**Mutational Landscape and Tumor Burden Assessed by Cell-Free DNA in Diffuse Large B-Cell Lymphoma**

**Supplemental Materials**

**Supplementary Methods**

1. Sample collection and processing
2. cfDNA extraction and quantification
3. Tumor genomic DNA/RNA extraction
4. Cell of origin determination
5. Library design for hybrid selection
6. Next generation sequencing
7. Bioinformatic analyses
8. Somatic *versus* germ line variants classification

**Sample collection and processing**

In all the cases, plasma samples for cell-free DNA (cfDNA) extraction were collected at diagnosis, before the start of treatment, using PAXgene Blood ccfDNA tubes (PreAnalytiX, Switzerland). In addition, in 45 selected cases genomic DNA from the paired diagnostic formalin-fixed paraffin-embedded (FFPE) tissue biopsy was obtained. cfDNA was extracted from 2-4 mL of plasma using the QIAamp circulating nucleic acid kit (Qiagen, Germany). Genomic DNA and RNA were isolated from FFPE diagnostic tissue biopsies using the AllPrep DNA/RNA FFPE Kit (Qiagen, Germany) according to the manufacturer’s instructions.

**cfDNA extraction and quantification**

Blood samples were centrifuged for 10 min at 1500g and subsequently for 1 minute at 20000 g. Plasma was aliquoted into 1mL in microtubes and stored at −80°C until extraction. cfDNA was extracted from 2-4 mL of plasma using the QIAamp circulating nucleic acid kit (Qiagen, Germany). The quantity and quality of the samples were assessed using Qubit High Sensitivity dsDNA (Thermo Fisher Scientific) and TapeStation (Agilent).

Levels of cfDNA are reported as haploid genome equivalents per mL of plasma (hGE/mL), determined as the product of total cell-free DNA concentration and the mean allele fraction of somatic alterations, and expressed as a base-10 logarithm (log hGE/mL). We selected the 2.5 log hGE/mL threshold for pretreatment cfDNA level according to previous publication1.

**Tumor genomic DNA/RNA extraction**

Genomic DNA and RNA were isolated from FFPE diagnostic tissue biopsies. Five 10 μm-thick sections per sample were used to extract RNA and DNA using the AllPrep DNA/RNA FFPE Kit (Qiagen, Germany) according to the manufacturer’s instructions.

**Cell of origin determination**

The molecular cell of origin of the tumors was established using 200ng of RNA to determine gene expression levels by means of the Lymph2Cx assay (Nanostring technologies, Seattle, WA). Samples were hybridized to the custom codesets at 65°C overnight (15.5 to 22.5 hours). nCounter digital analyzer was used for fully automated imaging and data collection. Samples were classified as GCB, ABC and Unclassified subtypes using the algorithm previously described2.

**Library design for hybrid selection**

A targeted sequencing gene panel, including 112 genes (target region: 388Kb) that are recurrently mutated in DLBCL and other mature B-cell tumors, was specifically designed for this project (Supplementary Table S1). The design of the panel was carried out using capture tools as follows: 1) genes mutated in >5% of mature B-cell tumors, 2) genes associated with resistance to chemotherapy in mature B-cell tumors.

**Next generation sequencing**

Mutation profiles were generated using a custom hybridization capture-based panel strategy (compatible with cfDNA and DNA from FFPE samples) and subsequent sequenced in a MiSeq instrument (Illumina). Libraries were performed using 15-30 ng of cfDNA and 150 ng of gDNA from FFPE samples, following the procedure indicated by the manufacturer recommendations. Targeted sequencing was performed using molecular-barcoded library adapters using the ThruPLEX Tag-seq kit (Takara) and a hybridization capture based method (SureSelectXT-Agilent Technologies). The quality of the libraries was determined using the Bioanalyzer high sensitivity DNA kit (Agilent) and quantified by PCR using the KAPA library quantification kit (KAPA Biosystems). Finally, the libraries were pooled and sequenced 2x130 bp in the MiSeq instrument.

**Bioinformatic analyses**

Next-generation sequencing data for variant calling was analyzed using an updated version of our in-house pipeline3. Briefly, raw reads were trimmed using the SurecallTrimmer (v4.0.1, AGeNT, Agilent). Alignment of the trimmed reads was performed using BWA-mem algorithm (v0.7.17), PCR or optical duplicates were marked using MarkDuplicates from Picard (RRID: SCR\_006525), and the base quality score recalibration was performed using GATK’s BaseRecalibrator and ApplyBQSR functions (RRID: SCR\_001876 v4.0). Variant calling was performed in parallel using VarScan2, Mutect2, VarDict, outLyzer, and freebayes. Only variants that were identified as “PASS” by at least 3 of the algorithms were considered. Finally, variants were annotated using snpEff/snpSift (v4.3t). Mutations were visually inspected on Integrative Genomics Viewer (IGV).

**Somatic *versus* germ line variants classification**

Variants reported in 1000 Genome Project, ExAC and/or gnomAD with a population frequency >1% were considered polymorphisms and automatically removed from the analysis. To further filter out non-recurrent polymorphisms, variants were only considered somatic if 1) they were not reported as germ line in our custom ICGC data base of 506 WGS/WES4; and were 2) truncating, or 3) predicted as potentially damaging by at least one of the following algorithms: CADD (phred score > 10), PolyPhen2 (score > 0.9), SIFT (score < 0.1) (RRID: SCR\_012813), and/or MutationAssessor (score > 2) (RRID: SCR\_005762). Next, we sequenced the non-tumoral DNA of 14 cases to assess the accuracy of our somatic predictor method. In these 14 cases, our 3-step algorithm had a sensitivity of 93% and a specificity of 70%.

**REFERENCES**

1 Kurtz DM, Scherer F, Jin MC, Soo J, Craig AFM, Esfahani MS et al. Circulating Tumor DNA Measurements As Early Outcome Predictors in Diffuse Large B-Cell Lymphoma. J Clin Oncol. 2018; 36: 2845–2853.

2 Scott DW, Wright GW, Williams PM, Lih C-J, Walsh W, Jaffe ES et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. Blood. 2014; 123: 1214–1217.

3 Nadeu F, Delgado J, Royo C, Baumann T, Stankovic T, Pinyol M et al. Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. Blood. 2016; 127: 2122–2130.

4 Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature. 2015; 526: 519–524.

**Figure S1. Consort diagram** **of patient flow through study**

DLBCL diagnosis

September 2016- March 2019

(n= 100)

Excluded (n= 21)

* Treatment started before sample extraction (n=13)
* Concomitant diagnosis of 2nd neoplasm (n= 2)
* Suboptimal sample (n= 6)

Library construction &

Sequencing (n=79)

No mutations observed/failed library construction (n= 10)

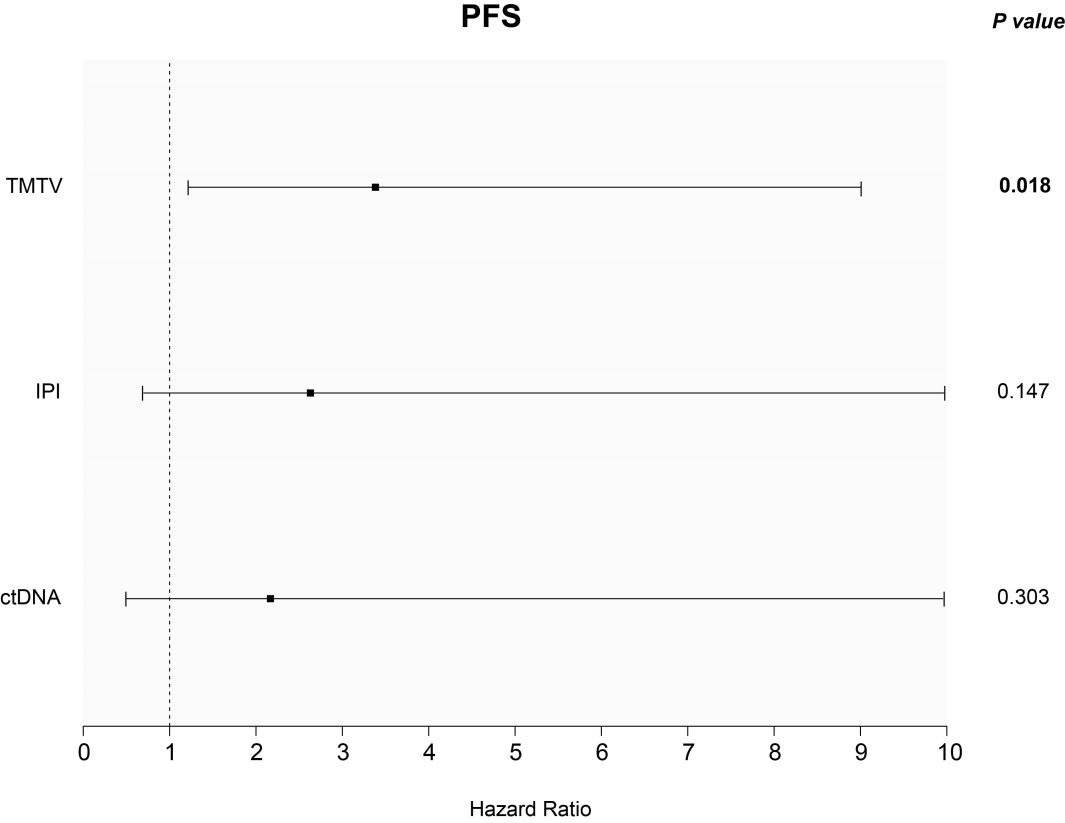
At least one mutation

detected (n= 69)

**Figure S2.** **Mutational profile in cfDNA of the 79 patients with DLBCL according to the genetic subtypes.**

****

**Figure S3. Results of the multivariate analyze for PFS**

****