Chapter 3

Data

This chapter provides an overview of the datasets that are central to this thesis. It also covers the generation of resources to make the data and associated analysis publicly-available.

3.1 Experimental overview

3.1.1 Drug-stimulus combinatorial pertubation assay and patient sample multi-omic profiling

This thesis centres on a dataset of 192 CLL patient samples subjected to functional and molecular profiling. A drug-stimulus combinatorial perturbation assay (refered to below as the screen) measured the effects of 17 cytokines and microenvironmental stimuli alone and in combination with 12 drugs, to investigate the influence on spontaneous and drug-induced apoptosis.

The screen was primarily performed by Peter-Martin Bruch in the Department of Medicine, University of Heidelberg, and published in the manuscript by Bruch and Giles et al. 2021. Drugs and stimuli were deposited first and patient samples second, using two 384-well plates per patient. Each plate contained stimulus, drug and drug - stimulus wells along with DMSO control wells. After 48 hours of incubation at 37°C, cell viability was assessed using the CellTiter-Glo Luminescent Cell Viability Assay. This method calculates the number of viable cells in a culture by determining the quantity of ATP present, which indicates the number of metabolically active cells. The drugs and stimuli tested in the screens are outlined in sections 3.3 and 3.4.

Multi-omics profiles for the patient samples were also available from the PACE repository (Oles et al. 2021), consisting of whole-exome sequencing, DNA-methylation, RNA-sequencing and copy number variant data. In addition, clinical follow-up data was also available for some patients, including LDT, TTT, TTFT and OS.

Collectively, these data enabled (i)characterisation of responses to microenvironmental stimuli, (ii) definition of functional patient subgroups, (iii) profiling of molecular determinants of drug and stimulus responses (iv) mapping of drug - stimulus and drug - stimulus - gene interactions, thus shedding light on the heterogeneity of CLL biology and drug response (Figure 3.1).

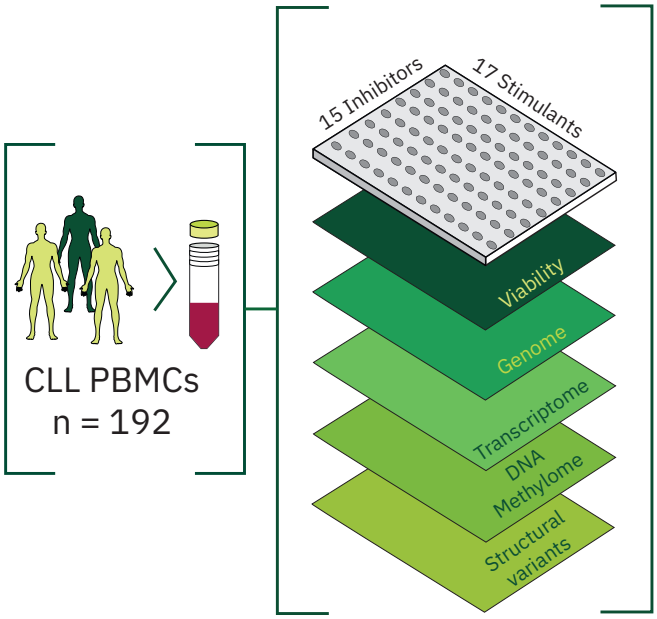


Figure 3.1: Schematic of experimental protocol. By combining 12 drugs and 17 stimuli, we systematically queried the effects of simultaneous stimulation and inhibition of critical pathways in CLL (n = 192). Integrating functional drug-stimulus response profiling with four additional omic layers, we identified 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.*

3.1.2 Additional datasets

A number of key findings emerged from the above data which warranted further investigation. In addition to the screen, this thesis makes use of a number of validatory datasets, from within our lab and from external sources. These are outlined in the Methods (Chapter 2) and in the relevant results chapters.

3.2 Data Processing

3.2.1 Processing the raw values obtained from the screen

This section describes the process of generating viability scores used in downstream analyses (following the same process as published in Bruch and Giles et al. 2021). Initially, raw luminescence measurements from the experiments were read in using custom-made `R` scripts and functions. Raw values represent the luminescence readout of the CellTiter-Glo Luminescent Cell Viability Assay, which is proportional to the amount of ATP present. The ATP levels are directly proportional to the number of viable cells in the well. Figure 3.2 shows the raw viability values from the DMSO-treated wells for each patient sample. The absolute values vary between patient samples, and between measuring dates, and thus the viability values required additional normalisation.

Each viability value was normalised to internal DMSO values of the same plate. Specifically, the mean of the two wells for each stimulus, drug or drug-stimulus treatment was divided by the median of the 50 DMSO negative control wells present on each plate, resulting in viability scores. [CHECK] A value of 0 indicated that all cells were killed by the treatment, and a value of >= 1 indicated the cells survived as well as or better than negative control.

Given the downstream analysis involved the use of linear modelling, I thus applied a log base e transformation to the viability scores. This generated log-transformed control-normalised viability scores which are used for the majority of the downstream analysis. Here 0 indicates that cells survived as well as negative control.

3.2.2 Quality control and data reproducibility

Next, several additional quality control steps were considered, with the aim of a) adjusting for any spatial effects on each screening plate, b) accounting for any batch effects

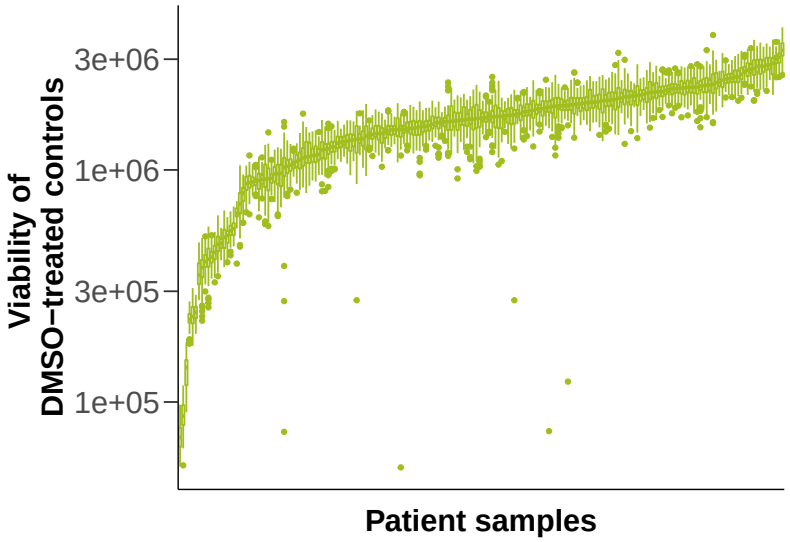


Figure 3.2: Boxplots of raw viability count data prior to normalisation and log transformation. For each of the 192 patient samples, there were 50 DMSO-treated wells. Viability represents the luminescence readout of the CellTiter-Glo Luminescent Cell Viability Assay, which is proportional to amount of ATP present.

between screening batches, c) testing data reproducibility.

First, the screening data for each well on each plate was plotted in a grid corresponding to the plate layout, to visualise whether viability values were affected by their position on the plate. We discussed adjusting for any position effect by fitting a surface to the negative control wells, to generate correction factors for each well for each plate. The resulting correction factor was then subtracted from each of the treatment wells. We decided that the position effect was not sufficient to warrant this adjustment, and continued the analysis with the unadjusted values. [CHECK]

With regards to batch effects, screening was performed on 16 separate days. We concluded that normalising data to the DMSO-treated wells on each plate sufficiently accounted for this. [CHECK]. Finally, with respect to data reproducibility, each plate was tested twice, such that the resulting viability scores are the mean of two duplicates [CHECK].

Thus, the data processing steps generated robust, log-transformed control-normalised viability scores that aim to capture the biological impact of the concomitant application of drugs and stimuli, outlined below.

3.3 Characteristics of drugs used in the screen

3.3.1 The panel of drugs

We aimed to generate a systematic study of the impact of soluble factors on therapies in CLL (Bruch and Giles et al. 2021). In addition, we were interested to dissect the effect of simultaneous inhibition and stimuli of critical pathways in CLL. To that end, a panel of 17 drugs was constructed, encompassing FDA-approved therapies for CLL, along with a number of drugs in clinical trial or laboratory compounds targeting pathways of interest (Figure 3.3). These included fludarabine (a frontline chemotherapeutic) and ibrutinib and idelalisib (newer BCR inhibitors).

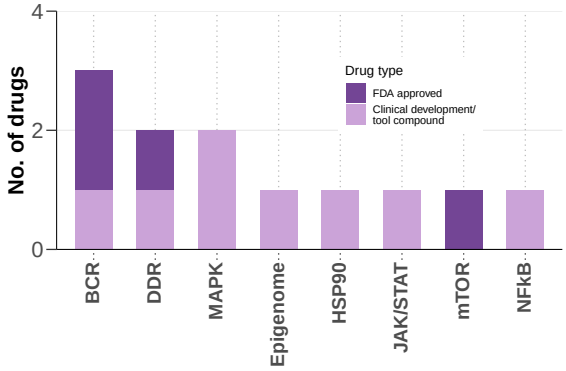


Figure 3.3: Bar plot of the drugs used in screen, indicating targets and licencing status. *Figure from Bruch and Giles et al. 2021).*

The number and concentration of drugs included was limited by the size of the plate. Thus there is minimal overlap between drug targets, and two concentrations were used for each drug. The choice of concentration was guided by the results of a previous drug screen in CLL patients, performed in our lab (Dietrich et al. 2017). The concentrations used were expected to reduce CLL viability without eliminating all cells. Drug concentrations are shown in Table 3.1.

3.3.2 Assessing drug response

To assess the quality of the drug response data, I quantified correlation coefficients for every drug pair (Bruch and Giles et al. 2021). Drugs were highly correlated if they shared identical target pathways, suggesting that the screen captures inter-individual differences in pathway dependencies, both sensitively and specifically. For example, BCR inhibitors ibrutinib, idelalisib, PRT062607 and selumetinib were all highly corre-

Table 3.1: Drug characteristics, including targets, concentrations and licencing status.

Drug	Main targets	Target category	Drug Group	Conc. 1	Conc. 2
Ibrutinib	BTK	BCR	FDA approved	500nM	50nM
Idelalisib	PI3K delta	BCR	FDA approved	500nM	50nM
Fludarabine	Purine analogue	DDR	FDA approved	2000nM	200nM
Nutlin-3a	MDM2	DDR	Clinical development/ tool compound	10000nM	1000nM
Selumetinib	MEK1/2	MAPK	Clinical development/ tool compound	1000nM	100nM
BAY-11-7085	NFkB	NFkB	Clinical development/ tool compound	2000nM	200nM
Everolimus	mTOR	mTOR	FDA approved	500nM	50nM
PRT062607	SYK	BCR	Clinical development/ tool compound	500nM	50nM
Pyridone-6	JAK1/2/3	JAK/STAT	Clinical development/ tool compound	500nM	50nM
Ralimetinib	p38 MAPK	MAPK	Clinical development/ tool compound	1500nM	150nM
Luminespib	HSP90	HSP90	Clinical development/ tool compound	200nM	20nM
I-BET 762	BRD2/3/4	Epigenome	Clinical development/ tool compound	1000nM	100nM

lated (Figure 3.4).

In addition, the individual drug response profiles (Figure 3.5) indicated that each of the drugs decreased CLL viability as expected, and in line with previous CLL drug screens performed in our lab (Dietrich et al. 2017).

3.4 Characteristics of stimuli used in the screen

3.4.1 The panel of stimuli

The stimuli selected for the screen, their associated targets and concentrations used are described in Table 3.2. They include 16 individual stimuli, plus HS-5 Culture Medium which encompasses the range of soluble factors secreted by the stromal cell line HS-5 (Bruch and Giles et al. 2021.) A number of studies have demonstrated the ability of various soluble factors to increase CLL viability or induce drug resistance *ex vivo* (see section 1.3.3. Guided by these observations, the panel of stimuli was selected so as to cover a range of key survival signals in CLL, aiming to minimise redundancy amongst the targeted pathways. The stimuli encompass a cross-section of the complex communication network between CLL cells and non-neoplastic cells, mediated by soluble factors within the tumour microenvironment. 3.6.

The stimuli activate a number of critical pathways in CLL, including BCR, TLR, JAK-STAT, NFκB and TGFβ. For more information on the importance of these pathways see section 1.1.4 and 1.3.3. Amongst these, the roles of BCR, IL4, sCD40L and TLR stimulation were of particular interest.

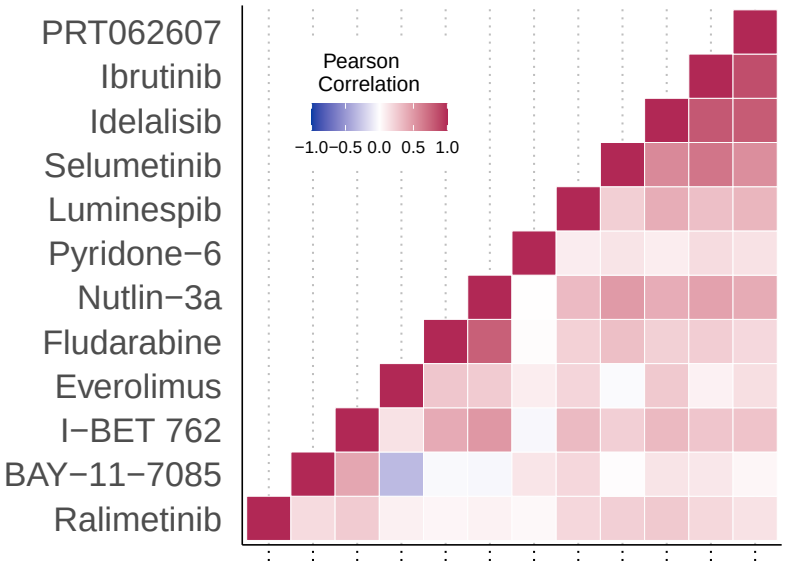


Figure 3.4: Heatmap of Pearson correlation coefficients of each pair of drugs, based on log transformed viability values. See Methods section 2.4.2. *Figure adapted from Bruch and Giles et al. 2021.*

3.4.2 Assessing stimulus response

Before beginning the downstream analysis, it was first important to determine whether heterogeneity of response to the stimuli could be caused by differences receptor expression, rather than some other cell-intrinsic feature (Bruch and Giles et al. 2021). This would guide our interpretation of the stimulus responses.

To determine this, we calculated Pearson correlation coefficients to compare log-transformed control - normalised viability values for each stimulus, with vst-transformed RNA counts of the matching receptor(s). All Pearson coefficients were less than 0.4, indicating that heterogeneity of response was not related to differential receptor expression.

3.5 Characteristics of patient samples used in the screen

3.5.1 Overview of the molecular profiles of the patient samples

Multi-omics profiles were available for the patient samples in the screen, taken from the Primary Cancer Cell Encyclopedia (PACE) (Oles et al. 2021). The PACE repository represents an initiative by our lab to characterise the multi-omic characteristics of primary tumour samples from leukemia and lymphoma patients. Patient multi-omic profiles

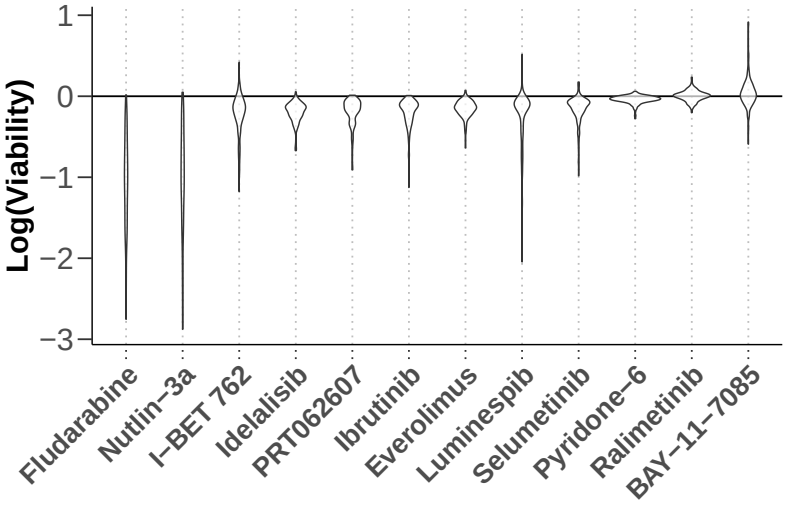


Figure 3.5: Log transformed control-normalised viability values for all drugs that were included in the screen after quality control. p values from Student's t-test.

included whole-exome sequencing, DNA-methylation, RNA-sequencing and copy number variant data. In addition, clinical information and follow-up data was also available for some patients, including their sex, IGHV status and LDT, TTT, TTFT and OS. Figure **??** summarises the molecular characteristics of the patient samples. For a summary of the number of samples with each mutation, see Appendix Table **??**.

The findings from this study have potential relevance in a clinical setting. Thus it was important to ensure that the distribution of genetic features amongst the cohort was representative of those observed in clinical practice. Other studies have determined the expected frequency of many recurrent genetic features in CLL (H. Döhner et al. 2000; Xose S. Puente et al. 2015; Landau et al. 2015) : the distribution of molecular lesions in our cohort is comparable to these (Bruch and Giles et al. 2021).

3.6 Publishing the data and assosciated analysis

Collectively, this dataset represents a valuable resource providing the ability to explore genetic, epigenetic and microenvironmental modulators of survival and drug response in a heterogeneous cohort, and how these relate to clinical outcomes. With CLL samples relatively simple to obtain compared to other cancers, this project (in collaboration with the PACE initiative (Oles et al. 2021)) represents a considerable dataset containing joint functional and molecular profiling of primary cancer samples.

Table 3.2: Stimuli names, targets pathways and concentrations.

Name	Concentration	Pathway
IL4	10 ng/ml	JAK/STAT
IL10	10 ng/ml	JAK/STAT
IL2	10 ng/ml	JAK/STAT
Resiquimod	1000 ng/ml	TLR 7/8
IL21	10 ng/ml	JAK/STAT
BAFF	250 ng/ml	NFkB
IL1	10 ng/ml	NFkB
sCD40L	1000 ng/ml	NFkB
soluble anti-IgM	20000 ng/ml	BCR
TGF1	10 ng/ml	MAPK
IL15	10 ng/ml	JAK/STAT
IL6	10 ng/ml	JAK/STAT
CpG ODN	1000 ng/ml	TLR 9
SDF-1	200 ng/ml	JAK/STAT
Interferon	5 ng/ml	NFkB
HS-5 CM	20 %	NA

An important goal of my work was thus to ensure that this dataset was both publicly available and accessible for a range of users, including medics and bioinformaticians. I aimed to ensure that our analysis was transparent and reproducible, and that others could explore the dataset for the purposes of their own research. To that end, I developed a shiny app to explore the screening dataset, and published all code and data from the manuscript Bruch and Giles et al. 2021 to an online git repository.

3.6.1 Shiny app

The shiny app can be found here. Figure 3.10 shows the home page. It consists of four tabs, covering the following:

- ***Drug and stimulus responses*** Explore drug - stimulus interactions and view log-transformed viabilities with single and combinatorial treatments (Figure 3.11).
- ***Effects of mutations on drug and stimulus responses*** Explore how drug and stimulus responses are modulated by genetic features and view log-transformed viability data stratified by mutations
- ***Genetic predictors of drug and stimulus responses*** Explore how drug and stimulus responses are modulated by genetic features with predictor profiles from section 5.1.2
- ***Genetic predictors of drug and stimulus interactions*** Explore how drug - stim-

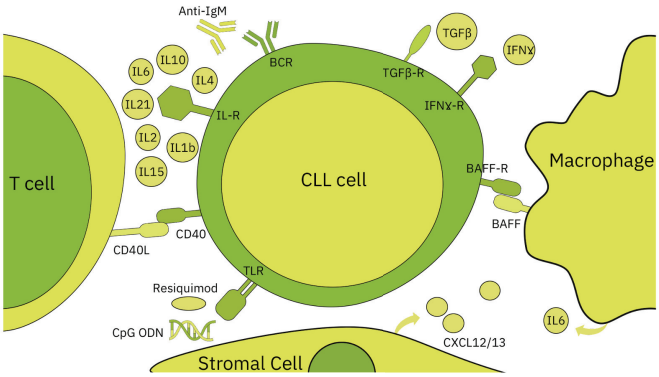


Figure 3.6: A selection of interactions between CLL cells and components of the microenvironment, covered by the screen. (See section 1.3.3 for more details on these signals). *Figure adapted from an original published in (**Wiestner2015?**) .*

ulus interactions are further modulated by mutations, and view predictor profiles from section 6.3.2

To generate the app, I first curated the individual datasets and ran each of the individual analyses required to generate the plots outline above. I then set up the four tab structure, and adapted the code required to generate plots in a dynamic manner. Finally, I worked on the aesthetics and interface of the app, to ensure that it was both professional and understandable. I then tested the app with number of colleagues, to ensure it was accessible and understandable by a range of users both familiar and unfamiliar with the project. I maintain the app on the university server.

3.6.2 Online code repository

The online repository can be found here.The repository consists of the data and executable transcripts to completely reproduce the analysis described in Bruch and Giles et al. 2021 (Figure 3.12).

To generate the repository, I first curated each of the individual datasets, including various objects containing the screening data, patient genetic meta data, ATACseq processed data, RNAseq counts, clinical data include LDT, TTT and OS data, plus the follow up data including lymph node IHV experiments, shRNA knockdown experiments, and additional stimulation and inhibition assays. As many of these data contain sensitive information on patients, I anonymised each object by updating the patient IDs and removing potential identifying features such as age.

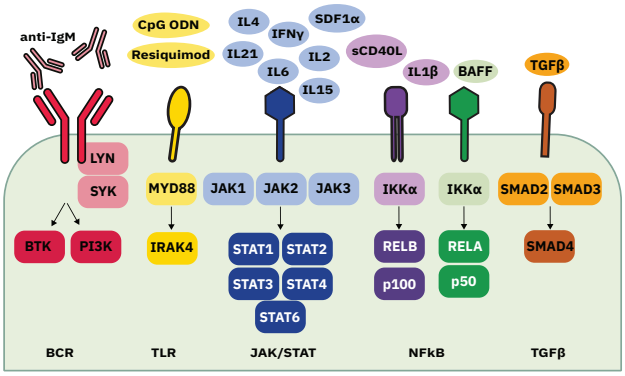
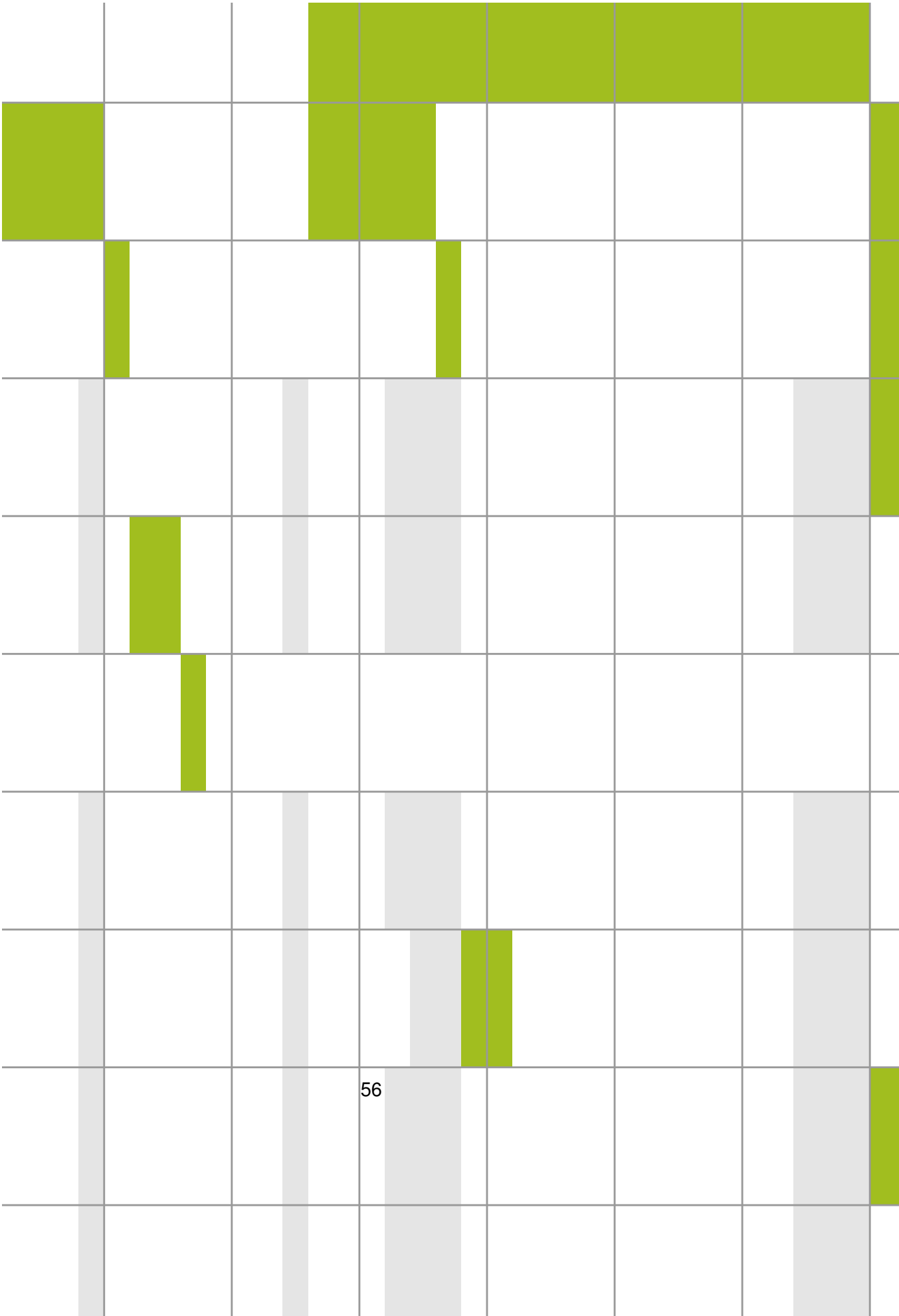


Figure 3.7: Overview of stimuli included in the screen and summary of their associated targets. HS-5 Culture Medium is omitted, as no specific target can be shown. *Figure and caption from Bruch & Giles et al. 2021.*

I next arranged the analysis into seven separate scripts, one for each figure, such that all individual sections can be rendered into a single html vignette outlining the entire analysis. I ensured that the code in each script was well-annotated and relatively simple to understand and to follow. I shared the code with several colleagues to receive feedback on coding style, and ensured that the analysis could be reproduced by others, and on different operating systems. I published the repository with the along with the preprint (Bruch and Giles et al. 2021).

These online resources are already beginning to serve as a community resource (e.g. J. Lu et al. (2021), Nature Cancer), as querying them enables researchers to test new hypotheses within minutes and may obviate the need for certain small-scale experiments.



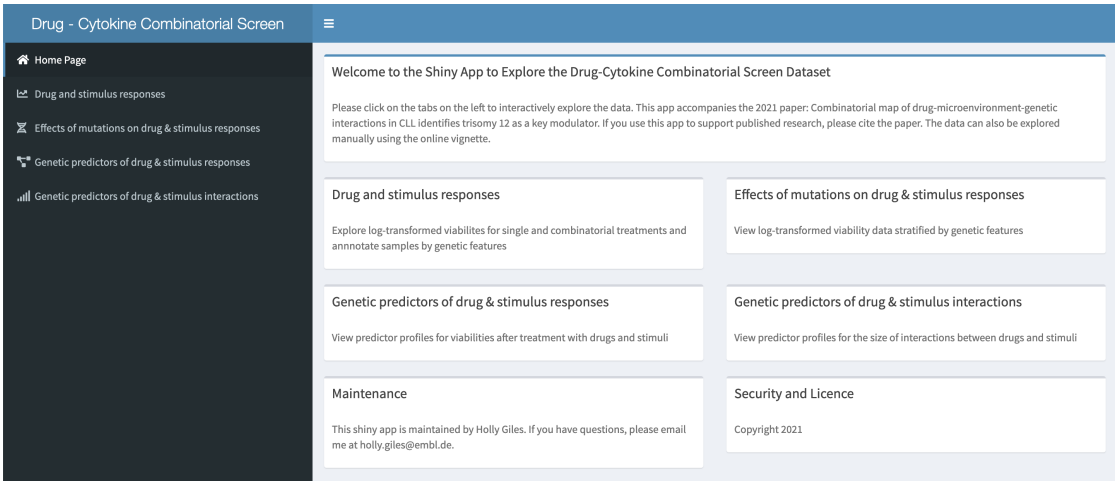


Figure 3.10: Home page of the shiny app accompanying this project. The app was published alongside Bruch and Giles et al. 2021.

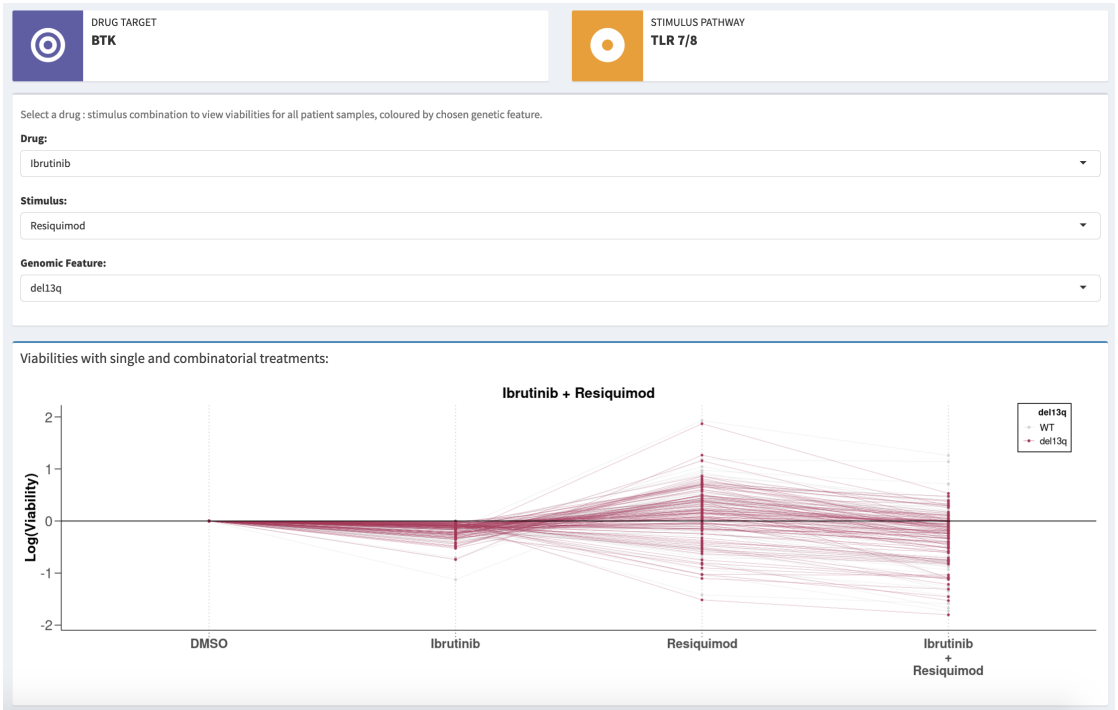


Figure 3.11: Image of one of the shiny tabs, in which the user can explore log-transformed viability values for different drug and stimulus treatments, stratified by genetic features.

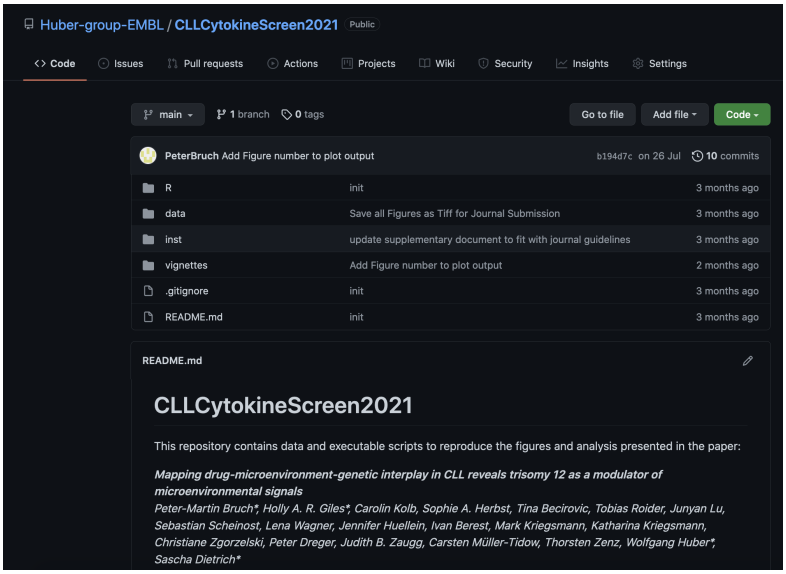


Figure 3.12: Interface of the online code repository