

## **Dissecting intratumor heterogeneity of nodal B cell lymphomas on the transcriptional, genetic, and drug response level**

Tobias Roider<sup>1-3</sup>, Julian Seufert<sup>4-5</sup>, Alexey Uvarovskii<sup>6</sup>, Felix Frauhammer<sup>6</sup>, Marie Bordas<sup>4,7</sup>, Nima Abedpour<sup>8</sup>, Marta Stolarczyk<sup>1</sup>, Jan-Philipp Mallm<sup>9</sup>, Sophie Rabe<sup>1-3,5,10</sup>, Peter-Martin Bruch<sup>1-3</sup>, Hyatt Balke-Want<sup>11</sup>, Michael Hundemer<sup>1</sup>, Karsten Rippe<sup>9</sup>, Benjamin Goeppert<sup>12</sup>, Martina Seiffert<sup>7</sup>, Benedikt Brors<sup>13</sup>, Gunhild Mechtersheimer<sup>12</sup>, Thorsten Zenz<sup>14</sup>, Martin Peifer<sup>8</sup>, Björn Chapuy<sup>15</sup>, Matthias Schlesner<sup>4</sup>, Carsten Müller-Tidow<sup>1-3</sup>, Stefan Fröhling<sup>10,16</sup>, Wolfgang Huber<sup>2,3</sup>, Simon Anders<sup>6\*</sup>, Sascha Dietrich<sup>1-3,10\*</sup>

<sup>1</sup> Department of Medicine V, Hematology, Oncology and Rheumatology, University of Heidelberg, Heidelberg, Germany,

<sup>2</sup> Molecular Medicine Partnership Unit (MMPU), Heidelberg, Germany,

<sup>3</sup> European Molecular Biology Laboratory (EMBL), Heidelberg, Germany,

<sup>4</sup> Bioinformatics and Omics Data Analytics, German Cancer Research Center (DKFZ), Heidelberg, Germany,

<sup>5</sup> Faculty of Biosciences, University of Heidelberg, Heidelberg, Germany,

<sup>6</sup> Center for Molecular Biology of the University of Heidelberg (ZMBH), Heidelberg, Germany,

<sup>7</sup> Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany,

<sup>8</sup> Department for Translational Genomics, University of Cologne, Cologne, Germany.

<sup>9</sup> Division of Chromatin Networks, German Cancer Research Center (DKFZ) & Bioquant, Heidelberg, Germany.

<sup>10</sup> Department of Translational Medical Oncology, National Cen

German Cancer Research Center (DKFZ), Heidelberg, Germany,

<sup>11</sup> Department I of Internal Medicine, University Hospital of Cologne, Cologne, Germany,

<sup>12</sup> Institute of Pathology, University of Heidelberg, Heidelberg, Germany,

<sup>13</sup> Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>14</sup> Department of Medical Oncology and Hematology, University of Zürich, Zürich, Switzerland,

<sup>15</sup> Clinic for Hematology and Medical Oncology, University Med

\* These authors contributed equally.

## **Corresponding Author**

Sascha Dietrich, MD PhD

Email address: sascha.dietrich@embl.de

Mailing address: Department of Medicine V, U

Im Neuenheimer Feld 410

69120 He

39 Fax: +49 6221 56 4049

40 **Contributions**

41 T.R., M.B., M.St., J.P.M., S.R. and P.M. B. performed experiments. T.R., J.S., A.U., F.F., M.B.,  
42 N.A., H.BW., M.P., M.Sc. and S.A. analyzed the data. T.R., M.H., K.R., B.G., M.Se., B.B.,  
43 G.M., C.MT., S.F., W.H., S.A. and S.D. interpreted the data. T.R., T.Z., S.F. and S.D. designed  
44 the study. T.R., J.S., K.R., B.C., M.Sc., W.H., S.A. and S.D wrote the paper

45

46 **Acknowledgements**

47 T.R. was supported by a physician scientist fellowship of the Medical Faculty of University  
48 Heidelberg. M.Se. was supported by a grant of the Deutsche Forschungsgemeinschaft (DFG).  
49 S.D. was supported by a grant of the Hairy Cell Leukemia Foundation, the Heidelberg Research  
50 Centre for Molecular Medicine (HRCMM) and an e:med BMBF junior group grant. For the  
51 data management we thank the Scientific Data Storage Heidelberg (SDS@hd) which is funded  
52 by the state of Baden-Württemberg and a DFG grant (INST 35/1314-1 FUGG). We thank  
53 Carolin Kolb (University Hospital Heidelberg) and Mareike Knoll (German Cancer Research  
54 Center Heidelberg) for their excellent technical assistance. We also thank the DKFZ Single-  
55 Cell Open Lab (scOpenLab) for the experimental assistance in terms of scRNA-seq. Also, this  
56 study was supported by the Heidelberg Center for Personalized Oncology (DKFZ-HIPO). We  
57 thank the DKFZ Omics IT and Data Management Core Facility (ODCF) and the DKFZ  
58 Genomics and Proteomics Core Facility (GPCF) for their technical support.

59 **Abstract**

60 Tumor heterogeneity encompasses both the malignant cells and their microenvironment. While  
61 heterogeneity between individual patients is well-known to affect the efficacy of anti-cancer  
62 drugs, most personalized treatment approaches do not account for intratumor heterogeneity. We  
63 addressed this issue by studying the heterogeneity of lymph node-derived B cell non-Hodgkin  
64 lymphoma (B-NHL) by single cell RNA-sequencing (scRNA-seq) and transcriptome-informed  
65 flow cytometry. We identified transcriptionally distinct malignant subclones and compared  
66 their drug response and genomic profiles. Malignant subclones of the same patient responded  
67 strikingly different to anti-cancer drugs *ex vivo*, which recapitulated subclone-specific drug  
68 sensitivity during *in vivo* treatment. Tumor infiltrating T cells represented the majority of non-  
69 malignant cells, whose gene expression signatures were similar across all donors, whereas the  
70 frequencies of T cell subsets varied significantly between the donors. Our data provide new  
71 insights into the heterogeneity of B-NHL and highlight the relevance of intratumor  
72 heterogeneity for personalized cancer therapies.

73 **Introduction**

74 The genomic and transcriptional landscape of many cancer entities has been catalogued over  
75 recent years, documenting the range of tumor heterogeneity between individual patients [1]. In  
76 addition, it has long been appreciated that tumors within each patient consist of diverse, but  
77 phylogenetically-related subclones [2]. Bulk sequencing studies of tumor cells have been  
78 conducted to infer the genetic spectrum of intratumor heterogeneity from variant allele  
79 frequencies of somatic mutations [3]. While important insights were gained from these studies,  
80 further characterization on the single cell level is needed to more accurately dissect the pathway  
81 and molecular properties associated with distinct subclones.

82 Neoplastic cells alone do not manifest a malignant disease, but attract a battery of non-  
83 malignant bystander cells, which support tumor cell growth and survival. The diversity and  
84 plasticity of the microenvironment constitutes another layer of heterogeneity, beyond the  
85 heterogeneity of the cancer cells themselves [4]. There is solid evidence that intratumor  
86 heterogeneity among malignant and non-malignant cells, and their interactions within the tumor  
87 microenvironment are critical to diverse aspects of tumor biology, response to treatment, and  
88 prognosis [5].

89 While bulk genomic tissue profiling has only a limited ability to reconstruct the complex  
90 cellular composition of tumors, single cell DNA-sequencing [6, 7] and RNA-sequencing  
91 (scRNA-seq) methods [8-11] have emerged as powerful tools to study intratumor heterogeneity  
92 and reconstruct the full picture of malignant and non-malignant cells. These technologies  
93 further enable researchers to identify rare cell types such as cancer stem cells [12] and  
94 circulating tumor cells [13, 14], or to follow clonal dynamics during cancer treatment [15].  
95 Most of these single cell studies have been used to describe distinct cell subpopulations on the  
96 transcriptional level, but their functional properties, such as drug response profiles, remain  
97 largely unexplored.

98 To address this, we used B cell non-Hodgkin lymphoma (B-NHL) as a model disease entity to  
99 dissect intratumor heterogeneity on the transcriptional, genetic, and functional (drug response)  
100 level. In parallel, we investigated the cellular heterogeneity of the B-NHL lymph node  
101 microenvironment. B-NHL are a heterogeneous group of hematologic malignancies that most  
102 frequently grow in the lymph node compartment. Almost half of all B-NHL are classified as  
103 diffuse large B cell lymphoma (DLBCL) or follicular lymphoma (FL) [16]. Transformation of  
104 indolent FL into aggressive DLBCL is observed in approximately 10% of all FL cases [17].  
105 Despite effective treatment options, 20-40% of B-NHL patients relapse multiple times and  
106 present with chemotherapy refractory disease [18, 19]. The response to single agent targeted

107 therapy in these patient cohorts is surprisingly low [20, 21]. Intratumor heterogeneity might be  
108 a key factor contributing to therapeutic failure and low success rate of these single agent  
109 targeted therapies [3]. Understanding subclonal drug response patterns would therefore be an  
110 important asset for designing more effective personalized lymphoma therapies.

111 To dissect the complex cellular composition of the malignant lymph node niche, we profiled  
112 transcriptomes of malignant and non-malignant cells derived from 12 different reactive or B-  
113 NHL lymph node biopsies. We further studied the variation of the cellular composition of the  
114 malignant lymph node niche by flow cytometry in a larger cohort of 41 patients. Among  
115 malignant cells, we identified transcriptionally distinct malignant subclones and characterized  
116 these subclones further by ex-vivo drug perturbation and genome sequencing. This revealed  
117 new insights into intratumor heterogeneity of B-NHL and demonstrated substantially different  
118 drug responses between malignant subclones in the same patient.

119

120 **Results**

121 ***Study outline***

122 We designed an experimental pipeline to dissect the heterogeneity of non-malignant and  
123 malignant lymph node-derived lymphocytes (Figure 1A). This involved first preparing single  
124 cell suspensions of B-NHL lymph node biopsies and performing scRNA-seq. These single cell  
125 transcriptomic data were then used to identify transcriptionally-distinct subclones by flow  
126 cytometry using distinguishing subclone-specific surface markers, and finally the subclones  
127 were functionally interrogated in drug perturbation assays with a comprehensive panel of 58  
128 drugs in five concentrations, and further characterized by whole genome (WGS) and/or exome  
129 sequencing (WES).

130

131 ***Dissecting the cellular composition of nodal B cell lymphomas***

132 We assayed single cell suspensions of a total of 12 samples: four germinal center-derived  
133 diffuse large B cell lymphoma (DLBCL) samples, of which two were transformed from FL  
134 (DLBCL1, DLBCL2, tFL1, tFL2), one non-germinal center-derived DLBCL (DLBCL3), four  
135 follicular lymphoma samples (FL1, FL2, FL3, FL4), and three reactive non-malignant lymph  
136 node sample (rLN1, rLN2, rLN3) by flow cytometry and droplet-based scRNA-seq  
137 (Supplementary Table 1). After removal of low-quality cells, we analyzed scRNA-seq profiles  
138 of 13,259 malignant and 9,296 non-malignant cells with an average sequencing depth of 1,409  
139 genes per cell.

140 First, we verified that the lymph node-derived single cell suspensions were representative for  
141 the cellular composition (B and T cells) of the lymphoma and its microenvironment in vivo.  
142 We used sections of paraffin-embedded samples of the same lymph nodes, which were formalin  
143 fixed directly after surgical excision and therefore represent the in vivo cellular composition,  
144 and quantified B and T cell frequencies by immunohistochemistry (IHC, Supplementary Figure  
145 1A). In parallel, we calculated B and T cell frequencies by flow cytometry and scRNA-seq in  
146 single cell suspensions (Supplementary Figure 1A, B). The frequencies of B and T cells derived  
147 from scRNA-seq correlated very well with the frequencies determined by flow cytometry  
148 ( $r = 0.97$ ,  $n = 12$ , Figure 1B) and IHC ( $r = 0.92$ ,  $n = 7$ , Figure 1C).

149 Next, we aimed to distinguish malignant from non-malignant B cells and delineate these  
150 populations in our single cell experiments. We took advantage of the fact that malignant B cell  
151 populations express only one type of immunoglobulin light chain (LC), either  $\kappa$  or  $\lambda$  [22]. We  
152 calculated the LC-ratio ( $\kappa/\lambda$ ) based on RNA expression of the genes *IGKC* (coding for the  
153 constant part of the  $\kappa$  LC) and *IGLC2* ( $\lambda$  LC) for each single B cell and color-coded this ratio

154 in a t-distributed stochastic neighbor embedding (t-SNE) plot (Figure 1D, E). In the malignant  
155 lymph nodes, we could either identify both a non-malignant and malignant or only malignant  
156 B cell clusters (Supplementary Figure 2). In contrast, reactive lymph node samples contained  
157 only non-malignant B cells (see method section for details).

158 We further evaluated the frequencies of these subsets in a larger cohort of 41 lymph node  
159 samples by flow cytometry, including those samples used for scRNA-seq. Both approaches  
160 showed very similar frequencies of these cell subsets ( $r = 0.97$ ,  $n = 12$ ,  
161 Supplementary Figure 3A, B). We found that the proportion of malignant cells was highly  
162 variable across samples. It ranged from 14.6 to 97.2 % (median 79.3 %,  $n = 9$ ) in DLBCL,  
163 23.7 to 85.4 % (median 79.9 %,  $n = 12$ ) in FL, 48.4 to 95.5 % (median 88.0 %,  $n = 4$ ) in mantle  
164 cell lymphoma, and 65.4 to 91.4 % (median 83.1 %,  $n = 7$ ) in chronic lymphocytic leukemia  
165 (Supplementary Figure 3C). This substantial cellular heterogeneity complicates bulk  
166 sequencing approaches of unsorted lymph node samples, and highlights the value of single cell  
167 sequencing to simultaneously study the full spectrum of malignant and non-malignant lymph  
168 node cells.

169

### 170 ***Characterization of lymph node-derived T cell populations***

171 T cells are key players of the host-specific tumor immunosurveillance [23]. B-NHL exhibit  
172 genetic immune escape strategies that can be targeted using current therapeutic strategies [24],  
173 including checkpoint inhibitors [25] and bispecific antibodies [26]. Notably, lymphoma cells  
174 can also orchestrate their tumor microenvironment so that certain T cell subsets support the  
175 growth and proliferation of the tumor cells [27]. Even though these subsets have been  
176 extensively studied by immunophenotyping, their transcriptional heterogeneity in B-NHL  
177 lymph nodes, in particular at the single cell level, still remains to be elucidated.

178 We combined single cell RNA expression profiles of T cells from all 12 donors and jointly  
179 visualized them by Uniform Manifold Approximation and Projection (UMAP), a dimension  
180 reduction algorithm alternative to t-SNE [28]. Many well-established surface markers, which  
181 are used to distinguish T cell subsets in flow cytometry studies, are insufficiently expressed on  
182 the scRNA-seq level. We therefore chose unsupervised clustering to partition T cells into  
183 transcriptionally distinct subsets, which were then annotated by differentially expressed marker  
184 genes. All T cells from either reactive or malignant lymph nodes distributed to only four major  
185 T cell subpopulations (Figure 2 A, B). Note, that clusters were not driven by patients or disease  
186 entity, suggesting only limited transcriptional heterogeneity across all donors. Apart from  
187 conventional T helper cells ( $T_H$ ; *CD4*, *IL7R*, *PLAC8*, *KLF2*) and regulatory T cells ( $T_{REG}$ ; *CD4*,

188 *IL2RA, FOXP3, ICOS*), we identified a third T helper cell population, which was characterized  
189 by overexpression of *PDCD1* (PD1), *ICOS*, *CXCR5*, *TOX*, *TOX2* and *CD200* (Figure 2C,  
190 Supplementary Table 2), suggesting a T follicular helper cell ( $T_{FH}$ ) phenotype [29-33]. In  
191 contrast to the diversity of T helper cells, we observed only one major cluster of cytotoxic T  
192 cells ( $T_{TOX}$ , *GZMK*, *CCL4/5*, *GZMA*, *NKG7*, *CD8A*). However, the frequencies of the four  
193 identified T cell subsets were highly variable between different B-NHL donors (Figure 2D).  
194 To study this variation in a larger cohort, we quantified the abovementioned T cell populations  
195 in 39 lymph node samples of DLBCL, FL, mantle cell lymphoma and chronic lymphocytic  
196 leukemia by flow cytometry using the most distinctive markers (CD3, CD4, CD8, CD25,  
197 FoxP3, ICOS, PD1), as seen in Figure 2C. The frequencies of all T cell subsets derived from  
198 scRNA-seq correlated well with the frequencies determined by flow cytometry ( $r = 0.69$ ,  
199  $n = 10$ , Figure 2E). We found that  $T_{FH}$  cells were significantly increased in FL (two-sided  
200 Wilcoxon test:  $p = 0.006$ , Figure 2F), and  $T_{REG}$  cell frequencies were significantly increased in  
201 malignant lymph nodes, compared to the reactive ones (two-sided Wilcoxon test:  $p$  values as  
202 indicated, Figure 2F).

203 Taken together, we demonstrated that T cells derived from malignant B-NHL lymph nodes are  
204 transcriptionally similar to those derived from non-malignant reactive lymph nodes. In contrast,  
205 the proportion of individual T cell subsets differed significantly between lymphoma entities and  
206 individual patients. This finding indicates that B-NHL shape their microenvironment by  
207 influencing the recruitment of certain T cell subpopulations, but have less effect on their  
208 transcriptional programs. Therefore, studying the frequencies of lymphoma infiltrating T cell  
209 subsets and their effect on the outcome after immunotherapies might be highly relevant for the  
210 development of biomarkers.

211

## 212 ***Identification of gene expression signatures driving B cell heterogeneity by scRNA-seq***

213 Next, we examined the heterogeneity of the malignant and non-malignant B cells. To gain a  
214 global overview of the gene expression pattern across all malignant and non-malignant B cells  
215 from the 12 different donors, we combined their single cell RNA expression profiles, clustered  
216 them jointly and visualized them by UMAP (Figure 3A, B).

217 Clustering partitioned the non-malignant B cells into two distinct subpopulations (C0-C1,  
218 Figure 3A). Among multiple differentially expressed genes between these two subsets  
219 (Supplementary Table 3), we found *IGHM* and *CD72* to be overexpressed in cluster C0, which  
220 characterizes naïve B cells [34], and *CD27* and *IGHG1* to be overexpressed in cluster C1, which  
221 characterizes memory B cells [35].

222 Of the eight transcriptionally distinct clusters formed by the malignant B cells (C2-C9, Figure  
223 3A), six exclusively contained cells of only one donor (Figure 3A, B). This suggests a higher  
224 degree of inter-patient heterogeneity for malignant than for non-malignant B cells. We  
225 performed a gene set enrichment analysis (GSEA) on the mean expression differences between  
226 each malignant B cell cluster and all non-malignant cells, which revealed multiple cluster-  
227 specific gene sets (Figure 3C). Germinal center (GC)-associated gene expression signatures  
228 were significantly enriched in all clusters except for cluster 6, which exclusively contained  
229 malignant B cells of DLBCL3. This finding supports the classification of all B-NHL cells as  
230 either GCB type DLBCL or FL, except for the remaining DLBCL3 sample, which was  
231 classified as a non-GCB type DLBCL based on the Hans-classifier (Supplementary  
232 Table 1) [36]. Individual clusters were characterized by oncogenic transcriptional programs,  
233 which indicated activation of oncogenic MYC or STK33 signaling (Figure 3D).

234 Inter-patient heterogeneity of B cell lymphomas also comprises their proliferative capacity,  
235 which can vary from very low in FL to very high in DLBCL. We determined the proportion of  
236 B cells in S, G<sub>2</sub> or M phase based on their single cell RNA profile (Supplementary Figure 4A)  
237 and observed a high correlation with flow cytometry- and IHC-based staining of Ki67 ( $R = 0.83$   
238 scRNA-seq to flow cytometry,  $R = 0.92$ , scRNA-seq to IHC, Supplementary Figure 4B).

239 In summary, these results indicate that inter-patient heterogeneity of malignant B cells,  
240 including their diverse proliferative activity, can be captured by the scRNA-Seq and can be  
241 linked to lymphoma-specific transcription signatures. Non-malignant B cells, however, had  
242 similar transcriptional profiles across different donors.

243

244 ***Decoding the crosstalk between T cells and malignant B cells in the lymph node  
245 microenvironment***

246 Above, we concluded that B cell lymphomas shape their microenvironment by modulating the  
247 frequency of different subsets of lymphoma infiltrating T cells. We now aimed to understand  
248 through which potential ligand-receptor interactions malignant B cells could benefit from their  
249 specific T cell microenvironment. For this purpose, we adopted a computational approach  
250 described by Vento-Tormo et al. [37] and analyzed 760 known ligand-receptor combinations  
251 (Supplementary Table 4) to identify the most significant interactions between malignant B cells  
252 and lymphoma infiltrating T cells within the lymph node microenvironment (Figure 4A).  
253 This analysis suggested that malignant B cells could receive costimulatory and coinhibitory  
254 signals by all four major T cell subsets, via CD80/CD86-CD28 and CD80/CD86-CTLA, while  
255 interactions via BCMA-BAFF, BAFF-R-BAFF and CD40-CD40LG could predominantly be

256 mediated by T<sub>H</sub> or T<sub>REG</sub> cells. Significant interaction scores of IL4-IL4R and IL4-IL13RA1  
257 were exclusively observed between T<sub>FH</sub> and malignant B cells, providing further evidence that  
258 T<sub>FH</sub> cells represent the most important source of IL4 production in B cell lymphoma [38]. This  
259 observation might be of clinical relevance because the IL4/IL4R interaction is discussed as  
260 potential resistance mechanism against Bruton's tyrosine kinase (BTK) inhibitors [39, 40]. In  
261 line with the current state of knowledge [41-43], we also observed strong interaction scores for  
262 T<sub>FH</sub> via IL21-IL21R with malignant B cells and via IL2-IL2R with other T cell subsets. This  
263 analysis supports the classification of T<sub>FH</sub> cell as one of the four main T cell subsets within the  
264 lymph node microenvironment and reveals that each subset may provide a distinct panel of  
265 stimuli to interact with malignant B cells.

266

#### 267 ***Dissecting transcriptional intratumor heterogeneity using multicolor flow cytometry***

268 Intratumor heterogeneity of nodal B cell lymphoma is a well-known phenomenon, however,  
269 most available studies infer intratumor heterogeneity from variant allele frequencies of genetic  
270 alterations corrected for purity, ploidy and multiplicity of local copy number [44, 45].  
271 Here, we aimed to investigate the genomic, transcriptomic and functional (drug response) layers  
272 of intratumor heterogeneity from single cells. Unsupervised clustering of scRNA-seq profiles  
273 of malignant and non-malignant B cells revealed that all malignant samples were composed of  
274 at least two or more transcriptionally distinct subclusters (Supplementary Figure 5). We aimed  
275 to validate scRNA-based clusters at the cellular level to understand if this clustering represents  
276 biologically and clinically relevant differences. Therefore, we selected three samples (FL4,  
277 tFL1, DLBCL1) based on the availability of material for follow-up studies. We inferred  
278 differentially expressed surface markers from single cell expression profiles and first validated  
279 the distinction of scRNA-based clusters by flow cytometry. In a second step, we cultured lymph  
280 node derived lymphocytes with 58 different drugs in 5 concentrations (Supplementary Table 5)  
281 and stained them with specific antibody combinations to assess their drug response profiles by  
282 flow cytometry. In a third step, we sorted subpopulation and performed genome sequencing for  
283 each subclone (tFL1, DLBCL1).

284

#### 285 ***Verifying five transcriptionally distinct clusters in follicular lymphoma sample***

286 The FL4 sample was collected at initial diagnosis. Based on single cell gene expression  
287 profiling, we identified five different B cell subpopulations (Supplementary Figure 6A). We  
288 aimed to validate all five clusters (C1 to C5) at the cellular level by flow cytometry and hence,  
289 we stained the differentially expressed surface markers CD44, CD24, CD22, CD27, kappa and

lambda light chain (encoded by *IGKC* and *IGLC2*, Supplementary Figure 6B). Using the ratio of *IGKC* and *IGLC2* (see Methods for details), we found benign B cells in C1, lambda-restricted malignant B cells in C2, and malignant B cells with only marginal expression of *IGKC* and *IGLC2* in C3 to C5 (Supplementary Figure 6C). The pattern of light chain expression could be perfectly comprehended using flow cytometry (Supplementary Figure 6D), enabling us to differentiate C1 versus C2 versus C3, C4 and C5. Cluster C3 could then be recognized by a high expression of CD44 (Supplementary Figure 6D, 6E). To further distinguish C4 and C5 among the CD44<sup>Low</sup> cells, we combined CD22, CD27 and CD24 and detected a subpopulation with CD22<sup>High</sup>, CD27<sup>High</sup> and CD24<sup>Low</sup>, which corresponded to the expression pattern of cluster C5 (Supplementary Figure 6F). This approach allowed us to proof all five scRNA-based clusters by flow cytometry with comparable frequencies.

To assess subclone-specific drug response, we stained for kappa and lambda light chains and focused on the two major populations (C2  $\cong$  lambda<sup>+</sup>, C3-C5  $\cong$  kappa/lambda<sup>-</sup>). We did not observe differential responses for the majority of targeted drugs, but we found that only the kappa/lambda<sup>-</sup> cluster was sensitive to chemotherapeutics (Supplementary Figure 6G). Interestingly, this patient received doxorubicin-based immunochemotherapy as first line treatment after sample collection and achieved only a partial remission.

307

308 ***The indolent and aggressive component of transformed follicular lymphoma exhibit a  
309 distinct transcriptional, genomic and drug response profile***

310 For the tFL1 sample, we detected three transcriptionally distinct clusters of B cells based on  
311 single cell RNA expression profiling (Figure 5A, B). Two clusters exclusively contained  
312 malignant B cells, and one cluster contained non-malignant B cells. We assessed the  
313 proliferative activity of both malignant populations based on their gene expression profiles, and  
314 observed that only one malignant cluster contained cells in S phase (Supplementary Figure 7A),  
315 with no cells in G<sub>2</sub> or M phase (Supplementary Figure 7B). This suggests that this cluster  
316 represents a proliferating, thus aggressive component of the transformed FL. We performed a  
317 GSEA on the mean expression differences between the two malignant clusters, which revealed  
318 that gene expression signatures associated with MYC, MTORC1, and the G<sub>2</sub>M transition [46]  
319 were significantly enriched in the presumptively aggressive malignant B cell subclone  
320 (Supplementary Figure 7D-F).

321 Among the genes differentially expressed between both subclones, we found *FCGR2B* (Figure  
322 5A), which encodes a surface receptor protein (CD32B), to be exclusively expressed in the  
323 presumptively indolent subclone. Thus, we confirmed the existence of three B cell populations

324 by flow cytometry (Figure 5C, see Supplementary Figure 8A for complete gating strategy).  
325 CD10 was strongly positive in both malignant B cell populations ( $CD32^{\text{High}}$ ,  $CD32^{\text{Low}}$ ), but not  
326 in non-malignant B cells.

327 As described above, we measured the ex vivo drug responses separately for each subclone  
328 (Figure 5D, E) and observed very different drug response profiles for the two malignant  
329 subclones. The BTK inhibitors, ibrutinib, acalabrutinib, and tirabrutinib, and the  
330 immunomodulatory imide drugs (pomalidomide, lenalidomide), were exclusively active in the  
331  $CD32^{\text{Low}}$  subclone, whereas HDAC inhibitors (panobinostat, romidepsin, vorinostat) were  
332 more active in the  $CD32^{\text{High}}$  subclone.

333 Based on CD32 and CD10 expression, we sorted the three B cell subclones by flow cytometry  
334 (Supplementary Figure 8A) and performed WES of peripheral blood-derived normal control  
335 DNA, whole tumor DNA, DNA of both malignant subclones, and DNA of the non-malignant  
336 B cell population. Copy number profiles of both malignant subclones were very different,  
337 including exclusive aberrations of chromosomes 3, 4, 6, 10, 12, 15, 18 and X (Supplementary  
338 Figure 7F). Only the  $CD32^{\text{Low}}$  subclone harbored a trisomy 12 (Figure 5F), which was  
339 confirmed by scRNA-seq data (Figure 5G). Trisomy 12 has been associated with a better  
340 response to B cell receptor (BCR) signaling inhibitors [47], which was consistent with our  
341 observation that this subclone was more responsive to these drugs (Figure 5D, E). We also  
342 detected 157 somatic single nucleotide variants (SNV) in exonic regions, of which 25 (15.9 %)  
343 or 24 (15.2%) were exclusively detected in the  $CD32^{\text{High}}$  or  $CD32^{\text{Low}}$  subclone, respectively  
344 (Figure 5H, I, Supplementary Table 6). However, the majority of somatic SNVs were equally  
345 represented in both subclones, indicating a phylogenetic relationship. We compared the allele  
346 count of all exonic SNVs between all three B cell populations and did not detect somatic SNV  
347 in healthy B cells (Figure 5J), which supports the validity of our sorting approach.

348 Taken together, scRNA-seq allowed us to identify different subclones within the same lymph  
349 node, which were genetically and functionally distinct in clinically-relevant aspects.

350

351 ***A subclone-specific copy number variation of MYC drives a distinct gene expression and***  
352 ***drug response program***

353 The DLBCL sample was collected from a patient with a chemotherapy refractory disease  
354 during progression, but before retreatment. Using scRNA-seq, we identified two distinct  
355 clusters of malignant B cells, which exhibited a high number of differentially expressed genes  
356 associated with diverse cellular programs (Figure 6A, B), such as BCR signaling (*PRKCB*,  
357 *NFKB1I*), cytokine signaling (*LGALS9*, *IFITM1*), MAPK signaling (*RGS13*, *FBLN5*) and

358 antigen processing (*PTPN22*, *SELL*, *CD48*). Among the differentially expressed genes, we  
359 found *CD48* and *SELL* (Supplementary Figure 9A, B), which encode for the surface markers  
360 *CD48* and *CD62L* respectively. Staining for *CD48* and *CD62L* by flow cytometry validated the  
361 existence of the two distinct subclones (Figure 6C). The proportions of both clusters  
362 ( $CD48^{\text{High}}CD62L^+$ ,  $CD48^{\text{Low}}CD62L^-$ ) calculated based on flow cytometry and scRNA-seq were  
363 comparable, indicating good concordance between RNA and protein expression. We measured  
364 again the ex vivo drug responses for each subclone (Figure 6D, E) and observed a strikingly  
365 different drug response profile between the two subclones: B cell receptor (BCR) signaling  
366 inhibitors (acalabrutinib, tirabrutinib, ibutinib, duvelisib, idelalisib, entospletinib) and CDK  
367 inhibitors were exclusively effective in the  $CD48^{\text{Low}}CD62L^-$  subclone, whereas Bromodomain  
368 and Extra-Terminal motif (BET) inhibitors (I-BET-762, OTX015), nucleoside analogues  
369 (cytarabine, fludarabine, cladribine) and vincristine were exclusively efficacious in the  
370  $CD48^{\text{High}}CD62L^+$  subclone.

371 We sorted viable tumor cells based on surface markers (*CD48*, *CD62L*, Supplementary Figure  
372 8B) and performed WGS on each subclone separately, as well as on the whole tumor sample.  
373 In total, we detected 240 non-synonymous SNV located in exonic regions (Supplementary  
374 Table 7), however, only 1 (0.4 %) or 5 (2.1 %) SNV were exclusively detected in the  
375  $CD48^{\text{Low}}CD62L^-$  or the  $CD48^{\text{High}}CD62L^+$  cluster, respectively (Figure 6F). We further  
376 compared CNV profiles of the two subclones and detected a number of differences: the  
377  $CD48^{\text{High}}CD62L^+$  cluster carried an additional copy of *MYC* (8q24, Figure 6G), which was  
378 reflected by increased *MYC* expression levels (Supplementary Figure 9C). The q arm of  
379 chromosome 14 harbored two copy number gains and one copy number loss in the  
380  $CD48^{\text{High}}CD62L^+$  cluster (Figure 6G). Moreover, chromosome X exhibited a copy number gain  
381 of the p arm in the  $CD48^{\text{High}}CD62L^+$  cluster, and a copy number loss of the q arm in the  
382  $CD48^{\text{Low}}CD62L^-$  cluster (Figure 6G).

383 Since pathologic activation of *MYC* renders cells sensitive to BET inhibitors [48, 49], we  
384 performed intracellular flow cytometry-based staining of *MYC* at baseline and after 24 hours  
385 incubation with and without the two BET inhibitors, I-BET-762 or OTX015. We confirmed the  
386 increased *MYC* expression level of the  $CD48^{\text{High}}CD62L^+$  subclone at baseline (Figure 6H,  
387 Supplementary Figure 9D), and, as expected, found that *MYC* was downregulated upon  
388 incubation with I-BET-762 and OTX015, but not upon incubation with the BTK inhibitor  
389 ibrutinib (Figure 6H, Supplementary Figure 9E-G).

390

391 ***In vivo retreatment confirms ex vivo prediction of subpopulation-specific drug response***

392 To exemplify the translational relevance of subclone-specific drug responses, we performed  
393 WES of DLBCL1 during the second relapse after retreatment with high-dose cytarabine. Based  
394 on ex vivo drug perturbation we had predicted that the CD48<sup>High</sup>CD62L<sup>+</sup> but not the  
395 CD48<sup>Low</sup>CD62L<sup>-</sup> subpopulation would respond to cytarabine (Figure 6D). We compared  
396 several synonymous SNV exclusive to the CD48<sup>High</sup>CD62L<sup>+</sup> subpopulation before retreatment  
397 and during second relapse, and observed that the cytarabine-sensitive subpopulation was  
398 successfully eradicated (Figure 6I). Due to the lack of sufficiently exclusive SNV in the  
399 resistant subclone, we took advantage of the loss of heterozygosity (LOH) on chromosome Xq  
400 (Figure 6G) to determine the aberrant fraction of cells harboring a loss of Xq before and after  
401 retreatment. We found that the fraction of chemotherapy-resistant cells, harboring the loss of  
402 Xq, increased from 72 % to 93 % (see methods section for details).

403 In summary, we dissected the intratumor heterogeneity of the DLBCL1 sample on the  
404 transcriptional, genomic, and drug response level. This clinically relevant example highlights  
405 the huge translational relevance of tumor subpopulations and their specific drug response  
406 profile for personalized cancer treatment.

407 **Discussion**

408 Intratumor heterogeneity poses a significant challenge for the clinical management of cancer  
409 patients. Advances of single cell technologies facilitated the profiling of intratumor  
410 heterogeneity at an unprecedented resolution [50]. Most of these studies comprehensively  
411 describe intratumor heterogeneity on the transcriptional level, but do not explore its functional  
412 consequences such as response or resistance to drugs. In this study, we address this limitation  
413 and identify transcriptionally distinct malignant subclones in B-NHL lymph node biopsies. We  
414 study differential drug response patterns of these subclones and genetic events which likely  
415 drove these differences.

416 Our analysis revealed the coexistence of up to four transcriptionally distinct subpopulations of  
417 malignant cells within individual B-NHL lymph node samples. This result recapitulates similar  
418 observations in follicular lymphoma [45], multiple myeloma [51] and other cancer entities [8,  
419 12, 52]. We and others attributed this heterogeneity to differentially enriched gene sets, which  
420 indicate, for instance, activity of *MYC*, proliferation, or germinal center experience. However,  
421 we went further and established a straightforward strategy to prove the coexistence of up to four  
422 different tumor subpopulations at the cellular level. We subsequently performed perturbation  
423 assays with a comprehensive panel of clinically relevant drugs and observed that tumor  
424 subclones within the same lymph node responded strikingly different both to targeted  
425 compounds, such as ibrutinib, but also chemotherapeutics. The study by de Boer and colleagues  
426 supports our observation by demonstrating that acute myeloid leukemia subclones, which were  
427 identified on the basis of 50 leukemia-enriched plasma membrane proteins, had distinct  
428 functional properties including a differential sensitivity to FLT3-inhibition driven by a  
429 subclonal FLT3-ITD mutation [53]. Most preclinical in vitro and in vivo drug screens do not  
430 address such clonal heterogeneity, which may explain the failure of numerous drug candidates  
431 in the clinic [47]. For a single patient, we even demonstrated that the ex vivo drug response  
432 profiling correctly predicted the treatment sensitivity of tumor subclones in vivo. The  
433 prospective identification of rational combinations of cancer drugs that effectively target co-  
434 existing tumor subclones separately could avoid the outgrowth of resistant tumor clones under  
435 therapeutic pressure of a single drug, and would thereby improve efficacy of cancer treatments.  
436 Our study addresses this limitation of many ex vivo drug perturbation studies, and, due to its  
437 unbiased approach to prospectively dissect the malignant substructure, it is also generalizable  
438 to other cancer entities. However, due to the limitation of lymph node derived primary cells,  
439 we have to acknowledge that we could not apply our approach to all samples. Further studies

440 are necessary to expand this approach and to address also spatial heterogeneity of malignant  
441 tumors.

442 Our approach enabled us to directly identify genetic factors that underlie the transcriptional and  
443 drug response differences between subclones. This distinguishes our work from a previous  
444 scRNA-seq study in FL, which indirectly compared allele frequencies of bulk WES with the  
445 size of transcriptionally distinct subclones [45]. The authors found a correlation between  
446 genomic alterations and subclonal fractions and concluded that somatic mutations are  
447 associated with transcriptional differences. These findings are in contrast to another study,  
448 which correlated subclusters derived from targeted single cell expression profiling of 91 genes  
449 with subclusters derived from single cell immunoglobulin heavy-chain (IGH) sequencing. In  
450 this study, the authors concluded that distinct gene expression clusters were not associated with  
451 subclones derived from IGH hypermutations [54]. While these studies provide only indirect  
452 evidence, we physically sorted tumor subclones and normal B cells, and performed WGS or  
453 WES separately for transcriptionally distinct lymphoma subpopulations. With regard to somatic  
454 mutations, we observed two different scenarios: in the DLBCL1 sample we identified almost  
455 no somatic SNVs to be exclusive for one or the other subclone, whereas in the tFL1 sample we  
456 found up to 15% exclusive somatic SNVs in each subclone. However, both examples represent  
457 scenarios where subclone specific drug profiles could not have been predicted by means of gene  
458 mutation sequencing. We further compared CNV profiles of the same tumor subclones, and  
459 found that all subclones harbored significantly different CNV profiles, suggesting that copy  
460 number alterations represent an important layer of genetic events which can drive differential  
461 gene expression programs and drug response profiles. Although our results support the general  
462 notion that genetic events drive subclone specific differences in drug response, they also  
463 highlight the difficulty to predict drug responses based on only genome sequencing in clinical  
464 practice. It might therefore be beneficial to obtain both genetic- and drug response profiles for  
465 personalized treatment decisions.

466 Exploring the heterogeneity of the immune microenvironment in B-NHL has the potential to  
467 better reveal how lymphomas shape their microenvironment and how lymphoma patients could  
468 be better stratified for the treatment with immunotherapies. T cells represented the largest non-  
469 malignant population in B NHL lymph node biopsies. We identified four major,  
470 transcriptionally distinct T cell subpopulations, which were annotated as cytotoxic T cells,  
471 regulatory T helper cells, conventional T helper cells and T follicular helper cells [29-33]. We  
472 measured the frequency of these T cell subsets in an extended cohort of malignant lymph node  
473 biopsies and found T follicular helper cells to be enriched in FL, which is in line with previous

474 flow cytometry-based studies [33, 55]. These T cell subsets displayed only limited  
475 transcriptional heterogeneity with less variability between lymph nodes compared to malignant  
476 cells. However, the frequencies of these T cell subsets varied significantly across donors, which  
477 suggests that B-NHL shape their microenvironment by regulating the recruitment of different  
478 T cell subsets. This observation might be of clinical relevance, because cold tumors with very  
479 few infiltrating T cells have been reported to respond less well to immunotherapies [56].  
480 Despite the rather small number of analyzed B-NHL patients, our study is of high clinical  
481 relevance. We demonstrated that the prospective identification of pre-existing transcriptionally  
482 distinct malignant subclones might be of diagnostic value to detect difficult to treat tumor  
483 subclones. In addition, our research establishes scRNA-seq as a new key technology for precise  
484 molecular profiling of relapsed and refractory nodal B cell lymphomas, and facilitates the  
485 design of new and molecularly-informed diagnosis and treatment strategies.

486 **Online Methods**

487 *Patients samples and lymph node procession*

488 Our study was approved by the Ethics Committee of the University of Heidelberg. Informed  
489 consent was obtained in advance. Immediately after the excision, the lymph node was cut in  
490 small pieces and put into Roswell Park Memorial Institute (Gibco) medium supplemented with  
491 10 % fetal bovine serum (FBS, Gibco), penicillin and streptomycin (Gibco) at a final  
492 concentration of 100 U/ml and 100 µg/ml and L-Glutamine (Gibco) at a final concentration of  
493 2 mM. After filtering by a 40 µm strainer, cells were washed once with phosphate-buffered  
494 saline (PBS, Gibco and put into RPMI medium (Gibco) medium supplemented with 20 % FBS  
495 (Gibco) and 10 % dimethyl sulfoxide (DMSO, Serva), and then cryopreserved in liquid nitrogen  
496 until further analysis. an

497

498 *Quantification of immunohistochemical staining*

499 Formalin fixed lymph node tissue were processed through the hospital's routine  
500 immunohistochemistry pipeline and thereby stained for CD3, PAX5 and Ki67 (all Ventana).  
501 After completion of diagnostics, the corresponding slides were scanned for a subset of patients  
502 ( $n = 7$ ). To quantify the frequencies of B and T cells, the open source software QuPath (v0.1.2)  
503 was used for PAX5 or CD3 stained slides according to the recommended workflow [57]. After  
504 detection of about 100.000 cells per slide, the measurements were exported and further analyzed  
505 using R. We visualized the intracellular signal of diaminobenzidine staining of all detected  
506 events in a histogram. For the staining of PAX5 and CD3 we observed two clear peaks for all  
507 samples and set a threshold in between. Cells with an intracellular signal of CD3 or PAX5  
508 greater than this threshold were regarded as T cells or B cells, respectively. The proportion of  
509 Ki67<sup>+</sup> cells was obtained from routine pathology reports.

510

511 *Surface and intracellular staining by flow cytometry*

512 As described above, lymph node derived cells were thawed and stained for viability using a  
513 fixable viability dye e506 (Thermo Fisher Scientific) and for different surface markers  
514 depending on the experimental setup. The following surface antibodies were used: anti-CD3-  
515 PerCP/Cy5.5, anti-CD3-APC, anti-CD19-BV421, anti-kappa-PE, anti-kappa-FITC, anti-  
516 lambda-PE/Dazzle, anti-CD22-APC, anti-CD24-BV785, anti-CD27-PE-Cy7, anti-CD32-PE,  
517 anti-CD44-PE, anti-CD48-PE, anti-CD62L-PE/Cy7, anti-CD10-APC-Cy7, anti-CD4-AF700,  
518 anti-CD8-FITC, anti-PD1-BV421 and anti-ICOS-PE/Dazzle (all Biolegend). In case of

519 subsequent intracellular staining, cells were fixed and permeabilized with the intracellular  
520 fixation/permeabilization buffer set (Thermo Fisher Scientific) and stained with anti-MYC-  
521 AF647 (Thermo Fisher Scientific), anti-FoxP3-AF647 (BD Biosciences), or adequate isotype  
522 controls (Thermo Fisher Scientific, BD Biosciences). Cells were then analyzed with an LSR  
523 Fortessa (BD Biosciences) and FACSDiva (BD Biosciences, Version 8)

524

525 *Estimating the proportion of malignant and non-malignant B cells by flow cytometry*

526 Staining for expression of the light chains (kappa, lambda) is a well-established tool to identify  
527 the accumulation of light chain restricted, malignant B cells [58]. Lymph node derived cells  
528 were stained as described above. In case of a kappa<sup>+</sup> or lambda<sup>+</sup> B cell population greater than  
529 80 %, we regarded this population as light chain restricted and therefore as malignant. We  
530 further assumed that the ratio of kappa<sup>+</sup> versus lambda<sup>+</sup> B cells among the potentially remaining  
531 non-malignant B cells is still balanced. Therefore, there must be roughly the same proportion  
532 of non-malignant B cells among those carrying the restricted type of light chain. This ends up  
533 in the following formula to estimate the proportion of malignant cells:

534 
$$\text{Proportion}_{\text{Bcells\_malignant}} \approx \text{Proportion}_{\text{Bcells\_restricted}} - \text{Proportion}_{\text{Bcells\_not.restricted}}$$

535 In addition, cells without detectable expression of kappa or lambda light chain on protein level  
536 were regarded as malignant cells because a loss of light chain expression is not observed in  
537 non-malignant lymph nodes [59].

538

539 *Single cell sample preparation and RNA sequencing*

540 After thawing, cells were washed to remove DMSO as quickly as possible. We used the dead  
541 cell removal kit (Miltenyi Biotec) for all samples to achieve a viability of at least 90%. The  
542 preparation of the single cell suspensions, synthesis of cDNA and single cell libraries were  
543 performed using the Chromium single cell v2 3' kit (10x Genomics) according to the  
544 manufacturer's instructions. Each was sequenced on one NextSeq 550 lane (Illumina).

545

546 *Subclone specific drug screening*

547 58 different drugs at 5 different concentrations (Supplementary Table 5) and a suitable number  
548 of DMSO controls were prepared in 384 well plates. DMSO concentration was kept equally at  
549 0.2 % in all wells. Lymph node cells were thawed in a 37°C water bath and DMSO containing  
550 freezing medium was removed as quickly as possible to reduce cytotoxic effects. Afterwards,  
551 lymph node cells were rolled for 3 hours in RPMI medium supplemented with penicillin and

streptomycin (Gibco) at a final concentration of 100 U/ml and 100 µg/ml, L-glutamine (Gibco) at a final concentration of 2 mM and with 10 % human AB male serum (Sigma). Cells were seeded at a cell count of 50,000 in 50 µl per well. After 48 hours, cells were washed once with staining buffer [PBS (Gibco) supplemented with 1% FBS and 0.5 % ethylenediaminetetraacetic acid (EDTA, Sigma Aldrich)]. Cells were subsequently stained with fixable viability dye e506 (Thermo Fisher Scientific), anti-CD3-APC, anti-CD19-BV421 and anti-CD48-PE, anti-CD62L-PE/Cy7 or anti-kappa-FITC, anti-lambda-PE/Dazzle, anti-CD10-APC/Cy7, anti-CD27-PE/Cy7, anti-CD32-PE (all Biolegend). After staining the microtiter plate was washed twice with staining buffer. Then, cells were fixed using paraformaldehyde at a final concentration of 2 % for 15 min at room temperature and washed with staining buffer. Fixed cells were analyzed with an LSR II and FACSDiva (BD Biosciences, Version 8) equipped with a high throughput sampler (HTS) system (BD Biosciences). Approximately 5,000 to 10,000 events were recorded per well. Flow cytometry data was analyzed using FlowJo software (Tree Star). The gating strategy is illustrated in Supplementary Figure 8. We ruled out that significant up- or downregulation of subclone-discriminating surface antigens confound subclone-specific drug response assessment by evaluating the fluorescence intensity of corresponding markers before and after drug treatment (Supplementary Figure 10).

569

#### 570 *Fluorescence-activated cell sorting of B cell subclones*

571 Lymph node cells were stained as described above. Sorting was performed at a FACS Aria  
572 Fusion (BD Biosciences). We sorted either for e506<sup>-</sup> CD3<sup>-</sup> CD19<sup>+</sup> CD48<sup>-</sup> CD62L<sup>-</sup> and e506<sup>-</sup>  
573 CD3<sup>-</sup> CD19<sup>+</sup> CD48<sup>-</sup> CD62L<sup>-</sup> (DLBCL1) or for e506<sup>-</sup> CD3<sup>-</sup> CD19<sup>+</sup> CD10<sup>-</sup>, e506<sup>-</sup> CD3<sup>-</sup> CD19<sup>+</sup>  
574 CD10<sup>+</sup> kappa<sup>+</sup>, CD32<sup>low</sup> and e506<sup>-</sup> CD3<sup>-</sup> CD19<sup>+</sup> CD10<sup>+</sup> kappa<sup>+</sup>, CD32<sup>high</sup> cells (tFL). The gating  
575 strategy is illustrated in Supplementary Figure 8. All relevant fractions were analyzed post-  
576 sorting to confirm a purity of at least 95 %.

577

#### 578 *Whole genome and whole exome sequencing*

579 DNA was extracted using the DNeasy mini kit (Qiagen) according to the manufacturers  
580 protocol, followed by quality control using gel electrophoresis and a TapeStation 2200 system  
581 (Agilent). Samples were prepared either for WGS or WES, as previously described [60]. Exome  
582 capturing was performed using SureSelect Human All Exon V5 in-solution capture reagents  
583 (Agilent). If samples were destined for WES on an Illumina HiSeq 2500 instrument, then 1.5  
584 µg genomic DNA were fragmented to 150 to 200 bp insert size with a Covaris S2 device, and  
585 250 ng of Illumina adapter-containing libraries were hybridized with exome baits at 65°C for

586 16 hours. If samples were destined for WES on an Illumina HiSeq 4000 instrument, then 200  
587 ng genomic DNA were fragmented to 300 bp insert size with a Covaris LE220 or E220 device,  
588 and 750 ng of adapter-containing libraries were hybridized with exome baits at 65°C for 16  
589 hours. If samples were destined for WGS on an Illumina HiSeq X instrument, then 100 ng of  
590 genomic DNA were fragmented to 450 bp insert size with a Covaris LE220 or E220 device,  
591 and libraries were prepared using the TruSeq Nano Kit (Illumina). On all platforms paired-end  
592 sequencing was carried out according to the manufacturer's recommendations, yielding read  
593 lengths of 101 bp (4000) or 151 bp (HiSeq X).

594

595 *Single cell RNA sequencing data processing*

596 The Cell Ranger analysis pipeline (v2.1, 10x Genomics) was used to demultiplex the raw base  
597 call files and to convert them into FASTQ files. FASTQ files were aligned to the reference  
598 genome (hg38) and filtered. Final numbers of cell barcodes, unique molecular identifiers (UMI)  
599 per cell, median genes and sequencing saturation are summarized in Supplementary Table 8.

600

601 *Filtering and normalizing single cell RNA sequencing data*

602 The R package Seurat [61] (v2.3.3) was used to perform quality control and normalization.  
603 Gene count per cell, UMI count per cell and the percentage of mitochondrial and ribosomal  
604 transcripts were computed using the functions of the Seurat package. Genes expressed in three  
605 cells or fewer were excluded from downstream analysis. Libraries with a percentage of  
606 mitochondrial transcripts greater than 5%, along with those with less than 200 genes were  
607 filtered out prior to further analysis. Since aggressive lymphomas displayed higher gene and  
608 UMI count, the upper limit was set with regard to each sample. Counts were adjusted for cell-  
609 specific sampling (“normalized”) using the LogNormalize function with the default scale factor  
610 of 10,000.

611

612 *Assessing the cell cycle state using scRNA-seq data*

613 The cell cycle state was assessed using the gene set and scoring system, described by Tirosh  
614 and colleagues [8]. Briefly, the S-Score and the G<sub>2</sub>M-Score were calculated based on a list of  
615 43 S phase-specific and 54 G<sub>2</sub> or M phase-specific genes. The calculation of the actual scores  
616 was performed using the CellCycleScoring function of the Seurat R package.

617

618 *Analysis of ligand-receptor interactions in scRNA-Seq data*

619 We used the CellPhoneDB database [37] as basis for potential cell-cell interactions, but  
620 expanded the list by important B to T cell interactions (Supplementary Table 4). To assess the  
621 significance of each interaction, we adapted a statistical framework recently described by  
622 Vento-Tormo and colleagues [37] to our purpose. Importantly, we considered only genes which  
623 were expressed in 5 % of at least one cell type.

624 Briefly, we performed pairwise comparisons between the different T and B cell subtypes for  
625 each ligand-receptor pair and sample. For each combination of two different cell types and each  
626 ligand-receptor-pair, we permuted the cluster labels of cells at least 1,000 times and determined  
627 the mean interaction score (mean expression of ligand in cell type A times mean expression of  
628 receptor in cell type B). A p value was determined by calculating the proportion of permuted  
629 interaction scores which were by hazard higher than the actual interaction score. All interactions  
630 were calculated sample-wise. To determine which interactions were most relevant across  
631 different samples, we calculated the mean interaction scores and combined the different p  
632 values using the Fisher's method. Then, p values were corrected using the Benjamini-Hochberg  
633 method. The R code is available on our GitHub repository (see code availability statement  
634 below).

635

636 *Combining data from different samples and batch correction*

637 After identification of the different cell types the data sets were split into non-B cells or B cells  
638 using the SubsetData function. Then the respective subsets were combined using the  
639 MergeSeurat function. Putative batch effects between two runs, were corrected by the mutual  
640 nearest neighbors (MNN) technique [62] which is implemented in the scran Bioconductor  
641 package (v1.10.2).

642

643 *Clustering and dimensionality reduction techniques*

644 SNN (Shared-nearest neighbor)-based clustering, t-SNE and UMAP visualization were  
645 performed using the FindClusters, RunTSNE and RunUMAP functions within the Seurat  
646 package [61]. Each of these were performed on the basis of a principal component analysis  
647 which was performed using the RunPCA function of the Seurat package. The same parameters  
648 were applied to all samples. UMAP was used instead of t-SNE for combined data sets because  
649 it is significantly faster than t-SNE and better preserves aspects of global structure in larger data  
650 sets [28]. Differentially expressed genes between the clusters were identified using the  
651 FindMarkers or FindAllMarkers functions within the Seurat package [61]. Differentially

652 expressed genes between malignant B cell clusters can be browsed interactively using an html  
653 file (see data sharing statement below).

654

655 *Gene set enrichment analysis*

656 Gene set enrichment analysis (GSEA) was performed using the GSEA java desktop application  
657 [63, 64] and the Molecular Signatures Database (MSigDB, v6.2) provided by the Broad Institute  
658 [63, 65]. Differentially expressed genes of two groups were used to determine significantly-  
659 enriched gene sets.

660

661 *WES and WGS data processing*

662 Alignment of sequencing read pairs and variant calling were performed as recently described  
663 [66]. Briefly, reads were mapped to human reference genome (hg19) with bwa-mem (version  
664 0.7.8, minimum base quality threshold set to zero [-T 0], remaining settings left to default) [67].  
665 Subsequently, reads were coordinate-sorted with bamsort (compression option set to fast) and  
666 duplicate read pairs were marked with bammarkduplicates (compression option set to best)  
667 (both part of biobambam package version 0.0.148).

668 SNV and indels in matched tumor normal pairs were identified using the internal DKFZ variant  
669 calling workflows based on samtools/bcftools 0.1.19 with additional custom filters (optimized  
670 for somatic variant calling by deactivating the pval-threshold in bcftools) and Platypus 0.8.1,  
671 respectively, as described previously [66]. Gene annotation of variants was done with  
672 Annovar [68]. The variants were annotated with dbSNP141, 1000 Genomes (phase 1), Gencode  
673 mapability track, UCSC High Seq Depth track, UCSC Simple-Tandem repeats, UCSC Repeat-  
674 Masker, DUKE-Excluded, DAC-Blacklist, UCSC Selfchain. These annotation tracks were used  
675 to determine a confidence score for each variant by a heuristic punishment scheme and only  
676 high confidence variants were kept for further analysis. In addition, variants with strong read  
677 biases according to the strand bias filter were removed.

678 Genomic structural rearrangements (SVs) were identified using the SOPHIA algorithm  
679 (unpublished, source code available at <https://bitbucket.org/utoprak/sophia/>). Briefly,  
680 supplementary alignments as produced by bwa-mem are used as indicators of potential  
681 underlying SVs. Candidates are filtered by comparing them to a background control set of  
682 sequencing data obtained using normal blood samples from a background population database  
683 of 3261 patients from published TCGA and ICGC studies as well as published and unpublished  
684 studies of the German Cancer Research Center (DKFZ).

685 Allele-specific CNV were detected using ACEseq (allele-specific copy number estimation from  
686 WGS) [69] for WGS data and CNVkit for WES data [70]. ACEseq determines absolute allele-  
687 specific copy numbers as well as tumor ploidy and tumor cell content based on coverage ratios  
688 of tumor and control as well as the B-allele frequency (BAF) of heterozygous single-nucleotide  
689 polymorphisms (SNPs). SVs called by SOPHIA were incorporated to improve genome  
690 segmentation.

691

#### 692 *Multi tumor comparison*

693 To compare multi tumor samples of the same donor, every SNV position in each sample was  
694 determined using samtools mpileup 1.6. At each of these SNV positions, the variant allele  
695 fraction was determined by calculating the ratio between the number of variant reads and the  
696 total coverage at that position. To correct the variant allele fraction for actual tumor cell content,  
697 a scaling factor was incorporated, comprising ploidy and total copy number (TCN) estimates  
698 obtained from ACEseq/CNVkit. Specifically, the scaling factor is obtained as the ratio between  
699 purity corrected number of alleles in the tumor (TCN\_tumor purity\_tumor) and purity corrected  
700 total number of alleles in the sample ((TCN\_tumor \* purity\_tumor) + 2 \* (1 - purity\_tumor)).

701

#### 702 *Aberrant cell fraction estimation from LOH*

703 To determine aberrant cell fractions, the minor allele-frequency (MAF, ratio between number  
704 of reads of minor allele and total coverage at given position) of single nucleotide polymorphism  
705 (SNP) was estimated for selected regions harboring a loss of heterozygosity (LOH) or a copy  
706 number neutral LOH (CN-LOH) in the tumor sample. Information on SNP location was  
707 received from matched-control SNV calling. To select heterozygous SNP, only SNP with a  
708 MAF  $\geq 0.3$  in the control were retained. Subsequently, MAF values of the selected SNP were  
709 calculated for the tumor samples. For exome samples, only SNP within the targeted capture  
710 regions were kept. The mean of the respective tumor MAF values was calculated and the  
711 aberrant cell fraction (ACF) was estimated as follows:

$$712 \quad ACF_{LOH} = 1 - 2 \cdot \text{mean}(MAF); \quad ACF_{CN-LOH} = \frac{1 - 2 \cdot \text{mean}(MAF)}{1 - \text{mean}(MAF)}$$

713

#### 714 *Data sharing statement*

715 The single cell expression data of merged B and T cell UMAP plots (Figure 2A/B and  
716 Figure 3A/B) are available for easy-to-use interactive browsing:

717 <https://www.zmbh.uni-heidelberg.de/Anders/scLN-index.html>.

718 The raw single cell count tables can be downloaded here doi.org/10.11588/data/VRJUNV. This  
719 link will be activated upon publication and is accessible without further restriction.  
720 Differentially expressed genes between B cell clusters can be browsed in an interactive html  
721 file (Supplementary File 1).

722

723 *Code availability statement*

724 R codes used for data analysis are available at our GitHub repository without further restriction  
725 ([www.github.com/DietrichLab/scLymphomaExplorer](https://www.github.com/DietrichLab/scLymphomaExplorer)).

726 **References**

- 727
- 728 1. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of  
729 genetic heterogeneity in cancer evolution. *Nature* 2013 09/18/online; **501**: 338.
- 730
- 731 2. Swanton C. Intratumor heterogeneity: evolution through space and time. *Cancer*  
732 *research* 2012; **72**(19): 4875-4882.
- 733
- 734 3. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present,  
735 and the Future. *Cell* 2017 Feb 9; **168**(4): 613-628.
- 736
- 737 4. Kim IS, Zhang XH. One microenvironment does not fit all: heterogeneity beyond cancer  
738 cells. *Cancer Metastasis Rev* 2016 Dec; **35**(4): 601-629.
- 739
- 740 5. Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human  
741 tumours: impact on clinical outcome. *Nat Rev Cancer* 2012 Mar 15; **12**(4): 298-306.
- 742
- 743 6. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, *et al.* Tumour evolution  
744 inferred by single-cell sequencing. *Nature* 2011 Apr 7; **472**(7341): 90-94.
- 745
- 746 7. Wang Y, Navin NE. Advances and applications of single-cell sequencing technologies.  
747 *Mol Cell* 2015 May 21; **58**(4): 598-609.
- 748
- 749 8. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, Treacy D, Trombetta JJ, *et al.*  
750 Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq.  
751 *Science* 2016 Apr 8; **352**(6282): 189-196.
- 752
- 753 9. Gao R, Kim C, Sei E, Foukakis T, Crosetto N, Chan LK, *et al.* Nanogrid single-nucleus  
754 RNA sequencing reveals phenotypic diversity in breast cancer. *Nat Commun* 2017 Aug  
755 9; **8**(1): 228.
- 756
- 757 10. Yuan J, Sims PA. An Automated Microwell Platform for Large-Scale Single Cell RNA-  
758 Seq. *Sci Rep* 2016 Sep 27; **6**: 33883.
- 759
- 760 11. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, *et al.* Quantitative single-  
761 cell RNA-seq with unique molecular identifiers. *Nat Methods* 2014 Feb; **11**(2): 163-  
762 166.
- 763
- 764 12. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, *et al.* Single-  
765 cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*  
766 2014 Jun 20; **344**(6190): 1396-1401.
- 767
- 768 13. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, *et al.* Full-length mRNA-  
769 Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat*  
770 *Biotechnol* 2012 Aug; **30**(8): 777-782.
- 771
- 772 14. Powell AA, Talasaz AH, Zhang H, Coram MA, Reddy A, Deng G, *et al.* Single cell  
773 profiling of circulating tumor cells: transcriptional heterogeneity and diversity from  
774 breast cancer cell lines. *PLoS One* 2012; **7**(5): e33788.
- 775

- 776 15. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, *et al.* Chemoresistance  
777 Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell*  
778 2018 May 3; **173**(4): 879-893 e813.
- 779
- 780 16. Teras LR, DeSantis CE, Cerhan JR, Morton LM, Jemal A, Flowers CR. 2016 US  
781 lymphoid malignancy statistics by World Health Organization subtypes. *CA Cancer J  
782 Clin* 2016 Sep 12.
- 783
- 784 17. Wagner-Johnston ND, Link BK, Byrtok M, Dawson KL, Hainsworth J, Flowers CR, *et  
785 al.* Outcomes of transformed follicular lymphoma in the modern era: a report from the  
786 National LymphoCare Study (NLCS). *Blood* 2015 Aug 13; **126**(7): 851-857.
- 787
- 788 18. Crump M, Neelapu SS, Farooq U, Van Den Neste E, Kuruvilla J, Westin J, *et al.*  
789 Outcomes in refractory diffuse large B-cell lymphoma: results from the international  
790 SCHOLAR-1 study. *Blood* 2017 Oct 19; **130**(16): 1800-1808.
- 791
- 792 19. Philip T, Armitage JO, Spitzer G, Chauvin F, Jagannath S, Cahn JY, *et al.* High-dose  
793 therapy and autologous bone marrow transplantation after failure of conventional  
794 chemotherapy in adults with intermediate-grade or high-grade non-Hodgkin's  
795 lymphoma. *N Engl J Med* 1987 Jun 11; **316**(24): 1493-1498.
- 796
- 797 20. Bartlett NL, Costello BA, LaPlant BR, Ansell SM, Kuruvilla JG, Reeder CB, *et al.*  
798 Single-agent ibrutinib in relapsed or refractory follicular lymphoma: a phase 2  
799 consortium trial. *Blood* 2018 Jan 11; **131**(2): 182-190.
- 800
- 801 21. Winter AM, Landsburg DJ, Mato AR, Isaac K, Hernandez-Ilizaliturri FJ, Reddy N, *et  
802 al.* A multi-institutional outcomes analysis of patients with relapsed or refractory  
803 DLBCL treated with ibrutinib. *Blood* 2017 Aug 9.
- 804
- 805 22. Horna P, Olteanu H, Kroft SH, Harrington AM. Flow cytometric analysis of surface  
806 light chain expression patterns in B-cell lymphomas using monoclonal and polyclonal  
807 antibodies. *Am J Clin Pathol* 2011 Dec; **136**(6): 954-959.
- 808
- 809 23. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles  
810 in cancer suppression and promotion. *Science* 2011 Mar 25; **331**(6024): 1565-1570.
- 811
- 812 24. Challa-Malladi M, Lieu YK, Califano O, Holmes AB, Bhagat G, Murty VV, *et al.*  
813 Combined genetic inactivation of beta2-Microglobulin and CD58 reveals frequent  
814 escape from immune recognition in diffuse large B cell lymphoma. *Cancer Cell* 2011  
815 Dec 13; **20**(6): 728-740.
- 816
- 817 25. Xu-Monette ZY, Zhou J, Young KH. PD-1 expression and clinical PD-1 blockade in B-  
818 cell lymphomas. *Blood* 2018 Jan 4; **131**(1): 68-83.
- 819
- 820 26. Sun LL, Ellerman D, Mathieu M, Hristopoulos M, Chen X, Li Y, *et al.* Anti-CD20/CD3  
821 T cell-dependent bispecific antibody for the treatment of B cell malignancies. *Sci Transl  
822 Med* 2015 May 13; **7**(287): 287ra270.
- 823
- 824 27. Os A, Burgler S, Ribes AP, Funderud A, Wang D, Thompson KM, *et al.* Chronic  
825 lymphocytic leukemia cells are activated and proliferate in response to specific T helper  
826 cells. *Cell Rep* 2013 Aug 15; **4**(3): 566-577.

- 827  
828 28. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, *et al.* Dimensionality  
829 reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 2018 Dec 3.  
830  
831 29. Schaeerli P, Willimann K, Lang AB, Lipp M, Loetscher P, Moser B. CXC chemokine  
832 receptor 5 expression defines follicular homing T cells with B cell helper function. *J  
833 Exp Med* 2000 Dec 4; **192**(11): 1553-1562.  
834  
835 30. Breitfeld D, Ohl L, Kremmer E, Ellwart J, Sallusto F, Lipp M, *et al.* Follicular B helper  
836 T cells express CXC chemokine receptor 5, localize to B cell follicles, and support  
837 immunoglobulin production. *J Exp Med* 2000 Dec 4; **192**(11): 1545-1552.  
838  
839 31. Dorfman DM, Shahsafaei A. CD200 (OX-2 membrane glycoprotein) is expressed by  
840 follicular T helper cells and in angioimmunoblastic T-cell lymphoma. *Am J Surg Pathol*  
841 2011 Jan; **35**(1): 76-83.  
842  
843 32. Weber JP, Fuhrmann F, Feist RK, Lahmann A, Al Baz MS, Gentz LJ, *et al.* ICOS  
844 maintains the T follicular helper cell phenotype by down-regulating Kruppel-like factor  
845 2. *J Exp Med* 2015 Feb 9; **212**(2): 217-233.  
846  
847 33. Yang ZZ, Grote DM, Ziesmer SC, Xiu B, Novak AJ, Ansell SM. PD-1 expression  
848 defines two distinct T-cell sub-populations in follicular lymphoma that differentially  
849 impact patient survival. *Blood Cancer J* 2015 Feb 20; **5**: e281.  
850  
851 34. Yamazaki T, Nagumo H, Hayashi T, Sugane K, Agematsu K. CD72-mediated  
852 suppression of human naive B cell differentiation by down-regulating X-box binding  
853 protein 1. *Eur J Immunol* 2005 Aug; **35**(8): 2325-2334.  
854  
855 35. Klein U, Rajewsky K, Kuppers R. Human immunoglobulin (Ig)M+IgD+ peripheral  
856 blood B cells expressing the CD27 cell surface antigen carry somatically mutated  
857 variable region genes: CD27 as a general marker for somatically mutated (memory) B  
858 cells. *J Exp Med* 1998 Nov 2; **188**(9): 1679-1689.  
859  
860 36. Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Delabie J, Ott G, *et al.*  
861 Confirmation of the molecular classification of diffuse large B-cell lymphoma by  
862 immunohistochemistry using a tissue microarray. *Blood* 2004 Jan 1; **103**(1): 275-282.  
863  
864 37. Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, *et  
865 al.* Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature*  
866 2018 Nov; **563**(7731): 347-353.  
867  
868 38. Shirota H, Klinman DM, Ito SE, Ito H, Kubo M, Ishioka C. IL4 from T Follicular Helper  
869 Cells Downregulates Antitumor Immunity. *Cancer Immunol Res* 2017 Jan; **5**(1): 61-71.  
870  
871 39. Aguilar-Hernandez MM, Blunt MD, Dobson R, Yeomans A, Thirdborough S, Larrayoz  
872 M, *et al.* IL-4 enhances expression and function of surface IgM in CLL cells. *Blood*  
873 2016 Jun 16; **127**(24): 3015-3025.  
874  
875 40. Peter-Martin B, Giles H, Kolb C, Rabe S, Roider T, Knoll M, *et al.* Systematic  
876 Investigation of Microenvironmental Drug Resistance Mechanisms in Chronic  
877 Lymphocytic Leukemia. *ASH Abstract Book* 2019.

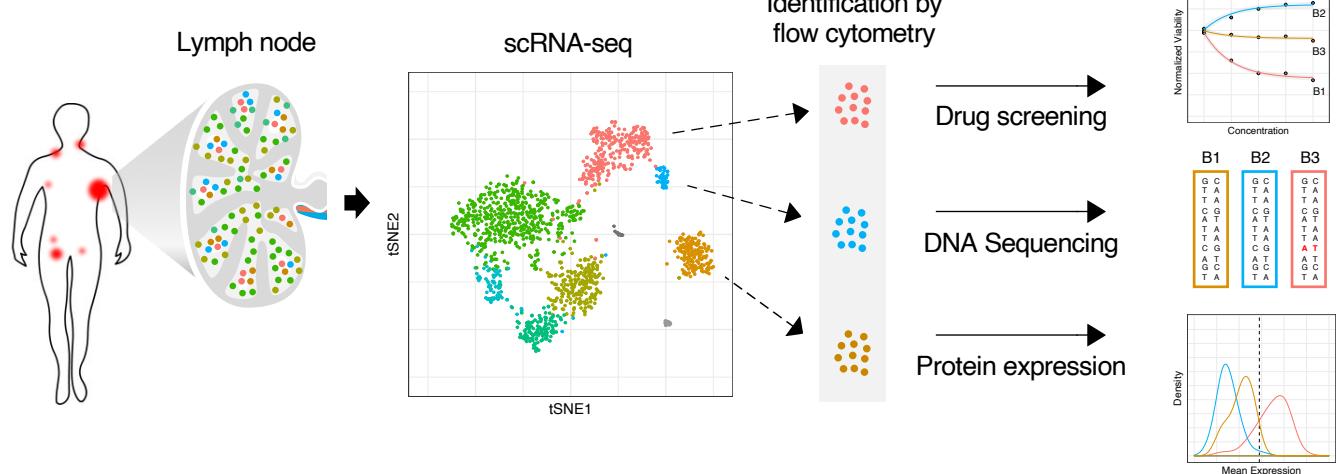
- 878  
879 41. Spolski R, Leonard WJ. IL-21 and T follicular helper cells. *Int Immunol* 2010 Jan;  
880 22(1): 7-12.  
881  
882 42. Gu-Trantien C, Migliori E, Buisseret L, de Wind A, Brohee S, Garaud S, *et al.* CXCL13-  
883 producing TFH cells link immune suppression and adaptive memory in human breast  
884 cancer. *JCI Insight* 2017 Jun 2; 2(11).  
885  
886 43. DiToro D, Winstead CJ, Pham D, Witte S, Andargachew R, Singer JR, *et al.* Differential  
887 IL-2 expression defines developmental fates of follicular versus nonfollicular helper T  
888 cells. *Science* 2018 Sep 14; 361(6407).  
889  
890 44. Chapuy B, Stewart C, Dunford AJ, Kim J, Kamburov A, Redd RA, *et al.* Molecular  
891 subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic  
892 mechanisms and outcomes. *Nat Med* 2018 May; 24(5): 679-690.  
893  
894 45. Andor N, Simonds EF, Czerwinski DK, Chen J, Grimes SM, Wood-Bouwens C, *et al.*  
895 Single-cell RNA-Seq of follicular lymphoma reveals malignant B-cell types and  
896 coexpression of T-cell immune checkpoints. *Blood* 2019 Mar 7; 133(10): 1119-1129.  
897  
898 46. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The  
899 Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015  
900 Dec 23; 1(6): 417-425.  
901  
902 47. Dietrich S, Oles M, Lu J, Sellner L, Anders S, Velten B, *et al.* Drug-perturbation-based  
903 stratification of blood cancer. *J Clin Invest* 2018 Jan 2; 128(1): 427-445.  
904  
905 48. Delmore JE, Issa GC, Lemieux ME, Rahl PB, Shi J, Jacobs HM, *et al.* BET  
906 bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* 2011 Sep 16;  
907 146(6): 904-917.  
908  
909 49. Chapuy B, McKeown MR, Lin CY, Monti S, Roemer MG, Qi J, *et al.* Discovery and  
910 characterization of super-enhancer-associated dependencies in diffuse large B cell  
911 lymphoma. *Cancer Cell* 2013 Dec 9; 24(6): 777-790.  
912  
913 50. Navin NE. The first five years of single-cell cancer genomics and beyond. *Genome Res*  
914 2015; 25(10): 1499-1507.  
915  
916 51. Ledergor G, Weiner A, Zada M, Wang S-Y, Cohen YC, Gatt ME, *et al.* Single cell  
917 dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma.  
918 *Nature Medicine* 2018 2018/12/01; 24(12): 1867-1876.  
919  
920 52. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, *et al.* Single-Cell  
921 Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and  
922 Neck Cancer. *Cell* 2017 Dec 14; 171(7): 1611-1624 e1624.  
923  
924 53. de Boer B, Prick J, Pruis MG, Keane P, Imperato MR, Jaques J, *et al.* Prospective  
925 Isolation and Characterization of Genetically and Functionally Distinct AML  
926 Subclones. *Cancer Cell* 2018 Oct 8; 34(4): 674-689 e678.  
927

- 928 54. Milpied P, Cervera-Marzal I, Mollichella M-L, Tesson B, Brisou G, Traverse-Glehen  
929 A, *et al.* Human germinal center transcriptional programs are de-synchronized in B cell  
930 lymphoma. *Nature Immunology* 2018 2018/09/01; **19**(9): 1013-1024.  
931
- 932 55. Byford ET, Carr M, Ladikou E, Ahearne MJ, Wagner SD. Circulating Tfh1 (cTfh1) cell  
933 numbers and PD1 expression are elevated in low-grade B-cell non-Hodgkin's  
934 lymphoma and cTfh gene expression is perturbed in marginal zone lymphoma. *PLoS  
935 One* 2018; **13**(1): e0190468.  
936
- 937 56. Li J, Byrne KT, Yan F, Yamazoe T, Chen Z, Baslan T, *et al.* Tumor Cell-Intrinsic  
938 Factors Underlie Heterogeneity of Immune Cell Infiltration and Response to  
939 Immunotherapy. *Immunity* 2018 Jul 17; **49**(1): 178-193 e177.  
940
- 941 57. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, *et  
942 al.* QuPath: Open source software for digital pathology image analysis. *Scientific  
943 Reports* 2017 2017/12/04; **7**(1): 16878.  
944
- 945 58. Ratech H, Litwin S. Surface immunoglobulin light chain restriction in B-cell non-  
946 Hodgkin's malignant lymphomas. *Am J Clin Pathol* 1989 May; **91**(5): 583-586.  
947
- 948 59. Kaleem Z, Zehnbauer BA, White G, Zutter MM. Lack of expression of surface  
949 immunoglobulin light chains in B-cell non-Hodgkin lymphomas. *Am J Clin Pathol* 2000  
950 Mar; **113**(3): 399-405.  
951
- 952 60. Heining C, Horak P, Uhrig S, Codo PL, Klink B, Hutter B, *et al.* NRG1 Fusions in  
953 KRAS Wild-Type Pancreatic Cancer. *Cancer Discov* 2018 Sep; **8**(9): 1087-1095.  
954
- 955 61. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell  
956 transcriptomic data across different conditions, technologies, and species. *Nat  
957 Biotechnol* 2018 Jun; **36**(5): 411-420.  
958
- 959 62. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-  
960 sequencing data are corrected by matching mutual nearest neighbors. *Nature  
961 Biotechnology* 2018 04/02/online; **36**: 421.  
962
- 963 63. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al.*  
964 Gene set enrichment analysis: a knowledge-based approach for interpreting genome-  
965 wide expression profiles. *Proc Natl Acad Sci U S A* 2005 Oct 25; **102**(43): 15545-15550.  
966
- 967 64. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, *et al.* PGC-  
968 1alpha-responsive genes involved in oxidative phosphorylation are coordinately  
969 downregulated in human diabetes. *Nat Genet* 2003 Jul; **34**(3): 267-273.  
970
- 971 65. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP.  
972 Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011 Jun 15; **27**(12):  
973 1739-1740.  
974
- 975 66. López C, Kleinheinz K, Aukema SM, Rohde M, Bernhart SH, Hübschmann D, *et al.*  
976 Genomic and transcriptomic changes complement each other in the pathogenesis of  
977 sporadic Burkitt lymphoma. *Nature Communications* 2019 2019/03/29; **10**(1): 1459.  
978

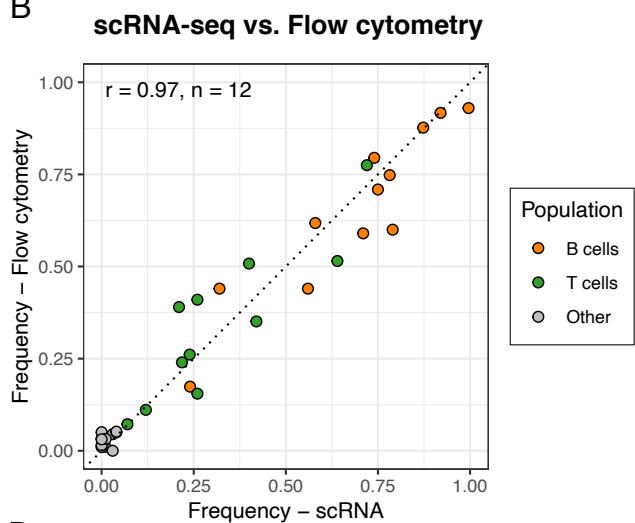
- 979 67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* The Sequence  
980 Alignment/Map format and SAMtools. *Bioinformatics* 2009 Aug 15; **25**(16): 2078-  
981 2079.
- 982
- 983 68. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants  
984 from high-throughput sequencing data. *Nucleic Acids Res* 2010 Sep; **38**(16): e164.
- 985
- 986 69. Kleinheinz K, Bludau I, Hübschmann D, Heinold M, Kensche P, Gu Z, *et al.* ACEseq  
987 – allele specific copy number estimation from whole genome sequencing. *bioRxiv* 2017:  
988 210807.
- 989
- 990 70. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number  
991 Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* 2016  
992 Apr; **12**(4): e1004873.

# Figure 1

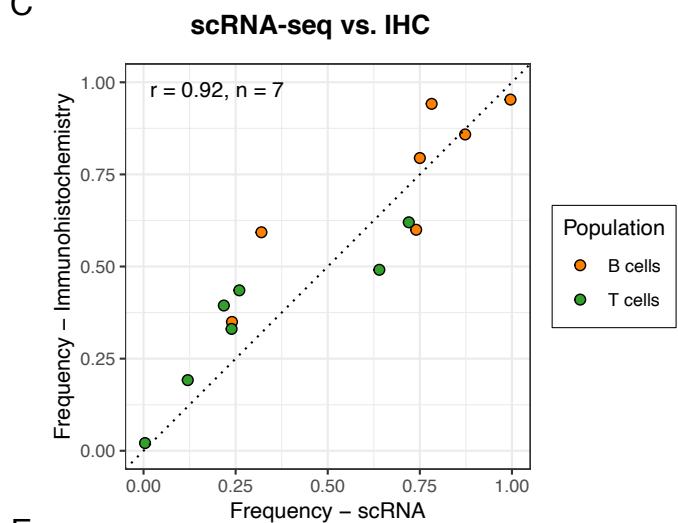
A



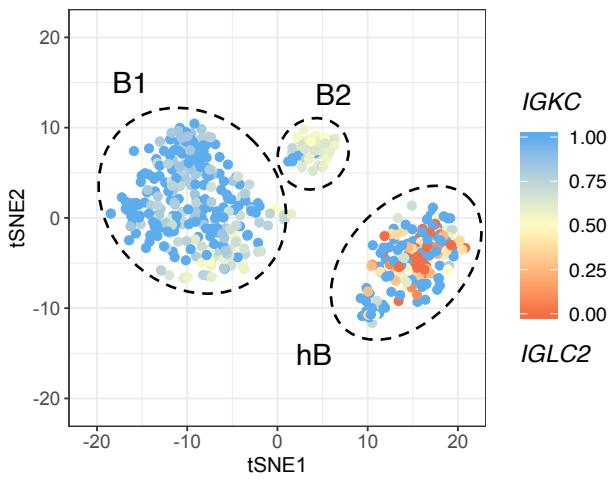
B



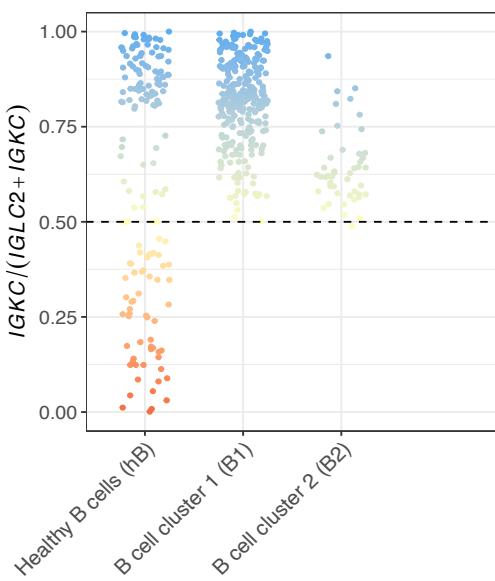
C



D



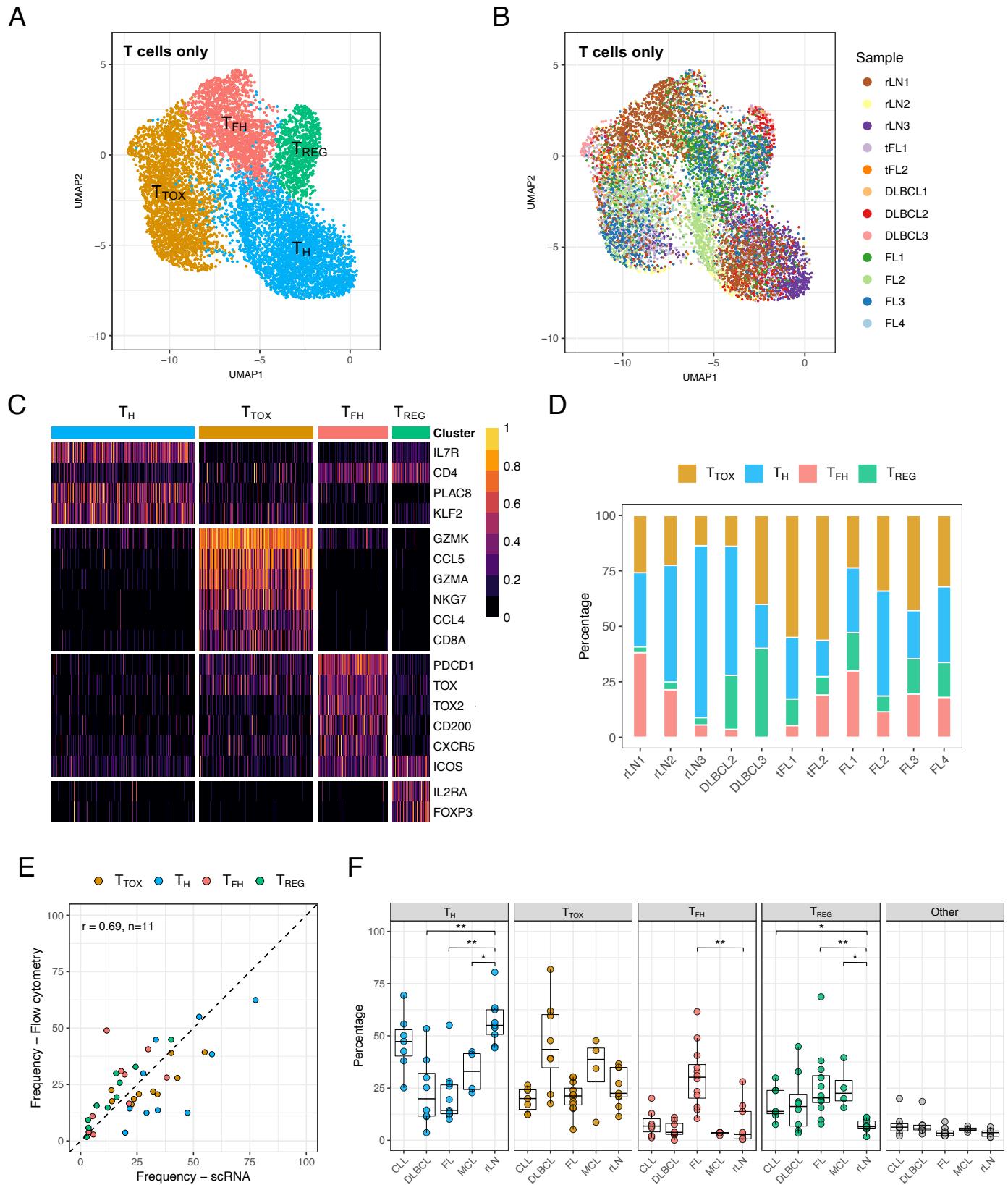
E



**Figure 1. Identification of cell types using scRNA-seq.**

A) Schematic overview of the study design. B and C) Lymph node-derived B and T cells were quantified by scRNA-seq, flow cytometry and immunohistochemistry (IHC) of paraffin tissue sections (see Supplementary Figure 1 for details). The frequencies of B and T cells were correlated for B) scRNA-seq and flow cytometry or C) scRNA-seq and IHC. Pearson's correlation coefficients ( $r$ ) and the number of samples included ( $n$ ) are given in the left top corner. D and E) Illustration of the strategy to identify malignant B cells. Single cell RNA expression profiles of B cells derived from the tFL1 sample were visualized by t-SNE. The different B cell clusters are circled and labeled with hB (healthy B cells), B1 (B cell cluster 1) and B2 (B cell cluster 2). For each single B cell we calculated the the kappa light chain (IGKC) fraction  $IGKC/(IGKC+IGLC2)$  (see color code D and E). If this IGKC-fraction was  $> 0.5$ , we classified a B cell as a kappa positive and if this ratio was below 0.5 we classified the B cell as a lambda positive. The percentage of B cells either expressing kappa or lambda per transcriptionally distinct B cell cluster was calculated. The non-malignant healthy B cell (hB) cluster contained approximately 50% kappa and 50% lambda expressing B cells while the tow malignant cluster (B1, B2) contained B cells homogeneously expressing the kappa light chain.

**Figure 2**

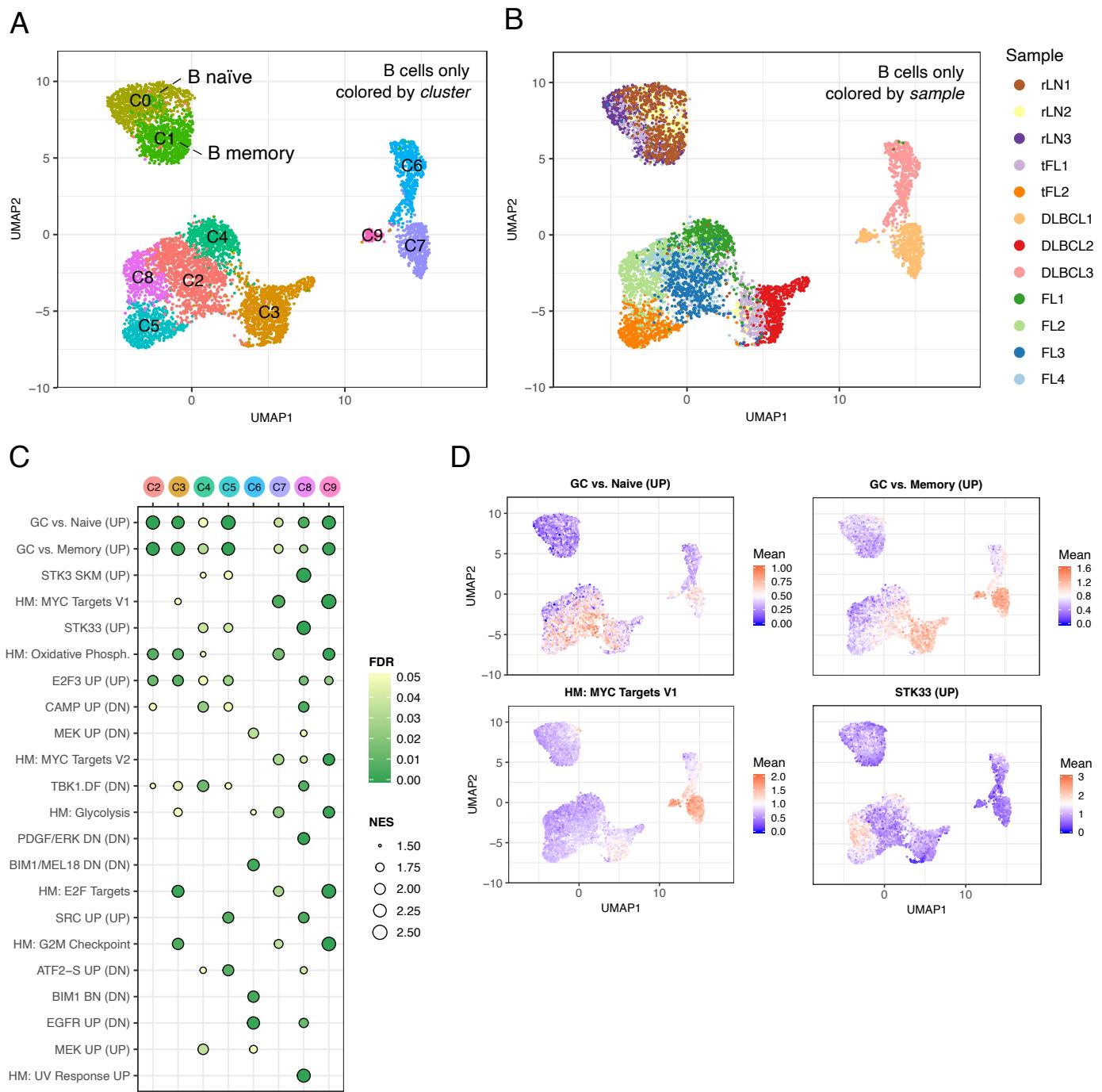


**Figure 2. The transcriptional heterogeneity of lymph node-derived T cells**

A and B) T cells from all samples were combined and jointly visualized using UMAP. Cells were colored with respect to their cluster of origin (A) or to their sample of origin (B). C) The heatmap shows differentially expressed genes, which were used to identify the T cell subsets: Cytotoxic T cells (T<sub>TOX</sub>), conventional T helper cells (T<sub>H</sub>), T follicular helper cells (T<sub>FH</sub>), regulatory T cells (T<sub>REG</sub>). Gene expression values were scaled to the

the maximum of each row. D) Stacked bar chart displaying the proportion of T cell subpopulations based on scRNA-seq identified in each sample. Note that the DLBCL1 sample is not shown here due to only five T cells identified in this sample. E and F) Single cell suspensions of lymph nodes derived from 39 different patients, including those passed to scRNA-seq, were characterized by flow cytometry. The four different T cell populations identified by single cell RNA-Seq were distinguished using the following marker panel: CD3, CD4, CD8, PD1, ICOS and FoxP3. Specifically,  $T_H$  were identified based on  $CD3^+CD4^+$  without the phenotype of  $T_{FH}$  or  $T_{REG}$ ;  $T_{TOX}$  were identified based on  $CD3^+CD8^+$ ;  $T_{FH}$  were identified based on  $CD3^+CD4^+ICOS^{High}PD1^{High}$ ; and  $T_{REG}$  were identified based on  $CD3^+CD4^+FoxP3^+$ . E) The frequencies based on flow cytometry were correlated with the frequencies based on scRNA-seq. Pearson's correlation coefficients ( $r$ ) and the number of samples included ( $n$ ) are given in the top left corner. Note that the DLBCL1 and tFL2 samples are not shown here due to the low number of T cells in the scRNA-seq data (DLBCL1) or the lack of material (tFL2). F) Frequencies for each subpopulation with regard to the sum of all T cells are shown. P values were calculated by the two-sided Wilcoxon's test comparing each Entity with rLN group, and corrected by Bonferroni method. Only significant differences are shown; \*\*  $\leqq$  p value  $\leqq 0.01$ , \*  $\leqq$  p value  $\leqq 0.05$ . SNN: Shared-nearest-neighbor-based. UMAP: Uniform Manifold Approximation and Projection. rLN: Reactive lymph node. MCL: Mantle cell lymphoma. FL: Follicular lymphoma. DLBCL: Diffuse large B cell lymphoma, CLL: Chronic lymphocytic leukemia.

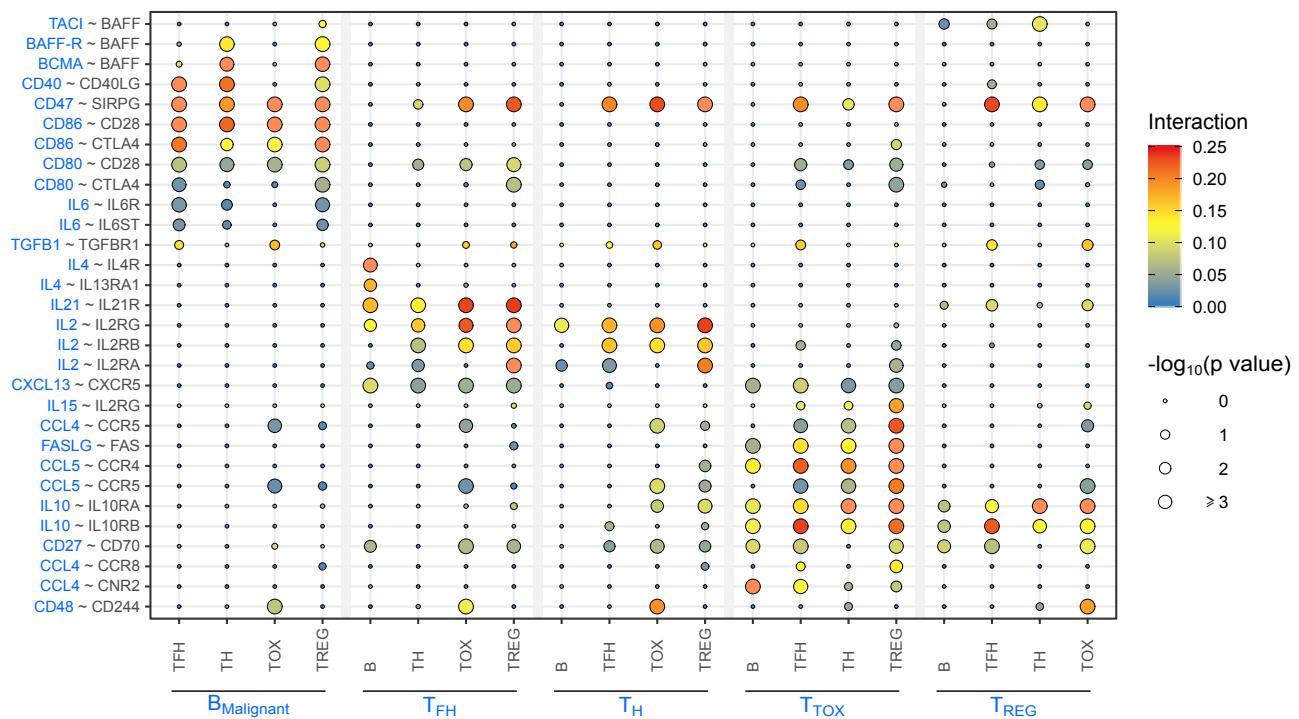
## Figure 3



**Figure 3. Gene expression signatures driving B cell heterogeneity.**

A and B) Single cell RNA expression profiles from all B cells were combined and jointly visualized using UMAP. Cells are colored either by SNN-based clusters (A) or by sample (B). C) A gene set enrichment analysis was performed separately for each malignant cluster (C3 to C9) versus all healthy B cells (C0, C1). The four most enriched gene sets per sample are shown. Columns refer to cluster. Circles are coded by color (nominal FDR) and size (NES). Gene sets with NES > 1.5 are shown. D) Cells in UMAP plot were colored by the mean expression of enriched genes for four representative gene expression signatures. UMAP: Uniform Manifold Approximation and Projection. SNN: Shared-nearest-neighbor. FDR: False-positive detection rate. NES: Normalized enrichment score.

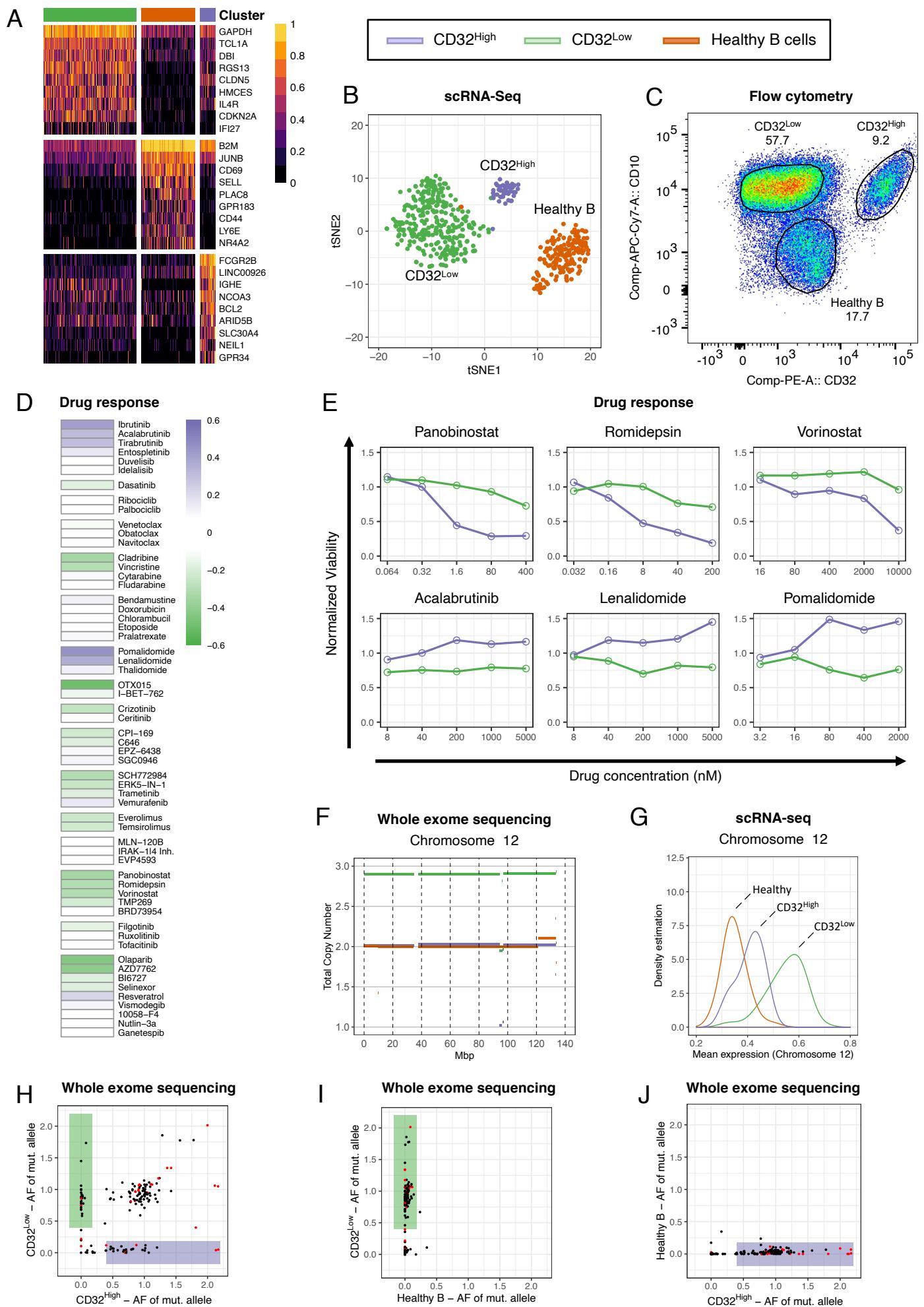
**Figure 4**



**Figure 4. Cellular crosstalk in B cell lymphoma within the lymph node microenvironment.**

Overview of most significant ligand-receptor interactions across all lymphoma samples, excluding DLBCL1 due to the low number of T cells. Circle size indicates negative  $\log_{10}$  of adjusted p values which were determined by permutation test (see Methods for details). Color scheme visualizes interaction scores which were calculated by the mean expression of molecule 1 (blue) in cell type A (blue) and the mean expression of molecule 2 (black) in cell type B (black). Protein names instead of gene names were used for TACI (*TNFRSF13B*), BAFF-R (*TNFRSF13C*), BCMA (*TNFRSF17*) and BAFF (*TNFSF13B*).

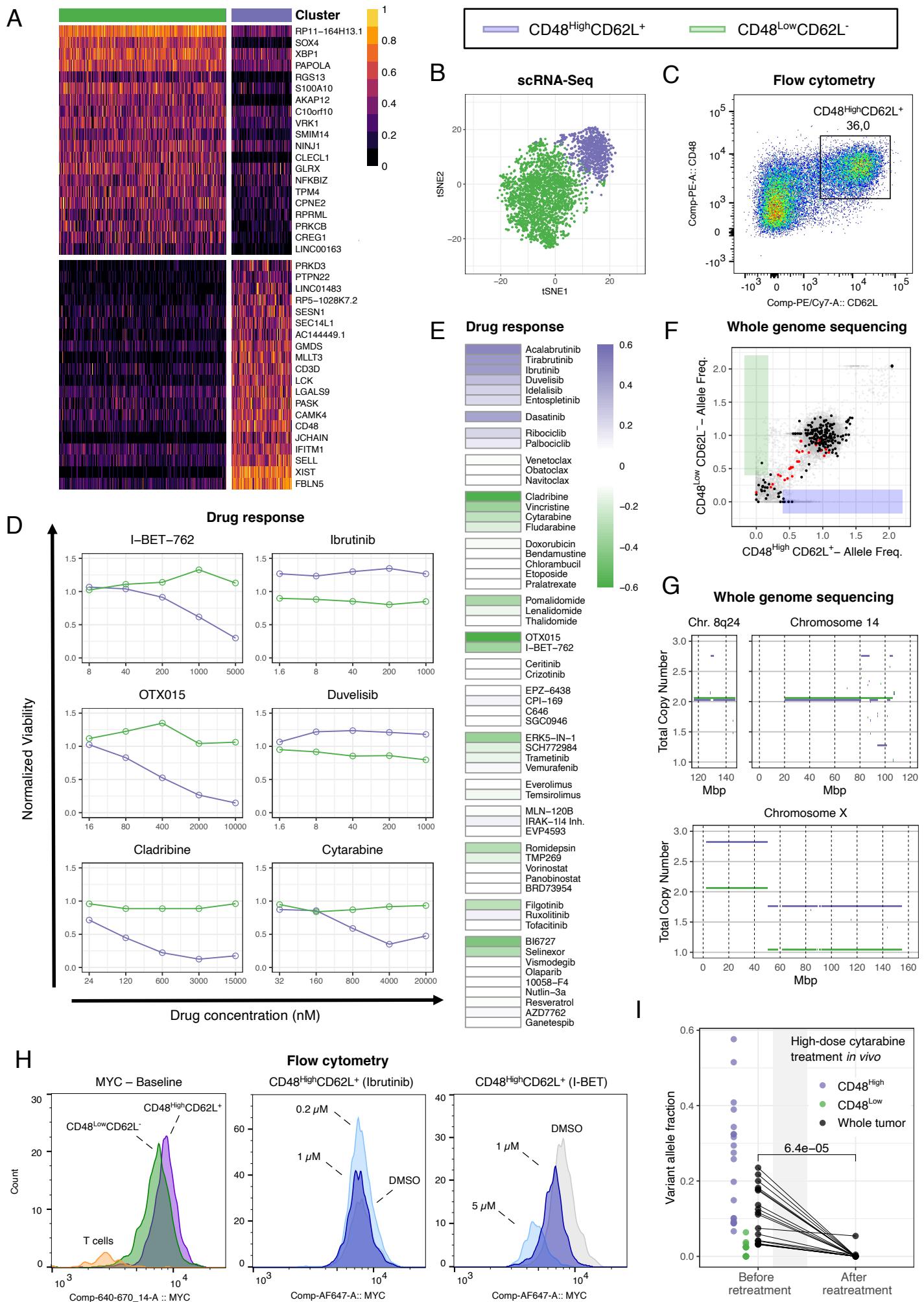
# Figure 5



### **Figure 5. In depth-analysis of sample tFL1**

A-B) Single cell RNA expression profiles of B cells derived from the tFL sample were subjected to SNN-based clustering. Three transcriptionally distinct clusters emerged. A) The heatmap illustrates the top 30 differentially expressed genes between all three identified clusters. Gene expression values were scaled to the maximum of each row. B) Clusters were colored and visualized in t-SNE projections of scRNA-seq expression profiles of malignant B cells. C) tFL1 derived lymph node cells were stained for viability, CD19, CD32, and CD10. The gates highlight three CD19<sup>+</sup> populations which correspond to the subclusters shown in panel B. D and E) Unsorted single cell suspensions from the tFL sample were incubated for 48 hours with 58 different drugs and five concentrations, and stained as described in panel C. Viability was normalized to vehicle control for each subpopulation separately. D) The mean difference of viabilities between the two malignant subclones is shown. White indicates that both malignant clones responded equally to this drug. Purple or green indicates that the viability of the CD32<sup>High</sup> or CD32<sup>Low</sup> subpopulation was superior. E) Six representative subclone specific responses to the following drugs are shown: panobinostat, vorinostat, romidepsin (HDAC inhibitors), acalabrutinib (BCR signaling inhibitor), lenalidomide and pomalidomide (immunomodulatory imide drugs). F) Whole exome sequencing was performed on FACS sorted CD32<sup>High</sup>-, CD32<sup>Low</sup>- and the non-malignant CD10- B cell subset. The line plot shows the total copy number estimation for chromosome 12 for all three sorted populations. The CD32<sup>Low</sup> clone harbors an additional copy of chromosome 12. G) Density curves of single cell expression values for all genes located on chromosome 12 are shown for each subclone. H-J) The scatter plots show the allele frequency (AF) of the mutated allele for exonic SNVs in both malignant subclones (H), in CD32<sup>Low</sup> versus healthy B cells (I) and in CD32<sup>High</sup> versus healthy B cells (J). Shaded purple or green boxes highlight SNV that are exclusive to one of the malignant B cell subclones. Red dots mark immunoglobulin-associated mutations. SNN: Shared-nearest-neighbor. HDAC: Histone deacetylase. BCR: B cell receptor. SNV: Single nucleotide variant.

# Figure 6



## Figure 6. In depth-analysis of sample DLBCL1

A-B) Single cell RNA expression profiles of malignant B cells derived from the DLBCL1 sample were subjected to SNN-based clustering. Two transcriptionally distinct clusters emerged. A) Top 40 differentially expressed genes between the two identified clusters are shown in the heatmap. Gene expression values were scaled to the maximum of each row. B) Clusters were colored and visualized in t-SNE projections of scRNA-seq expression profiles of malignant B cells. C) DLBCL1 derived lymph node cells were stained for viability, CD19, CD48 and CD62L (=SELL). The gate highlights a population which co-expresses CD62L and CD48, which represents the identified subclones. D and E) Lymph node derived cells from the DLBCL1 sample were incubated for 48 hours with 58 different drugs and 5 concentrations. Cells were stained as described in C. Viability was normalized to DMSO controls for each subpopulation separately. D) Six representative subclone-specific responses to the following drugs are shown: I-BET-762, OTX015 (BET inhibitors), ibrutinib, duvelisib (BCR signaling inhibitors), cytarabine and cladribine (chemotherapy). E) The mean difference of viabilities between the two subpopulations is shown. White indicates that both clusters responded equally to this drug. Purple or green indicates that the viability of the  $CD48^{\text{High}}CD62L^+$  or  $CD48^{\text{Low}}CD62L^-$  subpopulation was superior. F-G) Whole genome sequencing was performed on both FACS sorted populations ( $CD48^{\text{High}}CD62L^+$ ,  $CD48^{\text{Low}}CD62L^-$ ). F) The scatter plot shows the allele frequency (AF) of the mutated allele for non-synonymous exonic SNV in bold black and synonymous or intronic SNV in faded grey of both subclones. Shaded purple or green boxes highlight SNV that are exclusive to one or the other subclone. Red dots mark immunoglobulin-associated non-synonymous exonic mutations. G) Line plots show total copy number estimations for chromosome 8q24, 14 and X for both clones. H) DLBCL1 derived lymph node cells were incubated with DMSO control, I-BET-762 at two concentrations (1  $\mu\text{M}$ , 5  $\mu\text{M}$ ) or ibrutinib at two concentrations (0.2  $\mu\text{M}$ , 1  $\mu\text{M}$ ). At baseline and after 24 hours cells were harvested and stained for viability, CD19, CD3, CD48, CD62L and MYC or respective isotype control. Histograms show the fluorescence intensity of MYC at baseline for T cells,  $CD48^{\text{High}}CD62L^+$  and  $CD48^{\text{Low}}CD62L^-$  subclone, after 24 hours incubation with I-BET-762 and DMSO control or Ibrutinib and DMSO control. I) Shown are SNV with high variant allele frequencies in the  $CD48^{\text{High}}CD62L^+$  subpopulation (purple) and low or undetectable in  $CD48^{\text{Low}}CD62L^-$  subpopulation (green). Black circles show corresponding variant allele frequencies of whole tumor samples before and after retreatment with high-dose cytarabine. P value was calculated by the paired Wilcoxon-test.

SNN: Shared-nearest-neighbor. BET: Bromodomain and Extra-Terminal motif protein. BCR: B cell receptor. SNV: Single nucleotide variant. DMSO: Dimethyl sulfoxide.