Cancer pathway connectivity resolved by drug perturbation and RNA sequencing

Thi Huong Lan Do[1]* Caroline Lohoff[2,3,4]*, Ferris Jung[5], Sandra Kummer[1], Marcel Fabian Pohly[1,6,7], Sebastian Scheinost[8], Jennifer Huellein[3,8], Vladimir Benes[5], Wolfgang Huber[3,4,7]✉, Junyan Lu[2,3,4]✉, Thorsten Zenz[1,7]✉

[1] Department of Medical Oncology and Hematology, University Hospital Zurich and University of Zurich (UZH), Zurich, Switzerland
[2] Institute for Computational Biomedicine, Medical Faculty Heidelberg, Heidelberg University, Heidelberg, Germany
[3] Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
[4] Molecular Medicine Partnership Unit (MMPU), Heidelberg, Germany
[5] Genomic Core Facility, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
[6] Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland
[7] INTeRCePT consortium, The LOOP Zurich Medical Research Center, Zurich, Switzerland
[8] Department of Translational Medical Oncology, National Center for Tumor Diseases (NCT) Heidelberg and German Cancer Research Center (DKFZ), Heidelberg, Germany

* These authors contributed equally.
✉ Corresponding authors

Correspondence:

Thorsten Zenz, Department of Medical Oncology and Hematology, University Hospital and University of Zurich, Rämistrasse 100, CH-8091 Zurich, Switzerland. ✉ Thorsten.Zenz@usz.ch

Junyan Lu, Institute for Computational Biomedicine, Medical Faculty Heidelberg, Heidelberg University, Im Neuenheimer Feld 130, D-69120 Heidelberg, Germany. ✉ Junyan.Lu@uni-heidelberg.de

Wolfgang Huber, European Molecular Biology Laboratory, Meyerhofstraße 1, D-69117 Heidelberg, Germany. ✉ Wolfgang.Huber@embl.org

Abstract

Pathway inhibitors are a backbone of cancer treatment. The configuration of pathway dependencies varies from tumour to tumour. Better treatment for individual patients could be designed if the pathway wiring were readily measurable.

Here, we characterise the transcriptional responses of 116 lymphoma patient samples exposed to ten drug perturbations. We used factor analysis to decompose individual and shared drug effects, thereby generating a pathway connectivity map of chronic lymphocytic leukemia (CLL). The expression profiles of the major disease subgroups, defined by IGHV mutation status, became more similar to each other after BTK inhibition, consistent with B-cell receptor (BCR) signalling as a driver of their phenotypic difference. An even stronger convergence was observed with combined IRAK4 and BTK inhibition, indicating cooperation of BCR and toll-like receptor (TLR) signalling in CLL. We identified genetic aberrations (*BRAF*, *TP53*, deletion 17p, deletion 15q, trisomy 12) that modulated drug effects in CLL and constituted specific interaction patterns. IRAK4 inhibition effects depended on the presence of trisomy 12, a finding that suggests that the trisomy 12 driver event in CLL acts by gene dosage-dependent IRAK4 upregulation and amplification of the BCR/TLR cooperation.

Our results highlight the potential of systematic drug perturbation assays with transcriptome readout to map pathway interconnectivity and functionally annotate tumour drivers.

52 # Introduction

53 Chronic lymphocytic leukemia (CLL) is a common B-cell lymphoma in adults. More than 200
54 driver genes and chromosomal aberrations are known, which target a broad range of signalling
55 pathways including the B-cell receptor (BCR), DNA damage response and NOTCH pathway[1–
56 4]. A more precise understanding how this broad range of pathways cooperate and contribute
57 to the CLL phenotype should facilitate better treatments. Multi-omics profiling of CLL patient
58 samples has revealed major biological dimensions of CLL, such as the epigenetic cell-of-origin
59 fingerprint[5,6], the clinical and biological impact of IGHV mutations[7,8], autonomous BCR
60 signalling[9] and CLL proliferative drive (CLL-PD)[8,10–13]. Nonetheless, many common mutations
61 and copy number variants (CNVs), e. g. trisomy 12, have not yet been conclusively linked to a
62 function[10].

63 The ability to measure multi-omics data in large cancer cohorts has refined molecular
64 subgrouping, improved outcome prediction and treatments[12,14,15]. Recently, however, progress
65 has often focused on the discovery of rarer disease variants with the inherent downside that
66 findings only apply to a small proportion of patients[1,3]. Furthermore, clinical precision oncology
67 has largely relied on cancer genome sequencing and the identification of genetic mutations as
68 biomarkers[16,17], while other omics data have remained research tools. Among these, gene
69 expression profiling is particularly suited for the detailed characterisation of drug effects. This
70 concept has been applied in the “Connectivity Map”, which represents a landmark study
71 mapping over 80,000 genetic and chemical perturbations to gene expression signatures in 29
72 cancer cell lines[18,19], and has been widely used to identify drug effects, drug repurposing and
73 construct gene regulatory networks[18,20–22]. By using the perturbation signatures, transcriptional
74 consequences can specifically be linked to the targeted upstream cause and therefore facilitate
75 the understanding of cellular mechanisms[18].

76 In this study, we applied a set of 10 small-molecule drugs to a collection of 116 primary
77 tumours, to investigate how these drugs work, capture the wiring of targeted pathways and
78 identify modulators (i.e. mutations) of drug response.

Results

Shallow depth transcriptional profiles capture CLL disease dimensions and drug effects

In order to increase feasibility and cost-effectiveness of profiling hundreds to thousands of samples, we first compared gene expression from five CLL samples profiled with whole-transcript deep (TruSeq) and 3'-end-based library preparation with shallow depth RNA sequencing (QuantSeq[23]) (Suppl. Fig. 1a-d). Although the sequencing depth was reduced up to 30-fold (2.2-3.8 vs. 43-73 million reads per sample), QuantSeq recovered 71% of the genes detected by TruSeq with reproducible quantification (R = 0.82-0.85) (Suppl. Fig. 1e-g).

We selected a cohort of 108 chronic lymphocytic leukemia (CLL) and eight other B-cell Non-Hodgkin (B-NHL) lymphoma patient samples with available molecular profiles and clinical data (Fig. 1) and used QuantSeq to generate 1,086 (1,010 for CLL, 76 for B-NHL; 92% successfully prepared libraries) gene expression profiles after 48h of *ex-vivo* treatment with inhibitors used clinically (BTK inhibitor ibrutinib, PI3K inhibitor duvelisib, MEK inhibitor trametinib, mTOR inhibitor everolimus, XPO1 inhibitor selinexor), compounds targeting key signalling pathways (IRAK4 inhibitor compound 26, AKT inhibitor MK2206, MDM2 inhibitor and p53 activator nutlin-3a, BET inhibitor IBET762), and one combinatorial drug treatment (BTK and IRAK4 inhibition) (Fig. 1a, Suppl. Fig. 2a,b, Suppl. Table 1). IGHV mutational status, trisomy 12 and *TP53* mutation influenced viability of cells after drug treatment consistent with our prior findings[11] (Suppl. Fig. 2c-e).

We began our analysis by performing principal component analysis (PCA) to investigate which gene expression programmes, as characterised by factors (≙ principal components), were the main contributors of transcriptional variation in our data set. PCA of the DMSO-treated control samples showed that transcriptional variation between patients was predominantly driven by disease type, IGHV, methylation group[5,6] and trisomy 12 status (Suppl. Fig. 2f). This result was highly similar to results from studies using conventional RNA sequencing[13] and shows that we were able to capture the main biological disease dimensions after *ex-vivo* culture and markedly reduced sequencing depth (Suppl. Fig. 2g,h). PCA of all samples including drug treatments

107 revealed that the first component (PC1) captured a contribution of sample viability to the
108 transcriptional variation independent of perturbation (Fig. 2a,b, Suppl. Fig. 2i, Suppl. Fig. 3a-
109 c). This viability signature was enriched in pathways linked to a general upregulation of
110 hypoxia, p53 pathway, TNFα signalling via NFκB, and downregulation of MYC targets and
111 OXPHOS preferentially in drug-perturbed samples (Suppl. Fig. 3d). Variation driven by IGHV
112 status/methylation programme and trisomy 12 were reflected by PC2 and PC3 (Fig. 2c, Suppl.
113 Fig. 2i,j). We also used *t*-Distributed Stochastic Neighbour Embedding (*t*-SNE) to visualise
114 samples' gene expression similarities and found patterns that were dominated by inter-patient
115 differences and inherent molecular disease features (Fig. 2d-e). After subtracting the patient-
116 specific baseline signatures, the embeddings captured drug specific-effects (Fig. 2f, Suppl.
117 Fig. 2k).

118 Individual drug effects

119 To understand the effect of individual drugs, we extracted drug-specific transcriptional
120 signatures by performing differential gene expression analysis between drug- and control-
121 treated samples (Fig. 2g). We first assessed the response to nutlin-3a, a MDM2 inhibitor and
122 activator of the transcription factor p53[24]. We identified the p53 activation signature in primary
123 CLL cells, which consisted of a strong upregulation of well-known transcriptional p53 targets
124 and pathways including *BAX*, *MDM2* and *CDKN1A*/p21 (Fig. 2h-i). This response was blunted
125 in cases with *TP53* mutation and/or deletion in chromosome 17p (Fig. 2j).

126 Inhibition of PI3K and BTK elicited the strongest responses, as assessed by the number of
127 differentially expressed genes, in line with the role of BCR signalling as a key driver of CLL[7–9]
128 (Fig. 2g). BTK inhibition led to the downregulation of genes associated with B-cell activation
129 and function (*FCRL5*, *FILIP1L*, *MS4A1*/CD20, *CCL3*/*4*) and phosphatases (*PTPN6*, *DUSP4/5*)
130 (Fig. 2k,l). We also found downregulation of key B-cell transcription factors (*PAX5*, *IRF4*,
131 *IKZF3*) and members of the NFκB pathway (*REL*, *RELB, NFKB1),* suggesting a role of active
132 BCR signalling in the maintenance of the regulatory B-cell identity programme. This signature

133 overlapped with reported transcriptional signatures after BCR stimulation and *in-vivo* BTK
134 inhibitor treatment[25–27] (Suppl. Fig. 4a).

135 PI3K inhibition led to the downregulation of gene sets enriched in immune cell response,
136 cytokine signalling, interferon (IFN) response and BCR signalling, similar to the effects upon
137 BTK inhibition but more pronounced (Fig. 2l,m, Suppl. Fig. 4b,c). We investigated the
138 similarities and differences between BTK and PI3K inhibition by directly comparing the effects
139 of the two inhibitors against each other (Fig. 2m, Suppl. Fig. 4d). Genes only modulated by the
140 BTK inhibitor included components of B-cell development and function, endocytosis and
141 autophagy, and the IFN response pathway. PI3K inhibition exclusively downregulated genes
142 associated with IFN and cytokine signalling (Fig. 2n, Suppl. Fig. 4d). To understand whether
143 these differences were driven by a common upstream mediator, we inferred the transcription
144 factor activities from the differentially expressed genes[28] in which transcription factor activity of
145 STAT family members and in particular gene expression of *STAT1* was downregulated by PI3K
146 but not BTK inhibition (Fig. 2o, Suppl. Fig. 4e-f).

147 Inhibition of AKT modulated gene sets involved in antigen processing and presentation, IFN
148 signalling and suppression of MHC class II molecules similar to the PI3K inhibitor signature
149 (Suppl. Fig. 5a,b). Differentially expressed genes between PI3K and AKT inhibitor treatment
150 mostly consisted of BTK-dependent targets that were only suppressed upon PI3K inhibition
151 (Suppl. Fig. 5c). The mTOR inhibitor signature showed overlap with the AKT and PI3K inhibitor
152 signatures in the context of antigen processing and adaptive immune system (Suppl. Fig.
153 5b,d), and specifically included suppression of gene sets associated with lysosome,
154 cholesterol homeostasis and mTORC1 signalling (Suppl. Fig. 5b,g). The predominant
155 mechanisms of action (MoA) induced by the MEK inhibitor were cytokine signalling and IFN
156 response (Suppl. Fig. 5b,e-g).

157 Bromodomain and extra-terminal motif (BET) proteins are readers of acetylated histone marks
158 and regulate gene expression[29]. Inhibition of BET proteins led to upregulation of genes
159 associated with translation, ribosome and metabolism of RNA (Suppl. Fig. 5h,i). Selinexor is a

160 small-molecule that prevents the translocation of proteins from the nucleus to the cytoplasm
161 by inhibiting the nuclear exporter XPO1[30]. *XPO1* was among the top upregulated genes
162 together with solute carriers, nucleotide- and protein-binding adapters, and members of the
163 zinc finger protein family (Suppl. Fig. 5j,k). Gene expression of p53 was also increased after
164 XPO1 inhibitor treatment without induction of the transcriptional p53 activation signature
165 (Suppl. Fig. 5l). Inhibition of toll-like receptor (TLR) signalling with the IRAK4 inhibitor did not
166 lead to significant differential gene expression (Fig. 2g).

167 Connectivity of signalling and perturbation effects in CLL

168 The use of multiple drugs allowed us to understand how important signalling axes in CLL are
169 connected. For this, we constructed a pathway connectivity network as a bipartite graph in
170 which treatment and gene nodes are connected by edges when the perturbation led to
171 differential expression of that gene. We found that the BTK and PI3K inhibitors were strongly
172 linked through their common BCR-inhibiting MoA. The AKT inhibitor formed close connections
173 to the PI3K and mTOR inhibitors along the PI3K/AKT/mTOR pathway (Fig. 3a). We also
174 illustrated these relationships by comparing the perturbations directly against each other (Fig.
175 3b). Because drug perturbations typically do not affect a single gene in isolation, but
176 simultaneously modulate the expression of multiple related genes, we used a factorisation
177 technique, namely guided sparse factor analysis[31] (GSFA) to decompose the full set of drug-
178 induced differentially expressed genes into individual latent factors of co-regulated genes (Fig.
179 3c-e, Suppl. Fig. 6). We identified drug-specific effects upon inhibition of MDM2 (factor 18),
180 XPO1 (factor 12), BET (factor 16) and MEK (factor 14), but also found modules affected by
181 multiple drugs at variable intensities (Fig. 3c,d, Suppl. Fig. 6a,b). These shared axes of
182 pathway heterogeneity distinguished canonical BCR signalling (factor 8; BTK/PI3K inhibition)
183 from antigen processing (factor 17; PI3K/AKT inhibition), and the lysosome pathway (factor
184 20; mTOR/MEK inhibition) (Fig. 3c,e, Suppl. Fig. 6b,c). We also identified modules regulated
185 in opposing directions, which could indicate compensatory pathway activation, and link them
186 to IFN (factor 14; MEK/XPO1 vs. mTOR/BTK inhibition) and lysosomal signalling (factor 20;

187 MEK/mTOR vs. PI3K/BTK inhibition) (Fig. 3c-e). Next, we used linear regression to identify the
188 impact of the most prevalent disease aberrations and CNVs on individual patient factor
189 contributions. Mutation or deletion of *TP53* was negatively associated, while trisomy 12 was
190 positively associated with the response to the MDM2 inhibitor captured in factor 18 (Suppl. Fig.
191 6d). BTK and MEK inhibitor responses reflected in the factors 17 and 14, respectively, were
192 positively associated with unmutated IGHV and trisomy 12 (Suppl. Fig. 6e-f).

193 **Pathway dependency of known CLL drivers**

194 We aimed to understand how signalling activities contribute to signatures associated with
195 major CLL subtypes that affect clinical outcome. The proliferative drive of CLL (CLL-PD), linked
196 to prognosis and increased mTOR-MYC-OXPHOS activity, was significantly reduced upon
197 AKT, BET and mTOR inhibition, demonstrating a direct dependency of pro-proliferative
198 capacities of CLL cells on these pathways[12] (Fig. 3f). The IGHV status is the strongest
199 modulator of the transcriptional variation in CLL and a key determinant of outcome (Fig. 2c,
200 Suppl. Fig. 2f-j). Some well-known markers of IGHV status, i. e. *ZAP70* and *CD38*, were stable
201 after drug treatment, while other signature genes, i. e. *CCL3* and *CD72*, were dependent on
202 BCR activity, suggesting that active signalling maintains part of the IGHV signature (Suppl.
203 Fig. 7a,b). To test the impact of the drugs on the overall signature, we quantified the
204 differentially expressed genes between patients with unmutated and mutated IGHV with each
205 treatment and found that the signature size was moderately reduced by PI3K, mTOR and BTK
206 inhibition (Fig. 3g). This partial reversal was driven by downregulation of IGHV signature genes
207 in unmutated patients which enriched in the cholesterol and glucose metabolism, allograft
208 rejection, adipogenesis, mTORC1 signalling and IFN signalling gene sets (Suppl. Fig. 7c,d).
209 When the BTK and IRAK4 inhibitors were combined, we observed the most pronounced
210 collapse of the IGHV signature, with an enrichment of genes linked to ribosomes and
211 translation, in line with a more pronounced effect on these downstream consequences when
212 the BCR and TLR were simultaneously perturbed[32] (Fig. 3g, Suppl. Fig. 7c,e). To understand
213 to what extent BCR and TLR signalling cooperate in CLL, we assessed how the sum of the

214 individual BTK and IRAK4 inhibitor effects - with the expectation that there is no interaction -
215 compared to the observed effect from the BTK/IRAK4 inhibitor combination. We found that
216 genes affected by BTK, IRAK4 or BTK/IRAK4 inhibition showed stronger effects than what
217 would be expected from the sum of the individual effects (Suppl. Fig. 7f). Cooperation of TLR
218 and BCR signalling has been described in other B-NHL, where it is linked to activating gene
219 mutations targeting the pathway[33,34]. Our finding suggests a cooperation between TLR and
220 BCR signalling in CLL even in the absence of such a driver mutation.

221 Trisomy 12 occurs in 15%–20% of CLL patients and is considered an early event[26]. It is
222 associated with disease progression and a higher incidence of aggressive transformation, but
223 its molecular function is not well understood[35,36]. We repeated the same analysis with the
224 trisomy 12 signature. Trisomy 12 signature genes showed partial dependencies on BTK, PI3K
225 and mTOR (Fig. 3h, Suppl. Fig. 7g-i). Differential expression of signature genes encoded on
226 chromosome 12 was largely stable and not affected by disruption of signalling activities, in line
227 with a gene dosage-driven “passenger” signature with no driver role (Fig. 3h). In contrast, we
228 found more genes to discriminate trisomy 12 after MEK, BTK/IRAK4 and MDM2 inhibitor
229 treatment. These findings suggest a modulation of these pathways by trisomy 12 (Fig. 3h,
230 Suppl. Fig. 7j).

231 Disease drivers influence pathway connectivity in CLL

232 Epistasis describes a phenomenon of genetic interactions in which the phenotype of a mutation
233 is modified by the presence or absence of mutations in one or more genes in a non-additive
234 manner[37]. In this study, we adopted this concept and systematically tested for drug-gene
235 interactions in linear models and identified modifications in the transcriptional phenotype of a
236 chemically perturbed gene target in the presence of unmutated IGHV and the common disease
237 drivers trisomy 12, *BRAF*, deletion 17p, deletion 15q and *TP53* (Fig. 4a, Suppl. Fig. 8a). We
238 then grouped the observed phenotypes of the epistatic interactions based on the direction of
239 which perturbations mask, modify, or otherwise influence the expression of genes. We
240 identified genes upregulated by a given aberration which were downregulated in the presence

241 of both perturbation and driver to levels below baseline and vice-versa (inversion; n = 247).
242 Suppression of the driver phenotype in the presence of a perturbation was observed for 112
243 genes. We also observed synergy (n = 94), where drug effects are amplified and only
244 discernible in the presence of both aberration and drug treatment. Suppression of drug effects
245 in the presence of an aberration was found for 60 genes (Fig. 4b-c, Suppl. Fig. 8b-d). We used
246 these categories to understand the functional implications of disease drivers based on known
247 drug targets or mechanisms. A clear example for loss of function (LOF) in CLL was observed
248 with *TP53* mutations or deletion 17p, which showed impaired transcriptional activation of p53
249 targets, e. g. *MDM2*, *CDKN1A*, in response to MDM2 inhibitor treatment (≙ suppression of
250 drug effect) (Fig. 4d-f, Suppl. Data Fig. 8d). Consistent with the activation of the MAPK
251 pathway, *BRAF* mutations showed most interactions with MEK inhibition (synergy and driver
252 suppression type (Fig. 4d, Suppl. Fig. 8d)). We also identified so far unknown interactions
253 linking mTOR inhibition to deletion 15q (Fig. 4d, Suppl. Fig. 8d). For unmutated IGHV, we
254 observed interactions with mTOR predominantly of the driver and drug suppression type (Fig.
255 4d, Suppl. Fig. 8d). We also found important therapeutic targets in lymphoma, e. g. *CD19* and
256 *MS4A1*/CD20, in the synergy group. These findings suggest that expression of these targets
257 are dependent on a certain pathway in molecular subgroups which could serve as a basis for
258 designing new treatment strategies (Fig. 4g,h).

259 Most drug-gene interactions were observed for trisomy 12 (n = 314) characterised by inversion
260 and synergy with BTK, MEK, IRAK4 and combined BTK/IRAK4 inhibition (Fig. 4d,i-l, Suppl.
261 Fig. 8d). The effect of IRAK4 inhibition in the presence of trisomy 12 is remarkable as the
262 inhibitor's effect was undetectable when the complete cohort was analysed (Fig. 2g). We
263 identified most drug-gene interactions were modulated by trisomy 12 with IRAK4 (n = 114),
264 combined BTK/IRAK4 (n = 83) and BTK inhibition (n = 39) providing further evidence for
265 crosstalk and cooperative effects between BCR and TLR signalling in trisomy 12 (Fig. 4d,i-l).

266 To demonstrate the role of IRAK4 as a mediator of TLR signalling in CLL, we treated patient
267 samples with either IRAK4 inhibitor alone or in combination with the TLR9 agonist CpG. TLR9

268 stimulation led to increased viability in line with a pro-tumourigenic role of TLR signalling, which
269 was abrogated by IRAK4 inhibition (Suppl. Fig. 9a). We searched for focal gains in the CNV
270 profiles of our CLL cohort (n = 366) which may allow restricting the number of candidate genes
271 targeted by trisomy 12, and identified a CLL case with a focal gain between chr12:42,538,673
272 and chr12: 44,229,999. This gain included nine protein-coding genes including *IRAK4* (Fig. 5a-
273 c, Suppl. Fig. 9b). We clustered gene expression of CLL samples (n = 210) based on the
274 trisomy 12 baseline signature and restricted the analysis to genes not encoded on
275 chromosome 12 (642 out of 917 total trisomy 12 signature genes) to prevent clustering due to
276 gene dosage. The sample with the focal gain of *IRAK4* clustered with the trisomy 12 samples
277 demonstrating the phenotype resemblance (Fig. 5c).

278 Our findings suggest that trisomy 12 confers increased TLR signalling capacity driven by the
279 increased *IRAK4* expression from gene dosage and cooperates with BCR signalling.

280 Discussion

281 We performed a functional perturbation screening directly in cancer cells at scale to capture
282 drug-induced global gene expression changes with low-depth RNA sequencing. Prior studies
283 investigating perturbation-induced gene expression changes have exploited single-cell
284 technology (few patients, many cells e. g. MIX-seq[38], RASL-seq[39] or DRUG-seq[40]) and focused
285 on establishment of technical frameworks, but less on exploiting the signatures to dissect
286 disease biology and cellular processes. The Connectivity Map focused on the differences and
287 signatures induced by a large number of drugs in fewer cell lines[18,19]. By design, it therefore
288 has unique resolution on the drug effects and contributes to the understanding how drugs work.

289 Our approach complements the Connectivity Map by leveraging many primary tumours from
290 an individual disease, and resolving the effects of a limited number of clinically used drugs on
291 a broad and representative sampling of molecularly diverse tumours without immortalisation
292 artefacts. We exploit the drug effects on transcriptomes to resolve disease mechanisms as
293 well as inter-patient heterogeneity. This design allowed us to measure downstream effects of

294 signalling events even if effects are subtle. Our findings suggest that insights and hypotheses
295 from gene expression studies on primary tumours may be vastly increased by the analysis of
296 such perturbed samples as the ability to specifically target molecules and pathways grows.
297 Multiple clinical inhibitors targeting PI3K, BTK, mTOR and MEK, were used in this study as
298 optimal combination and clinical assignment to patients is a challenge, which cannot be easily
299 addressed in trials[41].

300 We used interaction testing to study the effect of gene mutations on the perturbed gene
301 expression, e.g. by LOF mutations (*TP53*) or GOF (trisomy 12, *BRAF*). A particularly notable
302 example is trisomy 12, present in approximately 20% of CLL patients. We identified IRAK4 as
303 a driver in CLL with trisomy 12, highlighting how perturbation transcriptomics can uncover
304 actionable targets within genetically defined subgroups. In line with our proposed mechanism
305 where IRAK4 is a key target, trisomy 12 has been found to be strongly associated with
306 response to TLR stimuli such as resiquimod (TLR7/8) and CpG-ODN (TLR9)[42]. Reid et al.
307 studied genes targeted by trisomy 12 in a human pluripotent cell line model and identified
308 IRAK4 inhibition to be potent against human trisomy 12-harboring CLL cells in patient-derived
309 xenografts[43].

310 Our experiments enable the deconvolution of drug effects by factor analysis, and facilitate the
311 transition to patient-specific pathway assembly read-outs by drug perturbation and global
312 expression analysis (Fig. 6). The identification of compensatory pathways (factor 14, 20) may
313 provide a basis for optimised combination treatment. While prior drug response perturbation
314 studies have hinted at some of these findings and structure[11], the use of a large set of tumours
315 and high-dimensional readout of gene expression gives a more comprehensive and systematic
316 picture of the connectivity of signalling pathways in CLL, where effects of inhibitors on
317 individual pathways can be deconvoluted and linked to other drugs. On the route to improved
318 precision medicine for individual cancer patients, the workflows established in our study
319 provide a framework to measure pathway connectivity of primary cancer cells using clinically
320 relevant drugs and adapting the treatment along the lines of measured pathway activity which

321 can be generated by the projection of the individual patients response compared to a larger
322 reference set of measurements such as the one presented here. While the precise strategy for
323 optimal drug combinations will need to be demonstrated, our data provides a roadmap for such
324 experimentation.

325 Drug-perturbed transcriptome profiling measures connectivity of cancer pathway
326 dependencies and functionally annotates mutations. As more specific drugs become available,
327 tumour pathway wiring could be fully resolved and improve precision oncology, moving closer
328 to treatments guided by measured pathway activity rather than inferred potential.

329 Material and Methods

330 Patient samples and study approval

331 Patient material was acquired in accordance with the Declaration of Helsinki and approval from
332 the local ethics committees (S-206/2011; S-356/2013, 2009-0062; 2019/01744), after written
333 informed consent. Detailed patient information is listed in Suppl. Table 1. DNA mutations,
334 CNVs, DNA methylation and gene expression were performed as described previously[11].

335 Isolation and cryoconservation of MNCs from whole blood

336 Mononuclear cells from peripheral blood were isolated using Ficoll-Paque (Cytiva) density
337 gradient centrifugation, cryopreserved in freezing medium (72.5% RPMI-1640 (Gibco), 20%
338 fetal bovine serum (FBS, Gibco), 7.5% DMSO (Sigma-Adrich)) and stored at -180°C in liquid
339 nitrogen.

340 Compounds

341 Small-molecules (except compound 26 (Calbiochem)) were purchased from Selleckchem and
342 reconstituted in DMSO at 10 mM.

343 In a pilot, we treated four CLL samples with concentrations based on results from multiple prior
344 drug screens[11,12] with average toxicity below 20% (Suppl. Methods Fig. 1a,b) and confirmed
345 drug-specific gene expression changes in line with the drugs' mechanisms of action (Suppl.

346 Methods Fig. 1c-h). For the screen, stocks were diluted to 200-fold of the final assay
347 concentrations with DMSO as “ready-to-use” aliquots and stored at -20°C.

348 *Ex-vivo* drug treatment and cell viability assessment

349 Primary samples were thawed and incubated in RPMI-1640 supplemented with 10% heat-
350 inactivated human serum from male AB clotted whole blood (Sigma-Aldrich), 1% 10,000 U/mL
351 penicillin/10,000 µg/mL streptomycin (P/S, Gibco) and 2% 200 mM L-glutamine (Gln, Gibco)
352 (RPMI-HS) on a roller mixer for 3h at room temperature to allow recovery from the thawing
353 process and support the removal of DMSO from the freezing medium. The medium was
354 replaced and 5 x 10^6 cells were treated with each drug (Suppl. Methods Fig. 1c) and cultured
355 at 37°C and 5% CO_2 for 48h. Cells were harvested, washed twice with cold phosphate-buffered
356 saline (Gibco), lysed in 600 µL TRI Reagent (Zymo Research) and stored at -80°C until RNA
357 extraction.

358 Sample viabilities of each sample (before and after perturbation) were assessed on the flow
359 cytometer LSR II Fortessa with FACSDiva v8.0.1 software (BD Biosciences) and analysed with
360 FlowJo v10.10.0 (FlowJo LLC) (Suppl. Method Fig. 2).

361 TLR9 stimulation and ATP-based viability assessment

362 For the 5x concentrated working dilution, a 5-point 5-fold serial dilution of IRAK4i was prepared
363 with a starting concentration of 50 µM (≙ 10 µM–0.016 nM final assay concentration). For the
364 negative control, 5 µL of 0.5% DMSO working dilution was pre-plated (≙ 0.1% final
365 concentration).

366 MNCs from 20 CLL patients were thawed and incubated for 3h on the roller mixer prior to
367 seeding as described above and reconstituted to 1 x 10^6/mL (1.25X) cell concentration. In
368 unstimulated conditions, 20 µL (≙ 20,000 cells) were added to each well containing the pre-
369 plated DMSO or IRAK4i. For the CpG stimulation, cells were supplemented with 1.25 µg/mL
370 of the TLR9 agonist ODN2006 (InvivoGen) before seeding to reach a final assay concentration
371 of 1 µg/mL. Cells were incubated at 37°C and 5% CO_2 for 48h. For the ATP-based viability
372 assessment, plates were spun down briefly before adding 7 µL CellTiter-Glo (Promega) to

373 each well. The reaction was incubated for 20 min at RT and luminescence measured on a
374 BioTek Synergy LX multimode microplate reader (Agilent). To obtain normalised viabilities,
375 intensity values were divided by the median of all DMSO values for each condition and
376 concentration for each patient.

377 RNA extraction

378 Frozen lysates were thawed at 37°C for 3 min. RNA was precipitated by thoroughly mixing 100
379 µL of 100% ethanol (Sigma-Aldrich) with the lysate. RNA was extracted with the Direct-zol
380 RNA MiniPrep kit (Zymo Research) including on-column DNase treatment (30U per sample)
381 for 15 min at room temperature according to the manufacturer's protocol and eluted in 25 µL
382 nuclease-free water (QIAGEN). The concentration was determined with Quant-it RiboGreen
383 RNA assay (ThermoFisher Scientific) on a BioTek Synergy LX multimode microplate reader.
384 RNA was stored at -80°C in 5 µL nuclease-free water until use for library preparation and RNA
385 sequencing.

386 Library preparation and 3’ RNA sequencing

387 For the pilot screen (4 patients) 200 ng of total RNA were used for cDNA library generation
388 with the QuantSeq 3′ mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen). Libraries from
389 42 samples were pooled and sequenced on the NextSeq500 (Illumina) using 75 bp and single-
390 end reads yielding 3.9 to 13.5 million reads. For the comparison between the library
391 preparation methods, baseline profiles of 5 patients generated by whole-transcript deep
392 sequencing (TruSeq, Illumina)[13], the same RNA material was sequenced in the pilot screen.
393 For the remaining 112 patients, cDNA libraries were prepared from 100 ng total RNA with the
394 QuantSeq 3′ mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen). Up to 48 libraries (22-
395 48) were pooled and sequenced on the HiSeq 4000 (Illumina) using 50 bp and single-end
396 reads yielding 2.3 to 21.7 million reads.

397 Processing of the 3'-RNA sequencing data

398 The same libraries from the first sequencing run were sequenced twice and the resulting reads
399 were pooled. The demultiplexed RNA sequencing FASTQ files were mapped to the ENSEMBL

400 human reference genome release 104 (Homo sapiens GRCh38.104) using STAR (version
401 2.7.9a) with default parameters. The aligned reads were summarised into per-gene counts
402 using htseq-count (version 0.9.1) with default parameters in union mode, counting only those
403 reads unambiguously mapped to a single gene. The resulting gene count matrix was imported
404 into R (version 4.5.1) for further processing. Given the shallow sequencing depth, the gene
405 count matrix exhibited higher sparsity compared to traditional bulk RNA-seq data. This sparsity
406 posed challenges for accurately estimating between-sample normalisation factors using
407 methods designed for deep RNA sequencing, such as DESeq2[44]. To address this,
408 scTransform[45], a method that applies regularised negative binomial regression to sparse gene
409 count matrices from single-cell sequencing data, was used to estimate the normalisation
410 factors. These factors were then supplied to DESeq2 (version 1.44) for variance-stabilising
411 transformation and differential expression analysis. For all the downstream analyses, only
412 protein-coding genes with minimum 10 counts in at least one sample were retained.

413 The baseline profiles were combined with profiles generated by 3'-end shallow depth
414 sequencing (QuantSeq, Lexogen) and normalised for library size before comparative analysis.

415 Data analysis

416 Statistics

417 Correlation trends were indicated by Pearson correlation (*R*) and linear regression lines with
418 95% confidence intervals. Density distributions of single values were visualised with sina plots
419 and summarised with Tukey-style box and whiskers plots. Statistical differences between
420 groups were computed with unpaired two-tailed Student's *t*-test when groups contained
421 samples from different patients, e. g. mutant vs. wild-type. Paired two-tailed Student's *t*-tests
422 were used when conditions of the same patients were compared. Nominal *P*-values were
423 corrected for multiple testing using the Benjamini-Hochberg procedure and indicated as false-
424 discovery rate (*FDR*).

425 **Dimensionality reduction and clustering analysis**

426 Dimensionality reduction including principal component analysis (PCA), *t*-distributed
427 Stochastic Neighbor Embedding (*t*-SNE) and hierarchical clustering were conducted on gene
428 counts after variance-stabilising transformation (*vst* function from DESeq2 package). PCA
429 (*prcomp* function from base R) and *t*-SNE (*tsne* function from M3C package[46], version 1.26)
430 were performed on the 1,000 most variable genes, unless indicated. Hierarchical clustering
431 was carried out using the Ward.D2 method based on Euclidean distances computed from the
432 centered and scaled expression values of genes selected from differential expression
433 analyses. The selection criteria are detailed in the figure legends. The ComplexHeatmap
434 package[47] (version 2.20) was used for visualisation. To analyse expression patterns after drug
435 treatment, patient-specific variations were adjusted in the gene expression matrix before
436 dimension reduction (*removeBatchEffect* function from limma package[48], version 3.60.6). The
437 application of this adjustment is noted in the respective figure legends.

438 **Differential expression analysis**

439 All differential expression analyses were conducted on the scTransform-adjusted gene count
440 matrix using the R package DESeq2[44] (versions 1.44.0 and 1.48.2).

441 To compute genes differentially expressed between drug- and DMSO-treated samples, patient
442 IDs were included as a covariate in the design formula to account for patient-specific variation.
443 Log2-fold changes were shrunken with the *lfcShrink* function using the apeglm[49] method.
444 Genes were filtered by significance (*FDR* < 0.5%) and effect size (|fold change| > 1.05).

445 p53 expression scores were downloaded from the Target Gene Regulation Database 2.0[50].

446 The CLL-PD score was computed from the Z-score gene count matrix using *CLLPDestimate*
447 from the mofaCLL package[12] (version 0.0.9).

448 The transcriptional signature of a disease dimension was defined as the set of differentially
449 expressed genes between molecular subgroups individually for each perturbation. To account
450 for the confounding effects of IGHV and trisomy 12 status, these variables were included in
451 the design formula. To compare the number of signature genes between perturbations, the

452 90% confidence interval was estimated using bootstrapping (1,000 times, sampling 90% of
453 patients while maintaining the ratio between wild-type and mutated/CNV patients).

454 **Drug-genotype interaction testing and epistasis clusters**

455 The interactions between genetic aberrations and drug effects were modeled by linear
456 interaction testing with DESeq2 using a design formula that accounts for aberration, patient ID
457 and treatment condition. Drug-gene interaction genes were filtered by *FDR* (< 10%). Median
458 and scaled expression was calculated for each gene and group (DMSO/wild-type, drug/wild-
459 type, DMSO/mut, drug/mut) and normalised against DMSO/wild-type as the reference.
460 Interaction types were grouped using soft Fuzzy C-Means Clustering (FCM) with the *cmeans*
461 function (centers = 20, iter.max = 1000, dist = “euclidian”, method = “cmeans”, m = 1.4) from
462 the e1071 package[51] (version 1.7.16). FCM was repeated three times for genes with probability
463 scores below 0.85 until every gene was assigned to a cluster with a probability score of at least
464 0.85, resulting in 60 pre-clusters. Pre-clusters with similar patterns were summarised to 20
465 final clusters using harder FCM (m = 1.1) and manual curation.

466 **Pathway connectivity networks**

467 Pathway connectivity networks of significant DEGs are shown as bipartite graphs in which
468 perturbation and DEGs denote the disjoint and independent sets of nodes and are connected
469 by the edges. The edges are weighted by the significance level of the drug effects. For the
470 connectivity network of interactions, nodes were filtered for aberrations with more than 5
471 mutated/CNV cases. The edges are weighted by the significance level of the interaction. Node
472 coordinates were computed using the network package[52] (version 1.19.0).

473 **Guided sparse factor analysis**

474 Guided Sparse Factor Analysis (GSFA) was conducted using the GSFA package (version
475 0.2.8)[31]. The gene count matrix was pre-processed by accounting for the patient-specific
476 variation and selecting the 5,000 genes with the largest deviance after deviance residual
477 transformation[53]. DMSO-treated samples were indicated as negative controls, and a
478 "mixture_normal" distribution was used as the sparse prior. Singular value decomposition was

479 employed to initialise the factor values, followed by 3,000 iterations of Gibbs sampling to train
480 the model. Posterior means were calculated using the last 1,000 iterations. All other
481 parameters were used with default settings.

482 The model with 20 factors was chosen for further analysis (Suppl. Method Fig. 3). Perturbations
483 and genes with association effect sizes above 0.25 and posterior inclusion probabilities (*PIP*)
484 above 0.95 were considered to have a significant effect on a particular factor.
485 Multiple linear regression was conducted with Lasso penalisation, leave-one-out cross-
486 validation, mean squared error loss, 200 iterations, mutations/CNVs as predictors and
487 centered and scaled individual drug effects as response vectors using the glmnet package[54]
488 (version 4.1-8).

489 **Pathway enrichment analysis**

490 Parametric analysis of gene set enrichment[55] was performed on the DEGs (ranked by *t*-
491 statistics) or genes associated with GSFA factors (ranked by factor loading). Human gene set
492 collections (version 2024.1) were downloaded from MSigDB (https://www.gsea-
493 msigdb.org/gsea/msigdb/human/collections.jsp). The decoupleR package (version 2.14.0)
494 was used to infer the impact on transcription factor activity by each drug treatment[28]. The filter
495 criteria for model input and results are indicated in the respective figure legends.

496 ## Data availability

497 All RNA sequencing data generated in this study will be available at European Genome-
498 Phenome Archive (EGA) upon publication. The sequencing data generated in our previous
499 studies and used in the current study, including whole-exome/whole-genome sequencing,
500 targeted-sequencing, DNA methylation profiling and RNA sequencing of baseline patient
501 samples, are available at EGA under the accession code EGAS00001001746. Processed
502 data, including RNA sequencing data, drug sensitivity profiling and sample metadata are
503 available in our project repository at GitHub (https://github.com/THLDo/CLL_Connectivity).
504 The RNA drug perturbation dataset can also be explored interactively via the Shiny app
505 available upon publication.

506 **Code availability**

507 The computational codes for reproducing all figures and results reported in this article will be
508 provided in our project repository at GitHub (https://github.com/THLDo/CLL_Connectivity)
509 under the GNU General Public License v3.0 upon publication.

510 References

511 1. Knisbacher, B. A. *et al.* Molecular map of chronic lymphocytic leukemia and its impact on
512 outcome. *Nat. Genet.* **54**, 1664–1674 (2022).
513 2. Nadeu, F., Diaz-Navarro, A., Delgado, J., Puente, X. S. & Campo, E. Genomic and
514 Epigenomic Alterations in Chronic Lymphocytic Leukemia. *Annu. Rev. Pathol. Mech. Dis.*
515 **15**, 149–177 (2020).
516 3. Robbe, P. *et al.* Whole-genome sequencing of chronic lymphocytic leukemia identifies
517 subgroups with distinct biological and clinical features. *Nat. Genet.* **54**, 1675–1689
518 (2022).
519 4. Burger, J. A. Treatment of Chronic Lymphocytic Leukemia. *N. Engl. J. Med.* **383**, 460–
520 473 (2020).
521 5. Oakes, C. C. *et al.* DNA methylation dynamics during B cell maturation underlie a
522 continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat. Genet.* **48**, 253–
523 264 (2016).
524 6. Kulis, M. & Esteller, M. DNA Methylation and Cancer. in *Advances in Genetics* vol. 70
525 27–56 (Elsevier, 2010).
526 7. Damle, R. N. *et al.* Ig V Gene Mutation Status and CD38 Expression As Novel Prognostic
527 Indicators in Chronic Lymphocytic Leukemia. *Blood* **94**, 1840–1847 (1999).
528 8. Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G. & Stevenson, F. K. Unmutated Ig
529 V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia.
530 *Blood* **94**, 1848–1854 (1999).
531 9. Minden, M. D. *et al.* Chronic lymphocytic leukaemia is driven by antigen-independent
532 cell-autonomous signalling. *Nature* **489**, 309–312 (2012).

533 10. Abruzzo, L. V. *et al.* Trisomy 12 chronic lymphocytic leukemia expresses a unique set of
534 activated and targetable pathways. *Haematologica* **103**, 2069–2078 (2018).
535 11. Dietrich, S. *et al.* Drug-perturbation-based stratification of blood cancer. *J. Clin. Invest.*
536 **128**, 427–445 (2017).
537 12. Lu, J. *et al.* Multi-omics reveals clinically relevant proliferative drive associated with
538 mTOR-MYC-OXPHOS activity in chronic lymphocytic leukemia. *Nat. Cancer* **2**, 853–864
539 (2021).
540 13. Lütge, A. *et al.* Subgroup-specific gene expression profiles and mixed epistasis in chronic
541 lymphocytic leukemia. *Haematologica* https://doi.org/10.3324/haematol.2022.281869
542 (2023) doi:10.3324/haematol.2022.281869.
543 14. Burd, A. *et al.* Precision medicine treatment in acute myeloid leukemia using prospective
544 genomic profiling: feasibility and preliminary efficacy of the Beat AML Master Trial. *Nat.*
545 *Med.* **26**, 1852–1858 (2020).
546 15. Irmisch, A. *et al.* The Tumor Profiler Study: integrated, multi-omic, functional tumor
547 profiling for clinical decision support. *Cancer Cell* **39**, 288–293 (2021).
548 16. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–
549 724 (2009).
550 17. The International Cancer Genome Consortium. International network of cancer genome
551 projects. *Nature* **464**, 993–998 (2010).
552 18. Lamb, J. *et al.* The Connectivity Map: Using Gene-Expression Signatures to Connect
553 Small Molecules, Genes, and Disease. *Science* **313**, 1929–1935 (2006).
554 19. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the
555 First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e17 (2017).
556 20. Ben Guebila, M. *et al.* GRAND: a database of gene regulatory network models across
557 human conditions. *Nucleic Acids Res.* **50**, D610–D621 (2022).
558 21. Cheng, J., Yang, L., Kumar, V. & Agarwal, P. Systematic evaluation of connectivity map
559 for disease indications. *Genome Med.* **6**, 540 (2014).
560 22. Musa, A. *et al.* A review of connectivity map and computational approaches in

561 pharmacogenomics. *Brief. Bioinform.* **19**, 506–523 (2018).
562 23. Moll, P., Ante, M., Seitz, A. & Reda, T. QuantSeq 3′ mRNA sequencing for RNA
563 quantification. *Nat. Methods* **11**, i–iii (2014).
564 24. Oren, M. & Prives, C. p53: A tale of complexity and context. *Cell* **187**, 1569–1573 (2024).
565 25. Herishanu, Y. *et al.* The lymph node microenvironment promotes B-cell receptor
566 signaling, NF-κB activation, and tumor proliferation in chronic lymphocytic leukemia.
567 *Blood* **117**, 563–574 (2011).
568 26. Landau, D. A. *et al.* Mutations driving CLL and their evolution in progression and relapse.
569 *Nature* **526**, 525–530 (2015).
570 27. Rendeiro, A. F. *et al.* Chromatin mapping and single-cell immune profiling define the
571 temporal dynamics of ibrutinib response in CLL. *Nat. Commun.* **11**, 577 (2020).
572 28. Badia-i-Mompel, P. *et al.* decoupleR: ensemble of computational methods to infer
573 biological activities from omics data. *Bioinforma. Adv.* **2**, vbac016 (2022).
574 29. Zaware, N. & Zhou, M.-M. Bromodomain biology and drug discovery. *Nat. Struct. Mol.*
575 *Biol.* **26**, 870–879 (2019).
576 30. Azmi, A. S., Uddin, M. H. & Mohammad, R. M. The nuclear export protein XPO1 — from
577 biology to targeted therapy. *Nat. Rev. Clin. Oncol.* **18**, 152–169 (2021).
578 31. Zhou, Y., Luo, K., Liang, L., Chen, M. & He, X. A new Bayesian factor analysis method
579 improves detection of genes and biological processes affected by perturbations in single-
580 cell CRISPR screening. *Nat. Methods* **20**, 1693–1703 (2023).
581 32. Kielbassa, K. *et al.* Ibrutinib sensitizes CLL cells to venetoclax by interrupting TLR9-
582 induced CD40 upregulation and protein translation. *Leukemia* **37**, 1268–1276 (2023).
583 33. Mandato, E. *et al.* MYD88L265P augments proximal B-cell receptor signaling in large B-
584 cell lymphomas via an interaction with DOCK8. *Blood* **142**, 1219–1232 (2023).
585 34. Phelan, J. D. *et al.* A multiprotein supercomplex controlling oncogenic signalling in
586 lymphoma. *Nature* **560**, 387–391 (2018).
587 35. Döhner, H. *et al.* Genomic Aberrations and Survival in Chronic Lymphocytic Leukemia.
588 *N. Engl. J. Med.* **343**, 1910–1916 (2000).

589 36. Fabbri, G. & Dalla-Favera, R. The molecular pathogenesis of chronic lymphocytic
590 leukaemia. *Nat. Rev. Cancer* **16**, 145–162 (2016).
591 37. Van De Haar, J. *et al.* Identifying Epistasis in Cancer Genomes: A Delicate Affair. *Cell*
592 **177**, 1375–1383 (2019).
593 38. McFarland, J. M. *et al.* Multiplexed single-cell transcriptional response profiling to define
594 cancer vulnerabilities and therapeutic mechanism of action. *Nat. Commun.* **11**, 4296
595 (2020).
596 39. Li, H., Qiu, J. & Fu, X. RASL-seq for Massively Parallel and Quantitative Analysis of
597 Gene Expression. *Curr. Protoc. Mol. Biol.* **98**, (2012).
598 40. Ye, C. *et al.* DRUG-seq for miniaturized high-throughput transcriptome profiling in drug
599 discovery. *Nat. Commun.* **9**, 4307 (2018).
600 41. O'Dwyer, P. J. *et al.* The NCI-MATCH trial: lessons for precision oncology. *Nat. Med.* **29**,
601 1349–1357 (2023).
602 42. Bruch, P. *et al.* Drug-microenvironment perturbations reveal resistance mechanisms and
603 prognostic subgroups in CLL. *Mol. Syst. Biol.* **18**, e10855 (2022).
604 43. Reid, J. C. *et al.* Human pluripotent stem cells identify molecular targets of trisomy 12 in
605 chronic lymphocytic leukemia patients. *Cell Rep.* **34**, 108845 (2021).
606 44. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
607 for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
608 45. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-
609 seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
610 46. John, C. R. *et al.* M3C: Monte Carlo reference-based consensus clustering. Preprint at
611 https://doi.org/10.1101/377002 (2018).
612 47. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in
613 multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
614 48. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing
615 and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
616 49. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count

617 data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092
618 (2019).
619 50. Fischer, M., Schwarz, R., Riege, K., DeCaprio, J. A. & Hoffmann, S. TargetGeneReg 2.0:
620 a comprehensive web-atlas for p53, p63, and cell cycle-dependent gene regulation. *NAR*
621 *Cancer* **4**, zcac009 (2022).
622 51. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. & Leisch, F. e1071: Misc Functions
623 of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien.
624 1.7-16 https://doi.org/10.32614/CRAN.package.e1071 (1999).
625 52. Butts, C. T. **network** : A Package for Managing Relational Data in *R. J. Stat. Softw.* **24**,
626 (2008).
627 53. Townes, F. W., Hicks, S. C., Aryee, M. J. & Irizarry, R. A. Feature selection and
628 dimension reduction for single-cell RNA-Seq based on a multinomial model. *Genome*
629 *Biol.* **20**, 295 (2019).
630 54. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear
631 Models via Coordinate Descent. *J. Stat. Softw.* **33**, (2010).
632 55. Kim, S.-Y. & Volsky, D. J. PAGE: Parametric Analysis of Gene Set Enrichment. *BMC*
633 *Bioinformatics* **6**, 144 (2005).
634 56. Landau, D. A. *et al.* The evolutionary landscape of chronic lymphocytic leukemia treated
635 with ibrutinib targeted therapy. *Nat. Commun.* **8**, 2185 (2017).

636 # Acknowledgements

637 T. Zenz was supported by the CRPP ‘Next Generation Drug Response Profiling for
638 Personalized Cancer Care’, the Swiss Cancer Research foundation (KFS-4439-02-2018), the
639 Monique-Dornonville-de-la-Cour Stiftung and the Helmut Horten Stiftung. M. F. Pohly, W.
640 Huber were supported by the INTeRCePT, and T. Zenz by the INTeRCePT and POLAR
641 (Uniscientia Stiftung) projects funded by The LOOP Zurich. J. Lu, C. Lohoff and W. Huber were
642 supported by the SMART-CARE project funded by the Federal Ministry of Research,
643 Technology and Space of Germany (161L0213 and 16LW0237). For access to infrastructure,

644 we thank M. Bawohl, M. Rechsteiner and A. Weber at the Department of Pathology and
645 Molecular Pathology, and the Functional Genomics Center Zurich. For access to the EMBL
646 Heidelberg HPC Cluster, we gratefully thank J. Pečar and EMBL IT Services.

647 **Contributions**

648 THLD, JL, WH, TZ conceptualised the study and designed the experiments. THLD, SK, MFP
649 and FJ conducted experiments. SS, JH, WH, JL and TZ were responsible for data curation.
650 CL, FJ and VB performed RNA sequencing and data pre-processing. THLD, CL, JL and TZ
651 performed data analysis and data visualisation. THLD, CL, JL, WH and TZ interpreted the data.
652 THLD, CL, JL, WH and TZ wrote the original draft of the manuscript. All authors reviewed and
653 edited the manuscript.

654 Ethics declarations

655 **Competing interests**

656 TZ received consulting fees from Astra-Zeneca, BeiGene, AbbVie, Janssen, Novartis, Lilly,
657 Roche, Bristol-Myers Squibb, and Gilead Sciences, and payment for lectures from Astra-
658 Zeneca, BeiGene, AbbVie, Janssen, Novartis, Lilly, Roche, Bristol-Myers Squibb, and Gilead
659 Sciences. All other authors declare no competing interests.

660

Figures

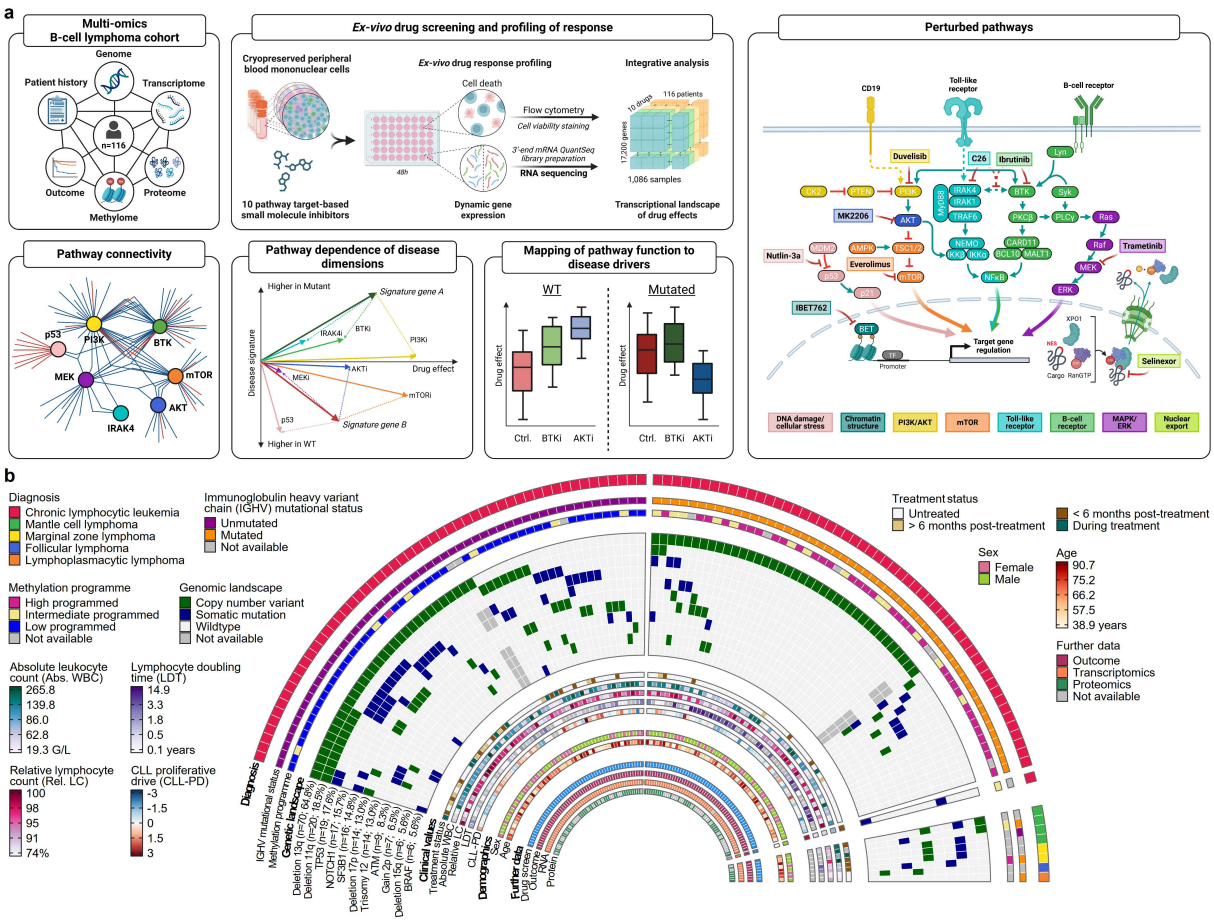


Fig. 1: Study overview

a, Study outline (https://BioRender.com/6i76n18). **b**, Genetic landscape of study cohort including recurrent genetic aberrations (> 5 cases), clinical and demographic parameters, and availability of additional molecular omics profiling data. Colours for continuous variables are scaled by minimum value, 25th, 50th, 75th percentile and maximum value. Numbers indicated in brackets show case number and frequency of patients with respective aberration for the CLL cohort.

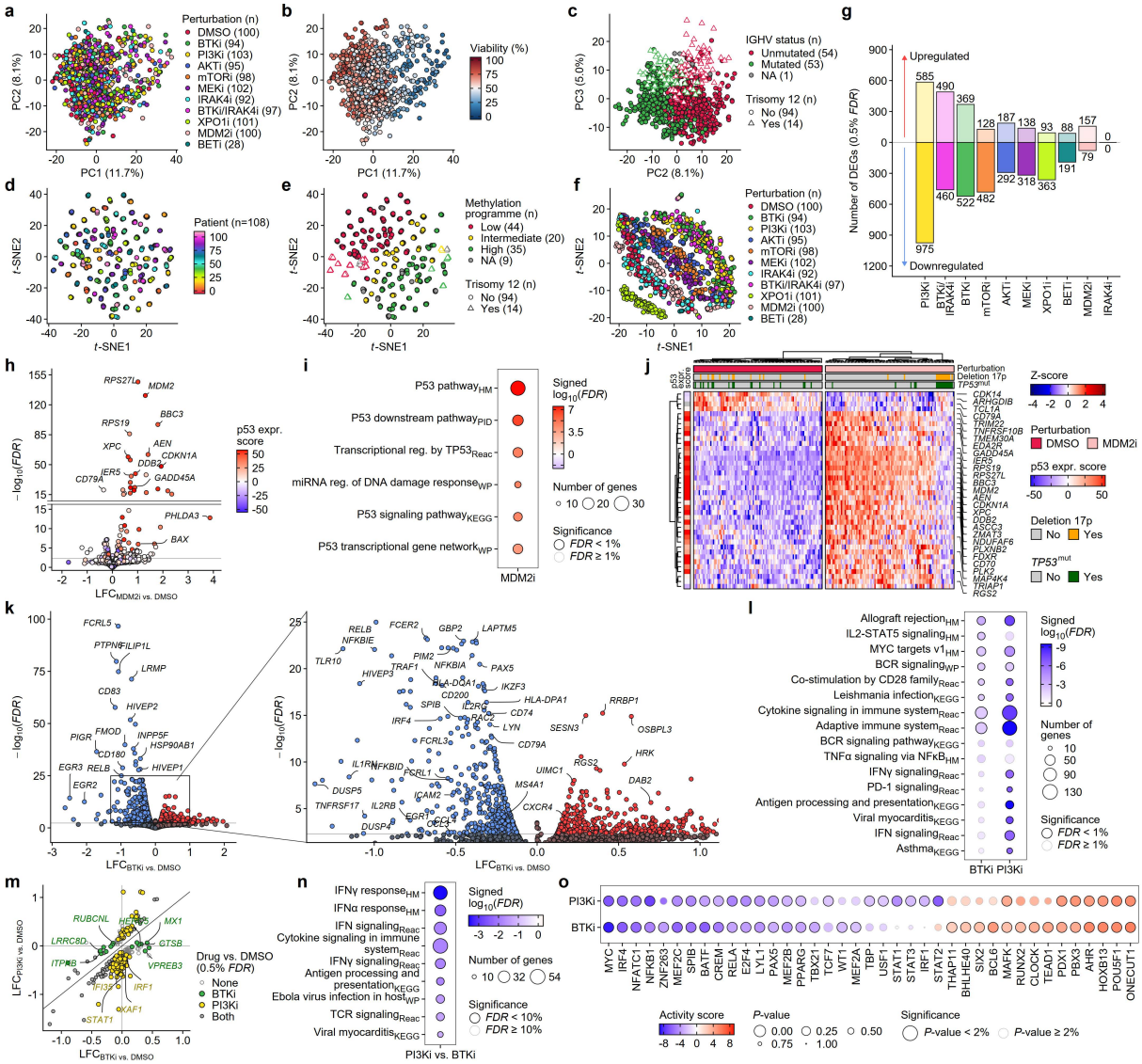


Fig. 2: Transcriptional landscape of drug effects in CLL

a, First two principal components of PCA based on the 1,000 most variant genes coloured by perturbation and **b**, absolute viability of each drug-treated sample measured by flow cytometry. **c**, PC2 and PC3 of PCA based on the 1,000 most variant genes coloured by IGHV and trisomy 12 status. **d**, *t*-SNE plot based on the 1,000 most variant genes coloured by patient and **e**, by methylation groups. Samples from the same patient form tight clusters. **f**, *t*-SNE plot based on the 1,000 most variant genes after adjustment for the patient effect coloured by perturbation. **g**, Number of differentially expressed genes (DEGs) between drug- and DMSO-treated samples (0.5% *FDR*, |fold change| > 1.05). Induction or suppression for each treatment are shown. **h**, Volcano plot summarising the differential gene expression after MDM2 inhibitor treatment. Genes are annotated according to p53 expression score[50]. **i**, GSEA with DEGs after

683 MDM2 inhibitor treatment (0.5% *FDR*) using hallmark and canonical pathways gene sets.
684 Gene sets were filtered by *FDR* (< 1%) and number of enriched genes (> 10). **j**, Heatmap with
685 hierarchical clustering showing patient effect-adjusted normalised expression of top 40 DEGs
686 after MDM2 inhibitor treatment. *TP53* mutation and deletion status are annotated. **k**, Volcano
687 plot and zoom-in summarising the differential gene expression after BTK inhibitor treatment. **l**,
688 GSEA with DEGs after BTK and PI3K inhibitor treatment (0.5% *FDR*) using hallmark and
689 canonical pathways gene sets. Gene sets were filtered by *FDR* (< 1%) and number of enriched
690 genes (> 10). Depicted are the top 10 gene sets for BTK and PI3K inhibition. **m**, Scatter plot
691 of drug effect sizes for DEGs between PI3K and BTK inhibitor treatment (10% *FDR*). Colours
692 and shapes depict in which comparisons the gene was differentially expressed (0.5% *FDR*).
693 **n**, GSEA with DEGs between PI3K and BTK inhibitor treatment (10% *FDR*) using hallmark and
694 canonical pathways gene sets. Gene sets were filtered by FDR (< 10%) and number of
695 enriched genes (> 10). **o**, Inferred transcription factor (TF) activity using all genes. TFs were
696 filtered by *FDR* (< 10%) and number of targets in TF regulon (> 30 target genes).

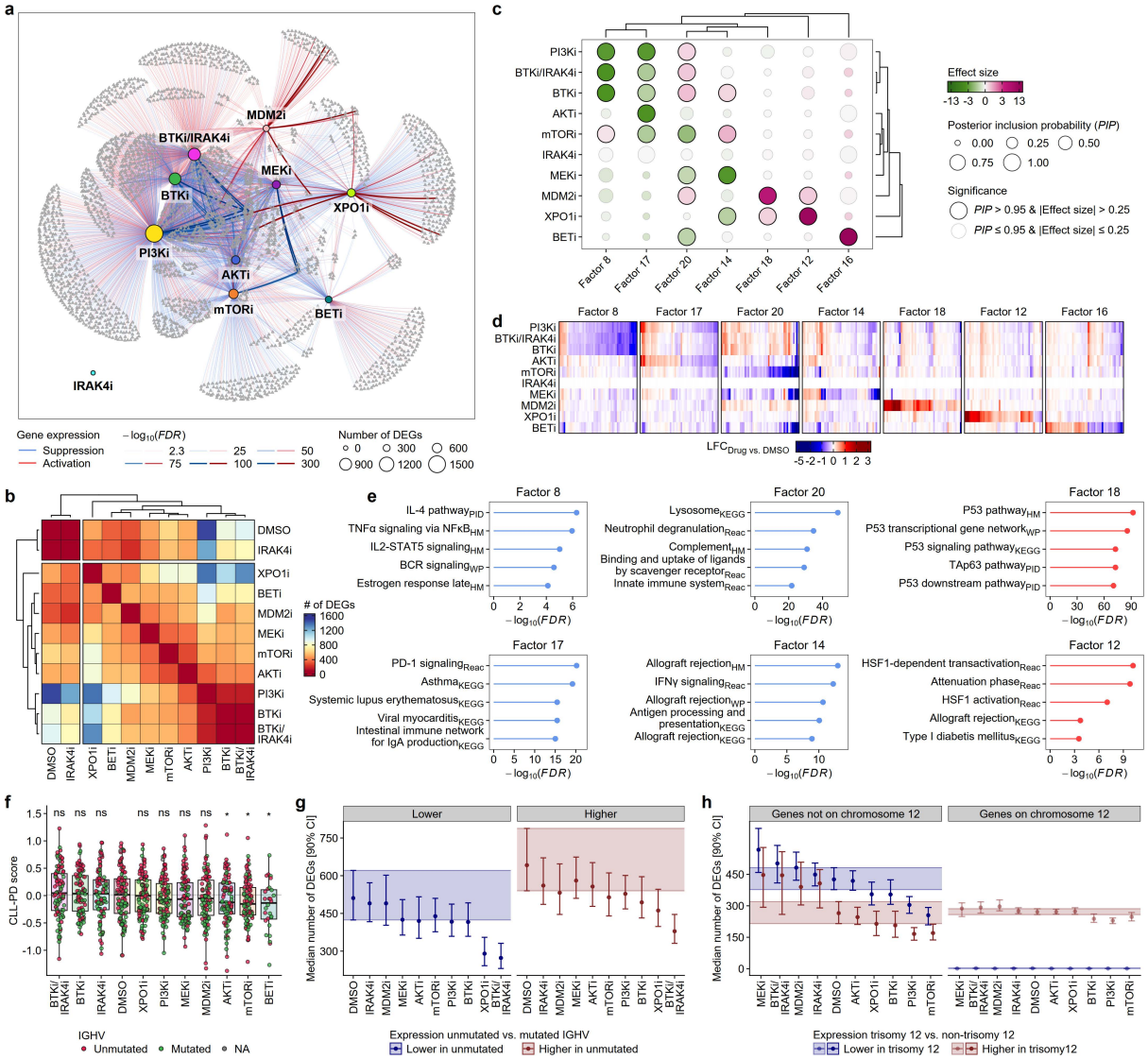


Fig. 3: Pathway connectivity in CLL

a, Pathway connectivity network shown as a bipartite graph. Genes and perturbations are connected if the perturbation led to differential expression of that gene. The total number of DEGs for each treatment is depicted by the size of the central nodes. Colour and thickness of edges represent whether genes are induced or suppressed, and significance of drug effect, respectively. **b**, Heatmap with hierarchical clustering showing number of DEGs after pairwise drug-drug comparison (0.5% *FDR*). **c**, Estimated effects of drug perturbations on co-regulated gene modules represented by latent factors. The posterior inclusion probability (*PIP*) denotes the probability that a perturbation affects the factor loading. Factors with significant loadings (*PIP* > 0.95, |effect size| > 0.25) are shown. The sign was inverted for selected factors to reflect the direction of the drug effects. **d**, Heatmap showing the drug effects for 50 genes with the

709 highest gene weights in each factor. **e**, GSEA with the gene weights for each factor using
710 hallmark and canonical pathways gene sets. Gene sets were filtered by *FDR* (< 1%) and
711 number of enriched genes (> 10). The sign of weights was reverted in accordance to (c). **f**,
712 CLL-PD scores computed from the transcriptomics data for each individual treatment.
713 Significance levels were calculated using two-tailed paired Student's *t*-test with DMSO as
714 reference and adjusted for multiple testing. Asterisks denote *FDR* < 10%. **g**, Median number
715 of DEGs and 90% confidence interval (CI) between samples with unmutated and mutated
716 IGHV status for each treatment at 10% *FDR* split into genes with lower or higher expression in
717 U-CLL compared to M-CLL. **h**, Median number of DEGs and 90% CI between samples with
718 and without trisomy 12 for each treatment at 10% *FDR* split into genes with lower or higher
719 expression in trisomy 12 compared to non-trisomy 12, and whether genes are encoded on
720 chromosome 12.

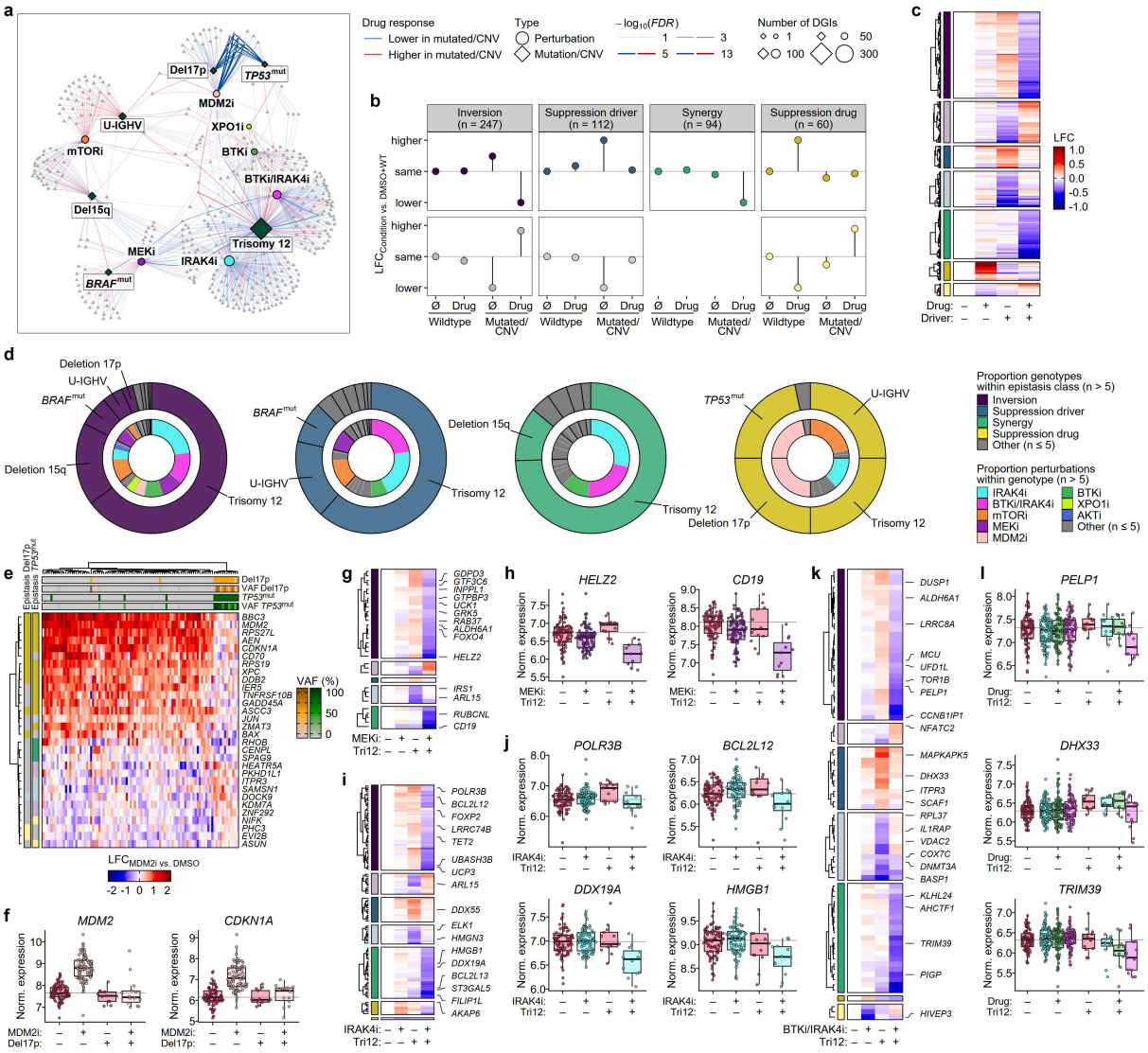
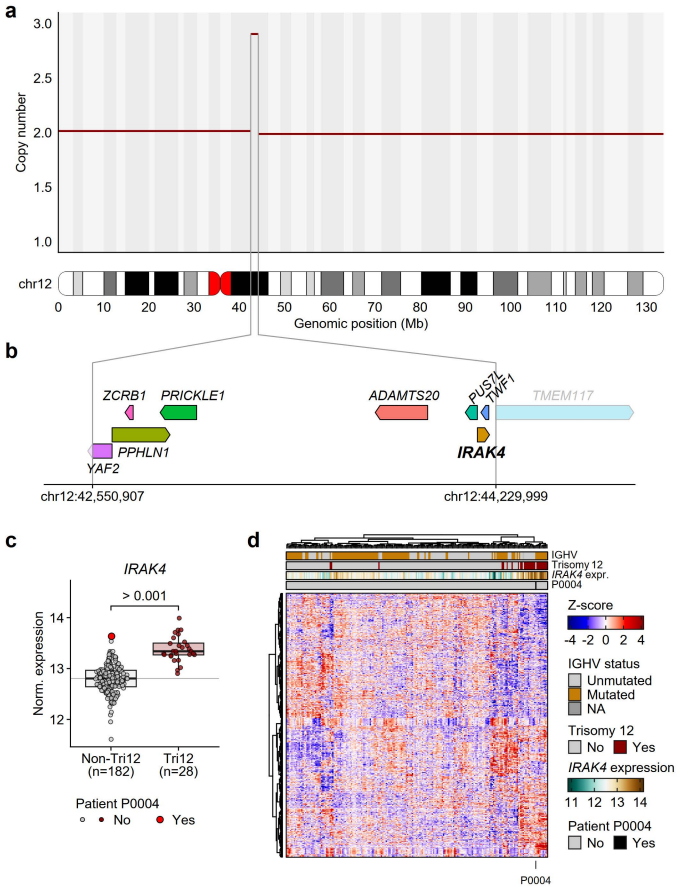


Fig. 4: Interaction of perturbation effects and disease drivers

a, Pathway connectivity network of significant drug-genotype interactions (DGIs) shown as a bipartite graph. Genes, drugs and disease aberration are connected if a differential transcriptional drug response was observed in the presence of the genetic driver (*FDR* < 10%). Drug-genotype pairs with at least 10 DGIs are displayed. **b**, Schematic classification of epistasis categories based on median values from all genes in the respective groups. **c**, Heatmap of relative gene expression of DGIs in the presence of either drug, genetic driver or both compared to WT/DMSO samples by epistasis class. **d**, Proportion of aberration contributing to each epistasis class (outer circle) and proportion of perturbations contributing to each aberration within each epistasis class (inner circle) with more than 5 DGIs. **e**, Heatmap visualising the effect sizes of genes with significant interaction between MDM2 inhibition and

733 del17p or *TP53*^{mut} (*FDR* < 10%). **f**, Effect of MDM2 inhibition on normalised expression of
734 *MDM2* and *CDKN1A* by deletion 17p status. **g**, Heatmap visualising the effect sizes of genes
735 with significant interaction between MEK inhibition and trisomy 12 (*FDR* < 10%). **h**, Normalised
736 expression of MEK inhibition on *HELZ2* and *CD19* by trisomy 12 status. **i**, Heatmap visualising
737 the effect sizes of genes with significant interaction between IRAK4 inhibition and trisomy 12
738 (*FDR* < 10%). **j**, Effect of IRAK4 inhibition on normalised expression of *POLR3B*, *BCL2L12*,
739 *DDX19A* and *HMGB1* by trisomy 12 status. **k**, Heatmap visualising the effect sizes of genes
740 with significant interaction between combined BTK/IRAK4 inhibition and trisomy 12 (*FDR* <
741 10%). **l**, Effect of DMSO (red), IRAK4 (cyan), BTK (green) and BTK/IRAK4 inhibition (pink) on
742 normalised expression of *PELP1*, *DHX33* and *TRIM39* by trisomy 12 status.



743

744 **Fig. 5: IRAK4 is a disease driver targeted by trisomy 12**

745 **a**, CNV plot with ideogram of chromosome 12 for patient P0004. The red line shows the copy
746 numbers across the chromosome. **b**, Genomic track displaying the protein-coding genes on
747 the amplified locus chr12:42,550,907-chr12: 44,229,999 (hg19). **c**, Gene expression of IRAK4
748 by trisomy 12 status for the CLL cohort from Lütge et al.[13]. The significance level was
749 calculated using two-tailed Student's *t*-test and shown as nominal *P*-value. **d**, Heatmap
750 visualising the scaled gene expression of trisomy 12 signature genes excluding the genes on
751 chromosome 12 (n = 642) for the CLL cohort (n = 210) from Lütge et al.[13]

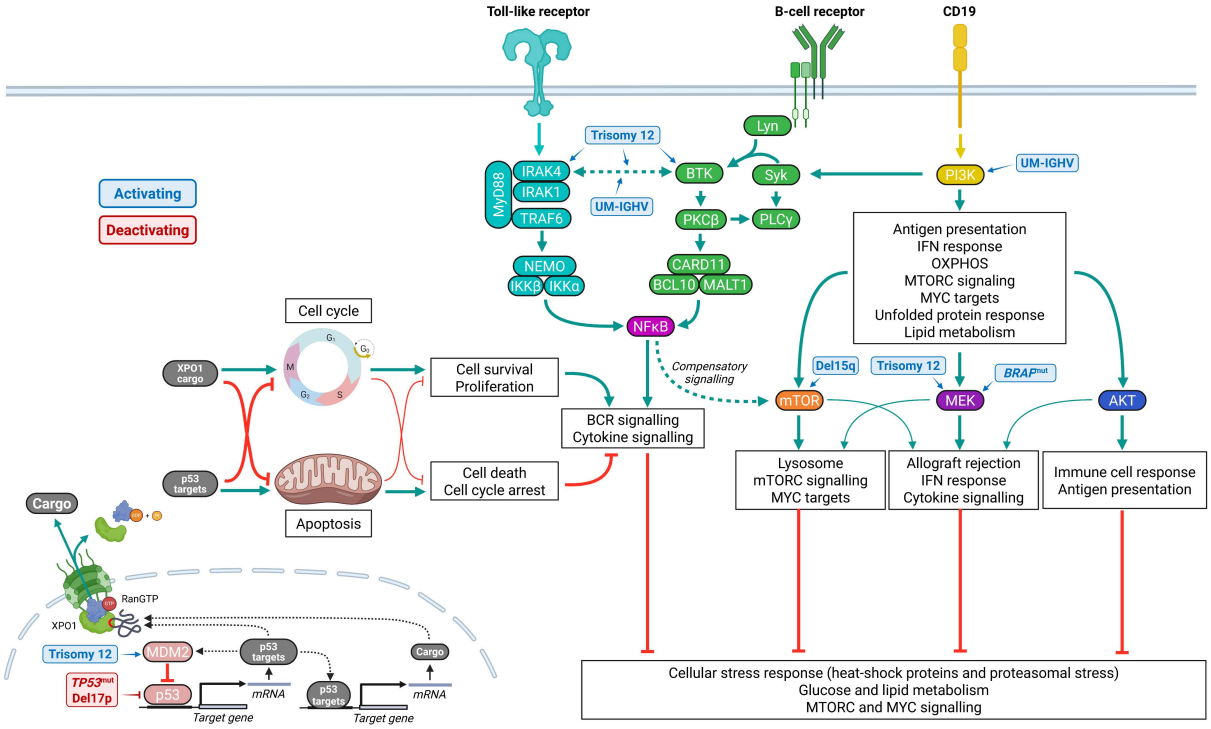
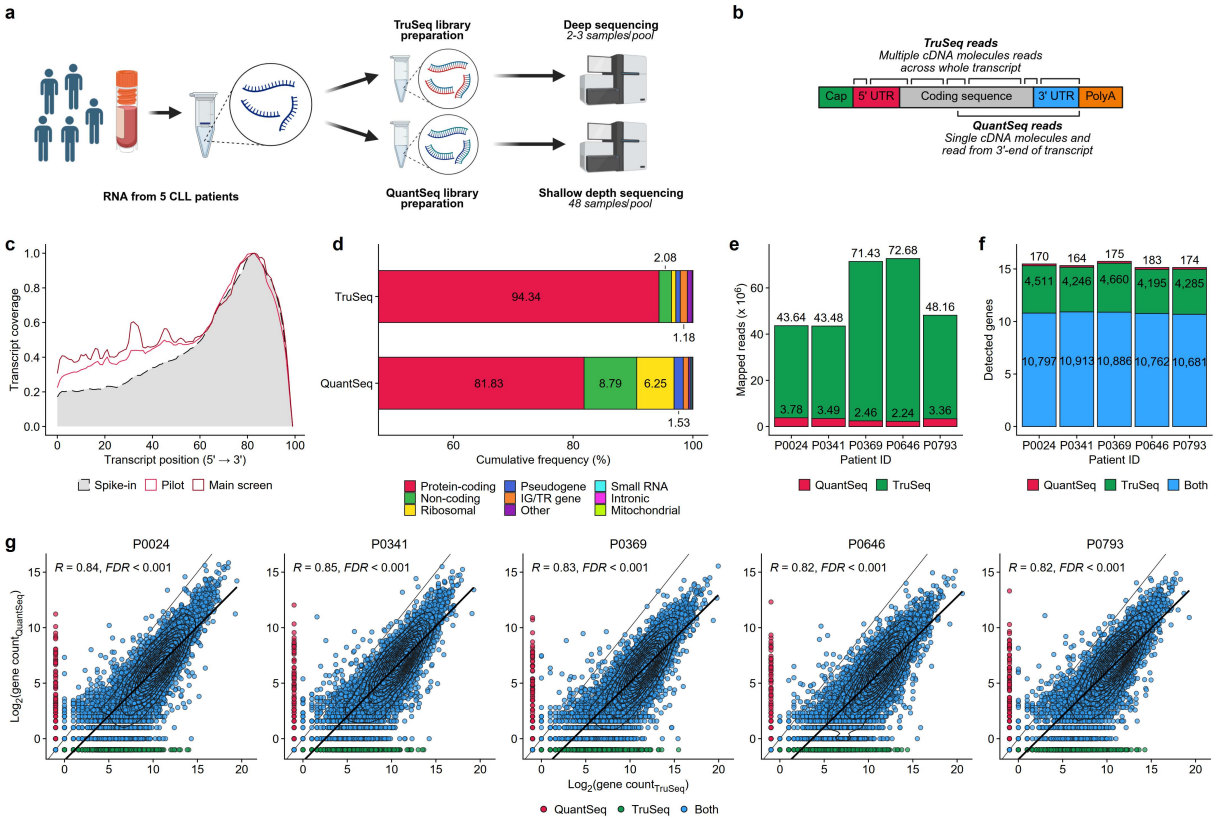


Fig. 6: Pathway diagram and dependency map in CLL

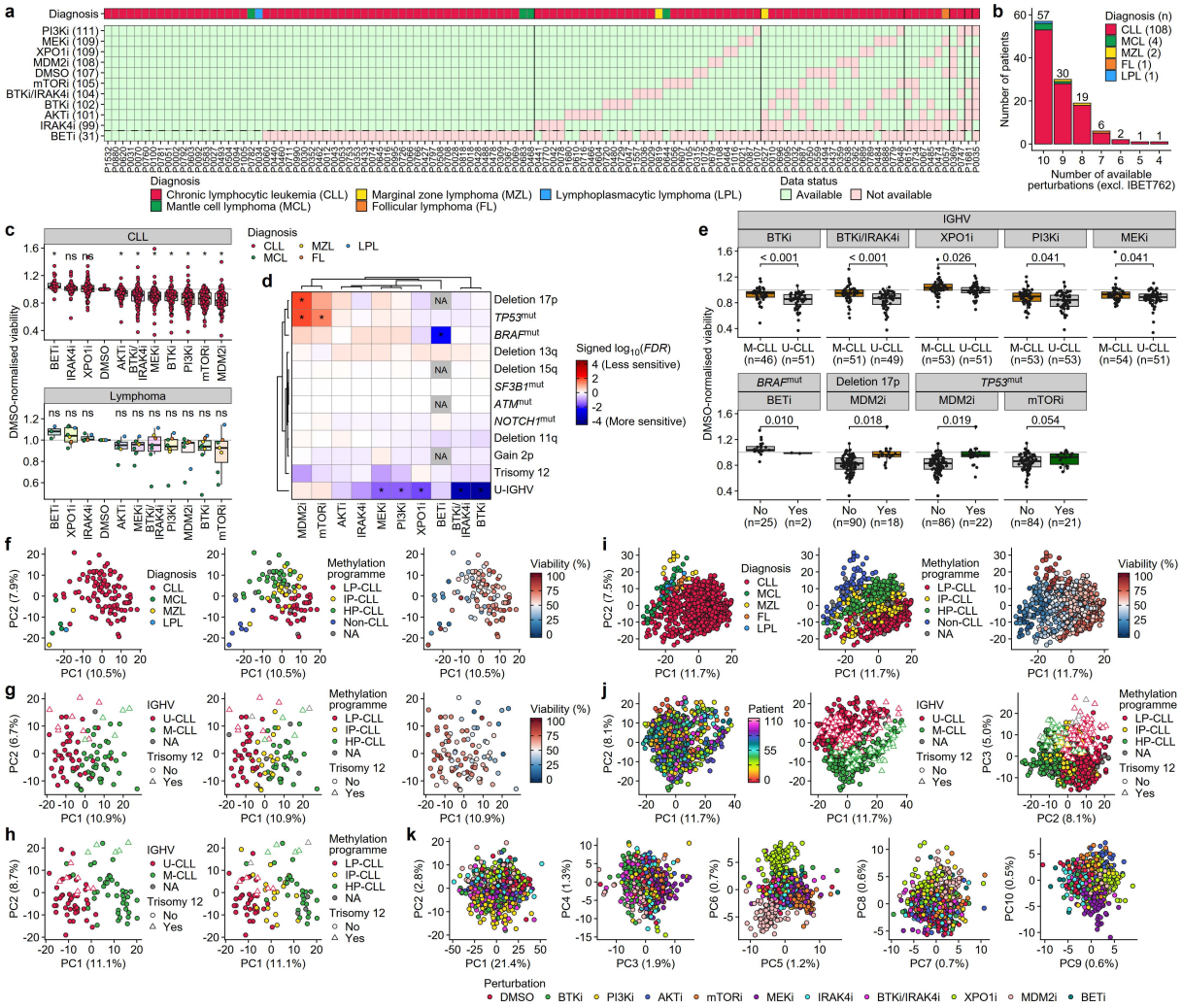
Pathway dependencies modulated by gene mutations/CNVs or IGHV status were identified by drug-genotype interaction analysis and depicted as activating (blue) or inhibitory (red) (https://BioRender.com/svmtjkd).

Supplementary figures



Suppl. Fig. 1: Comparison of shallow 3'-QuantSeq to regular coverage whole-transcript TruSeq sequencing

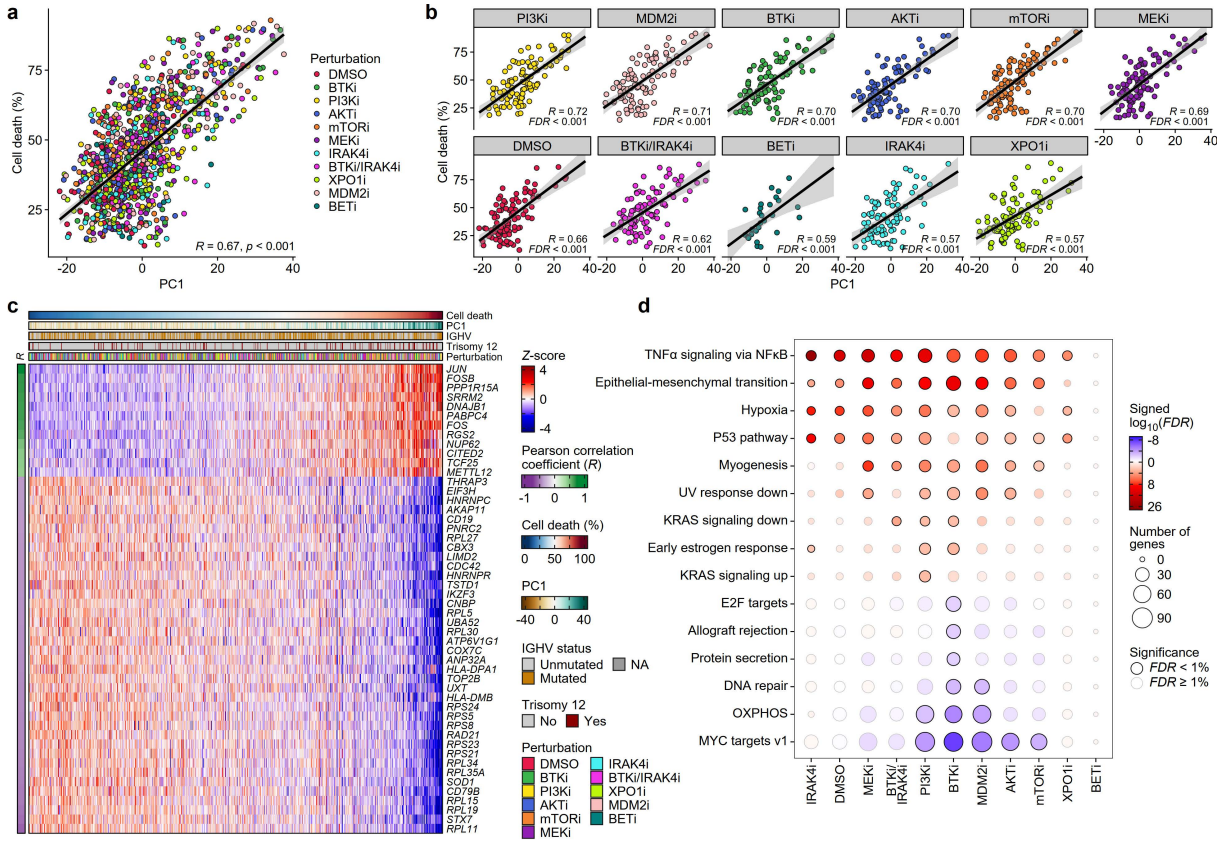
a, Overview of experimental workflow (https://BioRender.com/u4e6koa). **b**, Schematic comparison of read distribution along transcript between 3'-QuantSeq[23] and whole-transcript TruSeq (https://BioRender.com/sx3vfjw). **c**, Biased read coverage of transcript 3'-end for exemplary samples. **d**, Mapped biotypes, **e**, number of mapped reads, **f**, and number of detected genes by method. Colours depict whether genes were detected by both methods, only by TruSeq or only by QuantSeq. **g**, Correlation of log2 gene expression counts detected by both methods.



Suppl. Fig. 2: Overview viability effect of drugs and transcriptional landscape

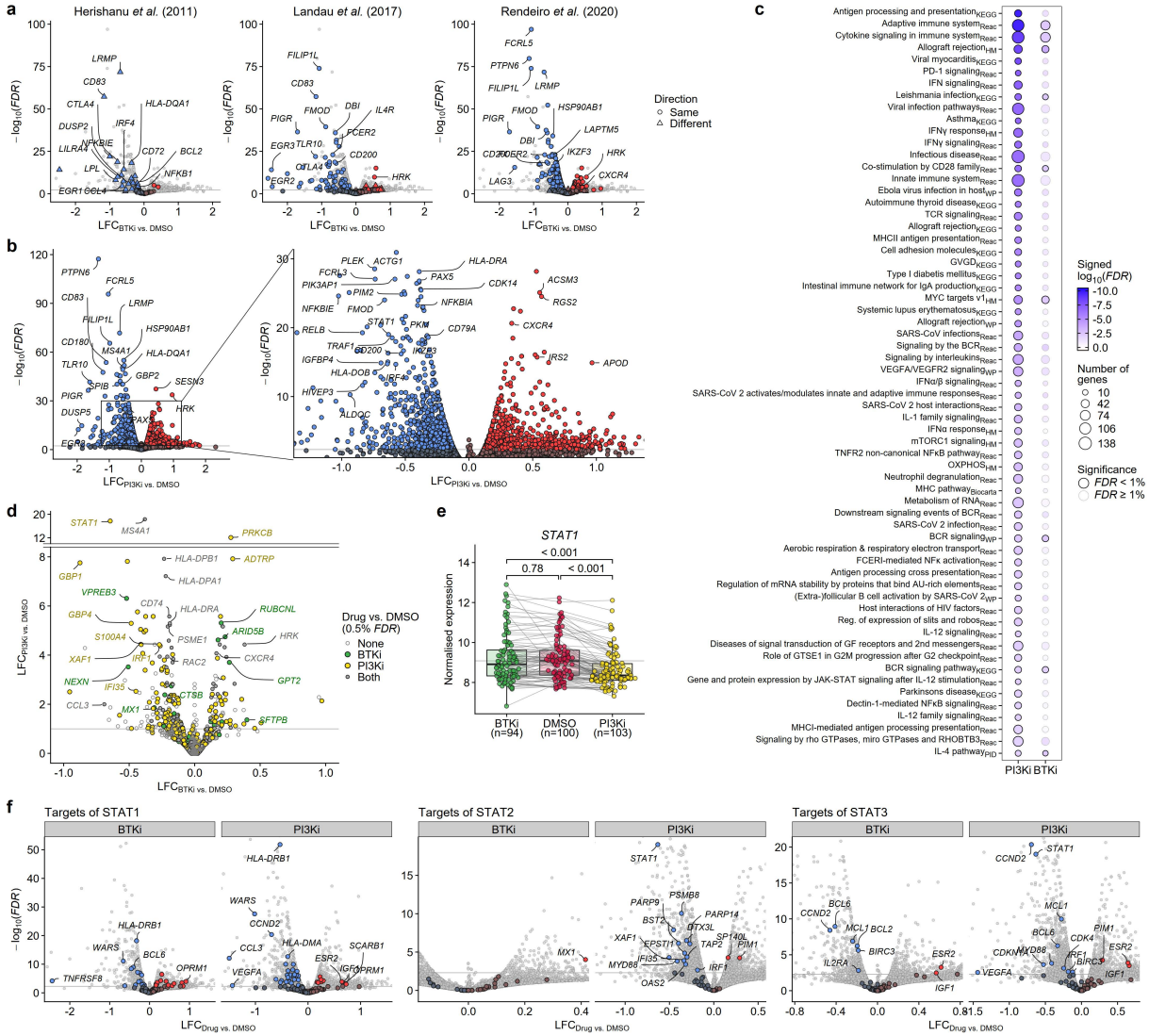
a, Overview of completeness and **b**, number of available treatment conditions by patient. **c**, DMSO-normalised sample viability after 48h treatment measured by flow cytometry and separated by diagnosis. Significance levels were calculated using two-tailed paired Student's *t*-test and adjusted for multiple testing. Asterisk denotes significant differences between drug- and DMSO-treated samples (*FDR* < 10%). **d**, Genomic association with drug response for CLL samples. Asterisk denotes associations with significant differences in drug response between patients with and without respective genetic aberrations with more than 5 cases (*FDR* < 10%). Significance levels were calculated using two-tailed Student's *t*-test and adjusted for multiple testing. **e**, Boxplots of individual patient data shown for top significant associations (*FDR* < 10%). Significance levels were calculated using two-tailed Student's *t*-test and adjusted for multiple testing. **f**, PCA based on 1,000 most variant genes from DMSO-treated CLL and

781 lymphoma samples (n = 107), **g**, only CLL samples (n = 100), and **h**, from overlapping CLL
782 patients from Lütge et al.[13] profiled by deep whole-transcript RNA sequencing (n = 95). **i**, PCA
783 based on 1,000 most variant genes from DMSO- and drug-treated CLL and lymphoma
784 samples (n = 1,086), **j**, only CLL samples (n = 1,010), **k**, and after adjustment for the patient
785 effect (n = 1,010).



Suppl. Fig. 3: Transcriptional signature associated with viability

a, Correlation of cell death (100% - sample viability) with PC1 for all samples and **b**, by treatment. Correlation trends are indicated by Pearson correlation coefficient (*R*) and adjusted for multiple testing. **c**, Heatmap showing normalised expression of top 50 correlated genes with cell death. Correlation trends are indicated by Pearson correlation coefficient (*R*). **d**, GSEA individually for each treatment using genes with significant correlations between expression and cell death (*FDR* < 0.5%) using hallmark gene sets and number of enriched genes (< 10). Gene sets were filtered by *FDR* (< 1%). Enrichments indicate positive (red) or negative (blue) associations with cell death.



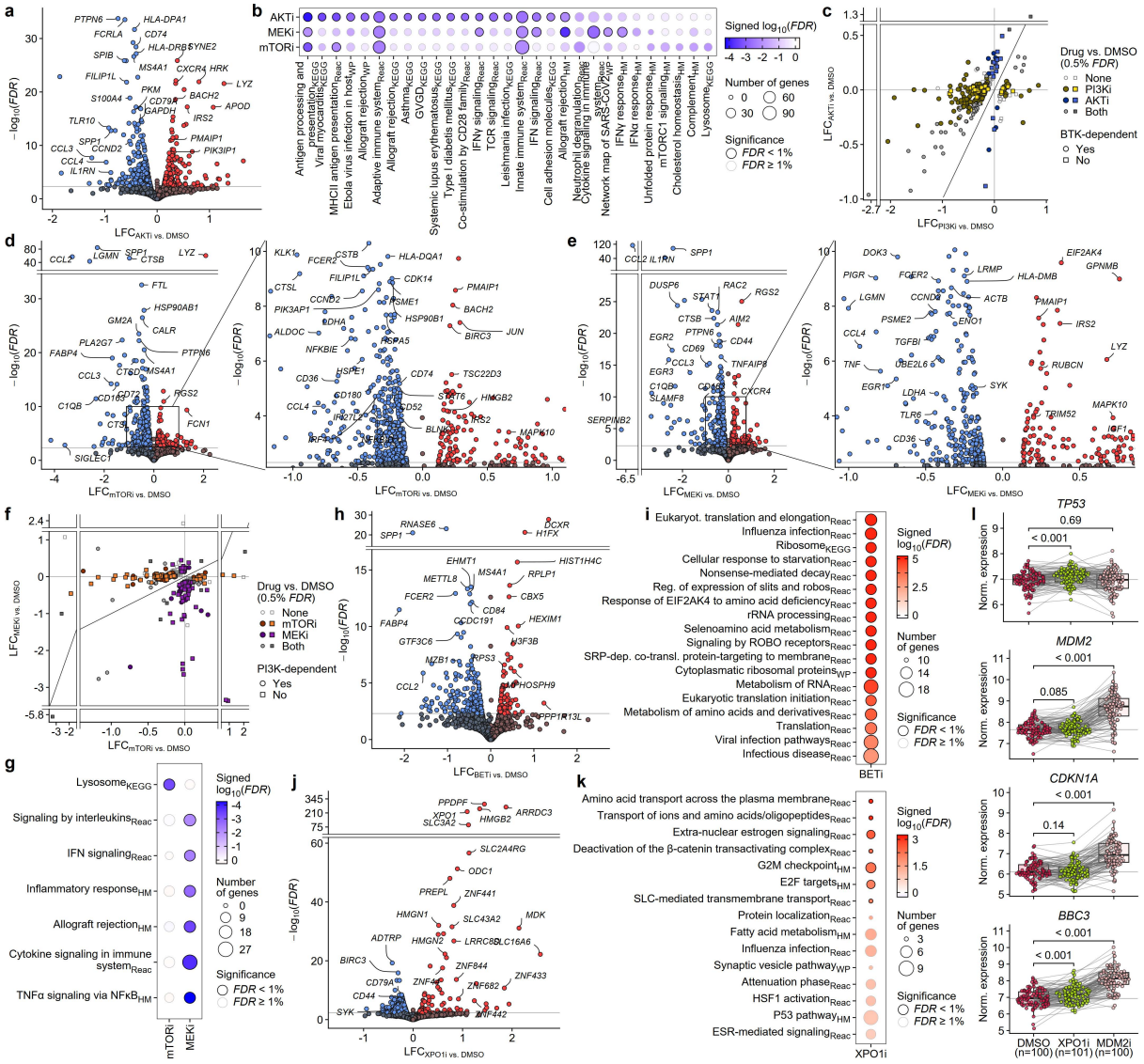
Suppl. Fig. 4: Drug effects of BTK and PI3K inhibitors on BCR inhibition

a, Projection of data from Herishanu et al.[25] (*ex-vivo* BCR activation signature), Landau et al.[56] and Rendeiro et al.[27] (*in-vivo* BTK inhibitor signature) into volcano plot of differential gene expression after BTK inhibition. Shapes depict whether the gene is regulated into the same or different direction. **b**, Volcano plot summarising the differential gene expression after PI3K inhibition. **c**, GSEA with DEGs after BTK and PI3K inhibitor treatment (0.5% *FDR*) using hallmark and canonical pathways. Gene sets were then filtered by *FDR* (< 1%) and number of enriched genes (> 10). Depicted are all significant gene sets for PI3K inhibition. **d**, Volcano plot summarising the differential gene expression between PI3K and BTK inhibition. Colours and shapes depict in which comparisons the gene was deregulated (0.5% *FDR*). **e**, Normalised *STAT1* expression in samples treated with DMSO, BTK or PI3K inhibitor. Significance levels

808 were calculated using two-tailed paired Student's *t*-test and shown as nominal *P*-values. **f**,

809 Projection of STAT1, STAT2 and STAT3 transcription factor target genes into volcano plot of

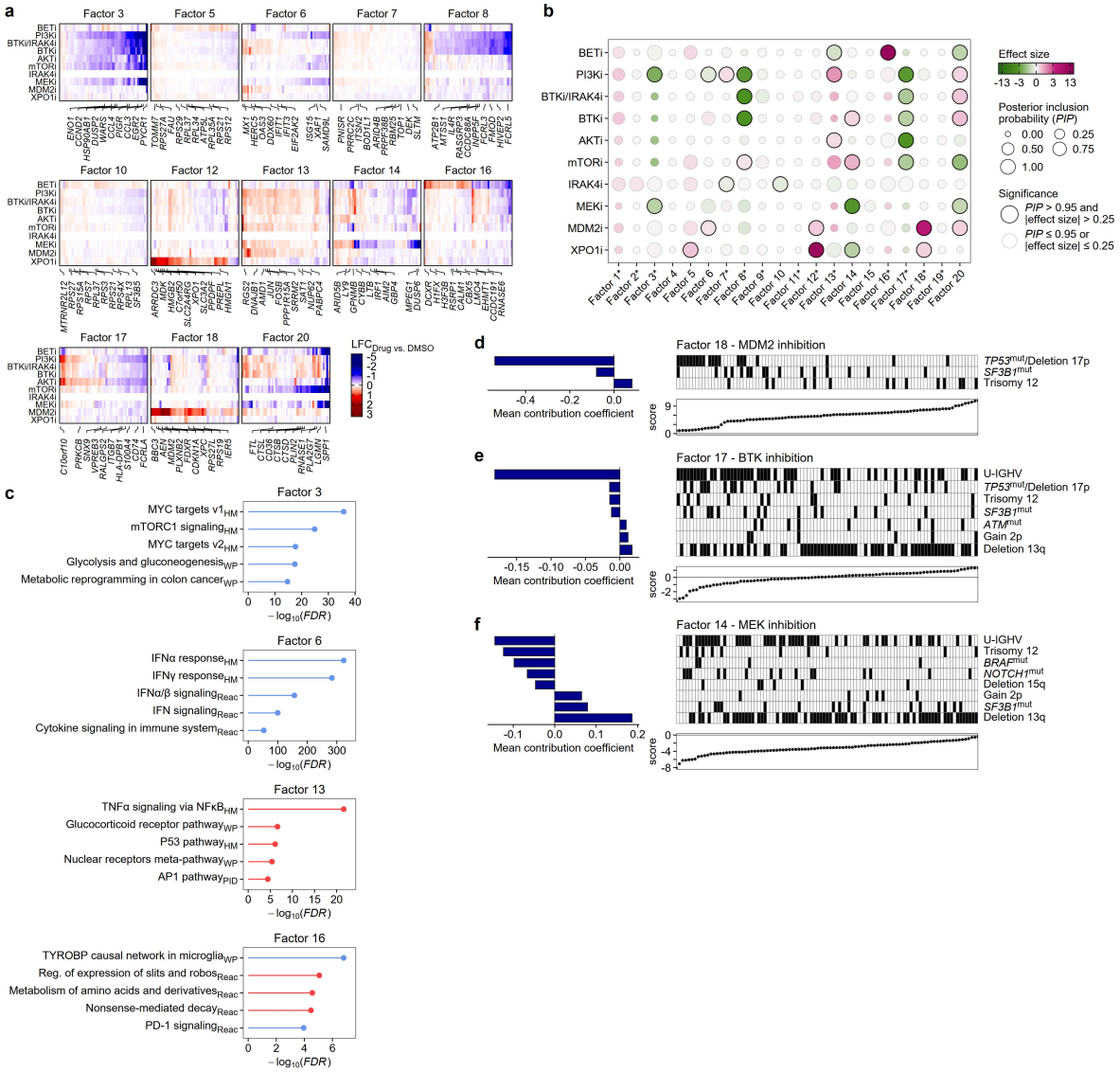
810 differential expression after BTK and PI3K inhibition.



Suppl. Fig. 5: Drug effects of AKT, mTOR, MEK, XPO1 and BET inhibitors

a, Volcano plot summarising gene expression after AKT inhibitor treatment. **b**, GSEA with DEGs after AKT, mTOR and MEK inhibitor treatment (0.5% *FDR*) using hallmark and canonical pathways gene sets. Gene sets were filtered by *FDR* (< 1%) and number of enriched genes (> 10). Depicted are all significant gene sets for AKT and the top 10 gene sets for MEK and mTOR inhibition. **c**, Scatter plot of drug effect sizes for DEGs between AKT and PI3K inhibitor treatment (0.5% *FDR*). Colours and shapes depict in which comparisons the gene was deregulated (0.5% *FDR*). **d**, Volcano plot summarising gene expression after mTOR and **e**, MEK inhibitor treatment. **f**, Scatter plot of drug effect sizes for DEGs between MEK and mTOR inhibitor treatment (0.5% *FDR*). Colours and shapes depict in which comparisons the gene was deregulated (0.5% FDR). **g**, GSEA with genes differentially expressed between MEK and

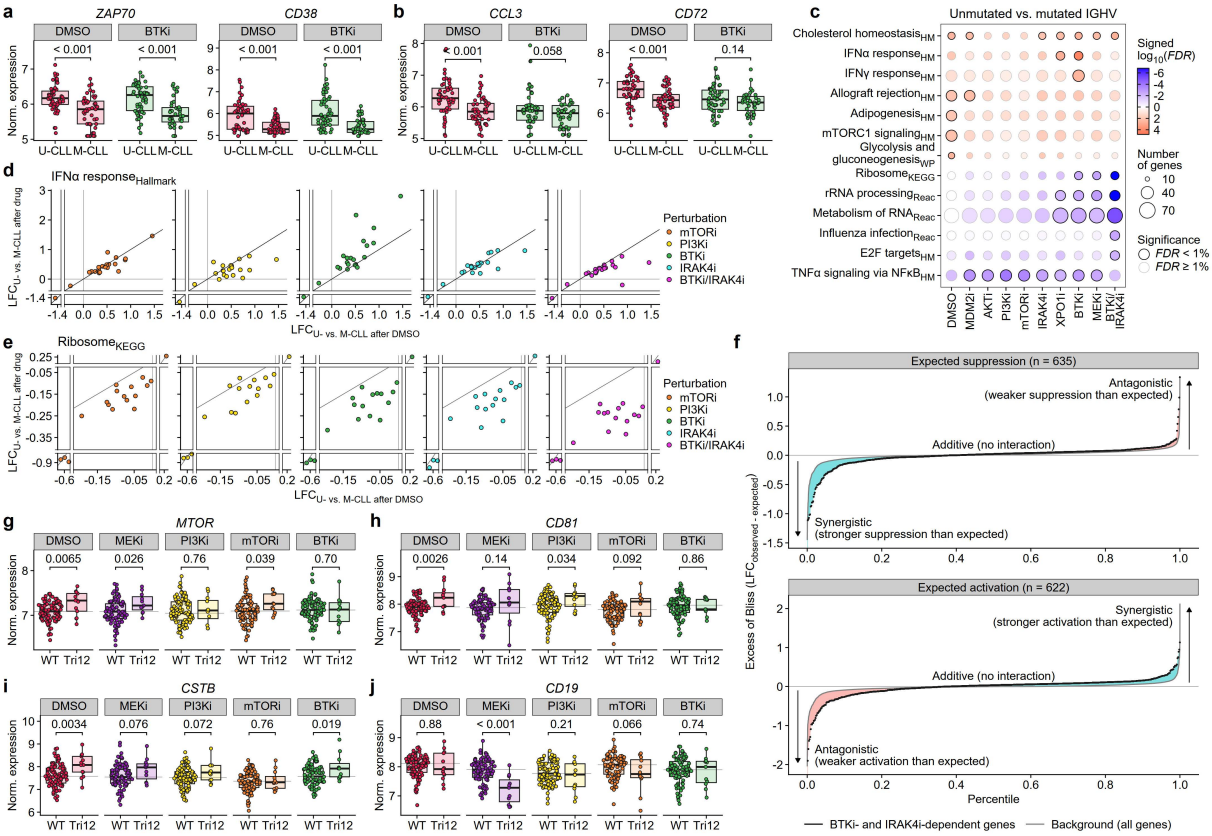
823 mTOR inhibition (0.5% *FDR*) using hallmark and canonical pathways gene sets. Gene sets
824 were filtered by *FDR* (< 1%) and number of enriched genes (> 10). **h**, Volcano plot summarising
825 gene expression after BET inhibitor treatment. **i**, GSEA with DEGs after BET inhibitor treatment
826 (0.5% *FDR*) using hallmark and canonical pathways gene sets. Gene sets were filtered by
827 *FDR* (< 1%) and number of enriched genes (> 10). **j**, Volcano plot summarising gene
828 expression after XPO inhibitor treatment. **k**, GSEA with DEGs after XPO1 inhibitor treatment
829 (0.5% *FDR*) using hallmark and canonical pathways gene sets. Gene sets were filtered by
830 *FDR* (< 1%) and number of enriched genes (≥ 3). **l**, Effect of XPO1 (green) and MDM2 inhibition
831 (rose) on the expression of *TP53* and the p53 target genes *MDM2*, *CDKN1A*, *BBC3*.
832 Significance levels were calculated using two-tailed paired Student's *t*-test and shown as
833 nominal *P*-values.



Suppl. Fig. 6: Guided sparse factor analysis of drug effects

a, Heatmap shows drug effects for 50 genes with the highest gene weights for factors with significant loadings. Top 10 genes with the highest gene weights for each factor are labelled. **b**, Estimated effects of drug perturbations on co-regulated gene modules represented by 20 latent factors. The posterior inclusion probability (*PIP*) denotes the probability that a perturbation affects the factor loading. Factors with significant loadings (*PIP* > 0.95, |effect size| > 0.25) are shown. Asterisks denote factors for which the sign was inverted to reflect the direction of the drug effects. **c**, GSEA with the gene weights for each factor using hallmark and canonical pathways gene sets. Gene sets were filtered by *FDR* (< 1%) and number of enriched genes (> 10). The sign of weights was reverted in accordance to (b). **d**, Influence of genetic disease drivers on the contribution of the MDM2 inhibitor effect in factor 18, **e**, BTK inhibitor

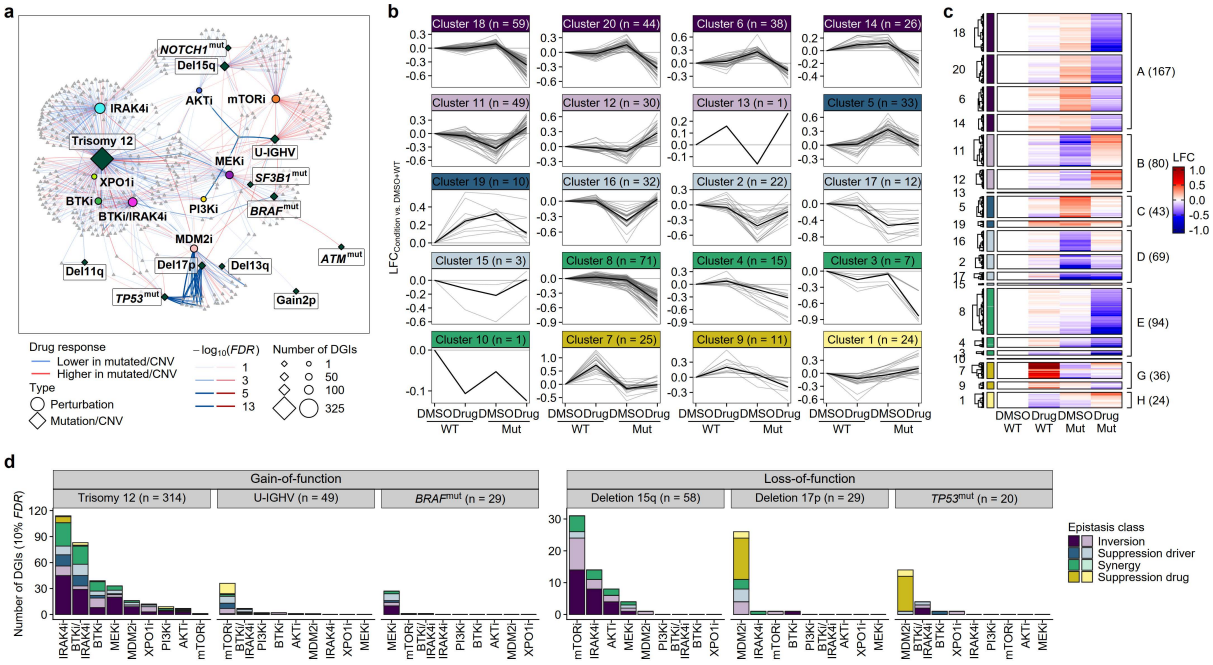
846 effect in factor 17, and **f**, MEK inhibitor effect in factor 14 computed by multiple linear
847 regression.



Suppl. Fig. 7: Pathway dependency of expression signatures and disease drivers

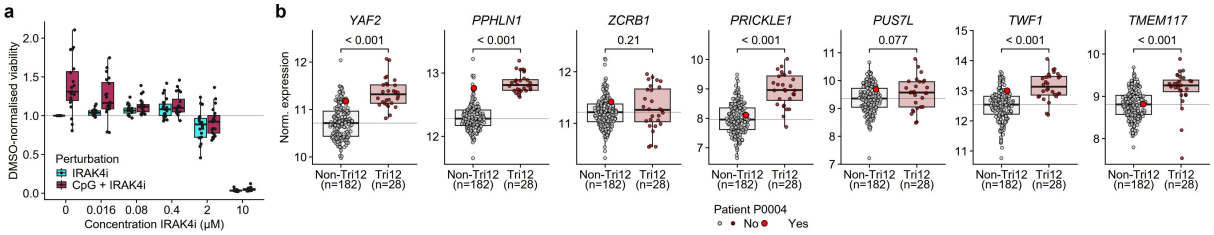
a, Normalised expression of *ZAP70 and CD38* by IGHV status in DMSO or BTK inhibitor-treated samples. **b**, Normalised expression of *CCL3* and *CD72* by IGHV status in DMSO or BTK inhibitor-treated samples. Significance levels were calculated using two-tailed Student's *t*-test and shown as nominal *P*-values. **c**, GSEA with genes differentially expressed between U- and M-CLL (10% *FDR* in > 90% of bootstrapping performed to increase stringency) for each treatment using hallmark and canonical pathways gene sets. Gene sets were filtered by *FDR* (< 1%) and number of enriched genes (> 10). **d**, Differential expression of IGHV signature genes within the IFN response and **e**, ribosome gene set with BTKi, PI3Ki, mTORi, IRAK4i and BTKi/IRAK4i compared to DMSO. **f**, Expected independent additive effect from single BTK and IRAK4 inhibition against the observed effect from the BTK/IRAK4 inhibitor combination for down- and up-regulated BTK, IRAK4 and BTK/IRAK4 inhibitor-dependent genes (black dots). The expected independent effect was calculated using the Bliss independence model ($LFC_{expected} = LFC_{BTKi} + LFC_{IRAK4i} - LFC_{BTKi} * LFC_{IRAK4i}$) and compared to the observed effect ($LFC_{BTKi/IRAK4i}$). The gray dots depict the background distribution of expected additive effect for

all genes (n_{Down} = 8,821, n_{Up} = 8,040). Shaded areas indicate increased synergistic (blue) or antagonistic (red) effects for BTK, IRAK4 and BTK/IRAK4 inhibitor-dependent genes compared to all genes. **g**, Normalised expression of dynamic BCR-dependent genes MTOR and **h**, CD81, and **i**, mTOR-dependent gene *CSTB* divided by trisomy 12 status in DMSO or drug-treated samples. **j**, Normalised expression of MEK-induced trisomy 12 signature gene *CD19* divided by trisomy 12 status in DMSO or MEK inhibitor-treated samples. Significance levels were calculated using two-tailed Student's *t*-test and shown as nominal *P*-values.



Suppl. Fig. 8: Interaction of perturbation effects and disease drivers

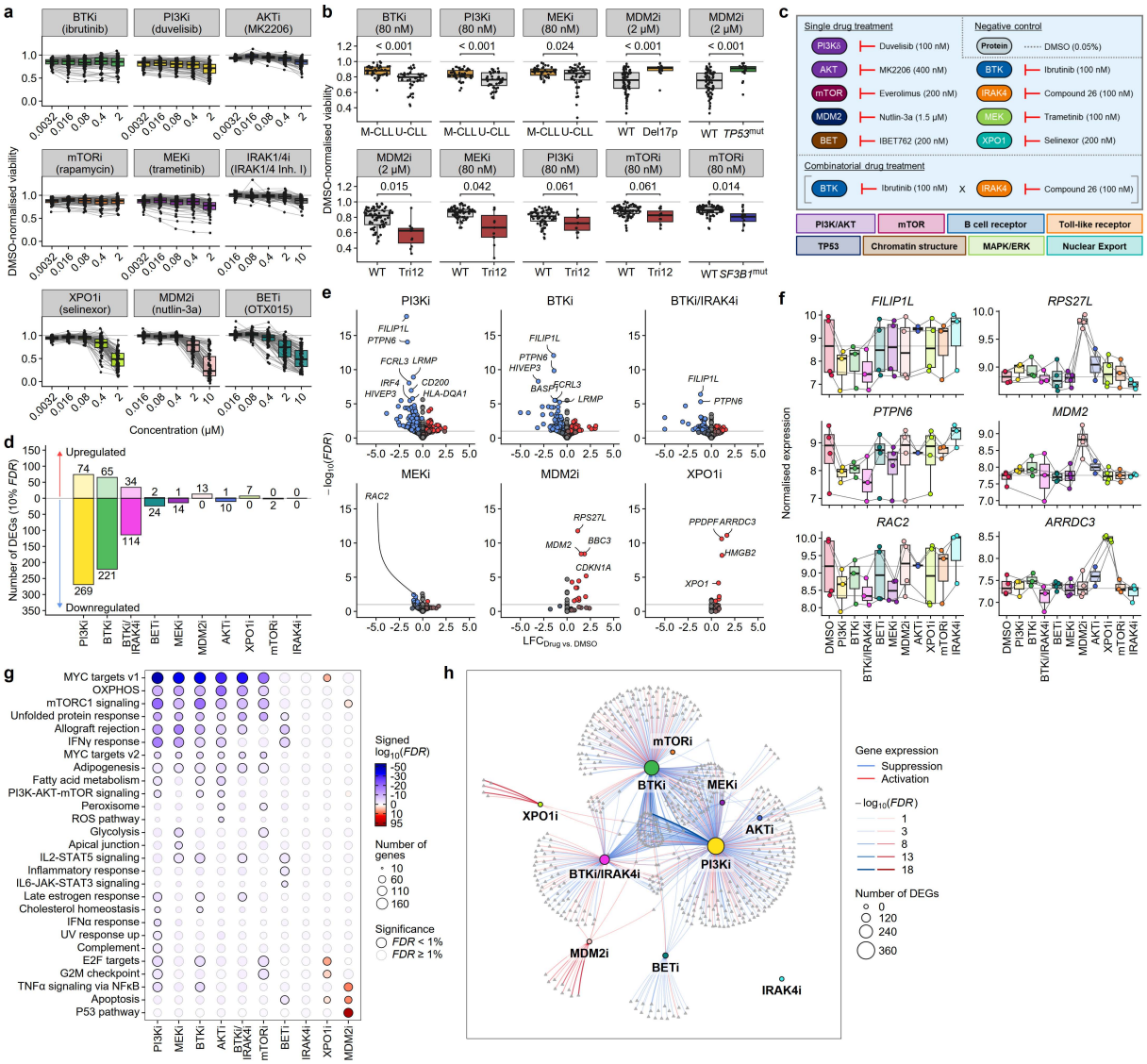
a, Pathway connectivity network of all DGI genes shown as a bipartite graph. Genes, drugs and disease drivers are connected if a differential drug response was observed in the presence of the mutation (*FDR* < 10%). **b**, Cluster assignment of DGIs based on gene expression in the presence of either drug, mutation/CNV or both compared to WT/DMSO samples by Fuzzy C-Means Clustering. The thick black line depicts the median LFC of all genes from the respective cluster. **c**, Heatmap of relative gene expression of DGIs in the presence of either drug, mutation/CNV or both compared to WT/DMSO samples by epistasis class. **d**, Number of target genes with significant DGIs for each perturbation, and all recurrent drivers with more than 5 cases by epistasis class (*FDR* < 10%).



Suppl. Fig. 9: IRAK4 is a disease driver targeted by trisomy 12

a, Dose-response curves of 20 patients treated with IRAK4 inhibitor and IRAK4 inhibitor combined with 1 μg/mL CpG-ODN for TLR9 stimulation. **b**, Gene expression of genes on the amplified locus chr12:42,538,673-chr12: 44,229,999 (hg19) by trisomy 12 status for the CLL cohort (n = 210) from Lütge et al.[13]. Significance levels were calculated using two-tailed Student's *t*-test and shown as nominal *P*-values.

Supplementary methods figures



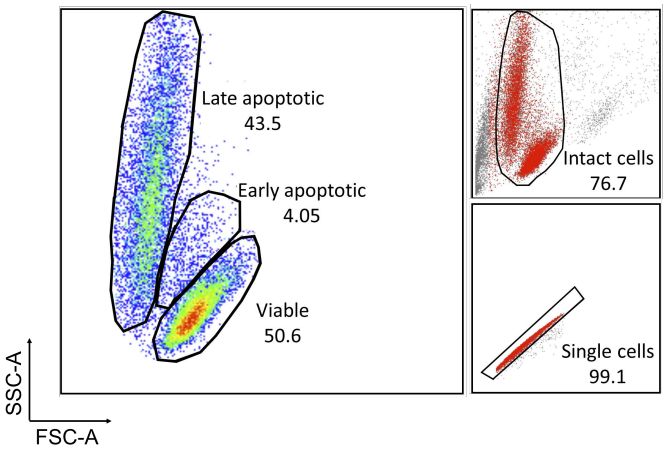
Suppl. Methods Fig. 1: Design and pilot drug screening using shallow QuantSeq

a, Dose-response curves of drug response in CLL patients (n = 88) (viability drug screen from Lu et al.[12]). **b**, Significant genetic association with drug response in CLL samples (*FDR* < 10%) for aberration with more than 5 cases. Significance levels were calculated using two-tailed Student's *t*-test and adjusted for multiple testing. **c**, Overview of treatments and concentrations. **d**, Number of DEGs between 4 drug- and DMSO-treated CLL samples. **e**, Volcano plot summarising the differential gene expression after PI3K, BTK, combined BTK/IRAK4, MEK, MDM2 and XPO1 inhibitor treatment. **f**, Gene expression of exemplary genes with treatment-specific down- and upregulation. **g**, GSEA with DEGs after each

900 treatment (all genes) using hallmark gene sets. Gene sets were filtered by *FDR* (< 1%) and
901 number of enriched genes (> 10). **h**, Pathway connectivity network shown as a bipartite
902 graph. Genes and drugs are connected if the drug perturbation leads to differential
903 expression (10% *FDR*).

904

905

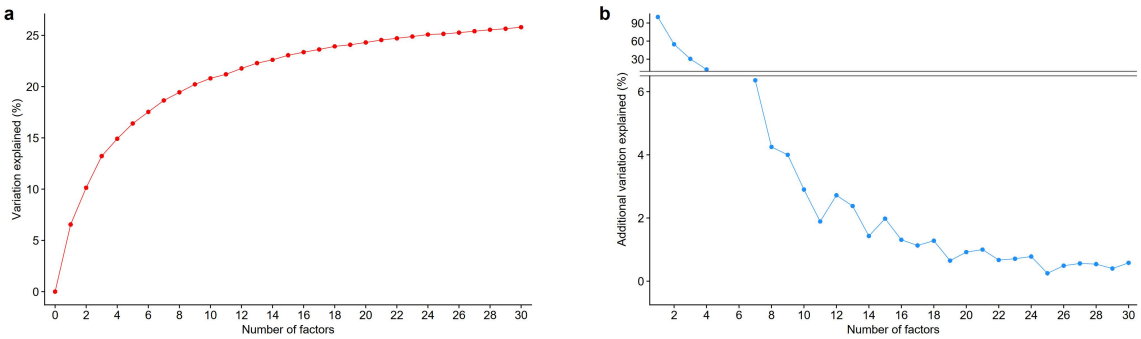


906

907 **Suppl. Methods Fig. 2 Gating strategy for flow cytometric assessment of cell viability**
908 Cells were pre-gated by excluding debris and doublet cells. Viable and dead cells were then
909 gated based on cell size (FSC-A) and granularity (SSC-A).

910

911



912

913 **Suppl. Methods Fig. 3 Guided sparse factor analysis model evaluation**
914 **a**, Proportion of variation explained by each GSFA model with 1 to 30 factors. **b**, Percentage
915 of additional variation explained by each GSFA model compared to the previous model with
916 the highest explained variation. The model with 20 factors was selected for further analysis.