SUPPLEMENTARY APPENDIX

Genomic alterations in high-risk chronic lymphocytic leukemia frequently affect cell cycle key regulators and NOTCH1-regulated transcription

Jennifer Edelmann,[1,2] Karlheinz Holzmann,[3] Eugen Tausch,[1] Emily A. Saunderson,[2] Billy M. C. Jebaraj,[1] Daniela Steinbrecher,[1] Anna Dolnik,[1] Tamara J. Blätte,[1] Dan A. Landau,[4,5] Jenny Saub,[1] Sven Estenfelder,[1] Stefan Ibach,[6] Florence Cymbalista,[7] Veronique Leblond,[8] Alain Delmer,[9] Jasmin Bahlo,[10] Sandra Robrecht,[10] Kirsten Fischer,[10] Valentin Goede,[10] Lars Bullinger,[1,11] Catherine J. Wu,[4] Daniel Mertens,[1], Gabriella Ficz,[2] John G. Gribben,[2] Michael Hallek,[10] Hartmut Döhner[1] and Stephan Stilgenbauer[1]

[1]Department of Internal Medicine III, Ulm University, Ulm, Germany; [2]Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, UK; [3]Genomics Core Facility, Ulm University, Ulm, Germany; [4]Department of Medical Oncology, Dana Farber Cancer Institute, Boston, MA, USA; [5]New York Genome Center, New York, NY, USA; [6]Wissenschaftlicher Service Pharma GmbH (WiSP), Langenfeld, Germany; [7]Service d'Hématologie Biologique, Hôpital Avicenne, Bobigny, France; [8]Service d'Hématologie, Hôpital Pitié-Salpêtrière, Paris, France; [9]Service d'Hématologie Clinique, CHU de Reims, Reims, France; [10]Department of Internal Medicine I, University of Cologne, Cologne, Germany and [11]Department of Hematology, Oncology and Tumor Immunology, Charité University Medicine Berlin, Campus Virchow Klinikum, Berlin, Germany

 doi:10.3324/haematol.2019.217307

Received: February 5, 2019.

Accepted: August 23, 2019.

Pre-published: August 29, 2019.

Correspondence: *JENNIFER EDELMANN* - jennifer.edelmann@uniklinik-ulm.de

Supplemental Information

Genomic alterations in high-risk chronic lymphocytic leukemia frequently affect cell cycle key regulators and NOTCH1-regulated transcription.

- Genomic alterations in high-risk CLL -

Jennifer Edelmann[1,2], Karlheinz Holzmann[3], Eugen Tausch[1], Emily A. Saunderson[2], Billy M. C. Jebaraj[1], Daniela Steinbrecher[1], Anna Dolnik[1], Tamara J. Blätte[1], Dan A. Landau[4,5], Jenny Saub[1] , Sven Estenfelder[1], Stefan Ibach[6], Florence Cymbalista[7], Veronique Leblond[8], Alain Delmer[9], Jasmin Bahlo[10], Sandra Robrecht[10], Kirsten Fischer[10], Valentin Goede[10], Lars Bullinger[1,11], Catherine J. Wu[4], Daniel Mertens[1], Gabriella Ficz[2], John G. Gribben[2], Michael Hallek[10], Hartmut Döhner[1], Stephan Stilgenbauer[1]

[1] Department of Internal Medicine III, Ulm University, Ulm, Germany

[2] Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, UK

[3] Genomics Core Facility, Ulm University, Ulm, Germany

[4] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, USA

[5] New York Genome Center, New York, NY 10013, USA

[6] Wissenschaftlicher Service Pharma GmbH (WiSP), Langenfeld, Germany

[7] Service d'Hématologie Biologique, Hôpital Avicenne, Bobigny, France

[8] Service d'Hématologie, Hôpital Pitié-Salpêtrière, Paris, France

[9] Service d'Hématologie Clinique, CHU de Reims, Reims, France

[10] Department of Internal Medicine I, University of Cologne, Cologne, Germany

[11] Department of Hematology, Oncology and Tumor Immunology, Charité-University Medicine Berlin, Campus Virchow Klinikum, Berlin, Germany

Corresponding author:

Jennifer Edelmann, MD, PhD; University Hospital of Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany

Phone: +49-731-500-45849

e-mail: jennifer.edelmann@uniklinik-ulm.de

Supplemental Methods

Single-nucleotide polymorphism array analysis

The present study included peripheral blood samples from 450 individuals enrolled on the CLL2O, CLL8 or CLL11 trial of the German CLL Study Group (GCLLSG). Genomic DNA was extracted [Allprep® DNA/RNA midi Kit, Quiagen®, Hilden, Germany] from frozen peripheral blood mononuclear cells (PBMCs) of CD19-positive and negative fractions and analysed on Affymetrix® 6.0 SNP arrays as outlined in the manufacturer's protocol [Affymetrix®, Santa Clara, CA, USA]. SNP genotype calls were generated by applying the birdseed algorithm in Genotyping Console version 4.0 [Affymetrix®] using at least 50 arrays in each analysis. DNA copy number and loss of heterozygosity (LOH) analyses were performed using reference alignment,[1] dChipSNP[2] and circular binary segmentation[3] as previously described.[4]

Segmentation was performed against intra-individual reference DNA in cases having a pure CD19-negative PBMC fraction [N=218] or against a pool of 10 gender-matched reference samples in cases lacking matched normal material [N=232]. Resulting segments with a window of 5 consecutive markers and mean log2-ratios of ≥0.2 and ≤-0.2 were visually inspected using dChipSNP to exclude inherited copy number variants and false calls due to noise or interbatch effects. The threshold for CNA detection was 10-15% affected cells for monoallelic lesions and the smallest CNA detected had a size of 4.276 kb. Lesions occurring in a tumor subclone with a clone size <25% were revised using the aroma.affymetrix software package[5] for an exact determination of segment boundaries.

Analysis for copy-neutral loss of heterozygosity (CN-LOH) was performed using dChipSNP.[2] According to prior experience, CN-LOH was considered to be a true lesion when containing homozygous SNPs in at least 20 consecutive markers with less than 10% intervening or conflicting calls.[6]

Pathway enrichment analysis using PathVisio

For pathway enrichment analysis with PathVisio, a permutation test was performed to calculate permuted p-values. Data was permuted 1000 times and a rank was

calculated of the actual z-score compared to permuted z- score. Pathways with a permuted p-value <0.05 were defined as significant.

Amplicon-based targeted next-generation and whole exome sequencing

Illumina® Design Studio [Illumina®,San Diego, CA, USA] was used to create custom amplicons with a size of 250 bp covering exons 2-11 of *TP53*, exon 34 of *NOTCH1*, exons 13-16 of *SF3B1*, and all coding regions of *MGA*, *SPEN*, *RBPJ*, *SNW1*, *CDKN2A*, and *MYC*. Library preparation was performed by using TruSeq® Custom Amplicon Assay Kit v1.5 [Illumina®] including extension and ligation steps between custom probes and adding of indices. Samples were pooled and loaded on a MiSeq® flowcell [Illumina®] in 48 sample batches and sequenced with MiSeq® Reagent Kit 500v2 [Illumina®] for paired end runs. Adapters and four 5' and 3' bp were trimmed using cutadapt[7] and reads were then aligned to the human reference genome GRCh37 using BWA-MEM.[8] Alignments were sorted and indexed using Picard[9] and locally realigned with GATK.[10] Resulting alignments were used for coverage calculation with BEDTools[11] as well as for pileup generation and variant calling with SAMtools[12] and VarScan2,[13] respectively. Variants were annotated by ANNOVAR and filtered to retain only non-synonymous mutations affecting exonic and splicing sites outside of polymer repeats that were either uncommon in the general population or associated with cancer, as based on dbSNP[14] and COSMIC annotations.[15]

Libraries for whole exome sequencing were constructed on an Agilent® platform and sequenced on an Illumina® HiSeq2000 or HiSeq2500 using 76 bp paired-end reads. Different software packages for bioinformatics analyses including demultiplexing, alignment to hg19 reference genome, variant calling and annotation were used as described elsewhere.[16]

At least ten nucleotides flanking each coding exon were included into the mutation analysis. The variant allele fraction (VAF) was required to be ≥0.07 and only nucleotide variants changing the amino acid sequence outside reported normal variants were scored as mutation.

The mean coverage taking all analyzed cases into account is provided below for each target gene. The provided range reflects the mean number of reads in cases with the lowest and the highest coverage.

TP53:	1131 reads mean	range: 182 – 2940 reads
NOTCH1:	1307 reads mean	range: 740 – 3334 reads
SF3B1:	1076 reads mean	range: 82 – 3072 reads
MGA:	951 reads mean	range: 440 – 1872 reads
SPEN:	757 reads mean	range: 385 – 1422 reads
RBPJ:	685 reads mean	range: 330 – 1438 reads
SNW1:	949 reads mean	range: 418 – 1874 reads
CDKN2A:	314 reads mean	range: 119 – 612 reads
MYC:	386 reads mean	range: 187 – 796 reads

Quantitative gene expression analysis

Complementary DNA (cDNA) was generated from RNA using High-Capacity RNA-to-cDNA Kit® [Thermo Fisher Scientific®, Waltham, MA, USA] and 100 ng cDNA were subsequently used in 20 µl qRT-PCR reactions with TaqMan® Gene Expression Assays [Applied Biosystems®, Foster City, CA, USA]. Gene expression levels of each target gene were calculated relative to an endogenous control gene and fold changes were calculated towards the median ΔCt-value of respective reference samples. *18S* or *ACTB* were used as endogenous control in 1:10 and/or 1:100 dilutions of the cDNA sample used for target gene analysis. Reactions were performed in triplicates on a QuantStudio™ 7 Flex System [Applied Biosystems®] using the standard thermal cycler protocol.

The following TaqMan® Gene Expression Assays were used:

HES1:	Hs00172878_m1
DTX1:	Hs01092201_m1
MGA:	Hs00861170_m1
SNW1:	Hs 00273351_m1
RBPJ:	Hs 00794653_m1
c-MYC:	Hs00153408_m1
CDKN2A, all transcript variants:	Hs 00923894_m1

CDKN2A, $p14^{ARF}$: Hs99999189_m1

CDKN2B, p15INK4B: Hs00793225_m1

Primers for *CCAT1* were custom designed according to Xiang et al., 2014 (ref. 17):

CCAT1, long + short isoform: Forward: ACTGCAGTAGGAAGGGAGCA

CCAT1, long + short isoform: Reverse: GTAAGCACTGGCCTTTCTGC

CCAT1, long isoform: Forward: CCACGTGCACATATTTGAATTG

CCAT1, long isoform: Reverse: TGCATTCCCTGCTTAATACTCA

Targeted DNA methylation analysis

After genomic DNA extraction, 250 ng of DNA was bisulfite converted following the manufacturer's instructions (EZ DNA Methylation-Lightning Kit, Zymo Research®, Irvine, CA, USA). Promoter regions of *CDKN2A* (both p14 and p16 alternative transcripts) and *CDKN2B* were amplified by PCR using bisulfite-specific primers:

p14_F: GTTGTTTATTTTTGGTGTTAAAGGG (GRCh38/hg38 chr9: 21994375-21994399);
p14_R: GTTGTTTATTTTTGGTGTTAAAGGG (GRCh38/hg38 chr9: 21994129-21994155);
p16_F: CTCTGGTGACCAGCCAGCCCCTCCT (GRCh38/hg38 chr9: 21974899-21974923);
p16_R: GCAAGGAAGGAGGACTGGGCTC (GRCh38/hg38 chr9: 21975149-21975170).
p15_F: GGCTGGCTCCCCACTCTGCCAGAG (GRCh38/hg38 chr9: 22009294-22009317);
p15_R: AGGGTAATGAAGCTGAGCCCAGG (GRCh38/hg38 chr9: 22009068-22009090).

Purified bisulfite-treated DNA was added to a mastermix (1x reaction buffer, 2 mM $MgCl_2$, 0.025 units/µl HotStarTaq (Qiagen®), 200 µM of each deoxynucleoside triphosphate (dNTP), 0.20 µM forward and reverse primers) and amplified using a thermocycler following the manufacturer's instructions. PCR products were purified and sent directly for Sanger Sequencing using the bisulfite-specific forward or reverse primer. This method can detect methylation at a single CpG from the cell population in the amplified region of ≥15% methylation (data not shown).

References Supplemental Methods

1 Pounds S, Cheng C, Mullighan C, Raimondi SC, Shurtleff S, Downing JR. Reference alignment of SNP microarray signals for copy number analysis of tumors. *Bioinformatics* 2009; 25: 315-321.

2 Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* 2004; 20: 1233-1240.

3 Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 2004; 5: 557-572.

4 Edelmann J, Holzmann K, Miller F, et al. High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. *Blood* 2012; 120: 4783-4794.

5 Bengtsson H, Irizarry R, Carvalho B, Speed TP. Estimation and assessment of raw copy numbers at the single locus level. *Bioinformatics* 2008; 24: 759-767.

6 Radtke I, Mullighan CG, Ishii M, et al. Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia. *Proc Natl Acad Sci U S A* 2009; 106: 12944-12949.

7 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMB:netjournal* 2011; 17.1.

8 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv*:1303. 2013:3997.

9 Picard Tools: Website. Accessed: 2015-08-11. URL: http://broadinstitute.github.io/picard/.

10 DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; 43: 491-498.

11 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010; 26:841-842.

12 Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; 25: 2078-2079.

13 Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012; 22: 568-576.

14 Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001; 29: 308-311.

15 Forbes SA, Beare D, Boutselakis H, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* 2017; 45: D777-D783.

16 Landau DA, Tausch E, Taylor-Weiner AN, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature* 2015; 526: 525-530.

17 Xiang JF, Yin QF, Chen T, Zhang Y, Zhang XO, Wu Z, et al. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res* 2014; 24: 513-531.

Supplementary Tables

See separate Excel-file for:

Supplementary Table S1

Patient and sample characteristics of high-risk CLL cases studied by SNP-array based DNA copy number and loss of heterozygosity analysis.

Supplementary Table S2

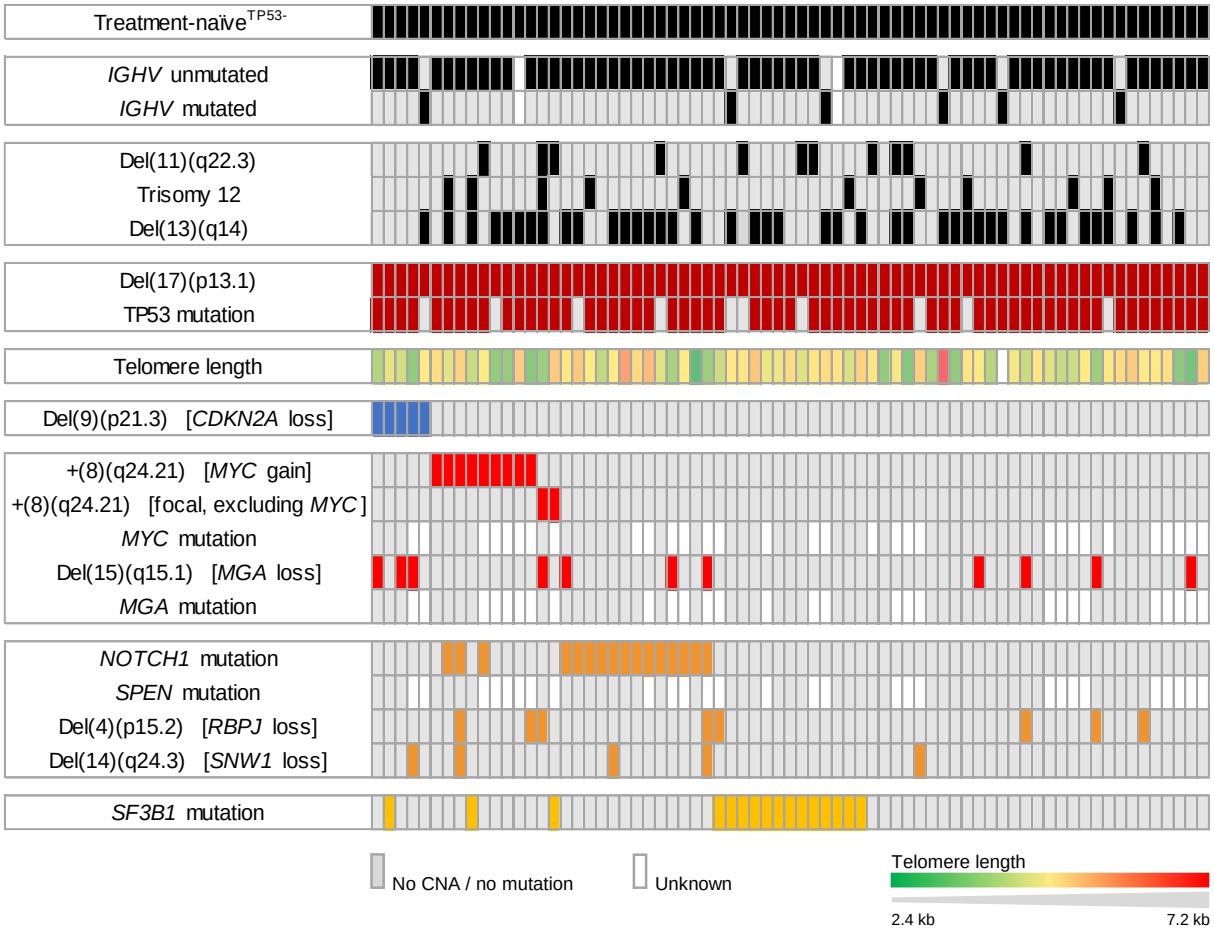
Observed copy number alterations (CNAs) in 146 high-risk CLL cases. CNAs are listed for each case by providing start and stop positions referring to the UCSC Genome browser, assembly March 2006, NCBI/hg18 (http://www.genome.ucsc.edu/) and providing the mean log2-ratio over each segment.

Supplementary Table S3

Sample characteristics of cases used for *CDKN2A*/*B* gene expression analysis. Gene expression levels of each target gene were calculated relative to ACTB expression levels and fold changes (FC) were calculated towards the median ΔCt-value of all reference samples [ΔΔCt-method].

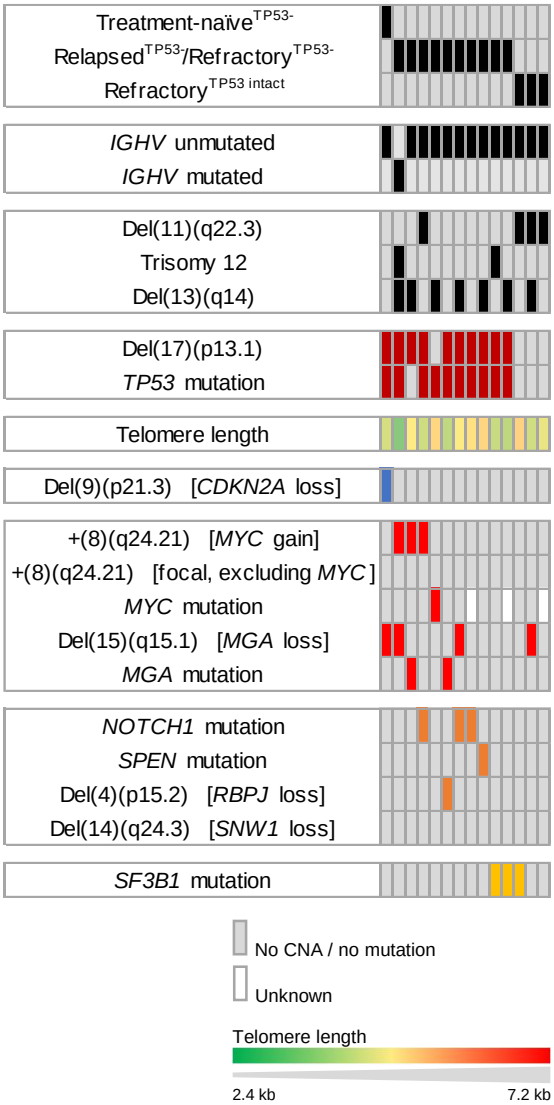
Supplementary Table S4

Sample characteristics of cases used for *HES1*, *DTX1* and *c-MYC* gene expression analysis. Gene expression levels of each target gene were calculated relative to *18S* expression levels fold changes (FC) were calculated towards the ΔCt-value of all reference samples [ΔΔCt-method].



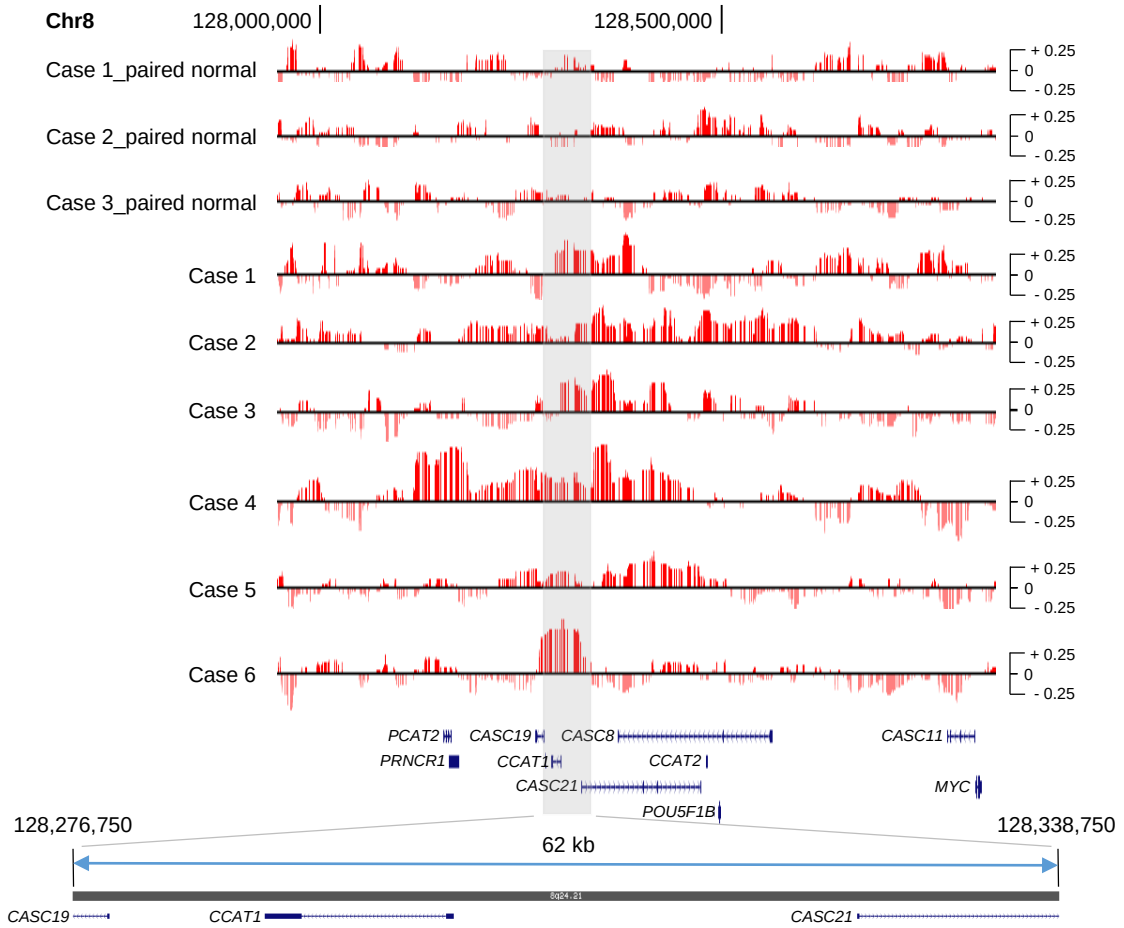
Supplementary Figure S1

Distribution of genetic alterations across treatment-naïve^{TP53-} cases.



Supplementary Figure S2

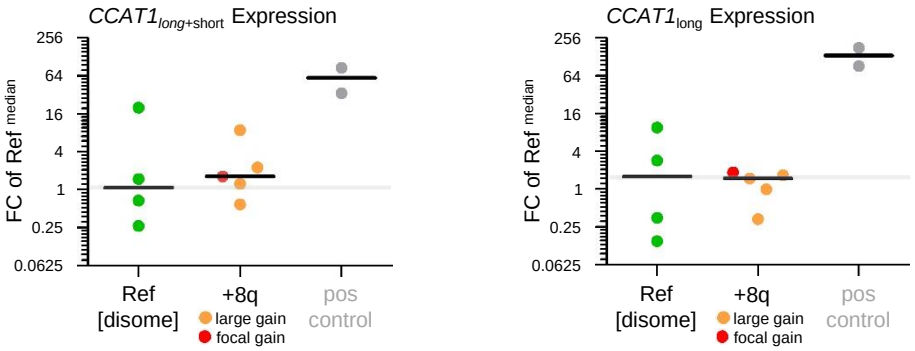
Distribution of genetic alterations across relapsed^{TP53-}/refractory cases developing Richter transformation during follow-up in the CLL2O trial.



Supplementary Figure S3

Minimally gained region in 8q24.21.

Raw log2-ratio, chromosome 8, cases with small focal gains in 8q24.21 displayed with the UCSC genome browser [hg18]. Red bars represent determined log-2 ratios of single probe sets sorted by their physical position along the chromosome.
The minimally gained region harbors three long non-coding RNA genes, namely *CASC19*, *CCAT1* and *CASC21*.

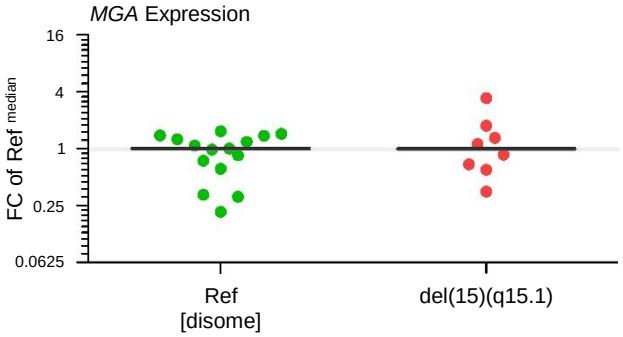


Supplementary Figure S4

Expression levels of both isoforms (left) and the long isoform of *CCAT1* in CLL with and without gain 8q.

Differences in the expression levels of *CCAT1* were not observed in a comparison of cases with 8q-gain and 8q-disome cases.
CCAT1$_{long+short}$ and *CCAT1*$_{long}$ gene expression levels were calculated relative to *ACTB* expression levels and fold changes (FC) were calculated towards the median ΔCt-value of all reference cases (Ref). Median expression levels within each sample group are highlighted. Cases with large 8q-gains affecting *CCAT1* and *c-MYC* gene loci are indicated in orange. One case with a focal gain in 8q24.21 affecting the *CCAT1* locus but not the *c-MYC* locus is indicated in red. 8q-gains were required to have a log2-ratio >0.75.
The two colorectal carcinoma cell lines HCT116 and HT29 with proven overexpression of *CCAT1* were used as positive control (pos control).[(1)]

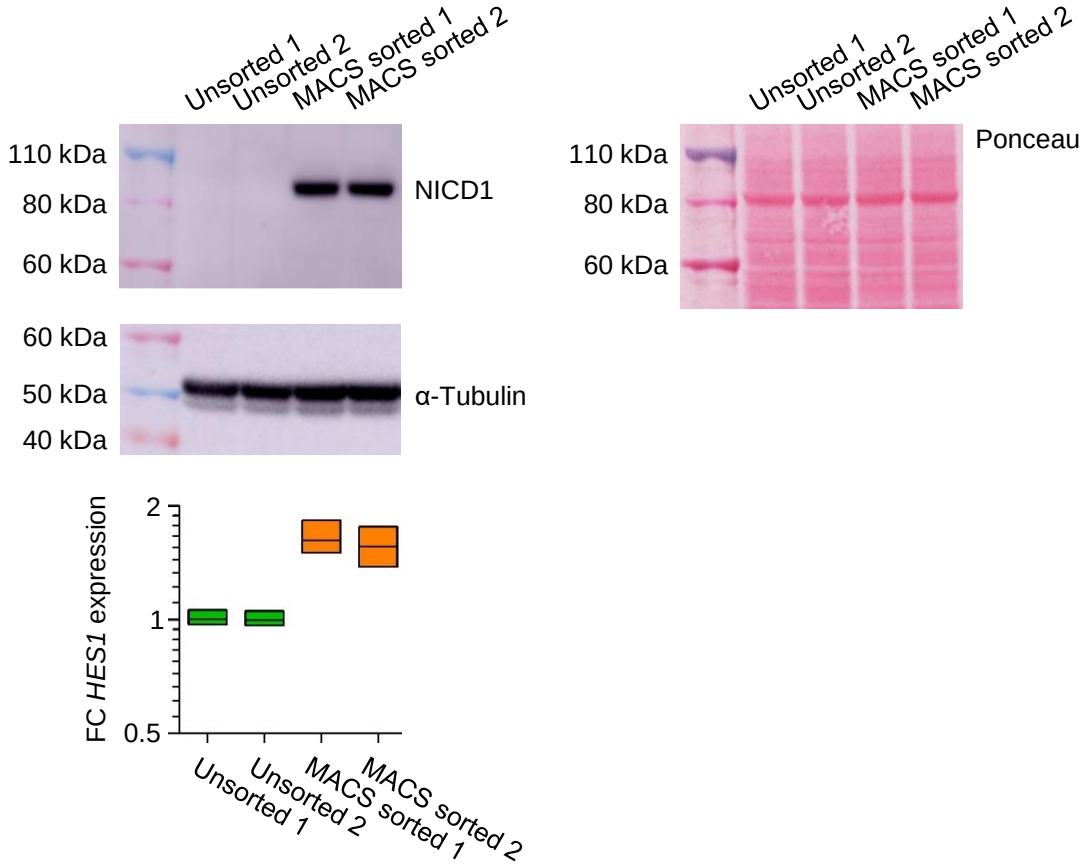
[(1)] Xiang JF, Yin QF, Chen T, et al. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. Cell Res. 2014;24(5):513-531.



Supplementary Figure S5

Expression levels of *MGA* in CLL with and without del(15)(q15.1).

No decreased *MGA* gene expression levels were observed in cases with monoallelic 15q-deletion. *MGA* gene expression levels were calculated relative to *ACTB* expression levels and fold changes (FC) were calculated towards the median ΔCt-value of all reference cases (Ref). Median expression levels within each sample group were highlighted. Only 15q-deleted cases with a log2-ratio <0.92 were included.



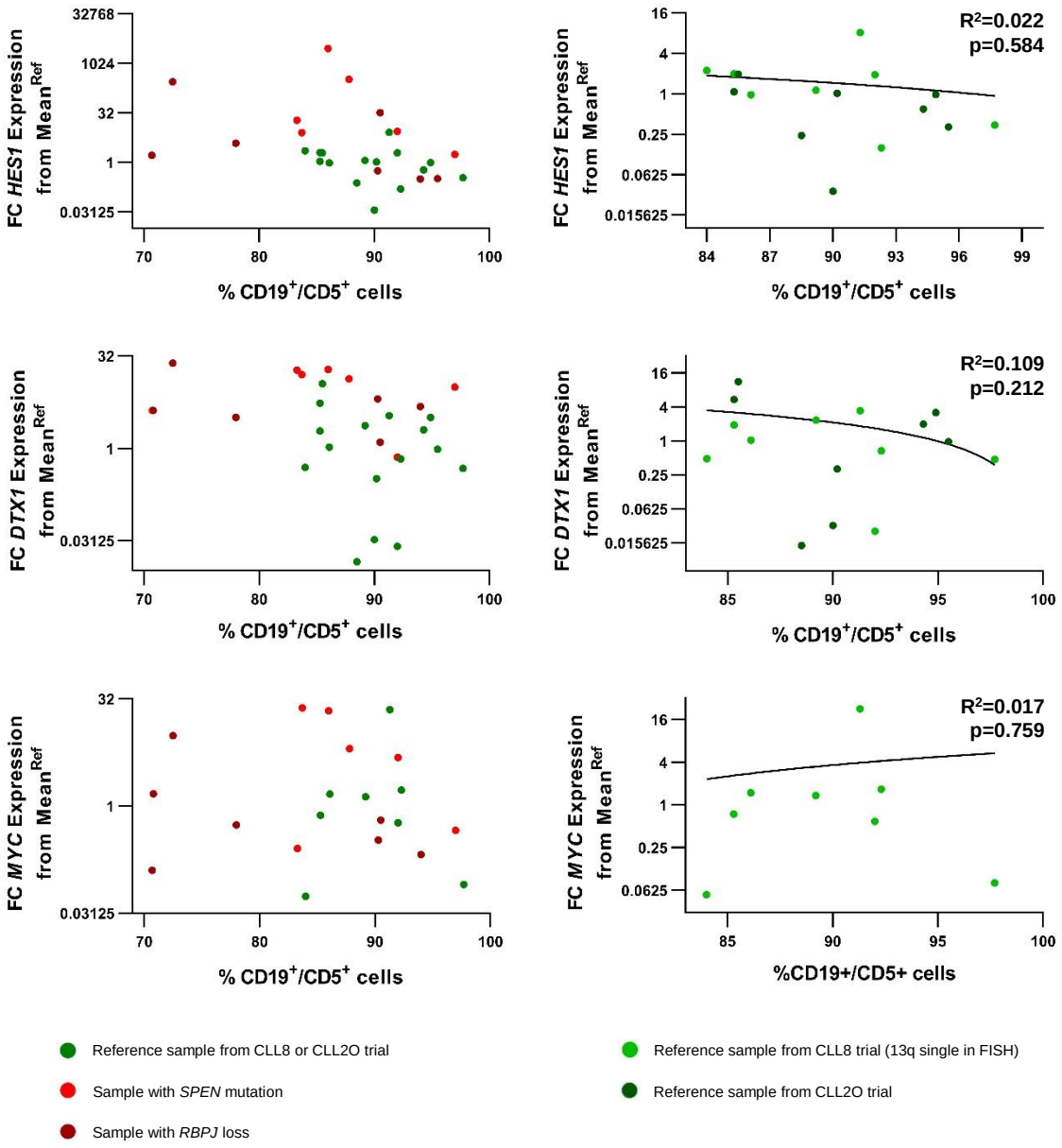
Supplementary Figure S6

EDTA-containing cell sorting buffer activates NOTCH1 signaling during cell sorting.

Sorting buffer containing 2 mM EDTA and 5% fetal bovine serum (FBS) was prepared according to the manufacturer's protocol [Miltenyi Biotec®, Bergisch Gladbach, Germany]. SU-DHL4 cells were incubated with anti-CD20 immunomagnetic beads [MACS, Miltenyi Biotec®] , for 15 minutes on ice at a concentration recommended by the manufacturer. Cells were then washed in sorting buffer, resuspended in sorting buffer and applied onto a MACS separation column. Three washing steps with sorting buffer were performed before cells were eluted in sorting buffer. Washing and centrifugation steps were performed at room temperature using ice-cold sorting buffer. Sorted cells and untreated control cells were pelleted and frozen at -80°C. The entire processing time for cell sorting was 50 to 60 minutes.

Whole protein lysates were obtained by using Qproteome® Mammalian Protein Prep Kit (Quiagen®). Western blot for the activated NOTCH1 intracellular domain (NICD1) was performed using 100 ng of total protein. Ponceau staining and α-tubulin were used as loading control. The NICD1 was only detectable in the two biological replicates obtained after cell sorting and was not detectable in the two biological replicates of unsorted control samples.

RNA was isolated with the Rneasy Mini Kit (Quiagen®) from two biological replicates of MACS sorted and unsorted control cells, respectively. Reverse transcription of RNA into cDNA was conducted with the High-Capacity RNA-to-cDNA Kit (Applied Biosystems®, Foster City, Ca, USA) and three technical replicates of cDNA were created. Gene expression of the NOTCH1 target gene *HES1* was analyzed by quantitative reverse transcription PCR (TaqMan Gene Expression Assay®, Applied Biosystems®). Gene expression levels of *HES1* were calculated relative to *ACTB* gene expression levels and fold changes (FC) were calculated relative to the mean ΔCt-value of all reference samples. Floating bars indicate mean FCs as well as highest and lowest FCs based on the three technical replicates for cDNA synthesis. *HES1* gene expression levels were higher in cells after they had undergone cell sorting. FCs towards expression levels in unsorted control cells ranged between 1.4 and 1.8 suggesting an activation of NOTCH1 regulated gene transcription during cell sorting.



Supplementary Figure S7

Expression levels of *HES1*, *DTX1* and *MYC* in relation to the percentage of CLL cells per sample.

Gene expression levels of *HES1*, *DTX1* and *MYC* did not correlate with the number of non-B-cells per sample.
HES1, *DTX1* and *MYC* gene expression levels were calculated relative to *ACTB* expression levels and fold changes (FC) were calculated towards the mean ΔCt-value of all reference cases (Ref). Fold changes were plotted against the percentage of $CD19^+/CD5^+$ CLL cells per sample and linear regression including all reference cases was carried out to rule out a systematic bias caused by the presence of non-B-cells in samples selected.