SUPPLEMENTAL MATERIAL

METHODS

Sequencing and Variant Calling

Duplex sequencing[1,2] was performed using the Twinstrand Duplex Sequencing Kit, according to the manufacturer's protocol (Twinstrand Biosciences, Seattle, WA). We utilized the TwinStrand DuplexSeq AML-29 Assay, which incorporates a targeted panel of 29 genes to detect single nucleotide variants, insertions and deletions (indels), and structural variants that are recurrently mutated in adult AML, CH, and myelodysplastic syndromes. The sequencing format was 150nt PE from the NovaSeq 6000 platform using S4-XP flow cells. Per sample, our average read depth was 15,564 (interquartile range, 14,893-16,308), and our average total raw read count was 2.72×10^8 (interquartile range, 2.52×10^8-2.94×10^8).

Error-corrected consensus sequence alignments to the GRCh38 reference genome in BAM file format were obtained using a customized DNAnexus platform.[3] Variant calling was performed using four variant callers: Mutect2 (GATK 4.1.0.0),[4] VarDict (1.7.0),[5] LoFreq (2.1.2),[6] and Varscan2 (2.4.2).[7] All variant callers were utilized in tumor-only mode with default settings. Variants were annotated according to VEP94, COSMIC v92, dbSNP (avsnp150, hg38), ANNOVAR (June 7, 2020, update), and Intervar (March 2019 update).[8]

Post-Processing Variant Filtering and Annotation

We applied a series of post-processing filters to remove possible false-positive variants and putative germline polymorphisms adapted from published methods.[9,10] In the initial phase, we utilized a panel of normal (PoN) samples (6 samples derived from healthy individuals aged 18-24 years) to filter sequencing artifacts. First, we excluded any variants detected in ≥ 2 PoN samples with a mean alternative read count of ≥ 10. Next, we

calculated exact nucleotide counts from BAM files in all patients and PoN samples using *getbasecounts* (1.4.0). We then conducted a Fisher's exact test for each mutation to compare the alternative and total reads of that mutation in an individual sample to the alternative and total reads across all PoN samples.[9] We excluded variants with a Bonferroni-corrected *P* value > 8.5×10^{-7}, which indicated insufficient evidence that the candidate mutations alternative and total reads counts statistically significantly differed from the PoN reference values. Third, we calculated the average VAF and standard deviation across all remaining variants in the PoN. We calculated a Z-score to compare the VAF in retained variants in patients to the reference PoN values. Variants were excluded if they had a Z-score of ≤ 1.96, indicating insufficient evidence that their VAF differed from that observed within the PoN.

We additionally excluded multiple nucleotide variants and variants with low mapping quality (i.e. no Varscan2 value of "Pass" or a mapping quality of at least 60 for other callers). Indels ≥ 10 base pairs (bp) were excluded except where detected by Pindel (0.2.5b9).[11] We required mutations detected by a single caller to have an alternative read count of ≥ 5 for single nucleotide variants and ≥ 10 for indels or, if detected by two or more callers, an alternative read count of ≥ 3. We did not employ a minimum VAF threshold for inclusion. Single-nucleotide deletions occurring within a homopolymer stretch with a repetition of ≥ 3 base pairs and single-nucleotide substitutions completing a stretch with a homopolymer length of ≥5 base pairs were removed via Integrative Genomics Viewer.[11] In order to filter out potential germline contributions we excluded variants if their minor allele frequency from either the 1000 Genomes (August 2015 update) or gnomAD (gnomad211) databases was > 0.01% or they were detected in > 10% of pretreatment patient samples, except where a variant had a minor allele frequency < 0.5% and was reported as somatic in the Catalogue of Somatic Mutations in Cancer (COSMIC, v92) ≥ 10 times. Finally, we excluded variants with a VAF of 0.45-0.55 or >0.95 to further filter potential germline variants.

SUPPLEMENTAL FIGURES

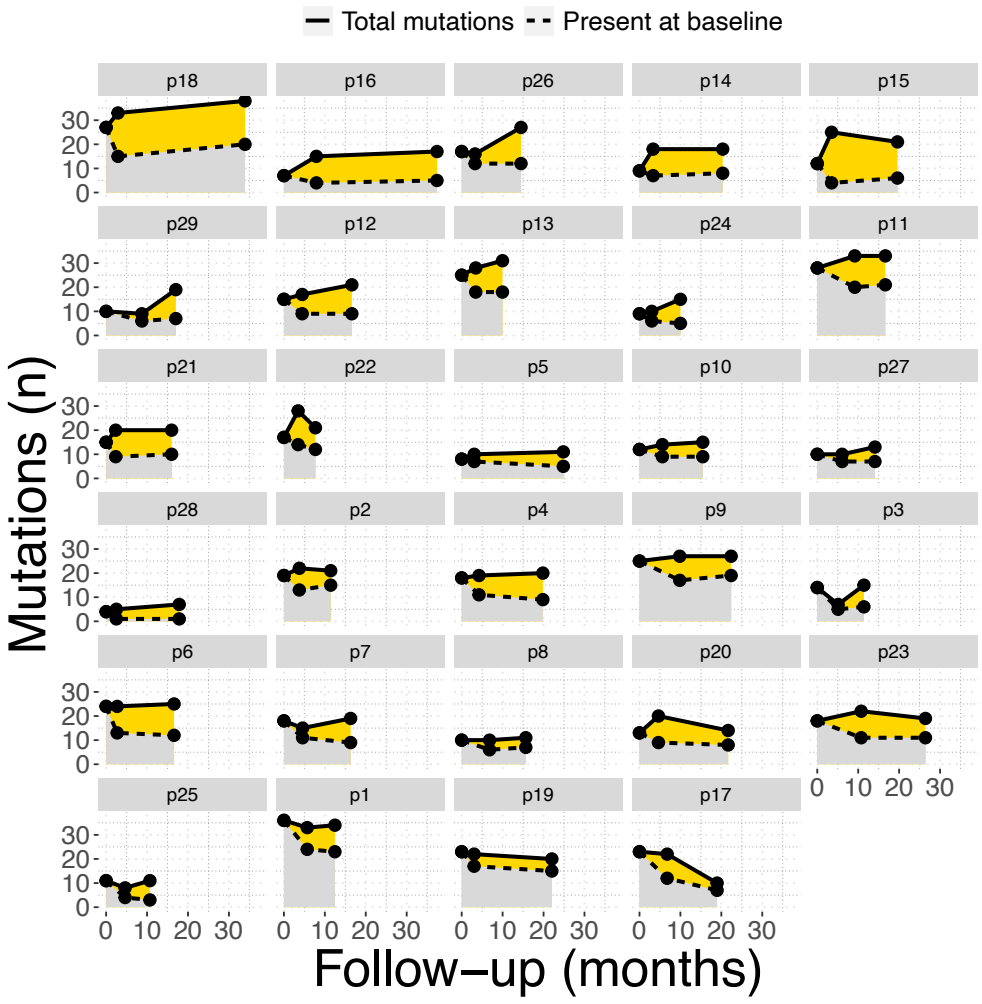
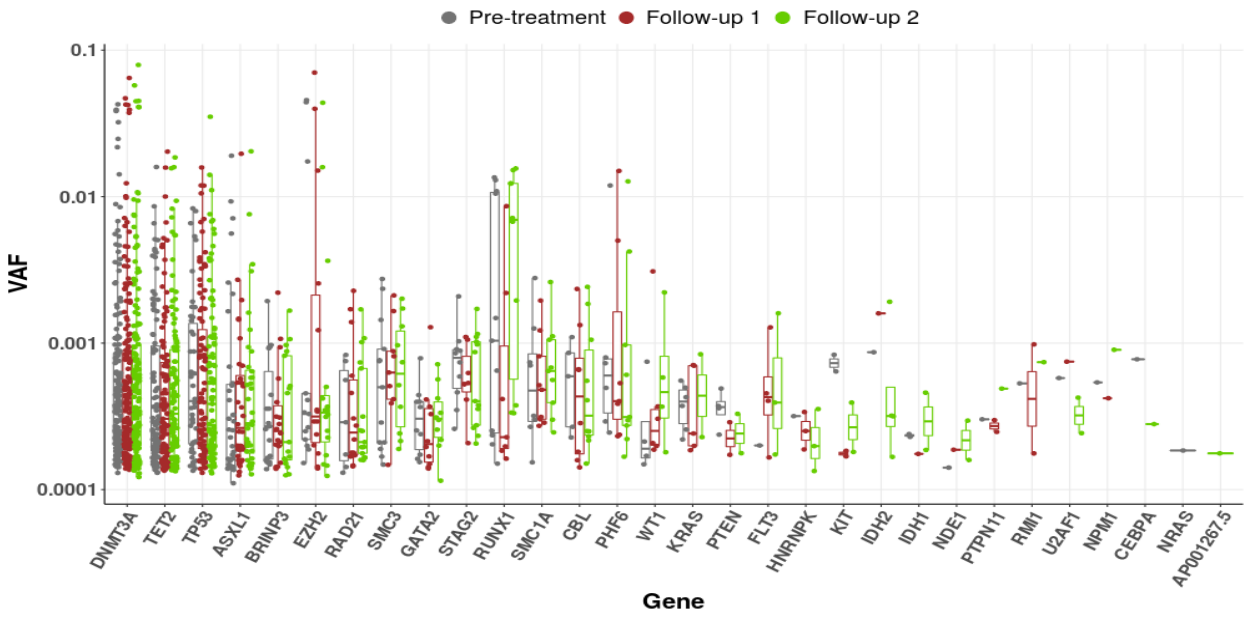


FIG 1. Clonal hematopoiesis mutations detected before treatment and at post-treatment follow-up timepoints by individual.

Line plot for each participant showing the total number of mutations detected at each timepoint (solid line) and the number of mutations detected before treatment that were subsequently detected at post-treatment timepoints (dashed line). The grey shaded region represents the number of unique mutations for each timepoint that were initially present at the baseline. The yellow shaded region represents the number of unique mutations for each timepoint that were not initially present at baseline.

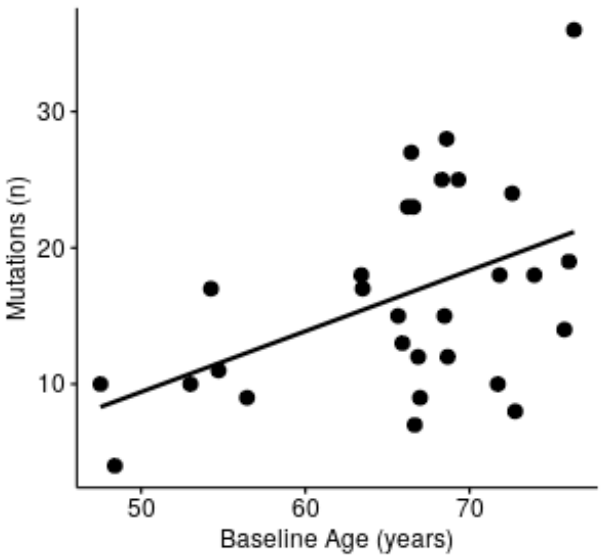
n, number.



SUPPLEMENTAL FIG 2. Variant allele fraction for identified clonal hematopoiesis mutations, stratified by gene and timepoint.

Y-axis presented on a log-10 scale.

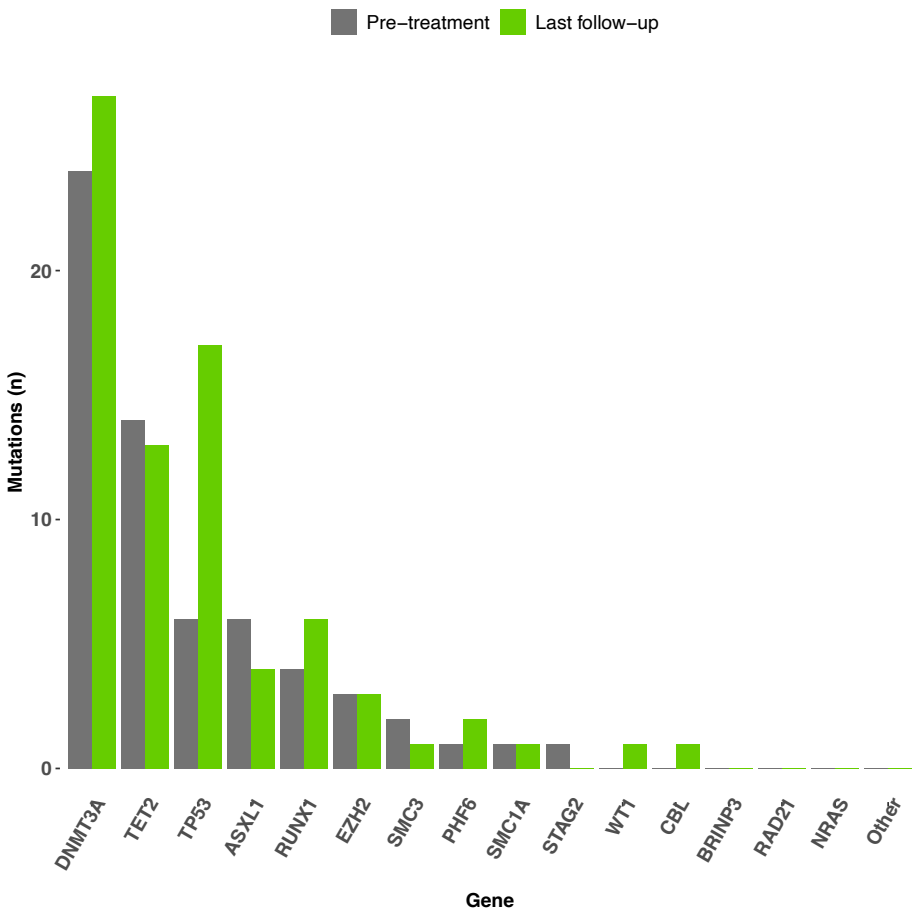
VAF, variant allele fraction.



SUPPLEMENTAL FIG 3. Association between the pre-treatment number of clonal hemopoiesis mutations and participant age at baseline.

Graph plots of the number of pre-treatment mutations per study participant (y-axis) and age at the time of pre-treatment blood draw (x-axis).

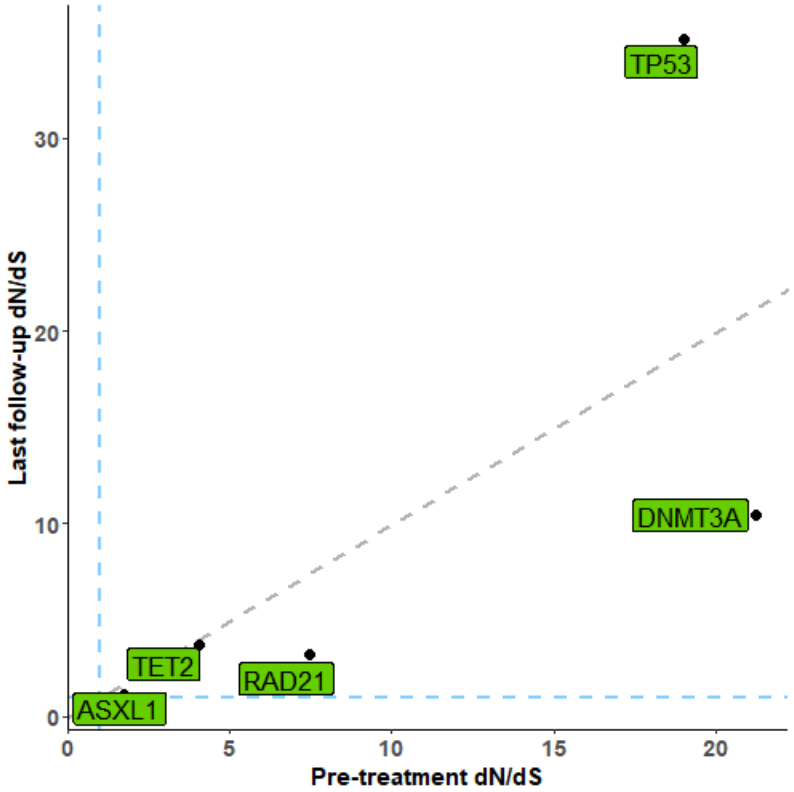
n, number.



SUPPLEMENTAL FIG 4. Clonal hematopoiesis mutations with variant allele fraction ≥ 0.2% detected before treatment and at post-treatment follow-up.

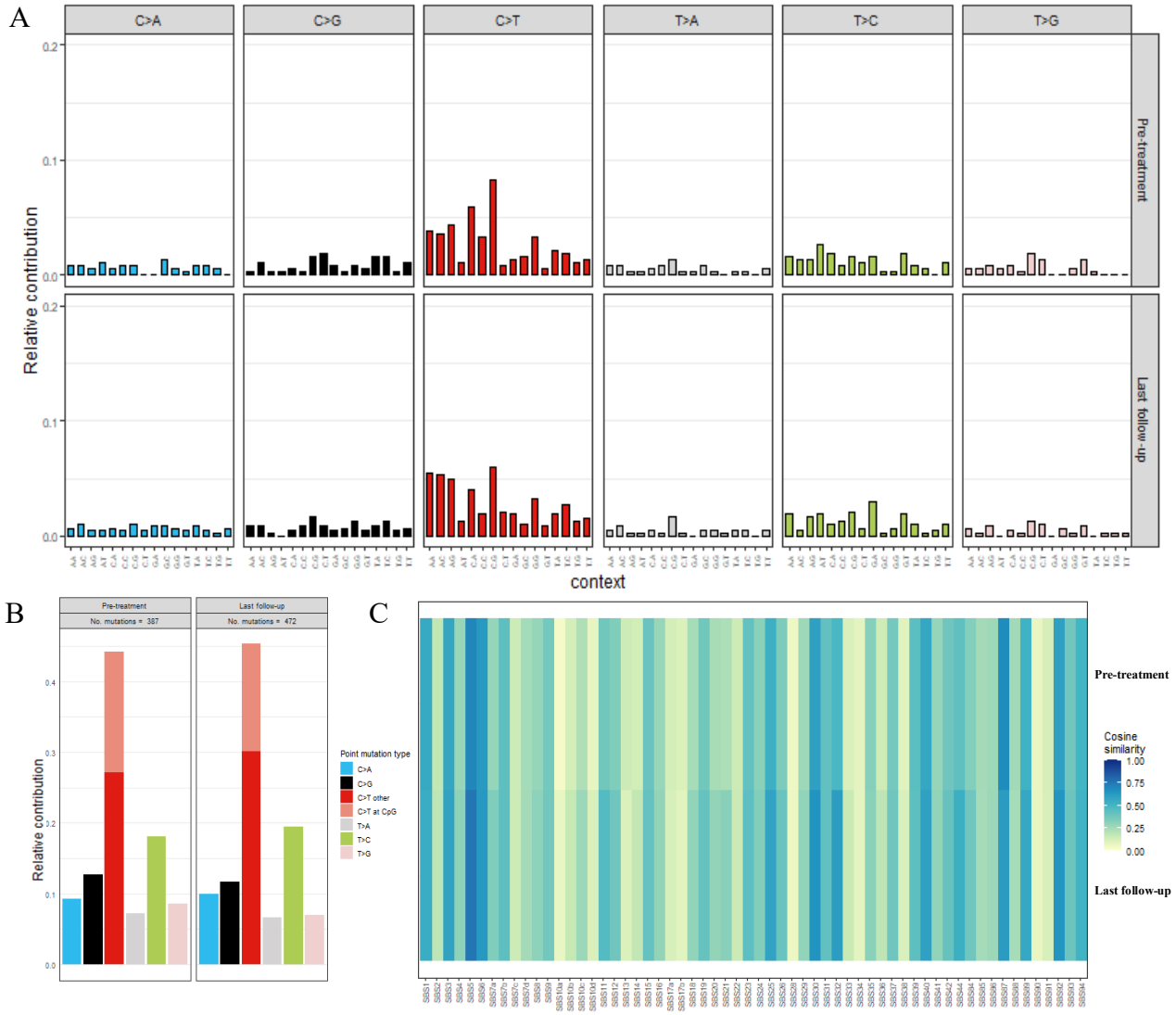
Bar plot of the number of total mutations in each gene detected prior to chemoradiation therapy and at last follow-up, stratified by gene.

n, number.



SUPPLEMENTAL FIG 5. Quantification of pre-treatment to post-treatment selection for missense mutations by gene.

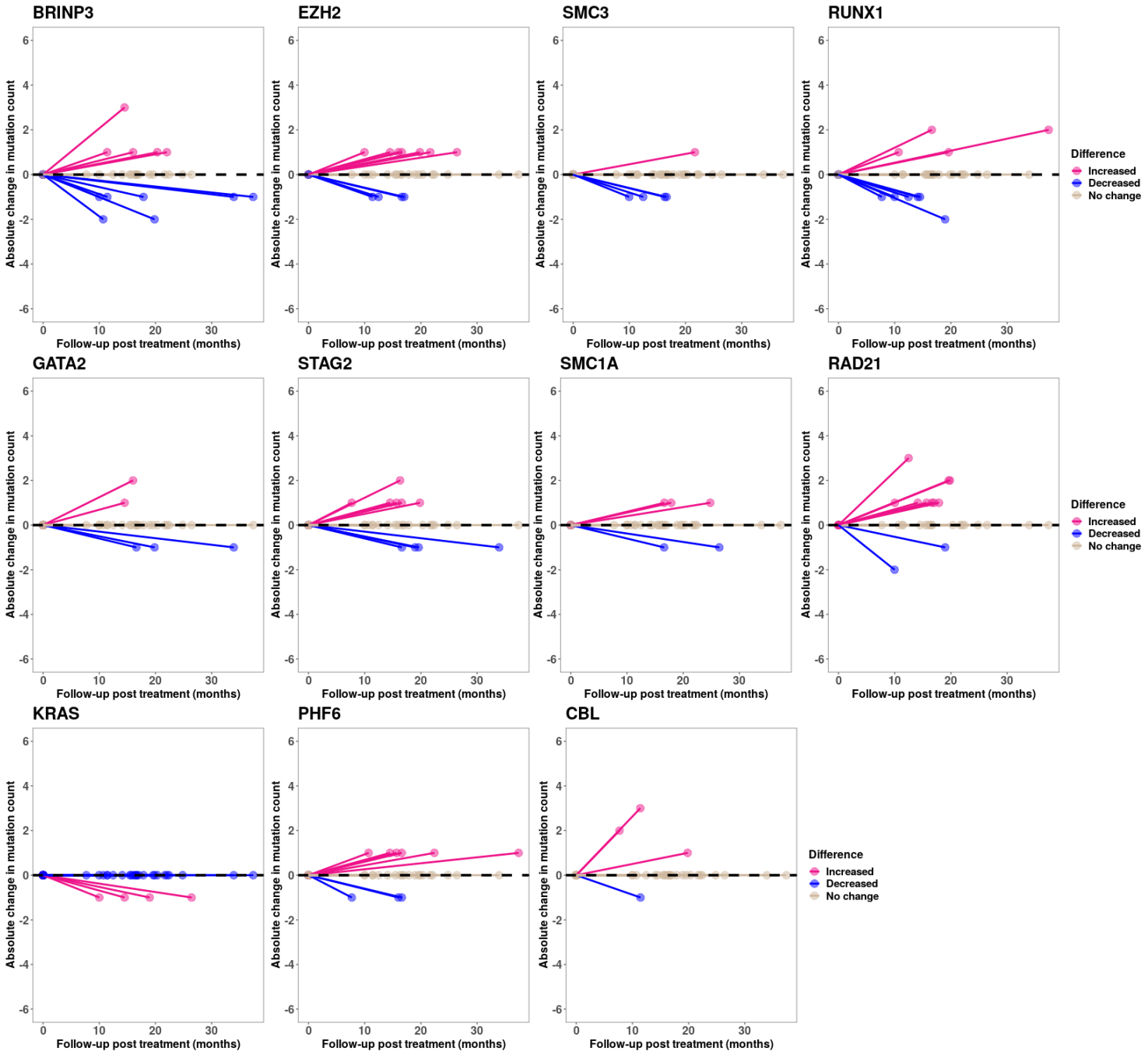
dN/dS represents the ratio of non-synonymous to synonymous substitutions to quantify the strength of the selection. The diagonal dashed line indicates no change in the dN/dS ratio from before treatment to last follow-up. Only genes with a multiple testing–corrected q-value of <0.10, which indicates the probability of selection, are presented.



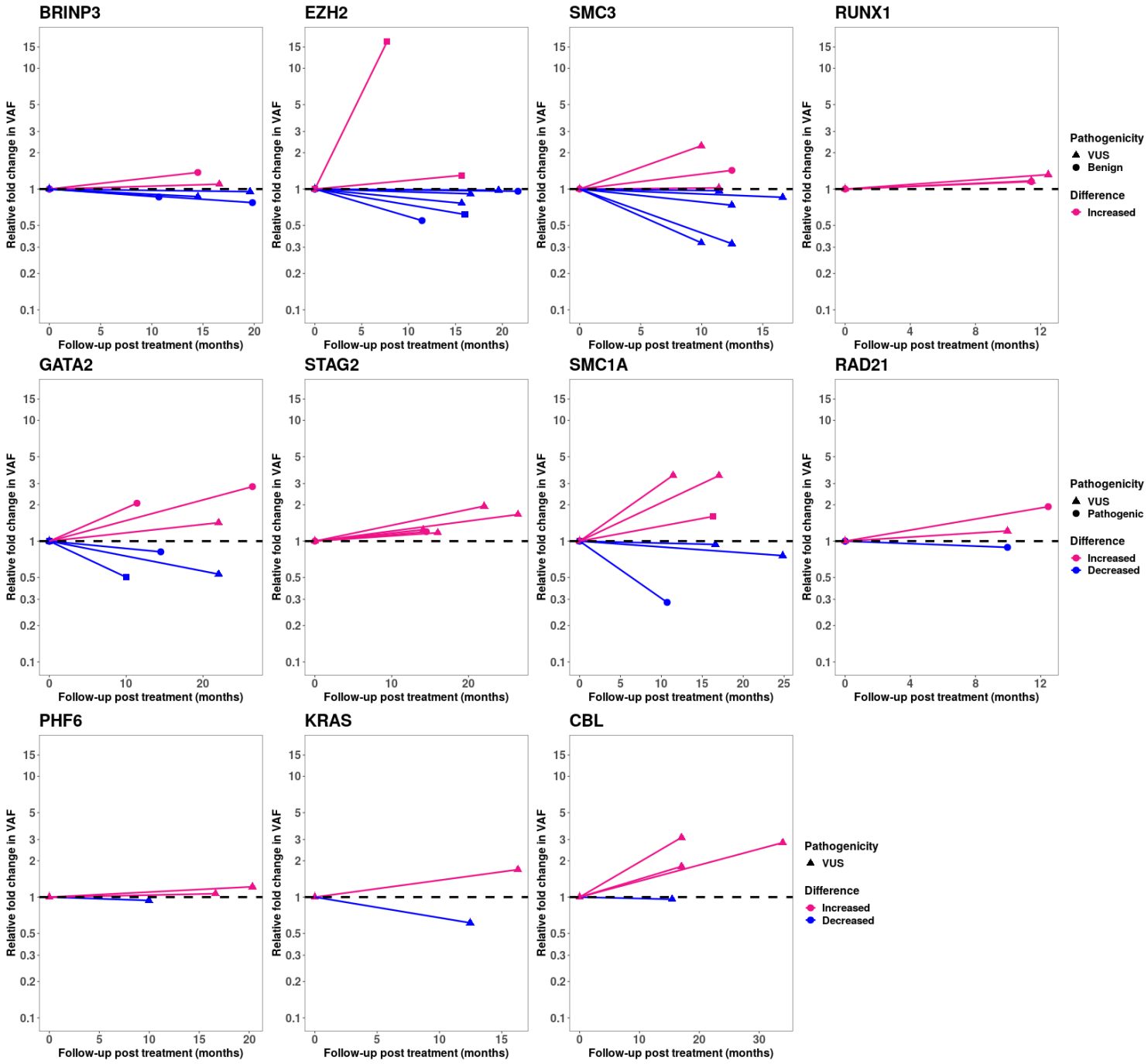
SUPPLEMENTAL FIG 6. Mutational signature analysis before treatment and at last follow-up timepoints.

(A) Mutational signature before treatment and at last follow-up. (B) Point mutation type distribution before treatment and at last follow-up. (C) Mutational signature catalogue analysis before treatment and at last follow-up. X-axis corresponds to the Catalogue Of Somatic Mutations In Cancer (COSMIC) mutational signatures (v3.3, June 2022).

A


BRINP3
EZH2
SMC3
RUNX1
GATA2
STAG2
SMC1A
RAD21
KRAS
PHF6
CBL
Absolute change in mutation count
Follow-up post treatment (months)
Difference
Increased
Decreased
No change

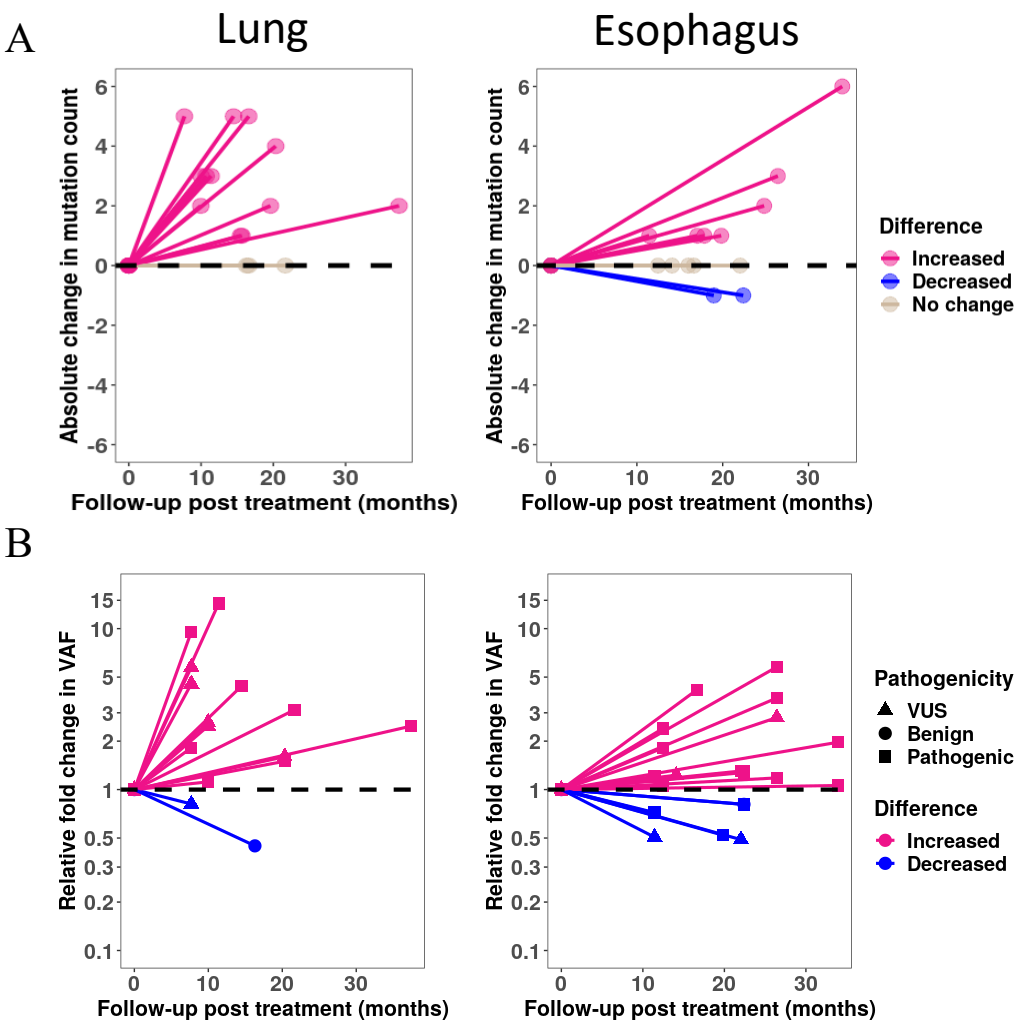
B



SUPPLEMENTAL FIG 7. Changes in variant count (A) and variant allele fraction (B) by gene from pre-treatment to last follow-up in additional genes.

(A) Line plot, where each line represents an individual study participant and shows the absolute change in the number of mutations from before treatment to last follow-up. (B) Line plot where each line represents an individual mutation and the relative fold change in variant allele fraction from before treatment to last follow-up.

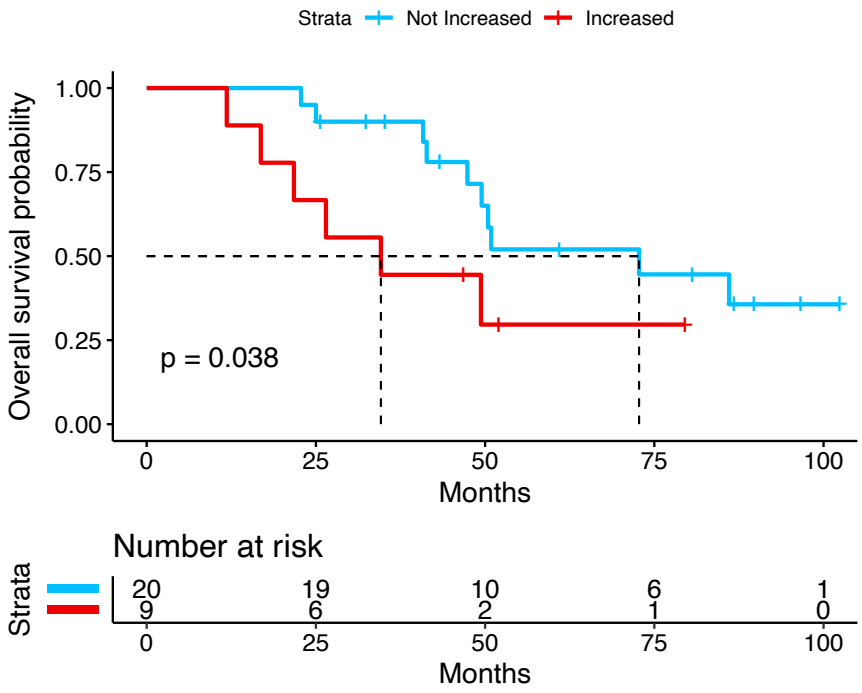
VAF, variant allele fraction; VUS, variant of uncertain significance.



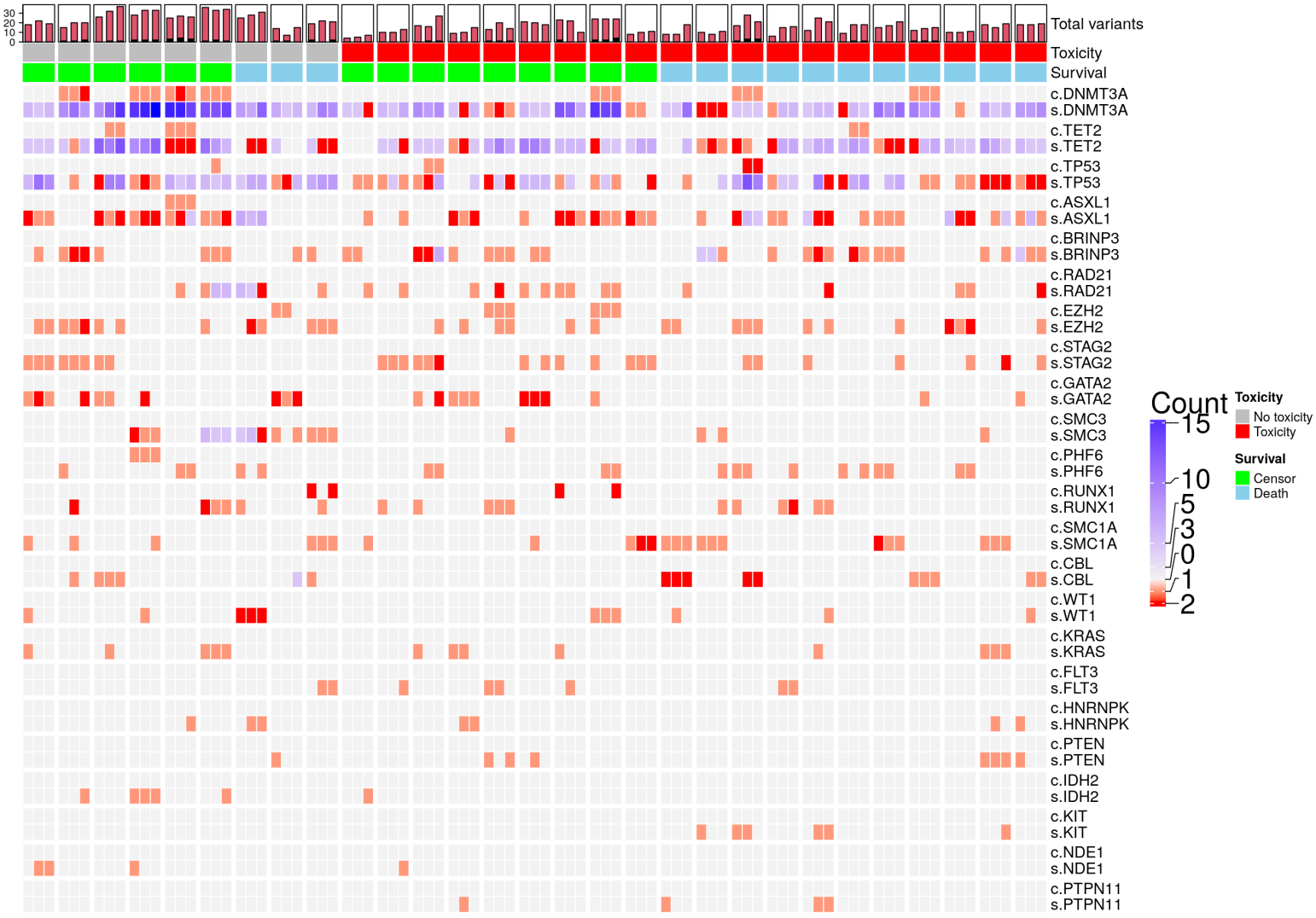
SUPPLEMENTAL FIG 8. Changes in variant count (A) and variant allele fraction (B) in *TP53* from before treatment to last follow-up, stratified by cancer type.

(A) Line plot where each line represents an individual study participant and shows the absolute change in the number of mutations from before treatment to last follow-up. (B) Line plot where each line represents an individual mutation and the relative fold change in variant allele fraction from before treatment to last follow-up.

VAF, variant allele fraction; VUS, variant of uncertain significance.



SUPPLEMENTAL FIG 9. Overall survival curves according to the absolute change in the number of *TP53* clonal hematopoiesis mutations with variant allele fraction ≥ 0.2% from before treatment to last follow-up.



SUPPLEMENTAL FIG 10. Participant-level summary of clinical outcomes and detected variants by gene and clone size.

Columns represent individual study participants with three sub-columns indicating pre-treatment and two post-treatment follow-up timepoints. c.Gene and s.Gene indicate clonal mutations (variant allele fractions ≥1%) and subclonal mutations (variant allele fraction <1%), respectively.

REFERENCES

1. Schmitt MW, Kennedy SR, Salk JJ, et al: Detection of ultra-rare mutations by next-generation sequencing. Proc Natl Acad Sci U S A 109:14508-13, 2012

2. Salk JJ, Schmitt MW, Loeb LA: Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. Nat Rev Genet 19:269-285, 2018

3. Zhivagui M, Hoda A, Valenzuela N, et al: DNA damage and somatic mutations in mammalian cells after irradiation with a nail polish dryer. Nat Commun 14:276, 2023

4. Cibulskis K, Lawrence MS, Carter SL, et al: Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 31:213-9, 2013

5. Lai Z, Markovets A, Ahdesmaki M, et al: VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res 44:e108, 2016

6. Wilm A, Aw PP, Bertrand D, et al: LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res 40:11189-201, 2012

7. Koboldt DC, Larson DE, Wilson RK: Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. Curr Protoc Bioinformatics 44:15 4 1-17, 2013

8. Li Q, Wang K: InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am J Hum Genet 100:267-280, 2017

9. Novetsky Friedman D, Chan ICC, Moskowitz CS, et al: Clonal hematopoiesis in survivors of childhood cancer. Blood Adv 7:4102-4106, 2023

10. Vlasschaert C, Mack T, Heimlich JB, et al: A practical approach to curate clonal hematopoiesis of indeterminate potential in human genetic data sets. Blood 141:2214-2223, 2023

11. Ye K, Schulz MH, Long Q, et al: Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25:2865-71, 2009