

# Multi-omics molecular phenotyping of childhood acute lymphoblastic leukemia cell lines reveals novel drug vulnerabilities

**Matthias Stahl**

Karolinska Institutet <https://orcid.org/0000-0002-0176-9386>

**Luay Aswad**

Karolinska Institute

**Isabelle Leo**

Karolinska Institute <https://orcid.org/0000-0002-7627-6690>

**Elena Kunold**

Karolinska Institute

**Frederik Post**

Karolinska Institute <https://orcid.org/0000-0003-0376-1326>

**Tom Erkers**

Karolinska Institute

**Nona Struyf**

Karolinska Institute

**Georgios Mermelekas**

Karolinska Institute

**Rubin Joshi**

Karolinska Institute

**Eva Gracia-Villacampa**

KTH Royal Institute of Technology

**Päivi Östling**

Karolinska Institute

**Olli Kalloniemi**

University of Helsinki

**Katja Pokrovskaja Tamm**

Department of Oncology-Pathology, Karolinska Institutet <https://orcid.org/0000-0001-6359-1256>

**Ioannis Siavelis**

Karolinska Institutet

**Janne Lehtiö**

Karolinska Institutet <https://orcid.org/0000-0002-8100-9562>

**Mattias Vesterlund**

Karolinska Institutet <https://orcid.org/0000-0001-9471-6592>

**Rozbeh Jafari** (✉ [rozbeh.jafari@ki.se](mailto:rozbeh.jafari@ki.se))

Karolinska Institute <https://orcid.org/0000-0002-3396-4709>

---

## Article

**Keywords:** Acute Lymphoblastic Leukemia, drug sensitivity and resistance testing, cell lines, proteomics, RNA-sequencing, tandem mass spectrometry.

**Posted Date:** April 29th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-443839/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# **Multi-omics molecular phenotyping of childhood acute lymphoblastic leukemia cell lines reveals novel drug vulnerabilities.**

## **Authors and affiliations**

Matthias Stahl<sup>1§</sup>, Luay Aswad<sup>1§</sup>, Isabelle Leo<sup>1</sup>, Elena Kunold<sup>1</sup>, Frederik Post<sup>1,5</sup>, Tom Erkers<sup>2</sup>, Nona Struyf<sup>2</sup>, Georgios Mermelekas<sup>1</sup>, Rubin Joshi<sup>1</sup>, Eva Gracia-Villacampa<sup>3</sup>, Päivi Östling<sup>2</sup>, Olli P. Kallioniemi<sup>2</sup>, Katja Pokrovskaja Tamm<sup>4</sup>, Ioannis Siavelis<sup>1</sup>, Janne Lehtiö<sup>1</sup>, Mattias Vesterlund<sup>1</sup>, Rozbeh Jafari<sup>1\*</sup>

<sup>1</sup> Clinical Proteomics Mass Spectrometry, Department of Oncology-Pathology, Karolinska Institutet, Science for Life Laboratory, Tomtebodavägen 23A, 171 65 Solna, Sweden

<sup>2</sup> Molecular Precision Medicine, Department of Oncology-Pathology, Karolinska Institutet, Science for Life Laboratory, Tomtebodavägen 23A, 171 65 Solna, Sweden

<sup>3</sup> Division of Gene Technology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH, Science for Life Laboratory, Tomtebodavägen 23A, 171 65 Solna, Sweden.

<sup>4</sup> Department of Oncology-Pathology, Karolinska Institutet, J6:20 BioClinicum, Akademiska stråket 1, 171 64 Solna, Sweden

<sup>5</sup> Institute of Plant Biology and Biotechnology, University of Muenster, Schlossplatz 7, 48149 Muenster, Germany

§ These authors contributed equally to this paper.

\* Corresponding Author and lead contact: Correspondence should be addressed to R.J.

Corresponding author and lead contact email: rozbeh.jafari@ki.se

## **Keywords**

Acute Lymphoblastic Leukemia, drug sensitivity and resistance testing, cell lines, proteomics, RNA-sequencing, tandem mass spectrometry.

## Acronyms

ALL: Acute Lymphoblastic Leukemia

ph: Philadelphia chromosome

DSRT: Drug Sensitivity and Resistance Testing

DSS: Drug Sensitivity Score

sDSS: Selective Drug Sensitivity Score

GSEA: Gene Set Enrichment Analysis

GO: Gene Ontology

BCP-ALL: B-cell Precursor Acute Lymphoblastic Leukemia

COG: Children's Oncology Group

DE: Differential expression

GC: Glucocorticoid

DEqMS: Differential expression quantitative mass spectrometry

LC-MS: Liquid chromatography mass spectrometry

RNAseq: Ribonucleic acid sequencing

MS: Mass spectrometry

CAR: Chimeric antigen receptor

PCA: Principal component analysis

BM: Bone marrow

PE: Peripheral effusion

PB: Peripheral blood

DAG: Diacylglycerol

HTS: High throughput sampler

PMA: Phorbol 12-myristate 13-acetate

TMT: Tandem mass tag

CLC: Consensus leukemic clusters

FDR: False discovery rate

MoA: Mechanism of action

HSCT: Hematopoietic stem cell transplantation

NES: Normalized enrichment score

## SUMMARY

Acute lymphoblastic leukemia (ALL) is the most common childhood cancer. Although standard-of-care chemotherapeutics are sufficient for most ALL cases, there are subsets of patients with poor response who relapse in disease. The biology underlying differences between subtypes and their response to therapy has only partially been explained by genetic and transcriptomic profiling. To characterize ALL subtypes and identify novel pharmacologic vulnerabilities, we performed comprehensive multi-omic analyses of 49 widely-used childhood ALL cell lines, using proteomics, transcriptomics, and pharmacoproteomic characterization. This enabled us to characterize the functional impact of genetic fusions and cellular differentiation states. The proteomics data revealed differences in spliceosome and p53 levels not evident in the transcriptomics data and with improved correlation to drug sensitivity. Focusing on BCP-ALL cell lines, we connected the genotype, molecular phenotype, and functional phenotype with drug response data on 528 oncology drugs. Here, we identified the DAG-analog Bryostatin-1 as a novel therapeutic candidate in the MEF2D-HNRNPUL1 fusion high-risk subtype, for which this drug activated pro-apoptotic ERK signaling associated with molecular mediators of pre-B cell negative selection. Our data also forms an interactive online resource with navigable proteomics, transcriptomics, and drug sensitivity profiles at <https://lehtio-lab.se/forall/>.

## INTRODUCTION

Acute lymphoblastic leukemia (ALL) accounts for ~30% of all cancers in children, making it the most common childhood cancer. Despite the clinical success of broad chemotherapy protocols and allogeneic hematopoietic stem cell transplantation (HSCT) <sup>1, 2</sup>, a considerable number of patients (15-20%) continue to experience poor survival outcomes because they are non-responsive to standard of care therapeutics and likely to relapse in disease <sup>3, 4</sup>. Additionally, approximately 80% of survivors of childhood ALL will experience a post-treatment life-threatening medical event by age 45, an effect hypothesized to result from the severity and duration of ALL treatment protocols <sup>5</sup>. Targeted therapy protocols have demonstrated potential to reduce the likelihood of post-treatment health complications, improve outcomes and address resistance to chemotherapy. These targeted approaches have been successfully implemented clinically in protocols giving tyrosine kinase inhibitors to Philadelphia chromosome positive ALL patients (NCT03911128) <sup>6</sup>. Recently, targeted treatment strategies have expanded to include immunotherapeutic agents such as antibody and chimeric antigen receptor (CAR) -T-cell therapies <sup>7</sup>. However, their adoption remains limited to ALL subtypes expressing specific surface antigens, and challenges in CAR-T production and the clinical emergence of adverse events presently limits them from attaining widespread application <sup>8, 9</sup>. Consequently, there is still an immediate need for novel therapeutic modalities for treatment of high-risk patients and patients who relapse.

Therefore, bridging the gap in therapeutic options for ALL will likely require development of both pharmacologic and cellular therapies; and this process will depend on deeper characterization of both the biomarkers and biology of non-responsive subtypes. While the genetic landscape of childhood ALL has been extensively studied by implementation of next generation genomic, transcriptomic, and epigenetic sequencing tools <sup>10, 11</sup>, somatic mutations can only partially explain the underlying biology and phenotype, and approximately 25% of

childhood ALL patients lack a detectable driving mutation<sup>12</sup>. Although genetic characterizations have improved risk stratification and hope of new molecular targets for therapy, the primary challenge of identifying novel and effective treatments for ALL is achieving more reliable clinical stratification that can improve therapeutic response<sup>13</sup>. Therefore, new approaches are needed to identify gaps between genetic and phenotypic properties of the disease.

Considering that regulation of protein abundance by post-transcriptional and post-translational mechanisms is necessary to maintain cellular fitness, and that these mechanisms have diverse impacts on the protein levels for each cellular component<sup>14</sup>, the proteome represents an ideal framework for understanding cellular phenotypes. Recent studies have demonstrated poor correlation between the proteomic and the genetic phenotypes of cancers<sup>15, 16, 17</sup>, including childhood ALL<sup>18</sup>. Although studies of tumors and human-derived cell lines<sup>19</sup> have demonstrated the importance of proteomic characterization, our recent characterization of ETV6-RUNX1 and hyperdiploid ALL is the only previous study that has used these methods to understand the biology of ALL<sup>18</sup>.

MEF2D fusions represent ~3.6% of childhood ALL cases, and patients experience 5-year event-free survival of 71%, identifying this patient subset as high-risk of relapse or disease progression in current therapy protocols<sup>20</sup>. Through proteomics-guided analysis of childhood ALL cell lines, we have generated a comprehensive resource of biomarkers and drug sensitivities, and identified a potential new therapeutic vulnerability to target the MEF2D-HNRNPUL1 rearranged high-risk subgroup. Our dataset demonstrates that proteomics and transcriptomics identify different phenotypic landscapes in childhood ALL cell lines, and we used proteomics to identify biological and phenotypic characteristics that suggest promising novel insights into both on-target and potential off-target drug activities. We provide a user-

friendly web application for exploration of the proteomic, transcriptomic, and drug sensitivity data described in this study, available at <https://lehtio-lab.se/forall>.

## RESULTS

### In-depth multi-omics profiling of childhood ALL cell lines

We performed in-depth profiling by examining protein and RNA levels as well as drug sensitivities (**Figure 1A**) of 51 readily available cell lines representing 29 B-lineage and 22 T-lineage cell lines with various cytogenetic backgrounds (**Figure 1B and Supp. Table 1**). Mass spectrometry (MS) based protein and peptide quantification for the cell lines (see methods) and subsequent analyses identified and quantified 279,351 peptides. These were assigned to 13,704 proteins originating from 12,446 genes at a false discovery rate (FDR) of 1%, with median protein sequence coverage of 46% and an overlap of 9,100 proteins (gene symbol centric) across all the 51 cell lines (**Figure 1A**). Multiplexing was achieved by peptide labeling using tandem mass tags (TMT), and in several of the TMT sets ( $n = 8$ ), a technical replica of the SEM cell line was used for analysis of technical variability (**Supp. Fig. 1A**). Biological replicates for selected cell lines ( $n = 32$ ) were generated approximately one year apart, and replicates grouped together in our analysis (**Supp. Fig. 1B**). Transcriptomic profiling was performed using ribosomal depleted total RNA for all cell lines ( $n = 51$ ) as well as for several biological replicates ( $n = 15$ ) from matching samples (**Figure 1B and Supp. Table 3**) with a minimum and median sequencing depth of 33M and 47M paired end reads respectively. Drug sensitivity and resistance testing (DSRT) was performed for the BCP-ALL cell lines ( $n = 25$ ) using a panel of 528 approved oncological drugs as well as investigational drugs in various clinical phases (**Figure 1A and Supp. Table 4**).

Principal component analysis of protein level profiles separated four major lineage-linked groups (**Figure 1C**). Cell lines from the T-lineage and B-lineage were clearly separated in this analysis, and the B-lineage cell lines were further distinguished into BCP-ALL and B-ALL cell



lines. Two cell lines (CCRF-SB and COG-LL-356h) were Epstein-Barr virus (EBV)-transformed B-cell lines, which comprised a fourth phenotypic group. To determine differentially expressed (DE) proteins across cell lines in our panel we performed quantitative proteomics comparison using DEqMS<sup>21</sup> to compare the different cytogenetic groups. At first, we performed quantitative analysis between the B-cell and T-cell lineages excluding the EBV-transformed B-cell lines with subsequent gene set enrichment analysis (GSEA), which correctly indicated B-cell-receptor or T-cell-receptor signaling pathways for each respective lineage (**Supp. Fig. 1C, Supp. Table 5**).

Stratified by cytogenetic background, our dataset was confirmed to reflect enrichment of known markers and pathways. Fusion proteins known to act as leukemia-driving lesions or signatures were enriched in our proteomics dataset, including PBX1, ROR1 and WNT16 for the TCF3-PBX1 fusion<sup>22</sup>, and ABL1 for BCR-ABL1, NUP214-ABL1, and SPK-ABL1 fusions (**Figure 1D**). Other known markers associated with specific translocations could also be detected, including IGF2BP1 and PIK3C3 for ETV6-RUNX1<sup>23, 24</sup> and MEIS1 for KMT2A/mixed-lineage leukemia<sup>25</sup> (**Figure 1D, Supp. Fig. 1D**).

### **Correlation analysis of ALL proteome and transcriptome**

Cellular phenotypes are controlled by diverse mechanisms that occur at multiple points following transcription, and we sought to quantify the impact of these mechanisms by performing a systems-level analysis of matched proteome and transcriptome profiles (n=66). We performed Spearman pairwise correlation for all mRNA-protein pairs (n=8981), and the median correlation coefficient was 0.55 (**Figure 2A**). Our previous study on clinical ALL samples<sup>18</sup> and other studies on solid cancers demonstrated lower Spearman correlation coefficients for this metric<sup>15, 16, 26</sup>, however, similar results have been reported for cell line studies<sup>19, 27</sup>. This could be as a result of clonal derivation of cell lines, or owed to the significant technical advantages of working with cell lines compared to the complexity in sample

acquisition, preparation, and cellular fitness requirements of clinical samples. To evaluate the functional impact of proteins detected at levels that diverged from their respective mRNA quantification, a list ranked by Spearman correlation coefficient was used to identify enriched KEGG pathways for the highest and lowest correlated pairs. Highly correlated mRNA-protein pairs belonged to specialized signal transduction pathways, while poorly correlating pairs belonged to house-keeping functions (**Figure 2B**).

Post-transcriptional fine-tuning of transcript stability and translational efficiency is known to be achieved by miRNAs <sup>28</sup>. Using mRNAs experimentally validated <sup>29</sup> to be targeted by miRNAs, we compared Spearman correlation coefficients of these miRNA targets relative to non-targets, and we observed that miRNA targets demonstrated significantly higher mRNA-protein correlations (**Supp. Fig. 2A**). These results confirm the relevance of miRNAs in regulating transcript abundance and target protein levels.

In contrast, molecular mechanisms that occur post-translationally could contribute to divergent mRNA-protein correlation. Evidence of post-translational buffering has been observed to have significant biological importance for members of supernumerary protein complexes <sup>30</sup>, which are theorized to be especially susceptible to degradation due to instability of orphaned components and dependence on precise stoichiometries for assembly and proper activity <sup>31</sup>. Using a dataset of protein complex members obtained from CORUM <sup>32</sup>, co-enrichment of complex members was quantified relative to random protein pairs. This demonstrated higher correlations at the protein level than the transcript level (**Figure 2C and Supp. Fig. 2B**). Previous studies of eukaryotic cells have indicated that post-translational regulation is most commonly observed at subcellular locations where protein complexes assemble <sup>31</sup>. By overlaying MS-based predicted localizations <sup>33</sup> on mRNA-protein correlations, we found that mitochondrial and nuclear proteins showed lower correlations than cytosolic and secretory

proteins (**Supp. Fig. 2C**). In contrast, the secretory protein group, consisting of vesicle and plasma membrane proteins, had higher mRNA-protein correlations<sup>18,33</sup>.

Degradation of protein complex members can occur via non-exponential ubiquitin-mediated degradation or exponential degradation<sup>34</sup>. These mechanisms differed in their degree of post-translational interference as assessed by mRNA-protein correlation in our dataset, and lower mRNA-protein correlations were observed for proteins following non-exponential degradation compared with proteins following exponential degradation (**Supp. Fig. 2D**). Interestingly, non-exponential degradation significantly differed from exponential degradation only for proteins not targeted by miRNAs (**Supp. Fig. 2E**) suggesting complementarity of buffering mechanisms for pervasive gene co-expression<sup>35</sup>.

In parallel with the above-mentioned general mechanisms, we detected unique molecular fingerprints on mRNA-protein ratios related to mutational status. For the REH cell line, which harbors a *GIN52* mutation (P19S)<sup>36</sup>, mRNA-protein correlations for GINS-complex members were significantly poorer than correlations for corresponding protein-mRNA pairs in other cell lines. This suggests that the mutation could have an impact on the regulation or function of the entire protein complex that is only evident at the proteome level. Our data implicates a potential impairment in complex formation, which would lead to collateral degradation of the other GINS-members<sup>37</sup> (**Figure. 2D**).

Considering markers that are frequently involved in ALL initiation and progression, *MTOR*, *AKT2* and *KMT2D* displayed a poor mRNA-protein correlation, and *TP53* showed no correlation (**Figure 2E**). This is in line with the previous observations that *KMT2D* truncating mutations impact protein level but not transcript level<sup>38</sup>. Markers from the B-, and T-cell receptor KEGG pathways showed a similar pattern with a wide range of mRNA-protein correlations (**Supp. Fig. 2F and G**). These data further support that transcript level analysis of

markers in ALL alone allows for imprecision in interpretations of protein-level abundance of markers.

### **Proteome and transcriptome of ALL display alternative subgroupings.**

Consensus clustering of the proteome using high variance proteins ( $n = 3282$ ) suggested seven distinct clusters. Hierarchical clustering including all samples and proteins stratified the cell lines into consensus leukemic clusters (CLC) (**Figure 3A and Supp. Fig. 3**). Although some clusters included multiple cell lines sharing the same cytogenetic subtype, the majority of the proteomics clusters could not be solely explained by shared cytogenetic background. CLC1 contained cell lines with ETV6-RUNX1, BCR-ABL1, KMT2A-MLLT1 and IGH-MYC gene fusions. CLC2 consisted primarily of TCF3-rearranged ALL cell lines (TCF3-PBX and TCF3-HLF) along with several cell lines with previously unspecified cytogenetic background. CLC3 contained 16 out of 22 T-ALL cell lines, again with various cytogenetic backgrounds and characterized by G2M checkpoint hallmark, higher spliceosome and higher E2F target activity when compared to the other six T-ALL cell lines that clustered in CLC7 (**Supp. Table 6**).

Given that co-transcriptional regulatory mechanisms have been implicated in the assembly and degradation of the spliceosome <sup>39</sup>, we performed DE analysis on the transcriptomics data between CLC3 and CLC7 followed by GSEA analysis, which failed to recapitulate the upregulation of the spliceosome (Normalized enrichment score (NES) 0.61,  $q = 1.0$ ) (**Figure 3B**). Altered splicing profiles and differential splicing have been implicated in drug and therapy resistance and in ALL <sup>40, 41, 42</sup>, and these results provide further support for the value of using the proteome to quantify key markers.

The B and T lineages are quite distinct cell types and to gain a better resolution on relative differences between the two different lineages, each lineage was subsequently clustered separately. This identified five clusters in the B lineage (B-CLC) and seven clusters in the T lineage (T-CLC) cell lines (**Figure 3C**). Additionally, we performed consensus clustering of

the transcriptomics data, which subdivided each lineage into four different subgroups (**Figure 3D**). In the transcriptomics data, KMT2A-rearranged subtypes (KMT2A-AFF1 and KMT2A-MLLT1) were classified to the same cluster (cluster 1), whereas the proteomics clustering divided these subtypes into two different proteomic clusters (B-CLC1 and B-CLC5), demonstrating that the proteomics analysis delivers an additional angle to reveal phenotypic differences, which may be clinically meaningful. We performed DEqMS analysis between these two KMT2A-rearranged subtypes and identified p53 among the top upregulated proteins in KMT2A-AFF1 cell lines (**Figure 3E, Supp. Table 7**) suggesting differences in the p53 regulatory network between the KMT2A cell lines.

#### **The proteome of ALL cell lines echoes distinct states in B-cell differentiation.**

Mechanisms driving B-lineage hematopoiesis are also implicated in leukemogenesis, and B-lineage markers are commonly used in clinical practice to characterize clinical ALL cases <sup>43</sup>. To evaluate whether biologically relevant lineage states were detected by proteome characterizations of our cell lines, we used established cellular markers of B-cell developmental stages <sup>44</sup>, and we identified corresponding cell states for each of the 25 BCP-ALL cell lines (**Figure 4A, 4B**). Together, relative protein abundance of DNMT, CD19, RAG1/RAG2, and VPREB1 by quantitative proteomics were assessed to immunotype the BCP-ALL cell lines. Three cell lines displayed ambiguous or transitional immunotyping (noted as “pre-B other”), including a BCR-ABL1 fusion cell line and two cell lines from KMT2A-rearranged subgroups. All other cell lines could be assigned to a corresponding cell state. This recapitulates clinical characteristics of samples with these rearrangements; BCR-ABL1 and KMT2A-rearranged leukemias have been previously demonstrated to display mixed lineage markers <sup>45</sup>. Among ambiguously immunotyped samples, we noted especially high abundance of SPN/CD43, a marker of hematopoietic stem cells of lymphoid lineage, which is expressed during development and downregulated during maturation, but can also be re-expressed following

activation <sup>46</sup> (**Figure 4A, 4B**). This indicates that hematopoietic-like traits could contribute to acquisition of an unconventional lineage, in line with theories of the hematopoietic origins of mixed lineage leukemias <sup>47</sup>.

When B-cell state assignments were compared across cytogenetic subtypes, some cytogenetic subtypes appeared to be confined to one B-cell state, while others occupied multiple cell states and proteomic clusters (**Figure 4C**). Uniform B-cell state assignment was identified for TCF3-PBX1 and MEF2D-HNRNPUL1 cell lines across technical and biological replicates, with all samples typed as pro-B or early-pre-B respectively, suggesting that these fusions could be linked to acquisition of an arrested cell state. These fusions were also associated with unified clustering assignments.

To explore how clinically and biologically defined lineage markers are associated with proteome-defined phenotypic subtypes, we examined relative levels of curated lineage markers across B-CLC clustering assignment (**Figure 4D, Figure 4E**). B-CLC2 and B-CLC3 were defined by cell lines displaying either pro-B or pre-B lineage traits, and also demonstrated high relative abundance of the B-lineage maintaining transcription factor PAX5. In contrast, B-CLC1 and B-CLC5 did not appear to be linked by a shared lineage phenotype. However, B-CLC1 demonstrated significant enrichment of FLT3, a hematopoietic trait with pathogenic effects on clinical outcome, commonly associated with leukemia-driving mutations in childhood ALL <sup>48</sup>. The leukemic marker MME (CD10) is also associated with pathogenic patient risk stratification <sup>49</sup>. B-CLC5 contained the cell lines most enriched for this protein (COG355, 380, MHH-CALL4). Additionally, B-CLC5 was characterized by abundance of the adhesion molecule CEACAM1, which induces a tolerogenic immune environment through interactions with TIM3 <sup>50</sup>, suggesting a common retained affinity for oncogenic immunomodulation <sup>51</sup>.

These observations indicate that both conventional lineage and oncogenic traits contribute to proteome-level differences in our cell line panel. Our phenotypic profiling supports current clinical practice in leukemia stratification and suggests that proteomics could be an effective new avenue to explore the drivers that contribute to pathogenic phenotypes.

#### **The drug sensitivity of BCP-ALL cell lines to a set of 528 oncology drugs.**

There is a demand for new ALL therapeutics to improve outcomes for high-risk patients and address disease relapse, and drug repurposing could help identify promising therapeutics that are safe and clinically viable. Given that the proteome represents a cumulative set of druggable target proteins, analysis of the proteome could reveal mechanistically relevant correlations to drug sensitivity, further elucidating the processes and markers that determine sensitivity to drugs. Therefore, we performed DSRT on the BCP-ALL cell lines (n=25) against a panel of 528 FDA approved and investigational oncology drugs at five concentrations. The selective drug sensitivity scores (sDSS) were obtained by normalizing drug sensitivity against drug response of normal bone marrow, as described in <sup>52</sup> (**Figure 5A and Supp. Table 8**).

Broad-spectrum cytotoxic drugs, *e.g.*, conventional chemotherapy drugs demonstrated high activity in a majority of the tested cell lines (**Figure 5B and Supp. Fig. 4A**). Interestingly, numerous inhibitors targeting kinases (*i.e.* Aurora kinases, PLK1 inhibitors) or binding to the anti-apoptotic protein BCL2 (*i.e.* venetoclax, navitoclax) demonstrated high activity across most tested cell lines regardless of cytogenetic subtype (**Supp. Fig. 4A**). Additionally, most cell lines were sensitive to the p53-MDM2 antagonists (*e.g.*, Idasanutlin, SAR405838) except for four cell lines, which were consistently resistant to these antagonists (**Supp. Fig. 4B and Supp. Table 8**), among them the KMT2A-AFF1 cell lines (ALL-PO and SEM) with suspected deregulated p53 pathway.

Glucocorticoids (GCs) are components of standard of care combination chemotherapy protocols for childhood ALL <sup>53, 54</sup>, and screening of patient cells *ex vivo* after diagnosis has

been shown to predict response or resistance to these therapeutics<sup>55</sup>. Most BCP-ALL cell lines demonstrated substantial sensitivity to dexamethasone, prednisolone, and methylprednisolone, except for six cell lines, among them the REH cell line (**Supp. Fig. 4A** and **Supp. Table 8**), which has previously been shown to be resistant to dexamethasone, in agreement with our data<sup>56</sup>. Sensitivity to GCs correlated very well with the glucocorticoid receptor (NR3C1) protein abundance (**Figure 5C**), consistent with associations between favorable patient response to GCs and basal expression levels of *NR3C1* in ALL and myelomas<sup>57,58</sup>. Anthracyclines are also widely used in childhood ALL therapy, and they demonstrated correlation with TOP2A and TOP2B (**Supp. Fig. 4C**). Correlation between favorable outcomes with anthracycline treatment and transcriptional expression of *TOP2A* and *TOP2B* in acute myeloid leukemia has been previously reported<sup>59</sup>. Together, these results demonstrate alignment with the activity profiles of clinically validated drugs and their associated mechanisms.

Identification of patients likely to be resistant or responsive to a targeted therapy is critical, and understanding the mechanism of action (MoA) of drugs is an effective first step to achieve this goal. Correlation between mRNA expression and drug sensitivities has been previously used to analyze MoA of compounds<sup>60</sup>. Limited datasets of Reverse-phase Protein Arrays (RPPA)<sup>61</sup> using ~230 proteins have also been used for protein drug correlations in cancer cell lines, including ALL cell lines (n=7), and these results identified that protein levels had stronger correlations with drug activity than corresponding mRNA levels. To investigate the association between the molecular markers in our dataset and the pharmacologic responses of our cell line panel, we performed pairwise drug sensitivity correlation analysis with our transcriptomics and proteomics data, using drugs with an sDSS score of  $\geq 8$  in at least one cell line (n=336) for the analysis. Overall, relative protein abundances assessed by sDSS score had a significantly ( $p=2.2e-16$ ) higher correlation than relative protein coding mRNA levels (**Supp. Fig. 4D**), in line with previous observations<sup>61</sup>. For each drug, which was effective (sDSS > 8) for one cell



line at minimum, a matrix of Pearson correlation values for proteins correlated to drug sensitivity was used as input for dimensionality reduction using UMAP<sup>62, 63</sup>. Conventional chemotherapy and epigenetic modifying drugs each demonstrated close association in UMAP space, illustrating similar patterns of protein abundance associated with drug sensitivity, which suggests conserved MoAs for the drugs in each respective drug class (**Figure 5D**). Similar patterns could be seen for many other drug target families such as ABL1, HDAC, BCL2, and PI3K inhibitors, further demonstrating a robust setting for MoA elucidation (**Supp. Fig. 4E**). For these drug classes, we examined outliers which were dispersed from their counterparts, and we noted that the selective class I HDAC inhibitor Tacedinaline was substantially dispersed from other class I HDAC targeting drugs. Additionally, TCF3-PBX1 fusion cell lines demonstrated resistance to Tacedinaline, despite responding to a majority of other HDAC inhibitors (**Supp. Table 8**). These observations suggest that further study is warranted to understand why Tacedinaline displays these distinctions from other drugs of its class. Close associations could be seen for drugs with different proposed molecular targets, such as Mubritinib, an ERBB2 inhibitor that correlated significantly with HIF1A inhibitor BAY-87-2243 ( $R = 0.69$ ,  $p = 1.19\text{e-}04$ ) and VCP/p97 inhibitor NMS-873 ( $R = 0.83$ ,  $p = 3.52\text{e-}07$ ), indicating a shared MoA. These drugs have a common alternative molecular target, respiratory complex I<sup>64, 65 66</sup>. Inhibiting the electron transport chain might lead to reduced ATP availability, which could explain the strong positive correlation of these drugs and abundance of proteasomal subunits (**Supp. Fig. 4F**), indicating ATP-consuming high protein turnover rates.

Using a list of reported putative drug target protein(s) for each drug, we ranked the drug sensitivity-protein abundance Pearson correlations of each drug and its putative drug target to examine the relationship between target abundance and drug response across the screened drug library (**Figure 5E**). Among the highly ranked drugs in this analysis, several tyrosine kinase

inhibitors (e.g. imatinib and bafetinib) and receptor tyrosine kinase inhibitors (e.g. sorafenib and quizartinib) correlated well with protein abundance of their putative targets ABL1 and FLT3 respectively (**Figure 5E and Supp. Fig. 4G**). Additionally, PARP1 inhibitors talazoparib, niraparib and olaparib correlated well with PARP1 levels (**Figure 5E**). MDM2 inhibitors positively correlated with MDM2 abundance while anticorrelating with the target of MDM2 degradation, p53 (**Figure 5E**). Thus, mechanistic insights into drug activity can be inferred using correlation analysis between drug sensitivity and baseline protein abundance.

The proteomics coupled with sDSS assessment also identified previously uncharacterized associations between protein abundance and drug sensitivity. The majority of the tested cell lines (n=22) demonstrated sensitivity to Rigosertib. Rigosertib, a pan-kinase inhibitor, has undergone late phase clinical trials for treatment of several hematopoietic malignancies (NCT01167166, NCT01928537, NCT01926587). Rigosertib has several proposed MoAs, such as being a RAS-mimetic compound <sup>67</sup> and an inhibitor of PLK1, which triggers a G2M transition <sup>68</sup>. Rigosertib has also been demonstrated to exhibit microtubule-destabilizing activity <sup>69</sup>, although this mechanism has been questioned amid criticism of drug purity levels <sup>69, 70</sup>. GSEA analysis of the sDSS-protein abundance correlation of Rigosertib identified enrichment of PLK1, AURORA\_B and cell cycle pathways (**Supp. Table 9**). This supports a PLK1 inhibition and cell cycle disruption MoA of Rigosertib in the BCP-ALL cell lines. Among the top correlating proteins with Rigosertib drug sensitivity was MASTL, a microtubule-associated serine/threonine kinase (**Supp. Fig. 4H**), suggesting that the debated microtubule destabilization effects could be a result of pleiotropic binding to MASTL.

Intriguingly, sensitivity to the competitive PDPK1 inhibitor OSU-03012 was negatively correlated with PDPK1 protein abundance (Pearson correlation = -.421, p=.0355). Additionally, no PDPK1 interacting proteins from the STRING network database correlated with drug activity (**Figure 5F**), and no significant correlation was detected for markers of

associated PDPK1 pathways such as MTOR or AKT signaling. Previous work has disputed the designation of OSU-03012 as a PDPK1 inhibitor, and identified that cell death following treatment with OSU-03012 occurred via induction of ER-stress signaling<sup>71</sup>. Additional studies identified changes in abundance and stability of heat shock proteins (HSPs) following OSU-03012 treatment, and suggested that HSPs were targets of OSU-03012<sup>72</sup>. Our dataset identified significant correlation with HSP90 (HSP90AB1, **Figure 5F**) in support of this hypothesis. However, among the most highly correlated proteins were many components of the chaperonin-containing TCP1-complex (**Figure. 5F**), which performs ATP-dependent folding of polypeptide chains and is implicated in ciliogenesis. Drug activity via functional interference with this complex would also be consistent with previous data describing induction of ER-stress following drug treatment, and these novel associations suggest that the chaperonin-containing TCP1-complex could be studied further as a potential target of OSU-03012, and a vulnerability axis in ALL. Additionally, this observation is especially interesting in the context of other ongoing clinical, preclinical and investigational applications of OSU-03012, which have identified the importance of observed changes in chaperone functionality and autophagy induced by drug treatment. This has included several disease indications, namely antifungal activity, antiviral activity, and prion clearance in neurodegeneration<sup>73, 74</sup>. We identified seven cell lines being sensitive to treatment with OSU-03012.

Another drug of interest is Tacrolimus, which demonstrated efficacy in eleven of the tested BCP-ALL cell lines. Tacrolimus is a macrolide lactone widely used clinically as an immunosuppressant due to its ability to inhibit calcineurin-mediated dephosphorylation of NFAT<sup>75</sup>. Previous characterizations have suggested Tacrolimus does not target this pathway by direct binding, but rather inhibits dephosphorylation of NFAT in complex with FKBP1A, a cis-trans prolyl isomerase belonging to the immunophilin protein family<sup>76</sup>. Despite the crucial role of FKBP1A in this mechanism of action, we found that the abundance of FKBP1A

negatively correlated with sensitivity to Tacrolimus ( $R = -0.4$ ,  $p = 0.048$ ). However, another member of the cis-trans prolyl isomerase family, FKBP10, strongly correlated with Tacrolimus sensitivity ( $R = 0.81$ ,  $p = 7.7\text{e-}05$ ) (**Figure 5G**). FKBP10 is known to bind to Tacrolimus, and although previous work has suggested that this interaction affects collagen assembly, the biological impact of the FKBP10/Tacrolimus interaction is under-investigated relative to other tacrolimus binding partners<sup>77, 78</sup>. This unexpected correlation suggests that FKBP10 merits further study with regard to its binding to Tacrolimus and activity profile. Collectively, our results demonstrate that integrated proteomics and drug sensitivity analysis is an especially well-suited approach to identify drug-target specificities and deciphering their MoA in ALL.

### **The phenotypic signature and drug sensitivity of MEF2D-rearranged cell lines**

Across our analysis of cytogenetic subtype, proteomic clustering, drug sensitivity, and cellular state, we observed striking similarities among a group of cell lines distinguished genetically by a *MEF2D-HNRNPUL1* fusion. *MEF2D* is a member of the myocyte-specific enhancer factor 2 (MEF2) family of transcription factors, which have key roles in a variety of malignancies<sup>79</sup> and in B-cell development<sup>80</sup>. In the context of childhood ALL, *MEF2D* has been found fused to at least six different fusion partners<sup>20, 81</sup>. The *MEF2D* rearranged cytogenetic subtype is associated with relatively poor clinical outcomes<sup>20, 81, 82, 83</sup>, and is also found in adult-onset ALL.

*MEF2D* rearrangements are associated with increased *HDAC9* levels and MEF2D transcriptional activity. HDAC9 is a class II histone deacetylase that regulates the activity of MEF2 transcription factors<sup>84</sup>. Two studies demonstrated the sensitivity of xenografted and primary culture samples carrying another *MEF2D* fusion, MEF2D-BLC9 to broad class I HDAC inhibitors<sup>20, 83</sup>.

MEF2D-HNRNPUL1 leukemias display a distinct gene expression signature<sup>20, 81</sup>, which showed substantial overlap with our data (**Supp. Fig. 5A, Supp. Table 10**). Phenotypically,

these cell lines clustered together when analyzed by transcriptomics or proteomics, and our cell state assessment immunotyped them uniformly as early pre-B cells (**Supp. Fig. 5B**), which is consistent with a recent study of a fourth MEF2D-HNRNPUL1 fusion cell line (KASUMI-7) derived from an adult patient <sup>85</sup>. DEqMS analysis identified MEF2C and HDAC9 among the top differentially abundant proteins in comparison to the other BCP-ALL cell lines (**Figure 6A, 6B, Supp. Table 11**), suggesting they may also be sensitive to the broad class I HDAC inhibitors. However, we found no significant difference in sensitivity compared to the other BCP-ALL cell lines (**Supp. Fig. 5C**). To investigate this further, we used TMP269, a potent class IIa HDAC inhibitor with an IC50 of 23 nM for HDAC9 <sup>86</sup>. Again, we found no inhibitory effect in any of the tested cell lines (**Supp. Fig. 5C**). This suggests that HDAC9 may not be a therapeutically targetable vulnerability, at least as a leukemic cell-intrinsic toxicity in patients carrying the MEF2D-HNRNPUL1 fusion.

In contrast to their lack of specific sensitivity to HDAC inhibitors, the three childhood ALL cell lines with the MEF2D-HNRNPUL1 lesion demonstrated uniquely potent and specific sensitivity to Bryostatin-1 ( $p = 5.6e-06$ ), a DAG-analog and protein kinase C (PKC) activator <sup>87</sup> (**Figure 6C**). DSRT analysis of the adult KASUMI-7 cell line also demonstrated significant sensitivity to Bryostatin-1 (sDSS = 24). By assessing proteomic markers with high Pearson correlations to sDSS score, we observed strong correlation between Bryostatin-1 toxicity and markers upregulated in the MEF2D-HNRNPUL1 cell lines (**Figure 6D**).

One of the positively correlating markers was RASGRP1 ( $R = 0.54$ ,  $p = 0.0064$ ), a calcium- and DAG-regulated guanine nucleotide exchange factor, which has been identified as a direct mediator of sensitivity to negative selection in pre-B cells <sup>88</sup>. Negative selection occurs at specific states of B-cell development, and it influences response to BCR stimulus by triggering programmed cell death after sustained, strong BCR signaling. In addition to limiting self-reactive mature B-cells, this mechanism has also been shown to limit the survival of

dysfunctional pre-B cells <sup>89, 90</sup>. In cells that are not vulnerable to negative selection, BCR stimulus can also contribute tonic survival signals, and therefore leukemias often demonstrate alterations in B-cell developmental progression that favor survival <sup>91</sup>. As cells that have expressed the pre-B cell receptor at the early pre-B stage, our *MEF2D* fusion cell lines are at a stage in development that could render them vulnerable to negative selection <sup>89</sup>. In support of an underlying manipulation of this negative selection pathway, we observed that MEF2D fusion cells had enriched abundance of the DAG degrading enzyme DGKH (**Figure 6A**).

Negative selection mimicry by pharmacologic methods has been applied to target pre-B cell leukemia carrying the BCR-ABL fusion <sup>92, 93</sup> as well as in B-cell lymphomas <sup>94</sup>, and has also been suggested for treating ALL <sup>95</sup>. The mechanism of negative selection mediated by RASGRP1 acts through calcium-regulated PKC delta activity: calcium-bound PKC delta phosphorylates RASGRP1 at S332, and this phosphorylation allows RASGRP1 to initiate pro-apoptotic ERK pathway activation <sup>88, 94</sup>. Thus, we evaluated whether Bryostatin-1, as a DAG-analog, works by activating this pro-apoptotic ERK signaling. Following 2 hrs of 100 nM Bryostatin-1 treatment, we observed a significant increase of abundance of phosphorylated ERK compared to DMSO treated controls (**Figure 6E**).

To validate that Bryostatin-1 acts on-target as a DAG analog, we evaluated whether MEF2D-HNRNPUL1 cell lines would be sensitive to another DAG analog, Phorbol myristate acetate (PMA). PMA also conferred potent toxicity in MEF2D-HNRNPUL1 cell lines (**Figure 6F**), and this effect was significantly more potent in MEF2D-HNRNPUL1 cell lines than other tested BCP-ALL cell lines (n=8 BCP-ALL cell lines, p = 8.8e-10), suggesting that the on-target effects of Bryostatin-1 are replicated by PMA. To further confirm that ERK signaling was a mediator of toxicity, we blocked ERK pathway (phosphorylation) by using the MEK inhibitors Trametenib, UO126, and Selumetinib, and the ERK inhibitor ERK 11e. In cells treated with 100nM Bryostatin-1 or 25nM PMA, concurrent dosing with 1uM of MEK or ERK inhibitors

significantly improved viable cell counts for all tested inhibitors (**Figure 6G, Supp. Fig. 5D**). With the exception of UO126, treatment with 1 $\mu$ M MEK or ERK inhibitors alone resulted in slight but significant reductions in viability for MEF2D-HNRNPUL1 fusion cell lines, further supporting that drugs targeting ERK phosphorylation specifically benefit viability in the context of correcting DAG-analog toxicity (**Supp. Fig 5E**).

These characterizations demonstrate that vulnerability to negative selection at the pre-B stage may be partially conserved in leukemic subsets and targetable by DAG analogs. Thus, further study of this therapeutic strategy in MEF2D-HNRNPUL1 fusion leukemias is merited. More broadly, this finding illustrates that proteomics coupled with drug screening is an analytical approach that can support robust and replicable identification of molecular mediators of drug toxicity.

## DISCUSSION

We here report an in-depth multi-omics layered analysis of 49 readily available childhood ALL cell lines encompassing more than 12,000 proteins and 19,000 protein-coding RNA transcripts as well as drug sensitivity to 528 oncology and investigational drugs. This represents the first proteomic analysis of readily available childhood ALL cell lines covering numerous cytogenetic subtypes, which complements our previous proteogenomic analysis<sup>18</sup> of the two most common subtypes (i.e. ETV6-RUNX1 and Hyperdiploid). Genomic profiling has been the major tool utilized for characterization of childhood ALL. However, protein levels have been demonstrated to have a more direct impact on cellular phenotypes, and our resource and analysis thus provides an improved insight into the biology of leukemia. In this study, we demonstrated that the mass spectrometry-based proteomics analysis of cell lines captures many features of ALL and subtype specific biology. Overall, our integrative data recapitulated the poor mRNA-protein correlation further highlighting the importance of addition of proteomic analysis in studying childhood ALL.

By utilizing this resource, we were able to identify novel phenotypic and biological insights. For the TCF3-PBX1 and MEF2D-HNRNPUL1 fusion samples, separation by cytogenetic subtype drove the clustering assignments when assessed using transcriptomics, but phenotypic similarities with other cytogenetic subtypes could be identified in the clustering designations obtained using proteomics data (**Figure 3D**). In the drug sensitivity screening, protein abundances showed higher correlations with drug sensitivity than protein coding mRNA expression, demonstrating the value of proteomics also in characterizing drug activity (**Figure 5E**). The drug sensitivity screening recapitulated clinically validated results, and conventional chemotherapeutic agents and GCs demonstrated broad activity across the majority of the BCP-ALL cell lines.

In addition, we performed drug sensitivity correlation analysis on a proteome-wide level, which identified correlations between clinically used therapeutics and abundance of their molecular targets. Tyrosine kinase inhibitors ranked as the highest correlated drug-target protein pairs. The correlation of drug sensitivities with proteins other than their intended protein targets (e.g. OSU-03012, Tacrolimus and Rigosertib) also provide an excellent basis for elucidating their true molecular targets in ALL. Beyond ALL, this analysis provides general insights that could be used to identify novel off-targets.

Notably, we identified a DAG analog and proposed PKC activator, Bryostatin-1, that demonstrated a phenotype- and subtype-specific efficiency in the MEF2D-HNRNPUL1 cell lines. Our observations demonstrate that Bryostatin-1 activates the proapoptotic PKC/RASGRP1/ERK signaling pathway in early pre-B cells, the progeny stage at which MEF2D-HNRNPUL1 cell lines are arrested. Bryostatin-1 has been investigated as a latency-reversing agent in T-cells for HIV treatment as well as cancer <sup>96,97</sup> and Alzheimer's <sup>98</sup> treatment and has successfully completed several phase I and II clinical trials <sup>99</sup>. Our finding provides an initial basis for further exploration and repurposing of Bryostatin-1 for this high-risk group.



Additionally, our identification of negative selection vulnerability in MEF2D-HNRNPUL1 fusion cases represents a biologically novel therapeutic approach, and as a result of its orthogonal mechanism it could be especially useful in cases of drug resistance and relapse or pursued in a combination therapy. Collectively our proteomics, transcriptomics, and pharmacoproteomics analysis of this childhood ALL cell line panel provide a rich resource for exploration and hypothesis generation.

### **Acknowledgements**

This study was supported by grants from the Swedish Childhood Cancer Foundation (R.J., grant reference TJ2016-0035, PR2016-0019 and PR2019-0025; M.S. TJ2019-0023; J.L. PR2019-0071), the Swedish Research Council (R.J., grant reference 2017-01653, J.L. 2019-04830), Felix Mindus Contribution to Leukemia research (R.J. and M.V.), Dr Åke Olsson Foundation for Hematological Research (R.J., grant reference 2017-00437), Cancer Society Stockholm and the King Gustaf V Jubilee Fund (R.J., grant reference 174182 and 194111, J.L. 181173), Magnus Bergvalls Stiftelse (R.J., grant reference 2017-02421 and 2016-01841). The authors would like to thank Audrey Anastasia, Maria Carmen-Hesselman and Jaromir Mikes for their assistance during the project. We would also like to thank Children's Oncology Group for providing us with cell lines from their repository. Sequencing was performed by the National Genomics Infrastructure (NGI) in Stockholm and SNP&SEQ Technology Platform in Uppsala. The authors also acknowledge SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. We also thank Shanghai Institute of Hematology for providing the RNA-seq data in this study. Last, we would like to pay our gratitude and our respects to the Prof. Dan Grandér for his support and guidance during initiation of this project, Prof. Dan Grandér passed away in 2017.

### **Author contributions**

R.J. conceived and coordinated the study and acquired funding and resources. E.K., E.G.V., G.M. and R.J. performed the LC-MS experiments with support from M.V.. E.K. and R.J. prepared the samples for RNA-seq with support from K.P.T.. T.E. and N.S. performed the DSRT experiments. J.L. provided LC-MS platform, O.P.K. and P.Ö. provided the DSRT platform. M.S. developed the analysis package and analyzed the LC-MS data with support from R.J. and I.L.. L.A. analyzed the RNA-seq data and developed the resource Shiny application. I.L., F.P. and R.Jo. performed the western blot and flow cytometry analyses. I.S. performed the subcellular and miRNA protein corr analysis. R.J. wrote the first draft with M.S., L.A. and I.L. All authors contributed to finalizing the manuscript and approved the final version.

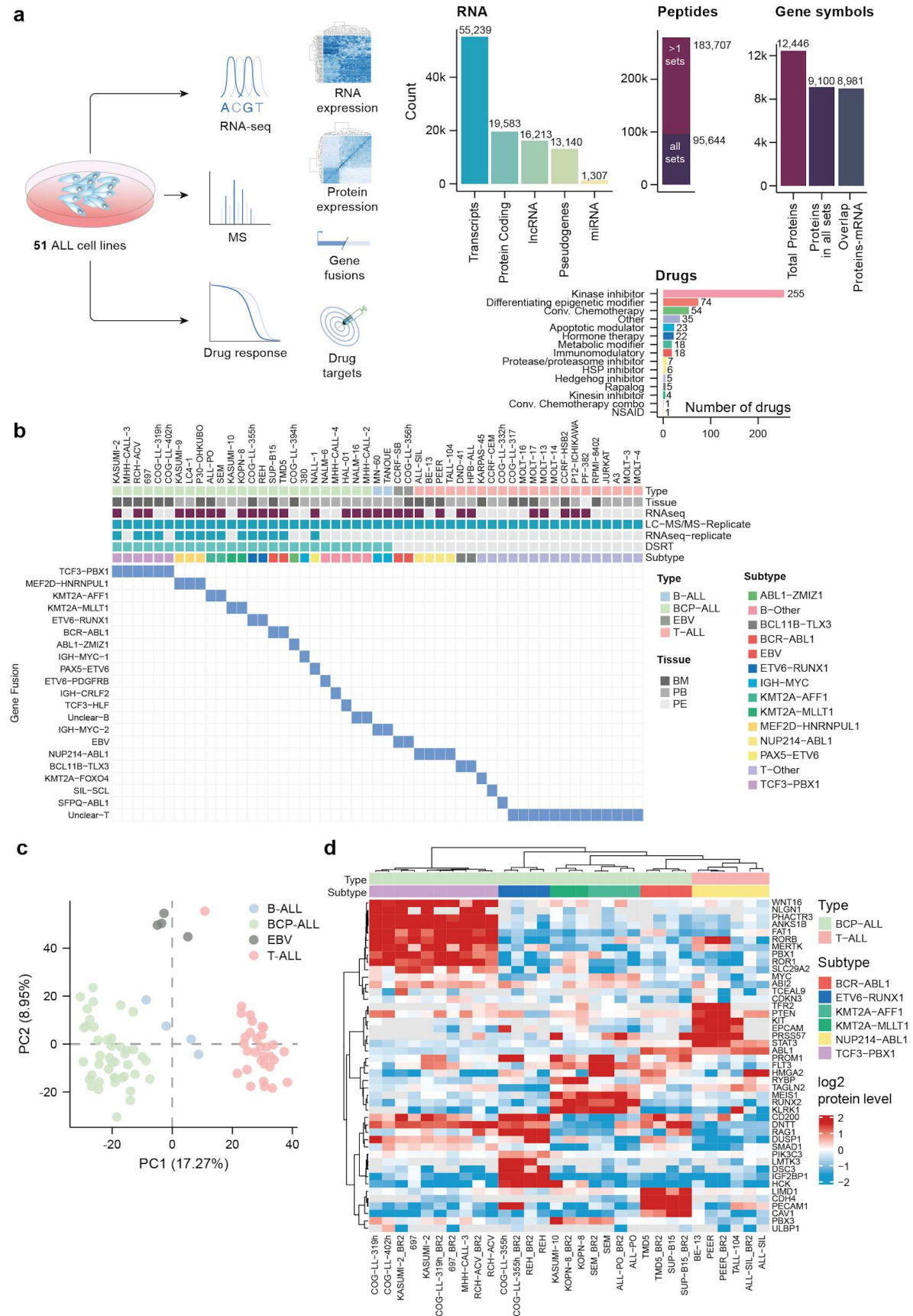
#### **Author information**

These authors contributed equally: Matthias Stahl, Luay Aswad.

#### **Conflict of interest**

The authors declare that they have no conflict of interest.

**Figure 1:**



**Figure 2:**

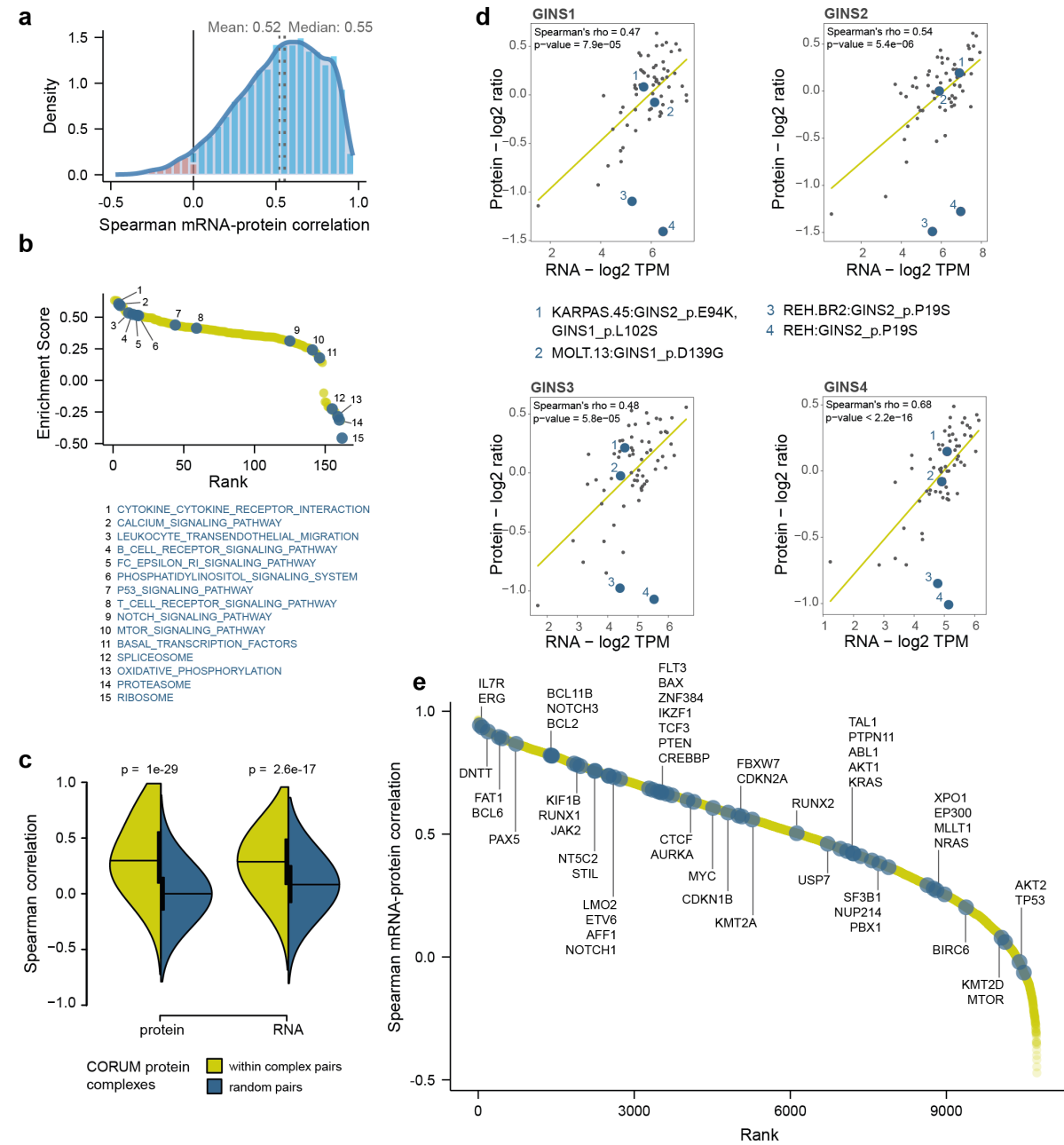
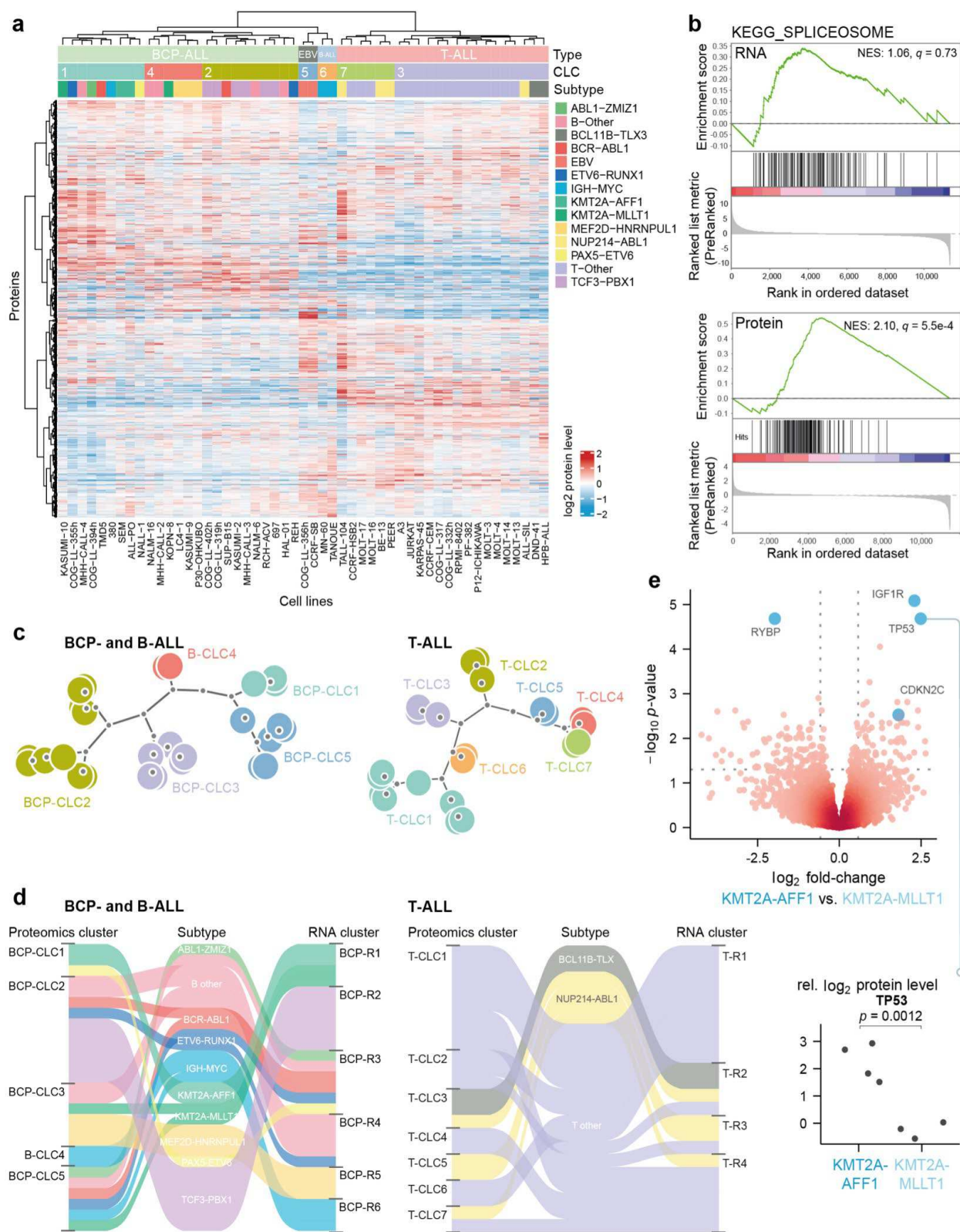
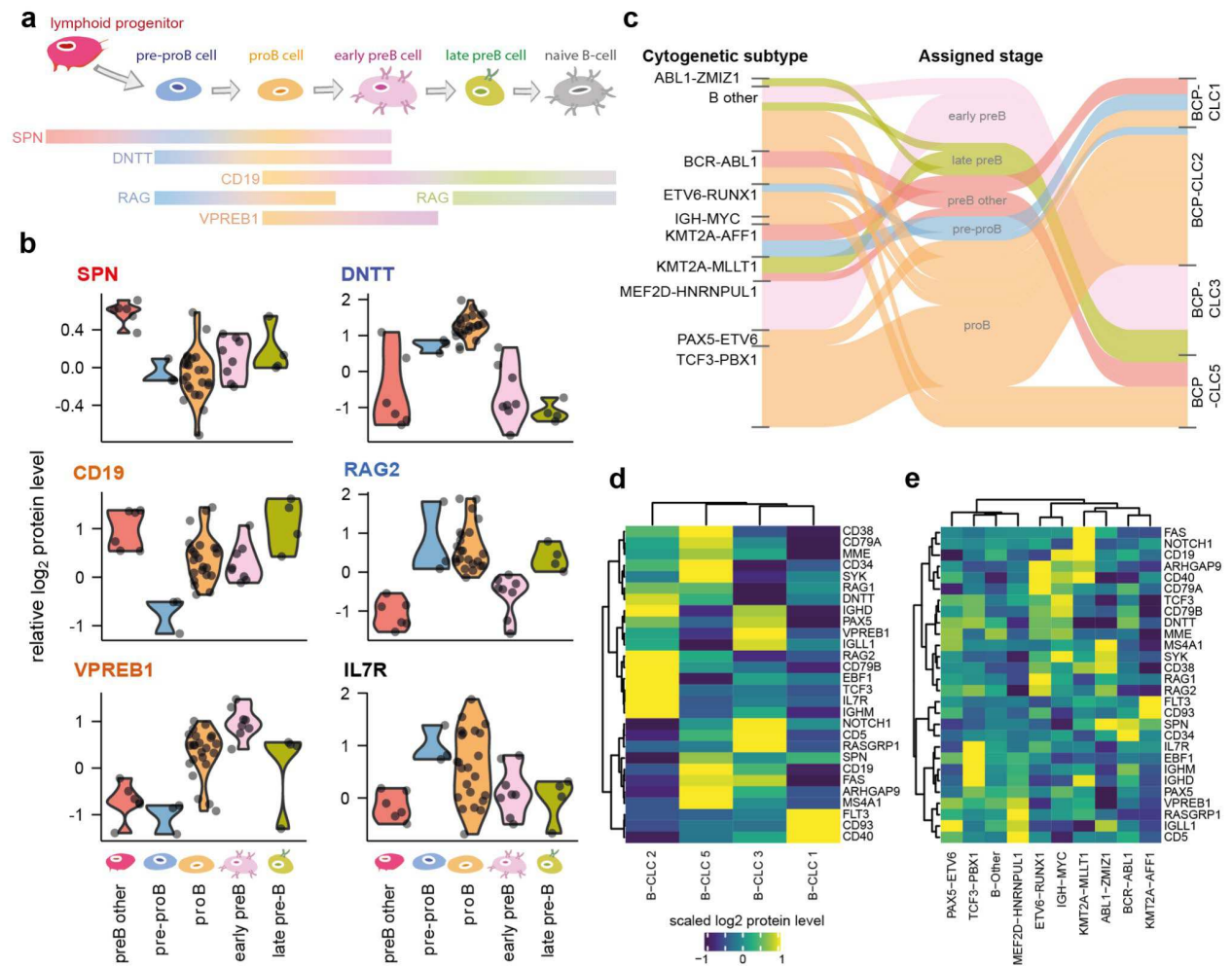


Figure 3:

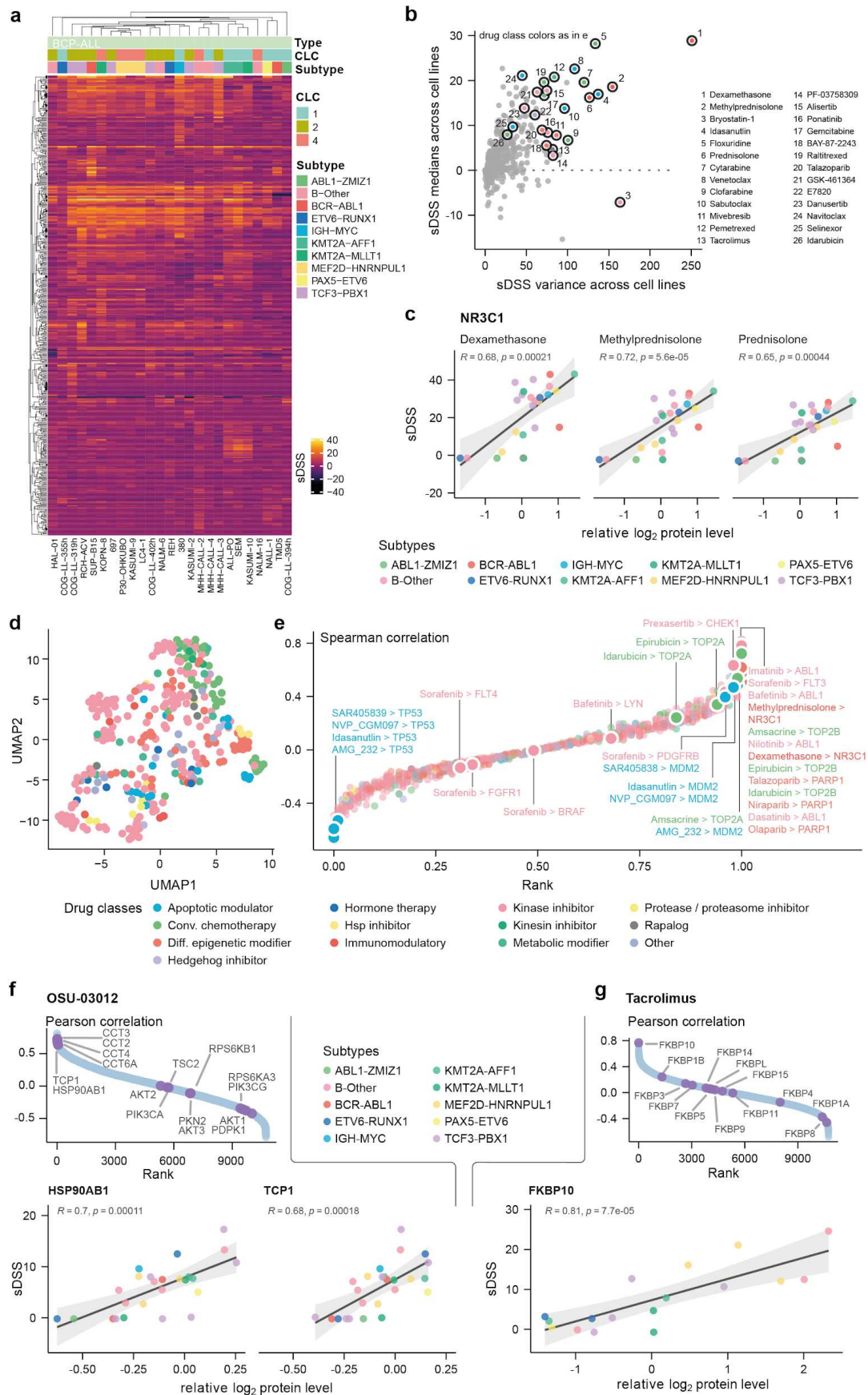




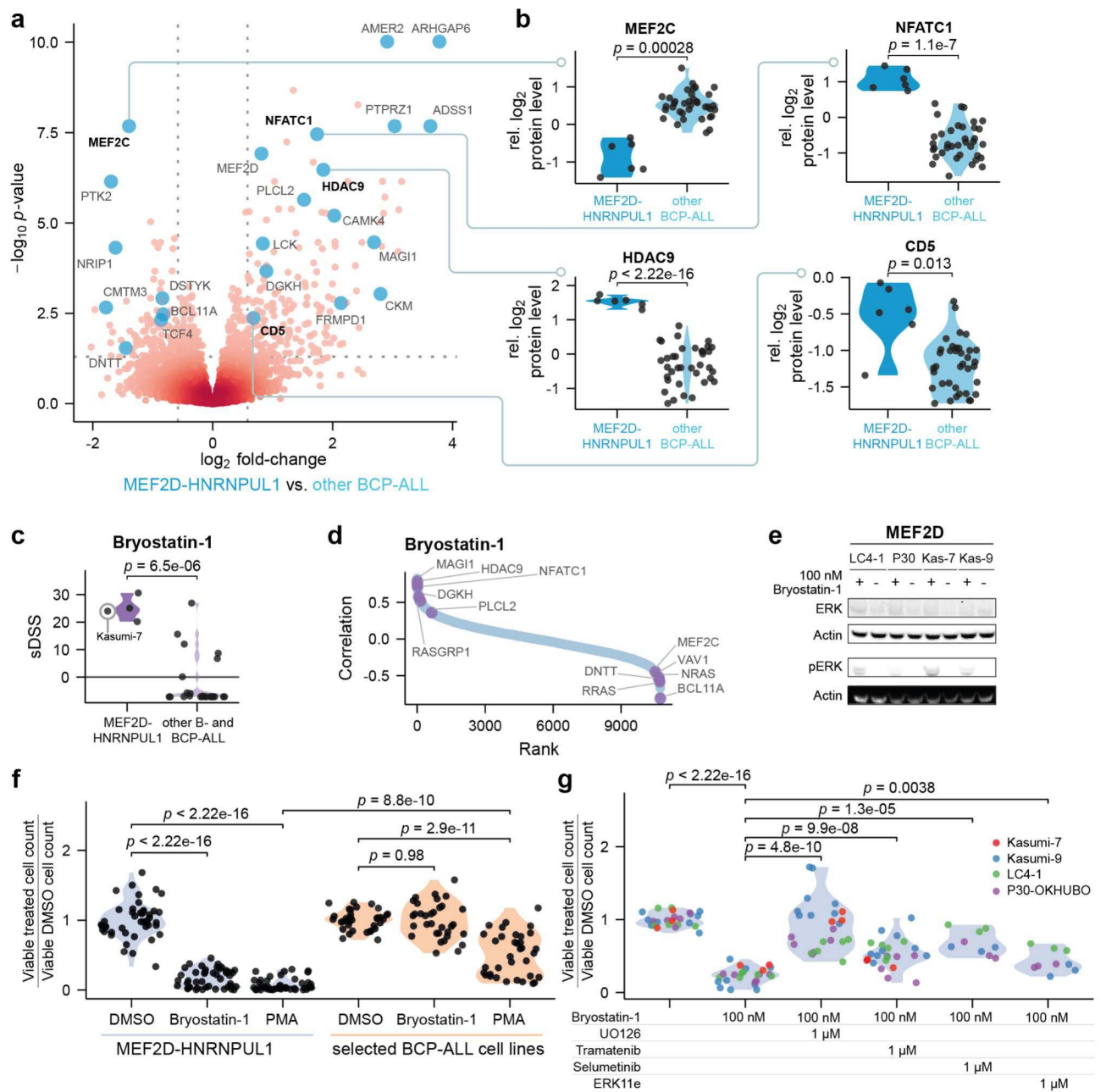
**Figure 4:**



**Figure 5:**



**Figure 6:**





## Figure Legends:

### Figure 1:

- A. Schematic workflow of childhood ALL cell lines data generation and analyses conducted in this study. Cells were profiled for protein level using mass spectrometry and mRNA expression using RNA-seq. A subset of the cells ( $n = 25$ ) was scanned for their sensitivity to 528 investigational and oncology drugs. Protein or RNA gene expression and drug sensitivity data were used for drug target analysis associated with specific subtypes.
- B. An overview of the most prominent molecular features of the 51 cell lines in this panel. BM: bone marrow, PB: peripheral blood, PE: peripheral effusion. DSRT: drug sensitivity and resistance testing.
- C. Principal component analysis (PCA) of the highly variable proteins in the 51 cell lines showed that the BCP-ALL, immature B-ALL and T-ALL cell lines clustered separately. The EBV transformed B-cells also clustered separately.
- D. Cell lines grouped by the most prominent features of the described cytogenetic using proteomics data.

### Figure 2:

- A. Density plot of Spearman rank order of gene symbol wise correlation of the mRNA-protein levels, with RNA or protein markers stratified by gene symbol. The mRNA-protein levels were positively correlated for 97% of the pairs across the entire matching panel ( $n=64$ ), with 78% of the pairs showing significant positive correlation ( $FDR \leq 0.01$ ).
- B. GSEA of the mRNA-protein level correlations for KEGG pathways showing that specialized signal transduction pathways were enriched among the highly correlating pairs, while pathways associated with the spliceosome, proteasome, and ribosomes belonged to the lowest mRNA-protein correlating pairs.
- C. Comparison of random pairwise correlations (blue) of quantitative proteomics data and mRNA levels across the 49 ALL cell lines and their replicates to known interaction pairs (yellow) from the CORUM database demonstrating higher correlations at the protein level.
- D. Relative log2 ratio protein levels versus log2 TPM mRNA levels of GINS complex members across cell lines. The p.D139G mutation in GINS1, p.E94K and p.P16S mutations in GINS2 are shown for annotated cell lines MOLT13, KARPAS-45 and REH respectively.
- E. Waterfall plot showing the ranked mRNA-protein correlations for selected markers that are frequently involved in or associated with leukemogenesis.

### Figure 3:

- A. Heatmap showing the hierarchical clustering of the total overlap of 9118 identified and quantified proteins across 51 cell lines.

- B. GSEA of the differentially expressed genes in CLC3 versus the remaining T-ALL cell lines using the proteomics and transcriptomic data showing significant enrichment of the spliceosome in the proteomics data (NES = 2.1, q-value =  $5.5 \times 10^{-4}$ ) but not in the transcriptomic data (NES = 1.06, q-value = 0.73). Only one biological replicate from the proteomics data was used to have matching numbers and an unbiased comparison to the transcriptomics data.
- C. Dendrogram of the consensus clustering of the B-lineage and T-lineage separately showing five CLC for B-lineage (BCP)-CLC) and seven CLC for T-lineage (T-CLC) cell lines.
- D. Sankey diagram depicting the connection nodes for degree of the proteomics CLCs and with the mRNA based CLCs, by cytogenetic subtype.
- E. Differentially abundant proteins in the KMT2A-AFF1 cell lines compared to the KMT2A-MLLT1 cell lines (top panel) and the TP53 levels in KMT2A-AFF1 cell lines compared to the KMT2A-MLLT1 cell lines (bottom panel).

**Figure 4:**

- A. Schematic diagram of the BCP-ALL cell line differentiation stages, illustrating the markers chosen during the subgrouping assignment by cell state. Relative log<sub>2</sub> protein levels of DNMT, CD19, RAG1/RAG2, and VPREB1 by quantitative proteomics were assessed to immunotype the B-cell precursor leukemia cell lines.
- B. Violin plots depicting the relative protein levels of relevant BCP markers across the BCP-ALL cell lines, grouped by assigned stage.
- C. Sankey diagram illustrating the distribution and connections of BCP-ALL cell lines by CLC assignment (right), B-cell precursor differentiation stage (center), and cytogenetic subtype (left)
- D. Heatmap illustrating the row-normalized log fold change of curated B-cell markers for BCP-ALL cell lines, grouped by B-CLC assignment. Hierarchical clustering was performed by Pearson Ward.D2 for rows and columns.
- E. Heatmap illustrating the row-normalized log fold change of curated B-cell markers for BCP-ALL cell lines, grouped by cytogenetic subtype. Hierarchical clustering was performed by Pearson Ward.D2 for rows and columns.

**Figure 5:**

- A. Heatmap depicting the sDSS of the 528 drugs from the DSRT across the BCP-ALL cell lines (n=25). The x-axis is the cell lines ordered by the rank from hierarchical clustering (Pearson Ward.D2) and the colors are the individual sDSS for each drug. The legends indicate the CLC and cytogenetic subtype of the cell lines. Higher sDSS indicates a more potent and effective drug.
- B. Scatter plot of the median versus variance of the sDSS for each drug across the tested cell lines. Selected drugs are highlighted and colored according to their drug class.
- C. The correlation between the sDSS for the glucocorticoids in the DSRT and relative protein level of the glucocorticoid receptor, NR3C1. The x-axis shows the relative log<sub>2</sub>

protein level of the protein and the y-axis shows the sDSS scores for each respective cell line. The Pearson correlation coefficient ( $R$ ) and p-values for the comparison are shown in each plot. The colors indicate the cytogenetic subtype of the cell lines.

- D. UMAP of the distribution of class-annotated drugs as a two-dimensional reduction. Drugs were correlated to gene symbol-based protein quantification data based on sDSS score, and Pearson correlation scores from this analysis were used to generate the UMAP. Each drug is annotated by its class.
- E. Scatter plot of the Pearson correlation versus rank for the sDSS of each drug and the relative log<sub>2</sub> protein level of its putative protein target. The drugs are colored by drug classes.
- F. Ranked OSU-03012 sDSS and protein level Pearson correlations and scatterplot of OSU-03012 sDSS and HSP90AB1 levels for each of the BCP-ALL cell lines. Pearson correlations show selected members of the TCP1 containing chaperonin complex, HSP90AB1, and PDPK1 interacting proteins from the STRING network database.  $R$  = Pearson correlation coefficient. p-values when calculated are shown within the concerned plot.
- G. Ranked Pearson correlations of Tacrolimus sDSS and highlighted proteins of the FKBP family and scatterplot of Tacrolimus sDSS and FKBP10 levels for each of the BCP-ALL cell lines.  $R$  = Pearson correlation coefficient. p-values when calculated are shown within the concerned plot.

**Figure 6:**

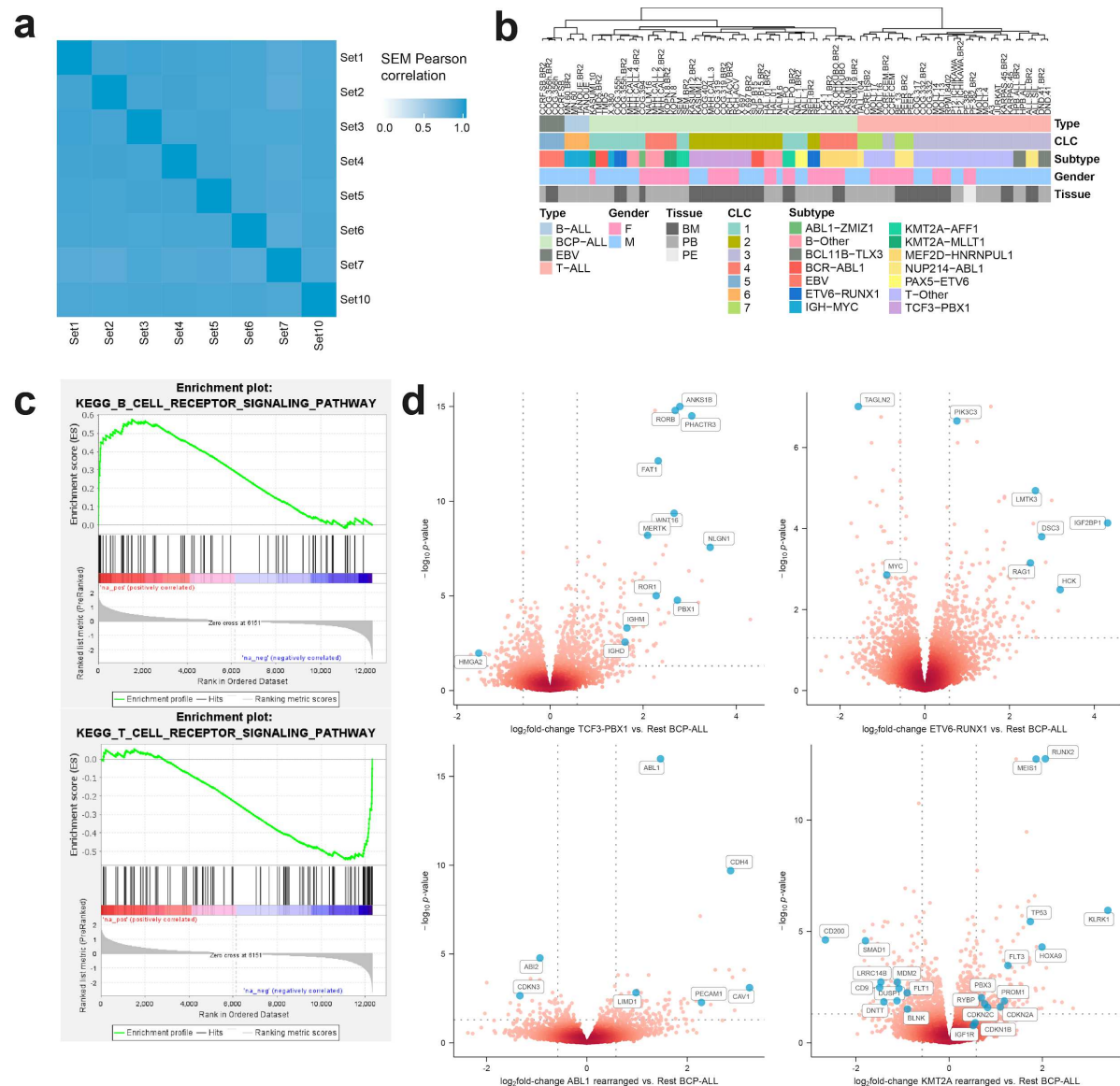
- A. Volcano plot of the differential protein levels in the MEF2D-HNRNPUL1 cell lines and their replicates ( $n = 6$ ) versus the rest of the BCP-ALL cell lines and replicates ( $n = 37$ ). The cut-off was set at  $p \leq 0.05$  and the fold change was set to  $\log_2(1.5)$ .
- B. Relative log<sub>2</sub> protein levels of HDAC9, CD5, NFATC1 and MEF2C in MEF2D-HNRNPUL1 cell lines compared to the remaining BCP-ALL cell lines. The p-value is indicated.
- C. The sDSS of Bryostatin-1 in the MEF2D-HNRNPUL1 cell lines ( $n = 4$ ) compared to the sDSS for the remaining BCP-ALL cell lines ( $n = 22$ ). The p-value is indicated above the bracket.
- D. Ranked Bryostatin-1 sDSS and protein level Pearson correlations with selected highlighted proteins in purple.
- E. Western blot of ERK1/2 and pERK1/2 (Thr202, Tyr204) of MEF2D-HNRNPUL1 fusion cell lines LC4-1, P30-OHKUBO (P30), Kasumi-7 (Kas-7), and Kasumi-9 (Kas-9) after 2-hour treatment with 100 nM Bryostatin-1.
- F. Viable cell quantification normalized to corresponding mean DMSO viable cell count of the four MEF2D-HNRNPUL1 fusion cell lines, treated with 100 nM of Bryostatin-1 or 25 nM PMA. Equal volume of DMSO was used as a control in triplicates, all wells were supplemented with DMSO to obtain equal volumes. BCP-ALL cell lines ALL-PO, REH, RCH-ACV, NALM6, SUP-B15, COG355, COG394, and MHH-CALL2, lacking the MEF2D-HNRNPUL1 fusion, were subjected to the same treatments. Viable cells were quantified by flow cytometry, excluding zombie aqua dyed non-viable cells.

Results are merged from five independent experiments. p-values were obtained from unpaired t-tests.

- G. Viable cell quantification normalized to corresponding mean DMSO viable cell count of the four MEF2D-HNRNPUL1 fusion cell lines, treated with 100 nM of Bryostatin-1 alone or in combination with 1uM MEK inhibitors UO126, Trametinib, or Selumetinib. Alternatively, to block ERK directly, 1uM ERK inhibitor ERK 11e was used, and equal volume of DMSO was used as a control in triplicates. All wells were supplemented to obtain equal DMSO volumes per well. Viable cells were quantified by flow cytometry, excluding zombie aqua dyed non-viable cells. Results are merged from three independent experiments. p-values were obtained from unpaired t-tests.

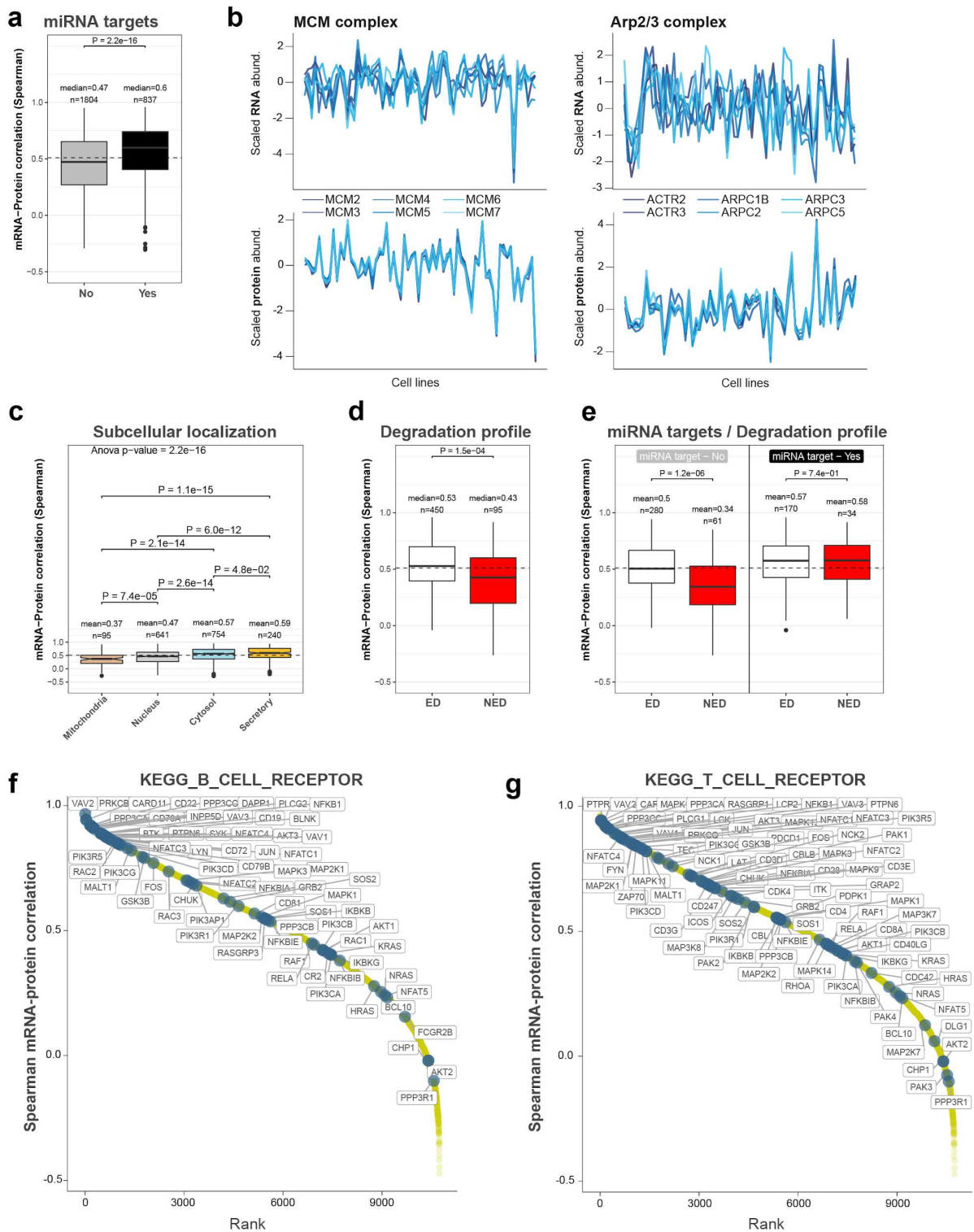
## Supplementary figures and legends:

Supp Figure 1:



- Pearson correlation of the technical replicates of SEM cell line (n=8) using the total overlapping proteins.
- Unsupervised Pearson Ward.D2 hierarchical clustering of the proteomics data and the biological replicates for selected cell lines (n = 32) that was generated approximately one year apart.
- GSEA plots of KEGG B-Cell (NES = 1.9, q-value = 0.0093); KEGG T-Cell (NES = -1.9, q-value = 0.0093) using DE genes between B-, and T lineage cell lines.
- Volcano plots of differentially abundant proteins for the TCF3-PBX1, ETV6-RUNX1, BCR-ABL1 and KMT2A-rearranged cell lines compared to the rest of the BCP-ALL cell lines. The cut-off was set at  $q \leq 0.01$  and the fold changes (log<sub>2</sub>) was set to more than 0.58. Selected proteins are annotated as blue circles.

Supp Figure 2:

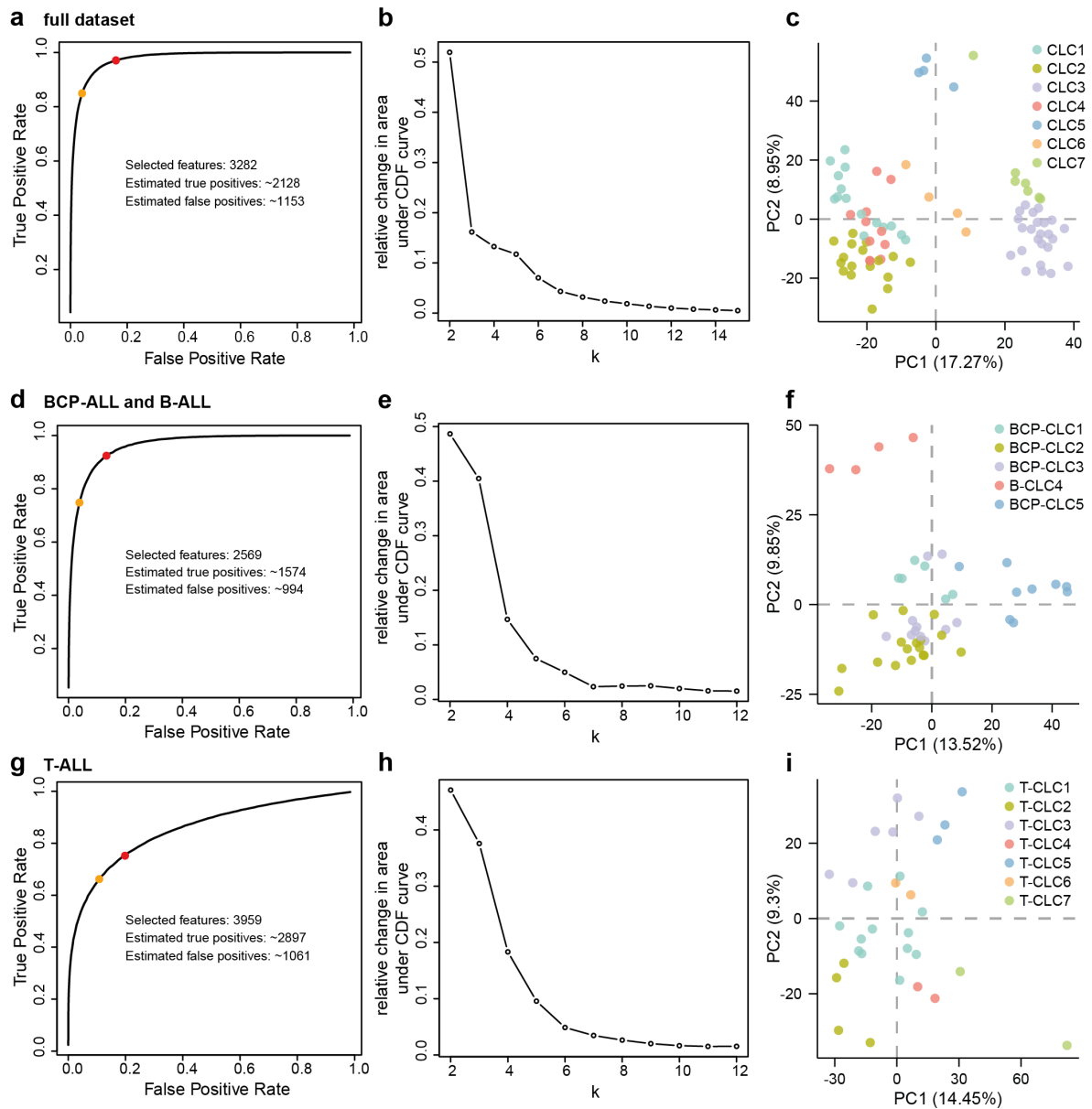


A. mRNA-protein Spearman correlation of miR- and non-miR-targeted protein complex members. The top and bottom of the box reflect first and third quartiles with the median being represented by the band inside each box. The whiskers are 1.5 times the interquartile range and data beyond this range are plotted as individual points as

outliers. The groups were compared using t-test with the p-values indicated above the bracket.

- B. Line chart of protein and mRNA levels across the cell line samples for MCM and Arp2/3 protein complexes.
- C. mRNA-protein Spearman correlation of protein complex members across subcellular locations (SubCellBarCode neighborhoods<sup>33</sup>). The top and bottom of the box reflect first and third quartiles with the median being represented by the band inside each box. The whiskers are 1.5 times the interquartile range and data beyond this range are plotted as individual points as outliers. The groups were compared using t-test with the p-values indicated above the bracket.
- D. mRNA-protein Spearman correlation of protein complex members with exponential (ED) and non-exponential (NED) degradation. The top and bottom of the box reflect first and third quartiles with the median being represented by the band inside each box. The whiskers are 1.5 times the interquartile range and data beyond this range are plotted as individual points as outliers. The groups were compared using t-test with the p-values indicated above the bracket.
- E. mRNA-protein Spearman correlation of protein complex members with exponential (ED) and non-exponential (NED) degradation grouped by miR-target membership. The top and bottom of the box reflect first and third quartiles with the median being represented by the band inside each box. The whiskers are 1.5 times the interquartile range and data beyond this range are plotted as individual points as outliers. The groups were compared using t-test with the p-values indicated above the bracket.
- F. Ranked mRNA-protein Spearman correlation of the genes in the KEGG\_B\_Cell\_Receptor pathway.
- G. Ranked mRNA-protein Spearman correlation of the genes in the KEGG\_T\_Cell\_Receptor pathway.

**Supp Figure 3:**

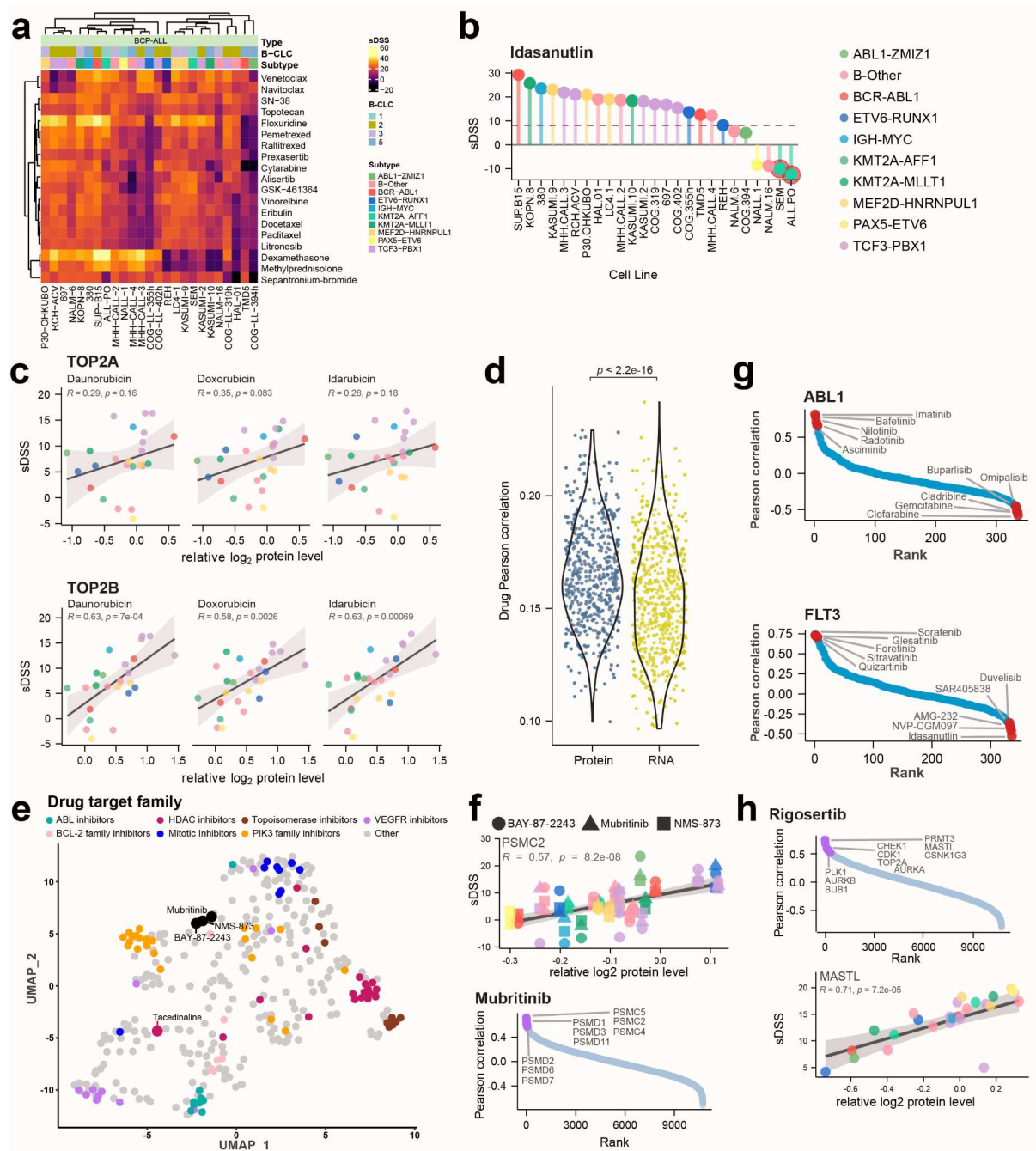


- A. Representative receiver operating characteristic (ROC) curve of the proteomics cluster analysis determining the highly variable proteins in all cell lines (orange: best quantile for separation; red: rounded quantile used due to reproducibility).
- B. Delta area curve of the consensus clustering analysis. The y-axis shows the relative change in area under the Cumulative Distribution Function (CDF) and the x-axis shows the number of clusters ( $k$ ).
- C. PCA plot of the first and second principal components of the determined CLC using highly variable proteins. The colors represent the seven different CLC.
- D. ROC curve of the proteomics cluster analysis determining the highly variable proteins in the B-lineage cell lines.



- E. Delta area curve of the consensus clustering analysis of the B-lineage cell lines. The y-axis shows the relative change in area under the Cumulative Distribution Function (CDF) and the x-axis shows the number of clusters ( $k$ ).
- F. PCA plot of the first and second principal components of the determined B-CLC using highly variable proteins.
- G. ROC curve of the proteomics cluster analysis determining the highly variable proteins in the T-lineage cell lines.
- H. Delta area curve of the consensus clustering analysis of the T-lineage cell lines. The y-axis shows the relative change in area under the Cumulative Distribution Function (CDF) and the x-axis shows the number of clusters ( $k$ ).
- I. PCA plot of the first and second principal components of the determined T-CLC using highly variable proteins.

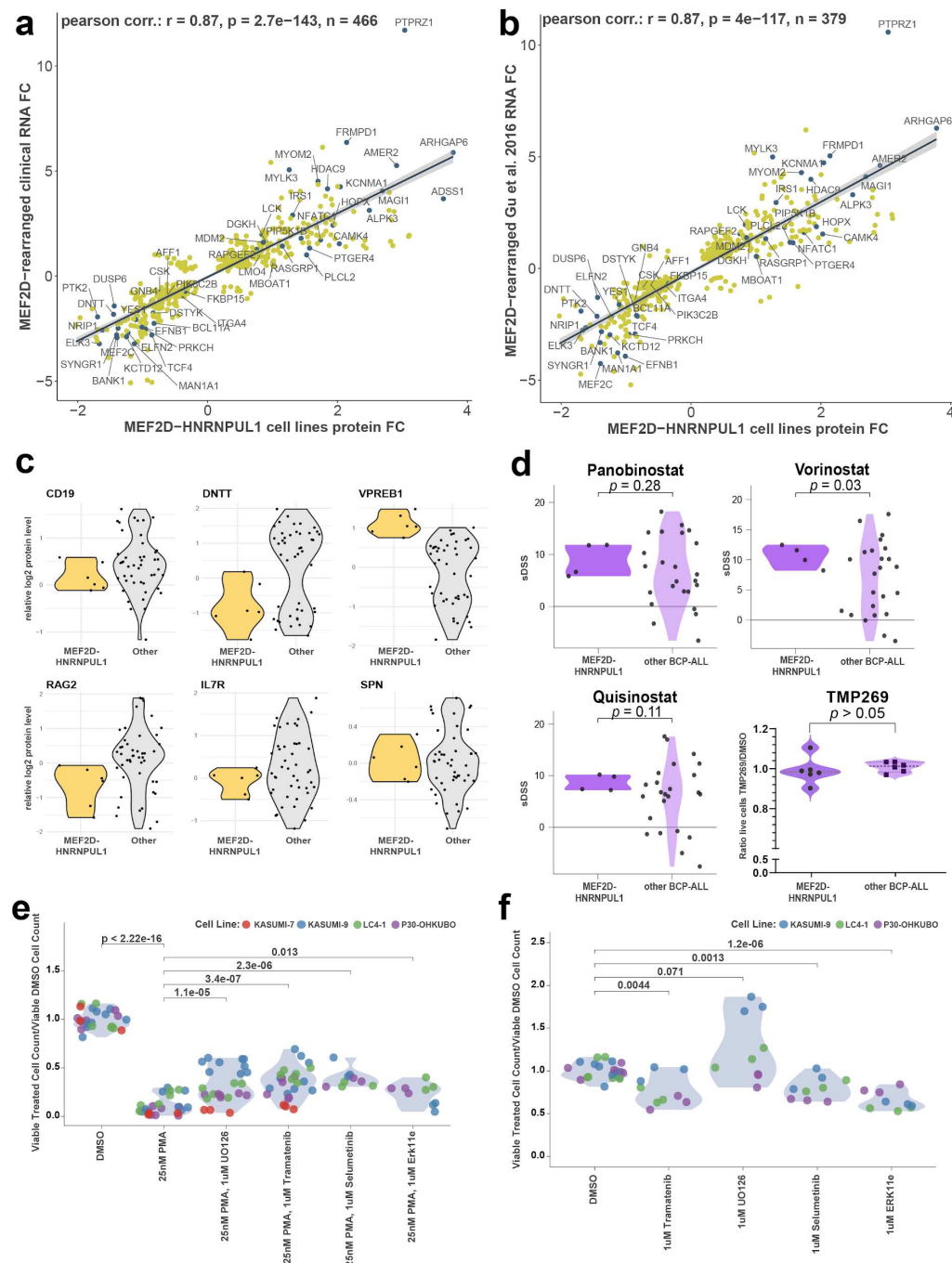
Supp Figure 4:



- A. Heatmap showing the sDSS of selected potent drugs in the DSRT across the BCP-ALL cell lines. The x-axis is the cell lines ordered by the rank from hierarchical clustering (Pearson Ward.D2) and the y-axis is the individual sDSS for each drug. The legends indicate the CLC and cytogenetic subtype of the cell lines.
- B. sDSS of the p53-MDM2 antagonist Idasanutlin across the tested cell lines. The red dashed line indicates the selected threshold of sDSS = 8. The Idasanutlin resistant KMT2A-AFF1 fusion cell lines (SEM and ALL-PO) are highlighted in red circles. The colors annotate the cytogenetic subtype of the cell lines illustrated at the bottom of the figure.

- C. The correlation between the sDSS for selected anthracyclines in the DSRT and relative protein level of TOP2A and TOP2B. The x-axis shows the relative log<sub>2</sub> protein level of the protein and the y-axis shows the sDSS scores for each respective cell line. The Pearson correlation coefficient (*R*) and p-values for the comparison are shown in each plot. The colors indicate the cytogenetic subtype of the cell lines.
- D. Comparison of Pearson average correlations of drug sensitivity per gene at the protein level (blue) or and the mRNA level (yellow).
- E. UMAP dimensional reduction of each drug, calculated using Pearson correlations between sDSS score and protein abundance. Each drug is annotated by specific targets.
- F. Scatterplot of BAY-87-2243, Mubritinib, and NMS-873 sDSS and PSMC2 protein levels (top panel) and ranked Pearson correlations of sDSS and protein levels for Mubritinib with highlighted selected proteasomal subunits (lower panel). *R* = Pearson correlation coefficient. p-values, when calculated, are shown within the concerned plot.
- G. Pearson correlation of drugs sensitivity and ABL1 and FLT3 protein levels, respectively. Top 5 positively and negatively correlating drugs are labeled in red.
- H. Characterization of Rigosertib susceptibility by Pearson correlation of sDSS and protein levels in Rigosertib responders (n=22), and Pearson correlation of sDSS and MASTL protein levels of the BCP-ALL cell lines. *R* = Pearson correlation coefficient. p-values when calculated are shown within the concerned plot.

Supp Figure 5:



A. Fold changes of differentially expressed mRNA in twenty clinical MEF2D-rearranged samples plotted against differentially abundant proteins in MEF2D-HNRNPUL1 cell lines. Only differentially expressed genes that passed a  $p$ -value  $\leq 0.01$  are shown in the scatter plot with selected highlighted genes. N indicates the number of genes passing this criteria.  $R$  = Pearson correlation coefficient.

B. Fold changes of differentially expressed mRNA samples from Gu et al. <sup>20</sup> plotted against differentially abundant proteins in MEF2D-HNRNPUL1 cell lines. Only differentially expressed genes that passed a  $p$ -value  $\leq 0.01$  are shown in the scatter plot

with selected highlighted genes.  $N$  indicates the number of genes passing this criteria.  $R$  = Pearson correlation coefficient.

- C. Immunophenotype of MEF2D-HNRNPUL1 cell lines, showing relative protein levels for relevant markers, plotted alongside the relative protein levels for all other B-lineage ALL cell lines.
- D. Comparison of drug sensitivity of selected HDAC inhibitors in MEF2D-HNRNPUL1 cell lines vs other BCP-ALL cell lines. The MEF2D-HNRNPUL1 cell lines were not significantly more sensitive to class I HDAC inhibitors from the DSRT experiments, the x-axis shows sDSS. The MEF2D-HNRNPUL1 cell lines were also not more sensitive to selective HDAC9 inhibitor (TMP269) at any of the six tested concentrations ranging from 0.0001 to 10 $\mu$ M in 10-fold dilution, the x-axis here shows the ratio of live cells between TMP269 and DMSO treated samples.
- E. Viable cell quantification normalized to corresponding mean DMSO viable cell count of MEF2D-HNRNPUL1 fusion cell lines KASUMI-9, P30-OHKUBO, LC41, and KASUMI 7, treated with 25 nM of PMA alone or in combination with 1 $\mu$ M MEK inhibitors UO126, Trametinib, or Selumetinib. Alternatively, to block ERK directly, 1 $\mu$ M ERK inhibitor ERK 11e was used, and equal volume of DMSO was used as a control in triplicates. All wells were supplemented to obtain equal DMSO volumes per well. Viable cells were quantified by flow cytometry per 15 $\mu$ L HTS collection, excluding zombie aqua dyed non-viable cells. Results are merged from  $n=3$  independent experiments. p-values were obtained from unpaired t-tests.
- F. Viable cell quantification normalized to corresponding mean DMSO viable cell count of MEF2D-HNRNPUL1 fusion cell lines KASUMI-9, P30-OHKUBO, and LC41, treated with 1 $\mu$ M MEK inhibitors UO126, Trametinib, or Selumetinib, or 1 $\mu$ M ERK inhibitor ERK 11e. Equal volume of DMSO was used as a control in triplicates. All wells were supplemented to obtain equal DMSO volumes per well. Viable cells were quantified by flow cytometry per 15 $\mu$ L HTS collection, excluding zombie aqua dyed non-viable cells. p-values were obtained from unpaired t-tests.

## **Supplementary information**

supplementary\_table\_1\_Cell\_line\_panel\_meta\_data

supplementary\_table\_2\_Dynamic\_range\_HiRIEF\_LCMSMS\_gradient\_lengths

supplementary\_table\_3\_RNAseq\_quality\_control\_data

supplementary\_table\_4\_DSRT\_drug\_meta\_data

supplementary\_table\_5\_Differentially\_abundant\_proteins\_B\_vs\_T\_DEqMS

supplementary\_table\_6\_GSEA\_of\_differential\_abundant\_proteins\_CLC3\_vs\_CLC7

supplementary\_table\_7\_Differentially\_abundant\_proteins\_AFF1\_vs\_MLLT1\_DEqMS

supplementary\_table\_8\_DSRT\_sDSS

supplementary\_table\_9\_GSEA\_Rigosertib\_sDSS-protein\_correlations\_GSEA

supplementary\_table\_10\_Differentially\_expressed\_mRNA\_in\_MEF2D-  
Rearranged\_Clinical\_samples\_edgeR

supplementary\_table\_11\_Differentially\_abundant\_proteins\_in\_MEF2D-  
HNRNPUL1\_DEqMS

# METHODS

## Cell cultivation

The 49 childhood acute lymphoblastic leukemia cell lines and two EBV transformed B-cell lines used in this study were obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany), from Children's Oncology Group Childhood Cancer Repository (Lubbock, TX, USA), from American Type Culture Collection (ATCC), Japanese Collection of Research Bioresources Cell Bank (JCRB), European Collection of Authenticated Cell Cultures (ECACC, England) and Banca Biologica e Cell Factory (San Martino, Italy). Roswell Park Memorial Institute (RPMI) 1640 (AQmedia, Sigma-Aldrich) supplemented with either 10% or 20% fetal bovine serum (FBS, Sigma-Aldrich), 20 mM HEPES (Gibco/Life Technologies), 1 mM sodium pyruvate (Sigma-Aldrich), 1x MEM non-essential amino acids (Sigma-Aldrich), and 1x Penicillin-Streptomycin (Sigma-Aldrich) was preferably used. For a few cell lines that were not growing well in RPMI we instead used Iscove's Modified Dulbecco's Medium (IMDM, Sigma-Aldrich) supplemented with 20% FBS or Nutrient Mixture F-10 Ham supplemented with 10% FBS. Detailed information on the provider of the cell lines and growth media can be found in **Supp. Table 1**. Cell lines were grown at 37°C and 5% CO<sub>2</sub> to a cell density of approximately 1-2 million cells/mL. Cells were harvested at 500 x g for 3 min and washed twice with Hank's Balanced Salt Solution (Gibco™ HBSS, no calcium, no magnesium, no phenol red). Aliquots of five million cells were saved for proteomics and transcriptomic analysis. Supernatants were tested for Mycoplasma by MycoAlert Mycoplasma detection kit (Lonza). All cell lines were authenticated by STR profiling (Eurofins Genomics, Ebersberg, Germany).

## Sample preparation for mass spectrometry

Samples were prepared using a modified version of the spin filter aided sample preparation (FASP) protocol<sup>100</sup>. A volume equivalent to 200 µg of protein was digested for each sample. Cell pellets were resuspended in a buffer containing 4% SDS, 25 mM HEPES pH 7.6 and 1mM DTT and lysed by heating to 95°C for 5 min and subsequent sonication (Bandelin). Cell debris were removed by centrifugation at 14,000g for 15 min. Total protein amount was estimated (Bio-Rad DC). For filter-aided sample preparation (FASP), 250 µg of protein sample was mixed with 1 mM DTT, 8 M urea, and 25 mM HEPES pH 7.6 in a centrifugation-filtering unit with a 10-kDa cut-off (Nanosep® Centrifugal Devices with Omega™ Membrane, 10 k). The

samples were then centrifuged for 15 min, 14.000g, followed by another addition of the 8 M urea buffer and centrifugation. Proteins were alkylated by 25 mM IAA, in 8 M urea, 25 mM HEPES pH 7.6 for 10 min, centrifuged, followed by two more additions and centrifugations with 4 M urea, 25 mM HEPES pH 7.6. Protein samples were digested on the filter, first by incubation the samples with Lys-C (Nordic Biolabs (Wako Chemicals GmbH)) overnight at 37°C and an enzyme:protein ratio of 1:50. In the second digestion step trypsin (Thermo Scientific) was added in 50 mM HEPES at an enzyme:protein ratio of 1:100 and incubated for another 8 h at 37°C. The addition of trypsin is repeated for a final over-night incubation. After digestion, the filter units were centrifuged for 15 min, 14.000 g, followed by another centrifugation with 50 µL MilliQ water. Peptides were collected and the peptide concentration was determined. The quality of digest was checked for every sample by LC-MS/MS analysis. Peptides were then dried in a speed-vac. 100 µg of peptides were resuspended in 100 mM TEAB pH 8.5 and labelled with isobaric TMT10-tags (Thermo Scientific) according to manufacturer's instructions, but for three hours. Labelling efficiency was determined by LC-MS/MS before mixing the samples. In total ten TMT10 sets were prepared, whereas some cell lines were analysed in duplicates to assess the influence of the biological variance. Each TMT set contained a pool for posterior dataset normalization that was composed of lysates of different cell lines and digested together with the individual samples. An overview of the sets and the pool composition is given in table **Supp. Table 1**. Individual samples for each TMT set were mixed and were purified by solid phase extraction using SPE strata-X-C columns (Phenomenex) and dried in a SpeedVac.

### High resolution isoelectric focusing (HiRIEF) of peptides

The prefractionation method was applied as previously described in <sup>101</sup>. Sample pools were subjected to peptide IEF-IPG (isoelectric focusing by immobilized pH gradient) in pI range 3 - 10 and 3.7 - 4.9. Dried peptide samples were dissolved in 250 µL rehydration solution containing 8 M urea, and allowed to adsorb to the gel bridge strip by swelling overnight. The 24 cm linear gradient IPG strips (GE Healthcare) were incubated overnight in 8 M rehydration solution containing 1% IPG pharmalyte pH 3 – 10 or 2.5 – 5, respectively (GE Healthcare). After focusing, the peptides were passively eluted into 72 contiguous fractions with MilliQ water/ 35% acetonitrile (CAN)/ 35% ACN + 0.1% formic acid (FA) using an in-house constructed IPG extractor robotics (GE Healthcare Bio- Sciences AB, prototype instrument)



into a 96-well plate (V-bottom, Greiner product #651201), which were then dried in a SpeedVac. The resulting fractions were dried and kept at -20°C.

### LC-MS/MS runs of the HiRIEF fractions

Online LC-MS was performed using a Dionex UltiMate™ 3000 RSLCnano System coupled to a Q-Exactive-HF mass spectrometer (Thermo Fisher Scientific). Each fraction was subjected to MS analysis. Samples were trapped on a C18 guard-desalting column (Acclaim PepMap 100, 75µm x 2 cm, nanoViper, C18, 5 µm, 100Å), and separated on a 50 cm long C18 column (Easy spray PepMap RSLC, C18, 2 µm, 100Å, 75 µm x 50 cm). The nano capillary solvent A was 95% water, 5% DMSO, 0.1% formic acid; and solvent B was 5% water, 5% DMSO, 95% acetonitrile, 0.1% formic acid. At a constant flow of 0.25 µl min<sup>-1</sup>, the curved gradient went from 2% B up to 40% B in **each fraction** as shown in the **Supp. Table 2**, followed by a steep increase to 100% B in 5 min.

FTMS master scans with 60,000 resolution (and mass range 300-1500 m/z) were followed by data-dependent MS/MS (35 000 resolution) on the top 5 ions using higher energy collision dissociation (HCD) at 30% normalized collision energy. Precursors were isolated with a 2 m/z window. Automatic gain control (AGC) targets were 1E6 for MS1 and 1E5 for MS2. Maximum injection times were 100 ms for MS1 and 100 ms for MS2. Dynamic exclusion was set to 30 s duration. Precursors with unassigned charge state or charge state 1 were excluded. An underfill ratio of 1% was used.

### Drug sensitivity and resistance testing of BCP-ALL cell lines

Drug sensitivity and resistance testing was performed as described previously<sup>52</sup>. Briefly, BCP-ALL cell lines were cultured as previously described, 10000 cells were then dispensed (Multidrop Combi, Thermo Fisher Scientific) in 25ul culture media into 384-well tissue culture plates (Corning). The cell lines were tested against 528 drugs and drug combinations in 5-fold dilutions across a ten-thousand-fold concentration range. The compounds were diluted in dimethyl sulfoxide or water where appropriate. Acoustic dispenser Echo® was used to plate the drugs (labcyte). Following incubation for 72h at 37°C and 5% CO<sub>2</sub>, cell viability was measured using CellTiter Glo (Promega). Data was collected on an Ensign (Perkin Elmer) system. Data on each plate were normalized to a plate-specific negative control (vehicle) and a positive control (100umol/L Benzalkonium chloride). Quality control and, selective drug-sensitivity score (sDSS) calculation<sup>102</sup> and data analysis was performed using Breeze

(breeze.fimm.fi)<sup>103</sup>. The sDSS are a modified area under the curve-based metric for assessing drug sensitivities. The drug sensitivity to TMP269 was carried out at six different concentrations ranging from 0.0001 to 10 $\mu$ M at 10-fold dilution using DMSO in the same 384-well format as the DSRT described above apart from the data analysis.

### Western Blotting

Cells were lysed in Cell Signaling Technologies lysis buffer supplemented with protease and phosphatase inhibitors (Halt™ Protease and Phosphatase Inhibitor Single-Use Cocktail, Thermo Fisher Scientific). Protein concentrations were determined using Bio-Rad DC assay (Bio-Rad). Proteins were denatured in SDS sample buffer (Thermo Fisher Scientific), resolved by SDS-PAGE using NuPAGE™ 4 to 12%, Bis-Tris Gel (Invitrogen™, Thermo Fisher Scientific), and transferred to Nitrocellulose membranes (Invitrogen™, Thermo Fisher Scientific). Afterward, the membranes were blocked with 5% non-fat dry milk in TBST (Thermo Fisher Scientific) and incubated with primary antibodies for the appropriate target. Following overnight primary incubation at 4C, blots were rinsed using TBST and incubated with the appropriate horseradish peroxidase (HRP)-conjugated secondary antibodies (lot 3208198, Abcam, cat no. AP127P for mouse anti-human ab and SCBT (sc-2005) for rabbit anti-human ab). All antibody incubations were diluted in 5% non-fat dry milk in TBST. Protein bands were developed with Clarity ECL Substrate Chemiluminescent HRP substrate (Bio-Rad) in a iBright CL1000 Imaging System (Invitrogen™, Thermo Fisher Scientific). Bands were quantified using the ImageJ software version 1.50i. Phospho-ERK1/2 (Thermo Fisher Scientific, cat. No 14-9109-80, RRID:AB\_2572925), ERK1/2 (Thermo Fisher Scientific, cat. No 13-6200, RRID:AB\_2533024), b-actin (Santa Cruz Biotechnology Cat# sc-47778 HRP, RRID:AB\_2714189) antibodies were used for Western blot to detect corresponding targets.

### Cell Viability Assessment by Flow Cytometry

Cell lines were cultured and diluted to plating density in RPMI 1640 (AQmedia, Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich), 20 mM HEPES (Gibco/Life Technologies), 1 mM sodium pyruvate (Sigma-Aldrich), 1x MEM non-essential amino acids (Sigma-Aldrich), and 1x Penicillin-Streptomycin (Sigma-Aldrich). All cell lines were diluted to a plating density of 500k cells/mL. Cells were treated with soluble compounds at the stated concentrations for 72 hours in standard tissue culture incubation conditions (37C, 5% CO<sub>2</sub>) in a 96-well sterile tissue culture plate (Corning). All drug treatments and DMSO

controls were brought to the same relative DMSO volume of 1:200. Following treatment, non-viable cells were stained using 1:500 Zombie Aqua Live Dead stain (ThermoFisher), diluted in PBS (Invitrogen) and added directly to the plated cells (1:2 volume). Cell staining was performed for 1.5 hours on ice, and during staining and all subsequent steps cells were protected from light using aluminum foil. Viable cell counts were obtained using a BD Biosciences LSRFortessa flow cytometer, and cells were collected in equal volumes per well using the high throughput sampler (HTS) plate reader. Gating and quantification was performed using the BD FACS Diva software, and gates were optimized to exclude noise by FSC-A/SSC-A, to exclude doublets by FSC-A/FSC-H, and to exclude dead cells positive in the BV510 channel. Drugs used in these experiments were: Bryostatin-1 (Chem Cruz), phorbol 12-myristate 13-acetate (PMA) (Sigma Aldrich), Trametinib (Cayman chemical), Selumetinib (Selleckchem), ERK 11e (Tocris Bioscience), and UO126 (Cayman chemical).

### Analysis of LC-MS/MS runs

Orbitrap raw MS/MS files were converted to mzML format using msConvert from the ProteoWizard tool suite <sup>104</sup>. Spectra were then searched using MSGF+ (v10072) <sup>105</sup> and Percolator (v2.08) <sup>106</sup>, where search results from 8 subsequent fractions were grouped for Percolator target/decoy analysis. All searches were done against the human protein subset of Ensembl 99 in the Galaxy platform <sup>107</sup>. MSGF+ settings included precursor mass tolerance of 10 ppm, fully-tryptic peptides, maximum peptide length of 50 amino acids and a maximum charge of 6. Fixed modifications were TMT-10plex on lysines and peptide N-termini, and carbamidomethylation on cysteine residues, a variable modification was used for oxidation on methionine residues. Quantification of TMT-10plex reporter ions was done using OpenMS project's IsobaricAnalyzer (v2.0) <sup>108</sup>. PSMs found at 1% FDR (false discovery rate) were used to infer gene identities.

Protein quantification by TMT10plex reporter ions was calculated using TMT PSM ratios to the entire sample set (all 10 TMT-channels) and normalized to the sample median. The median PSM TMT reporter ratio from peptides unique to a gene symbol was used for quantification. Protein false discovery rates were calculated using the picked-FDR method using gene symbols as protein groups and limited to 1% FDR. The eight technical replicates of the SEM cell lines were combined by taking the protein-wise median levels.

### RNA sequencing and transcriptome analysis

Total RNAs were extracted from aliquots harvested at the same time point as proteomics samples using the RNeasy Mini Kit (Qiagen), following manufacturer's instructions using the option of adding  $\beta$ -mercaptoethanol in the lysis step and using DNAase. RNA concentration and quality were determined using Qubit and RNA Assay kit (Thermo Fisher Scientific) and Bioanalyzer with RNA Nano Chips (Agilent Technologies). RNA libraries were prepared with TruSeq Stranded total RNA RiboZero Kit (Illumina) for ribosomal depletion at the sequencing facilities. Sequencing of the libraries (Paired-end 2 x 150 bp) were performed in three different batches by NovaSeq6000 S2 platform (Illumina) at the National Genomics Infrastructure (NGI) in Stockholm and SNP&SEQ Technology Platform in Uppsala. Basic quality control of the sequencing data and reads was performed by the facilities using a standard quality control pipeline (average quality per base = 36  $\pm$  0.7) ([https://github.com/NationalGenomicsInfrastructure/ngi\\_pipeline](https://github.com/NationalGenomicsInfrastructure/ngi_pipeline)). The reads were preprocessed for adapter and quality-based trimming using cutadapt <sup>109</sup> and then mapped to the human reference genome GRCh38 (gencode.v31.p12 primary assembly) using STAR aligner <sup>110</sup> with enabled chimeric reads detection. All samples showed a good percentage of uniquely mapped reads (average 91%). The mapped reads were summarized/quantified at the gene level using featureCounts <sup>111</sup> using gencode.v31 comprehensive gene annotation. The combined counts of all samples were filtered and adjusted for technical variation due the sequencing batches using the ComBat-seq method implemented in the sva R package <sup>112</sup>. The adjusted genes counts were normalized using the edgeR package <sup>113</sup> using Trimmed Mean of M-values (TMM) method <sup>114</sup>. Transcriptomic data for leukemia samples from pediatric patients were obtained from St. Jude (EGAS00001001952) Cloud upon Data Access Agreement (DAC) approval. Two additional RNA-seq samples of MEF2D-HNRNPUL1 subtype were obtained from Shnaghail institute of haematology and were added to St. Jude cohort and processed accordingly using the same RNA-seq analysis pipeline. The reads from bam files were re-aligned to gencode.v31.p12 human genome then the processing steps and annotation files were used identical to the cell lines RNA-seq data processing steps.

### Differential expression analysis of transcriptomics data

RNA-based differential expression analyses were performed using edgeR <sup>113</sup>. FeatureCounts was used to assemble a raw counts matrix, and edgeR used this counts matrix to perform differential expression analysis based on negative binomial distribution. The two-groups

differential expression analysis has been implemented in the data portal for both cell lines and clinical EGA data.

### Consensus clustering of transcriptomics data

Consensus clustering was performed using ConsensusClusterPlus R package (version 1.50) on highly variable RNA within T and B lineage cells independently. The highly variable RNAs were chosen using Median absolute deviation (MAD). Using Pearson correlation for distance measurement and ward.D2 as agglomerative clustering method, optimal number of clusters ( $n = 4$ ) chosen for T- and ( $n = 6$ ) for B- cell types using elbow method. RNA and protein clusters were compared using Sankey plot implemented in the ggalluvial R package (<https://cran.r-project.org/web/packages/ggalluvial/index.html>).

### Identification of highly variable proteins

Identification of highly variable proteins were performed as previously described in <sup>26</sup> with slight modifications. To define proteins with variation between the cell lines we calculated a modified “quantile” standard deviation for each protein by ignoring the lowest and highest values for each protein. Then, the distribution of the modified standard deviation was modeled using a mixture of Gaussian distributions and we used an expectation maximization method (EM) to estimate the different mixture components using the package mixtools, version 1.2.0. The EM process converged in a two distribution solution, which we assumed to represent the highly variable and the unmodulated proteins. Using this model, we estimated the number of highly variable proteins and selected a modified standard deviation threshold, which optimized the number of highly variable minus unmodulated proteins. As the EM process inevitably produces slightly different thresholds every time it is executed, we performed ten iterations and rounded the mean of the iterations to the lower 0.5 in order to have a reproducible solution.

### Clustering of proteomics data

We submitted three different data sets to the clustering procedure: the full panel of cell lines, B cells (BCP-ALL, B-ALL) only, and T cells (T-ALL) only. For each of these datasets we identified the set of highly variable proteins as detailed above. We then marked samples that exhibited no Pearson correlation of 0.5 or greater to any other sample using all proteins with valid values only. Sample replicates were excluded. In order to identify the optimal number of

clusters, we ran the consensus cluster algorithm using the R package ConsensusClusterPlus, version 1.52.0, applying the following parameters: pFeature = 0.8, pItem = 0.8, reps = 2000, clusterAlg = hc. We filtered the dataset for highly variable proteins and non-outlier samples and clustered with different distance measures (Pearson, Spearman) and linkages (average, ward.D2, complete). The number of clusters was determined by the elbow method and the delta area of the cumulative distribution function. After having determined the number of clusters, we reestablished the original datasets with all proteins and the outlier samples and performed hierarchical clustering using 1 - Pearson correlation as distance measure and ward.D2 as linkage method. Biological sample replicates were assigned to the cluster of their parents.

### Differential abundance analysis of proteomics data

Differential abundance of the proteomics data was performed using DEqMS, version 1.6.0<sup>21</sup>. When comparing samples within a specific ALL lineage (B, T), we only selected samples belonging to the same lineage. For each specific comparison, we first stratified the samples to be compared into two different groups. The remaining samples were passed in as a combined third group (“other”). Each of the three groups needed to have at least one valid quantification value for a respective protein. In parallel, for each protein we calculated the sum of quantified PSMs per TMT set. Only sets contributing to the actual DEqMS analysis were taken into account. Then we computed the minimum PSM count across all of these sets while ignoring zero and NA values. If a protein had no quantified PSM, it was excluded from the analysis. This final PSM count per protein was then used to build the empirical Bayes statistics. The significance cut-off was set to an adjusted p-value of 0.01 and fold-changes to log<sub>2</sub>(1.5)-fold difference in abundance.

### Correlation analyses

Correlation between mRNA and protein levels for each gene was performed using overlapping genes (n=8981) between the transcriptomic and proteomics data using 64 matched samples. Correlation between the mRNA and protein abundance values for each of these gene was determined using Spearman’s  $\rho$  correlation method. Correlation analysis between sDSS and protein abundance of all quantified proteins (n=12446) for 25 BCP-ALL cell lines (total n=43 including biological replicates) were also performed using Pearson Rank Correlation.

### Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA, <https://www.gsea-msigdb.org/gsea/index.jsp>) of gene lists from DEqMS, EdgeR and drug sensitivity correlation analysis were performed using the GSEA v4.1.0 software against priori-defined gene sets available at available from Molecular Signatures DataBase (MolSigDB, <https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>). The priori defined hallmark, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Pathway Interaction Database gene sets were used. Weighted enrichment statistics was used for gene sets sizes between 15 and 500, running with 1000 permutations.

### Complex regulation analyses

To investigate protein complex regulation, we used mRNA-protein correlations of all cell lines for CORUM complexes <sup>32</sup> (Complete Dataset) as calculated above. We excluded proteins of spliceosome, Nop56p-associated pre-rRNA complex, ribosome, and proteasome in order to avoid bias of large complexes. Subcellular localizations (neighborhood level) were downloaded from <sup>33</sup>. For miRNA analysis, we downloaded miRNA-mRNA interaction data from <sup>29</sup> excluding ‘weak’ and ‘non-functional’ interactions. Degradation profiles were extracted from supplementary table S4 from <sup>34</sup>. Welch’s t-test and ANOVA were used for two- and multi-level comparisons, respectively, after converting correlations to z-scores. For GINS complex analysis, mutation dataset was downloaded from DepMap portal. Fitted lines were plotted using the ‘lm’ method in the geom\_smooth function.

### UMAP of drug-protein correlations

Using the Pearson correlation coefficient results calculated as described above, a 2D matrix of values was assembled representing Pearson correlations to gene symbol-centric protein quantification values for each drug. Drugs were limited to only the screen results meeting a minimum sDSS threshold of 8 for at least one cell line (n=336 drugs, n=9786 proteins). A scaled, centered matrix was generated and used as input for a PCA using the Seurat package in R (<https://cran.r-project.org/web/packages/Seurat/index.html>). This PCA calculation was used to generate an elbow plot and a jack straw plot, also using functions in the Seurat package. Assessed by these metrics, the number of principal component inputs for UMAP initiation was chosen to be 18. Using the uwot package (<https://cran.r-project.org/web/packages/uwot/index.html>), the Pearson correlation values were input for calculation of UMAP embedding for two dimensions, performed using the following key parameters: n\_neighbors = 25, local\_connectivity = 1, spread = 3.5, min\_dist = .05, metric =

"euclidean", pca = 18, init = "normlaplacian", nn\_method = "annoy". The 2D UMAP was added to the Seurat package as a dimensional reduction and plotted as a ggplot2 object (<https://cran.r-project.org/web/packages/ggplot2/index.html>) using the DimPlot function in Seurat.

### Statistical analysis of flow cytometry data

All quantification was performed using the BDFacsDiva software (BD Biosciences). For each experiment, mean DMSO treated counts were obtained across technical replicates (per cell line, per day of experiment), and this mean value was used to normalize counts to a ratio of viable treated cell count divided by viable mean DMSO cell count. Normalization operations, statistical tests, and data plotting was performed in R using the packages ggplot2 (<https://cran.r-project.org/web/packages/ggplot2/index.html>), and ggpubr (<https://cran.r-project.org/web/packages/ggpubr/index.html>). p-values were obtained from an unpaired t-test between groups performed using the stat\_compare\_means function in ggpubr. Significance was validated in un-normalized raw counts data.

## **Data and code availability**

Processed data, raw data and code will be available upon publication in a peer-reviewed journal.



## References:

1. Bassan R, Hoelzer D. Modern therapy of acute lymphoblastic leukemia. *J Clin Oncol* **29**, 532-543 (2011).
2. Hunger SP, *et al.* Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the children's oncology group. *J Clin Oncol* **30**, 1663-1669 (2012).
3. Nguyen K, *et al.* Factors influencing survival after relapse from acute lymphoblastic leukemia: a Children's Oncology Group study. *Leukemia* **22**, 2142-2150 (2008).
4. Oskarsson T, *et al.* Relapsed childhood acute lymphoblastic leukemia in the Nordic countries: prognostic factors, treatment and outcome. *Haematologica* **101**, 68-76 (2016).
5. Bhakta N, *et al.* The cumulative burden of surviving childhood cancer: an initial report from the St Jude Lifetime Cohort Study (SJLIFE). *Lancet* **390**, 2569-2582 (2017).
6. Biondi A, *et al.* Imatinib treatment of paediatric Philadelphia chromosome-positive acute lymphoblastic leukaemia (EsPhALL2010): a prospective, intergroup, open-label, single-arm clinical trial. *The Lancet Haematology* **5**, e641-e652 (2018).
7. Lee DW, *et al.* T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: a phase 1 dose-escalation trial. *Lancet* **385**, 517-528 (2015).
8. Maude SL, Barrett D, Teachey DT, Grupp SA. Managing cytokine release syndrome associated with novel T cell-engaging therapies. *Cancer J* **20**, 119-122 (2014).
9. Seimetz D, Heller K, Richter J. Approval of First CAR-Ts: Have we Solved all Hurdles for ATMPs? *Cell Med* **11**, 2155179018822781 (2019).
10. Li J, *et al.* Emerging molecular subtypes and therapeutic targets in B-cell precursor acute lymphoblastic leukemia. *Front Med*, (2021).
11. Pui CH, Yang JJ, Bhakta N, Rodriguez-Galindo C. Global efforts toward the cure of childhood acute lymphoblastic leukaemia. *Lancet Child Adolesc Health* **2**, 440-454 (2018).
12. Iacobucci I, Mullighan CG. Genetic Basis of Acute Lymphoblastic Leukemia. *J Clin Oncol* **35**, 975-983 (2017).
13. Mullighan CG, *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758-764 (2007).
14. Hausser J, Mayo A, Keren L, Alon U. Central dogma rates and the trade-off between precision and economy in gene expression. *Nat Commun* **10**, 68 (2019).
15. Mertins P, *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55-62 (2016).
16. Zhang B, *et al.* Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382-+ (2014).
17. Petralia F, *et al.* Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer. *Cell* **183**, 1962-1985 e1931 (2020).
18. Yang M, *et al.* Proteogenomics and Hi-C reveal transcriptional dysregulation in high hyperdiploid

- childhood acute lymphoblastic leukemia. *Nat Commun* **10**, 1519 (2019).
19. Nusinow DP, *et al.* Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **180**, 387-402 e316 (2020).
  20. Gu Z, *et al.* Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia. *Nat Commun* **7**, 13331 (2016).
  21. Zhu Y, *et al.* DEqMS: A Method for Accurate Variance Estimation in Differential Protein Expression Analysis. *Mol Cell Proteomics* **19**, 1047-1057 (2020).
  22. Karvonen H, *et al.* Wnt5a and ROR1 activate non-canonical Wnt signaling via RhoA in TCF3-PBX1 acute lymphoblastic leukemia and highlight new treatment strategies via Bcl-2 co-targeting. *Oncogene* **38**, 3288-3300 (2019).
  23. Polak R, *et al.* Autophagy inhibition as a potential future targeted therapy for ETV6-RUNX1-driven B-cell precursor acute lymphoblastic leukemia. *Haematologica* **104**, 738-748 (2019).
  24. Stoskus M, Vaitkeviciene G, Eidukaite A, Griskevicius L. ETV6/RUNX1 transcript is a target of RNA-binding protein IGF2BP1 in t(12;21)(p13;q22)-positive acute lymphoblastic leukemia. *Blood Cells Mol Dis* **57**, 30-34 (2016).
  25. Kumar AR, *et al.* A role for MEIS1 in MLL-fusion gene leukemia. *Blood* **113**, 1756-1758 (2009).
  26. Johansson HJ, *et al.* Breast cancer quantitative proteome and proteogenomic landscape. *Nat Commun* **10**, 1600 (2019).
  27. Wu L, *et al.* Variation and genetic control of protein abundance in humans. *Nature* **499**, 79-82 (2013).
  28. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. *Nature* **455**, 64-71 (2008).
  29. Chou CH, *et al.* miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res* **46**, D296-D302 (2018).
  30. Juszkievicz S, Hegde RS. Quality Control of Orphaned Proteins. *Mol Cell* **71**, 443-457 (2018).
  31. Taggart JC, Zauber H, Selbach M, Li GW, McShane E. Keeping the Proportions of Protein Complex Components in Check. *Cell Syst* **10**, 125-132 (2020).
  32. Giurgiu M, *et al.* CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res* **47**, D559-D563 (2019).
  33. Orre LM, *et al.* SubCellBarCode: Proteome-wide Mapping of Protein Localization and Relocalization. *Mol Cell* **73**, 166-182 e167 (2019).
  34. McShane E, *et al.* Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. *Cell* **167**, 803-815 e821 (2016).
  35. Kustatscher G, Grabowski P, Rappsilber J. Pervasive coexpression of spatially proximal genes is buffered at the protein level. *Mol Syst Biol* **13**, 937 (2017).
  36. Ghandi M, *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503-508 (2019).
  37. Choi JM, Lim HS, Kim JJ, Song OK, Cho Y. Crystal structure of the human GINS complex. *Genes Dev* **21**, 1316-1321 (2007).
  38. Huang C, *et al.* Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. *Cancer Cell*, (2021).

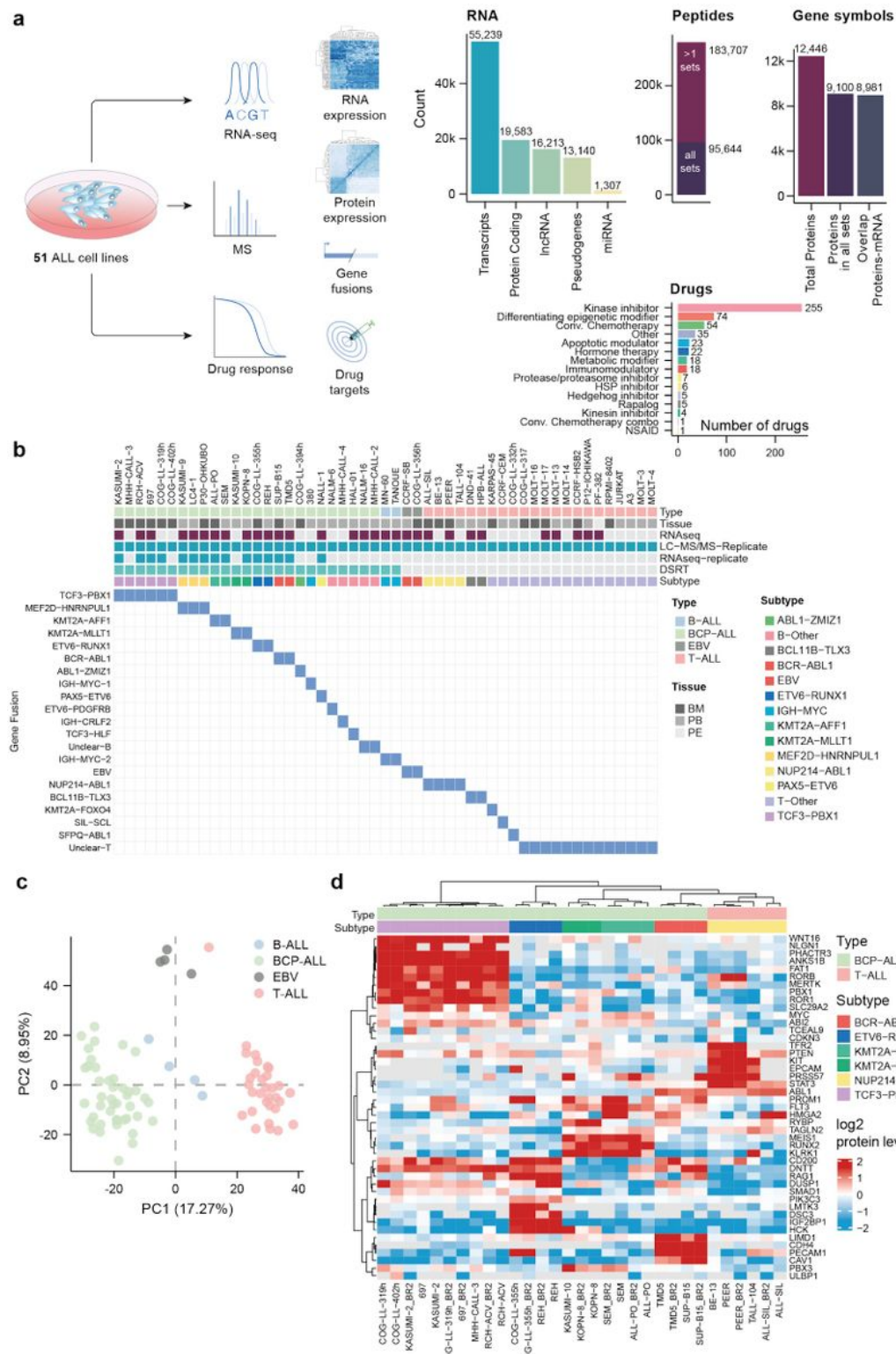
39. Herzel L, Ottoz DSM, Alpert T, Neugebauer KM. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* **18**, 637-650 (2017).
40. Sciarrillo R, *et al.* Glucocorticoid Resistant Pediatric Acute Lymphoblastic Leukemia Samples Display Altered Splicing Profile and Vulnerability to Spliceosome Modulation. *Cancers (Basel)* **12**, (2020).
41. Sotillo E, *et al.* Convergence of Acquired Mutations and Alternative Splicing of CD19 Enables Resistance to CART-19 Immunotherapy. *Cancer Discov* **5**, 1282-1295 (2015).
42. Black KL, *et al.* Aberrant splicing in B-cell acute lymphoblastic leukemia. *Nucleic Acids Res* **47**, 1043 (2019).
43. Campos-Sanchez E, Toboso-Navasa A, Romero-Camarero I, Barajas-Diego M, Sanchez-Garcia I, Cobaleda C. Acute lymphoblastic leukemia and developmental biology: a crucial interrelationship. *Cell Cycle* **10**, 3473-3486 (2011).
44. Hardy RR, Kincade PW, Dorshkind K. The protean nature of cells in the B lymphocyte lineage. *Immunity* **26**, 703-714 (2007).
45. Vardiman JW, *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937-951 (2009).
46. Gulley ML, Ogata LC, Thorson JA, Dailey MO, Kemp JD. Identification of a murine pan-T cell antigen which is also expressed during the terminal phases of B cell differentiation. *J Immunol* **140**, 3751-3757 (1988).
47. So CW, Karsunky H, Passegue E, Cozzio A, Weissman IL, Cleary ML. MLL-GAS7 transforms multipotent hematopoietic progenitors and induces mixed lineage leukemias in mice. *Cancer Cell* **3**, 161-171 (2003).
48. Armstrong SA, *et al.* FLT3 mutations in childhood acute lymphoblastic leukemia. *Blood* **103**, 3544-3546 (2004).
49. Gleissner B, *et al.* CD10- pre-B acute lymphoblastic leukemia (ALL) is a distinct high-risk subgroup of adult ALL associated with a high frequency of MLL aberrations: results of the German Multicenter Trials for Adult ALL (GMALL). *Blood* **106**, 4054-4056 (2005).
50. Huang YH, *et al.* CEACAM1 regulates TIM-3-mediated tolerance and exhaustion. *Nature* **517**, 386-390 (2015).
51. Blaesche F, *et al.* Leukemia-induced dysfunctional TIM-3+CD4+ bone marrow T cells increase risk of relapse in pediatric B-precursor ALL patients. *Leukemia* **34**, 2607-2620 (2020).
52. Pemovska T, *et al.* Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. *Cancer Discov* **3**, 1416-1429 (2013).
53. Inaba H, Greaves M, Mullighan CG. Acute lymphoblastic leukaemia. *Lancet* **381**, 1943-1955 (2013).
54. Pui CH, Evans WE. Treatment of acute lymphoblastic leukemia. *N Engl J Med* **354**, 166-178 (2006).
55. Kaspers GJL, *et al.* In Vitro Cellular Drug Resistance and Prognosis in Newly Diagnosed Childhood Acute Lymphoblastic Leukemia. *Blood* **90**, 2723-2729 (1997).
56. Laane E, *et al.* Cell death induced by dexamethasone in lymphoid leukemia is mediated through initiation of autophagy. *Cell Death Differ* **16**, 1018-1029 (2009).
57. Pui CH, Ochs J, Kalwinsky DK, Costlow ME. Impact of treatment efficacy on the prognostic value of glucocorticoid receptor levels in childhood acute lymphoblastic leukemia. *Leuk Res* **8**, 345-350 (1984).

58. Shuo Ma, *et al.* Glucocorticoid Receptor Expression Correlates with Clinical Outcome in Myeloma Patients Treated with Glucocorticoid-Containing Regimens. *Blood* **112**, (2008).
59. Song JH, *et al.* High TOP2B/TOP2A expression ratio at diagnosis correlates with favourable outcome for standard chemotherapy in acute myeloid leukaemia. *Br J Cancer* **107**, 108-115 (2012).
60. Rees MG, *et al.* Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol* **12**, 109-116 (2016).
61. Li J, *et al.* Characterization of Human Cancer Cell Lines by Reverse-phase Protein Arrays. *Cancer Cell* **31**, 225-239 (2017).
62. Corsello SM, *et al.* Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. *Nat Cancer* **1**, 235-248 (2020).
63. Becht E, *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*, (2018).
64. Baccelli I, *et al.* Mubritinib Targets the Electron Transport Chain Complex I and Reveals the Landscape of OXPHOS Dependency in Acute Myeloid Leukemia. *Cancer Cell* **36**, 84-99 e88 (2019).
65. Ellinghaus P, *et al.* BAY 87-2243, a highly potent and selective inhibitor of hypoxia-induced gene activation has antitumor activities by inhibition of mitochondrial complex I. *Cancer Medicine* **2**, 611-624 (2013).
66. Bouwer MF, Hamilton KE, Jonker PB, Kuiper SR, Louters LL, Looyenga BD. NMS-873 functions as a dual inhibitor of mitochondrial oxidative phosphorylation. *Biochimie* **185**, 33-42 (2021).
67. Athuluri-Divakar SK, *et al.* A Small Molecule RAS-Mimetic Disrupts RAS Association with Effector Proteins to Block Signaling. *Cell* **165**, 643-655 (2016).
68. Gumireddy K, *et al.* ON01910, a non-ATP-competitive small molecule inhibitor of Plk1, is a potent anticancer agent. *Cancer Cell* **7**, 275-286 (2005).
69. Jost M, *et al.* Pharmaceutical-Grade Rigosertib Is a Microtubule-Destabilizing Agent. *Mol Cell* **79**, 191-198 e193 (2020).
70. Baker SJ, *et al.* A Contaminant Impurity, Not Rigosertib, Is a Tubulin Binding Agent. *Mol Cell* **79**, 180-190 e184 (2020).
71. Park MA, *et al.* OSU-03012 stimulates PKR-like endoplasmic reticulum-dependent increases in 70-kDa heat shock protein expression, attenuating its lethal actions in transformed cells. *Mol Pharmacol* **73**, 1168-1184 (2008).
72. Booth L, *et al.* AR-12 Inhibits Chaperone Proteins Preventing Virus Replication and the Accumulation of Toxic Misfolded Proteins. *J Clin Cell Immunol* **7**, 454 (2016).
73. Abdulrahman BA, *et al.* The celecoxib derivatives AR-12 and AR-14 induce autophagy and clear prion-infected cells from prions. *Sci Rep* **7**, 17565 (2017).
74. Chan JF, *et al.* The celecoxib derivative kinase inhibitor AR-12 (OSU-03012) inhibits Zika virus via down-regulation of the PI3K/Akt pathway and protects Zika virus-infected A129 mice: A host-targeting treatment strategy. *Antiviral Res* **160**, 38-47 (2018).
75. Crabtree GR, Olson EN. NFAT Signaling: Choreographing the Social Lives of Cells. *Cell* **109**, S67-S79 (2002).
76. Griffith JP, *et al.* X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex. *Cell* **82**, 507-522 (1995).

77. Knuppel L, *et al.* FK506-binding protein 10 (FKBP10) regulates lung fibroblast migration via collagen VI synthesis. *Respir Res* **19**, 67 (2018).
78. Kolos JM, Voll AM, Bauder M, Hausch F. FKBP Ligands-Where We Are and Where to Go? *Front Pharmacol* **9**, 1425 (2018).
79. Di Giorgio E, Hancock WW, Brancolini C. MEF2 and the tumorigenic process, hic sunt leones. *Biochim Biophys Acta Rev Cancer* **1870**, 261-273 (2018).
80. Herglotz J, *et al.* Essential control of early B-cell development by Mef2 transcription factors. *Blood* **127**, 572-581 (2016).
81. Ohki K, *et al.* Clinical and molecular characteristics of MEF2D fusion-positive B-cell precursor acute lymphoblastic leukemia in childhood, including a novel translocation resulting in MEF2D-HNRNPH1 gene fusion. *Haematologica* **104**, 128-137 (2019).
82. Liu YF, *et al.* Genomic Profiling of Adult and Pediatric B-cell Acute Lymphoblastic Leukemia. *EBioMedicine* **8**, 173-183 (2016).
83. Suzuki K, *et al.* MEF2D-BCL9 Fusion Gene Is Associated With High-Risk Acute B-Cell Precursor Lymphoblastic Leukemia in Adolescents. *J Clin Oncol* **34**, 3451-3459 (2016).
84. Zhang CL, McKinsey TA, Chang S, Antos CL, Hill JA, Olson EN. Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell* **110**, 479-488 (2002).
85. Tsuzuki S, *et al.* Targeting MEF2D-fusion Oncogenic Transcriptional Circuitries in B-cell Precursor Acute Lymphoblastic Leukemia. *Blood Cancer Discovery*, (2020).
86. Lobera M, *et al.* Selective class IIa histone deacetylase inhibition via a nonchelating zinc-binding group. *Nat Chem Biol* **9**, 319-325 (2013).
87. Mutter R, Wills M. Chemistry and clinical biology of the bryostatins. *Bioorg Med Chem* **8**, 1841-1860 (2000).
88. Limnander A, *et al.* STIM1, PKC- $\delta$  and RasGRP set a threshold for proapoptotic Erk signaling during B cell development. *Nat Immunol* **12**, 425-433 (2011).
89. Keenan RA, *et al.* Censoring of autoreactive B cell development by the pre-B cell receptor. *Science* **321**, 696-699 (2008).
90. Melchers F. Checkpoints that control B cell development. *J Clin Invest* **125**, 2203-2210 (2015).
91. Mullighan CG. New strategies in acute lymphoblastic leukemia: translating advances in genomics into clinical practice. *Clin Cancer Res* **17**, 396-400 (2011).
92. Chen Z, *et al.* Signalling thresholds and negative B-cell selection in acute lymphoblastic leukaemia. *Nature* **521**, 357-361 (2015).
93. Shojaee S, *et al.* Erk Negative Feedback Control Enables Pre-B Cell Transformation and Represents a Therapeutic Target in Acute Lymphoblastic Leukemia. *Cancer Cell* **28**, 114-128 (2015).
94. Stang SL, *et al.* A proapoptotic signaling pathway involving RasGRP, Erk, and Bim in B cells. *Exp Hematol* **37**, 122-134 (2009).
95. Müschen M. Autoimmunity checkpoints as therapeutic targets in B cell malignancies. *Nature Reviews Cancer* **18**, 103-116 (2018).
96. Kollár P, Rajchard J, Balounová Z, Pazourek J. Marine natural products: Bryostatins in preclinical and clinical studies. *Pharm Biol* **52**, 237-242 (2014).

97. Wu R, Chen H, Chang N, Xu Y, Jiao J, Zhang H. Unlocking the Drug Potential of the Bryostatin Family: Recent Advances in Product Synthesis and Biomedical Applications. *Chemistry – A European Journal* **26**, 1166-1195 (2020).
98. Sun MK, Nelson TJ, Alkon DL. Towards universal therapeutics for memory disorders. *Trends Pharmacol Sci* **36**, 384-394 (2015).
99. Raghuvanshi R, Bharate SB. Preclinical and Clinical Studies on Bryostatins, A Class of Marine-Derived Protein Kinase C Modulators: A Mini-Review. *Curr Top Med Chem* **20**, 1124-1135 (2020).
100. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).
101. Branca RM, *et al.* HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat Methods* **11**, 59-62 (2014).
102. Yadav B, *et al.* Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. *Sci Rep* **4**, 5193 (2014).
103. Potdar S, *et al.* Breeze: an integrated quality control and data analysis application for high-throughput drug screening. *Bioinformatics* **36**, 3602-3604 (2020).
104. Holman JD, Tabb DL, Mallick P. Employing ProteoWizard to Convert Raw Mass Spectrometry Data. *Curr Protoc Bioinformatics* **46**, 13 24 11-19 (2014).
105. Kim S, Pevzner PA. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* **5**, 5277 (2014).
106. Granholm V, Kim S, Navarro JC, Sjolund E, Smith RD, Kall L. Fast and accurate database searches with MS-GF+Percolator. *J Proteome Res* **13**, 890-897 (2014).
107. Boekel J, *et al.* Multi-omic data analysis using Galaxy. *Nat Biotechnol* **33**, 137-139 (2015).
108. Sturm M, *et al.* OpenMS - an open-source software framework for mass spectrometry. *BMC Bioinformatics* **9**, 163 (2008).
109. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).
110. Dobin A, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
111. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
112. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* **2**, lqaa078 (2020).
113. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
114. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).

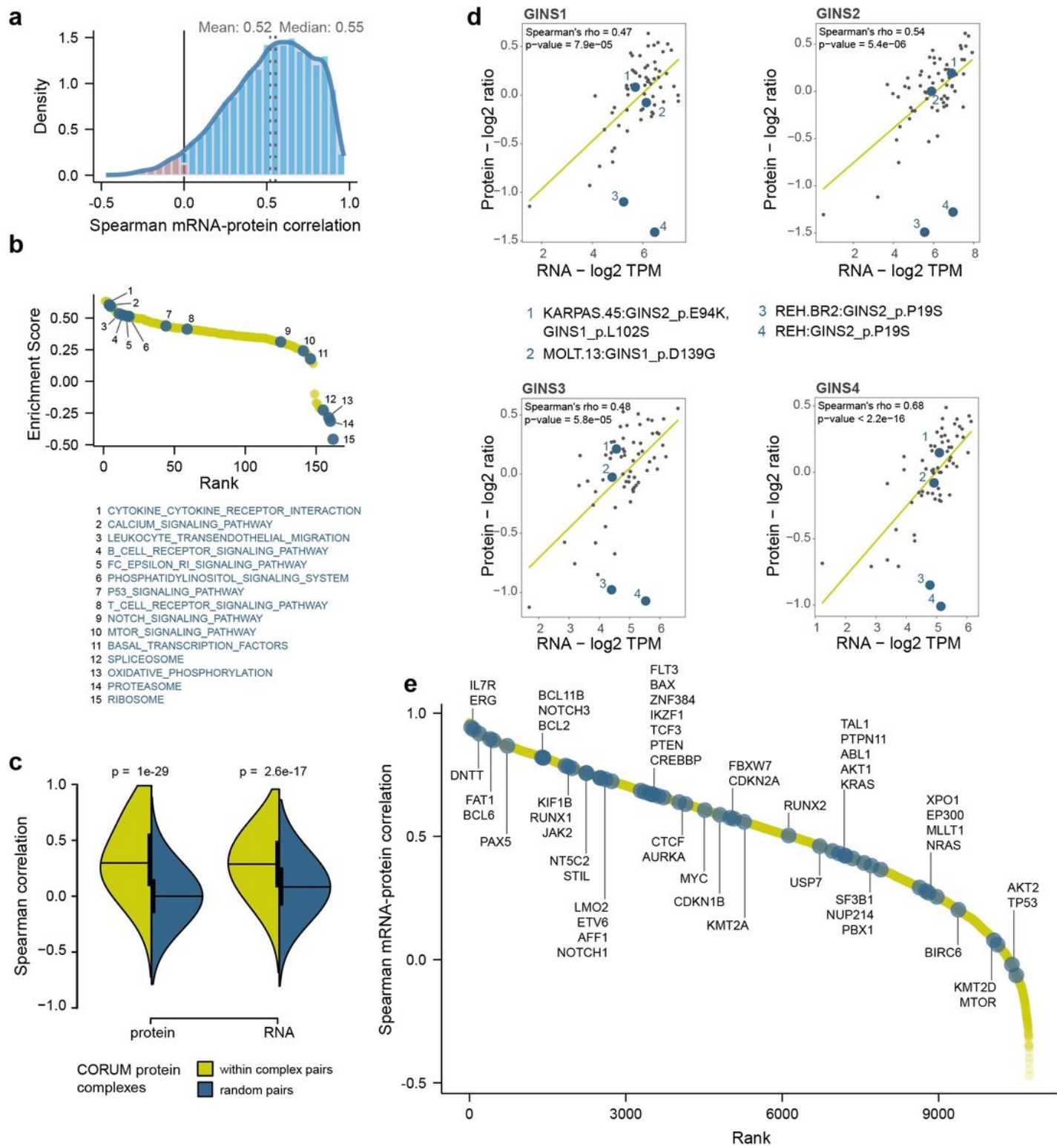
## Figures



### Figure 1

A. Schematic workflow of childhood ALL cell lines data generation and analyses conducted in this study. Cells were profiled for protein level using mass spectrometry and mRNA expression using RNA-seq. A subset of the cells (n = 25) was scanned for their sensitivity to 528 investigational and oncology drugs.

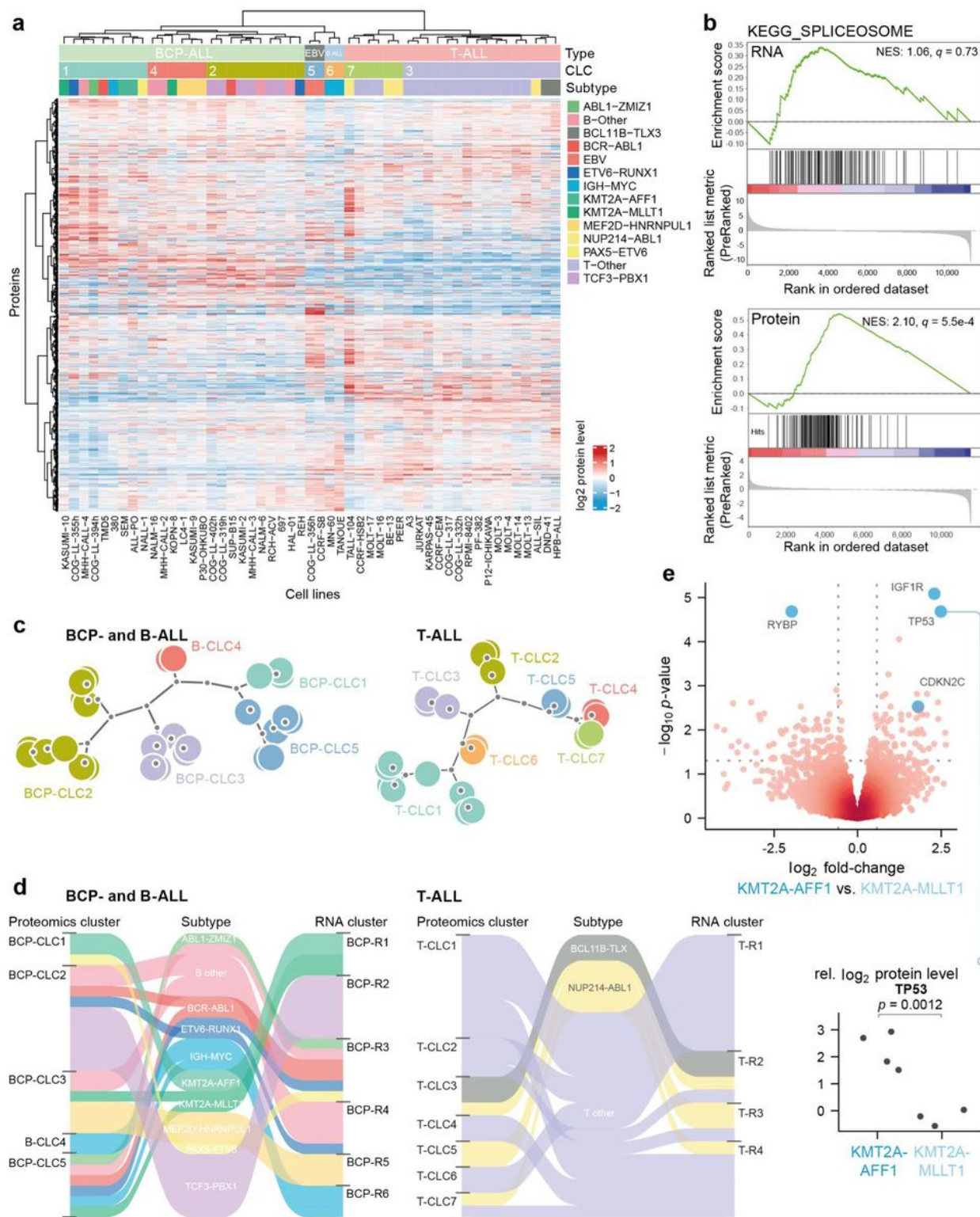
Protein or RNA gene expression and drug sensitivity data were used for drug target analysis associated with specific subtypes. B. An overview of the most prominent molecular features of the 51 cell lines in this panel. BM: bone marrow, PB: peripheral blood, PE: peripheral effusion. DSRT: drug sensitivity and resistance testing. C. Principal component analysis (PCA) of the highly variable proteins in the 51 cell lines showed that the BCP-ALL, immature B-ALL and T-ALL cell lines clustered separately. The EBV transformed B-cells also clustered separately. D. Cell lines grouped by the most prominent features of the described cytogenetic using proteomics data.





## Figure 2

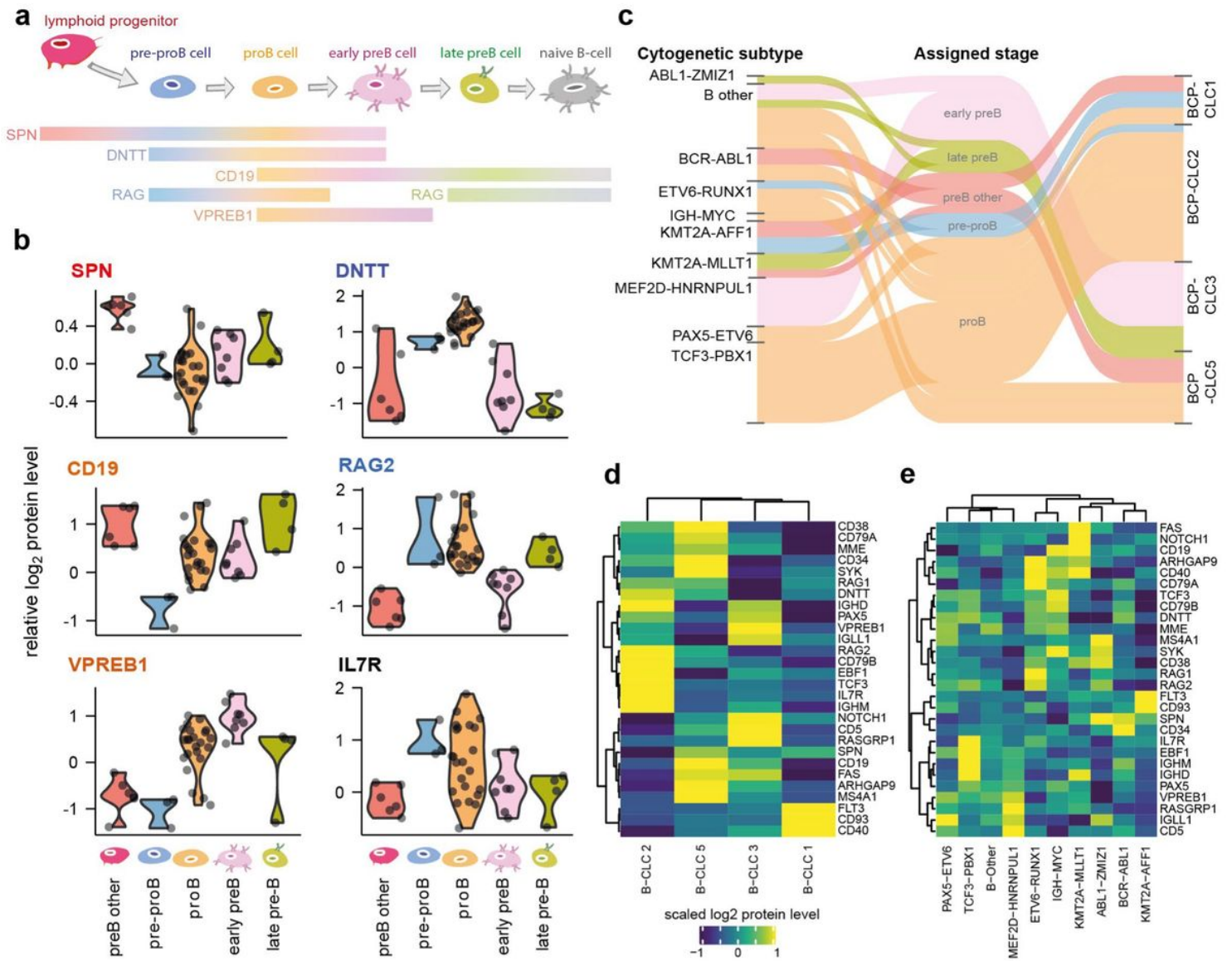
A. Density plot of Spearman rank order of gene symbol wise correlation of the mRNA-protein levels, with RNA or protein markers stratified by gene symbol. The mRNA-protein levels were positively correlated for 97% of the pairs across the entire matching panel (n=64), with 78% of the pairs showing significant positive correlation ( $\text{FDR} \leq 0.01$ ). B. GSEA of the mRNA-protein level correlations for KEGG pathways showing that specialized signal transduction pathways were enriched among the highly correlating pairs, while pathways associated with the spliceosome, proteasome, and ribosomes belonged to the lowest mRNA-protein correlating pairs. C. Comparison of random pairwise correlations (blue) of quantitative proteomics data and mRNA levels across the 49 ALL cell lines and their replicates to known interaction pairs (yellow) from the CORUM database demonstrating higher correlations at the protein level. D. Relative log<sub>2</sub> ratio protein levels versus log<sub>2</sub> TPM mRNA levels of GINS complex members across cell lines. The p.D139G mutation in GINS1, p.E94K and p.P16S mutations in GINS2 are shown for annotated cell lines MOLT13, KARPAS-45 and REH respectively. E. Waterfall plot showing the ranked mRNA–protein correlations for selected markers that are frequently involved in or associated with leukemogenesis.



**Figure 3**

A. Heatmap showing the hierarchical clustering of the total overlap of 9118 identified and quantified proteins across 51 cell lines. 32 B. GSEA of the differentially expressed genes in CLC3 versus the remaining T-ALL cell lines using the proteomics and transcriptomic data showing significant enrichment of the spliceosome in the proteomics data (NES = 2.1,  $q\text{-value} = 5.5e-4$ ) but not in the transcriptomic data (NES = 1.06,  $q\text{-value} = 0.73$ ). Only one biological replicate from the proteomics data was used to have

matching numbers and an unbiased comparison to the transcriptomics data. C. Dendrogram of the consensus clustering of the B-lineage and T-lineage separately showing five CLC for B-lineage (BCP)-CLC) and seven CLC for T-lineage (T-CLC) cell lines. D. Sankey diagram depicting the connection nodes for degree of the proteomics CLCs and with the mRNA based CLCs, by cytogenetic subtype. E. Differentially abundant proteins in the KMT2A-AFF1 cell lines compared to the KMT2A-MLLT1 cell lines (top panel) and the TP53 levels in KMT2A-AFF1 cell lines compared to the KMT2A-MLLT1 cell lines (bottom panel).



**Figure 4**

A. Schematic diagram of the BCP-ALL cell line differentiation stages, illustrating the markers chosen during the subgrouping assignment by cell state. Relative log<sub>2</sub> protein levels of DNTT, CD19, RAG1/RAG2, and VPRED1 by quantitative proteomics were assessed to immunotype the B-cell precursor leukemia cell lines. B. Violin plots depicting the relative protein levels of relevant BCP markers across the BCP-ALL cell lines, grouped by assigned stage. C. Sankey diagram illustrating the distribution and connections of BCP-ALL cell lines by CLC assignment (right), B-cell precursor differentiation stage (center), and cytogenetic subtype (left) D. Heatmap illustrating the row-normalized log fold change of curated B-cell markers for



BCP-ALL cell lines, grouped by B-CLC assignment. Hierarchical clustering was performed by Pearson Ward.D2 for rows and columns. E. Heatmap illustrating the row-normalized log fold change of curated B-cell markers for BCP-ALL cell lines, grouped by cytogenetic subtype. Hierarchical clustering was performed by Pearson Ward.D2 for rows and columns.

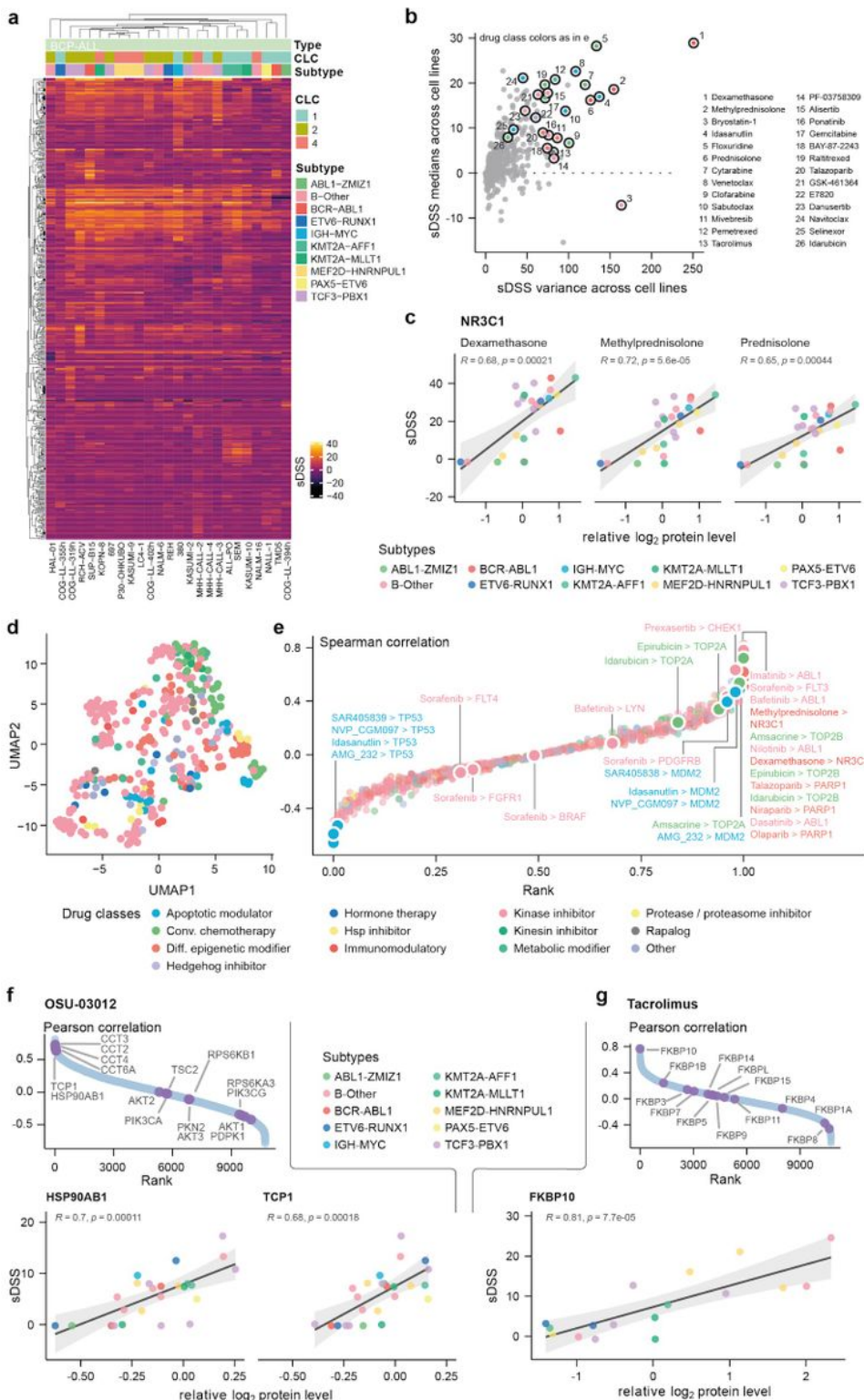
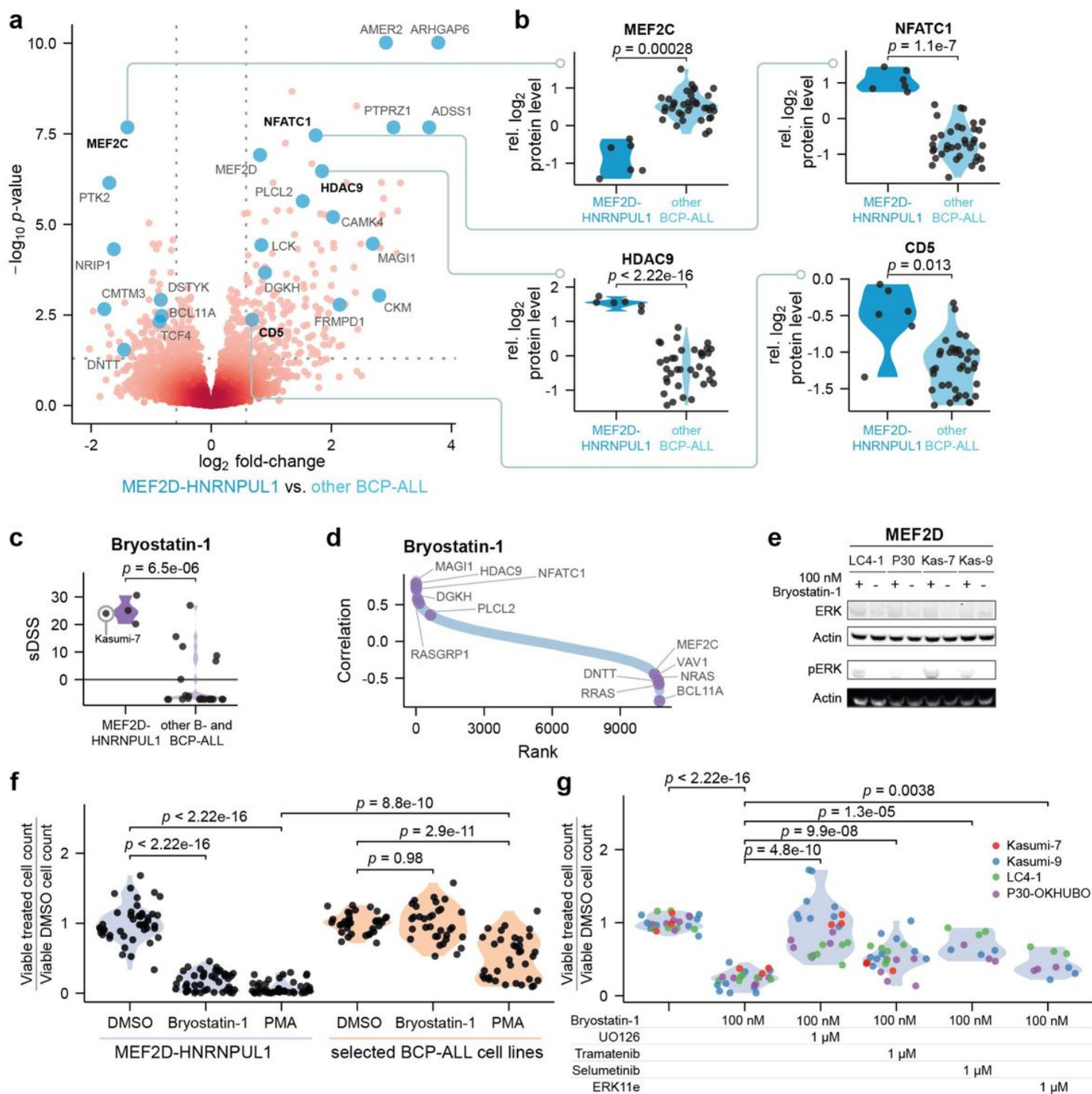


Figure 5

A. Heatmap depicting the sDSS of the 528 drugs from the DSRT across the BCP-ALL cell lines (n=25). The x-axis is the cell lines ordered by the rank from hierarchical clustering (Pearson Ward.D2) and the colors are the individual sDSS for each drug. The legends indicate the CLC and cytogenetic subtype of the cell lines. Higher sDSS indicates a more potent and effective drug. B. Scatter plot of the median versus variance of the sDSS for each drug across the tested cell lines. Selected drugs are highlighted and colored according to their drug class. C. The correlation between the sDSS for the glucocorticoids in the DSRT and relative protein level of the glucocorticoid receptor, NR3C1. The x-axis shows the relative log<sub>2</sub> protein level of the protein and the y-axis shows the sDSS scores for each respective cell line. The Pearson correlation coefficient (R) and p-values for the comparison are shown in each plot. The colors indicate the cytogenetic subtype of the cell lines. D. UMAP of the distribution of class-annotated drugs as a two-dimensional reduction. Drugs were correlated to gene symbol-based protein quantification data based on sDSS score, and Pearson correlation scores from this analysis were used to generate the UMAP. Each drug is annotated by its class. E. Scatter plot of the Pearson correlation versus rank for the sDSS of each drug and the relative log<sub>2</sub> protein level of its putative protein target. The drugs are colored by drug classes. F. Ranked OSU-03012 sDSS and protein level Pearson correlations and scatterplot of OSU-03012 sDSS and HSP90AB1 levels for each of the BCP-ALL cell lines. Pearson correlations show selected members of the TCP1 containing chaperonin complex, HSP90AB1, and PDPK1 interacting proteins from the STRING network database. R = Pearson correlation coefficient. p-values when calculated are shown within the concerned plot. G. Ranked Pearson correlations of Tacrolimus sDSS and highlighted proteins of the FKBP family and scatterplot of Tacrolimus sDSS and FKBP10 levels for each of the BCP-ALL cell lines. R = Pearson correlation coefficient. p-values when calculated are shown within the concerned plot.



**Figure 6**

A. Volcano plot of the differential protein levels in the MEF2D-HNRNPUL1 cell lines and their replicates ( $n = 6$ ) versus the rest of the BCP-ALL cell lines and replicates ( $n = 37$ ). The cut-off was set at  $p \leq 0.05$  and the fold change was set to  $\log_2(1.5)$ . B. Relative  $\log_2$  protein levels of HDAC9, CD5, NFATC1 and MEF2C in MEF2D-HNRNPUL1 cell lines compared to the remaining BCP-ALL cell lines. The p-value is indicated. C. The sDSS of Bryostatin-1 in the MEF2D-HNRNPUL1 cell lines ( $n = 4$ ) compared to the sDSS for the remaining BCP-ALL cell lines ( $n = 22$ ). The p-value is indicated above the bracket. D. Ranked Bryostatin-1

sDSS and protein level Pearson correlations with selected highlighted proteins in purple. E. Western blot of ERK1/2 and pERK1/2 (Thr202, Tyr204) of MEF2D-HNRNPUL1 fusion cell lines LC4-1, P30-OHKUBO (P30), Kasumi-7 (Kas-7), and Kasumi-9 (Kas-9) after 2-hour treatment with 100 nM Bryostatin-1. F. Viable cell quantification normalized to corresponding mean DMSO viable cell count of the four MEF2D-HNRNPUL1 fusion cell lines, treated with 100 nM of Bryostatin-1 or 25 nM PMA. Equal volume of DMSO was used as a control in triplicates, all wells were supplemented with DMSO to obtain equal volumes. BCP-ALL cell lines ALL-PO, REH, RCH-ACV, NALM6, SUP-B15, COG355, COG394, and MHH-CALL2, lacking the MEF2D-HNRNPUL1 fusion, were subjected to the same treatments. Viable cells were quantified by flow cytometry, excluding zombie aqua dyed non-viable cells. Results are merged from five independent experiments. p-values were obtained from unpaired t-tests. G. Viable cell quantification normalized to corresponding mean DMSO viable cell count of the four MEF2D-HNRNPUL1 fusion cell lines, treated with 100 nM of Bryostatin-1 alone or in combination with 1uM MEK inhibitors U0126, Trametinib, or Selumetinib. Alternatively, to block ERK directly, 1uM ERK inhibitor ERK 11e was used, and equal volume of DMSO was used as a control in triplicates. All wells were supplemented to obtain equal DMSO volumes per well. Viable cells were quantified by flow cytometry, excluding zombie aqua dyed non-viable cells. Results are merged from three independent experiments. p-values were obtained from unpaired t-tests.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarytable1Celllinepanelmeta.xlsx](#)
- [supplementarytable2DynamicrangeHiRIEFLCMSMS.xlsx](#)
- [supplementarytable3RNAseqQC.xlsx](#)
- [supplementarytable4DSRTMeta.xlsx](#)
- [supplementarytable5BvsTDEqMS.xlsx](#)
- [supplementarytable6CLC3vsCLC7GSEA.xlsx](#)
- [supplementarytable7AFF1vsMLLT1DEqMS.xlsx](#)
- [supplementarytable8DSRTsDSS.xlsx](#)
- [supplementarytable9RigosertibsDSSprotein correlationGSEA.xlsx](#)
- [supplementarytable10MEF2DRearrangedClinicaledgeR.xlsx](#)
- [supplementarytable11MEF2DHNRNPUL1DEqMS.xlsx](#)