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 3.3.1) (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 Prolifing 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 NfκB pathways and a smaller group encompassing IL4 and TLR stimuli.

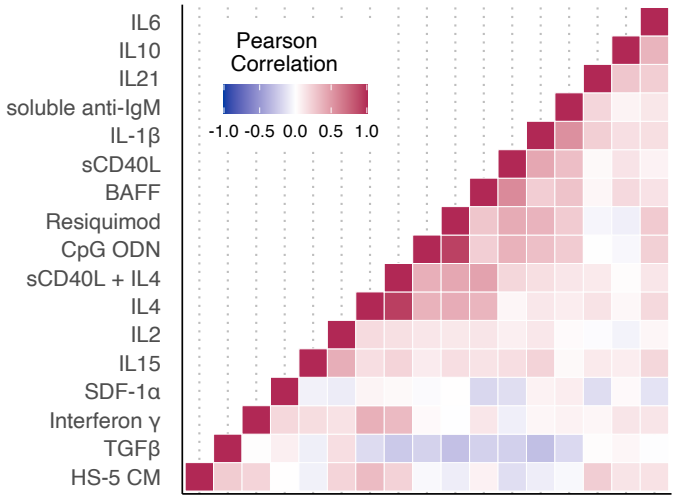

Figure 4.1: Heatmap of Pearson correlation coefficients of each pair of stimuli, based on log transformed viability values. See Methods section 2.5.2. *Figure adapted from Bruch et al. (2021) .*

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).

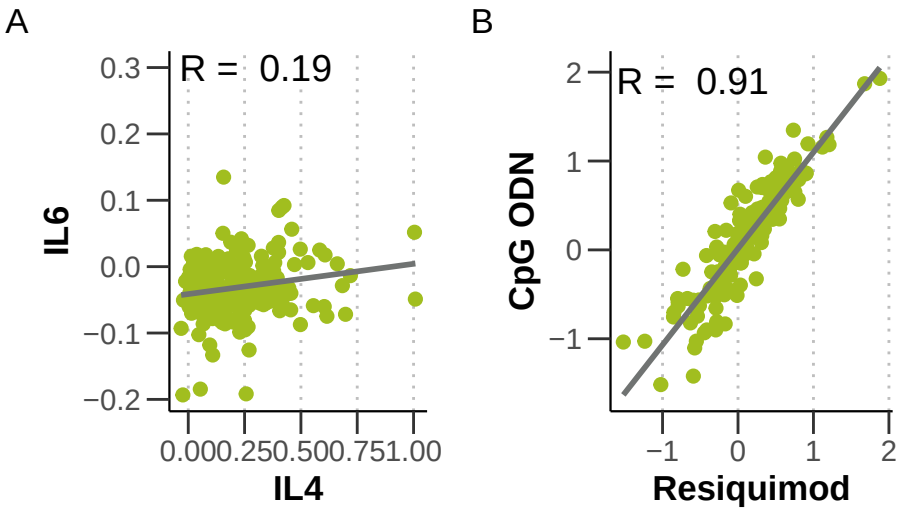

Figure 4.2: Scatter plot of log-transformed viability values, normalised to DMSO controls, for (A) treatment with JAK-STAT agonists IL4 and IL6 and (B) treatment with TLR agonists CpG ODN and Resiquimod.

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, TGFβ and TLR 7/8/9 agonists CpG ODN and

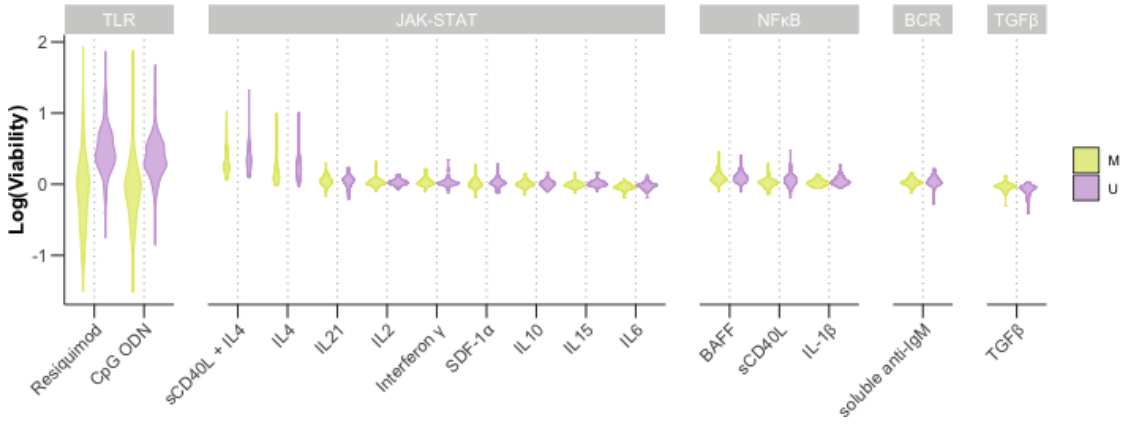

Figure 4.3: Log transformed viabilities after treatment with each stimulus. Stimuli are grouped by corresponding pathway, and responses are stratified by IGHV status. *Figure adapted from Bruch et al. (2021) .*

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.

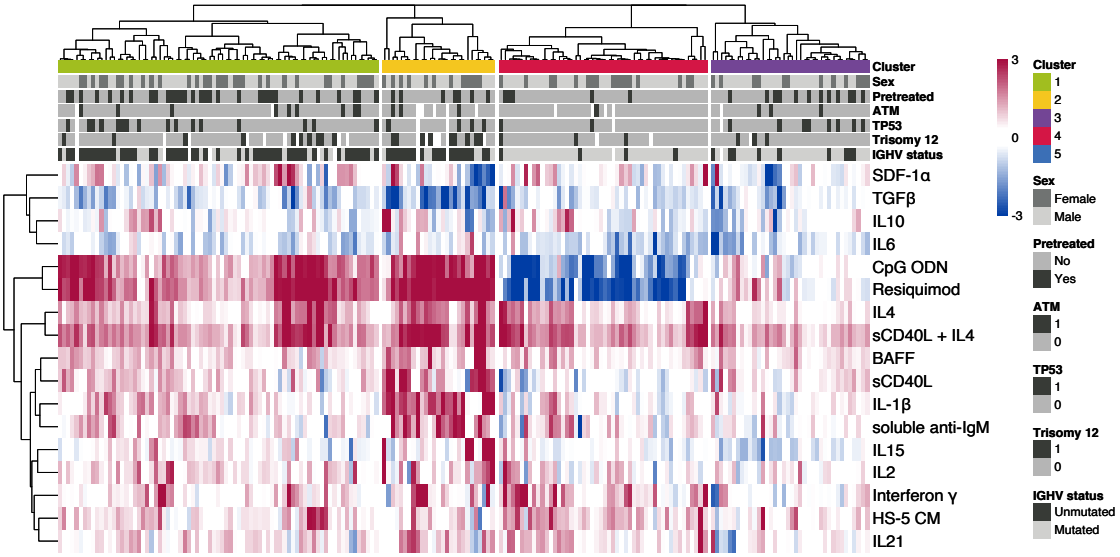

Figure 4.4: The heatmap matrix shows viability measurements for 192 samples (columns) and 17 stimuli (rows). The data are shown z-scores of log-transformed, control-normalised viability values. The colour bars to the right show sample annotations. Consensus clustering was used to define column tree layout, 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).

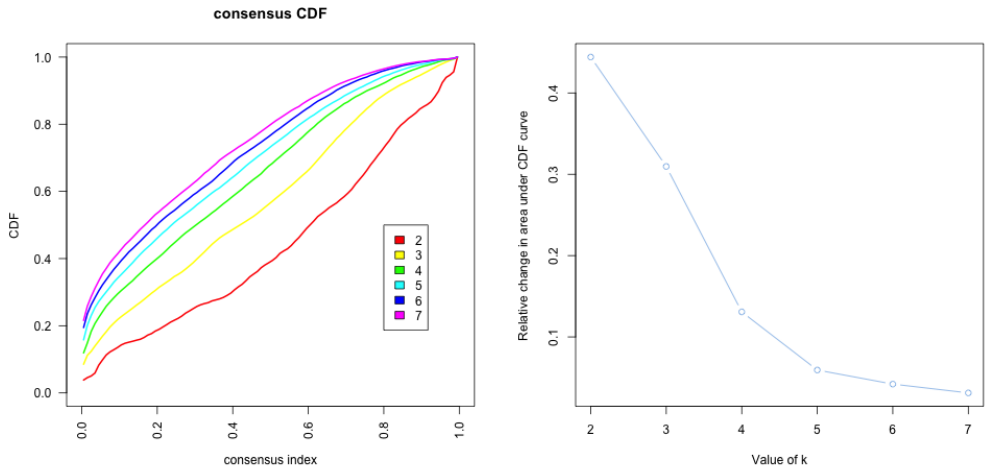

Figure 4.5: (left) CDFs of the consensus matrices for k = 2 - 7, as indicated in the legend, estimated using 100 bin histogram. (right) Relative change in area under the CDF curve, for k = 2 - 7, to compare k with k - 1. In the case of k = 2, there is no k - 1, so the total area is plotted. Line shows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

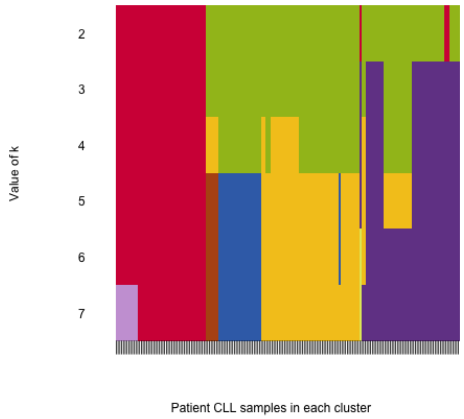

Figure 4.6: Assignment of patient samples (columns), to each cluster, for k = 1 - 7 (rows) to demonstrate stability of cluster membership. Cluster colour for k = 4 match those in heatmap in 4.4. See Methods section 2.5.2.

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).

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

[3] 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.

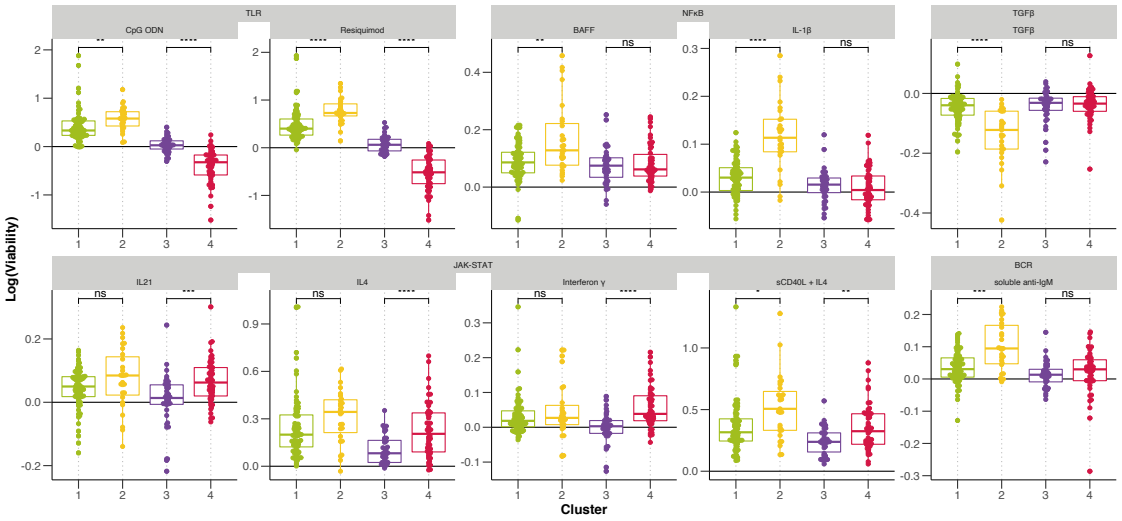

Figure 4.7: Log-transformed normalised viability values stratified by cluster, for each stimulus. Stimuli activating the same pathway are grouped together. P-values from Student's t-test. *Figure adapted from Bruch et al. (2021).*

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

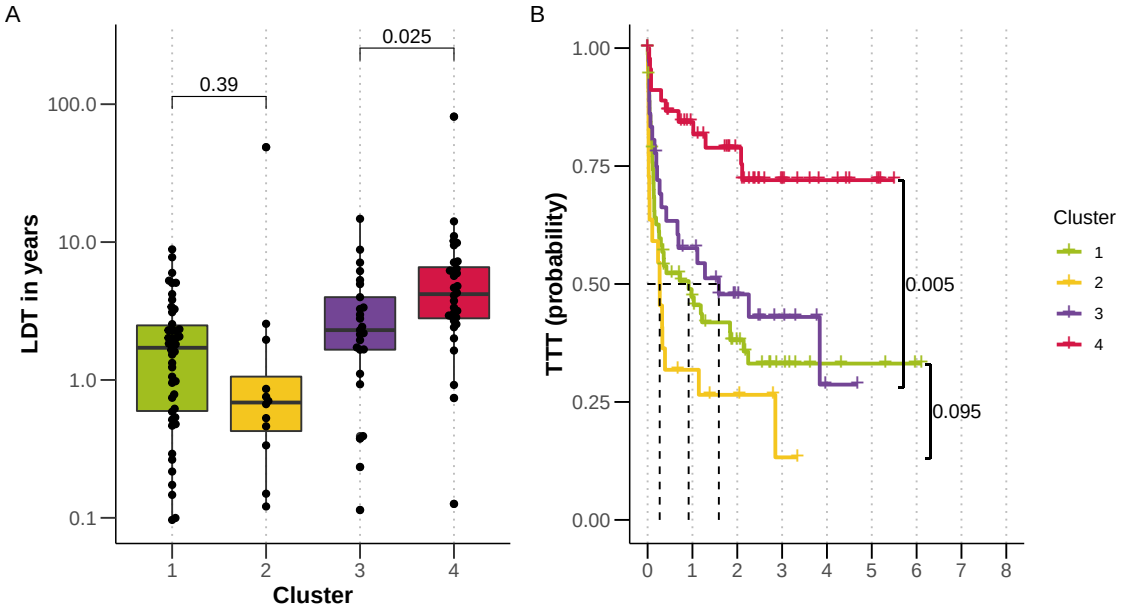

Figure 4.8: (A) LDT stratified by cluster, p-values from Student's t-test. (B) Kaplan-Meier curves to show TTT for each cluster. p-values from univariate Cox proportional hazard models comparing IGHV-U enriched C1 with C2, and IGHV-M enriched C3 with C4. See Methods sections **??** and 2.5.2. *Figure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

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 λmin.

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

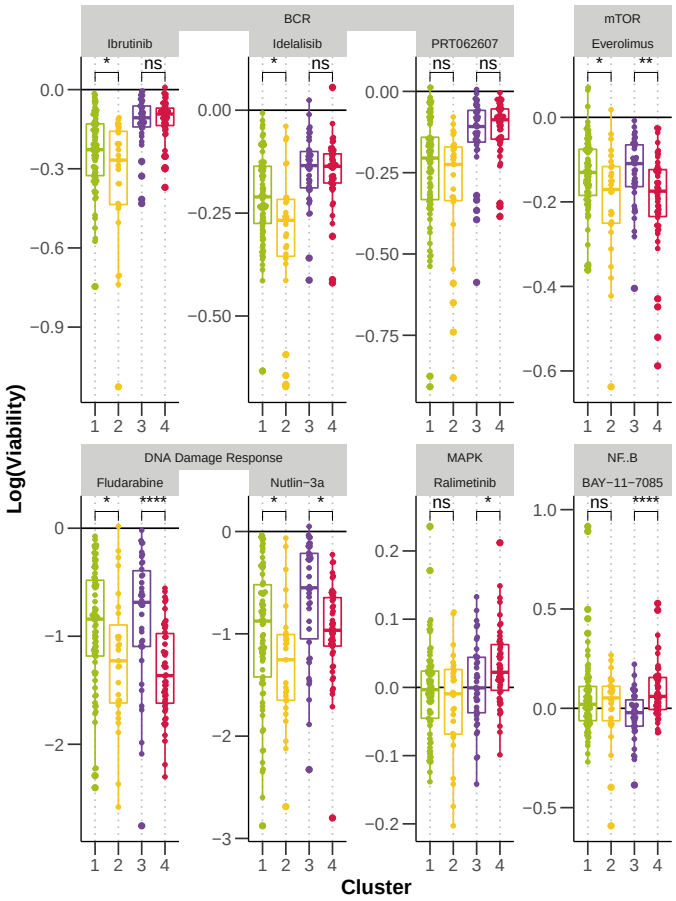

Figure 4.9: Log-transformed normalised viability values, stratified by cluster, for each drug. Drugs targeting the same pathway are grouped together. P-values from Student's t-test.

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,

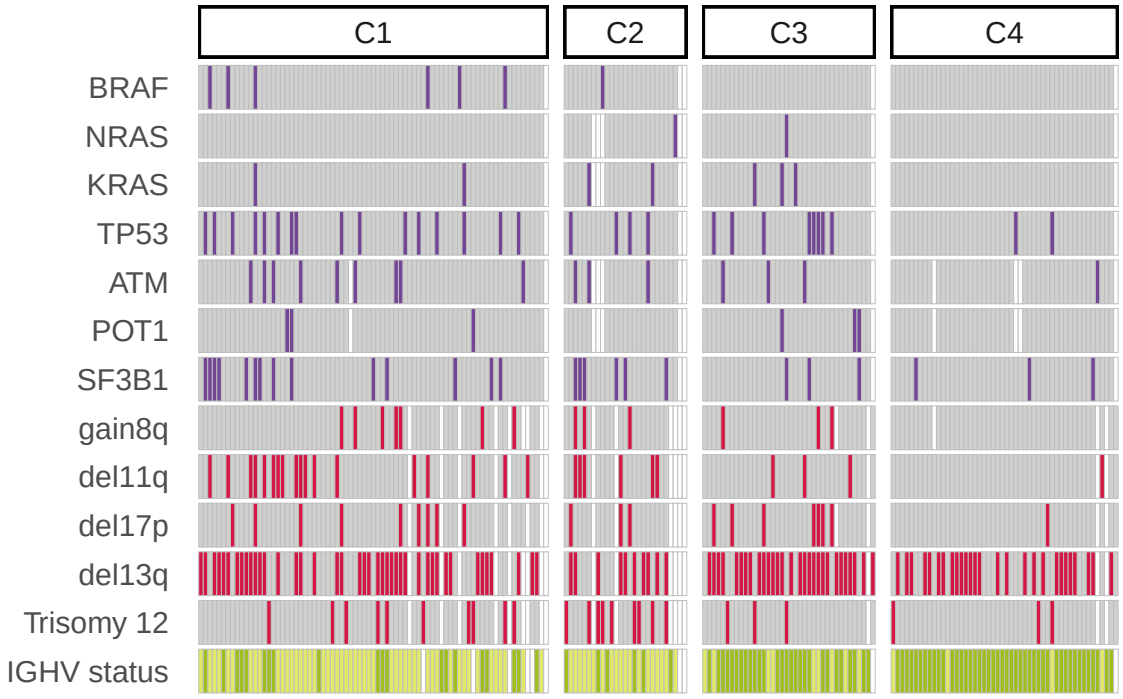

Figure 4.10: Distribution of selected genetic features (rows) within each cluster for all patient samples (columns). White indicates patient sample is not annotated for that feature.

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

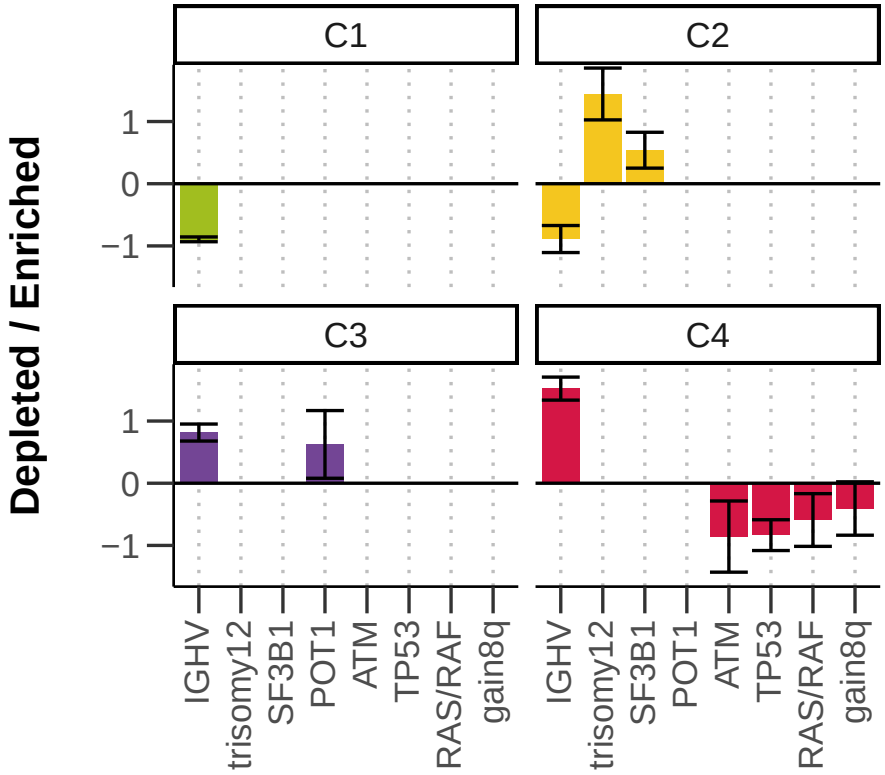

Figure 4.11: Enrichment and depletion of genetic features within each cluster, quantified using lasso-penalised regression. x axis shows genetic features, y axis indicates value and sign of coefficient assigned to feature, for each cluster (positive coefficients are enriched in the cluster, negative 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.

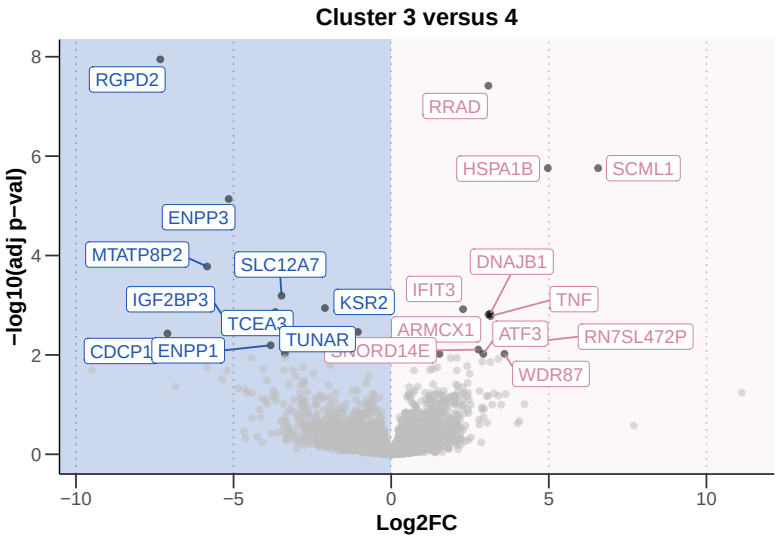

Figure 4.12: Volcano plot of differentially expressed genes between C3 and C4. x axis indicates log2 fold change values, calculated using the `DESeq2` package (**DESeq2?**), y axis gives corresponding -log10(BH-adjusted p value). Darker grey points are labelled where adjusted p < 0.01. See Methods section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

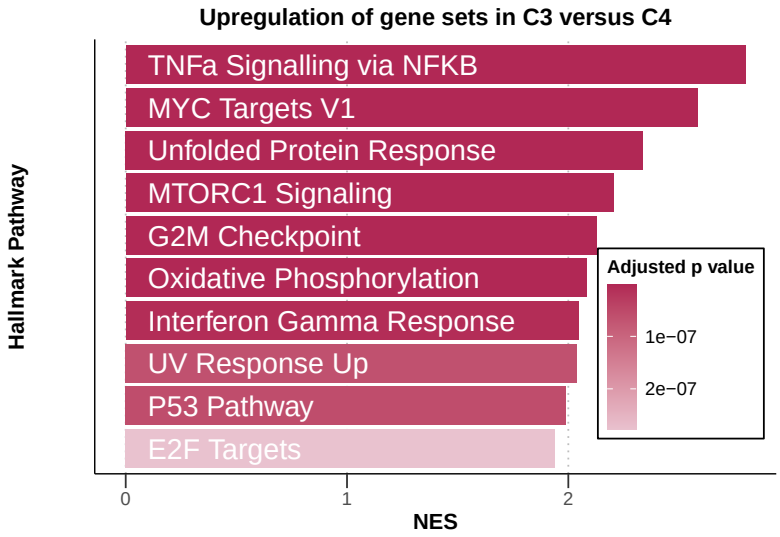

Figure 4.13: Bar plot showing gene set enrichment to compare gene expression of samples in C3 and C4. Normalised enrichment scores (NES) are shown for top ten pathways with smallest adjusted p-value. Gene ranking performed based on Wald statistic, calculated using the `DESeq2` package. See Methods section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

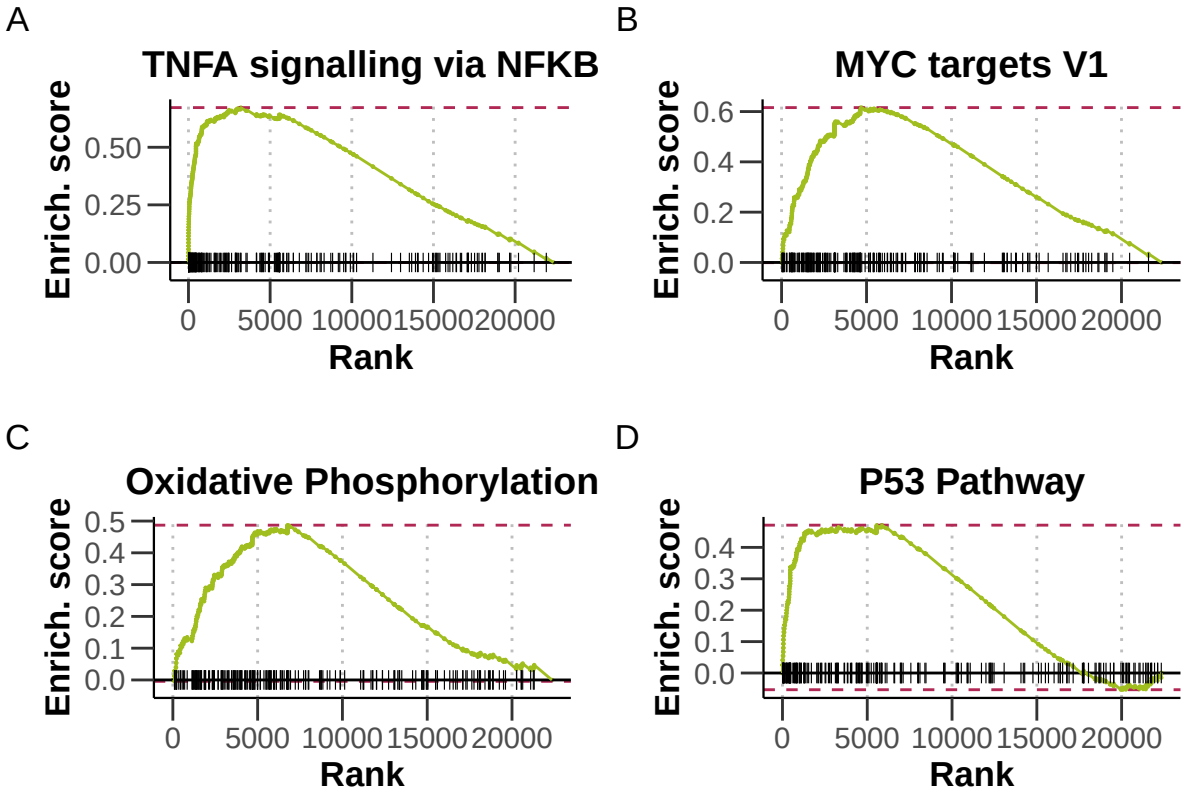

Figure 4.14: Enrichment plot of selected pathways. Gene set enrichment analysis (GSEA) was performed with the Hallmark gene sets from the GSEA Molecular Signatures Database. Wald statistic was used to rank the genes. The green curve corresponds to the Enrichment Score curve, which is the running sum of the weighted enrichment score obtained from GSEA software. See Methods section 2.5.2. *Figure and caption from Bruch & Giles et al. 2021.*