Cell free circulating tumor DNA in cerebrospinal fluid detects and monitors central nervous system involvement of B-cell lymphomas

Sabela Bobillo,[1*] Marta Crespo,[1*] Laura Escudero,[2*] Regina Mayor,[2] Priyanka Raheja,[1] Cecilia Carpio,[1] Carlota Rubio-Perez,[2] Bárbara Tazón-Vega,[1] Carlos Palacio,[1] Júlia Carabia,[1] Isabel Jiménez,[1] Juan. C. Nieto,[1] Julia Montoro,[1] Francisco Martínez-Ricarte,[3] Josep Castellví,[4] Marc Simó,[5] Lluis Puigdefàbregas,[1] Pau Abrisqueta,[1] Francesc Bosch[1#] and Joan Seoane[2,6#]

[1]Laboratory of Experimental Hematology, Department of Hematology, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron University Hospital (HUVH), Universitat Autònoma de Barcelona, Department of Medicine; [2]Translational Research Program, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron University Hospital (HUVH), Universitat Autònoma de Barcelona; [3]Department of Neurosurgery, Universitat Autònoma de Barcelona, Hospital Vall d'Hebron; [4]Department of Pathology, Vall d'Hebron University Hospital, Universitat Autònoma de Barcelona; [5]Department of Nuclear Medicine, Vall d'Hebron University Hospital, Universitat Autònoma de Barcelona and [6]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain and CIBERONC

**SB, MC and LE contributed equally as co-first authors.*

#FB and JS contributed equally as co-senior authors.

 doi:10.3324/haematol.2019.241208

Received: October 22, 2019.

Accepted: February 18, 2020.

Pre-published: February 20, 2020.

Correspondence: *JOAN SEOANE* - jseoane@vhio.net

FRANCESC BOSCH - fbosch@vhio.net

SUPPLEMENTARY MATERIAL

PATIENTS ANS METHODS:

Patients

Nineteen patients diagnosed with the following conditions were included: restricted CNS lymphomas n=6: PCNSL, n=1; SCNSL, n=5 [diffuse large B cell lymphoma (DLBCL) n=4, mantle cell lymphoma n=1, Waldeström macroglobulinemia n=1]; systemic lymphoma with concomitant CNS involvement, n=1 (mantle cell lymphoma, n=1); and systemic lymphoma without CNS disease, n=12 [DLBCL, n=7; high-grade B cell lymphoma, n=3; Burkitt lymphoma, n=2].

DNA sequencing and mutation genomic analysis

SureSelect Human All Exon V5 (Agilent Technologies) was used to perform whole exome enrichment for the Illumina paired-read sequencing platform. Genomic DNA (200ng) was sheared on a Covaris™ E210 and purified/size selected with AMPure XP beads (Agencourt, Beckman Coulter). The sheared DNA was end-repaired, 3´ adenylated and ligated to Agilent sequencing adaptors. The adaptor-modified DNA was amplified in pre-capture 10 PCR cycles using Herculase II fusion DNA polymerase (Agilent Technologies). The PCR product was quality controlled on the Agilent 2100 Bioanalyzer 7500 chip (Agilent Technologies) to confirm size range (200 to 350bp) and quantity and hybridized for 24h on 65ºC with the SureSelect Capture Library. The hybridization mix was washed and the eluate was post-capture PCR amplified (12 cycles) in order to add the index tags using SureSelectXT Indexes for Illumina. The final library size and the concentration was determined on Agilent 2100 Bioanalyzer 7500 chip. The libraries were sequenced on HiSeq2500 (Illumina) in paired-end mode with a read length of 2x100bp using TruSeq SBS Kit v4 (Illumina). Each sample was sequenced in a fraction of a sequencing v4 flow cell lane, following the manufacturer's protocol. Image analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA 1.18.66.3) and followed by generation of FASTQ sequence files.

The tumor mutations of the 300 gene panel were QC checked by FastQC [1] and then mapped to the human reference genome (Hg19) by bwa (v. 0.7.12) [2] with default

settings. Alignment files (BAM format) were processed using Picard (v.1.110) [3] to add read groups and remove duplicates. The resulting BAM files were processed using SAMtools (v. 1.2) [4] and the Genome Analysis ToolKit (GATK) (v. 3.2.0) [5] Variants were called by VarScan (v2.4.3) [6] with the following parameters: minimum variant allele frequency (VAF) of 5%, a minimum coverage of 8 reads and a minimum of 7 reads that confirm the mutation with a p-value below 0.05. Annotation of the vcf files was performed with ANNOVAR [7]. Next, variants were filtered for false positive, germline and non-protein-affecting mutations through an in-house algorithm.

WES reads were mapped to Hg19 using the GEM3 toolkit [8]. Alignment files (BAM format), containing only properly paired, uniquely mapping, reads; were processed using Picard (v.1.110) [3] to add read groups and remove duplicates. The resulting BAM files were processed using SAMtools (v. 1.2)[4] and the GATK (v. 3.2.0) [5]. Somatic tumor variants were called by Mutect2 and Strelka [9, 10] doing a matched normal analysis. Only those mutations reported by both callers were considered for further analyses. We manually revised the alignment of certain mutations of interest and rescued two mutations (in *ID3* and *MYC*), which were only reported by Strelka and two additional ones, in *TCF3*, which were located in an exon-flank region.

The biological relevance of all mutations was annotated through Cancer Genome Interpreter (CGI; v1907) [11], which was ran considering "Lymphoma" as input cancer type. Genomic coordinates were used as input for WES mutations while HGVS protein was used as input for the mutations coming from the gene mutation panel. We considered as driver mutations all mutations predicted by CGI as drivers (either known or predicted driver tier 1 and tier 2), see Tamborero et al 2018 [11] for details on the method.

For each mutation the Variant Allele Frequency (VAF) was adjusted for purity and ploidy. Purity was estimated by Vall d'Hebron Hospital Lymphoma specialist pathologists. The ploidy of each gene in the 300 gene panel was estimated from the calculation of copy number alterations (CNA), through the tool CNVkit [6] with the PSCBS segmentation method [9, 10]. On the other hand, the ploidy of the WES was estimated from the calculation of CNAs through Control-FREEC [12]. Next, p-values were adjusted for multiple-testing correction (FDR Benjamini-Hochberg), CNAs with q-values < 0.05 were considered significant.

The adjusted VAF, named Cancer Allelic Fraction (CAF), was defined as:

$$CAF = VAF \times \frac{ploidy \times purity\,(1 - purity)}{purity \times ploidy}$$

TABLES

Table S1. Mutations identified across 16 lymphoma patients

Table S1A. List of the mutations, in protein coordinates, identified through the 300 gene panel across 11 patients.

Table S1B. List of the mutations, in genomic and protein coordinates, identified through WES across 5 patients.

Table S1C. Biological relevance annotation of the mutations across 16 lymphoma patients

References:

1. "FastQC." n.d. FastQC: Quality Control Tool for High Throughput Sequence Data. . In. 2019.
2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25: 1754-1760.
3. "Picard Tools - By Broad Institute.". In. 2019.
4. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011; 27: 2987-2993.
5. McKenna A, Hanna M, Banks E et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20: 1297-1303.
6. Reble E, Castellani CA, Melka MG et al. VarScan2 analysis of de novo variants in monozygotic twins discordant for schizophrenia. Psychiatr Genet 2017; 27: 62-70.
7. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010; 38: e164.
8. Marco-Sola S, Ribeca P. Efficient Alignment of Illumina-Like High-Throughput Sequencing Reads with the GEnomic Multi-tool (GEM) Mapper. Curr Protoc Bioinformatics 2015; 50: 11.13.11-20.
9. Cibulskis K, Lawrence MS, Carter SL et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013; 31: 213-219.
10. Saunders CT, Wong WS, Swamy S et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 2012; 28: 1811-1817.
11. Tamborero D, Rubio-Perez C, Deu-Pons J et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med 2018; 10: 25.
12. Boeva V, Popova T, Bleakley K et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 2012; 28: 423-425.