SUPPLEMENTARY APPENDIX

Multiclonal complexity of pediatric acute lymphoblastic leukemia and the prognostic relevance of subclonal mutations

Željko Antić,[1*] Jiangyan Yu,[1,2*] Simon V. van Reijmersdal,[1,2] Anke van Dijk,[2] Linde Dekker,[1] Wouter H. Segerink,[1] Edwin Sonneveld,[1,3] Marta Fiocco,[1,4,5] Rob Pieters,[1,3] Peter M. Hoogerbrugge,[1,3] Frank N. van Leeuwen,[1] Ad Geurts van Kessel,[2] Esmé Waanders[1,6] and Roland P. Kuiper[1,6]

[1]Princess Máxima Center for Pediatric Oncology, Utrecht; [2]Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen; [3]Dutch Childhood Oncology Group, Utrecht; [4]Medical Statistics, Department of Biomedical Data Science, Leiden University Medical Center, Leiden; [5]Mathematical Institute, Leiden University and [6]Department of Genetics, University Medical Center Utrecht, Utrecht, the Netherlands

**ŽA and JY contributed equally as co-first authors.*

 doi:10.3324/haematol.2020.259226

Received: May 13, 2020.

Accepted: October 23, 2020.

Pre-published: November 5, 2020.

Correspondence: *ROLAND P. KUIPER* - r.kuiper@prinsesmaximacentrum.nl

Supplementary Data

Multiclonal complexity of pediatric acute lymphoblastic leukemia and the prognostic relevance of subclonal mutations

Željko Antić[1,±], Jiangyan Yu[1,2,±], Simon V. van Reijmersdal[1,2], Anke van Dijk[2], Linde Dekker[1], Wouter H. Segerink[1], Edwin Sonneveld[1,3], Marta Fiocco[1,4,5], Rob Pieters[1,3], Peter M. Hoogerbrugge[1,3], Frank N. van Leeuwen[1], Ad Geurts van Kessel[2], Esmé Waanders[1,6], Roland P. Kuiper[1*]

1. Princess Maxima Center for Pediatric Oncology, Utrecht, The Netherlands
2. Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands
3. Dutch Childhood Oncology Group, Utrecht, The Netherlands
4. Medical Statistics, Department of Biomedical Data Science, Leiden University Medical Center, Leiden, The Netherlands.
5. Mathematical Institute, Leiden University, The Netherlands.
6. Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands

± These authors contributed equally to this work.

* Corresponding author

SUPPLEMENTARY MATERIAL AND METHODS

smMIP-based sequencing and variant calling

In order to accurately detect subclonal alterations in diagnosis samples, a total of 166 smMIP probes was designed to cover the hotspot regions of the genes *CREBBP*, *PTPN11*, *NT5C2,* and *WHSC1*, and coding regions of *TP53, KRAS* and *NRAS*, seven genes that are frequently mutated in relapsed ALL (Supplementary Table 4). All genomic regions of interest were covered by at least two probes, preferably covering both the sense and antisense strands. smMIP-based sequencing was performed as previously described using paired-end sequencing on an Illumina NextSeq 500 Desktop Sequencer (Illumina, CA, USA), after which smMIP-based consensus variant calling was performed using SeqNext software (JSI) version 4.2.5, as previously described[1]. Mutant allele frequencies were corrected based on the blast percentage of the sample. We achieved an average on-target read depth of 22,985 raw reads per probe (Supplementary Figure 1A). After removing random errors present in less than 70% of raw reads, consensus reads were formed from reads with the same unique molecular identifier. We achieved an average on-target depth of 308 (median 141) unique-capture-based consensus reads per probe (Supplementary Figure 1B), with multiple probes overlapping on hotspot regions[1]. After exclusion of the variants called from repetitive regions, poorly performing probes, variants called from less than 2 independent probes and variants called by less than 5 unique reads in one consensus read, a total of 7,836 variants remained with an average on-target depth of 1,419 consensus reads. We further filtered out variants present in an in-house database of the Radboud University Medical Center (Nijmegen, the Netherlands) containing exomes from 20,000 individuals[2] and variants predicted as non-pathogenic (synonymous, phyloP <2.5, CADD score < 15) (Supplementary Table 3). For the final list of variants, a correction of the mutant allele frequency was made based on the percentage of blast cells determined at the time of diagnosis, which was high for the majority of cases (>70% for 93% of cases). In a low number of cases an aneuploidy required a correction of mutant allele frequency for

copy number as well, which involved *WHSC1* (17 mutations), *TP53* (10 mutations), *KRAS* (6 mutations), *NRAS* (5 mutations), *PTPN11* (3 mutations) and *CREBBP* (1 mutation) (Supplementary Table 5).

IKZF1 deletion detection

IKZF1 status was assessed using the Multiplex ligation probe assay (MLPA) SALSA P335 ALL-IKZF1 kit (MRC-Holland, The Netherlands), according to manufacturer's instructions and as described before[3,4]. Additionally, *IKZF1* 4-7 deletions were assessed using real time quantitative PCR. Primers covering the breakpoint clusters in introns 3 and 7 were designed using Primer3 software version 0.4.0. Quantitative PCR (qPCR) was performed using an IQ SYBR Green supermix (Biorad, CA, USA) according to manufacturer's instructions. The primer sequences used for qPCR were: 5'-CTCCCAGCCCATAGGGTATAA-3' (forward) and 5'-GTTAAATAAAGAACCCTCAGGCATT-3' (reverse). The sensitivity of the qPCR assay was tested using dilution series of a sample with a high tumor load (96% blasts) and a full-clonal *IKZF1* del4-7 (detected by MLPA). All qPCR reactions were performed in duplicate, and the percentage of cells with *IKZF1* exon 4-7 deletions was calculated based on the dilution series of the control sample, and a correction was made based on the percentage of blast cells determined at the time of diagnosis. For every sample with a clonal or subclonal *IKZF1* exon 4-7 deletion, PCR products were sequenced with both forward and reverse primers using Sanger sequencing, after which the sequences were mapped to the reference genome (hg19) to determine the exact breakpoint positions and unique interstitial sequences (Supplementary Figure 1C and Supplementary Table 6).

Defining clonal and subclonal alterations

Previous studies have used allele frequency (AF) thresholds to define mutations as being present in a minor subclonal or major clone (clonal) ranging between 20% and 30%[5-8]. In this study, we used an AF threshold of 25% to separate major clonal from subclonal alterations, since mutations below 25% represent a minor cell fraction. The same threshold was used for defining clonal *IKZF1* 4-7 deletions, which correlates with deletions that are detectable using MLPA.

Cox regression analysis

Multivariate Cox regression model was estimated including all covariates significant in the univariate model, as well as age at diagnosis, gender and MRD as clinically relevant covariates. Multivariable model including combined ALL9 and ALL10 cohorts was stratified based on the treatment protocol. Proportional hazard assumption was checked by visual inspection of Schoenfeld Residuals. The score test was used to test violation of the proportional hazard assumption for each variable[9]. We did not identify violation of the proportional hazard assumption for any of the tested covariates. Potential multicolinearity was inspected using variance inflation factor (VIF). Univariate interactions were inspected for each of the tested covariates and reported in the supplementary table 11.

REFERENCES

1. Yu J, Antic Z, van Reijmersdal SV, Hoischen A, Sonneveld E, Waanders E, et al. Accurate detection of low-level mosaic mutations in pediatric acute lymphoblastic leukemia using single molecule tagging and deep-sequencing. Leuk Lymphoma. 2017 Oct 23:1-10.

2. Neveling K, Feenstra I, Gilissen C, Hoefsloot LH, Kamsteeg EJ, Mensenkamp AR, et al. A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. Hum Mutat. 2013 Dec;34(12):1721-6.

3. Kuiper RP, Waanders E, van der Velden VH, van Reijmersdal SV, Venkatachalam R, Scheijen B, et al. IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL. Leukemia. 2010 Jul;24(7):1258-64.

4. Waanders E, van der Velden VH, van der Schoot CE, van Leeuwen FN, van Reijmersdal SV, de Haas V, et al. Integrated use of minimal residual disease classification and IKZF1 alteration status accurately predicts 79% of relapses in pediatric acute lymphoblastic leukemia. Leukemia. 2011 Feb;25(2):254-8.

5. Ma X, Edmonson M, Yergeau D, Muzny DM, Hampton OA, Rusch M, et al. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. Nat Commun. 2015;6:6604.

6. Spinella J-F, Richer C, Cassart P, Ouimet M, Healy J, Sinnett D. Mutational dynamics of early and late relapsed childhood ALL: rapid clonal expansion and long-term dormancy. Blood Adv. 2018;2(3):177-88.

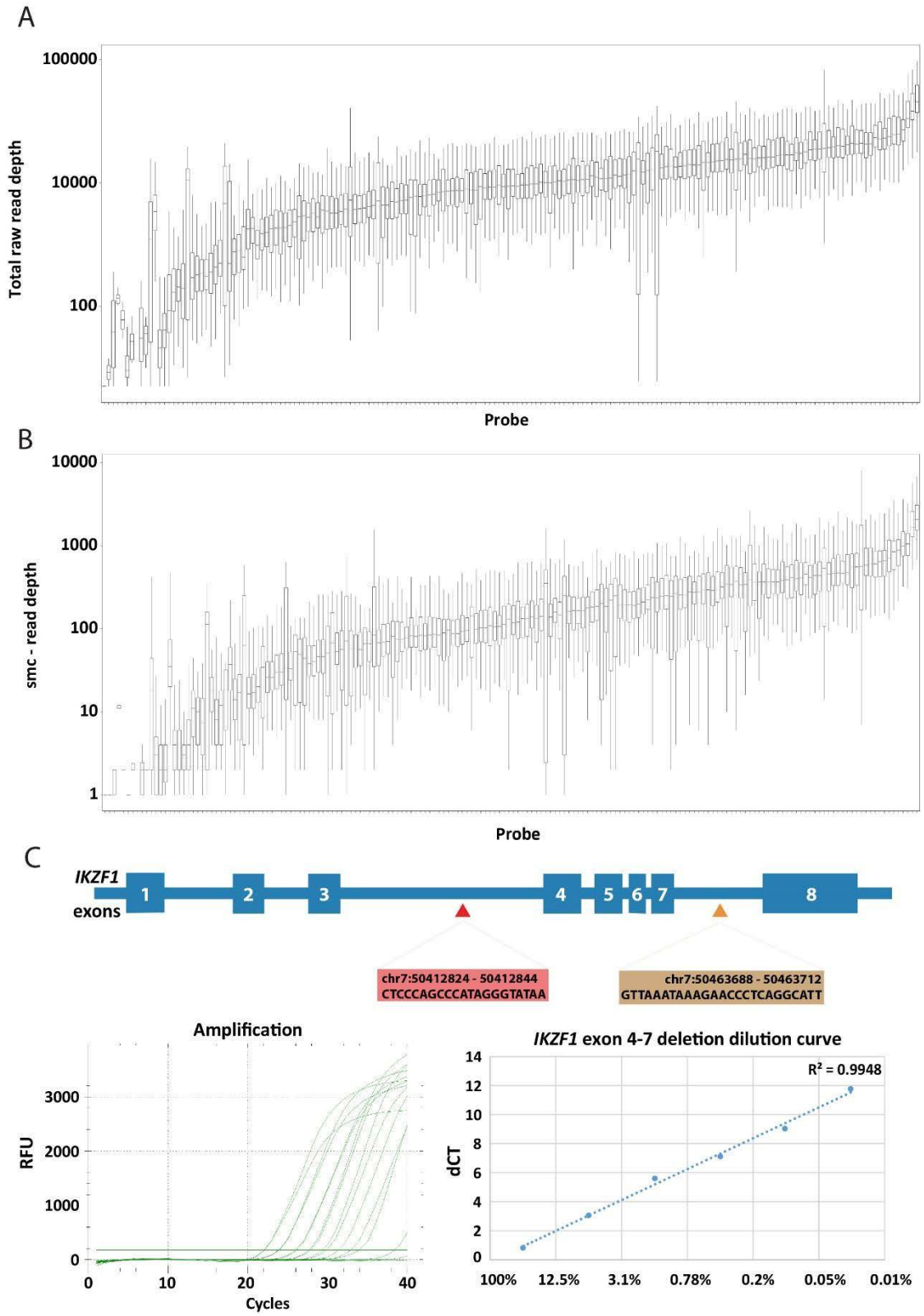
7. Jerchel IS, Hoogkamer AQ, Ariës IM, Steeghs EMP, Boer JM, Besselink NJM, et al. RAS pathway mutations as a predictive biomarker for treatment adaptation in pediatric B-cell precursor acute lymphoblastic leukemia. Leukemia. 2018;32(4):931-40.

8. Agraz-Doblas A, Bueno C, Bashford-Rogers R, Roy A, Schneider P, Bardini M, et al. Unraveling the cellular origin and clinical prognostic markers of infant B-cell acute lymphoblastic leukemia using genome-wide analysis. Haematologica. 2019;104(6):1176-88.

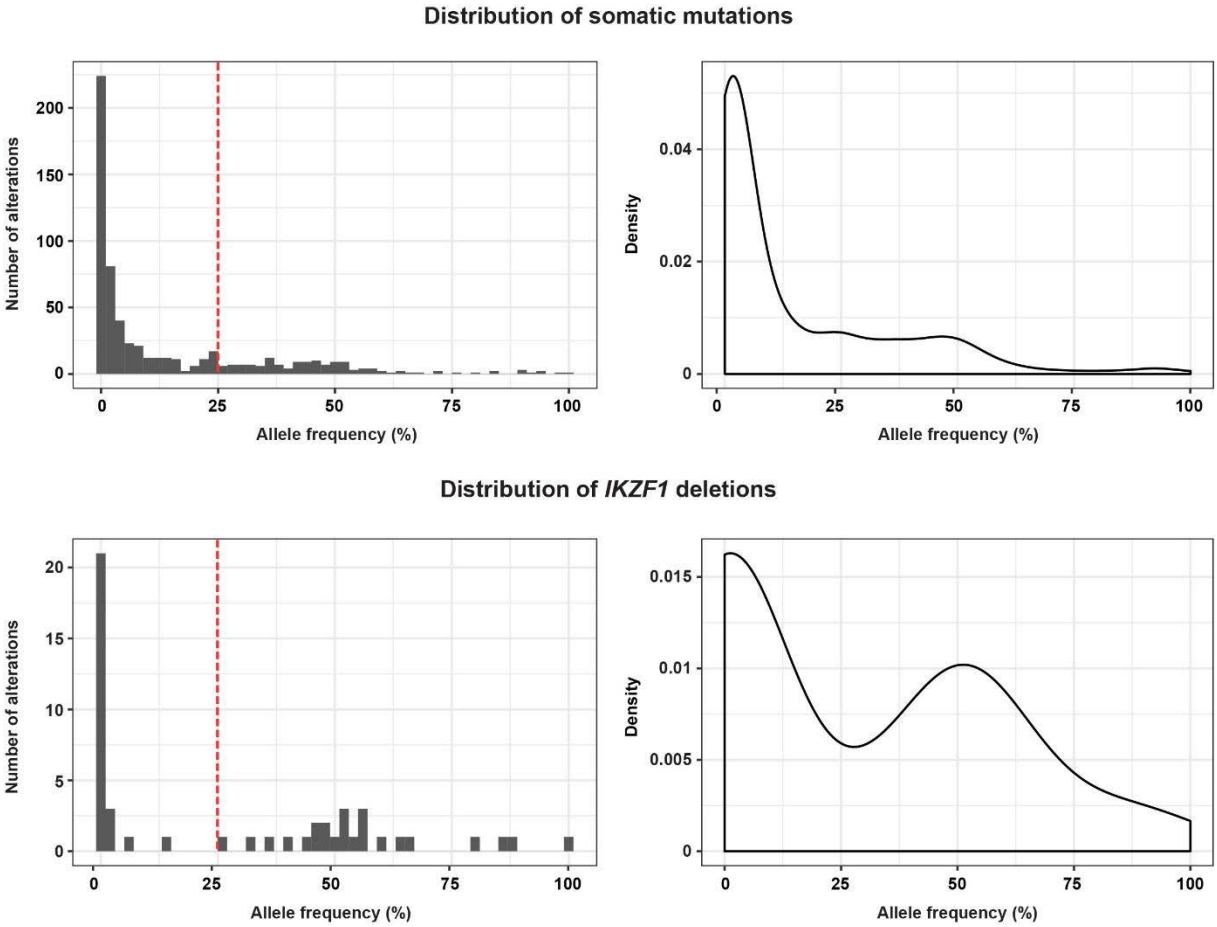
9. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 1994;81(3):515-26.

SUPPLEMENTARY FIGURES

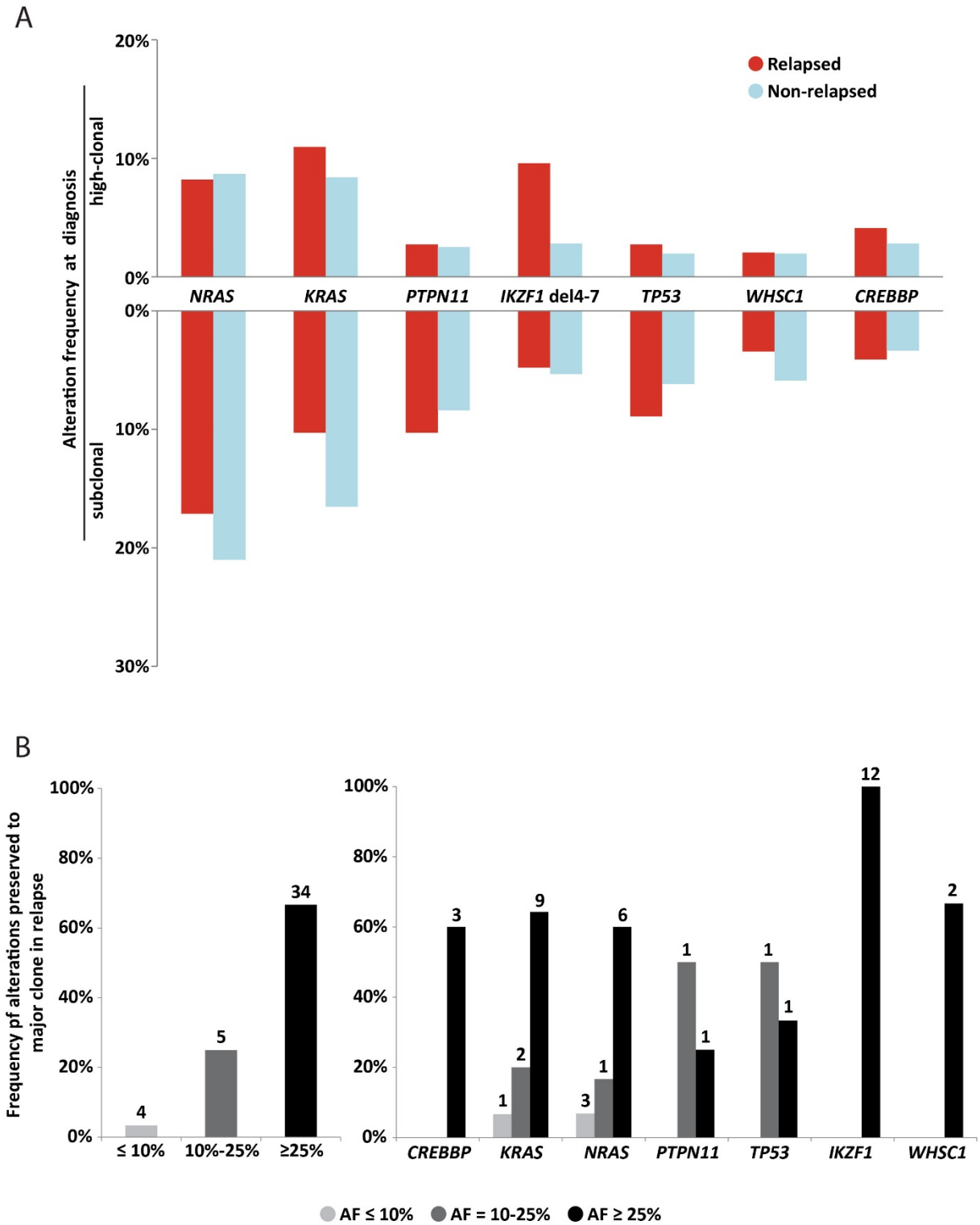
List of supplementary figures



Supplementary Figure 1 – **(A)** Box plot showing total raw read depth per probe. **(B)** Box plots showing unique-capture-based consensus reads (smc) depth per probe. **(C)** Schematic representation of the *IKZF1* gene, indicating the position of the common breakpoint cluster in introns 3 and 7 (top panel). The bottom left panel shows a representative example of qPCR amplification curves of dilution series and the bottom right panel displays the dilution curve from the same experiment indicating high correlation (R2 = 0.9948) for quantifiable samples.



Supplementary Figure 2 – Histogram (left panel) and density plot (right panel) showing the distribution of the mutant allele frequency in cases with mutations and *IKZF1* 4-7 deletions. Allele frequency (AF) of 25% was chosen as the threshold for clonal versus subclonal mutations (indicated in red). Mutations above this threshold are present in more than 50% of the cells and thus always represent a major clone. As depicted in these panels the majority of (subclonal) mutations and deletions have AF below 10%.



Supplementary Figure 3 – **(A)** Bar plot showing alteration frequencies at diagnosis for relapsed (red) and non-relapsed cases (blue). Major clone alterations are shown upwards, while subclonal are presented downwards. **(B)** Bar plot indicating fraction of alterations preserved to major clones in available relapse samples for all genes tested (left panel) and per gene (right panel). Samples are grouped based on the alteration allele frequency detected in diagnosis. Absolute numbers for each group are shown on the top of the bars (supplementary table 7).

SUPPLEMENTARY TABLES

Supplementary tables legends

Supplementary Table 1 – Clinical characteristics of patients included in the study.

Supplementary Table 2 – Comparison between total ALL10 cohort and the representative selection used in this study.

Supplementary Table 3 – Overview of all mutations detected in this study per patient

Supplementary Table 4 – Sequence of smMIP probes used in this study.

Supplementary Table 5 – Overview of all the cases in which, based on karyotype, estimated mutant allele frequency was affected by aneuploidies.

Supplementary Table 6 – Overview of all *IKZF1* exon 4-7 deletions detected using breakpoint-spanning PCR and Sanger sequencing in diagnosis samples.

Supplementary Table 7 – Distribution of detected alterations based on allele frequency.

Supplementary Table 8 – Frequency of subclonal alterations at diagnosis in ALL subtypes.

Supplementary Table 9 – Frequency of relapse in total primary diagnosis cohort (n = 503) based on mutation status in eight relapse-associated genes.

Supplementary Table 10 – MRD status in cases with *IKZF1* 4-7 deletions

Supplementary Table 11 – Alteration frequency at diagnosis of non-relapsed (NR) and relapsed (R) cases in representative cohorts of ALL9 and ALL10 and Cox regression analysis.

Supplementary Table 12 – Distribution of *ERG* deletions in B-other cases with *IKZF1* 4-7 deletions

Supplementary Table 13 – Frequency of the alterations detected at diagnosis in cases that relapsed and frequency of the preserved alterations in relapsed clones for available relapse samples.

Supplementary Table 14 – Status of mutations detected at diagnosis in relapse samples.