

**Supplementary Information to Genomic and transcriptomic changes complement each other in the
pathogenesis of sporadic Burkitt Lymphoma**

López, Kleinheinz, Aukema, Rohde, Bernhart, Hübschmann, Wagener, Toprak et al.. 2018

Table of contents:

1. Supplementary Note 1:

Full list of members of the Deutsche Krebshilfe Network Project

"Molecular Mechanisms in Malignant Lymphoma" (MMML) (alphabetical order)

:3

2. Supplementary Figures

:4-16

3. Supplementary Table

:17

4. References

:18

Supplementary Note 1.

Full list of members of the MMML

Pathology group and analytes preparation: Thomas F.E. Barth¹, Heinz-Wolfram Bernd², Sergio B. Cogliatti³, Alfred C. Feller², Martin L. Hansmann⁴, Michael Hummel⁵, Wolfram Klapper⁶, Dido Lenze⁵, Peter Möller¹, Hans-Konrad Müller-Hermelink⁷, German Ott⁷, Andreas Rosenwald⁷, Harald Stein⁵, Monika Szczepanowski⁶, Hans-Heinrich Wacker⁶.

Genetics group: Thomas F.E. Barth¹, Petra Behrmann⁸, Peter Daniel¹⁰, Judith Dierlamm⁸, Eugenia Haralambieva⁷, Lana Harder¹¹, Paul-Martin Holterhus¹², Ralf Küppers¹³, Dieter Kube¹³, Peter Lichter¹⁴, Jose I. Martín-Subero¹¹, Peter Möller¹, Eva M. Murga-Peñas⁹, German Ott⁷, Christiane Pott¹⁶, Armin Pscherer¹⁵, Andreas Rosenwald⁷, Carsten Schwaenen¹⁷, Reiner Siebert¹⁸, Heiko Trautmann¹⁶, Martina Vockerodt¹⁹, Swen Wessendorf¹⁶.

Bioinformatics group: Stefan Bentink²⁰, Hilmar Berger²¹, Dirk Hasenclever²¹, Markus Kreuz²¹, Markus Loeffler²¹, Maciej Rosolowski²¹, Rainer Spang²⁰.

Project coordination: Benjamin Stürzenhofecker¹⁴, Lorenz Trümper¹⁴, Maren Wehner¹⁴.

Steering committee: Markus Loeffler²¹, Reiner Siebert¹⁸, Harald Stein⁵, Lorenz Trümper¹⁴.

¹Institute of Pathology, University Hospital of Ulm, Ulm, Germany;

²Institute of Pathology, University Hospital Schleswig-Holstein Campus Lübeck, Lübeck, Germany;

³Institute of Pathology, Kantonsspital St. Gallen, St. Gallen, Switzerland;

⁴Institute of Pathology, University Hospital of Frankfurt, Frankfurt, Germany;

⁵Institute of Pathology, Campus Benjamin Franklin, Charité–Universitätsmedizin Berlin, Berlin, Germany;

⁶Institute of Hematopathology, University Hospital Schleswig-Holstein Campus Kiel/ Christian-Albrechts University Kiel, Kiel, Germany;

⁷Institute of Pathology, University of Würzburg, Würzburg, Germany;

⁸Cytogenetic and Molecular Diagnostics, Internal Medicine III, University Hospital of Ulm, Ulm, Germany;

⁹University Medical Center Hamburg-Eppendorf, Hamburg, Germany;

¹⁰Department of Hematology, Oncology and Tumor Immunology, University Medical Center Charité, Berlin, Germany;

¹¹Institute of Human Genetics, University Hospital Schleswig-Holstein Campus Kiel/Christian-Albrechts University Kiel, Kiel, Germany;

¹²Division of Pediatric Endocrinology and Diabetes, Department of Pediatrics, University Hospital Schleswig-Holstein Campus Kiel / Christian-Albrechts University Kiel, Kiel, Germany;

¹³Institute for Cell Biology (Tumor Research), University of Duisburg-Essen, Essen, Germany;

¹⁴Department of Hematology and Oncology, Georg-August University of Göttingen, Göttingen, Germany;

¹⁵German Cancer Research Center (DKFZ), Heidelberg, Germany;

¹⁶Second Medical Department, University Hospital Schleswig-Holstein Campus Kiel/ Christian-Albrechts University Kiel, Kiel, Germany;

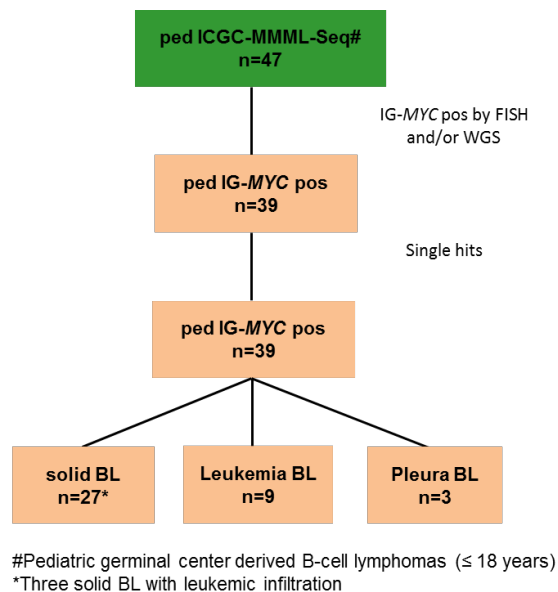
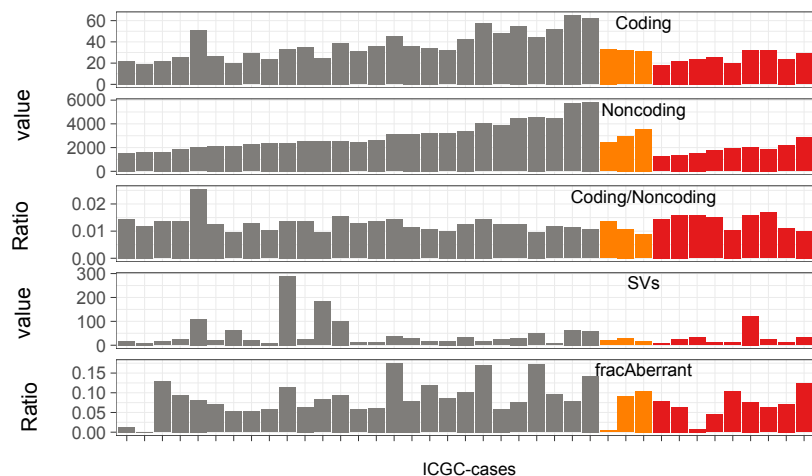
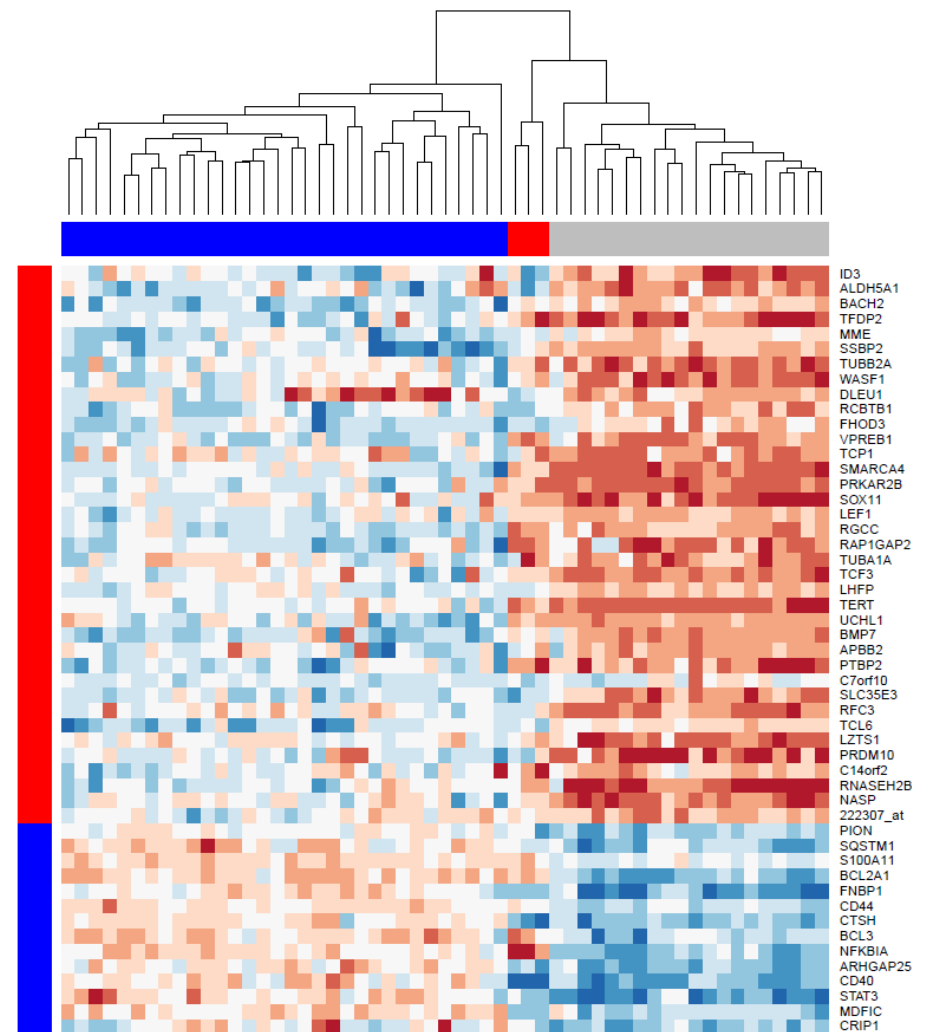
¹⁷Cytogenetic and Molecular Diagnostics, Internal Medicine III, University Hospital of Ulm, Ulm, Germany;

¹⁸Institute of Human Genetics, University of Ulm and University Hospital of Ulm, Ulm, Germany

¹⁹Department of Pediatrics I, Georg-August University of Göttingen, Göttingen, Germany;

²⁰Institute of Functional Genomics, University of Regensburg, Regensburg, Germany;

²¹Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany.

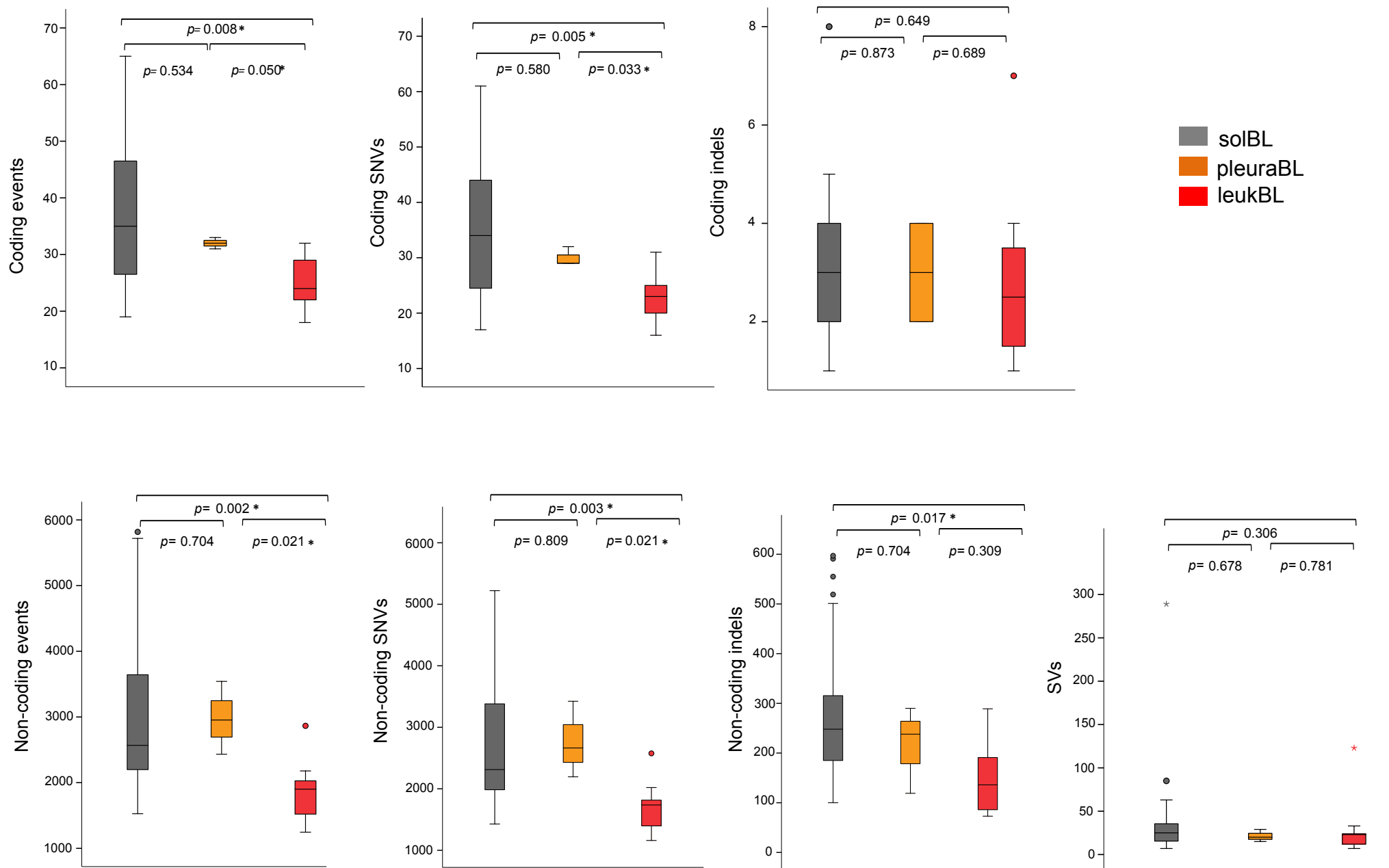
a**c****b**

Supplementary Figure 1. Characteristics of the cohort and number of events detected by WGS.

a Inclusion criteria of the BL cases of the ICGC MMML-Seq cohort analyzed.

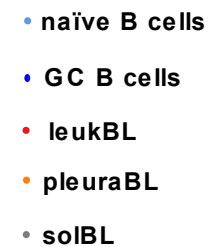
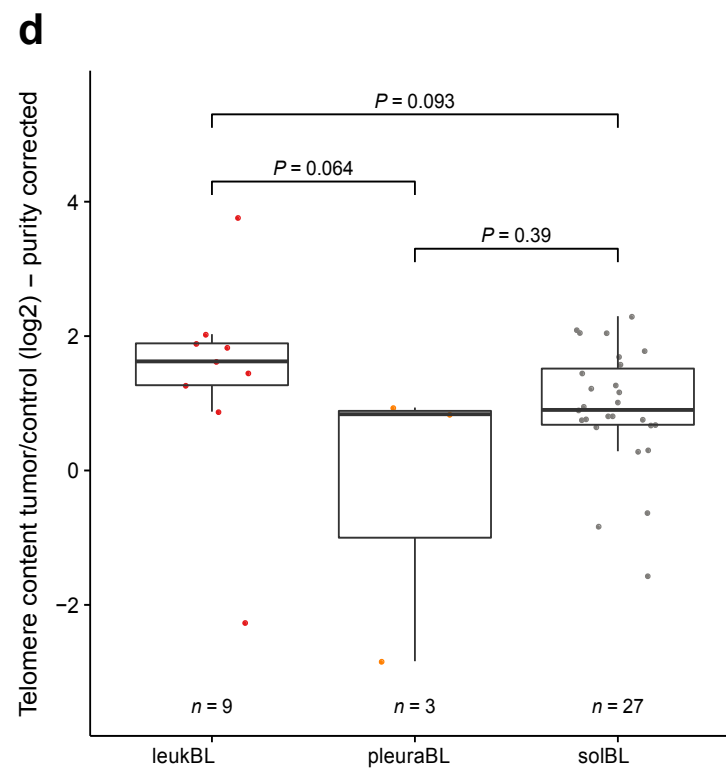
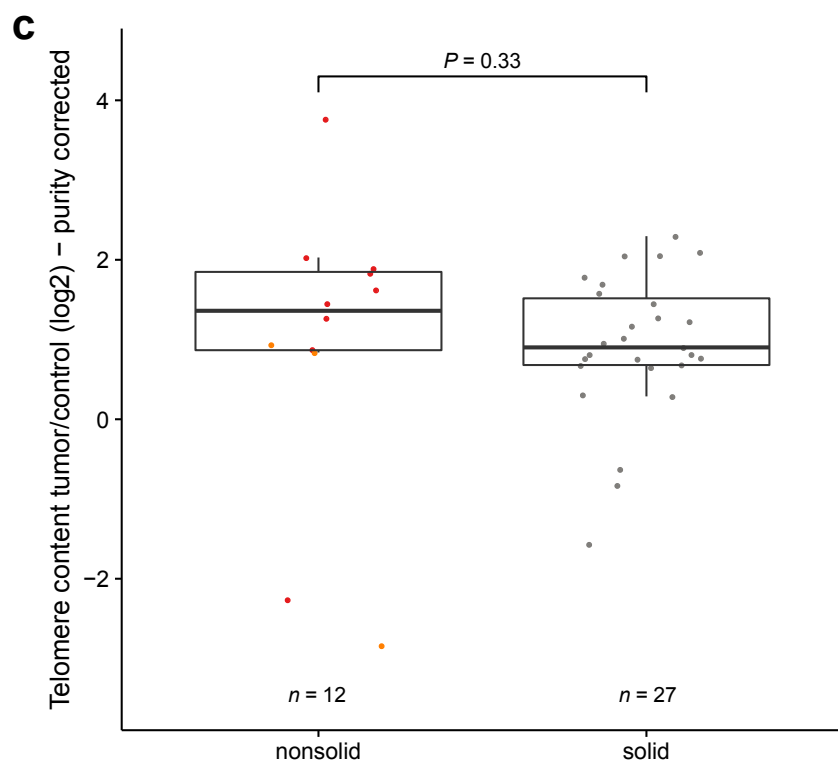
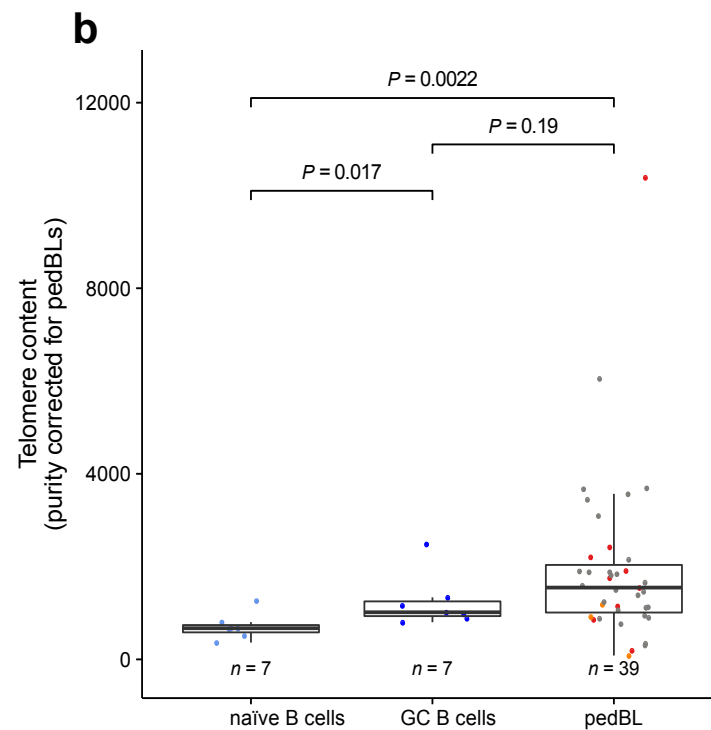
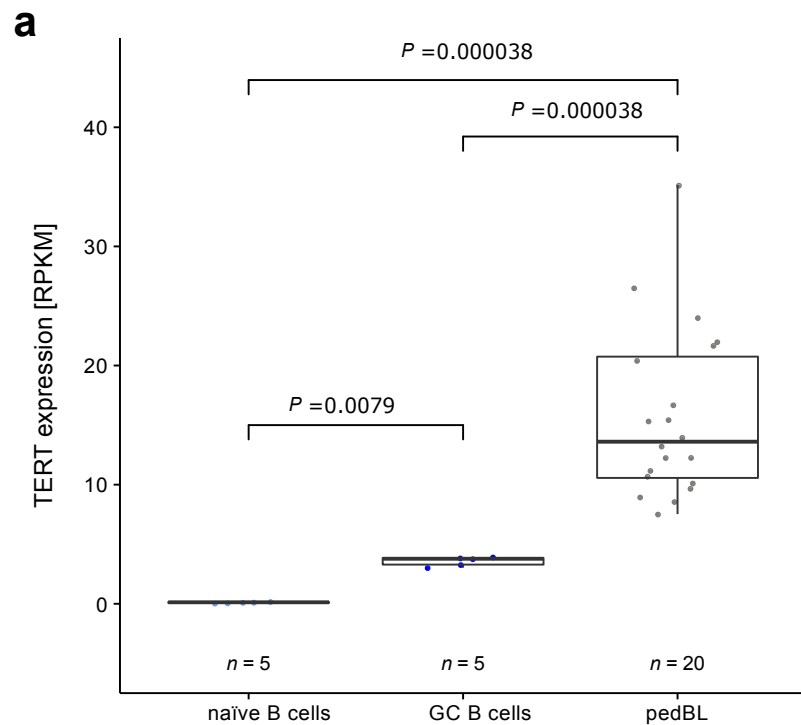
b The heatmap comprises all $n=51$ genes of the mBL vs. non-mBL expression classifier. Patients of the ICGC MMML-Seq cohort with available transcriptome data were clustered hierarchically and labelled: prototypic non-BL cases (blue), leukBL cases (red), and solBL cases (gray). Genes upregulated in BL are marked in red, genes downregulated in BL are marked in blue.

c Total number of coding and non-coding events (SNVs, indels, SVs) and fraction of aberrations in BL cohort (grey: solBL, orange: pleuraBL, and red: leukBL).



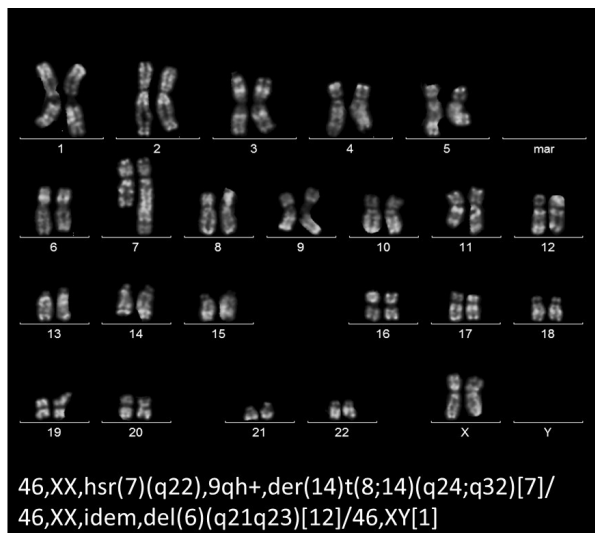
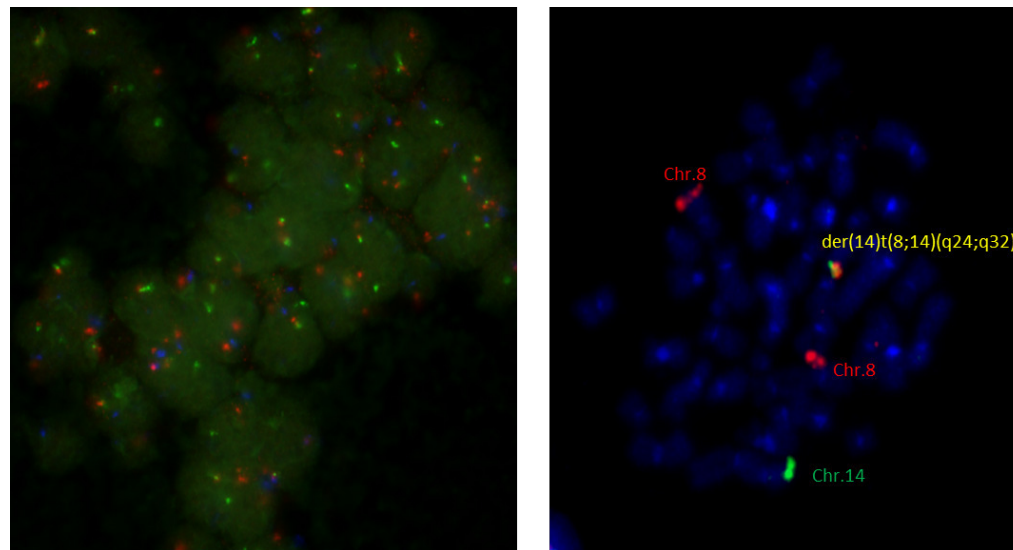
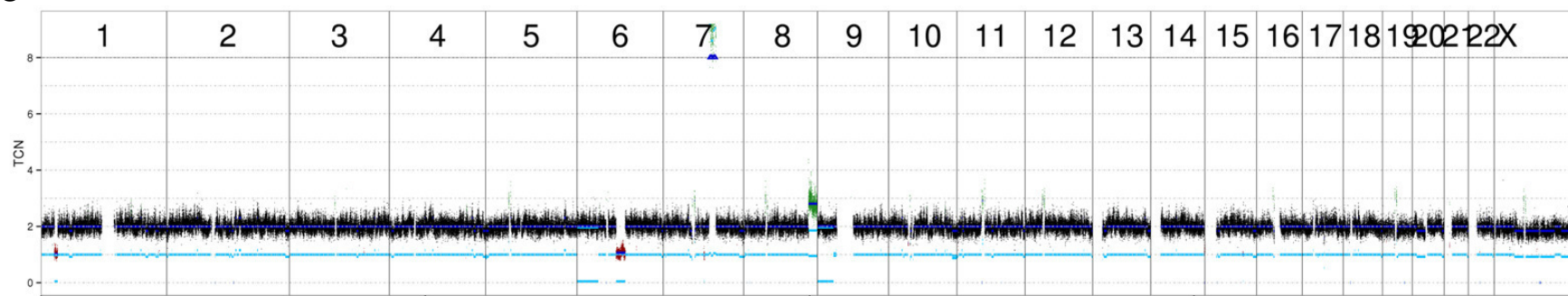
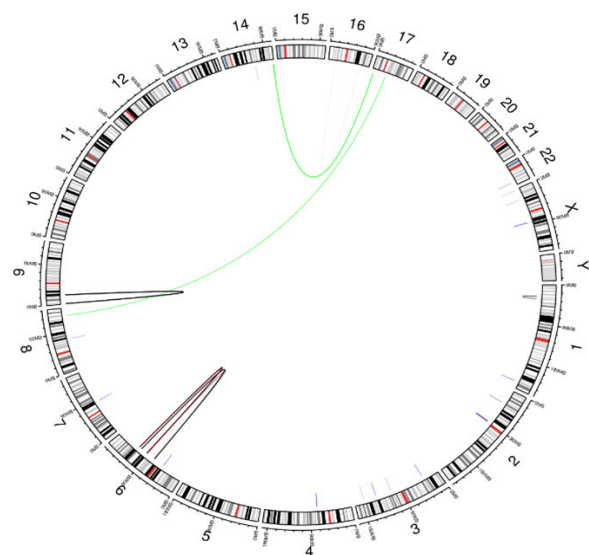
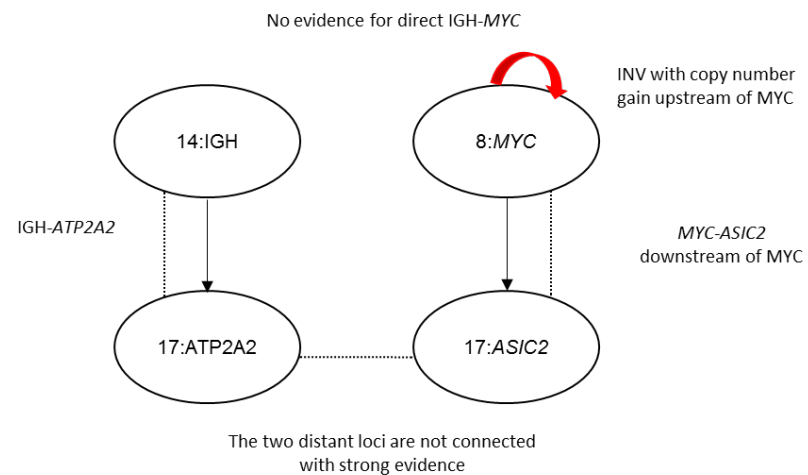
Supplementary Figure 2. Box-plot of genomic events detected by WGS in BL cohort.

The upper panel shows the total number of coding events, coding SNVs, and coding indels (upper part), distributed over the solBL (grey), pleuraBL (orange), leukBL (red). The lower panel shows the total number of non-coding events, non-coding SNVs, non-coding indels, as well as SVs across the cohort (labeled as described above). Statistical analyses were performed using Kruskal-Wallis test for all the variables described. Significance considered as p-value ≤ 0.05 is indicated by an asterisk (*).



Supplementary Figure 3. TERT expression and telomere content

- a** TERT transcript expression in the pedBL cohort (only considering solBL) is significantly higher compared to the expression in normal naive and germinal center (GC) B cell populations.
- b** Telomere content in pedBL cohort as compared to normal naïve B cells and normal germinal center B cells showing no significantly elevated telomere content in BL as compared to GC B cells. Naive B cells do have a significantly lower telomere content as compared to both, normal GC B cells and pedBL.
- c** Telomere content in non-solid BL versus solid BL, showing no significant difference between both groups.
- d** Telomere content across the BL subgroups (leukBL, solBL, and pleuraBL) showing no significant differences between the subgroups.

a**b****c****d****e**

Supplementary Figure 4. Integrative analysis of a complex *MYC-ASIC2*-IGH rearrangement in case 4177434.

a Karyotype showing the der(14)t(8;14)(q24;q32) as consequence of an unbalanced translocation between one IGH locus on chromosome 14 and one *MYC* locus in chromosome 8.

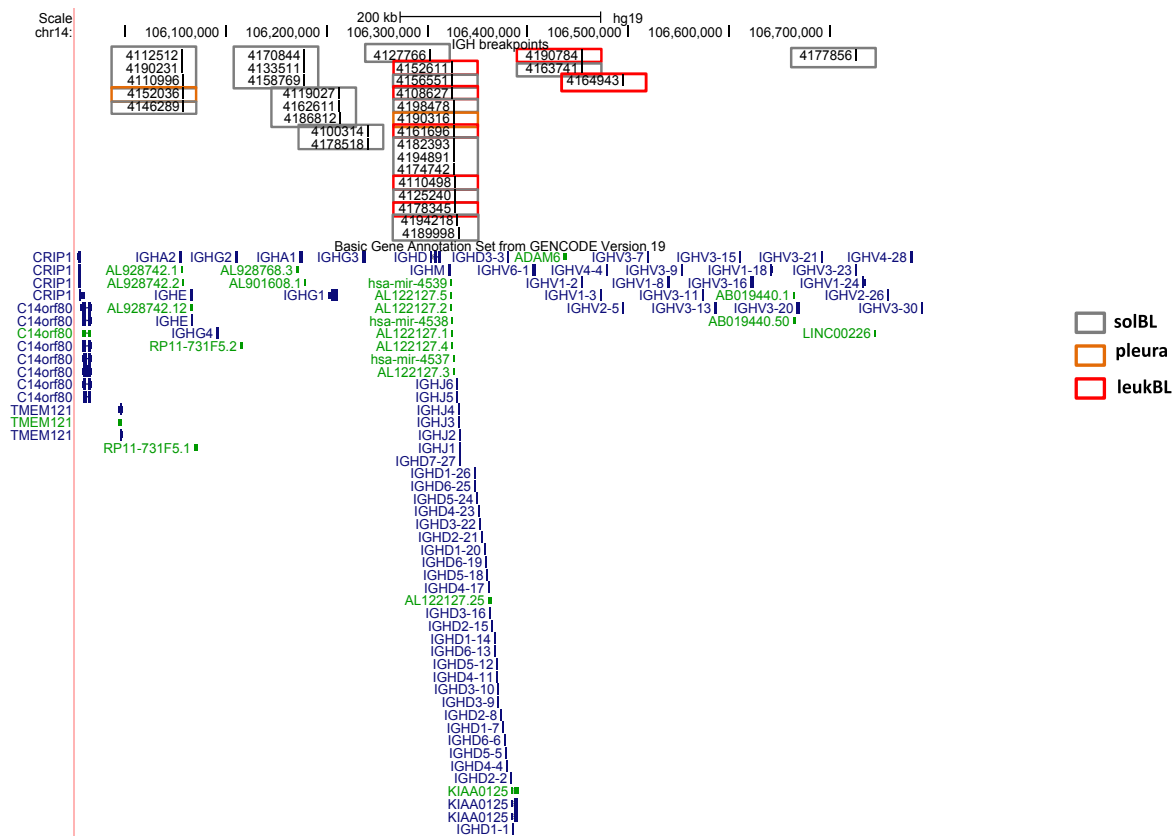
b IGH-*MYC* dual-color, dual fusion probe in interphase nuclei (left) and metaphase (right) cells showing one fusion (yellow) signal in der(14)t(8;14)(q24;q32), two red signals in chromosomes 8 and 1 green signal in chromosome 14.

c Imbalance profile displaying gains (in green) and losses (in red) for each chromosome. Remarkably, the amplicon in chromosomal region 7q includes *CDK6* gene.

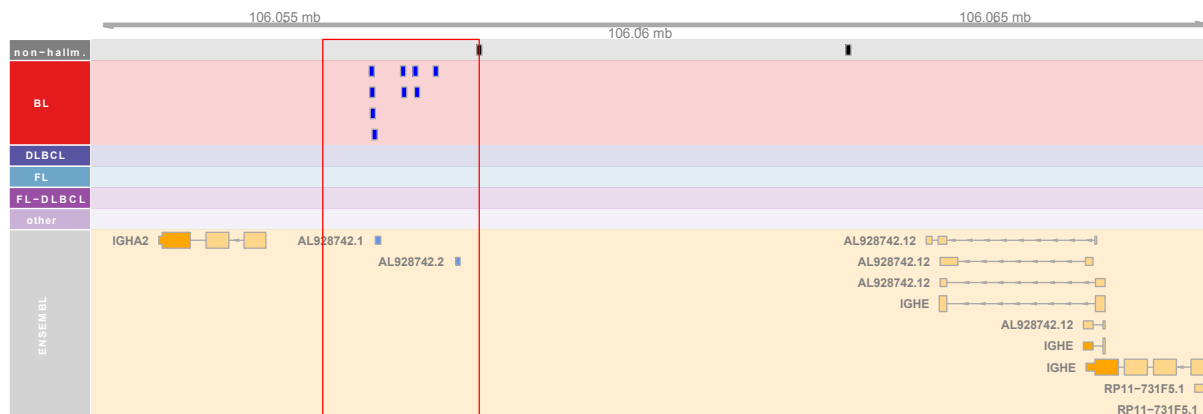
d Circle plot indicating structural variants. The green lines represent translocation events, blue lines deletions, red lines duplications, black lines inversions.

e Schematic of the rearrangement events: A three-way translocation involving the IGH-*ASIC2*-*MYC* genes were detected by WGS. A breakpoint downstream of *MYC* detected by WGS is fused to the *ASIC2* gene locus in chromosome 17. Subsequently, the *ATP2A2* gene also located in chromosome 17 is translocated to the IGH locus.

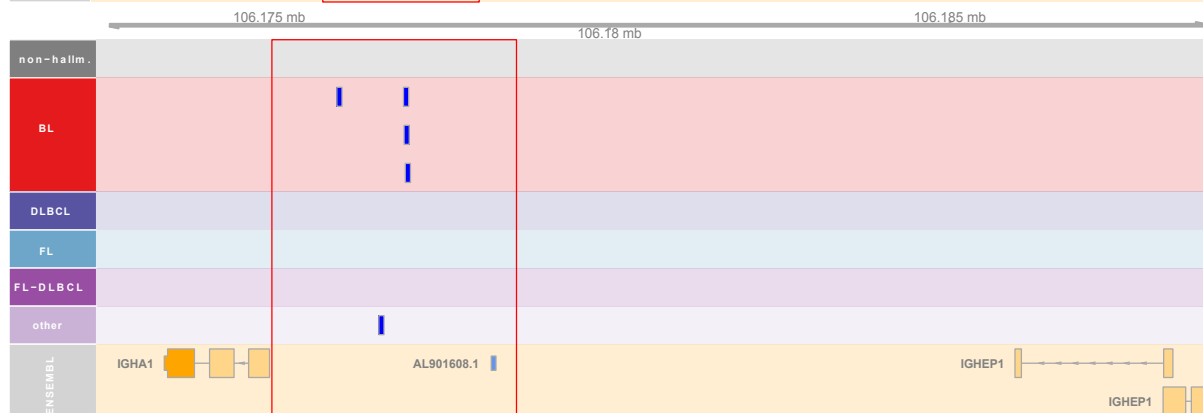
a



b



C

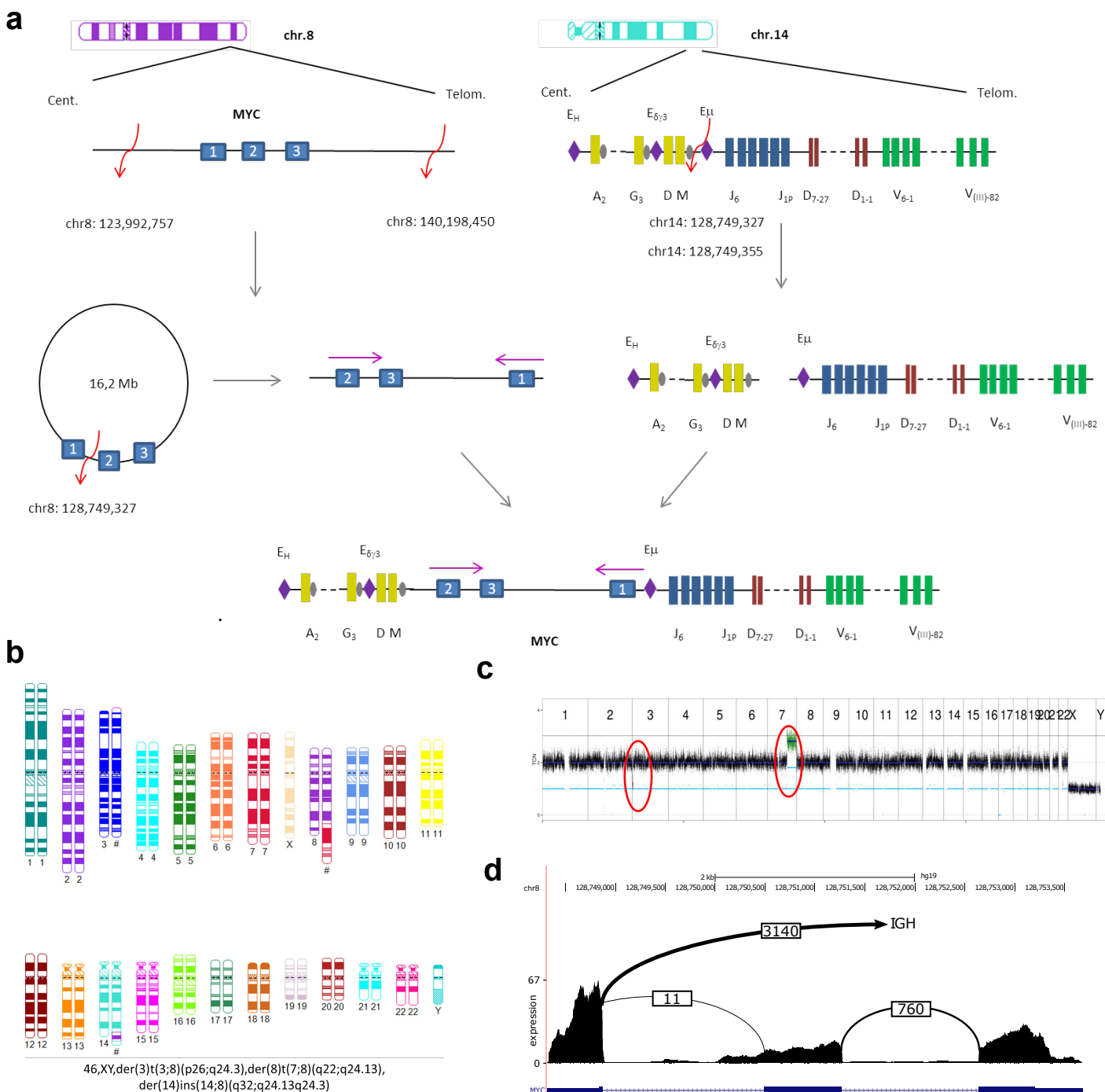


Supplementary Figure 5. Distribution of breakpoints within the IGH locus.

a Display of the breakpoints within the IGH locus for each case. The IGH-breakpoints of the solBL (grey) and the pleuraBL (orange) were located downstream of the IGHM region, the leukBL (red) had the breakpoints in IGHM or downstream of IGHM region.

b Genomic view showing the IGHA2 region including the switch $\alpha 2$ region (according to ENSEMBL) and displaying the breakpoint positions (some cases harbor more than one breakpoint) in this genomic regions across the ICGC MMML-Seq cohort, including DLBCL, FL, FL-DLBCL, and others like DH-BL lymphoma (Hübschmann et al., under revision). The red square defines the switch region.

c Genomic region showing the IGHA1 region (switch $\alpha 1$) (according to ENSEMBL) and displaying the distribution of the breakpoints (some cases containing more than one breakpoint) in this genomic region across the ICGC MMML-Seq cohort (as describe above). The red square highlights the switch region.



Supplementary Figure 6. Genomic and transcriptomic data of the BL case 4125240

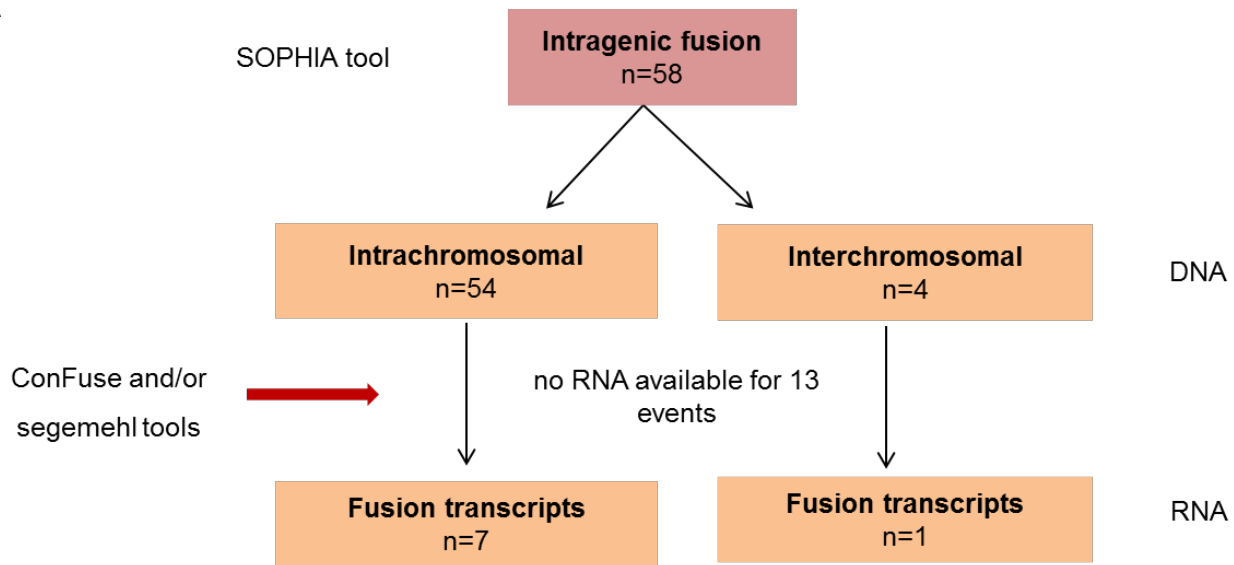
a Representation of an insertion event of *MYC* into the *IGH* locus. Two breaks, indicated by red arrows upstream and downstream of the *MYC* gene in case 4125240 generated a circular DNA fragment containing the *MYC* gene, which was subsequently opened between exon 1 and 2. The opened fragment contained exon 1 of *MYC* in opposite direction of exons 2 and 3. Insertion of this fragment into chromosome 14 brought *MYC* under the regulation of the enhancers of the *IGH* locus and lead to the independent expression of exon 2 and exon 3 via $E_{\delta\gamma 3}$, and from exon 1, which is regulated via E_{μ} enhancer.

b Karyotype predicted using the Cydas online program (<http://www.cydas.org/>) and based on FISH and structural variant analysis.

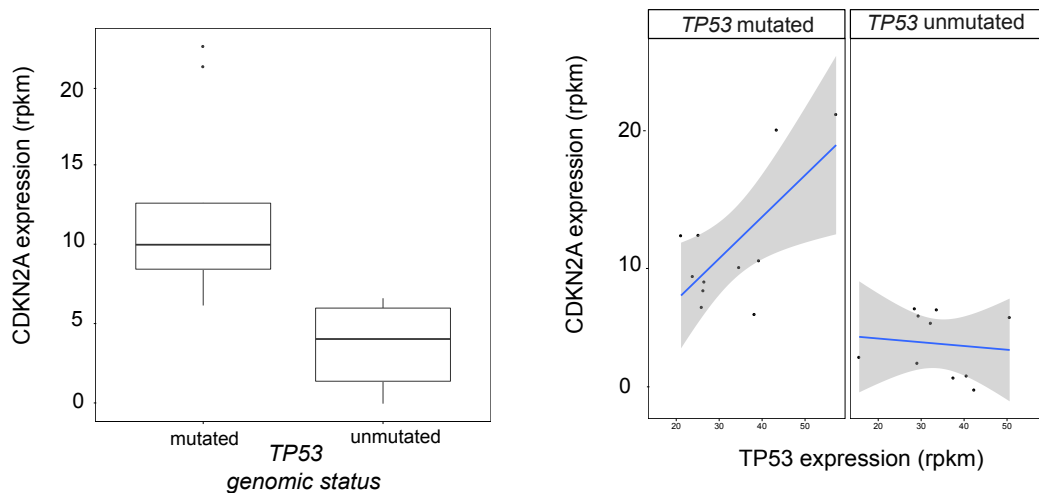
c Imbalance profile from WGS data, showing the loss of 3p and gain of 7q (highlighted with the red circles).

d Sashimi plot showing the number of splice reads between exon 1 and exon 2, exon 2 and exon 3, and exon 1 and *IGH* locus. The low number of splice reads between exon 1 and exon 2 corroborated the data observed by WGS data.

a



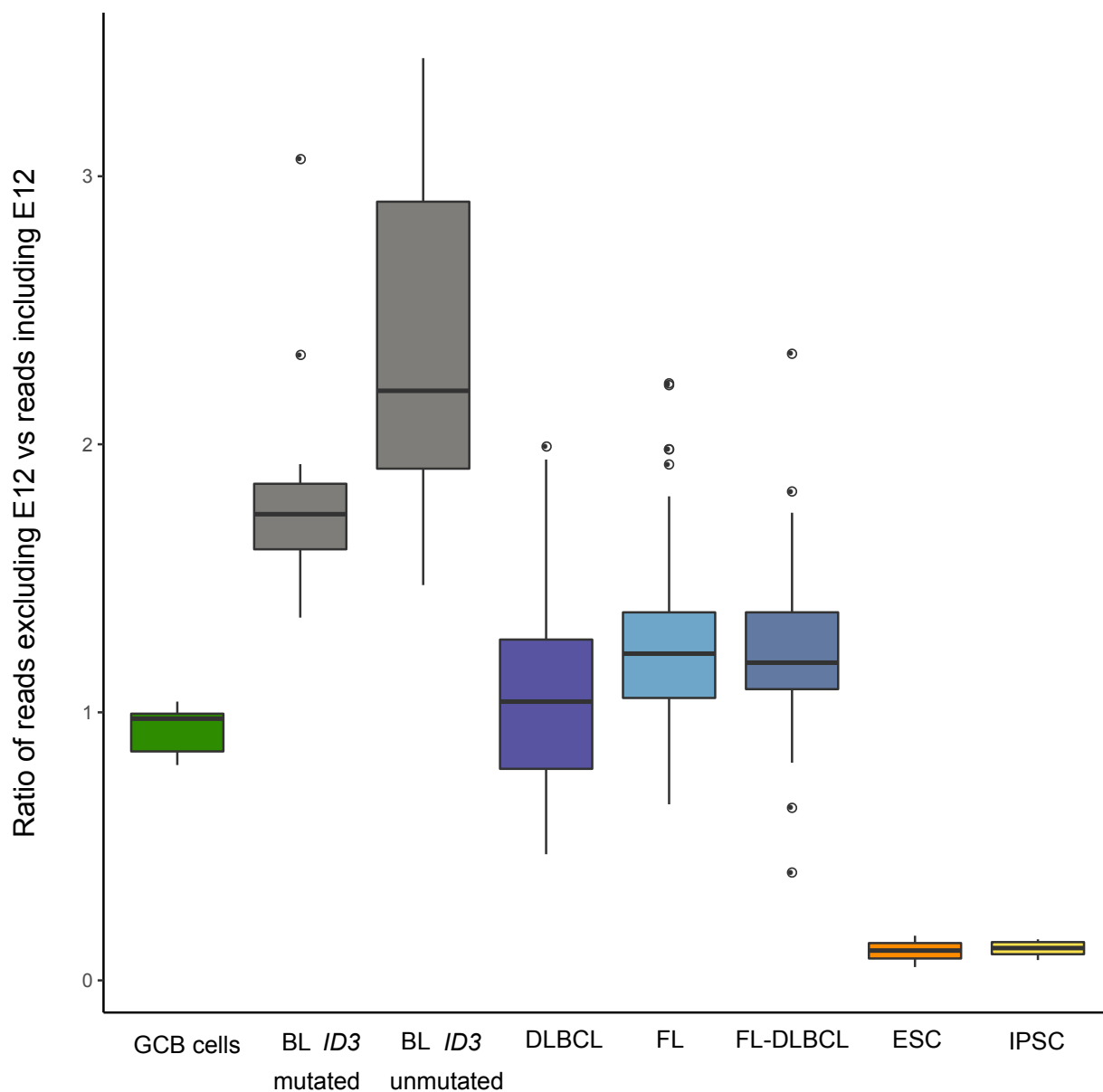
b



Supplementary Figure 7. RNA-seq analysis

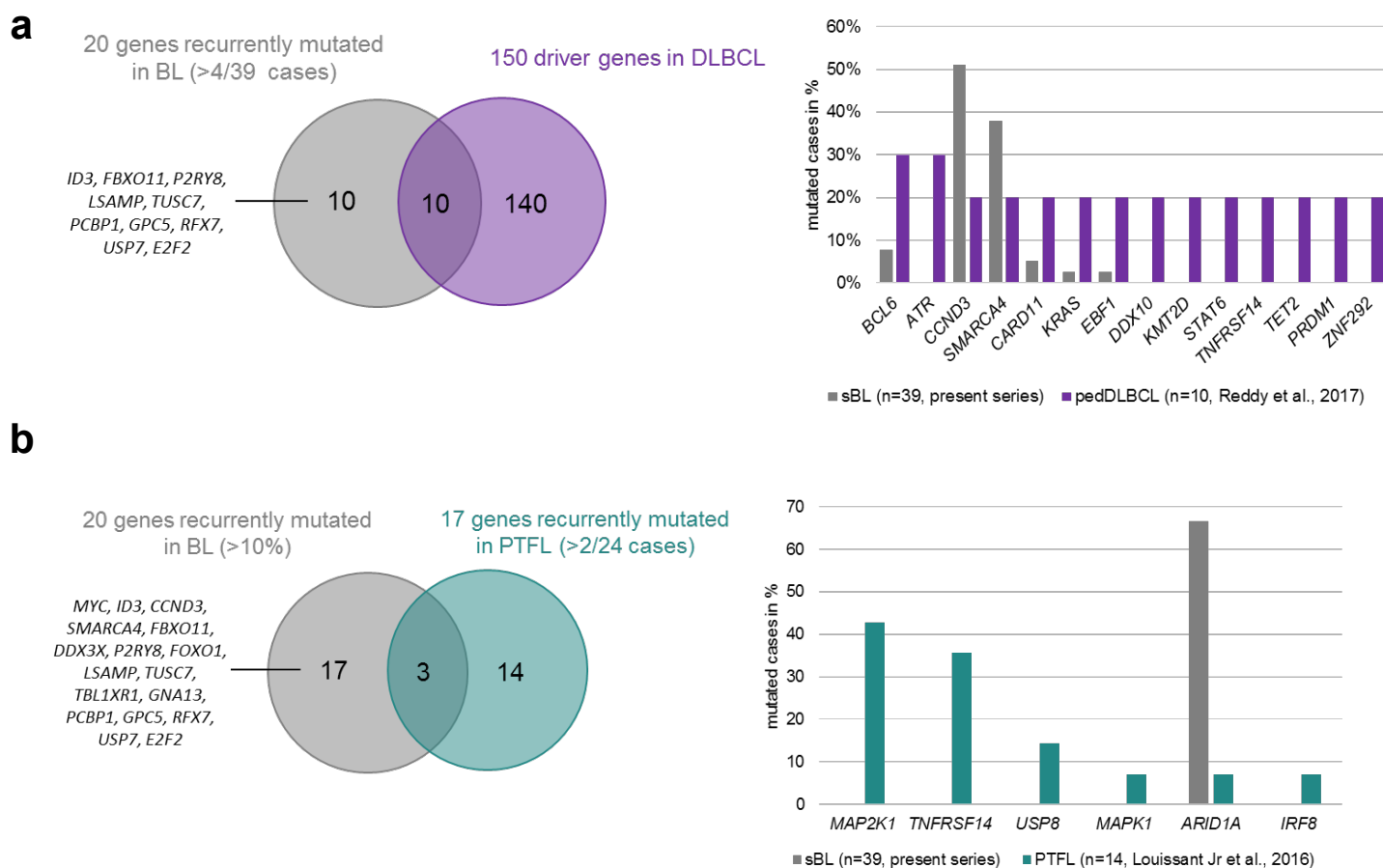
a Schematic work-flow showing the single steps applied for the detection of fusion genes and transcripts. The SOPHIA tool was applied to detect intragenic fusions at the genomic level using WGS data. After that, two independent tools, ConFuse and/or segemehl were applied to search for fusion transcripts at transcriptomic level using RNA-Seq data at the genomic fusions.

b Expression of CDKN2A and TP53 with regard to the *TP53* mutation status in solBL. The box plot on the left displays the CDKN2A expression (RPKM) on the Y axis related to the genomic mutations status of *TP53* on the X axis (mutated n=11, unmutated n= 10). The expression of CDKN2A is higher in solBL cases with *TP53* mutations than without ($p<0.0001$) (Wilcoxon rank sum test). The plot on the right depicts the correlation of TP53 expression (displayed as RPKM on the X axis) and CDKN2A expression (RPKM, Y axis), again separated by TP53 mutation status. Only in cases with *TP53* mutations exhibit a positive correlation ($p=0.021$) (Pearson test).



Supplementary Figure 8. TCF3 splicing

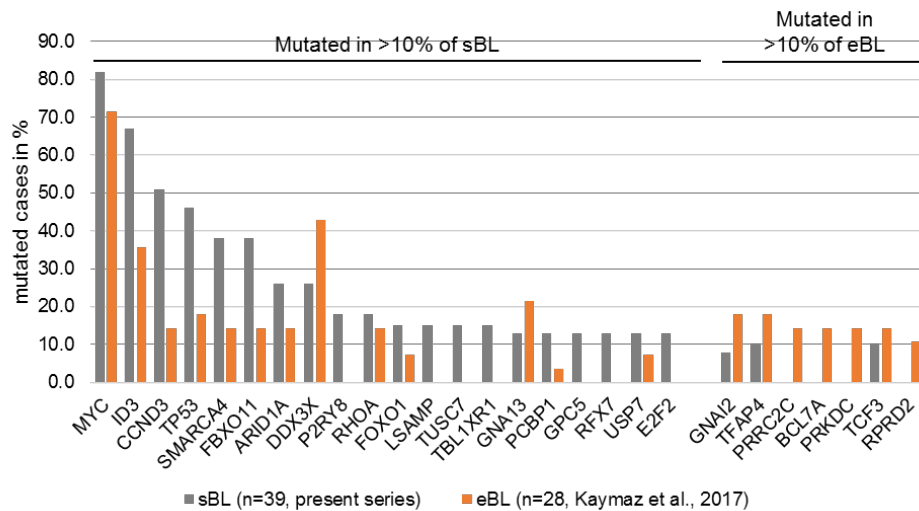
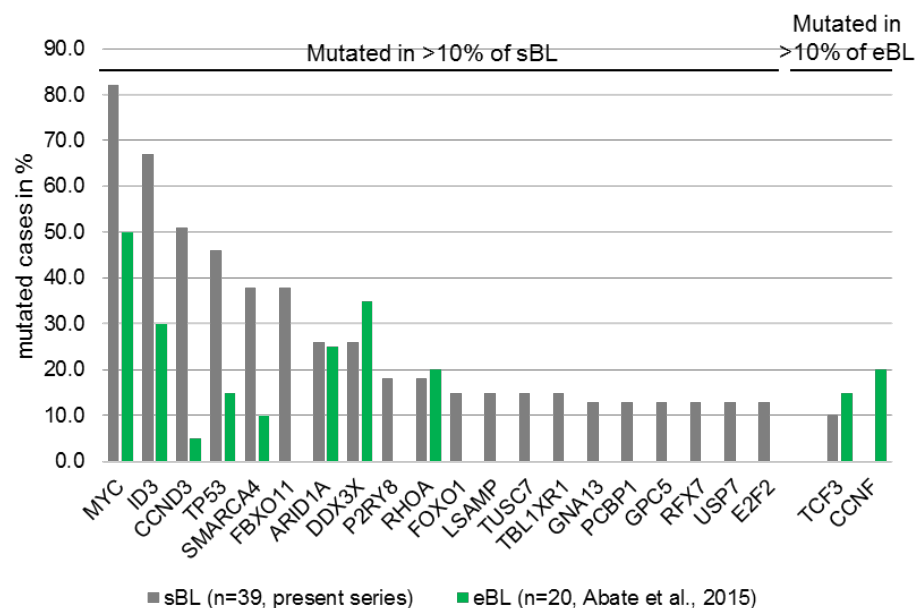
Boxplot showing the ratio of spliced reads that are directly spliced to exon E47 vs spliced reads that are spliced to exon E12. While half of the spliced reads splice to E47 in GCB cells (green), most of the BL cases have roughly two thirds of the spliced reads splicing to E47 (grey boxes) (all BL vs GCB cells, $p < 0.001$; see Figure 5b for comparison of ID mutated and unmutated BL). A higher fraction of reads excluding E12 than in GCB cells (green) is also observed, though only as a trend, in the non-BL group ($p = 0.052$), reaching significance in FL ($n = 85$, light blue box, $p = 0.008$) and FL-DLBCL ($n = 17$, bluish-grey box, $p = 0.031$), whereas no significant difference to GCB cells is present in DLBCL ($n = 72$, dark blue dark box, $p = 0.444$). In contrast, in the ESC ($n = 4$, orange) and iPSC ($n = 6$, yellow) the splice reads predominately splice to the E12 exon ($p = 0.024$ and $p = 0.008$, respectively, for comparison to GCB cells). For analyses of DLBCL, FL and FL-DLBCL, in which *ID3* mutations are rare (see Supplementary Figure 9), RNA-seq data of the ICGC MMML-Seq consortium (unpublished), which have been generated using the same sequencing pipeline like the one here applied to BL, were mined. For the analyses of ESC and iPSC we mined data from Friedi et al., 2014¹ and used the alignment tool STAR for mapping splice read. For statistical analysis the Wilcoxon rank sum test was applied and p-values were adjusted for multiple testing according to Benjamini-Hochberg.



Supplementary Figure 9. Comparison of the mutational landscape of pediatric sBL vs pediatric diffuse large B-cell lymphoma (DLBCL) and pediatric-type follicular lymphoma (PTFL)

a Comparison of recurrently mutated genes in pediatric sBL (n=20) to driver genes (n=150) identified in an overall cohort of 1001 DLBCL by Reddy et al., 2017². This cohort contains a subset of 10 pediatric DLBCL (≤ 18 years at diagnosis). On the left side, the Venn diagram shows that only 10 of the 20 recurrently mutated genes in sBL are reported as driver genes in DLBCL. Of those 150 driver genes, only 14 are recurrently mutated in pediatric DLBCL, i.e. affected in $\geq 2/10$ pediatric DLBCL cases. The diagram on the right side shows the relative amount of sBL and pediatric DLBCL cases being mutated in those 14 genes. Eight of these DLBCL driver genes recurrently mutated in pediatric DLBCL are not at all mutated in our sBL cohort and four in less than 10% of the sBL cases.

b Comparison of the mutational landscape of pediatric sBL to mutations in 17 genes identified by Louissaint A Jr et al.³ as frequently mutated in 24 PTFL (according to Figure 2A in Louissaint A Jr et al., 2016). This cohort contains a subset of 14 PTFL ≤ 18 years at diagnosis. On the left side, the Venn diagram shows that only 3 of the 20 frequently mutated genes in sBL are recurrently mutated in PTFL. Of the 17 frequently mutated genes in PTFL, only 6 are mutated in PTFL ≤ 18 years at diagnosis, i.e. affected in at least 1/14 PTFL ≤ 18 years at diagnosis. The diagram on the right side shows the relative amount of sBL and PTFL ≤ 18 years at diagnosis being mutated in those 6 genes. Five of these 6 genes are not at all mutated in our sBL cohort.

a**b**

Supplementary Figure 10. Comparison of pediatric sBL vs eBL

a/b Comparison of the frequently mutated genes in sBL (grey) and eBL based on analyses of eBL published by Kaymaz et al. 2017⁴ (a, orange) as well as Abate et al., 2015 (b, green). Only genes which were mutated in more than 10% of sBL (left sides) or in more than 10% of eBL (right side) of the respective studies were included. The comparison shows that overall sBL and eBL share a comparable mutational landscape with regard to affected genes, but with some variation in the percentages of cases affected per gene. Whereas eBL harbor a lower amount of cases with mutations in genes like *MYC*, *ID3*, *CCND3* or *SMARCA4*, they show a higher frequency of mutations in genes like *DDX3X*, *TCF3* or *GNAI2*. These findings are in line with the observations published Kaymaz et al.⁴ and Abate et al.⁵.

Supplementary Table 1. Clinical and biological features of the cohort.

The table gives an overview of the main clinical characteristics of the cohort and a comparison with previously published cohort from B-NHL a population in BFM trials.

		Overall B-NHL population in NHL- BFM trials*	Whole cohort in this study (n=39)	Subset solBL (n=27)	Subset leukBL (n=9)	Subset pleuraBL (n=3)
Median age at diagnosis, years [range]		8.4 [1-18]	8 [2-18]	8 [2-18]	7 [4-10]	11 [7-14]
Male; n (%)		873 (87%)	32 (82%)	23 (85%)	7 (78%)	2 (67%)
St Jude	B-AL	211 (21%)	10 (26%)	1 (4%)	9 (100%)	0 (0%)
	Stage I	90 (9%)	1 (2%)	1 (4%)	0 (0%)	0 (0%)
	Stage II	221 (22%)	3 (8%)	3 (11%)	0 (0%)	0 (0%)
	Stage III	402 (40%)	22 (56%)	19 (70%)	0 (0%)	3 (100%)
	Stage IV	80 (8%)	3 (8%)	3 (11%)	0 (0%)	0 (0%)
CNS involvement (%)		80 (8%)	4 (10%)	0 (0%)	3 (33%)	0 (0%)
BM involvement (%)		261 (26%)	11/39 (28%)	2/27 (7%)	9/9 (100%)	0/3 (0%)
Events (n)		129 (13%)	2 (6%)	2 (8%)	0	0

*as published by Burkhardt et al, 2005, British Journal of Haematology, doi:10.1111/j.1365-2141.2005.05735.x

n= number; solBL: solid BL; leukBL: leukemia BL; pleuraBL: pleura BL

CNS = central nervous system

BM = bone marrow

B-AL= Burkitt leukemia

B-NHL= Mature B cell non-Hodgkin lymphoma; NHL-BFM=Non-Hodgkin lymphoma Berlin-Frankfurt-Münster trial

References

1. Friedli, M. *et al.* Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res.* **24**, 1251-1259 (2014).
2. Reddy, A. *et al.* Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* **171**, 481-494.e15 (2017).
3. Louissaint, A., Jr *et al.* Pediatric-type nodal follicular lymphoma: a biologically distinct lymphoma with frequent MAPK pathway mutations. *Blood* **128**, 1093-1100 (2016).
4. Kaymaz, Y. *et al.* Comprehensive Transcriptome and Mutational Profiling of Endemic Burkitt Lymphoma Reveals EBV Type-Specific Differences. *Mol. Cancer. Res.* **15**, 563-576 (2017).
5. Abate, F. *et al.* Distinct Viral and Mutational Spectrum of Endemic Burkitt Lymphoma. *PLoS Pathog.* **11**, e1005158 (2015).