PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

# Inferring tumor-specific cancer dependencies through integrating *ex vivo* drug response assays and drug-protein profiling

**Alina Batzilla**[1,2☯], **Junyan Lu**[1,3,4☯*], **Jarno Kivioja**[5], **Kerstin Putzker**[6], **Joe Lewis**[6], **Thorsten Zenz**[5*], **Wolfgang Huber**[1,4*]

**1** Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany, **2** Faculty of Biosciences, Heidelberg University, Heidelberg, Germany, **3** Medical Faculty Heidelberg, Heidelberg University, Heidelberg, Germany, **4** Molecular Medicine Partnership Unit (MMPU), Heidelberg, Germany, **5** Department of Medical Oncology and Hematology, University Hospital Zürich and University of Zürich, Zürich, Switzerland, **6** Chemical Biology Core Facility, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

☯ These authors contributed equally to this work.
* junyan.lu@uni-heidelberg.de (JL); thorsten.zenz@usz.ch (TZ); wolfgang.huber@embl.org (WH)





**Citation:** Batzilla A, Lu J, Kivioja J, Putzker K, Lewis J, Zenz T, et al. (2022) Inferring tumor-specific cancer dependencies through integrating *ex vivo* drug response assays and drug-protein profiling. PLoS Comput Biol 18(8): e1010438. https://doi.org/10.1371/journal.pcbi.1010438

**Editor:** James Gallo, University at Buffalo - The State University of New York, UNITED STATES

**Received:** April 18, 2022

**Accepted:** July 26, 2022

**Published:** August 22, 2022



**Data Availability Statement:** All data and code used in this study are in the repository: https://github.com/Huber-group-EMBL/DepInfeR_workflow. A website of html reports is available at: https://www.huber.embl.de/users/jlu/depInfeR/index.html. We provide an R/Bioconductor package, DepInfeR (https://bioconductor.org/packages/release/bioc/html/DepInfeR.html), for users to estimate cancer protein dependence coefficient using their own datasets.

## Abstract

The development of cancer therapies may be improved by the discovery of tumor-specific molecular dependencies. The requisite tools include genetic and chemical perturbations, each with its strengths and limitations. Chemical perturbations can be readily applied to primary cancer samples at large scale, but mechanistic understanding of hits and further pharmaceutical development is often complicated by the fact that a chemical compound has affinities to multiple proteins. To computationally infer specific molecular dependencies of individual cancers from their *ex vivo* drug sensitivity profiles, we developed a mathematical model that deconvolutes these data using measurements of protein-drug affinity profiles. Through integrating a drug-kinase profiling dataset and several drug response datasets, our method, DepInfeR, correctly identified known protein kinase dependencies, including the EGFR dependence of HER2+ breast cancer cell lines, the FLT3 dependence of acute myeloid leukemia (AML) with *FLT3*-ITD mutations and the differential dependencies on the B-cell receptor pathway in the two major subtypes of chronic lymphocytic leukemia (CLL). Furthermore, our method uncovered new subgroup-specific dependencies, including a previously unreported dependence of high-risk CLL on Checkpoint kinase 1 (CHEK1). The method also produced a detailed map of the kinase dependencies in a heterogeneous set of 117 CLL samples. The ability to deconvolute polypharmacological phenotypes into underlying causal molecular dependencies should increase the utility of high-throughput drug response assays for functional precision oncology.

## Author summary

As survival and proliferation of cancer cells depend on molecular aberrations that can be highly specific to cancer types and individual tumors, identifying such dependence is

**Funding:** JL and WH were funded by the German Federal Ministry of Education and Research (CompLS Project MOFA under grant agreement 031L0171A; SMART-CARE under grant agreement 031L0212E and 161L0213). TZ was funded by the University Hospital Zürich ("CRPP-Next Generation Drug Response Profiling for Personalized Cancer Care" project), the Swiss Cancer Research foundation (KFS-4439-02-2018), and the Monique-Dornonville-de-la-Cour Stiftung. JK was funded by postdoctoral fellowships from the Sigrid Juselius Foundation and the Finnish Cultural Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

pivotal to designing individualized tumor therapy. Chemical perturbations, through screening of bioactive compounds using primary cancer cells, provide an important tool for identifying tumor-specific dependencies. However, many chemical compounds bind multiple proteins, which complicates interpreting screening results and pinpointing the phenotype-causing target. To overcome this challenge and increase the utility of drug screening approaches for functional precision medicine, we developed a computational framework, DepInfeR, to identify tumor-specific dependencies on druggable proteins through integrating two sources of information: drug sensitivity assays and drug-protein affinity profiling. Our approach correctly identifies known kinase dependencies, which validates our approach. Furthermore, by integrating a newly generated drug screening dataset on primary tumor samples, we discovered a previously unreported survival dependence on Checkpoint kinase 1 (CHEK1) by a molecular subgroup of chronic lymphocytic leukemia samples, highlighting the clinical potential of our method.

This is a *PLOS Computational Biology* Methods paper.

## Introduction

The sustained proliferation or apoptosis resistance of cancer cells rely on abnormal pathway activities, often resulting from genetic or epigenetic alterations. The molecular processes underlying these activities, however, can vary greatly across cancer types, and even between tumors from different patients with the same cancer type. Such heterogeneity of driver dependencies can lead to differential treatment responses. A key aim of precision cancer therapy is to exploit core vulnerabilities of each individual tumor.

A powerful, scalable, and systematic approach to identify cancer-specific dependencies uses genetic perturbations such as RNAi and CRISPR/Cas9 systems to knock down or knock out a specific gene in cancer cells and observe the subsequent effect on cell growth or survival [1–4]. While RNAi and CRISPR/Cas9 screens can be applied for nearly any gene in cell line models, their use remains challenging in primary tumor samples [5,6], which poses an important limitation for clinical research. This is the case especially for indolent diseases such as chronic lymphocytic leukemia (CLL). Furthermore, significant discrepancies may exist between the effects of a genetic perturbation and targeted drug inhibition of the encoded protein [7]. This is potentially because a drug might quantitatively inhibit the enzymatic activity of a protein but leave other functions, such as scaffolding, unaffected [8], whereas the genetic perturbation completely disrupts all functions. Off-target effects [9] and side effects of the CRISPR/Cas9 system, such as inducing cellular DNA damage response [10,11], may further complicate analysis.

A major goal of cancer dependency mapping is to propose druggable targets for cancer treatment. Chemical perturbation experiments, which rely on high-throughput screening of bioactive compounds on cancer cells, provide an attractive complementary approach to genetic perturbations due to their high translational value for personalized oncology [12,13]. Chemical perturbations are readily applicable to primary tumor models, and thus can identify patient- and tumor-specific dependencies. Drug treatment of primary tumor cells also better approximate the *ex vivo* cellular effects. By design, all identified protein dependencies are "targetable". However, a major challenge is the polypharmacological nature of small molecular compounds, as many of them show a broad range of off-target effects, which obstructs the identification of the protein targets related to the desired phenotype [14].

To improve the utility of high-throughput drug sensitivity data for functional precision medicine, we developed a computational method that integrates two experimentally accessible input data matrices, the drug sensitivity profiles of cancer cell lines or primary tumors *ex vivo* (X), and the drug affinities of a set of proteins (Y), to infer information that is not directly observed, the matrix of protein dependencies of each of the cancer samples ($\hat{\beta}$). We achieve this by regularized multivariate linear regression. This framework assigns a "dependence coefficient" to each protein and each sample, and therefore may be used to gain causal understanding of functional consequences of genomic aberrations in a heterogenous disease, as well as to guide the choice of pharmacological intervention for a specific cancer type, subtype, or a patient. We called this computational framework DepInfeR and implemented it as an R/Bioconductor package (https://bioconductor.org/packages/release/bioc/html/DepInfeR.html).

As similar *ex vivo* assays can be readily performed on cell lines and primary samples, with established platforms and ready to use datasets for both drug sensitivity profiling [13,15–17] and drug-protein profiling [14], we expect our approach to be adaptable to different diseases.

## Results

We aimed to infer the extent and direction of the dependence of individual tumor samples on individual proteins. To perform this, we used two available data types: 1) a drug-protein profiling matrix, which reports binding affinities of drugs to proteins, and 2) a drug sensitivity matrix (Fig 1A), which reports the effects of each drug on cell viability across a set of tumor samples. The effects are measured as dimensionless ratios, relative cell viability compared to DMSO (dimethyl sulfoxide) control. Although our method is designed to employ any type of drug-protein profiling matrix, extant datasets focus on protein kinase inhibitors [18,19]. Protein kinases also represent one of the most important drug target classes for cancer treatment [20]. Here, we used the kinase inhibitor profiling dataset by Klaeger *et al.* [14], who immobilized 243 kinase inhibitors on kinobeads and detected presence and abundance of kinases binding after exposition of the beads to cell lysate by mass spectrometry proteomics. They reported their results in terms of $K_d$ values, which we converted into binding scores in the range between 0 and 1 (see Material and Methods). The data show a large variety in the degree of selectivity among kinase inhibitors: while some bind specifically to only a handful of kinases, many bind dozens (Fig 1B). Similarly, many kinases are bound by multiple drugs (Fig 1C). Altogether, the picture implied by the data of Klaeger *et al.* represents an intricate, spread out network of drug-kinase binding with few one-to-one pairs or clear drug-kinase clusters (S1 Fig).

The objective of our approach is to use this network to infer the kinases that a particular sample depends on for viability by overlaying the network with the effects of the drugs on the sample. This task can be posed in terms of multivariate multi-response regression involving three matrices: the known drug-protein binding matrix, a known drug-sample sensitivity matrix—below we consider three different instances of such a matrix—and the unknown protein dependence matrix (Fig 1D). While no useful inference can be expected if there is only data for a single or a small number of samples, the regression becomes increasingly identifiable as the number of samples increases. The matrix of fitted coefficient values indicates to what extent the viability of each sample depends on each protein. High positive coefficients imply strong reliance of a particular sample on a particular protein for survival, and therefore a high vulnerability towards inhibition of the protein. Conversely, proteins with coefficients close to zero are less essential for the cell's immediate survival. Negative coefficients indicate that the viability phenotype benefits from inhibition of the protein. As the regression problem is high-dimensional, we used L1 (LASSO) regularization with cross-validation to achieve

A

| Datasets | Number of samples | Number of drugs | Number of proteins (input) | Number of proteins (selected) |
|---|---|---|---|---|
| GDSC | 126 | 66 | 118 | 9 |
| BeatAML | 421 | 61 | 112 | 15 |
| EMBL2016 | 131 | 85 | 136 | 24 |

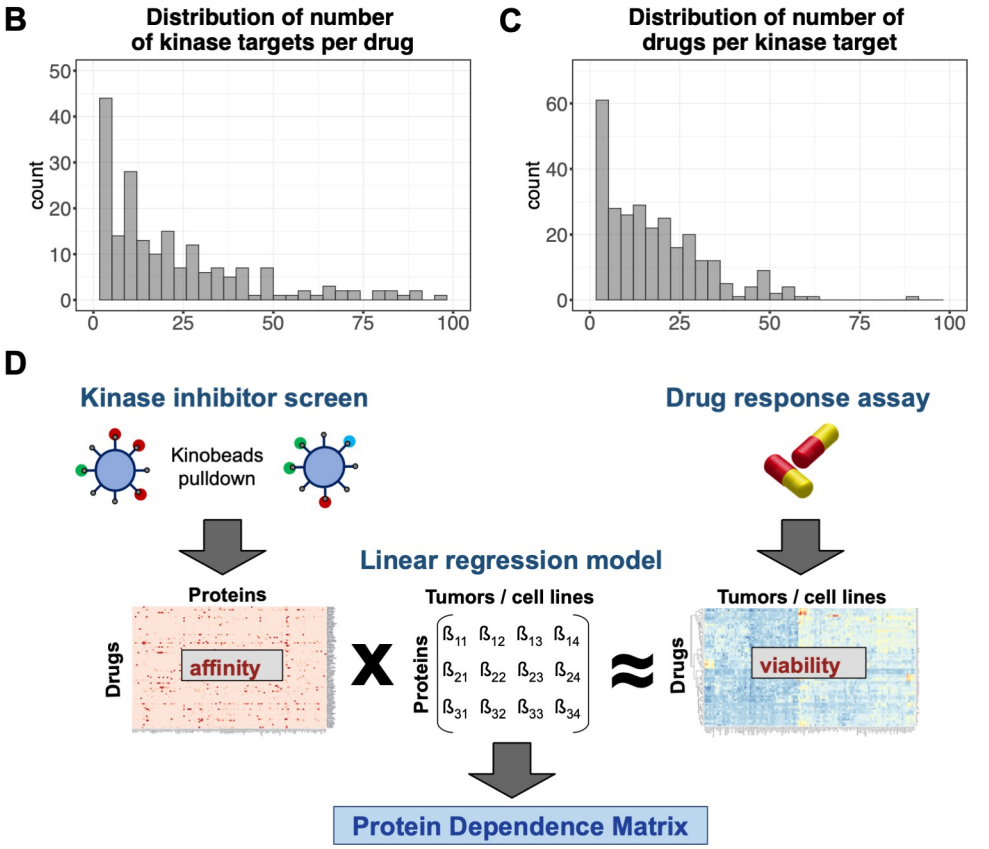



**Fig 1. Principle of the protein dependence inference framework.** (A) Summary of tumor samples and drugs in the three datasets used in this study, and the number of proteins with non-zero coefficients in the inferred protein dependence matrix. (B) Distribution of the number of kinases bound by each drug (kd < 1000nM) and (C) distribution of the number of drugs binding each kinase from the kinobeads profiling data used in our analysis. (D) Inference of the protein dependence matrix using a multivariate multi-response regression model with L1 regularization. A known drug-protein affinity matrix (independent variables) and a known drug-effect matrix (response variables) are used as input for the model to infer the unknown protein dependence matrix.

https://doi.org/10.1371/journal.pcbi.1010438.g001

parsimonious and stable regression results and select the most predictive features. Moreover, we collapsed sets of nearly collinear independent variables (i.e., proteins with highly similar drug binding profiles) into groups represented by one of the variables (S2 Fig and S1–S3 Tables). We named our method as DepInfeR and implemented an R package with the same name to perform the data preprocessing, the multivariate regression and various downstream analyses and visualizations.

As the second data source, we used three separate drug sensitivity datasets: one from cancer cell line collections and two from drug sensitivity studies of primary blood cancer samples (AML and CLL).

## Inferred protein dependence profiles reflect differential inter- and intra-cancer type dependencies in cell lines

As a first test of DepInfeR, we applied it to the cancer cell line drug sensitivity data from the GDSC database [13]. We used the subset of data for cell lines derived from AML, diffuse large B-cell lymphoma (DLBCL), acute lymphocytic leukemia (ALL), and breast carcinoma (BRCA). We concatenated these data into a single matrix of relative cell line viabilities (n = 126) under the effect of the different drugs (n = 66). Besides the relative viabilities, the matrix did not contain explicit information on cancer type or any other properties of the cell lines. We inferred the protein dependence matrix (S4 Table) and visualized it with a heatmap with rows and columns ordered in accordance with hierarchical clustering (S3 Fig). The four different cancer types were separated in these visualizations, indicating that they have distinctive protein dependence profiles (Fig 2A and 2B). The breast cancer cell lines further split into two groups, which corresponded to their HER2 status—an important prognostic marker based on the assessment of the expression level of the epidermal growth factor receptor 2 (HER2/EGFR) [21,22]. DepInfeR correctly detected the dependence of HER2-positive breast cancer cell lines on EGFR (Epidermal Growth Factor Receptor) signaling (Fig 2A and 2B). Irrespective of their HER2 status, the breast cancer cell lines showed higher vulnerability to the inhibition of JAK1 (Janus Kinase 1) than the leukemia cell lines (Fig 2A). ALL and AML cell lines showed a significantly higher dependence on the receptor tyrosine kinase FLT3 (Fig 2A), in line with the known importance of mutated *FLT3* in AML and ALL [23–25].

DepInfeR identified further, more subtle protein dependence profile differences that were associated with genetic aberrations of cell lines within one cancer type (Fig 2C). *NRAS* mutation status was associated with the dependence coefficient of multiple kinases. Its positive association with MAP2K2 (Fig 2D) reflects the higher activity of the MAPK/ERK kinase pathway in tumors driven by *NRAS* mutation [26]. Conversely, decreased sensitivity of *NRAS* mutated cell lines to the inhibition of the tyrosine-protein kinases RET and FLT3 (Fig 2D) is consistent with the fact that these two proteins are upstream activators of the RAS pathway: their inhibition is expected to be of little consequence if RAS is constitutively activated by *NRAS* mutation [27–29].

## Inferring cancer dependencies using *ex vivo* drug sensitivity profiling on primary leukemic samples

**DepInfeR method correctly infers kinase dependencies in AML.** We next applied DepInfeR to a dataset of primary tumor samples, the BeatAML data, which report the *ex vivo* drug sensitivities of 672 tumor specimens from 562 AML patients [12] (S5 Table). The inferred protein dependence coefficients indicated significantly higher dependence on FLT3 for tumors with *FLT3*-ITD mutation, a known proliferative driver in AML [30], consistent with the findings by Tyner *et al*. and in line with biological expectation (S4 Fig). An additional positive correlate of *FLT3*-ITD was the inferred dependence on the Lck kinase. This association was not described by Tyner *et al*. *2018*, but supports cooperative roles of Lck activity with *FLT3*-ITD mutation as suggested previously [31]. We also identified the vulnerability of *KRAS* and *NRAS* mutated tumors to MAPK-pathway inhibition (S4 Fig).

**The protein dependence profiles of primary CLL samples recapitulate major molecular subgroups.** Having established the feasibility of protein dependence inference on these existing datasets, we proceeded to generate a dataset tailored to the study of lymphocytic leukemias. We selected samples from 131 patients diagnosed with three prevalent leukemia types: CLL (n = 117), MCL (mantle cell lymphoma, n = 7), and T-PLL (T cell prolymphocytic leukemia, n = 7). This study design enabled us to consider entity specific properties of MCL and T-PLL,

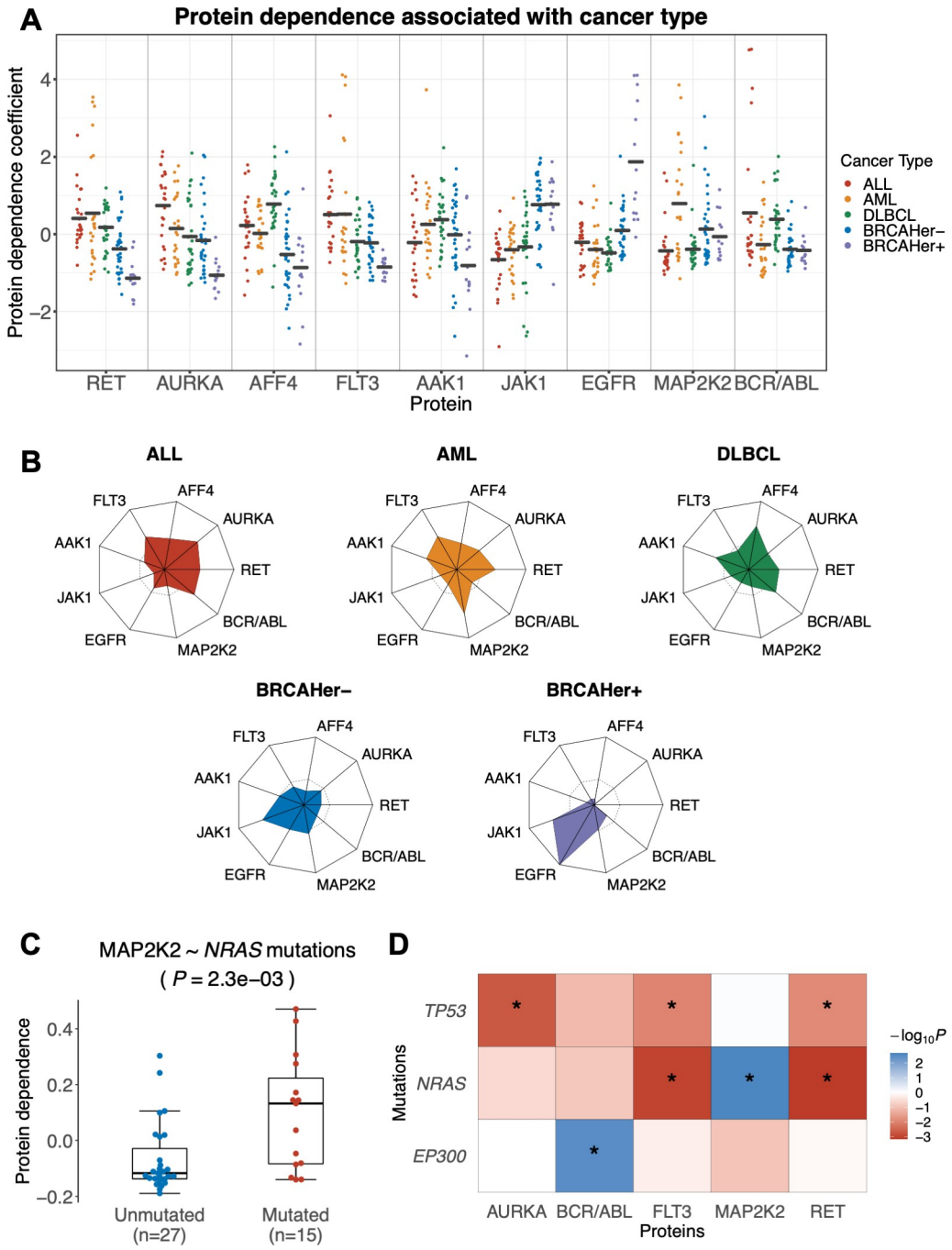



**Fig 2. Results of protein dependence inference on the GDSC dataset.** (**A**) Shown are the data for nine kinases for which we found significant differences in the protein dependence coefficients between cancer types (one-way ANOVA, BH adjusted p-value $< 0.05$, Fold Change $> 0.1$). Within each panel, each point corresponds to a cell line. The coefficients were centered and scaled to obtain a per-protein z-score, and the points are grouped and colored by cancer type (ALL, red; AML, orange; DLBCL, green, BRCAHer-, blue; BRCAHer+, purple). (**B**) Radar plots of protein dependence coefficients across the different cancer types. Dashed line represents a protein dependence coefficient of zero. (**C**) Association between *NRAS* mutation status and MAP2K2 dependence. Association testing was performed using Student's t-test (two-sided, with equal variance). (**D**) A heatmap showing the -log10(*P*-values) with signs determined by direction of fold changes of the associations between mutational background of the cell lines and protein dependence coefficients (Student's t-test). Blue: associations with higher dependence coefficients in mutated cases; red: lower dependence coefficients in mutated cases; stars indicate the associated pass 10% FDR control.

https://doi.org/10.1371/journal.pcbi.1010438.g002

and to further resolve the considerable intra-entity heterogeneity of CLL with respect to molecular factors including the mutational status of immunoglobulin heavy chain variable region (IGHV) genes, trisomy 12 status, and *TP53* mutations [16,32]. We measured drug sensitivity phenotypes of these tumors towards 85 kinase inhibitors and applied DepInfeR to infer the protein dependence map (S5 Fig and S6 Table).

We observed that kinases from the cell cycle regulation machinery, such as CDK6, CDK17, and CCNT1, had uniformly high dependence coefficients across samples, reflecting the high toxicity of inhibitors targeting cell cycle (S5 Fig). By contrast, kinases from specific signaling pathways, such as the Bruton's tyrosine kinase (BTK) from the B cell receptor (BCR) pathway, showed varied coefficients among samples (S5 and S6 Figs).

The T-PLL samples were clearly separated from CLL and MCL by their characteristically low dependence on the BTK (S6 Fig), which is consistent with the role of BTK as a main pathway component of BCR signaling and major driver of B-cell leukemias [33].

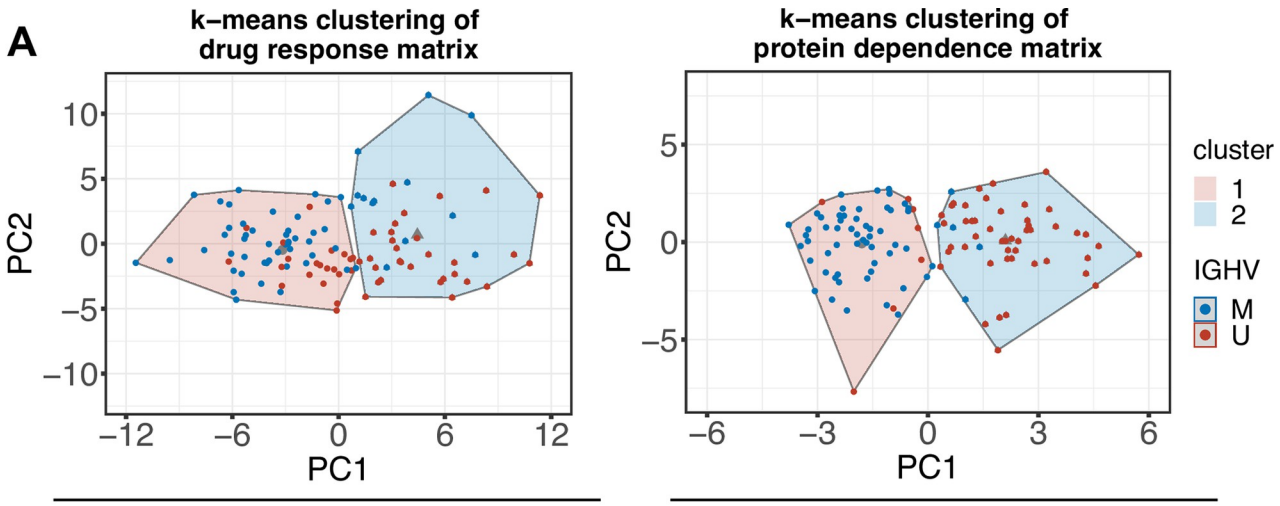
Next, we zoomed in to study the relationship between the molecular heterogeneity of our CLL samples with the protein dependencies. In a hierarchical clustering of the samples based on their protein dependence profiles, the most striking pattern was the separation between CLLs with mutated IGHV status (M-CLL) and unmutated IGHV status (U-CLL) (S5 Fig). The IGHV mutational status reflects the cell of origin in CLL and remains an important prognostic factor: U-CLL shows significantly faster disease progression and worse clinical outcome than M-CLL [34].

Clustering and other forms of pattern detection on the samples can be done on the drug sensitivity data directly, in addition to the protein dependence map. We aimed to compare the two approaches by assessing the results of k-means clustering with the Rand index [35]. We chose k = 2 in order to differentiate between U-CLL and M-CLL. While both approaches broadly separated the two subgroups, clustering based on the protein dependence map was clearer and biologically more interpretable with a Rand index of 0.62 compared to 0.15 for the drug sensitivity profiles (Fig 3A).

We looked more specifically into the associations identified between genetic features of CLL patients and the protein dependence coefficients. Most associations were found between IGHV status and protein dependence coefficients (Fig 3B): as previously described, U-CLL samples showed a higher dependence on BCR signaling components such as BTK, INPPL1, or MAPK/ERK pathway components (Fig 3B and 3C) [16,36]. The inferred protein dependence coefficient of the Salt-inducible kinase 2 (SIK2) also showed a significant association with the IGHV status (Fig 3B and 3C). Notably, several studies suggest SIK2 (and other Salt-inducible kinases) playing a role in the metabolism of B-cell lymphoma [37] and AML [38].

Trisomy 12—an important cytogenetic driver presents in around 15% of CLL [39]—was associated with higher dependence on MAP2K2 (Fig 3B and 3C). This association recapitulates a previously reported relationship between trisomy 12 and the MAPK/ERK pathway [16].

**Protein dependence inference uncovers increased dependence of U-CLL on the cell cycle checkpoint pathway.** Having validated the CLL protein dependence map with several associations to known molecular stratifiers of CLL, we looked into novel patterns. The most notable finding was the higher dependence of U-CLL on the checkpoint kinase CHEK1 (Fig 4A)–a major component of the Chk signaling pathway—compared to M-CLL. A previous study reported higher sensitivity of U-CLL to several CHEK1 inhibitors and hypothesized that this was due to off-target effects of the CHEK1 inhibitors towards components of the BCR-pathway, which are known to have differential dependence between U-CLL and M-CLL [16]. DepInfeR enabled us to further disentangle the differential effect of CHEK1 inhibitors and suggests that it is indeed due to higher dependence of U-CLL on the CHEK1 protein, rather than being due to off-target effects of the used CHEK1 inhibitors. Based on the kinobeads




**Cross tabulation: drug response**

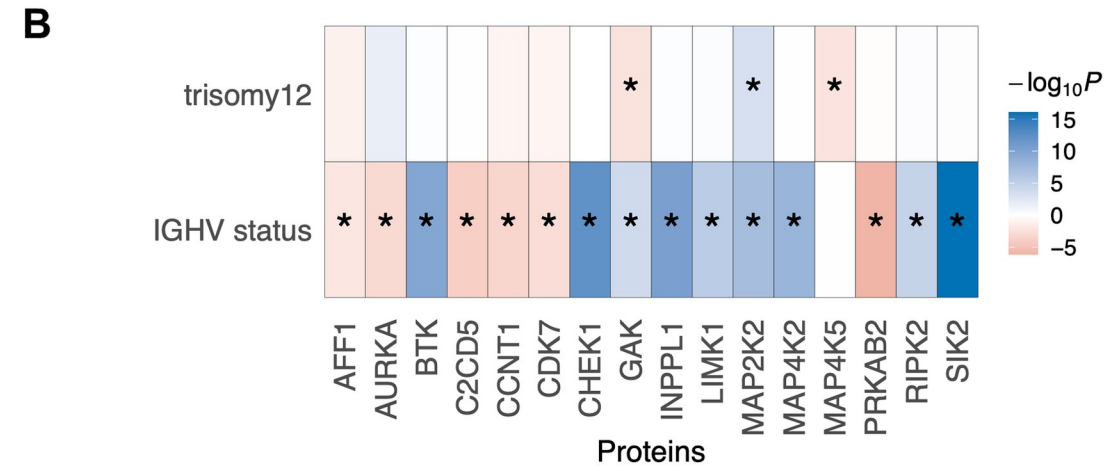
Rand index: 0.146

| IGHV | Cluster1 | Cluster2 |
|---|---|---|
| U | 22 | 33 |
| M | 47 | 13 |

**Cross tabulation: protein dependence**

Rand index: 0.623

| IGHV | Cluster1 | Cluster2 |
|---|---|---|
| U | 7 | 48 |
| M | 55 | 5 |



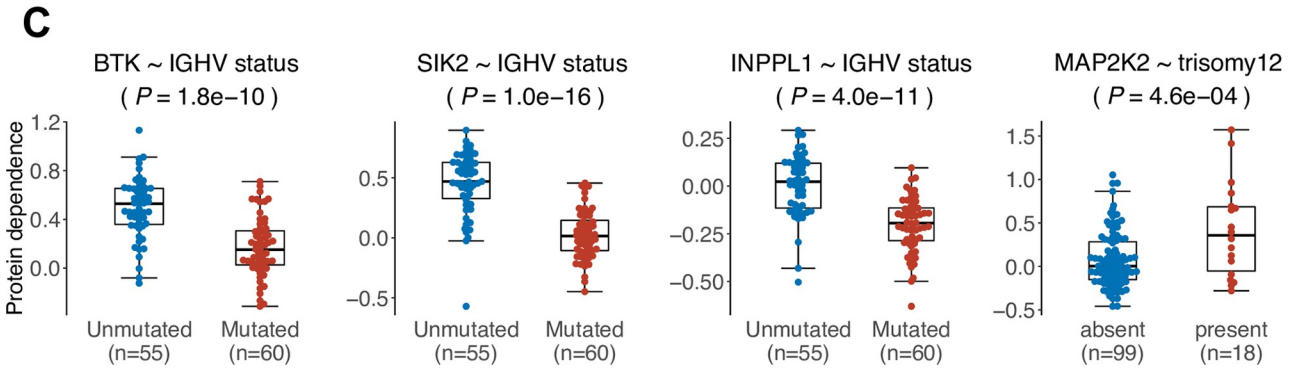



**Fig 3. Results of protein dependence inference on primary CLL samples.** (**A**) PCA visualization of the CLL samples according to protein dependence matrix (right) and drug sensitivity matrix (left). The points are colored by IGHV mutational status (mutated: blue, unmutated: red), and the results of k-means clustering (k = 2) is indicated by the shading. The cross-tabulation and Rand indices show that the protein dependence matrix-based clustering is more consistent with IGHV mutational status, a known strong stratifying factor in CLL biology (Rand index = 0.623), than the raw drug sensitivity matrix (Rand index = 0.146). (**B**) A heatmap showing the -$\log_{10}$(*P*-values) with signs determined by direction of fold changes of the associations between mutational background of the cell lines and protein dependence coefficients (Student's t-test). Blue: associations with higher dependence coefficients in trisomy 12 positive / U-CLL; red: higher dependence coefficients in Trisomy12 negative / M-CLL; stars indicate the associated pass 10% FDR control. (**C**) Examples of associations, visualized in beeswarm plots: associations between IGHV mutational status and SIK2 / INPPL1 / BTK dependence and association between trisomy 12 and MAP2K2 dependence. Association testing was performed using Student's t-test (two-sided, with equal variance) and Benjamini-Hochberg correction was applied.

https://doi.org/10.1371/journal.pcbi.1010438.g003

dataset, four out of seven CHEK1 inhibitors (kd < 1000nM) also target BCR pathway components (BTK, SYK, LYN) (Fig 4B). When looking at the effect of the other three drugs, we found a significant association with the IGHV status for two (Rabusertib, MK-8776) of the three CHEK1 inhibitors without any BCR pathway off-targets (Fig 4C). The third CHEK1 inhibitor (PF-3758309) targets 54 different kinases and binds CHEK1 only very moderately (Kd = 210.19 nM). In combination with our DepInfeR results, this indicates that the differential effect of CHEK1 inhibitors is due to the inhibition of CHEK1 itself and not only due to the BCR pathway off-target effect of these drugs.

The association of U-CLL to CHEK1 pathway was further corroborated by an analysis of differential gene expression between M-CLL and U-CLL patient samples from the data of Ferreira *et al*. (European Genome-phenome Archive; http://www.ebi.ac.uk/ega/, EGAS00001000374). Amongst the most upregulated pathways in U-CLL (S7 Fig) were DNA-damage response and cell cycle checkpoint-related pathways (G2/M checkpoints, DNA repair, E2F transcription factor targets, UV response) [40], and we found several genes upregulated that are directly involved in CHEK1-mediated checkpoint control (e.g. *CDC25A*, *CDK1*, *CCNE1*) (S8 Fig).

Moreover, we found that BCR stimulation by anti-IgM antibodies in CLL cells led to a significant upregulation of pathways related to DNA-damage response and cell cycle checkpoint, including UV response, DNA repair, E2F transcription factor targets, and G2/M checkpoints (Fig 4C). This finding is in line with previous studies showing that CHEK1 plays a role in the survival of BCR-activated and developing B-cells [41,42].

Taken together, our results suggest a stronger dependence of U-CLL on CHEK1-mediated checkpoint control. This might be either a direct effect of the increased BCR pathway activity in U-CLL or a result of increased replication stress in this more aggressive CLL subtype. Targeted CHEK1 inhibition may therefore be especially beneficial for U-CLL.

## Discussion

We developed a computational framework, DepInfeR, that integrates drug response assays and drug-protein affinity data to infer the sample-specific protein dependencies, which reflect how much the survival of the cancer cells (from a specific cell line or patient sample) depends on a certain protein.

We contrast some of the methodological choices of DepInfeR to a recent proposal by Wang et al., who pursued a similar aim by taking a drug-protein binding graph using the mean effect of all drugs connected to a given protein in the graph as a measure of dependence on it (termed *target addiction score* in their study) [43]. In contrast, our approach takes a more quantitative approach by representing protein-drug affinities with a matrix whose elements take values between 0 and 1, and if quantitative measurements such as $K_d$ values are available, uses these by mapping them to the range [0,1] with a sigmoid function. Moreover, we use a set of linear equations to relate protein dependencies (not directly observed), drug-protein affinities (data) and cell viabilities (data) and use multivariate regression with regularization to solve for the

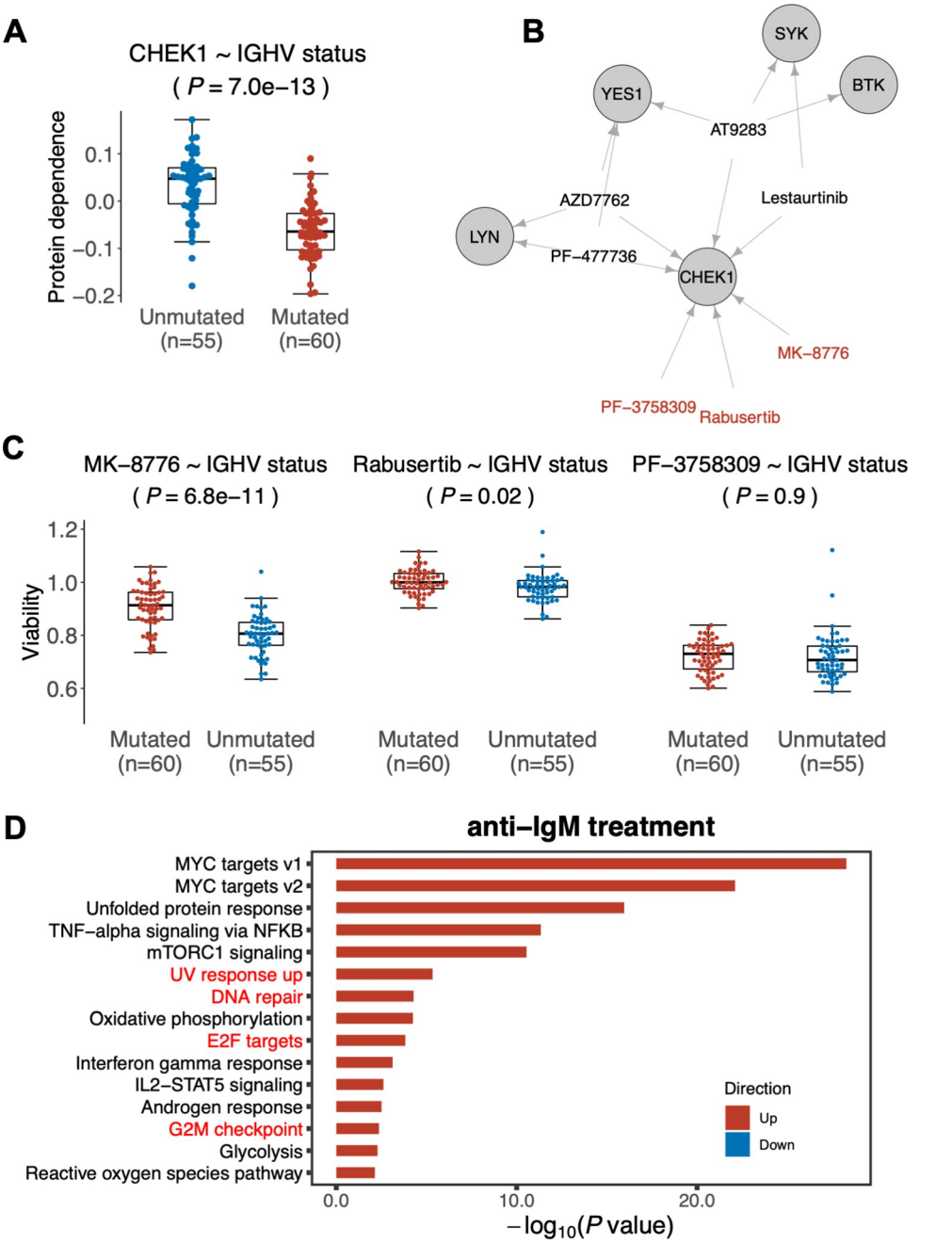



**Fig 4. Association of CHEK1 dependence with IGHV status in CLL.** (**A**) Beeswarm plot of the CHEK1 protein dependence values in all samples, visualizing the increased dependence on CHEK1 in U-CLL tumors. Association testing was performed using Student's t-test test (two-sided, with equal variance) and Benjamini-Hochberg correction was applied. **(B)** A network illustration of the off-target effect of CHEK1 inhibitors that involves BCR components (BTK, SYK, YES1 and LYN). Only the high confidence pairs in the kinobeads dataset are considered. The drugs that only target CHEK1 are colored in red. (**C**) Beeswarm plots showing the effect of three CHEK1-specific inhibitors in U-CLL and M-CLL samples. Association testing was performed using Student's t-test test (two-sided, with equal variance) and Benjamini-Hochberg correction was applied. (**D**) Differentially expressed hallmark gene sets in CLL cells after BCR pathway stimulation through anti-IgM treatment (ArrayExpress ID: E-GEOD-39411). Highlighted upregulated hallmark gene sets are associated with DNA-damage response and cell cycle checkpoint.

https://doi.org/10.1371/journal.pcbi.1010438.g004

quantities of interest. The regularization step provides sparsity—many of the estimated effects are zero—which can facilitate interpretation and reduce noise, and is consistent with the notion that not every drug-protein binding event will contribute to the observed phenotype of the drug. Our approach stops short of more realistic biophysical modeling of the drug-protein binding kinetics: this would be an interesting direction for future research.

We validated our approach by showing that it could accurately recapitulate known protein dependencies of different cancer entities as well as of their molecular subtypes, both in data from cancer cell line compendia and from primary patient samples. We then made a novel finding, the high dependence of CLL with unmutated IGHV status (U-CLL) on the Checkpoint kinase CHEK1. This protein is an important component of the cellular DNA-damage and cell cycle checkpoint response and has been targeted in cancer therapeutics, especially for MYC-expressing cancers [44,45]. The critical role of CHEK1 in rapidly proliferating, developing, and BCR-activated normal B-cells was previously noted [41,42]. Consistent with our results, the drug screen by Dietrich *et al*. found high sensitivity of U-CLL to two CHEK1 inhibitors, AZD7762 and PF47736[16]. However, they hypothesized that this phenomenon was due to an off-target effect of the two compounds. Both of them have binding affinity towards kinases in the B-cell receptor (BCR) pathway, which is known to be more active in U-CLLs than M-CLLs. However, our results, in combination with these previous observations, suggest an on-target effect of CHEK1 inhibitors behind their higher efficacy in U-CLLs. We propose a model where U-CLL tumor cells, which often exhibit increased tonic or antigen triggered BCR signaling, experience more stress-related signals due to higher metabolic activity or increased replication stress. This might prime cancer cells towards a highly activated CHEK1-mediated cell cycle checkpoint control (which remains active in *in vitro* culture). As a result, U-CLL cells are more vulnerable to CHEK1 inhibition, driving these cells towards apoptosis. Our results suggest targeted CHEK1 inhibition as a potential therapeutic strategy for U-CLL.

The inference accuracy of our method relies on the size and quality of both the drug-protein profiling matrix and of the *ex vivo* tumor-drug response matrix. Generally, any increase in the number of drugs, tumor samples or proteins is beneficial. The data preprocessing steps, such as the transformation of drug-protein binding affinity matrix and the summarization of the drug dose-response values, may also impact the final inference results. In addition, collinearities in the drug-protein profiling matrix, which arise if several proteins have similar binding affinity profiles to all tested drugs, can pose challenges to the model fitting, as such proteins are indistinguishable in terms of the data visible to the model. Our method adopts the cluster/proxy set method to deal with collinear features [46]. Such protein clusters are statistical constructs, there is no assumption that the proteins have related biological functions. Therefore, if our method selects such a protein cluster as an explanatory variable, we recommend the use of additional information, be it prior knowledge or additional follow-up experiments, to disentangle the cluster and identify the causal protein. For example, BTK and YES1 share similar drug binding profiles based on the kinobeads dataset from Klaeger et al., and this cluster was selected as an explanatory variable in our analysis of the EMBL2016 dataset. Without the cluster/proxy set method, while the results for other selected kinases are largely identical (S9 Fig), our method would more or less randomly select between either of these two kinases (S10 Fig). However, our prior understanding of CLL biology strongly suggests BTK as the binding partner that is causally responsible for the drug response phenotypes. More systematic and automated treatment of statistical questions of identifiability, uncertainty, and stability of the estimated protein dependence map would be of interest for future work.

While our present analyses were limited to kinases and kinase inhibitors, we expect that the approach would be analogously applicable should protein-drug binding matrices for other protein classes become available.

In conclusion, we present a method to infer protein dependence maps from drug sensitivity data. Among the possible uses of this method, we envisage both, the discovery of novel disease stratifications by their characteristic dependencies on specific proteins, and improved understanding of existing disease stratifications in terms of differential protein dependencies. Nevertheless, such stratifications gain potential therapeutic value by being directly associated with druggable targets.

## Materials and methods

### Ethics statement

The use of patient tumor samples in this study was approved by the Ethics Committee Heidelberg (University of Heidelberg, Germany; approval S-206/2011 and S-356/2013). Patients who donated tumor material provided written informed consent prior to study.

### Processing of the drug response assay datasets

**The Genomics of Drug Sensitivity in Cancer (GDSC) project data.** The GDSC1 dataset of drug sensitivities was downloaded from the Genomics of Drug Sensitivity in Cancer Project (https://www.cancerrxgene.org) [13], separately for different cancers: diffuse large B-cell lymphoma (DLBCL, n = 30), acute lymphocytic leukemia (ALL, n = 25), acute myeloid leukemia ((L)AML, n = 24) and breast carcinoma (BRCA, n = 47). As a drug sensitivity measure, the dataset includes the area under the curve (AUC), defined as the area under the sigmoid-fit dose-response curve (five-point or nine-point) [47]. AUC values were converted into *z*-scores, by computing mean and standard deviation for each cell line, across the drugs. These *z*-scores were used as the drug sensitivity matrix Y, including sensitivity data for 66 drugs. Cell lines with more than 24 (≙ 35%) missing values were removed, resulting in a total of 126 cell lines. The missForest imputation method from the missForest package (version 1.4) [48] was used to impute missing values in the drug sensitivity matrix for the remaining cell lines.

The cancer cell lines were annotated with their cancer types and mutation status obtained from the GDSC project page. Additionally, HER2 status of breast cancer cell lines were annotated based on information from prior publications [49,50].

**BeatAML data.** The BeatAML dataset [12] contains *ex vivo* drug sensitivity data of 528 tumor specimens collected from AML patients. The data were downloaded from the National Cancer Institute's CTD2 Network (BeatAML_Waves1_2). AUC from a seven-point dose-response curve was used as described by Tyner *et al*. 2018 and the *z*-scores of the AUC-values for each cell line across all drugs were used as the response value for generating the drug sensitivity matrix *Y*, including sensitivity data for 61 drugs. Samples with more than 15 (≙ 24%) missing values were removed, resulting in a total number of 421 samples. Cut-off selection and missing value imputation were done as described for the GDSC1 dataset.

***Ex vivo* drug response assay on primary leukemic tumor samples.** We performed *ex vivo* drug response profiling on primary tumor samples from CLL (n = 117), mantle cell lymphoma (MCL, n = 7), or T-cell prolymphocytic leukemia (T-PLL, n = 7) patients. We termed this drug screen dataset as the EMBL2016 dataset. The samples were annotated with available genomic patient metadata. To obtain a single drug by tumor samples matrix, relative cell viabilities (compared to treatment with DMSO control) were averaged from 9 drug concentrations. Subsequently, the z-score of these values for each patient across all drugs was used as the response value for generating drug sensitivity matrix *Y*, including sensitivity data for 85 drugs. A list of compounds used in each drug screen dataset included in our study is provided in S7 Table.

## Processing of the kinase inhibitor profiling data

The kinase inhibitor profiling data were obtained from the Supplementary S2 Table of Klaeger *et al*. [14]. The data were subset for drugs overlapping with those used in either of the drug response assays described above by matching the drug names. The drug synonym annotation was considered and naming variations were manually curated. The data were arranged in a matrix $X$ of dissociation constants ($K_d$) for each drug-protein pair. The $K_d$ values were transformed with a sigmoid function into the range [0, 1]. We used $f(x) = (\arctan((-\log_{10}(x) + 2) * 3) + \pi/2) / \pi$, with x being the raw $K_d$ values. Values of $f(x)$ close to 1 correspond to evidence of drug-protein binding, values close to 0 to absence of such evidence. The graph of the function $f$ and the distribution of values before and after the transformation are shown in S11 Fig.

As some kinases had similar ligand profiles, and since this drug-target matrix was intended to be the design matrix in a multivariate regression model, highly similar kinases were "collapsed" in order to achieve model identifiability and parameter estimate stability. To this end, we computed the cosine similarity between each pair of kinases, applied hierarchical clustering, and identified subtrees in the cluster dendrogram in which all kinases had a cosine similarity $> 0.8$. For each such cluster, the kinase with the largest sum of transformed $K_d$ values within the cluster was used as a cluster name, while the other kinases were also recorded and retained for case-by-case follow-up. The collapsed kinase sets are provided in S2 Fig and S1–S3 Tables. Note that this clustering step is purely technical and the proteins or kinases in the same cluster can have different biological functions. Therefore, if the selected proteins are from clusters, the users of our method are recommended to examine the other cluster members and interpret the results with the help of prior knowledge.

For the analysis of the BeatAML and EMBL2016 datasets, for which the gene expression data are also available, the drug-protein profiling matrix was additionally filtered by removing kinases that were not or only weakly expressed in the tumor samples, as follows. In the gene-level count matrix of the RNA sequencing data of the tumor samples, the 80% quantile of its counts across samples was computed for each gene, and a gene was considered not or only weakly expressed if this value was less than 10. Thus, 20 of the 377 kinases were removed from the downstream analysis for the EMBL2016 dataset, and 14 of the 380 kinases in the BeatAML analysis. The final dimensions of the input matrices X were: GDSC: 66 drugs × 118 proteins; BeatAML: 61 drugs × 112 proteins; EMBL2016: 85 drugs × 131 proteins (Fig 1A).

## Regularized multivariate regression model

Regularized multivariate regression was used to obtain the protein dependence coefficients. We fit the model

$$Y = X * \beta + \epsilon, \tag{1}$$

where $X$ is the processed drug-protein binding matrix (drugs × proteins), $Y$ is the processed drug sensitivity matrix (drugs × samples), β is the unknown protein dependence matrix (samples × proteins), and $\epsilon$ is a matrix of residuals (drugs x samples). Given the data for $X$ and $Y$, we computed the model fit for β by minimizing $\epsilon$ using a multi-response Gaussian linear model (family = "mgaussian") with L1-penalty (lasso) on β using the *glmnet* package (*version 3.0–2*) [51] with the mixing parameter $\alpha = 1$, and λ shared among samples.

We repeated the following 100 times and computed the element-wise medians of the 100 β-matrices to obtain the final estimate $\hat{\beta}$: determine an optimal penalty parameter λ that minimizes the cross-validated mean squared error using the function *cv.glmnet* from the glmnet package with three-fold cross-validation and fit the model with the optimal choice of λ. As a quality measure, the coefficient of determination $R^2$ from the final ß estimate was used.

## Principal Component Analysis (PCA) visualization

PCA analysis of $\hat{\beta}$ was performed using the R function *prcomp* from the *stats* package, which centers and scales the columns (corresponding to samples) of the matrix for equal variance. Visualization graphics were made using the *factoextra* package (*version 1.0.6*) [52].

## *k*-means clustering and Rand index calculation

*k*-means clustering analysis was performed on $\hat{\beta}$ using the *eclust* function from the *factoextra* package, which centers and scales the columns of the matrix. Class information (five cancer types in the case of the GDSC data, IGHV status for the EMBL2016 CLL data) was withheld from the clustering method, but used to set the number of clusters *k*. 50 different random starting configurations (*nstart*) were attempted within the *eclust* function and the best configuration was chosen. As a measure of concordance of the resulting clusters and the 'ground truth' classes, the adjusted Rand index [35] was calculated using the *cluster.stats* function from the *fpc* package (*version 2.2–9*) [53]. *k*-means clustering of the drug sensitivity matrix was performed analogously.

## Differential gene expression analysis

To identify gene expression signatures related to IGHV status, an RNA sequencing dataset from the European Genome-phenome Archive (EGA; http://www.ebi.ac.uk/ega/, accession: EGAS00001000374) was analyzed [54]. Differential expression testing was performed using the *DESeq2* package (*version 1.28.1*) [55]. Genes with overall low counts (sum across samples $< 10$) were excluded from the analysis. The *camera* method was used to test for differential expression of hallmark gene sets from the MSigDB database v7.2 [56,57], using the implementation in the eponymous function in the *limma* package (*version 3.44.3*) [58].

To identify gene expression signatures related to B-cell stimulation, we analyzed gene expression microarray dataset of anti-IgM triggering in primary CLL samples (E-GEOD-39411) [59]. The *limma* package was used to perform variance stabilizing normalization and differential expression testing, and the *camera* method was used to test for differential expression of Hallmark gene sets as above.

## Other statistical methods

Association tests were performed using Student's *t*-test (two-sided, with equal variance) and ANOVA test. The Benjamini-Hochberg method was applied to the *P* values to account for multiple testing.

# Supporting information

**S1 Fig. A network plot showing the drug-kinase binding landscape measured by the kinobeads assay.** Only the high confidence drug-kinase pairs are included. Drugs are shown as blue circles and kinases are shown as red squares. No clear structure can be observed directly from the drug-kinase network.
(PDF)

**S2 Fig. Network plots showing the clusters of kinases that have similar drug binding profiles (cosine similarity $> 0.8$).** For each drug screen dataset, only the kinases that were selected by our protein dependence inference algorithm and share similar drug binding profiles with other kinases are shown here. The full list is provided in S1–S3 Tables.
(PDF)

**S3 Fig. Heatmap visualization of the protein dependence values for selected kinases, inferred from the GDSC drug screen dataset.** Red indicates high dependence value and blue indicates low dependence value. In the column annotations, cell lines with certain genomic variations are colored by black. Columns are ordered by hierarchical clustering based on Euclidean distance.
(PDF)

**S4 Fig. Boxplots showing the significant associations between the genomic variations and the protein dependence values inferred from the BeatAML drug screen dataset.** *P* values are from two-sided Student's t-test with equal variance.
(PDF)

**S5 Fig. Heatmap visualization of the protein dependence values for selected kinases inferred from the EMBL2016 drug screen dataset.** Red indicates high dependence value and blue indicates low dependence values. The column annotations of the heatmap indicate the disease types, genomic variations and epigenetic subtypes of the primary blood cancer samples. Columns are ordered by hierarchical clustering based on Euclidean distance.
(PDF)

**S6 Fig. Association between the inferred dependence on BTK and the disease types, based on the EMBL2016 drug screen dataset.** *P* value was calculated by one-way ANOVA test. To avoid potential bias due to the unbalanced sample size or unequal variance, nonparametric Kruskal-Wallis test was also performed and the resulting *P* value is 0.00019.
(PDF)

**S7 Fig. Enrichment analysis for genes differentially expressed between IGHV mutated and unmutated CLL samples in the ICGC-CLL cohort (EGA accession ID: EGAS00001000374).** "Up" (colored by red") indicates the gene sets enriched for the genes up-regulated in U-CLL samples compared to M-CLL samples. The enrichment analysis was performed by using CAMERA against the Cancer Hallmark gene sets from the Molecular Signature Database (MSigDB). Only the gene sets passed 5% FDR are shown. The gene sets that have been shown to be related to CHEK1 functions are colored in red.
(PDF)

**S8 Fig. Differential expression of three downstream genes regulated by CHEK1 pathway in IGHV unmutated and mutated CLL samples from the ICGC-CLL cohort.** *P* values were calculated by DESeq2.
(PDF)

**S9 Fig. Comparing the protein dependence values inferred from the models with (x-axis) and without (y-axis) clustering and combining similar targets (EMBL2016 dataset).**
(PDF)

**S10 Fig. Comparing the inferred dependence values of YES1 and BTK from the models with and without clustering and combining similar targets (EMBL2016 dataset).**
(PDF)

**S11 Fig. Plot of the transformation function of Kd values from the kinobeads assay.**
(PDF)

**S1 Table. The full list of clusters of kinases that have similar drug binding profiles for the GDSC dataset.**
(XLSX)

**S2 Table. The full list of clusters of kinases that have similar drug binding profiles for the BeatAML dataset.**
(XLSX)

**S3 Table. The full list of clusters of kinases that have similar drug binding profiles for the EMBL2016 dataset.**
(XLSX)

**S4 Table. Inferred protein dependence values for the GDSC dataset.**
(XLSX)

**S5 Table. Inferred protein dependence values for the BeatAML dataset.**
(XLSX)

**S6 Table. Inferred protein dependence values for the EMBL2016 dataset.**
(XLSX)

**S7 Table. Compounds used in the drug screen datasets included in our study.**
(XLSX)

## Author Contributions

**Conceptualization:** Junyan Lu, Wolfgang Huber.

**Data curation:** Junyan Lu, Jarno Kivioja, Kerstin Putzker.

**Formal analysis:** Alina Batzilla, Junyan Lu.

**Funding acquisition:** Thorsten Zenz, Wolfgang Huber.

**Investigation:** Jarno Kivioja, Kerstin Putzker, Joe Lewis, Thorsten Zenz.

**Resources:** Alina Batzilla, Junyan Lu.

**Supervision:** Wolfgang Huber.

**Writing – original draft:** Alina Batzilla, Junyan Lu.

**Writing – review & editing:** Junyan Lu, Jarno Kivioja, Thorsten Zenz, Wolfgang Huber.

## References

1. McFarland JM, Ho ZV, Kugener G, Dempster JM, Montgomery PG, Bryan JG, et al. Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. Nature Communications. 2018; 9: 4610. https://doi.org/10.1038/s41467-018-06916-5 PMID: 30389920
2. Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. Nature. 2019; 568: 511–516. https://doi.org/10.1038/s41586-019-1103-9 PMID: 30971826
3. Cowley GS, Weir BA, Vazquez F, Tamayo P, Scott JA, Rusin S, et al. Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. Scientific data. 2014; 1: 140035–140035. https://doi.org/10.1038/sdata.2014.35 PMID: 25984343
4. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a Cancer Dependency Map. Cell. 2017; 170: 564–576.e16. https://doi.org/10.1016/j.cell.2017.06.010 PMID: 28753430
5. Laustsen A, Bak RO. Electroporation-Based CRISPR/Cas9 Gene Editing Using Cas9 Protein and Chemically Modified sgRNAs. In: Luo Y, editor. CRISPR Gene Editing: Methods and Protocols. New York, NY: Springer New York; 2019. pp. 127–134. https://doi.org/10.1007/978-1-4939-9170-9_9

6. Mandal PK, Ferreira LMR, Collins R, Meissner TB, Boutwell CL, Friesen M, et al. Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. Cell stem cell. 2014; 15: 643–652. https://doi.org/10.1016/j.stem.2014.10.004 PMID: 25517468
7. Gonçalves E, Segura-Cabrera A, Pacini C, Picco G, Behan FM, Jaaks P, et al. Drug mechanism-of-action discovery through the integration of pharmacological and CRISPR screens. Molecular Systems Biology. 2020; 16. https://doi.org/10.15252/msb.20199405 PMID: 32627965
8. Weiss WA, Taylor SS, Shokat KM. Recognizing and exploiting differences between RNAi and small-molecule inhibitors. Nature chemical biology. 2007; 3: 739–744. https://doi.org/10.1038/nchembio1207-739 PMID: 18007642
9. Morgens DW, Wainberg M, Boyle EA, Ursu O, Araya CL, Tsui CK, et al. Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. Nat Commun. 2017; 8: 15178. https://doi.org/10.1038/ncomms15178 PMID: 28474669
10. Fellmann C, Gowen BG, Lin P-C, Doudna JA, Corn JE. Cornerstones of CRISPR–Cas in drug discovery and therapy. Nature Reviews Drug Discovery. 2017; 16: 89–100. https://doi.org/10.1038/nrd.2016.238 PMID: 28008168
11. Munoz DM, Cassiani PJ, Li L, Billy E, Korn JM, Jones MD, et al. CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but Generate False-Positive Hits for Highly Amplified Genomic Regions. Cancer Discovery. 2016; 6: 900. https://doi.org/10.1158/2159-8290.CD-16-0178 PMID: 27260157
12. Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, et al. Functional genomic landscape of acute myeloid leukaemia. Nature. 2018; 562: 526–531. https://doi.org/10.1038/s41586-018-0623-z PMID: 30333627
13. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Research. 2012; 41: D955–D961. https://doi.org/10.1093/nar/gks1111 PMID: 23180760
14. Klaeger S, Heinzlmeir S, Wilhelm M, Polzer H, Vick B, Koenig P-A, et al. The target landscape of clinical kinase drugs. Science (New York, NY). 2017; 358: eaan4368. https://doi.org/10.1126/science.aan4368 PMID: 29191878
15. Lu J, Cannizzaro E, Meier-Abt F, Scheinost S, Bruch P-M, Giles HAR, et al. Multi-omics reveals clinically relevant proliferative drive associated with mTOR-MYC-OXPHOS activity in chronic lymphocytic leukemia. Nat Cancer. 2021; 2: 853–864. https://doi.org/10.1038/s43018-021-00216-6 PMID: 34423310
16. Dietrich S, Oleś M, Lu J, Sellner L, Anders S, Velten B, et al. Drug-perturbation-based stratification of blood cancer. The Journal of Clinical Investigation. 2017; 128: 427–445. https://doi.org/10.1172/JCI93801 PMID: 29227286
17. Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A Landscape of Pharmacogenomic Interactions in Cancer. Cell. 2016; 166: 740–754. https://doi.org/10.1016/j.cell.2016.06.017 PMID: 27397505
18. Wang Y, Ma H. Protein kinase profiling assays: a technology review. Drug Discov Today Technol. 2015; 18: 1–8. https://doi.org/10.1016/j.ddtec.2015.10.007 PMID: 26723886
19. Goldstein DM, Gray NS, Zarrinkar PP. High-throughput kinase profiling as a platform for drug discovery. Nat Rev Drug Discov. 2008; 7: 391–397. https://doi.org/10.1038/nrd2541 PMID: 18404149
20. Attwood MM, Fabbro D, Sokolov AV, Knapp S, Schiöth HB. Trends in kinase drug discovery: targets, indications and inhibitor design. Nat Rev Drug Discov. 2021; 20: 839–861. https://doi.org/10.1038/s41573-021-00252-y PMID: 34354255
21. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. 1987; 235: 177–182. https://doi.org/10.1126/science.3798106 PMID: 3798106
22. Ahn S, Woo JW, Lee K, Park SY. HER2 status in breast cancer: changes in guidelines and complicating factors for interpretation. J Pathol Transl Med. 2020; 54: 34–44. https://doi.org/10.4132/jptm.2019.11.03 PMID: 31693827
23. Annesley CE, Brown P. The Biology and Targeting of FLT3 in Pediatric Leukemia. Frontiers in Oncology. 2014; 4: 263. https://doi.org/10.3389/fonc.2014.00263 PMID: 25295230
24. Griffith M, Griffith OL, Krysiak K, Skidmore ZL, Christopher MJ, Klco JM, et al. Comprehensive genomic analysis reveals FLT3 activation and a therapeutic strategy for a patient with relapsed adult B-lymphoblastic leukemia. Exp Hematol. 2016; 44: 603–613. https://doi.org/10.1016/j.exphem.2016.04.011 PMID: 27181063

25. Armstrong SA, Mabon ME, Silverman LB, Li A, Gribben JG, Fox EA, et al. FLT3 mutations in childhood acute lymphoblastic leukemia. Blood. 2004; 103: 3544–3546. https://doi.org/10.1182/blood-2003-07-2441 PMID: 14670924

26. Raaijmakers MIG, Widmer DS, Narechania A, Eichhoff O, Freiberger SN, Wenzina J, et al. Co-existence of BRAF and NRAS driver mutations in the same melanoma cells results in heterogeneity of targeted therapy resistance. Oncotarget. 2016; 7: 77163–77174. https://doi.org/10.18632/oncotarget.12848 PMID: 27791198

27. McMahon CM, Ferng T, Canaani J, Wang ES, Morrissette JJD, Eastburn DJ, et al. Clonal Selection with RAS Pathway Activation Mediates Secondary Clinical Resistance to Selective FLT3 Inhibition in Acute Myeloid Leukemia. Cancer Discov. 2019; 9: 1050–1063. https://doi.org/10.1158/2159-8290.CD-18-1453 PMID: 31088841

28. Jhiang SM. The RET proto-oncogene in human cancers. Oncogene. 2000; 19: 5590–5597. https://doi.org/10.1038/sj.onc.1203857 PMID: 11114739

29. Köthe S, Müller JP, Böhmer S-A, Tschongov T, Fricke M, Koch S, et al. Features of Ras activation by a mislocalized oncogenic tyrosine kinase: FLT3 ITD signals through K-Ras at the plasma membrane of acute myeloid leukemia cells. J Cell Sci. 2013; 126: 4746–4755. https://doi.org/10.1242/jcs.131789 PMID: 23943874

30. Daver N, Schlenk RF, Russell NH, Levis MJ. Targeting FLT3 mutations in AML: review of current knowledge and evidence. Leukemia. 2019; 33: 299–312. https://doi.org/10.1038/s41375-018-0357-9 PMID: 30651634

31. Marhäll A, Kazi JU, Rönnstrand L. The Src family kinase LCK cooperates with oncogenic FLT3/ITD in cellular transformation. Scientific reports. 2017; 7: 13734. https://doi.org/10.1038/s41598-017-14033-4 PMID: 29062038

32. Zenz T, Mertens D, Küppers R, Döhner H, Stilgenbauer S. From pathogenesis to treatment of chronic lymphocytic leukaemia. Nat Rev Cancer. 2010; 10: 37–50. https://doi.org/10.1038/nrc2764 PMID: 19956173

33. Hendriks RW, Yuvaraj S, Kil LP. Targeting Bruton's tyrosine kinase in B cell malignancies. Nat Rev Cancer. 2014; 14: 219–232. https://doi.org/10.1038/nrc3702 PMID: 24658273

34. Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, et al. Ig V Gene Mutation Status and CD38 Expression As Novel Prognostic Indicators in Chronic Lymphocytic Leukemia: Presented in part at the 40th Annual Meeting of The American Society of Hematology, held in Miami Beach, FL, December 4–8, 1998. Blood. 1999; 94: 1840–1847. https://doi.org/10.1182/blood.V94.6.1840

35. Rand WM. Objective Criteria for the Evaluation of Clustering Methods. Journal of the American Statistical Association. 1971; 66: 846–850. https://doi.org/10.1080/01621459.1971.10482356

36. Guo A, Lu P, Galanina N, Nabhan C, Smith SM, Coleman M, et al. Heightened BTK-dependent cell proliferation in unmutated chronic lymphocytic leukemia confers increased sensitivity to ibrutinib. Oncotarget. 2016; 7: 4598–4610. https://doi.org/10.18632/oncotarget.6727 PMID: 26717038

37. Nagel S, Leich E, Quentmeier H, Meyer C, Kaufmann M, Zaborski M, et al. Amplification at 11q23 targets protein kinase SIK2 in diffuse large B-cell lymphoma. Leuk Lymphoma. 2010; 51: 881–891. https://doi.org/10.3109/10428191003699878 PMID: 20367563

38. Tarumoto Y, Lu B, Somerville TDD, Huang Y-H, Milazzo JP, Wu XS, et al. LKB1, Salt-Inducible Kinases, and MEF2C Are Linked Dependencies in Acute Myeloid Leukemia. Mol Cell. 2018; 69: 1017–1027.e6. https://doi.org/10.1016/j.molcel.2018.02.011 PMID: 29526696

39. Döhner H, Stilgenbauer S, Benner A, Leupolt E, Kröber A, Bullinger L, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. N Engl J Med. 2000; 343: 1910–1916. https://doi.org/10.1056/NEJM200012283432602 PMID: 11136261

40. Verlinden L, Vanden Bempt I, Eelen G, Drijkoningen M, Verlinden I, Marchal K, et al. The E2F-regulated gene Chk1 is highly expressed in triple-negative estrogen receptor /progesterone receptor /HER-2 breast carcinomas. Cancer Res. 2007; 67: 6574–6581. https://doi.org/10.1158/0008-5472.CAN-06-3545 PMID: 17638866

41. Schoeler K, Jakic B, Heppke J, Soratroi C, Aufschnaiter A, Hermann-Kleiter N, et al. CHK1 dosage in germinal center B cells controls humoral immunity. Cell Death & Differentiation. 2019; 26: 2551–2567. https://doi.org/10.1038/s41418-019-0318-5 PMID: 30894677

42. Schuler F, Weiss JG, Lindner SE, Lohmüller M, Herzog S, Spiegl SF, et al. Checkpoint kinase 1 is essential for normal B cell development and lymphomagenesis. Nature Communications. 2017; 8: 1697. https://doi.org/10.1038/s41467-017-01850-4 PMID: 29167438

43. Wang T, Gautam P, Rousu J, Aittokallio T. Systematic mapping of cancer cell target dependencies using high-throughput drug screening in triple-negative breast cancer. Computational and Structural Biotechnology Journal. 2020; 18: 3819–3832. https://doi.org/10.1016/j.csbj.2020.11.001 PMID: 33335681

44. Höglund A, Nilsson LM, Muralidharan SV, Hasvold LA, Merta P, Rudelius M, et al. Therapeutic Implications for the Induced Levels of Chk1 in Myc-Expressing Cancer Cells. Clinical Cancer Research. 2011; 17: 7067–7079. https://doi.org/10.1158/1078-0432.CCR-11-1198 PMID: 21933891

45. Ferrao PT, Bukczynska EP, Johnstone RW, McArthur GA. Efficacy of CHK inhibitors as single agents in MYC-driven lymphoma cells. Oncogene. 2012; 31: 1661–1672. https://doi.org/10.1038/onc.2011.358 PMID: 21841818

46. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography. 2013; 36: 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

47. Vis DJ, Bombardelli L, Lightfoot H, Iorio F, Garnett MJ, Wessels LF. Multilevel models improve precision and speed of IC50 estimates. Pharmacogenomics. 2016; 17: 691–700. https://doi.org/10.2217/pgs.16.15 PMID: 27180993

48. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics. 2011; 28: 112–118. https://doi.org/10.1093/bioinformatics/btr597 PMID: 22039212

49. Dai X, Cheng H, Bai Z, Li J. Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. Journal of Cancer. 2017; 8: 3131–3141. https://doi.org/10.7150/jca.18457 PMID: 29158785

50. Jernström S, Hongisto V, Leivonen S-K, Due EU, Tadele DS, Edgren H, et al. Drug-screening and genomic analyses of HER2-positive breast cancer cell lines reveal predictors for treatment response. Breast cancer (Dove Medical Press). 2017; 9: 185–198. https://doi.org/10.2147/BCTT.S115600 PMID: 28356768

51. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software. 2010; 33: 1–22. https://doi.org/10.18637/jss.v033.i01 PMID: 20808728

52. Kassambara A, Mundt F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package version 1.0.6. -. 2019;

53. Hennig C. Package 'fpc.' 2015.

54. Ferreira PG, Jares P, Rico D, Gómez-López G, Martínez-Trillos A, Villamor N, et al. Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. Genome research. 2014; 24: 212–26. https://doi.org/10.1101/gr.152132.112 PMID: 24265505

55. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014; 15: 550–550. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

56. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005; 102: 15545. https://doi.org/10.1073/pnas.0506580102 PMID: 16199517

57. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell systems. 2015; 1: 417–425. https://doi.org/10.1016/j.cels.2015.12.004 PMID: 26771021

58. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015; 43: e47–e47. https://doi.org/10.1093/nar/gkv007 PMID: 25605792

59. Vallat L, Kemper CA, Jung N, Maumy-Bertrand M, Bertrand F, Meyer N, et al. Reverse-engineering the genetic circuitry of a cancer cell with predicted intervention in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2013; 110: 459–464. https://doi.org/10.1073/pnas.1211130110 PMID: 23267079