# Clinical impact of small subclones harboring *NOTCH1*, *SF3B1* or *BIRC3* mutations in chronic lymphocytic leukemia

*Silvia Rasi,[1*] Hossein Khiabanian,[2*§] Carmela Ciardullo,[1] Lodovico Terzi-di-Bergamo,[1] Sara Monti,[1] Valeria Spina,[1] Alessio Bruscaggin,[1] Michaela Cerri,[1] Clara Deambrogi,[1] Lavinia Martuscelli,[1] Alessandra Biasi,[1] Elisa Spaccarotella,[1] Lorenzo De Paoli,[1] Valter Gattei,[3] Robin Foà,[4] Raul Rabadan,[2] Gianluca Gaidano,[1§] and Davide Rossi[1§]*

*[1]Division of Hematology, Department of Translational Medicine, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy; [2]Center for Topology of Cancer Evolution and Heterogeneity, Department of Biomedical Informatics and Center for Computational Biology and Bioinformatics, Columbia University, New York, NY, USA; [3]Clinical and Experimental Onco-Hematology, Centro di Riferimento Oncologico, Aviano, Italy; and [4]Hematology, Department of Cellular Biotechnologies and Hematology, Sapienza University, Rome, Italy*

*Correspondence: rossidav@med.unipmn.it*
*doi:10.3324/haematol.2015.136051*

**SUPPLEMENTARY APPENDIX CONTENTS**

# CLINICAL IMPACT OF SMALL SUBCLONES HARBORING *NOTCH1, SF3B1* OR *BIRC3* MUTATIONS IN CHRONIC LYMPHOCYTIC LEUKEMIA

Silvia Rasi,[1*] Hossein Khiabanian,[2*] Carmela Ciardullo,[1] Lodovico Terzi-di-Bergamo,[1] Sara Monti,[1] Valeria Spina,[1] Alessio Bruscaggin,[1] Michaela Cerri,[1] Clara Deambrogi,[1] Lavinia Martuscelli,[1] Alessandra Biasi,[1] Lorenzo De Paoli,[1] Valter Gattei,[3] Robin Foà,[4] Raul Rabadan,[2] Gianluca Gaidano,[1§] Davide Rossi[1§]

*S.R. and H.K. equally contributed
[§]G.G. and D.R. equally contributed

[1]Division of Hematology, Department of Translational Medicine, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy; [2]Center for Topology of Cancer Evolution and Heterogeneity, Department of Biomedical Informatics and Center for Computational Biology and Bioinformatics, Columbia University, New York, USA; [3]Clinical and Experimental Onco-Hematology, Centro di Riferimento Oncologico, Aviano, Italy; [4]Hematology, Department of Cellular Biotechnologies and Hematology, Sapienza University, Rome, Italy.

# SUPPLEMENTARY METHODS

**Ultra-deep-NGS of *NOTCH1, SF3B1* and *BIRC3***

Ultra-deep-NGS of the *NOTCH1* (genomic region hg19 chr9:139390616-139390809; RefSeq NM_017617.2), *SF3B1* (genomic regions hg19 chr2:198267401-198267564, chr2:198267222-198267424, chr2:198266704-198266869; RefSeq NM_012433.2), and *BIRC3* (genomic regions hg19 chr11:102201667-102201850, chr11:102201812-102201997, chr11:102207617-102207837; RefSeq NM_001165.3) mutation hotspots [1] was performed using the 454 chemistry and was based on amplicon libraries. Sequences for *NOTCH1, SF3B1*, and *BIRC3* were retrieved from the UCSC Human Genome database using the corresponding mRNA accession number as a reference. Sequence specific PCR primers, located ~10 bp upstream or downstream to target exon boundaries were designed in the Primer 3 program (http://frodo.wi.mit.edu/primer3/) and filtered using UCSC in silico PCR to exclude pairs yielding more than a single product. In total, *NOTCH1, SF3B1, BIRC3* regions of interest were covered by 7 sequence specific primer pairs, each flanked by MID-labeled Primer A and Primer B sequences to barcode the samples. In each ultra-deep-NGS experiment, a total of 63 amplicons, corresponding to *NOTCH1, SF3B1, BIRC3* regions of interest of 9 distinct patients, were amplified from genomic DNA by using a high fidelity Taq polymerase (FastStart High fidelity PCR System, Roche Diagnostics). PCR products were then individually purified using Agencourt AMPure XP beads (Beckman Coulter) and quantified using the Quant-iT PicoGreen dsDNA kit (Invitrogen). Corresponding patient-specific amplicon pools were generated by combining each of the amplicons in an equimolar ratio for each patient sample. The pools were diluted to a concentration of $1\times10^6$ molecules per µl and processed using the GS Junior Series Lib-A method (Roche Diagnostics). Forward (A beads) and reverse (B beads) reactions were carried out using 5,000,000 beads per emulsion oil tube. The copy per bead ratio used was 1.1:1. The amplification reaction, breaking of the emulsions and enrichment of beads carrying amplified DNA were performed using the workflow as recommended by the manufacturer. Finally, the obtained amplicon library was loaded on a PicoTiterPlate (PTP) and subjected to ultra-deep-NGS on the Genome Sequencer Junior instrument (454 Life Sciences).

**Bioinformatic approach to call subclonal *NOTCH1, SF3B1,* and *BIRC3* variants out of the background error noise**

The following bioinformatic workflow was established to call subclonal *NOTCH1, SF3B1*, and *BIRC3* variants from ultra-deep-NGS experiments in patient samples.[2] First, for each patient, the sequencing reads were mapped to the reference genome hg19 using the Burrows-Wheeler Aligner (BWA) long-read alignment tool, and sites that differed from the reference *NOTCH1*, *SF3B1*, or *BIRC3* sequence were identified. Variants overlapping the primers or neighboring them by one nucleotide, and variants mapping within homopolymeric tracts (for *SF3B1* and *BIRC3*), that specifically suffer from errors in 454 NGS, were excluded from the analysis. To fit for the parameters of the negative binomial distribution, we used the remaining variants with a mean quality phred score >27 and a frequency <0.05. After correcting for multiple hypotheses according to the Bonferroni test, we obtained error distributions and depth thresholds for the discovery of true missense substitutions, deletions, and insertions. Next, we obtained error distributions and depth thresholds (without a multiple hypothesis correction) for the identification of each discovered significant variant, after excluding them from those used for fitting of the parameters. Finally, a list of variants that passed the discovery and identification depth thresholds was created.

**Allele-specific PCR**

Separate AS-PCR was designed for each subclonal *NOTCH1, SF3B1*, and *BIRC3* variant, and the assay was tailored at a sensitivity of $10^{-3}$ (i.e detection of one mutant allele/1000 wild type alleles). As AS-PCR positive controls, pGEM-*NOTCH1,* pGEM-*SF3B1*, and pGEM-*BIRC3* plasmids containing each one of the subclonal *NOTCH1, SF3B1*, and *BIRC3* variants were generated by site directed mutagenesis.[2] Briefly, genomic DNA was isolated from healthy donor PBMC and *NOTCH1* (genomic region hg19 chr9:139390616-139390809), *SF3B1* (genomic regions hg19 chr2:198267401-198267564, chr2:198267222-198267424, chr2:198266704-198266869), and *BIRC3* (genomic regions hg19 chr11:102201667-102201850, chr11:102201812-102201997, chr11:102207617-102207837) were amplified using a high fidelity Taq polymerase (Pfu DNA polymerase, Promega). *NOTCH1, SF3B1,* and *BIRC3* amplicons were subjected to Sanger sequencing to rule out clonal mutations. *NOTCH1, SF3B1,* and *BIRC3* amplicons were then cloned into the pGEM vector. Primers containing the index *NOTCH1, SF3B1*, and *BIRC3* variant were used to introduce the variant into the pGEM-*NOTCH1*, pGEM-*SF3B1*, and pGEM-*BIRC3* plasmids by high fidelity circular PCR amplification (Pfu DNA polymerase, Promega). pGEM-*NOTCH1,* pGEM-*SF3B1*, and pGEM-*BIRC3* plasmids were then transformed into JM109 competent E. coli cells for high-yield plasmid production. The introduction of the index *NOTCH1, SF3B1*, and *BIRC3* mutation into the pGEM-*NOTCH1,* pGEM-*SF3B1*, and pGEM-*BIRC3* plasmids was confirmed by Sanger sequencing. Wild type *NOTCH1* (genomic region hg19 chr9:139390616-139390809), *SF3B1* (genomic regions hg19 chr2:198267401-198267564, chr2:198267222-198267424, chr2:198266704-198266869), and *BIRC3* (genomic regions hg19 chr11:102201667-102201850, chr11:102201812-102201997, chr11:102207617-102207837) cloned into pGEM plasmids served as negative controls for the AS-PCR. The analytical sensitivity of our

AS-PCR assays was estimated to be ≥0.1% by subjecting to AS-PCR serial dilutions (1:2; 1:4; 1:8; 1:16; 1:32; 1:64; 1:128; 1:256; 1:512; 1:1024; 1:2048) of mutated plasmid DNA into wild type plasmid DNA.

**Mutation analysis by Sanger sequencing**

The mutation hotspots [1] of *NOTCH1* (genomic region hg19 chr9:139390616-139390809; RefSeq NM_017617.2), *SF3B1* (genomic regions hg19 chr2:198267401-198267564, chr2:198267222-198267424, chr2:198266704-198266869; RefSeq NM_012433.2) ,and *BIRC3* (genomic regions hg19 chr11:102201667-102201850, chr11:102201812-102201997, chr11:102207617-102207837; RefSeq NM_001165.3) genes were analyzed by PCR amplification and Sanger sequencing of high molecular weight genomic DNA. All PCR primers and conditions are available upon request. Purified amplicons were subjected to conventional DNA Sanger sequencing using the ABI PRISM 3100 Genetic Analyzer (Applied Biosystems), and compared to the corresponding germline sequences using the Mutation Surveyor Version 4.0.5 software package (SoftGenetics) after automated and/or manual curation. Of the evaluated sequences, 99% had a Phred score of 20 or more and 97% had a score of 30 or more. Candidate variants were confirmed from both strands on independent PCR products. The following databases were used to exclude known germline variants: Human dbSNP Database at NCBI (Build 136) (http://www.ncbi.nlm.nih.gov/snp); Ensembl Database (http://www.ensembl.org/index.html); The 1000 Genomes Project (http://www.1000genomes.org/); five single-genome projects available at the UCSC Genome Bioinformatics resource (http://genome.ucsc.edu/). Synonymous variants, previously reported germline polymorphisms and changes present in the matched normal DNA were removed from the analysis.

***TP53* mutation analysis**

Mutation analysis of the *TP53* gene was performed by ultra-deep-NGS as previously described.[2]

***IGHV-IGHD-IGHJ* rearrangement analysis**

PCR amplification of *IGHV-IGHD-IGHJ* rearrangements was performed on HMW genomic DNA using IGHV leader primers or consensus primers for the IGHV FR1 along with appropriate IGHJ genes, as previously described.[3] PCR products were directly sequenced with the BigDye Terminator v1.1 Ready Reaction Cycle Sequencing kit (Applied Biosystems) using the ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). Sequences were analyzed using the IMGT databases and the IMGT/V-QUEST tool (version 3.2.17, Université Montpellier 2, CNRS, LIGM, Montpellier, France).

**Fluorescence in situ hybridization (FISH)**

Probes used for FISH analysis were: *i)* LSID13S319 (13q14 deletion), CEP12 (trisomy 12), LSIp53 (17p13/*TP53* deletion), and LSIATM (11q2-q23/*ATM* deletion) (Abbott, Rome, Italy); and *ii)* the RP11-177O8 (*BIRC3*) BAC clone. The labeled BAC probe was tested against normal control metaphases to verify the specificity of the hybridization. For each probe, at least 400 interphase cells with well-delineated fluorescent spots were examined. Nuclei were counterstained with 4',6'-diamidino-2-phenylindole (DAPI) and antifade reagent, and signals were visualized using an Olympus BX51 microscope (Olympus Italia, Milan, Italy). The presence of 13q14 deletion, trisomy 12, 11q22-q23 deletion, 17p13 deletion, and *BIRC3* deletion was scored when the percentage of nuclei with the abnormality was above our internal cut off (5%, 5%, 7%, 10%, and 10% respectively), defined as the mean plus 3 standard deviations of the frequency of normal control cells exhibiting the abnormality.

**Proportional hazard regression**

The proportional hazard assumption was assessed by plotting the smoothed Schoenfeld residuals against time.[4] Possible interactions were tested as reported.[5] The bias corrected c-index and calibration slope of the Cox model were calculated through the .632 bootstrap method (1000 resamplings).[5-8] The heuristic shrinkage estimator was calculated using the formula (model likelihood ratio $\chi^2$) - (number of degree of freedom in the model)/(model likelihood ratio $\chi^2$). This approach provides an estimate of prediction accuracy of the Cox model to protect against overfitting.[5]

**Statistical power**

The statistical power of the sample size was calculated by using the Cox.NIS of the TrialSize (v1.3) package of R.

## SUPPLEMENTARY FIGURE LEGENDS

**Supplementary Figure 1. Experimental validation of subclonal mutations identified by ultra-deep next generation sequencing.** Representation of the variant frequency of exemplificative subclonal *NOTCH1* **(A)**, *SF3B1* **(B),** and *BIRC3* **(C)** mutations of very low allelic abundance. The first bar of the graphs shows the variant allele frequency in the discovery ultra-deep-NGS experiment. The second and the third bars show the variant allele frequency in independent ultra-deep-NGS validation experiments. The number of mutated reads out of the total number of reads covering the variant position is reported. Conventional agarose-gel electrophoresis of the AS-PCR products validating the *NOTCH1* **(D)**, *SF3B1* **(E)**, and *BIRC3* **(F)** mutations of very low allelic abundance. After AS-PCR for the mutant allele, a mutation-specific band is amplified from the patient samples and from the mutated plasmid DNA (positive control). No bands are amplified from the wild type plasmid DNA and the wild type genomic DNA from a healthy donor (negative controls), thus confirming the specificity of the assay. Due to their low clonal abundance, the subclonal *NOTCH1* **(G)**, *SF3B1* **(H)**, and *BIRC3* **(I)** mutations are not detectable by conventional Sanger sequencing. Asterisks point to the positions of the subclonal variants.

**Supplementary Figure 2. Clonal representation of *NOTCH1*, *SF3B1*, and *BIRC3* mutations. (A)** Box-plot representing the variant allele frequency of *NOTCH1*, *SF3B1*, and *BIRC3* mutations detected by ultra-deep-NGS. The bottom and top of the box are the 25th and 75th percentiles. The band inside the box is the median. The ends of the whiskers are the minimum and maximum of the data. p value by Mann-Whitney after Bonferroni correction for multiple comparisons. **(B)** Prevalence of *NOTCH1*, *SF3B1*, and *BIRC3* mutations detected by Sanger sequencing (in red) and missed by Sanger sequencing (in grey) because of low clonal abundance. p value by Chi-square after Bonferroni correction for multiple comparisons.

**Supplementary Figure 3. Maximally selected rank statistics for the identification of a best cut-off level in the *NOTCH1*, *SF3B1* and *BIRC3* mutation clonal abundance to predict CLL overall survival (OS).** Maximally selected rank statistics for the identification of a best cut-off level in the *NOTCH1* **(A)**, *SF3B1* **(B)**, and *BIRC3* **(C)** mutation clonal abundance to predict CLL OS. The black line indicates the absolute standardized log-rank statistics for each level of mutation abundance. The red line indicates the statistical threshold. Comparison of OS from CLL diagnosis between patients harboring mutations represented below the variant allele frequency (VAF) threshold (yellow line), cases harboring mutations above the VAF threshold (red line), and cases harboring wild type sequence (blue line) of the *NOTCH1* **(A)**, *SF3B1* **(B)**, and *BIRC3* **(C)** genes. p, p values by log-rank test.

**Supplementary Figure 4. Maximally selected rank statistics for the identification of a best cut-off level in the *TP53* mutation clonal abundance to predict CLL overall survival.** The black line indicates the absolute standardized log-rank statistics for each level of *TP53* mutation abundance. The red line indicates the statistical threshold.

**Supplementary Figure 5. Longitudinal analysis of clonal evolution in CLL patients harboring small mutated subclones.** Graphical illustration of the kinetics of the mutated populations in CLL patients who have been longitudinally investigated by ultra-deep-NGS. The x axis represents time and the y-axis represents allele frequency. *NOTCH1*, *SF3B1*, and *BIRC3* mutations are represented by color-coded circles. The size of the circles is proportional to the allele frequency of the lesion. Arrows indicate the time point at which tumor samples were collected. The relationship between sample collection and treatments is also indicated. W&W, watch and wait; CLB, chlorambucil; FC, fludarabine, cyclophosphamide; FCR, fludarabine, cyclophosphamide, rituximab; BR, bendamustine, rituximab; FCM, fludarabine, cyclophosphamide, mitoxantrone, RDHAOX, rituximab, dexamethasone, high dose cytarabine, cisplatin; BA, bendamustine, alemtuzumab; A, alemtuzumab; RCHOP, rituximab, cyclophosphamide, doxorubicine, vincristine, prednisone; RCVP, rituximab, cyclophosphamide, vincristine, prednisone; CR, complete response according to the IWCLL-NCI criteria; PR, partial response according to IWCLL-NCI criteria; Richter, Richter syndrome.

**Supplementary Table 1. Characteristics of the CLL series**

| **Characteristics** [a] | **N** | **%** |
|---|---|---|
| Age >70 years | 158 | 52 |
| Male | 163 | 53.6 |
| Binet A | 240 | 78.9 |
| Binet B | 37 | 12.2 |
| Binet C | 27 | 8.9 |
| *IGHV* identity ≥98% [b] | 106 | 34.9 |
| Stereotyped VH CDR3 [b] | 67 | 22 |
| 13q14 deletion | 156 | 51.3 |
| Trisomy 12 | 64 | 21.1 |
| 11q22-q23 deletion | 24 | 7.9 |
| 17p13 deletion | 29 | 9.5 |
| *TP53* mutations [c] | 46 | 15.1 |

[a] *IGHV*, immunoglobulin heavy variable gene; CDR3, complementarity determining region 3
[b] 5 patients lacked productive *IGHV-IGHD-IGHJ* rearrangements
[c] Detected by ultra-deep-NGS

**Supplementary Table 2. Ultra-deep-NGS primers**

| | | | |
|---|---|---|---|
| *NOTCH1* | chr9:139390616-139390809 | forward | 5'-CGTATCGCCTCCCTCGCGCCATCAG-MID-CACTATTCTGCCCCAGGAGA-3' |
| | | reverse | 5'-CTATGCGCCTTGCCAGCCCGCTCAG-MID-CAGTCGGAGACGTTGGAATG-3' |
| *SF3B1* | chr2:198267401-198267564 | forward | 5'-CGTATCGCCTCCCTCGCGCCATCAG-MID-CCAACTCATGACTGTCCTTTCTT-3' |
| | | reverse | 5'-CTATGCGCCTTGCCAGCCCGCTCAG-MID-TGCCAGGACTTCTTGCTTTT-3' |
| *SF3B1* | chr2:198267222-198267424 | forward | 5'-CGTATCGCCTCCCTCGCGCCATCAG-MID-CCCTGGGCATTCCTTCTTTA-3' |
| | | reverse | 5'-CTATGCGCCTTGCCAGCCCGCTCAG-MID-GAGTCCAGTCTGGGCAACAT-3' |
| *SF3B1* | chr2:198266704-198266869 | forward | 5'-CGTATCGCCTCCCTCGCGCCATCAG-MID-TTGGGGCATAGTTAAAACCTG-3' |
| | | reverse | 5'-CTATGCGCCTTGCCAGCCCGCTCAG-MID-AAATCAAAAGGTAATTGGTGGA-3' |
| *BIRC3* | chr11:102201667-102201850 | forward | 5'-CGTATCGCCTCCCTCGCGCCATCAG-MID-CCTTTCAATTTTTGATTTCATTTCA-3' |
| | | reverse | 5'-CTATGCGCCTTGCCAGCCCGCTCAG-MID-TTCTCTCCAGTTGCTAGGATTTTT-3' |
| *BIRC3* | chr11:102201812-102201997 | forward | 5'-CGTATCGCCTCCCTCGCGCCATCAG-MID-ATTAATGCTGCCGTGGAAAT-3' |
| | | reverse | 5'-CTATGCGCCTTGCCAGCCCGCTCAG-MID-AGACTGATATCAAATCCTTATGAAAAT-3' |
| *BIRC3* | chr11:102207617-102207837 | forward | 5'-CGTATCGCCTCCCTCGCGCCATCAG-MID-TGAAGAAGCAAACTGCCTTTTAT-3' |
| | | reverse | 5'-CTATGCGCCTTGCCAGCCCGCTCAG-MID-AAAGTTTAGACGATGTTTTGGTTC-3' |

**Supplementary Table 3. List of subclonal *NOTCH1* mutations identified by ultra-deep-NGS**

| ID | cDNA position | Amino acid position | Mutation type | Variant frequency | Variant depth | Position depth | Genomic position (hg19) | Ref | Var |
|---|---|---|---|---|---|---|---|---|---|
| 4938 | c.7451C>T | p.T2484M | missense | 1.413 | 40 | 2830 | 139390743 | G | A |
| 5076 | c.7544_7545delCT | p.P2515fs*4 | indel | 9.334 | 134 | 1441 | 139390649-139390650 | AG | - |
| 5182 | c.7544_7545delCT | p.P2515fs*4 | indel | 3.506 | 55 | 1569 | 139390649-139390650 | AG | - |
| 5842 | c.7544_7545delCT | p.P2515fs*4 | indel | 3.304 | 56 | 1710 | 139390649-139390650 | AG | - |
| 5842 | c.7510C>T | p.Q2504* | nonsense | 1.404 | 24 | 1710 | 139390684 | G | A |
| 10642 | c.7544_7545delCT | p.P2515fs*4 | indel | 1.415 | 23 | 1661 | 139390649-139390650 | AG | - |
| 10666 | c.7504C>T | p.Q2502* | nonsense | 5.182 | 174 | 3358 | 139390690 | G | A |
| 10666 | c.7544_7545delCT | p.P2515fs*4 | indel | 2.173 | 73 | 3358 | 139390649-139390650 | AG | - |
| 11178 | c.7544_7545delCT | p.P2515fs*4 | indel | 2.761 | 56 | 2046 | 139390649-139390650 | AG | - |
| 11772 | c.7544_7545delCT | p.P2515fs*4 | indel | 1.476 | 27 | 1829 | 139390649-139390650 | AG | - |
| 13691 | c.7544_7545delCT | p.P2515fs*4 | indel | 2.414 | 36 | 1491 | 139390649-139390650 | AG | - |
| 14379 | c.7544_7545delCT | p.P2515fs*4 | indel | 7.936 | 163 | 2054 | 139390649-139390650 | AG | - |
| 14633 | c.7544_7545delCT | p.P2515fs*4 | indel | 6.085 | 85 | 1405 | 139390649-139390650 | AG | - |
| 17843 | c.7544_7545delCT | p.P2515fs*4 | indel | 6.011 | 116 | 1930 | 139390649-139390650 | AG | - |

**Supplementary Table 4. List of subclonal *SF3B1* mutations identified by ultra-deep-NGS**

| ID | cDNA position | Amino acid position | Mutation type | Variant frequency | Variant depth | Position depth | Genomic position (hg19) | Ref | Var |
|---|---|---|---|---|---|---|---|---|---|
| 3715 | c.1996A>G | p.K666E | missense | 1.706 | 39 | 2286 | 198267361 | T | C |
| 3716 | c.2223G>C | p.K741N | missense | 1.919 | 27 | 1407 | 198266709 | C | G |
| 3981 | c.2110A>T | p.I704F | missense | 2.589 | 53 | 2047 | 198266822 | T | A |
| 3981 | c.1866G>T | p.E622D | missense | 5.502 | 91 | 1654 | 198267491 | C | A |
| 3981 | c.1874G>T | p.R625L | missense | 1.692 | 28 | 1655 | 198267483 | C | A |
| 3981 | c.1996A>G | p.K666E | missense | 3.725 | 78 | 2094 | 198267361 | T | C |
| 3981 | c.1986C>G | p.H662Q | missense | 2.292 | 48 | 2094 | 198267371 | G | C |
| 3981 | c.1866G>C | p.E622D | missense | 2.600 | 43 | 1654 | 198267491 | C | G |
| 4047 | c.2111T>C | p.I704T | missense | 2.221 | 73 | 3287 | 198266821 | A | G |
| 4294 | c.2110A>T | p.I704F | missense | 7.741 | 196 | 2532 | 198266822 | T | A |
| 5564 | c.1866G>T | p.E622D | missense | 3.301 | 76 | 2302 | 198267491 | C | A |
| 5965 | c.2098A>G | p.K700E | missense | 9.294 | 474 | 5100 | 198266834 | T | C |
| 5965 | c.2110A>T | p.I704F | missense | 4.630 | 236 | 5097 | 198266822 | T | A |
| 6480 | c.2098A>G | p.K700E | missense | 2.849 | 83 | 2913 | 198266834 | T | C |
| 6698 | c.2098A>G | p.K700E | missense | 0.535 | 24 | 4484 | 198266834 | T | C |
| 7222 | c.2098A>G | p.K700E | missense | 5.040 | 138 | 2738 | 198266834 | T | C |
| 7561 | c.1996A>C | p.K666Q | missense | 5.517 | 79 | 1432 | 198267361 | T | G |
| 8525 | c.1866G>T | p.E622D | missense | 1.288 | 21 | 1631 | 198267491 | C | A |
| 8662 | c.2098A>G | p.K700E | missense | 1.390 | 23 | 1655 | 198266834 | T | C |
| 9485 | c.2098A>G | p.K700E | missense | 1.770 | 32 | 1808 | 198266834 | T | C |
| 10761 | c.2111T>A | p.I704N | missense | 1.831 | 84 | 4588 | 198266821 | A | T |
| 12969 | c.1988C>T | p.T663I | missense | 15.415 | 195 | 1265 | 198267369 | G | A |
| 17842 | c.2098A>G | p.K700E | missense | 2.142 | 38 | 1774 | 198266834 | T | C |
| 17850 | c.2098A>G | p.K700E | missense | 17.002 | 406 | 2388 | 198266834 | T | C |
| 17862 | c.2098_2100del3 | p.K700delAGT | indel | 4.117 | 59 | 1441 | 198266830-198266832 | ACT | - |

**Supplementary Table 5. List of subclonal *BIRC3* mutations identified by ultra-deep-NGS**

| ID | cDNA position | Amino acid position | Mutation type | Variant frequency | Variant depth | Position depth | Genomic position (hg19) | Ref | Var |
|---|---|---|---|---|---|---|---|---|---|
| 3370 | c.1639delC | p.Q547fs*21 | indel | 1.657 | 20 | 1207 | 102207657 | C | - |
| 3703 | c.1639delC | p.Q547fs*21 | indel | 1.802 | 48 | 2664 | 102207657 | C | - |
| 3703 | c.1294insG | p.R432fs*6 | indel | 3.736 | 145 | 3881 | 102201942 | - | G |
| 3714 | c.1660delG | p.E554fs*14 | indel | 1.348 | 21 | 1558 | 102207678 | G | - |
| 3714 | c.1639delC | p.Q547fs*21 | indel | 1.412 | 22 | 1558 | 102207657 | C | - |
| 3722 | c.1168delT | p.F390fs*6 | indel | 0.244 | 20 | 8207 | 102201816 | T | - |
| 3975 | c.1281_1285del5 | p.I427fs*9 | indel | 1.412 | 51 | 3653 | 102201929-102201933 | AAGGG | - |
| 4048 | c.1281_1285del5 | p.I427fs*9 | indel | 2.154 | 71 | 3296 | 102201929-102201933 | AAGGG | - |
| 4174 | c.1192A>G | p.T398A | missense | 0.339 | 22 | 6494 | 102201840 | A | G |
| 4788 | c.1639delC | p.Q547fs*21 | indel | 0.817 | 22 | 2694 | 102207657 | C | - |
| 4788 | c.1286delA | p.E429fs*18 | indel | 0.817 | 25 | 3060 | 102201934 | A | - |
| 4788 | c.1641_1647del7 | p.Q547fs*19 | indel | 0.737 | 19 | 2694 | 102207659-102207665 | ATTGCGG | - |
| 4831 | c.1281_1285del5 | p.I427fs*9 | indel | 0.857 | 29 | 3454 | 102201929-102201933 | AAGGG | - |
| 4833 | c.1176delA | p.R392fs*4 | indel | 0.248 | 19 | 7660 | 102201824 | A | - |
| 5726 | c.1688_1690del3 | p.K563delAAG | indel | 2.620 | 42 | 1603 | 102207706-102207708 | AAG | - |
| 5825 | c.1802C>A | p.T601K | missense | 5.894 | 124 | 2104 | 102207820 | C | A |
| 5889 | c.1281delA | p.I427fs*20 | indel | 0.638 | 24 | 3763 | 102201929 | A | - |
| 5889 | c.1298_1299delAA | p.E433fs*4 | indel | 0.837 | 31 | 3764 | 102201946-102201947 | AA | - |
| 5956 | c.1639delC | p.Q547fs*21 | indel | 1.027 | 32 | 3116 | 102207657 | C | - |
| 6288 | c.1636insA | p.E546fs*13 | indel | 3.031 | 112 | 3695 | 102207654 | - | A |
| 8762 | c.1663_1664delAG | p.R555fs*3 | indel | 0.57 | 19 | 3332 | 102207681-102207682 | AG | - |
| 9630 | c.1281delA | p.I427fs*20 | indel | 0.63 | 20 | 3177 | 102201929 | A | - |
| 9630 | c.1639delC | p.Q547fs*21 | indel | 0.791 | 20 | 2529 | 102207657 | C | - |
| 9630 | c.1285delG | p.E429fs*18 | indel | 0.724 | 23 | 3177 | 102201933 | G | - |
| 10320 | c.1639delC | p.Q547fs*21 | indel | 0.963 | 24 | 2492 | 102207657 | C | - |
| 11815 | c.1639delC | p.Q547fs*21 | indel | 1.364 | 22 | 1613 | 102207657 | C | - |
| 14281 | c.1639delC | p.Q547fs*21 | indel | 2.582 | 79 | 3060 | 102207657 | C | - |
| 14281 | c.1690G>T | p.E564* | nonsense | 0.98 | 30 | 3060 | 102207708 | G | T |
| 14495 | c.1639delC | p.Q547fs*21 | indel | 0.842 | 22 | 2613 | 102207657 | C | - |
| 17847 | c.1199_1200delAG | p.Q400fs*11 | indel | 0.95 | 53 | 5581 | 102201847-102201848 | AG | - |

**Supplementary Table 6. Univariate and multivariate analysis of OS [a]**

| | Univariate analysis | | | | Multivariate analysis | | | |
|---|---|---|---|---|---|---|---|---|
| | **HR** | **LCI** | **UCI** | **p** | **HR** | **LCI** | **UCI** | **p** |
| **Age ≤70 years** | - | - | - | | - | - | - | |
| **Age >70 years** | 3.33 | 2.20 | 5.06 | <.001 | 3.63 | 2.30 | 5.72 | <.001 |
| **Binet A** | - | - | - | | - | - | - | |
| **Binet B-C** | 2.13 | 1.57 | 3.14 | <.001 | 1.66 | 1.04 | 2.63 | .031 |
| ***IGHV* homology <98%** | - | - | - | | - | - | - | |
| ***IGHV* homology ≥98%** | 1.74 | 1.20 | 2.52 | .003 | 1.20 | 0.77 | 1.88 | .411 |
| **No *TP53* disruption** | - | - | - | | - | - | - | |
| ***TP53* disruption [b]** | 2.91 | 1.93 | 4.37 | <.001 | 2.71 | 1.75 | 4.19 | <.001 |
| **No trisomy 12** | - | - | - | | - | - | - | |
| **Trisomy 12** | 1.61 | 1.08 | 2.41 | .019 | 1.45 | 0.90 | 2.33 | .123 |
| **No 11q22-q23 deletion** | - | - | - | | - | - | - | |
| **11q22-q23 deletion** | 2.30 | 1.29 | 4.07 | .004 | 1.56 | 0.58 | 4.18 | .377 |
| ***NOTCH1* wt** | - | - | - | | - | - | - | |
| **Small subconal *NOTCH1* mutation** | 0.63 | 0.19 | 2.06 | .450 | 0.61 | 0.15 | 2.49 | .493 |
| **Sanger sequencing positive *NOTCH1* mutation** | 2.14 | 1.33 | 3.45 | .002 | 2.07 | 1.16 | 3.68 | .013 |
| ***SF3B1* wt** | - | - | - | | - | - | - | |
| **Small subclonal *SF3B1* mutation** | 1.37 | 0.66 | 2.83 | .389 | 1.92 | 0.91 | 4.07 | .087 |
| **Sanger sequencing positive *SF3B1* mutation** | 2.69 | 1.44 | 5.04 | .002 | 2.14 | 1.09 | 4.21 | .026 |
| ***BIRC3* wt** | - | - | - | | - | - | - | |
| **Small subclonal *BIRC3* mutation** | 1.47 | 0.80 | 2.69 | .205 | 0.80 | 0.40 | 1.58 | 0.534 |
| **Sanger sequencing positive *BIRC3* mutations** | 2.76 | 1.20 | 6.30 | .016 | 1.21 | 0.41 | 3.59 | 0.724 |
| **No *BIRC3* deletion** | - | - | - | | - | - | - | |
| ***BIRC3* deletion** | 3.376 | 1.69 | 6.74 | <.001 | 1.71 | 0.52 | 5.58 | .374 |

[a] OS, overall survival; HR, hazard ratio; LCI, 95% lower confidence interval; UCI, 95% upper confidence interval; *IGHV*, immunoglobulin heavy variable gene; wt, wild type
[b] *TP53* mutations detected by ultra-deep-NGS and/or 17p13 deletion
Total number of patients included in the multivariate analysis: 299; Events: 116; 5 patients lacked productive *IGHV-IGHD-IGHJ* rearrangements

## SUPPLEMENTARY REFERENCES


1. Foà R, Del Giudice I, Guarini A, Rossi D, Gaidano G. Clinical implications of the molecular genetics of chronic lymphocytic leukemia. *Haematologica*. 2013;98(5):675-85.
2. Rossi D, Khiabanian H, Spina V, et al. Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood*. 2014;123(14):2139-47.
3. Rossi D, Rasi S, Spina V, et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. Blood 121:1403-1412, 2013
4. Schoenfeld D. Partial residuals for the proportional hazard regression model. Biometrika 69:239-241, 1982.
5. Harrell FE Jr, Lee K, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15:361-387, 1996.
6. Efron B, Tibshirani R. Improvements on cross-validation: the .632_bootstrap method. JASA 92:548-560, 1997
7. van Howelingen JC, le Cessie S. Predictive value of statistical models. Stat Med 8:1303-1325, 1990.

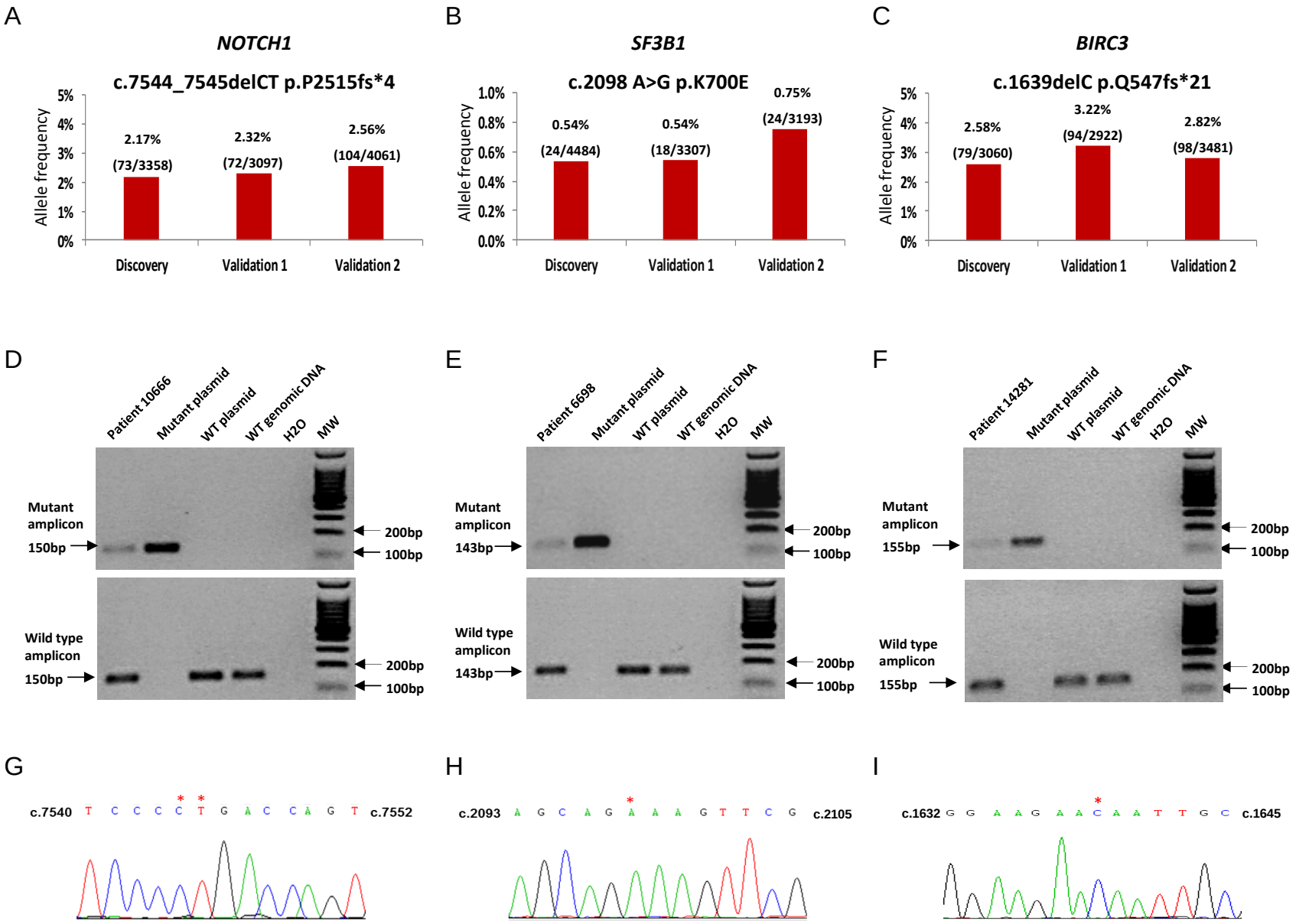

A
NOTCH1
c.7544_7545delCT p.P2515fs*4
Allele frequency
5%
4%
3%
2%
1%
0%
2.17%
(73/3358)
2.32%
(72/3097)
2.56%
(104/4061)
Discovery
Validation 1
Validation 2
B
SF3B1
c.2098 A>G p.K700E
Allele frequency
1.0%
0.8%
0.6%
0.4%
0.2%
0.0%
0.54%
(24/4484)
0.54%
(18/3307)
0.75%
(24/3193)
Discovery
Validation 1
Validation 2
C
BIRC3
c.1639delC p.Q547fs*21
Allele frequency
5%
4%
3%
2%
1%
0%
2.58%
(79/3060)
3.22%
(94/2922)
2.82%
(98/3481)
Discovery
Validation 1
Validation 2
D
Patient 10666
Mutant plasmid
WT plasmid
WT genomic DNA
H2O
MW
Mutant amplicon
150bp
200bp
100bp
Wild type amplicon
150bp
200bp
100bp
E
Patient 6698
Mutant plasmid
WT plasmid
WT genomic DNA
H2O
MW
Mutant amplicon
143bp
200bp
100bp
Wild type amplicon
143bp
200bp
100bp
F
Patient 14281
Mutant plasmid
WT plasmid
WT genomic DNA
H2O
MW
Mutant amplicon
155bp
200bp
100bp
Wild type amplicon
155bp
200bp
100bp
G
c.7540 T C C C C T G A C C A G T c.7552
H
c.2093 A G C A G A A A G T T C G c.2105
I
c.1632 G G A A G A A C A A T T G C c.1645

Figure 1S

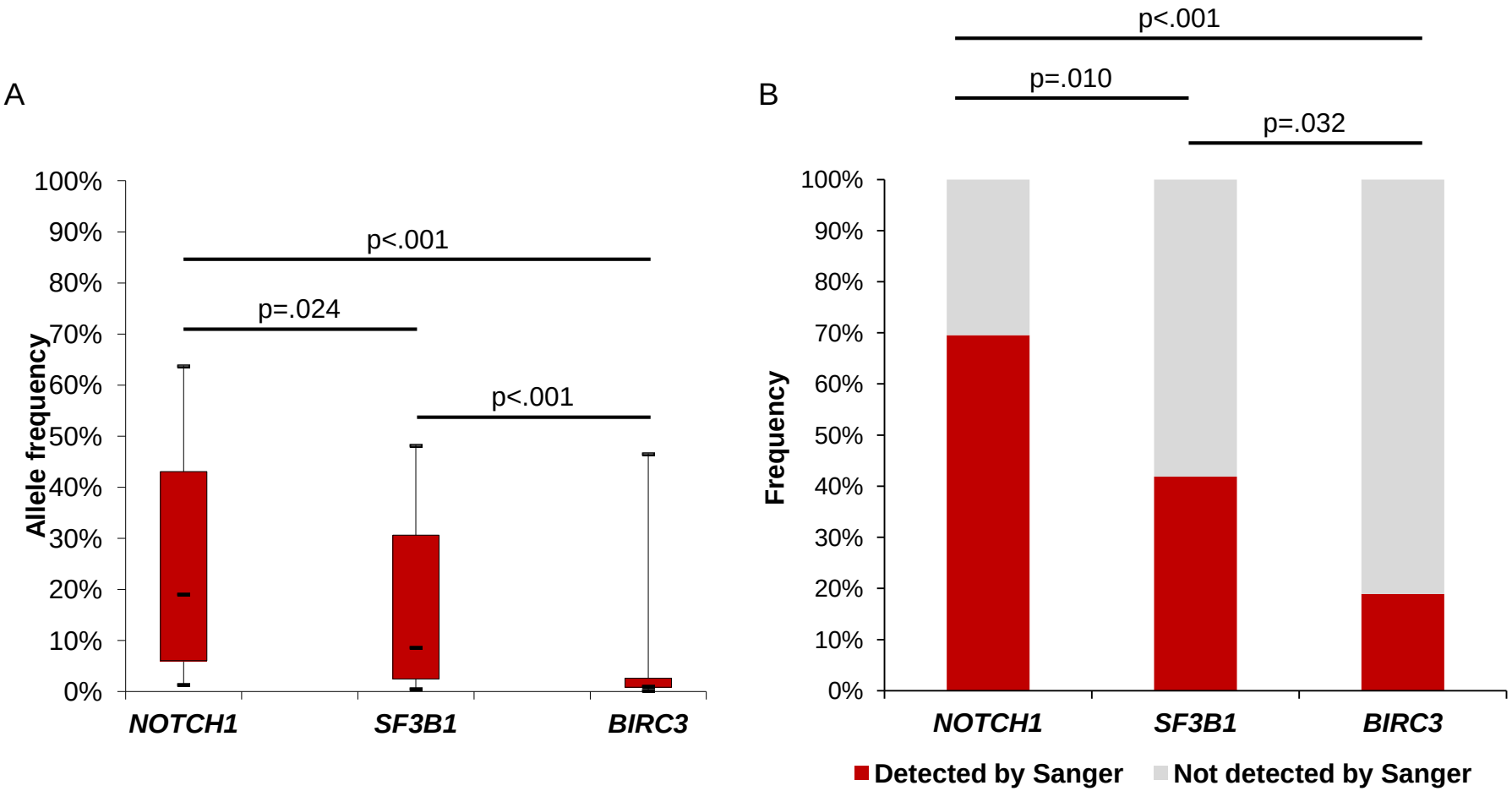

A
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Allele frequency
p<.001
p=.024
p<.001
NOTCH1
SF3B1
BIRC3
B
p<.001
p=.010
p=.032
Frequency
NOTCH1
SF3B1
BIRC3
Detected by Sanger
Not detected by Sanger

Figure 2S

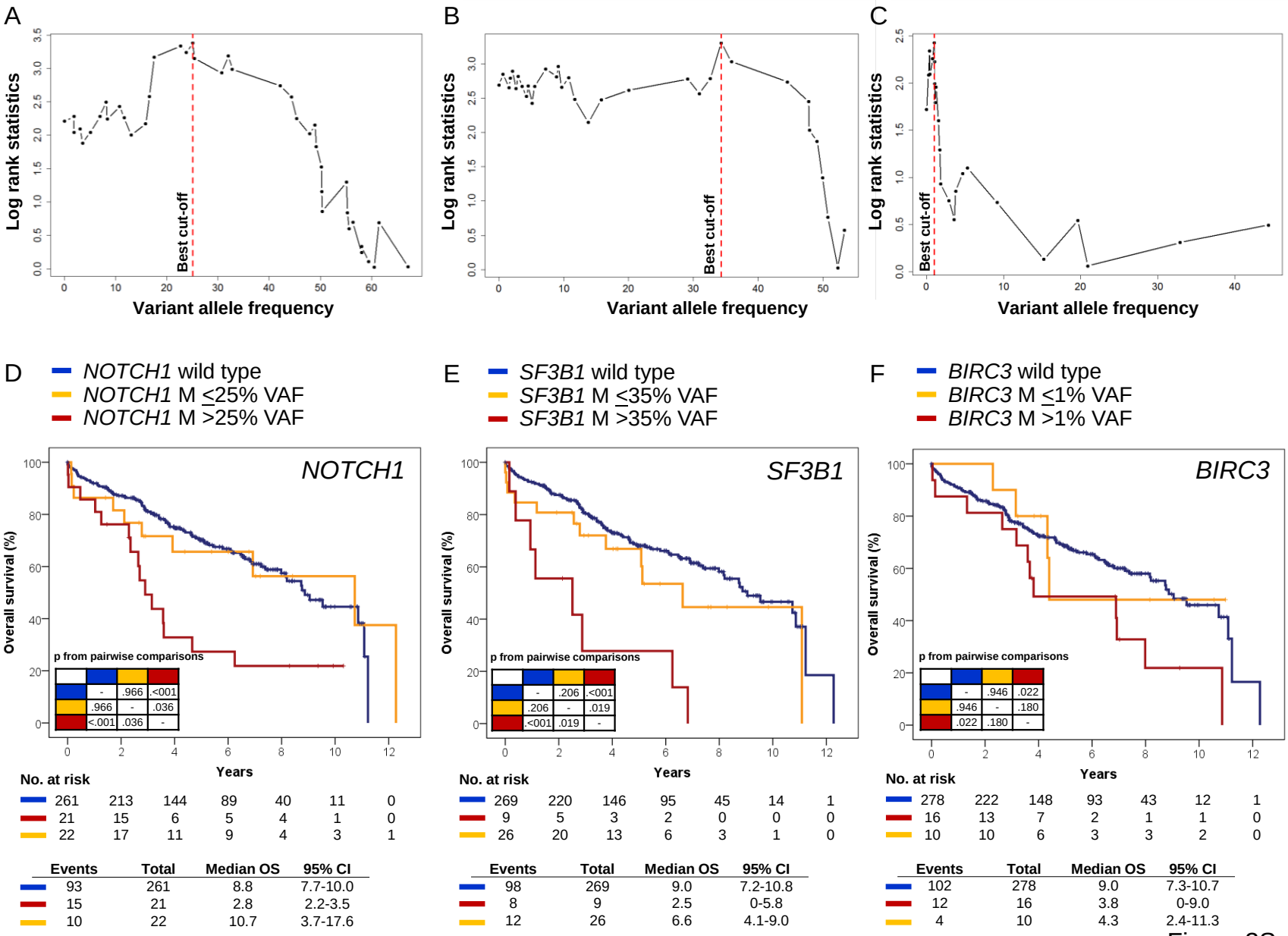

A
B
C
Log rank statistics
Best cut-off
Variant allele frequency
D
E
F
NOTCH1 wild type
NOTCH1 M ≤25% VAF
NOTCH1 M >25% VAF
SF3B1 wild type
SF3B1 M ≤35% VAF
SF3B1 M >35% VAF
BIRC3 wild type
BIRC3 M ≤1% VAF
BIRC3 M >1% VAF
NOTCH1
SF3B1
BIRC3
Overall survival (%)
Years
p from pairwise comparisons
No. at risk
Events
Total
Median OS
95% CI

Figure 3S

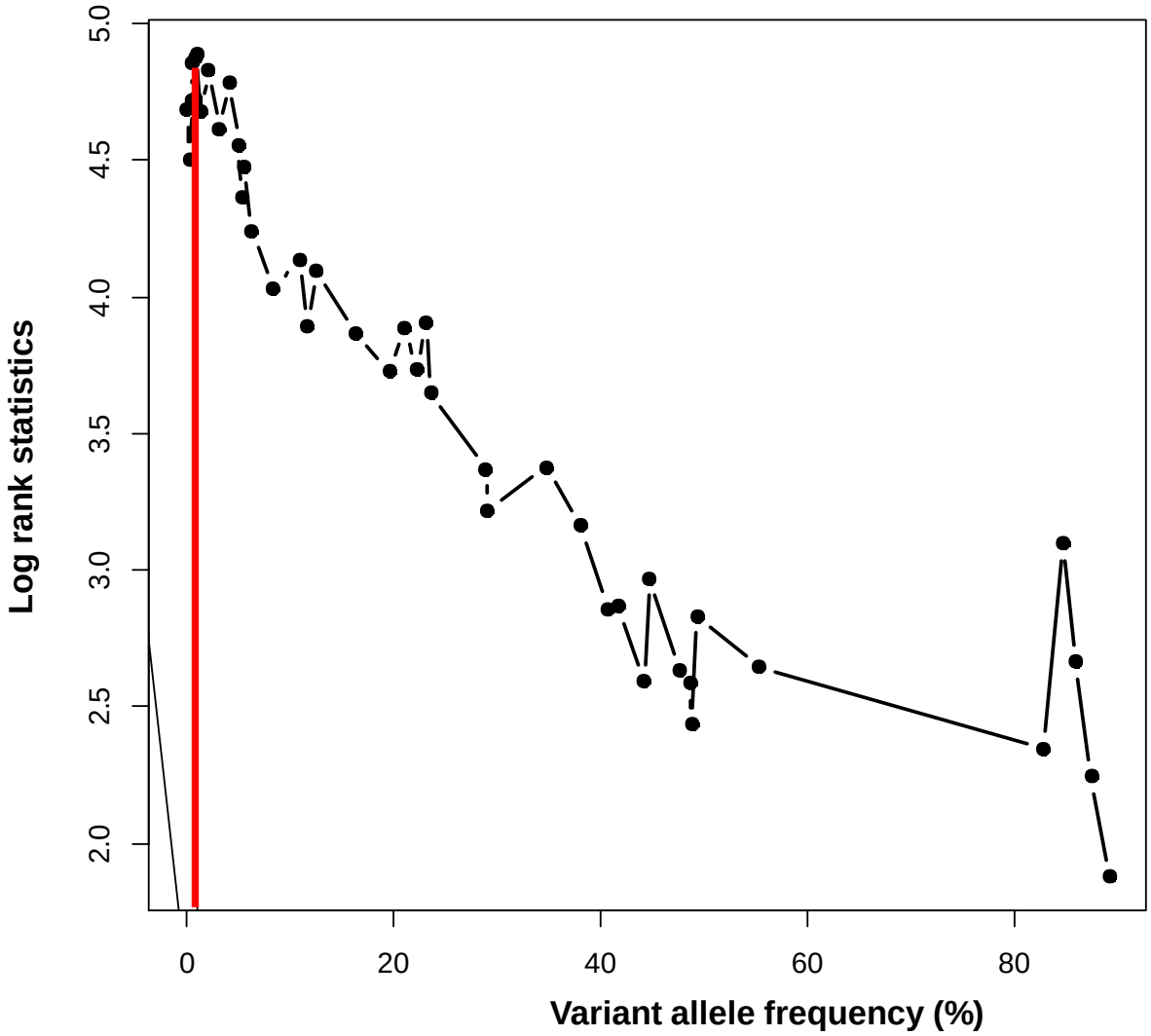
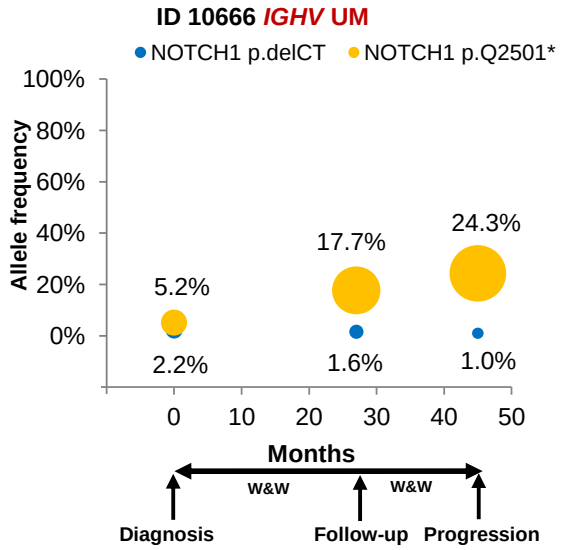
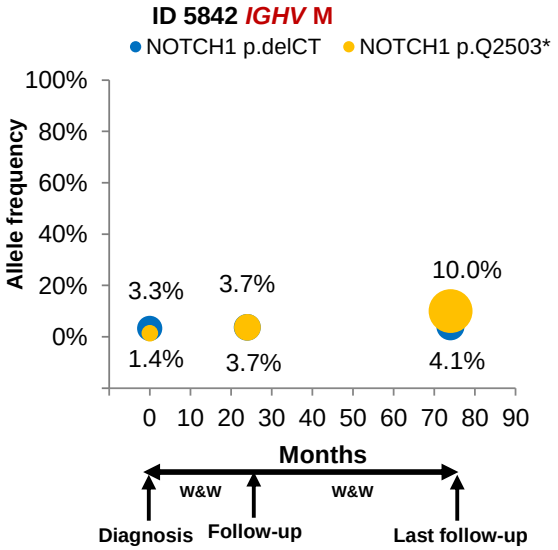
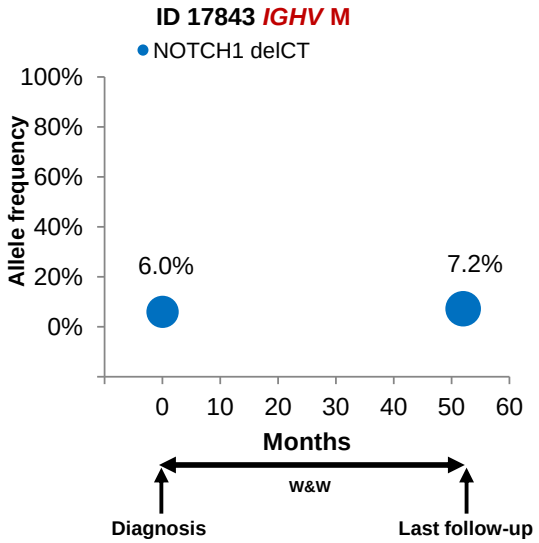

Log rank statistics
5.0
4.5
4.0
3.5
3.0
2.5
2.0
0
20
40
60
80
Variant allele frequency (%)

Figure 4S


ID 10666 IGHV UM
NOTCH1 p.delCT
NOTCH1 p.Q2501*
Allele frequency
100%
80%
60%
40%
20%
0%
5.2%
2.2%
17.7%
1.6%
24.3%
1.0%
0
10
20
30
40
50
Months
W&W
W&W
Diagnosis
Follow-up
Progression


ID 5842 IGHV M
NOTCH1 p.delCT
NOTCH1 p.Q2503*
Allele frequency
100%
80%
60%
40%
20%
0%
3.3%
1.4%
3.7%
3.7%
10.0%
4.1%
0
10
20
30
40
50
60
70
80
90
Months
W&W
W&W
Diagnosis
Follow-up
Last follow-up


ID 17843 IGHV M
NOTCH1 delCT
Allele frequency
100%
80%
60%
40%
20%
0%
6.0%
7.2%
0
10
20
30
40
50
60
Months
W&W
Diagnosis
Last follow-up

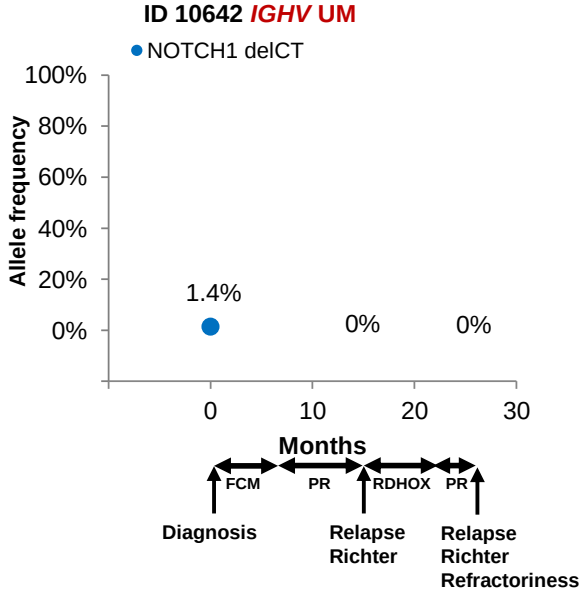
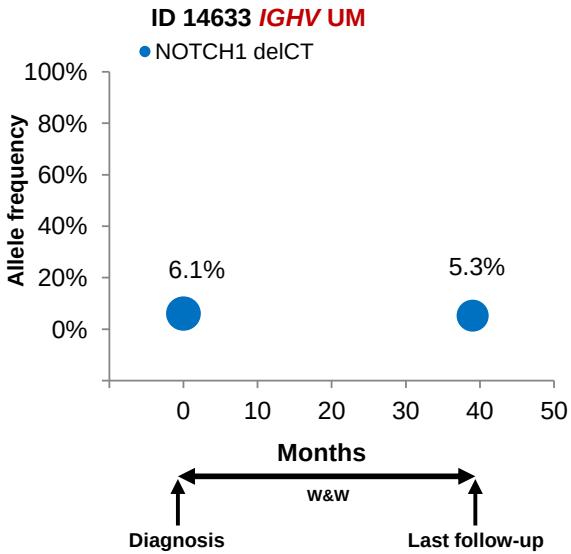
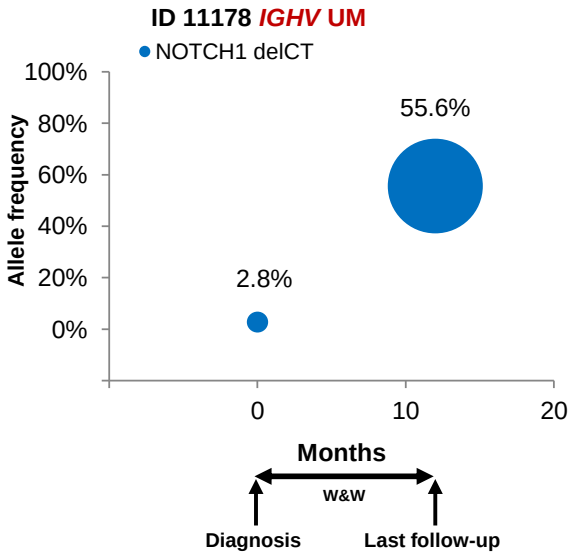

ID 10642 IGHV UM
NOTCH1 delCT
Allele frequency
100%
80%
60%
40%
20%
0%
1.4%
0%
0%
0
10
20
30
Months
FCM
PR
RDHOX
PR
Diagnosis
Relapse
Richter
Relapse
Richter
Refractoriness

Figure 5S


ID 14633 IGHV UM
NOTCH1 delCT
100%
80%
60%
40%
20%
0%
Allele frequency
6.1%
5.3%
0
10
20
30
40
50
Months
W&W
Diagnosis
Last follow-up


ID 11178 IGHV UM
NOTCH1 delCT
100%
80%
60%
40%
20%
0%
Allele frequency
55.6%
2.8%
0
10
20
Months
W&W
Diagnosis
Last follow-up

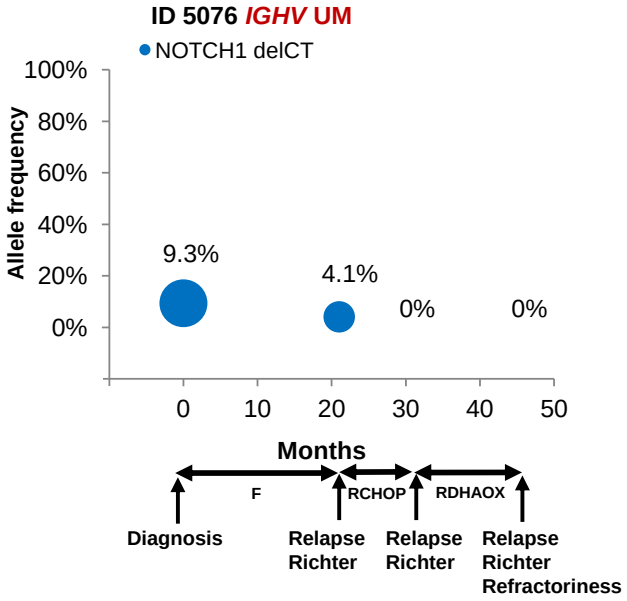
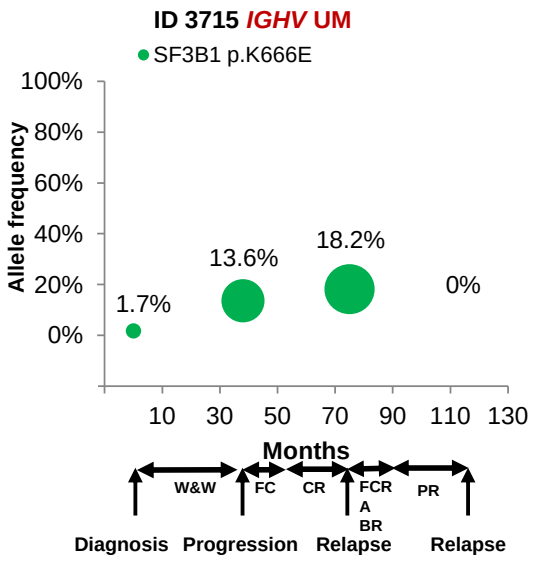
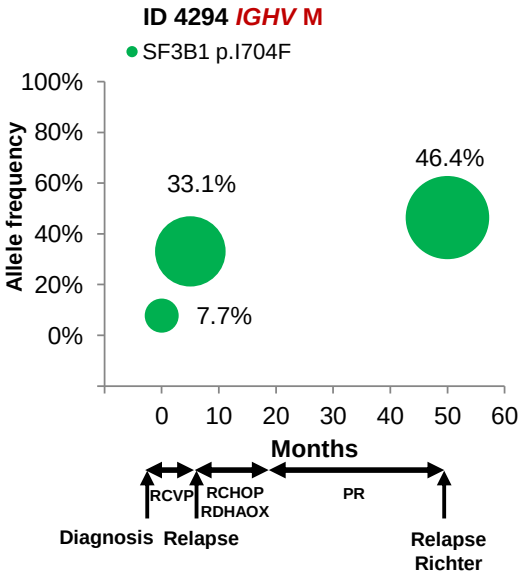
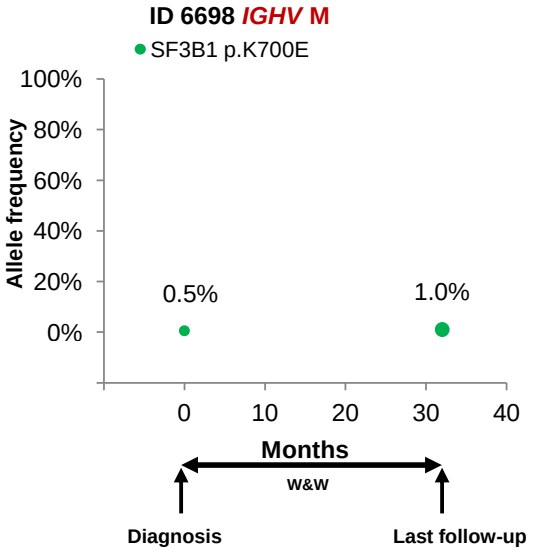

ID 5076 IGHV UM
NOTCH1 delCT
100%
80%
60%
40%
20%
0%
Allele frequency
9.3%
4.1%
0%
0%
0
10
20
30
40
50
Months
F
RCHOP
RDHAOX
Diagnosis
Relapse Richter
Relapse Richter
Relapse Richter Refractoriness

Figure 5S


ID 3715 IGHV UM
SF3B1 p.K666E
Allele frequency
100%
80%
60%
40%
20%
0%
1.7%
13.6%
18.2%
0%
10 30 50 70 90 110 130
Months
W&W
FC
CR
FCR
A
BR
PR
Diagnosis
Progression
Relapse
Relapse


ID 4294 IGHV M
SF3B1 p.I704F
Allele frequency
100%
80%
60%
40%
20%
0%
33.1%
7.7%
46.4%
0 10 20 30 40 50 60
Months
RCVP
RCHOP
RDHAOX
PR
Diagnosis
Relapse
Relapse
Richter


ID 6698 IGHV M
SF3B1 p.K700E
Allele frequency
100%
80%
60%
40%
20%
0%
0.5%
1.0%
0 10 20 30 40
Months
W&W
Diagnosis
Last follow-up

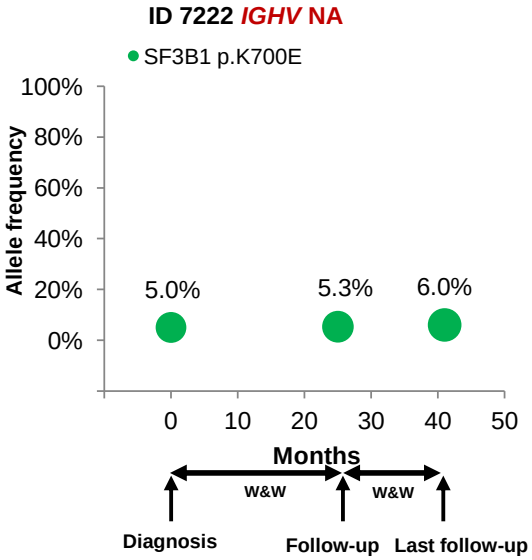
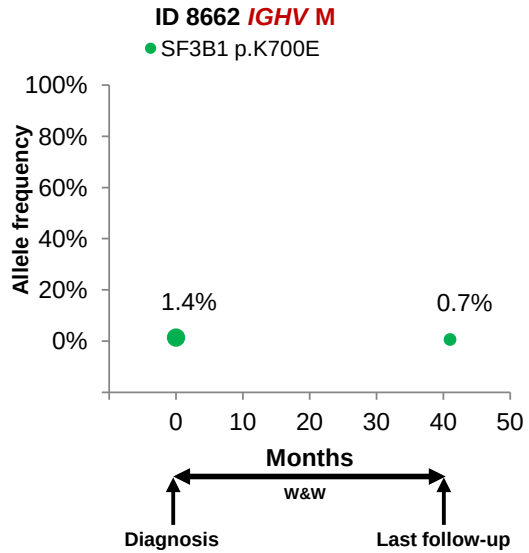
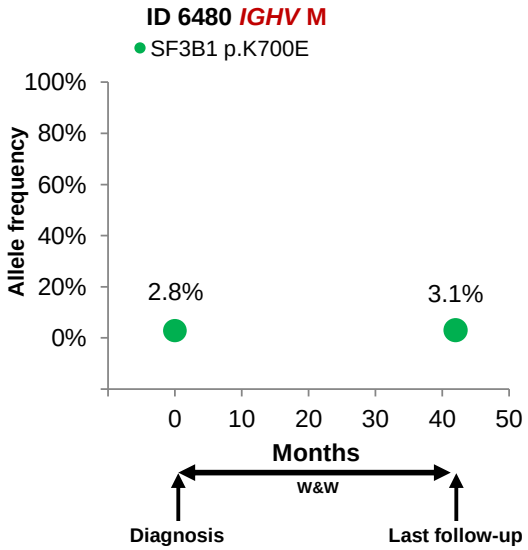
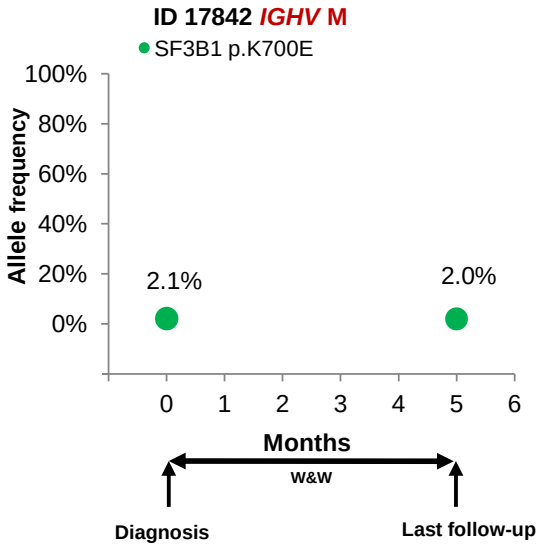

ID 7222 IGHV NA
SF3B1 p.K700E
Allele frequency
100%
80%
60%
40%
20%
0%
5.0%
5.3%
6.0%
0 10 20 30 40 50
Months
W&W
W&W
Diagnosis
Follow-up
Last follow-up

Figure 5S


ID 8662 IGHV M
SF3B1 p.K700E
100%
80%
60%
40%
20%
0%
Allele frequency
1.4%
0.7%
0
10
20
30
40
50
Months
W&W
Diagnosis
Last follow-up


ID 6480 IGHV M
SF3B1 p.K700E
100%
80%
60%
40%
20%
0%
Allele frequency
2.8%
3.1%
0
10
20
30
40
50
Months
W&W
Diagnosis
Last follow-up


ID 17842 IGHV M
SF3B1 p.K700E
100%
80%
60%
40%
20%
0%
Allele frequency
2.1%
2.0%
0
1
2
3
4
5
6
Months
W&W
Diagnosis
Last follow-up

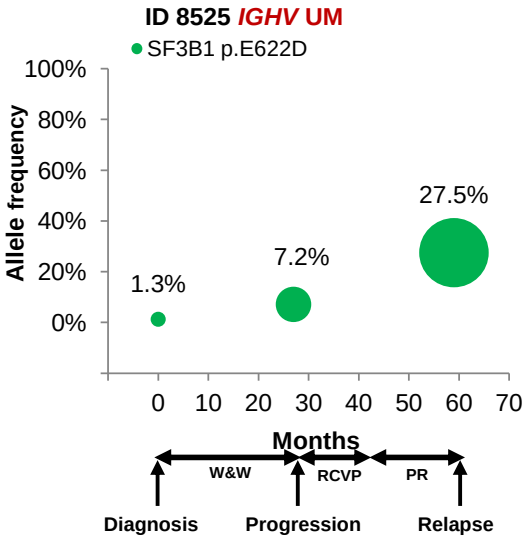
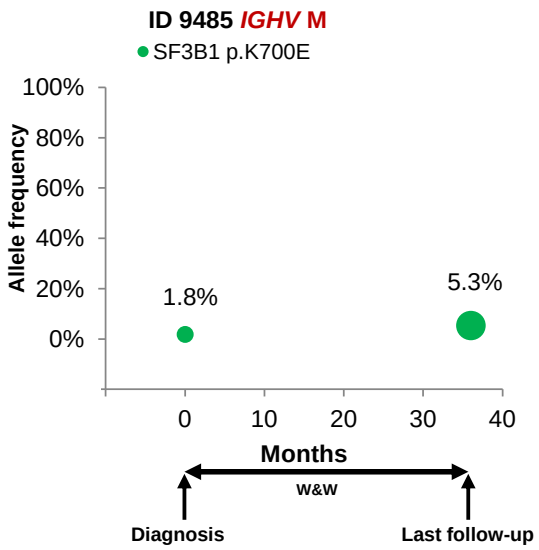
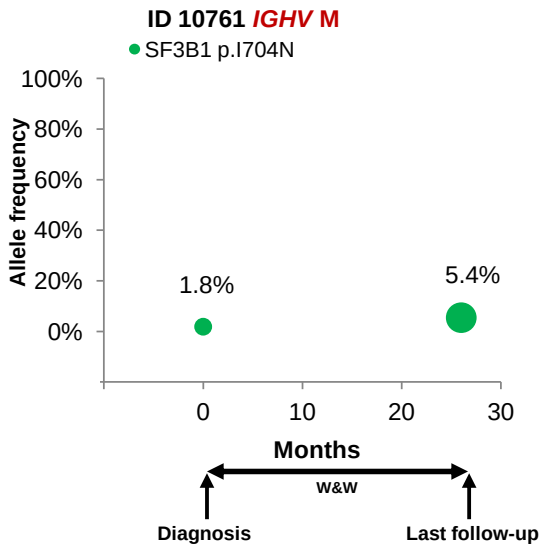
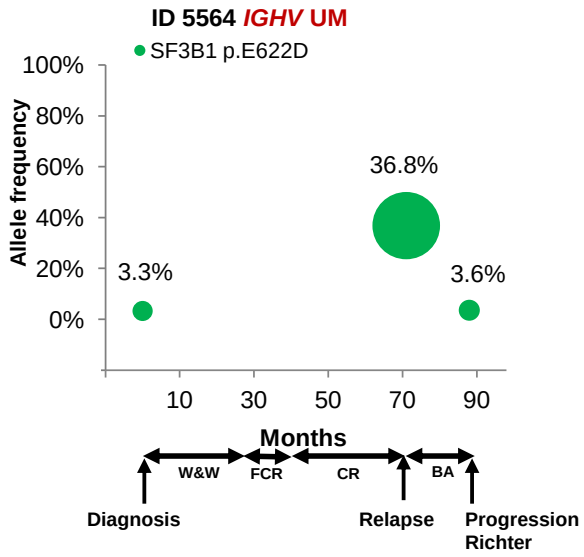

ID 8525 IGHV UM
SF3B1 p.E622D
100%
80%
60%
40%
20%
0%
Allele frequency
1.3%
7.2%
27.5%
0
10
20
30
40
50
60
70
Months
W&W
RCVP
PR
Diagnosis
Progression
Relapse

Figure 5S


ID 9485 IGHV M
SF3B1 p.K700E
Allele frequency
100%
80%
60%
40%
20%
0%
1.8%
5.3%
0
10
20
30
40
Months
W&W
Diagnosis
Last follow-up


ID 10761 IGHV M
SF3B1 p.I704N
Allele frequency
100%
80%
60%
40%
20%
0%
1.8%
5.4%
0
10
20
30
Months
W&W
Diagnosis
Last follow-up


ID 5564 IGHV UM
SF3B1 p.E622D
Allele frequency
100%
80%
60%
40%
20%
0%
3.3%
36.8%
3.6%
10
30
50
70
90
Months
W&W
FCR
CR
BA
Diagnosis
Relapse
Progression
Richter

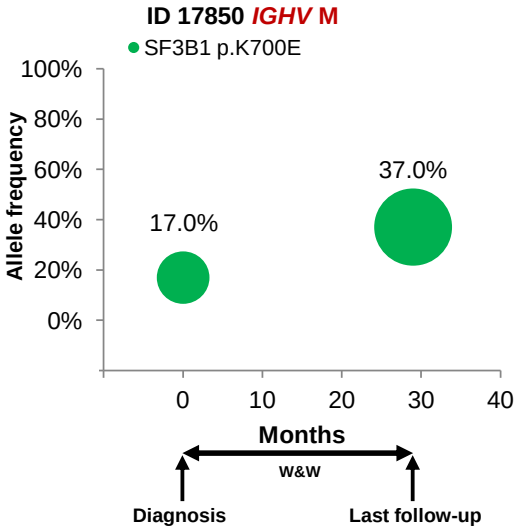
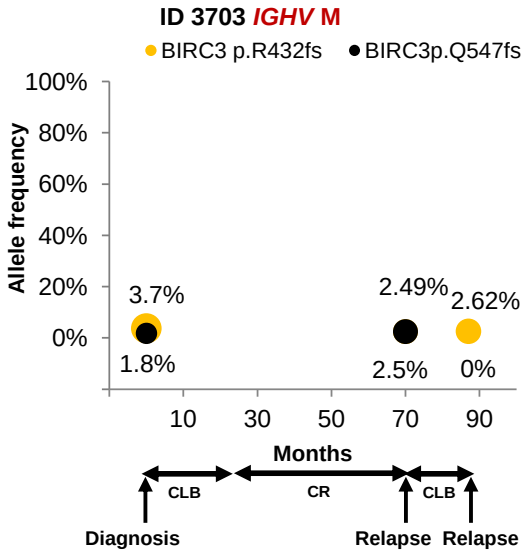
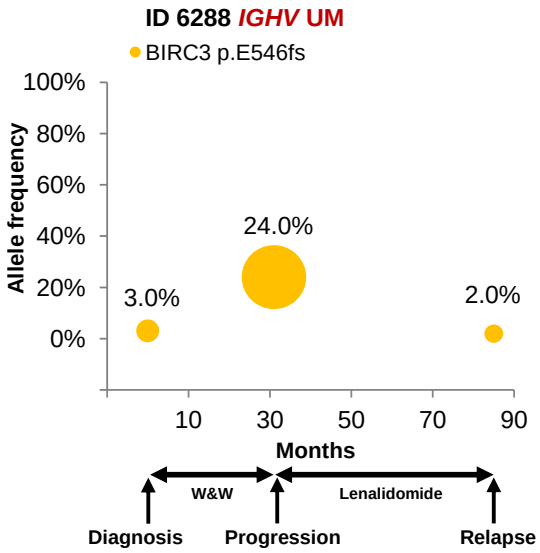
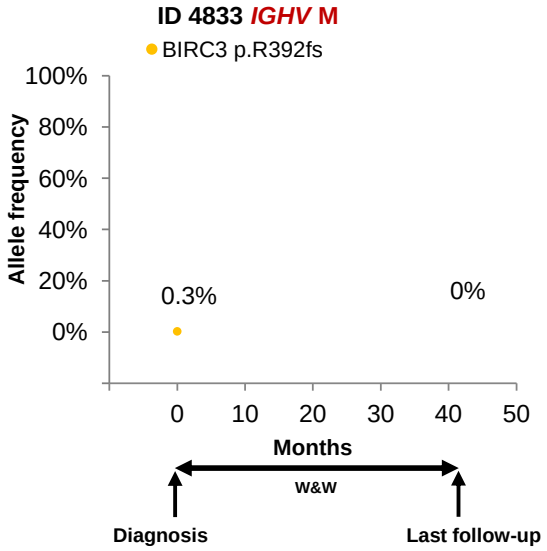

ID 17850 IGHV M
SF3B1 p.K700E
Allele frequency
100%
80%
60%
40%
20%
0%
17.0%
37.0%
0
10
20
30
40
Months
W&W
Diagnosis
Last follow-up

Figure 5S


ID 3703 IGHV M
BIRC3 p.R432fs
BIRC3p.Q547fs
Allele frequency
100%
80%
60%
40%
20%
0%
3.7%
1.8%
2.49%
2.5%
2.62%
0%
10
30
50
70
90
Months
CLB
CR
CLB
Diagnosis
Relapse
Relapse


ID 6288 IGHV UM
BIRC3 p.E546fs
Allele frequency
100%
80%
60%
40%
20%
0%
3.0%
24.0%
2.0%
10
30
50
70
90
Months
W&W
Lenalidomide
Diagnosis
Progression
Relapse


ID 4833 IGHV M
BIRC3 p.R392fs
Allele frequency
100%
80%
60%
40%
20%
0%
0.3%
0%
0
10
20
30
40
50
Months
W&W
Diagnosis
Last follow-up

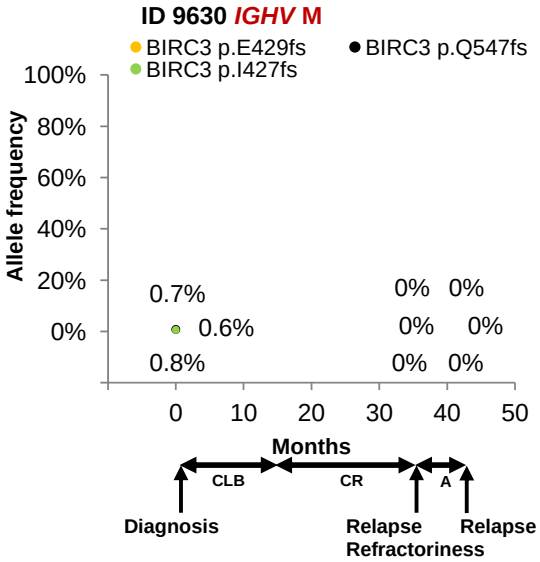
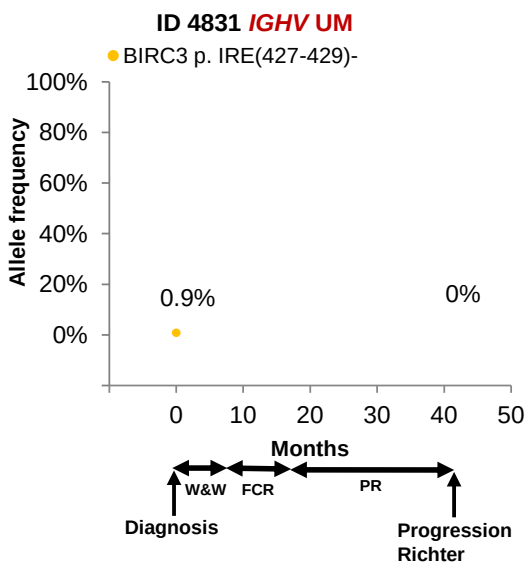

ID 9630 IGHV M
BIRC3 p.E429fs
BIRC3 p.Q547fs
BIRC3 p.I427fs
Allele frequency
100%
80%
60%
40%
20%
0%
0.7%
0.6%
0.8%
0% 0%
0% 0%
0% 0%
0
10
20
30
40
50
Months
CLB
CR
A
Diagnosis
Relapse
Refractoriness
Relapse

Figure 5S


ID 4831 IGHV UM
BIRC3 p. IRE(427-429)-
100%
80%
60%
40%
20%
0%
Allele frequency
0.9%
0%
0
10
20
30
40
50
Months
W&W
FCR
PR
Diagnosis
Progression
Richter

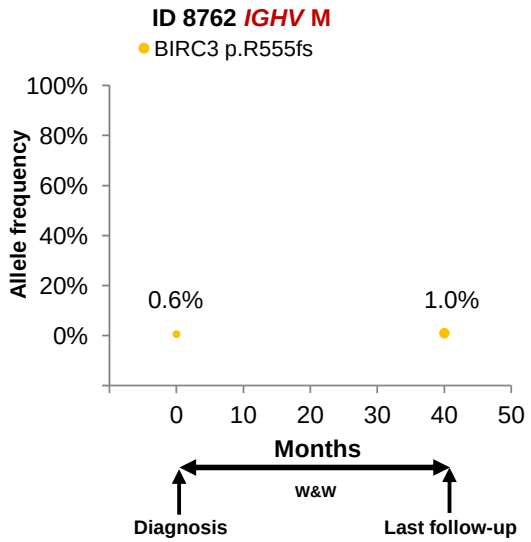

ID 8762 IGHV M
BIRC3 p.R555fs
100%
80%
60%
40%
20%
0%
Allele frequency
0.6%
1.0%
0
10
20
30
40
50
Months
W&W
Diagnosis
Last follow-up

Figure 5S