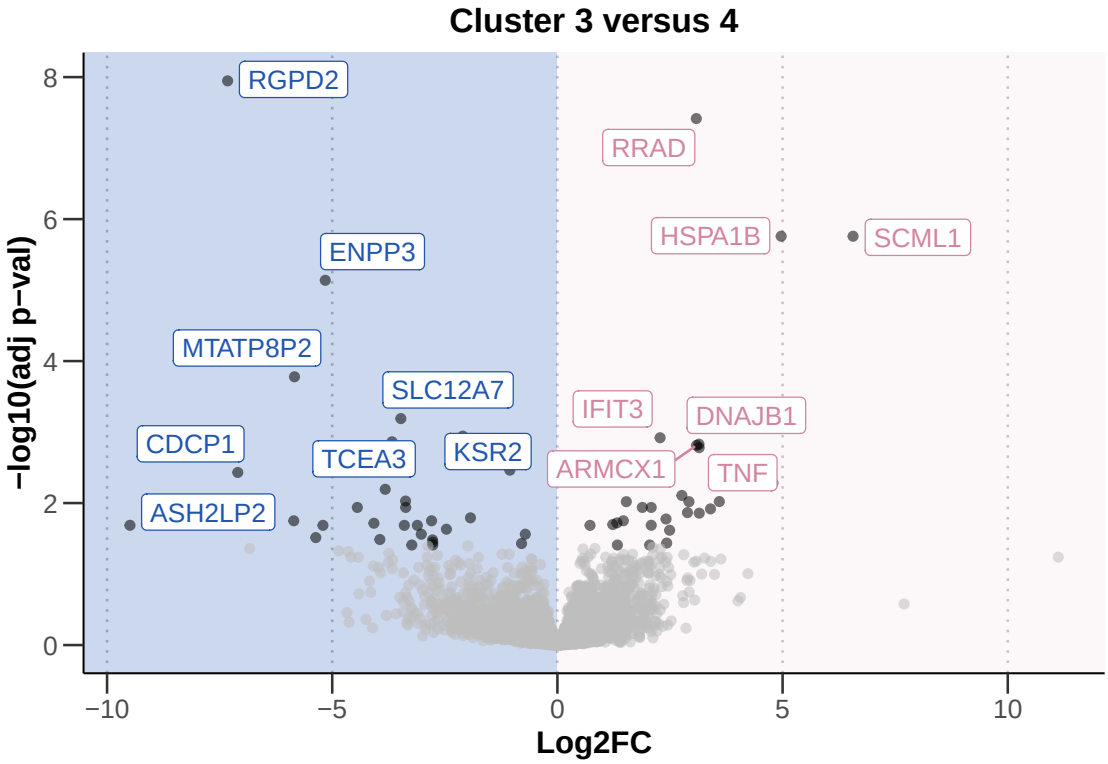

Figure 4.12: Volcano plot of differentially expressed genes between C3 and C4. X axis indicates log2 fold change values, calculated using the `DESeq2` package (Love, Huber, and Anders 2014), y axis gives corresponding -log10(adjusted p value). P values adjusted using BH method. Genes are labeled where ajdusted $p < 0.05$. Figure from Bruch & Giles et al. 2021.

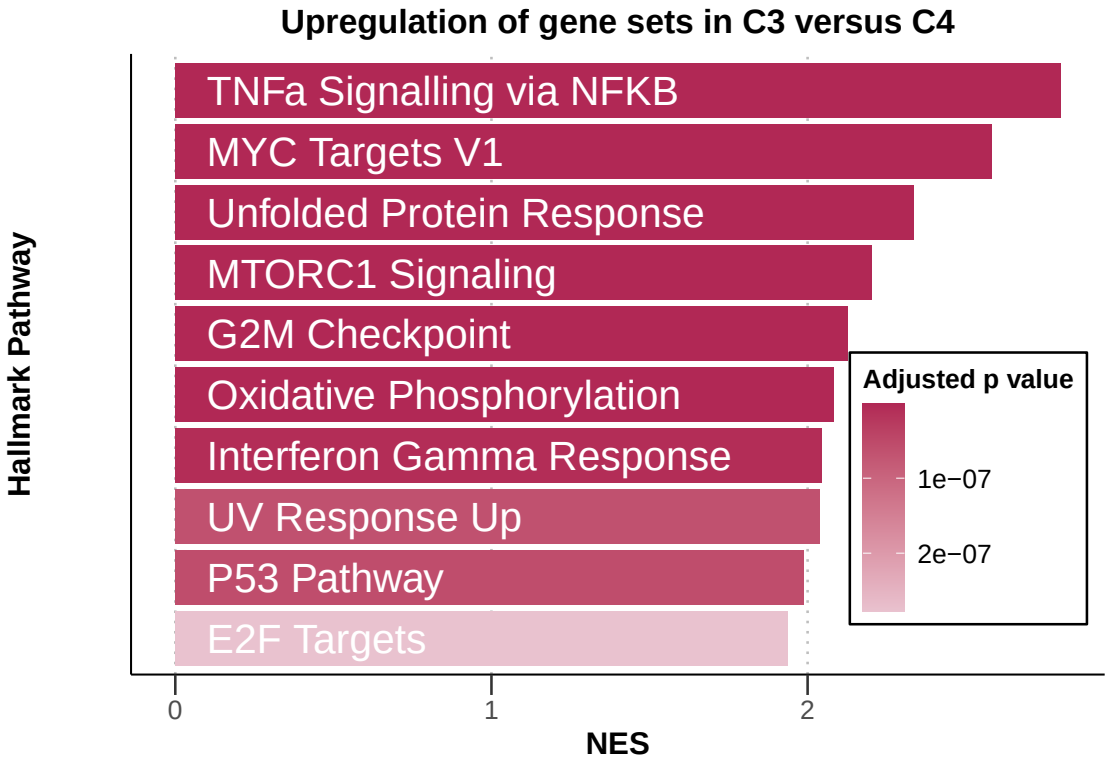

Figure 4.13: Gene set enrichment analysis (GSEA) to compare expression of genes in samples from C3 and C4 reveals upregulation of Hallmark pathways involved in microenvironmental signalling, stress response, metabolism and proliferation in C3. Normalised enrichment scores (NES) are shown for top 10 most significant pathways upregulated in C3 versus C4. Bars coloured according to adjusted p-value. Genes are ranked based on Wald statistic, calculated using the Deseq2 package and GSEA performed using the fgsea algorithm. Figure from Bruch & Giles et al. 2021.

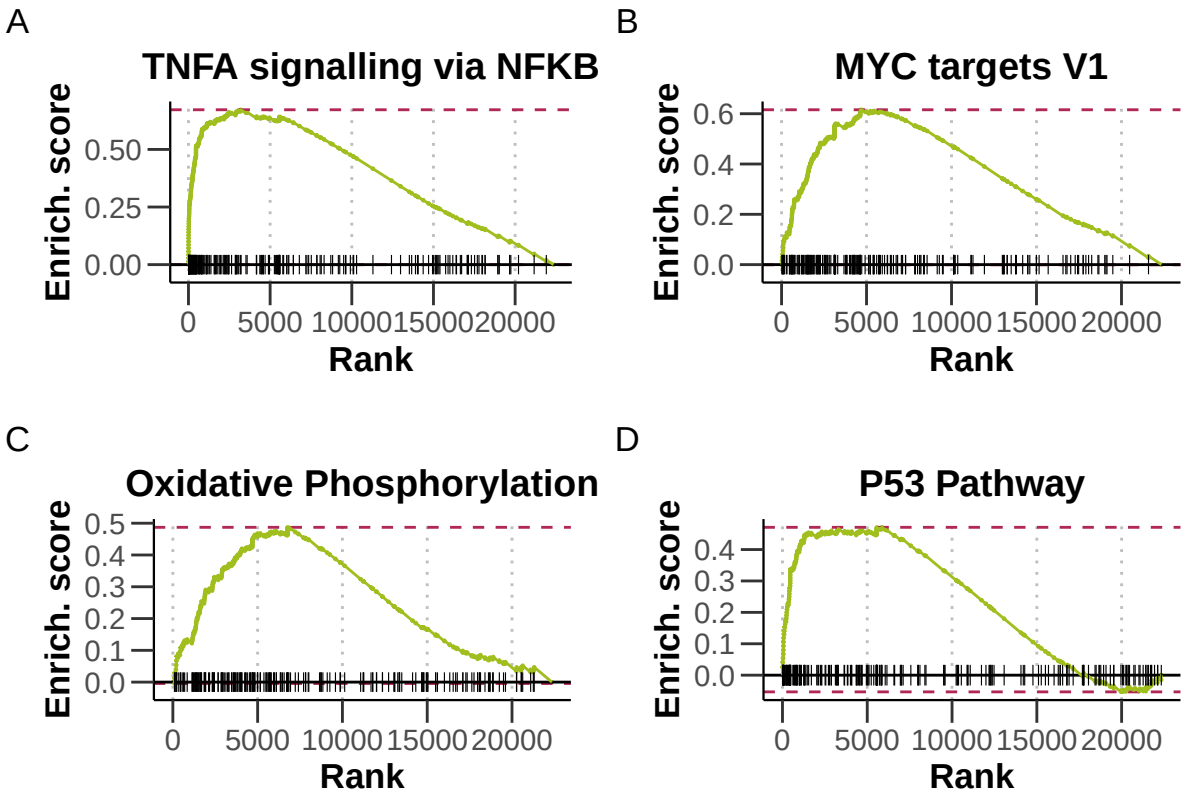

Figure 4.14: Enrichment plots of selected pathways. Gene set enrichment analysis (GSEA) was performed with the Hallmark gene sets from the GSEA Molecular Signatures Database. Wald statistic was used to rank the genes. The green curve corresponds to the Enrichment Score curve, which is the running sum of the weighted enrichment score obtained from GSEA software. Figure from Bruch & Giles et al. 2021

Chapter 5

Results

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

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

$$f_y(\lambda) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\lambda j} = (2\pi)^{-1} \left(\gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j \cos(\lambda j) \right) \tag{5.1}$$

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}`.

5.1 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 6

Results

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

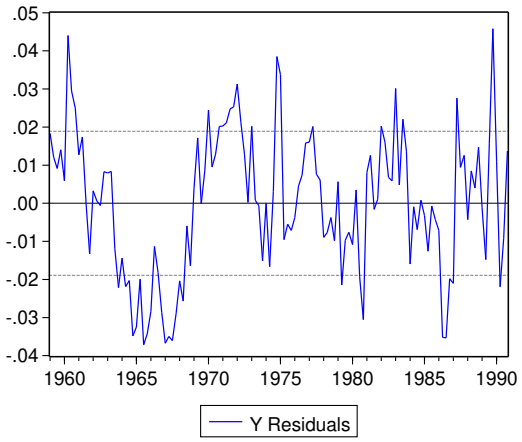

Figure 6.1: Estimated residuals from model XXX3. ...

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

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

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

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

6.1 Review of Results

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

Chapter 7

Results

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

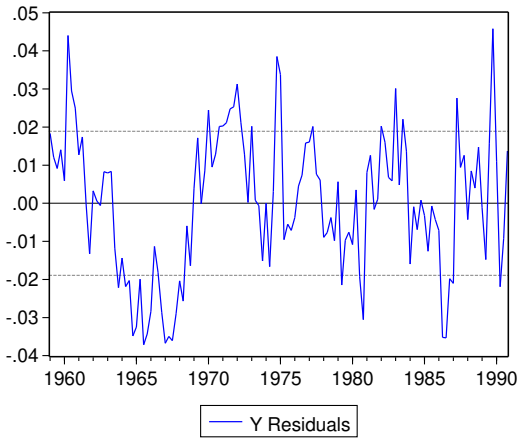

Figure 7.1: Estimated residuals from model XXX4. ...

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

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

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

Conclusion

- Give a short summary of what has been done and what has been found.
- Expose results concisely.
- Draw conclusions about the problem studied. What are the implications of your findings?
- Point out some limitations of study (assist reader in judging validity of findings).
- Suggest issues for future research.

References

10 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.

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.

Love, Michael I, Wolfgang Huber, and Simon Anders. 2014. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

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

Here goes the appendix!

Figures

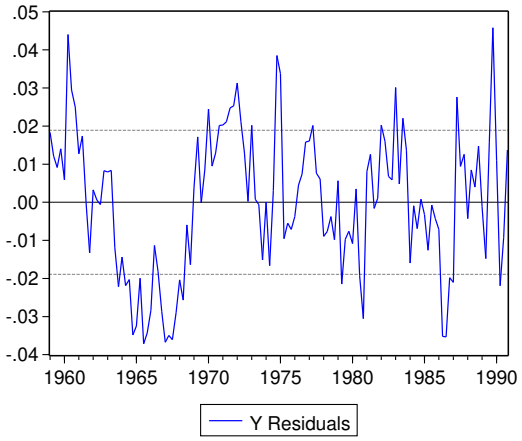

Figure 1: Estimated residuals (2) from model XXX. ...

Tables

Table 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