Supplemental Appendix

Supplement to: Tanaka T, Morita K, Takahashi K, et al. Clonal Dynamics and clinical implications of Post-Remission Clonal Hematopoiesis in Acute Myeloid Leukemia (AML)

Contents

Supplemental Method

Patients and treatment protocol

We retrospectively studied 176 AML patients who attained morphological CR after either intensive induction or low-intensity chemotherapy (Supplemental Table S1). Among those, we studied in patients with previously untreated 164 AML who attained morphological CR at day 30 after intensive chemotherapy (n=126), including idarubicin plus cytarabine based regimen, with clofarabine (n=60, ClinicalTrials.gov Identifier:NCT01025154), with cladribine (n=19, ClinicalTrials.gov Identifier:NCT02115295), with fludarabine (n=36, ClinicalTrials.gov Identifier:NCT01289457), with cladribine and sorafenib (n=9), with SGI-110 (n=1) (SGI-110 plus Idarubicin per protocol 2013-0843), and CPX-351 (n=1) , or low-intensity chemotherapy, including clofarabine plus low dose cytarabine (n=23, ClinicalTrials.gov Identifier:NCT00778375), cladribine plus low dose cytarabine (n=13, ClinicalTrials.gov Identifier: NCT01515527), decitabine (n=1), Azacitidine plus venetoclax (n=1).

Targeted capture DNA sequencing of 295 genes and detection of high-confidence somatic mutations

Genomic DNA was extracted from diagnostic and post-remission bone marrow aspirate samples by using an Autopure extractor (QIAGEN/Gentra, Valencia, CA). DNAs were fragmented and bait-captured in solution, as previously described according to the manufacturer's protocols[1]. Captured DNA libraries were then sequenced by using a HiSeq 2000 sequencer (Illumina, San Diego, CA) with 76 base-pair (bp) paired-end reads. Raw sequencing data from the Illumina platform were converted to a fastq format and aligned to the reference genome (hg19) using the Burroughs-Wheeler Aligner (BWA)[2]. The aligned BAM files were subjected to mark duplication, realignment, and recalibration using Picard and GATK (https://www.broadinstitute.org/gatk/guide/best-practices?bpm=DNAseq). Preprocessed BAM files were then analyzed to detect single nucleotide variants (SNVs) and small insertions and deletions (indels) using MuTect[3] and Pindel[4] algorithms, respectively, against normal virtual sequences developed in-house. We modified an approach described by Papaemmanuil et al. to identify high-confidence driver mutations in the bone marrow samples without matched germline control[5]. First, variants with low-quality supporting sequencing data were filtered out. Specifically, variants matching one or more of the following criteria were considered of low quality and therefore filtered out from further analysis: 1) tumor coverage < 15x and 20x for SNVs and indels, respectively, 2) tumor allele frequency < 5%, and 3) normal allele frequency ≥1% and 0% for SNVs and indels, respectively. Second, only variants that would introduce an obvious protein-coding change were kept for further analysis. Specifically, variants with an ANNOVAR annotation of nonsynonymous, stop-gain, stop-loss, splicing, frameshift insertion, frameshift deletion, nonframeshift insertion, or non-frameshift deletion were considered able to introduce an apparent protein-coding change and were therefore kept for further analysis. Third, common polymorphisms were removed to reduce a load of possible germline contamination due to the absence of matched normal control. Specifically, a series of public variant databases, including the 1000 Genome Database (http://www.1000genomes.org/), ESP6500 Database (http://evs.gs.washington.edu/EVS/), dbSNP ver.132 (http://www.ncbi.nlm.nih.gov/SNP/),

and Exome Aggregation Consortium database (http://exac.broadinstitute.org/), were used. Variants with a population frequency of 0.14% or more in any of the databases were considered possible germline polymorphisms and were therefore removed from further analysis. Finally, a hierarchical classification system was developed to assign a confidence level for each remaining variant to facilitate the identification of putative somatic driver mutations. Specifically, each variant was classified on the basis of the following hierarchical order and was assigned a confidence level corresponding to its rank in the system: 1) Confirmed somatic mutation based on the COSMIC database (version 81), 2) loss-of-function mutation such as splicing, stop gain, stop-loss, and a frameshift mutation in well-characterized tumor suppressor genes, 3) recurrent variant that resides within three amino acids away from a confirmed somatic mutation according to the COSMIC database (version 81), 4) variant that resides within three amino acids away from a confirmed somatic mutation according to the COSMIC database (version 81), 5) variant that was predicted to be damaging by in silico function prediction algorithms, and 6) variant with unknown significance. The first two groups were considered a high-confidence tier, while the remaining were considered low-confidence. The final annotated variant list was passed out of the pipeline and was further analyzed by manual inspection and literature mining in order to identify high-confidence somatic mutations.
Based on our previous analysis, the sequencing platform had an estimate of sequencing sensitivity of 1%[6] .

Statistical analysis

Categorical variables were compared with the Fisher exact test. Continuous variables were compared by Student t-test. A Kaplan-Meier plot was used to visualize survival distributions and differences were assessed by a log-rank test. Gray's method was used for the cumulative incidence of relapse analysis. R (version 3.1.4) and EZR[7] were used for statistical analysis.

Supplemental Figure and Legend

Figure S1
Box plot showing the number of baseline gene mutations per post-CR CH status.

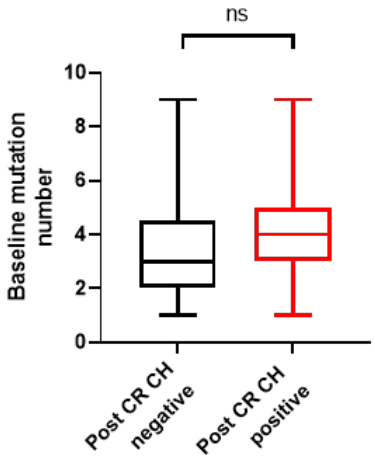


Figure S2
Mutation distribution of post-CR CH compared to CHIP in previous normal population analysis (Jaiswal et al. NEJM 2014).

Bar graph showing the frequency of mutations per a total of genes detected as CHIP in a healthy population (blue bar)[8] or as post-CR CH in our cohort (red bar). * P < 0.05.

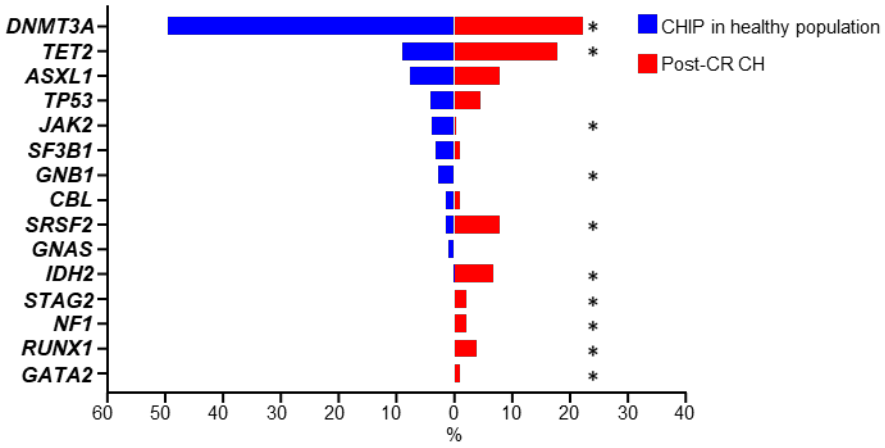
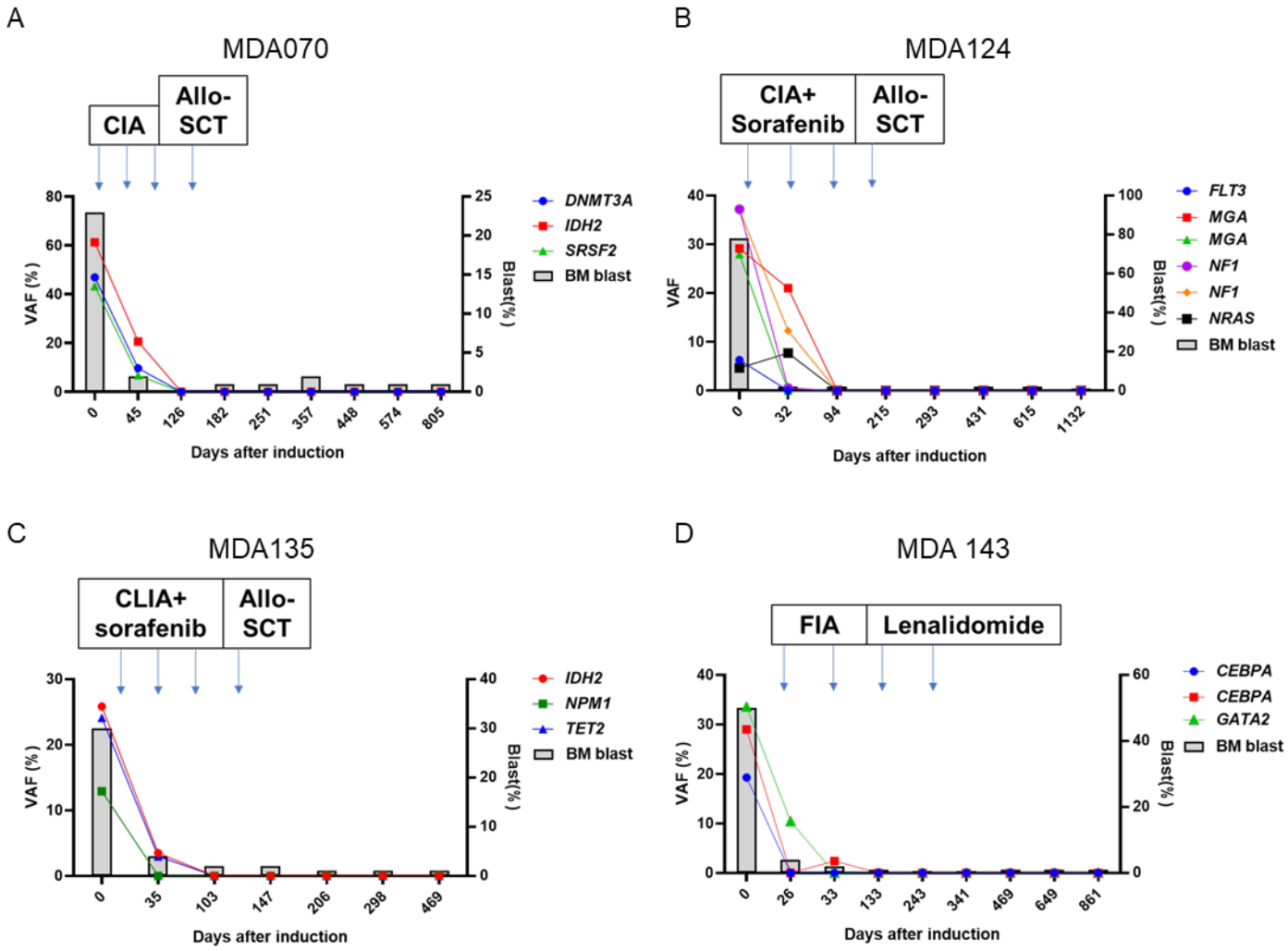


Figure S3
Representative cases (A; MDA070, B; MDA124, C; MDA135, D; MDA 143) whose post-CR CH are cleared following consolidation therapy before allo-SCT.



Abbreviation: VAF; variant allele frequency, CIA; clofarabine, idarubicin, and cytarabine, CLIA; cladribine, idarubicin, and cytarabine, FIA; fludarabine, idarubicin, and cytarabine, Allo-SCT; allogeneic stem cell transplant

Figure S4
Variant allele frequency (VAF) fluctuation during the consolidation or maintenance therapies per post-CR CH genes.

(**A**) *ASXL1*, (**B**) *DNMT3A*, (**C**) *IDH2*, (**D**) *RUNX1*, (**E**) *SRSF2*, (**F**) *TET2*, (**G**) *TP53*, (**H**) *STAG2*

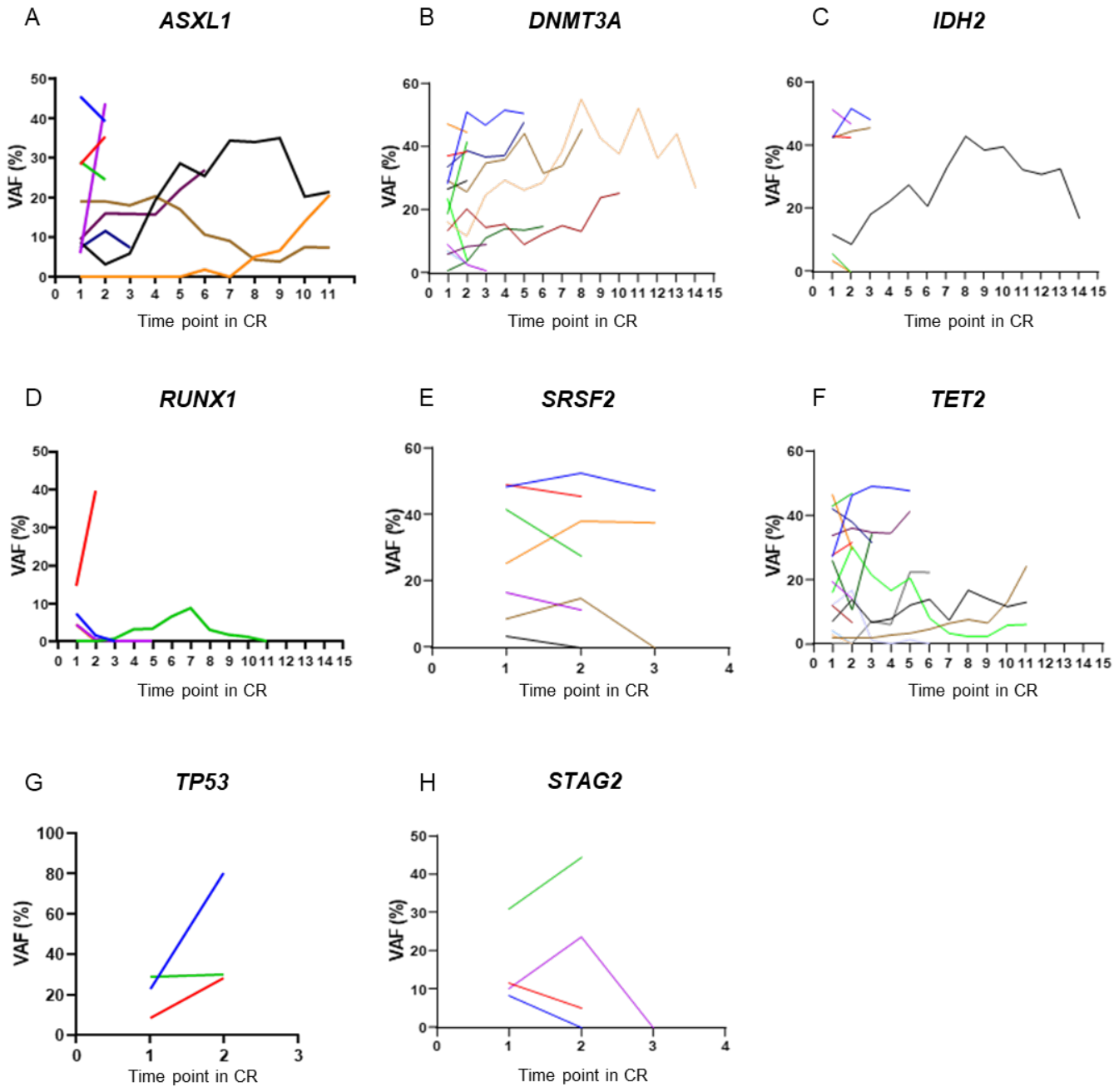
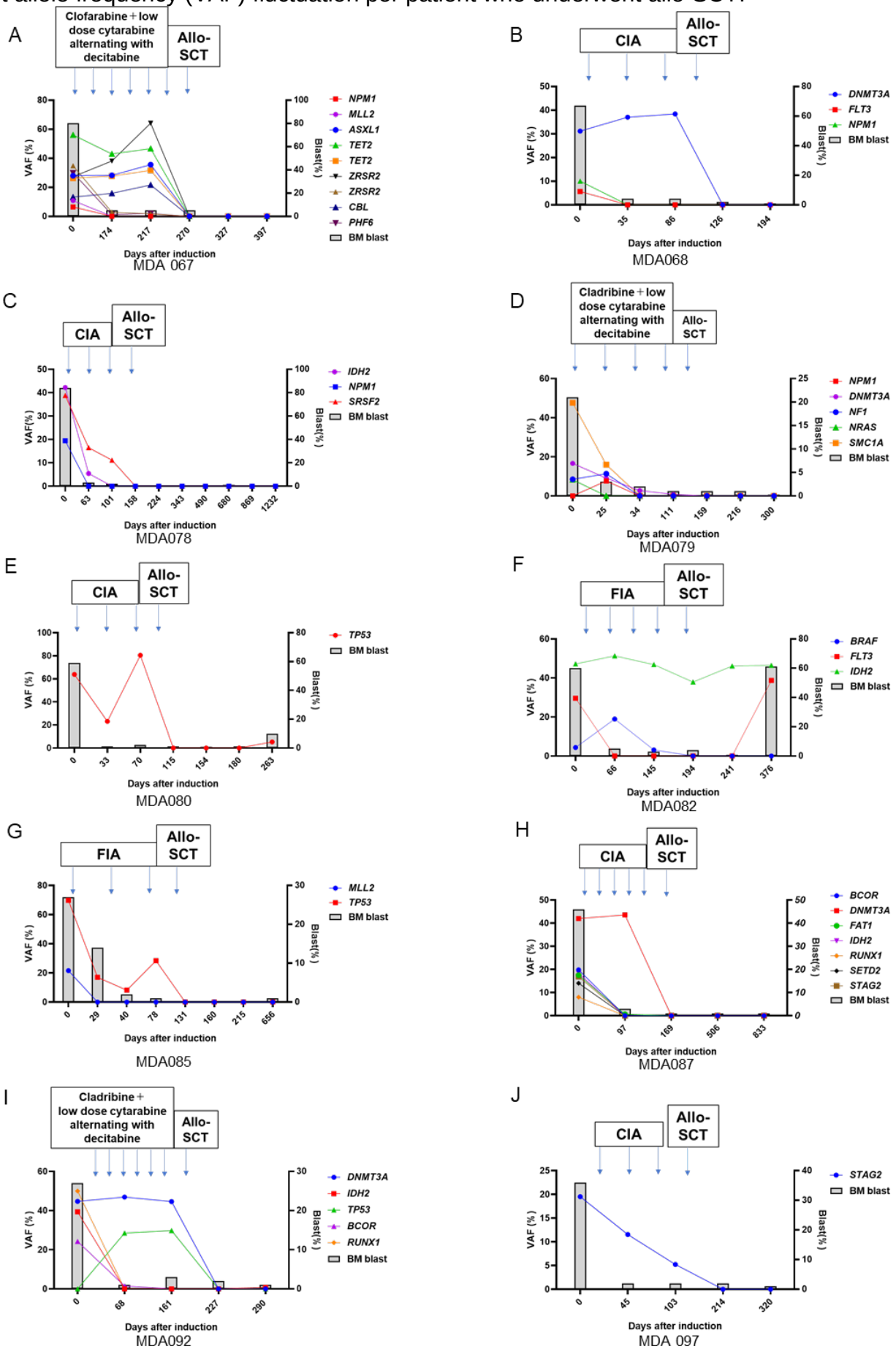
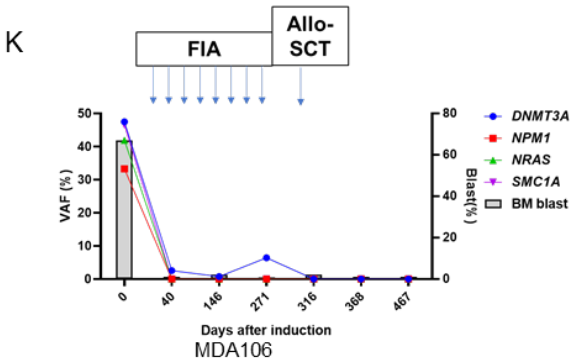
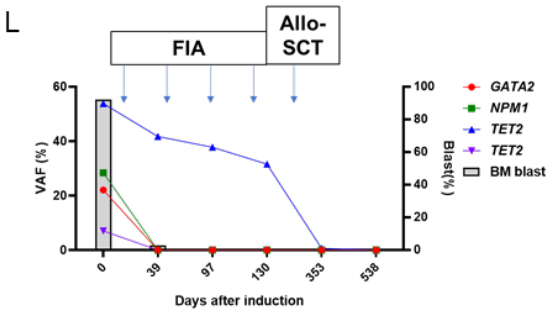


Figure S5
Variant allele frequency (VAF) fluctuation per patient who underwent allo-SCT.

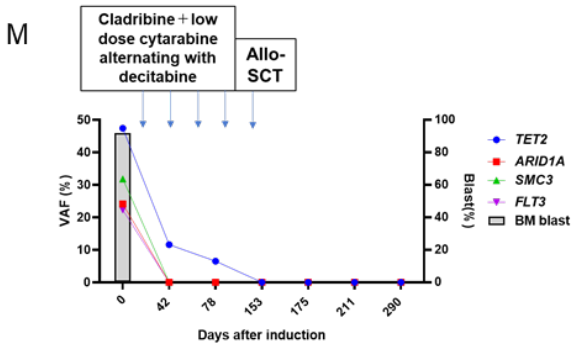




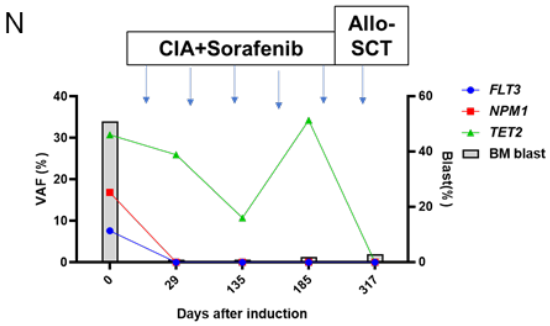
MDA106



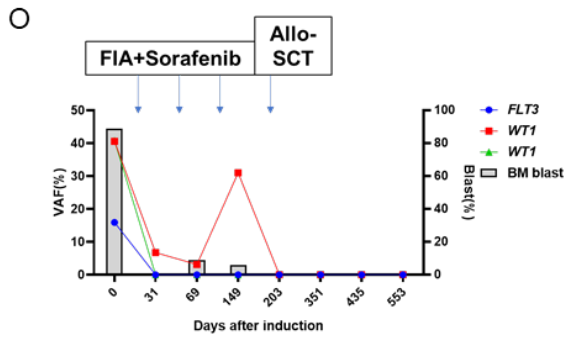
MDA108



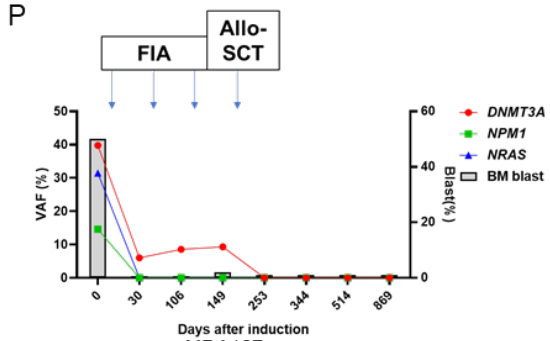
MDA113



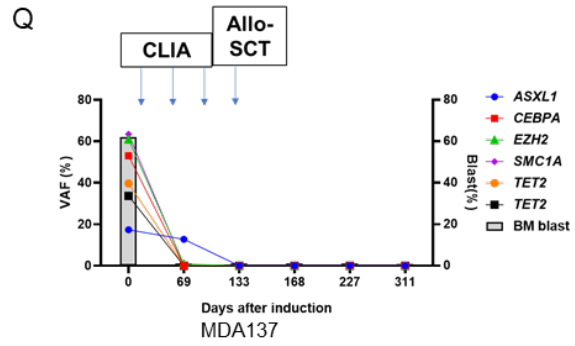
MDA114



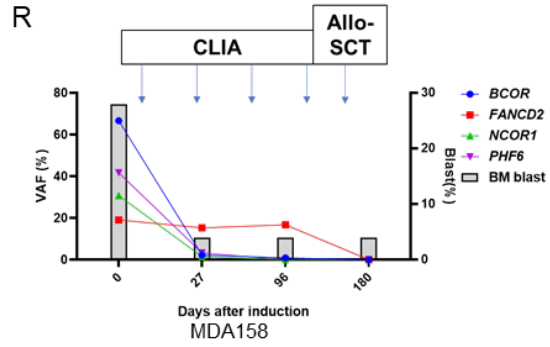
MDA126



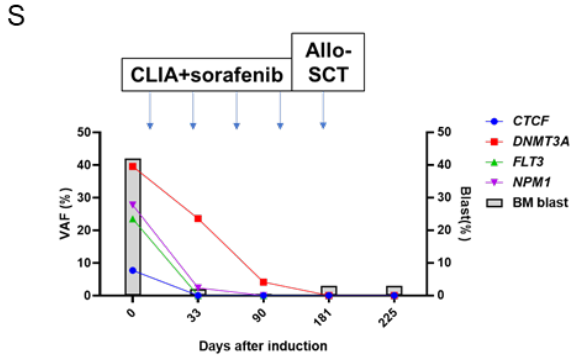
MDA127



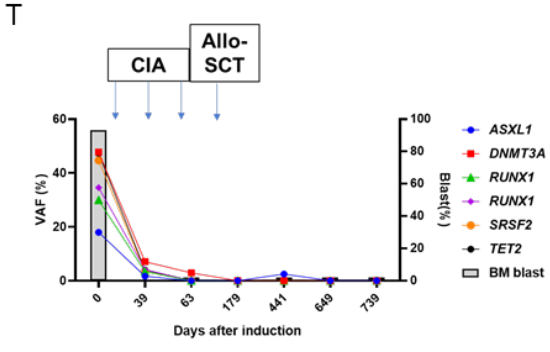
MDA137



MDA158

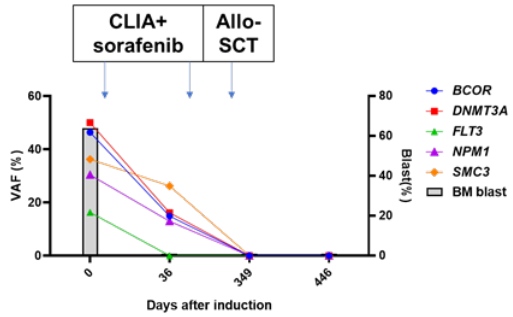


MDA161



MDA166

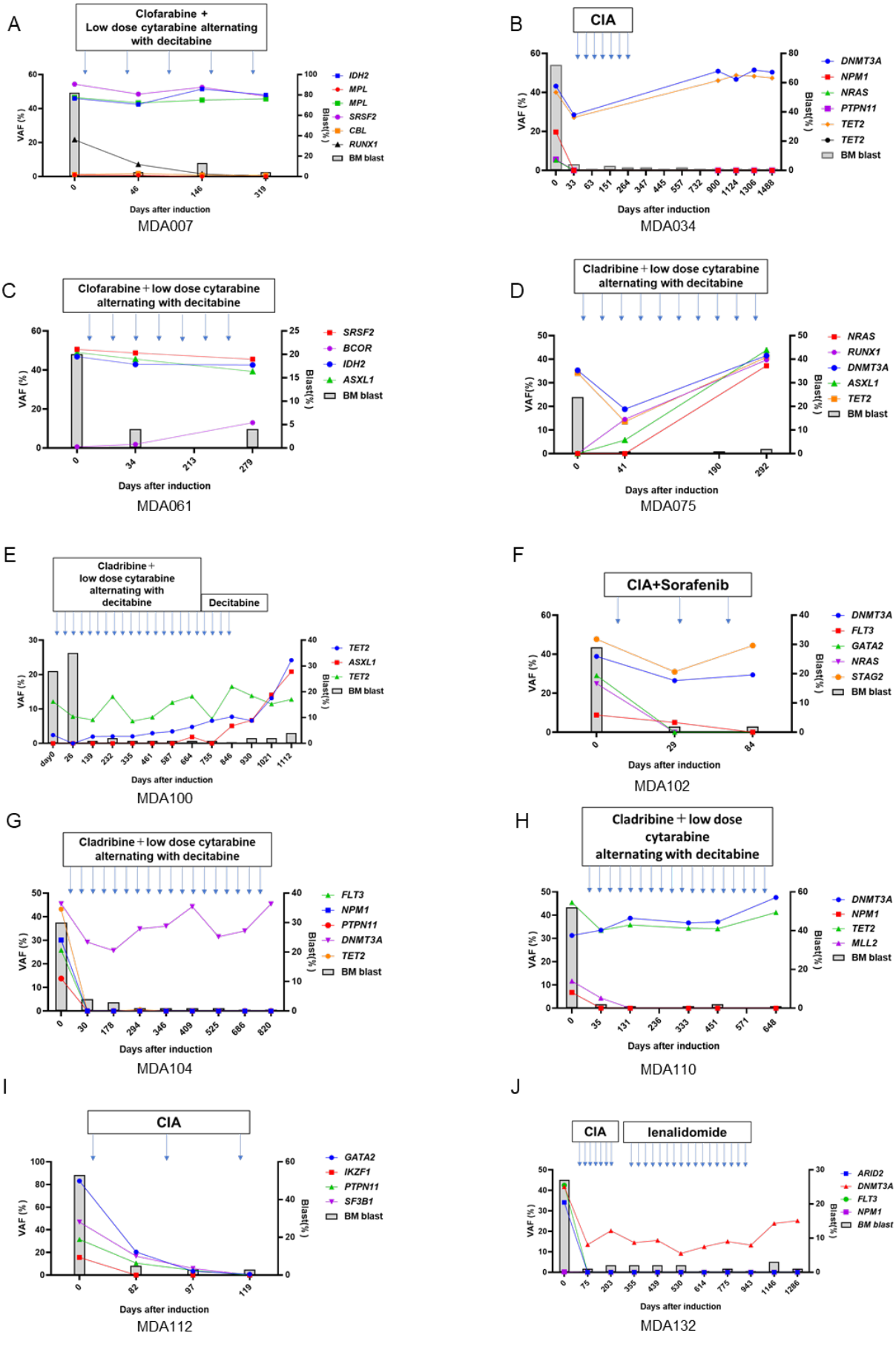
U

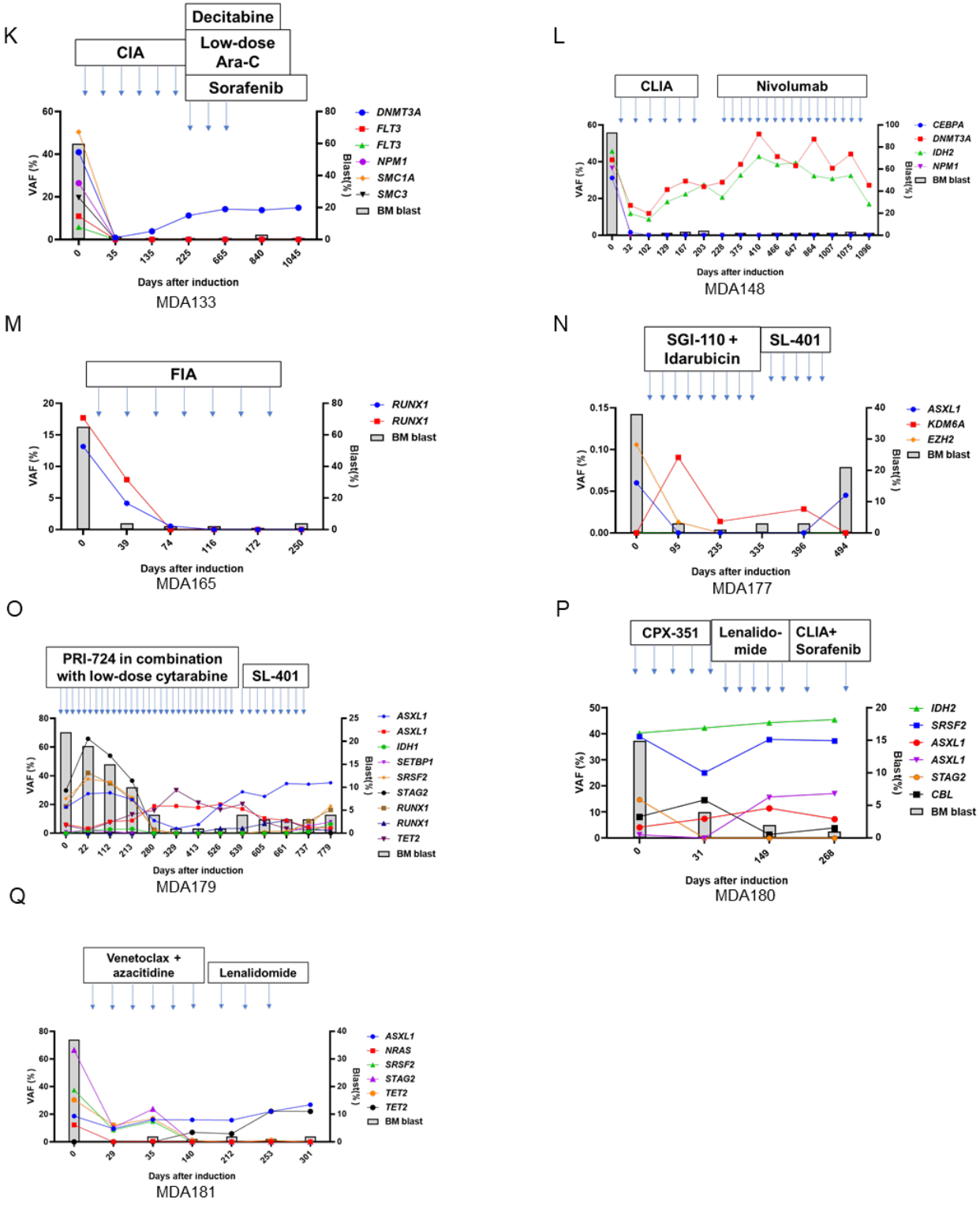


MDA169

Abbreviation: CIA; clofarabine, idarubicin, and cytarabine, CLIA; cladribine, idarubicin, and cytarabine, FIA; fludarabine, idarubicin, and cytarabine, Allo-SCT; allogeneic stem cell transplant.

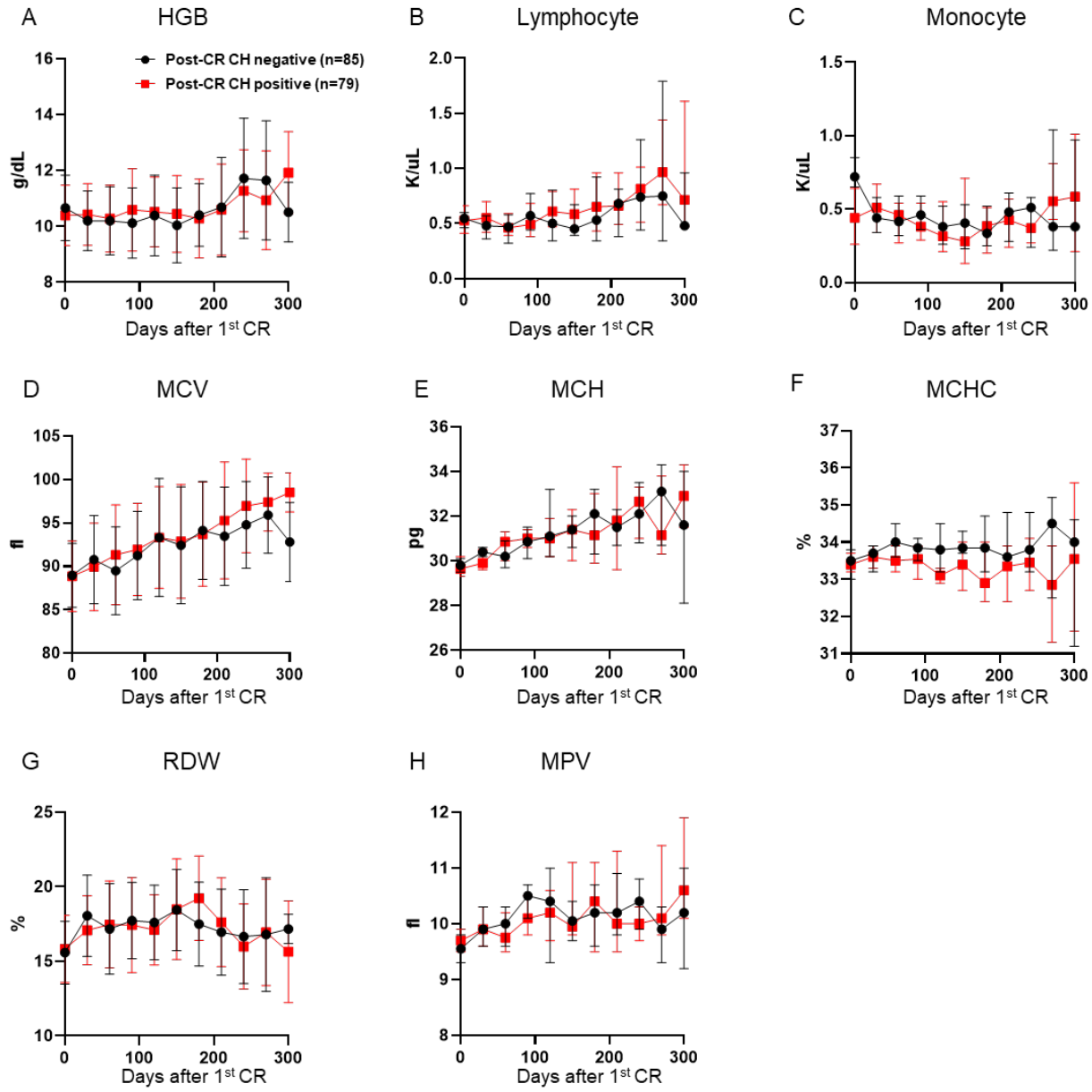
Figure S6
Variant allele frequency (VAF) fluctuation per patient who did not undergo allo-SCT or post-CR CH was not cleared by chemotherapy.





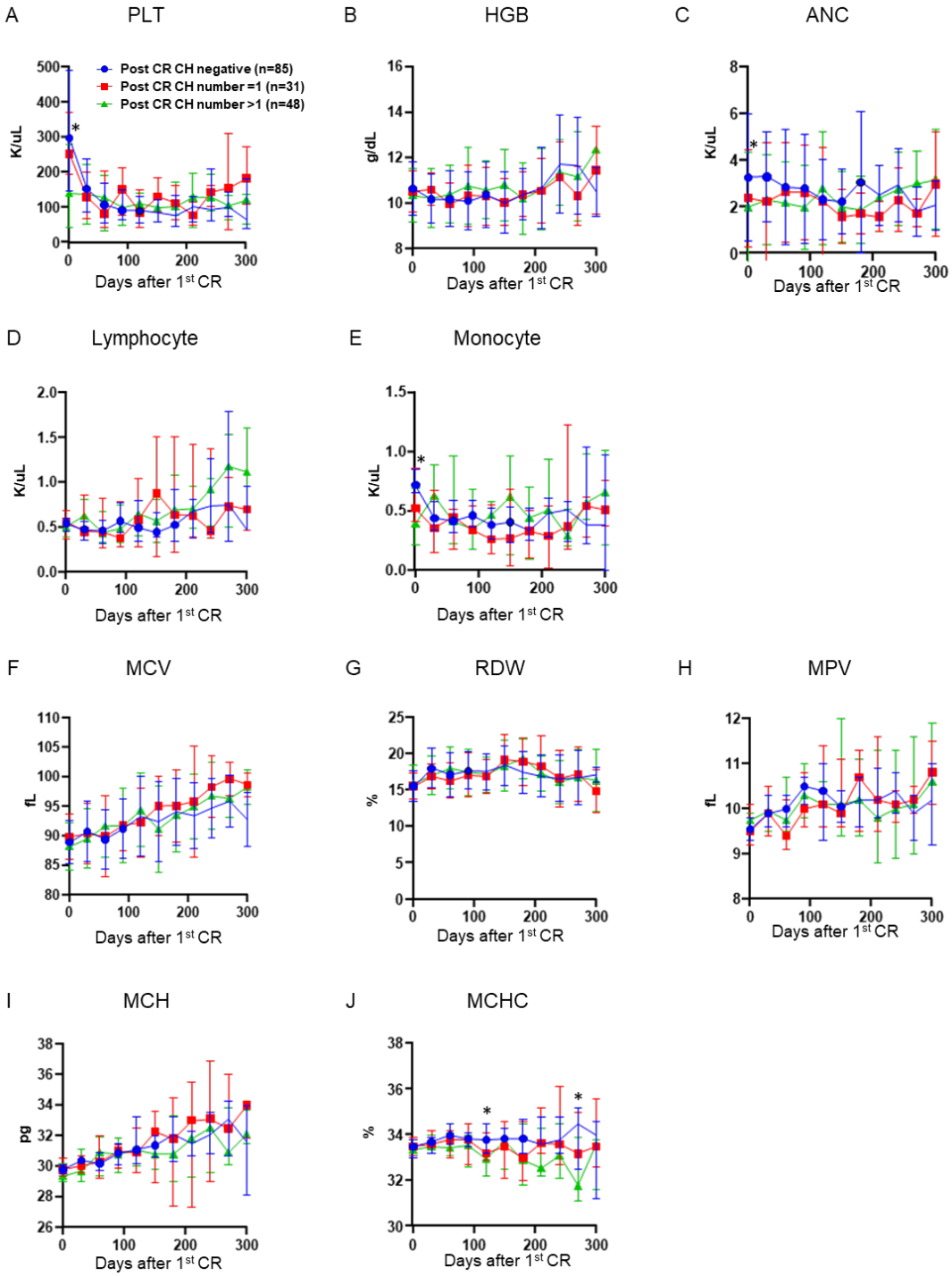
Abbreviation: CIA; clofarabine, idarubicin, and cytarabine, CLIA; cladribine, idarubicin, and cytarabine, FIA; fludarabine, idarubicin, and cytarabine, Ara-C; cytarabine.

Figure S7
The long-term trend of peripheral blood counts in patients with post-CR CH (red line) without post-CR CH (black line).



Peripheral blood (PB) cell counts were checked every about 30 days at the recovery phase. PB data were censored when the patients had a relapse or allogeneic stem cell transplantation. (**A**) HGB (**B**) Lymphocyte, (**C**) Monocyte, (**D**) MCV, (**E**) MCH, (**F**) MCHC, (**G**) RDW, (**H**) MPV. * $P < 0.05$.
Abbreviation: HGB; hemoglobin, MCV; mean cell volume, RDW; red cell distribution width, MPV; mean platelet volume, MCH; mean corpuscular hemoglobin, MCHC; mean corpuscular hemoglobin concentration.

Figure S8
The long-term trend of peripheral blood counts in patients based on the number of mutations in post-CR CH (0: blue line, 1: red line, 1>: green line).

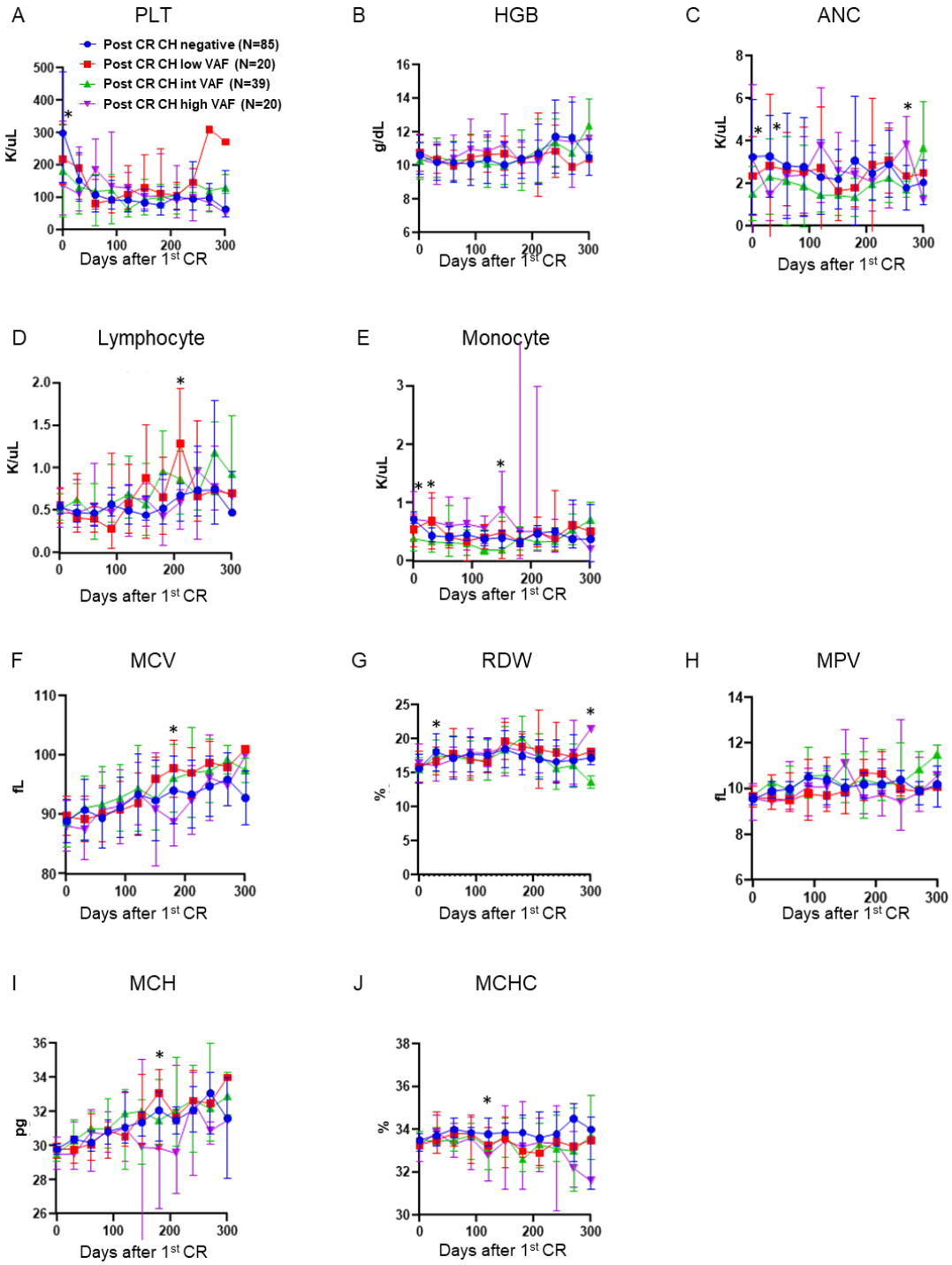


Peripheral blood (PB) cell counts were checked every about 30 days at the recovery phase. PB data were censored when the patients had a relapse or allogeneic stem cell transplantation. (**A**) PLT, (**B**) HGB, (**C**) ANC, (**D**) lymphocyte, (**E**) Monocyte, (**F**) MCV, (**G**) RDW, (**H**) MPV, (**I**) MCH, (**J**) MCHC. * $P < 0.05$.

Abbreviation: PLT; platelets, HGB; hemoglobin, ANC; absolute neutrophil counts, MCV; mean cell volume, RDW; red cell distribution width, MPV; mean platelet volume, MCH; mean corpuscular hemoglobin, MCHC; mean corpuscular hemoglobin concentration.

Figure S9

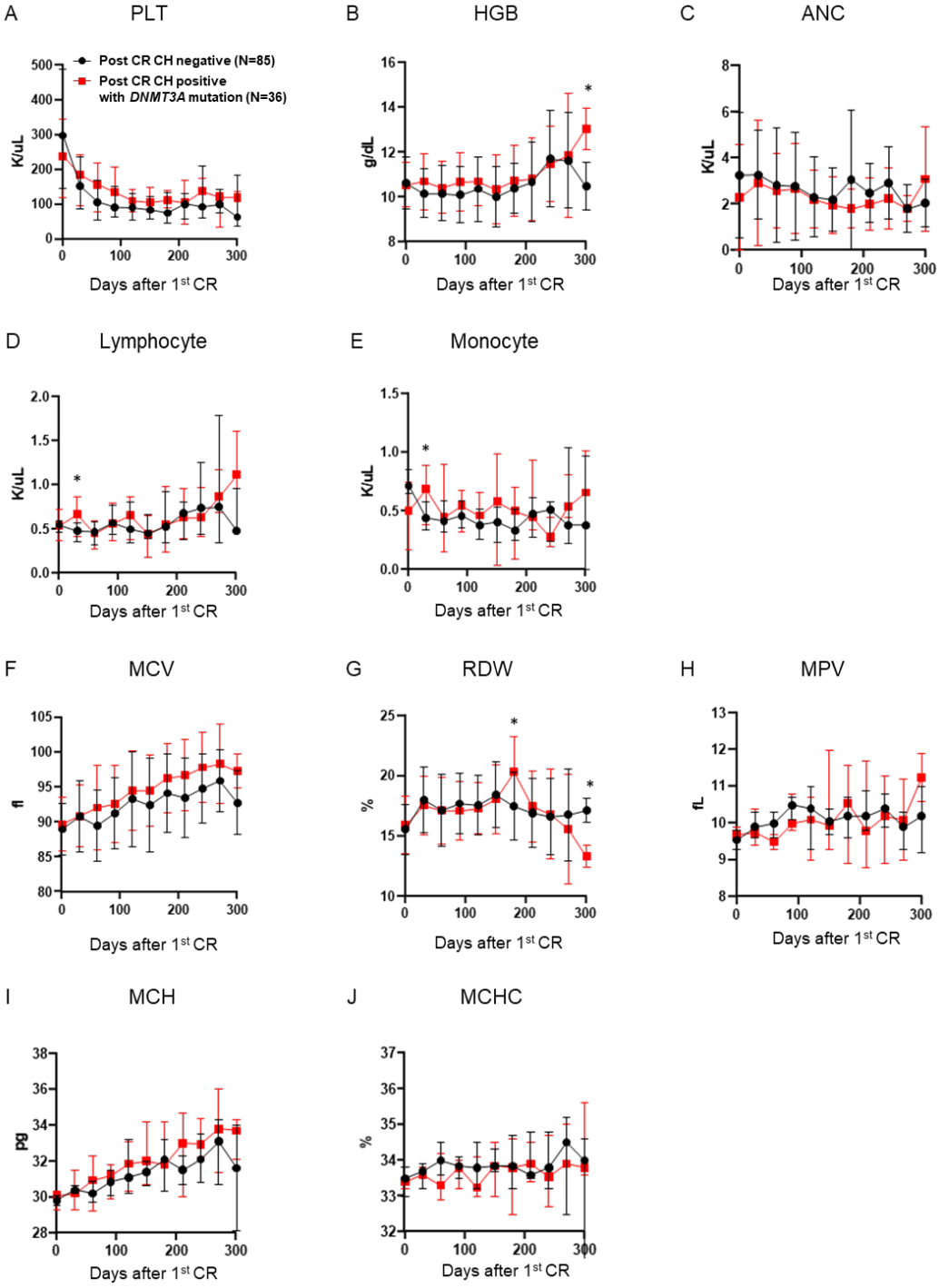
Comparison of the long-term trend of peripheral blood counts stratified by the variant allele frequency (VAF) of post-CR CH (negative: blue line, low VAF (1st quartile): red line, intermediate VAF (2nd and 3rd quartile): green line, high VAF (4th quartile): purple line).



Peripheral blood (PB) cell counts were checked every about 30 days at the recovery phase. PB data were censored when the patients had a relapse or allogeneic stem cell transplantation. (**A**) PLT, (**B**) HGB, (**C**) ANC, (**D**) lymphocyte, (**E**) Monocyte, (**F**) MCV, (**G**) RDW, (**H**) MPV, (**I**) MCH, (**J**) MCHC. * $P < 0.05$.

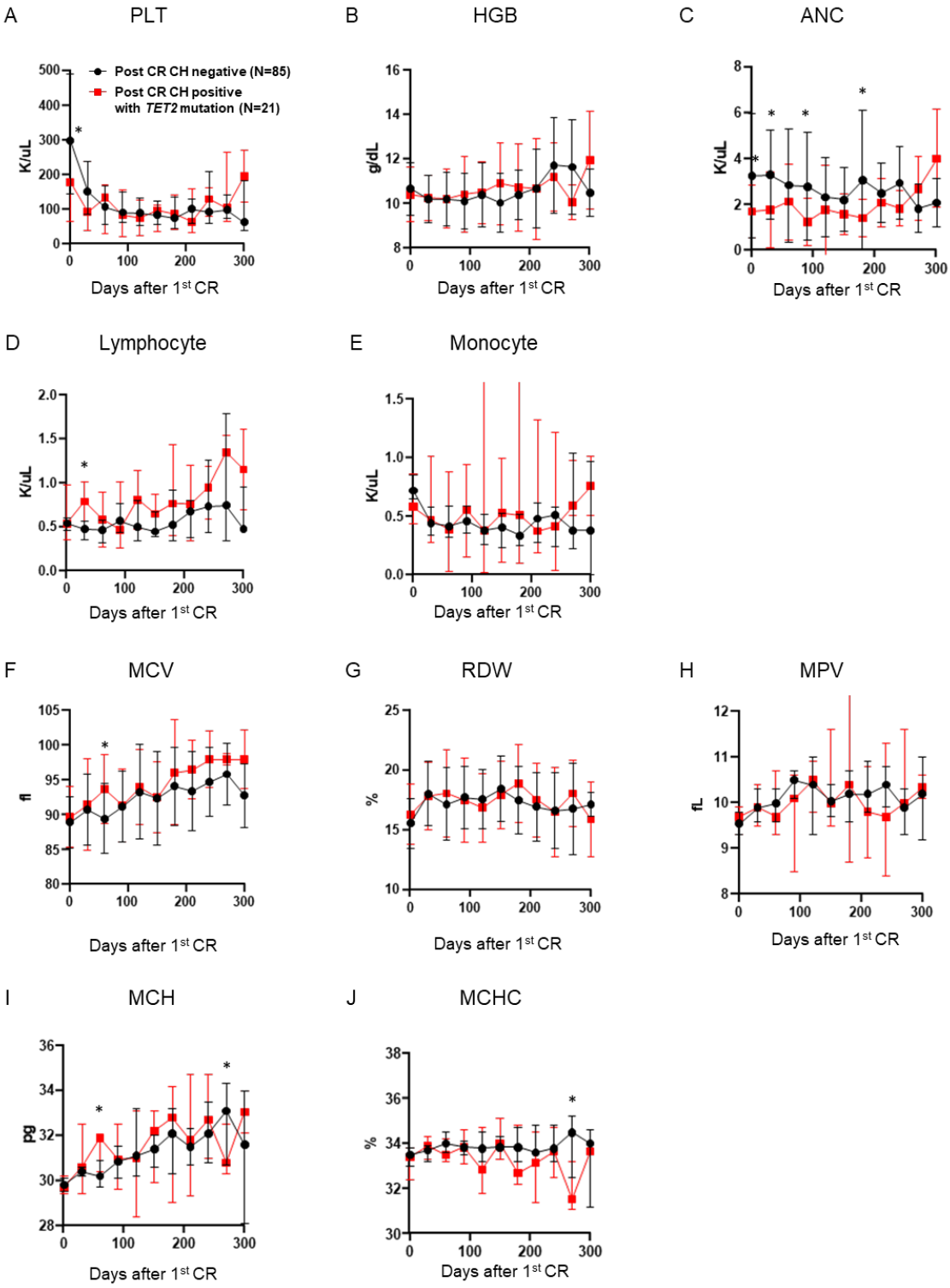
Abbreviation: PLT; platelets, HGB; hemoglobin, ANC; absolute neutrophil counts, MCV; mean cell volume, RDW; red cell distribution width, MPV; mean platelet volume, MCH; mean corpuscular hemoglobin, MCHC; mean corpuscular hemoglobin concentration.

Figure S10
The long-term trend of peripheral blood counts in patients with post-CR CH with *DNMT3A* mutation (Post-CR CH negative: black line, Post-CR CH positive with *DNMT3A* mutation: red line).



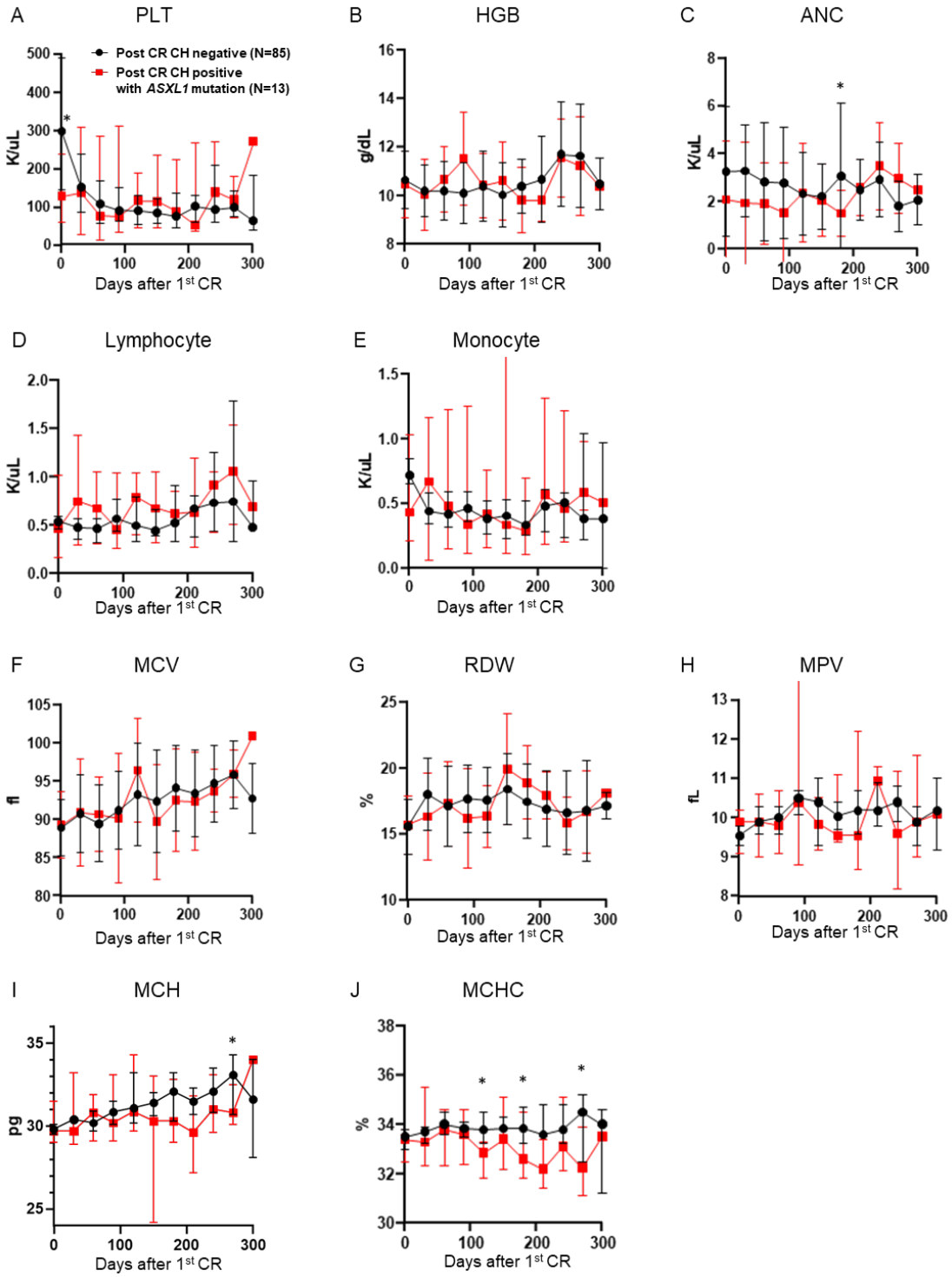
Peripheral blood (PB) cell counts were checked every about 30 days at the recovery phase. PB data were censored when the patients had a relapse or allogeneic stem cell transplantation. (**A**) PLT, (**B**) HGB, (**C**) ANC, (**D**) lymphocyte, (**E**) Monocyte, (**F**) MCV, (**G**) RDW, (**H**) MPV, (**I**) MCH, (**J**) MCHC. * $P < 0.05$.
Abbreviation: PLT; platelets, HGB; hemoglobin, ANC; absolute neutrophil counts, MCV; mean cell volume, RDW; red cell distribution width, MPV; mean platelet volume, MCH; mean corpuscular hemoglobin, MCHC; mean corpuscular hemoglobin concentration.

Figure S11
The long-term trend of peripheral blood counts in patients with post-CR CH with *TET2* mutation (Post-CR CH negative: black line, Post-CR CH positive with *TET2* mutation: red line).



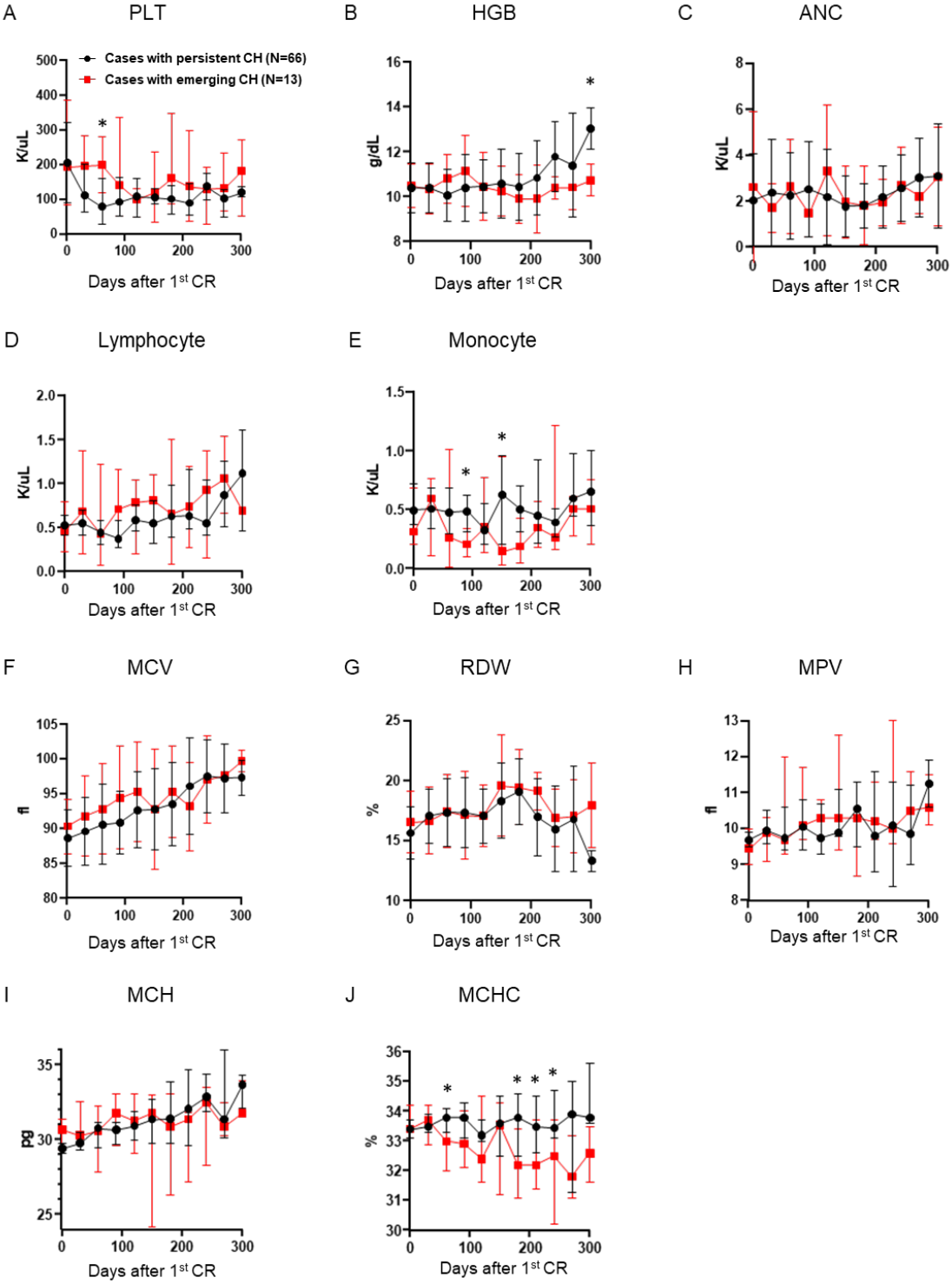
Peripheral blood (PB) cell counts were checked every about 30 days at the recovery phase. PB data were censored when the patients had a relapse or allogeneic stem cell transplantation. (**A**) PLT, (**B**) HGB, (**C**) ANC, (**D**) lymphocyte, (**E**) Monocyte, (**F**) MCV, (**G**) RDW, (**H**) MPV, (**I**) MCH, (**J**) MCHC. * $P < 0.05$.
Abbreviation: PLT; platelets, HGB; hemoglobin, ANC; absolute neutrophil counts, MCV; mean cell volume, RDW; red cell distribution width, MPV; mean platelet volume, MCH; mean corpuscular hemoglobin, MCHC; mean corpuscular hemoglobin concentration.

Figure S12
The long-term trend of peripheral blood counts in patients with post-CR CH with *ASXL1* mutation (Post-CR CH negative: black line, Post-CR CH positive with *ASXL1* mutation: red line).



Peripheral blood (PB) cell counts were checked every about 30 days at the recovery phase. PB data were censored when the patients had a relapse or allogeneic stem cell transplantation. (**A**) PLT, (**B**) HGB, (**C**) ANC, (**D**) lymphocyte, (**E**) Monocyte, (**F**) MCV, (**G**) RDW, (**H**) MPV, (**I**) MCH, (**J**) MCHC. * $P < 0.05$.
Abbreviation: PLT; platelets, HGB; hemoglobin, ANC; absolute neutrophil counts, MCV; mean cell volume, RDW; red cell distribution width, MPV; mean platelet volume, MCH; mean corpuscular hemoglobin, MCHC; mean corpuscular hemoglobin concentration.

Figure S13
The long-term trend of peripheral blood counts in patients with emerging CH (red line) and with persistent CH (black line).



Peripheral blood (PB) cell counts were checked every about 30 days at the recovery phase. PB data were censored when the patients had a relapse or allogeneic stem cell transplantation. (**A**) PLT, (**B**) HGB, (**C**) ANC, (**D**) lymphocyte, (**E**) Monocyte, (**F**) MCV, (**G**) RDW, (**H**) MPV, (**I**) MCH, (**J**) MCHC. * $P < 0.05$.

Abbreviation: PLT; platelets, HGB; hemoglobin, ANC; absolute neutrophil counts, MCV; mean cell volume, RDW; red cell distribution width, MPV; mean platelet volume, MCH; mean corpuscular hemoglobin, MCHC; mean corpuscular hemoglobin concentration.

Figure S14
(**A**) Cumulative incidence of relapse (CIR) by the number of mutations as post-CR CH (0: black line, 1: red line, more than 1: green line). (**B**) CIR stratified by the variant allele frequency (VAF) of post-CR CH (negative: blue line, low VAF (1st quartile): green line, intermediate VAF (2nd-3rd quartile): red line, high VAF (4th quartile): black line). (**C**) CIR comparing between emerging post-CR CH versus persistent CH (emerging CH: black line, persistent CH: red line). All data were censored when patients proceeded with allo-SCT

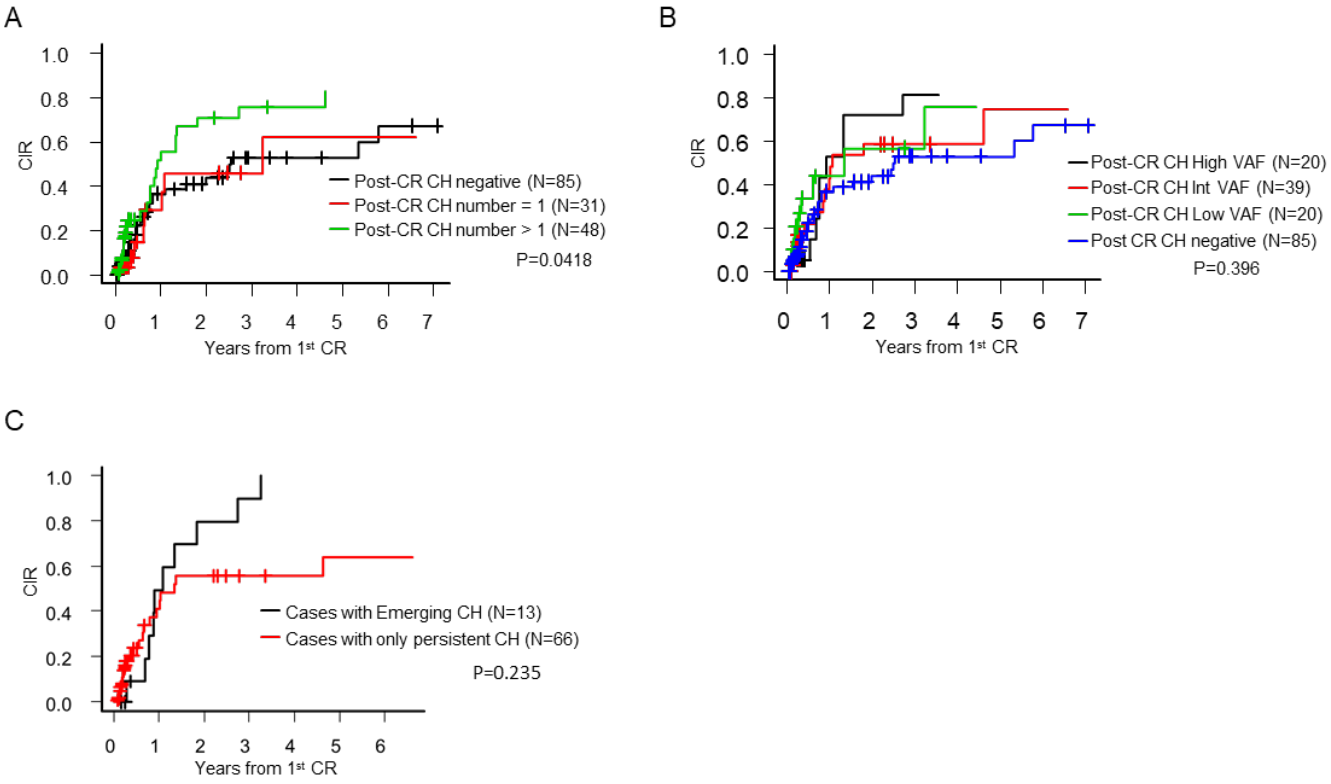
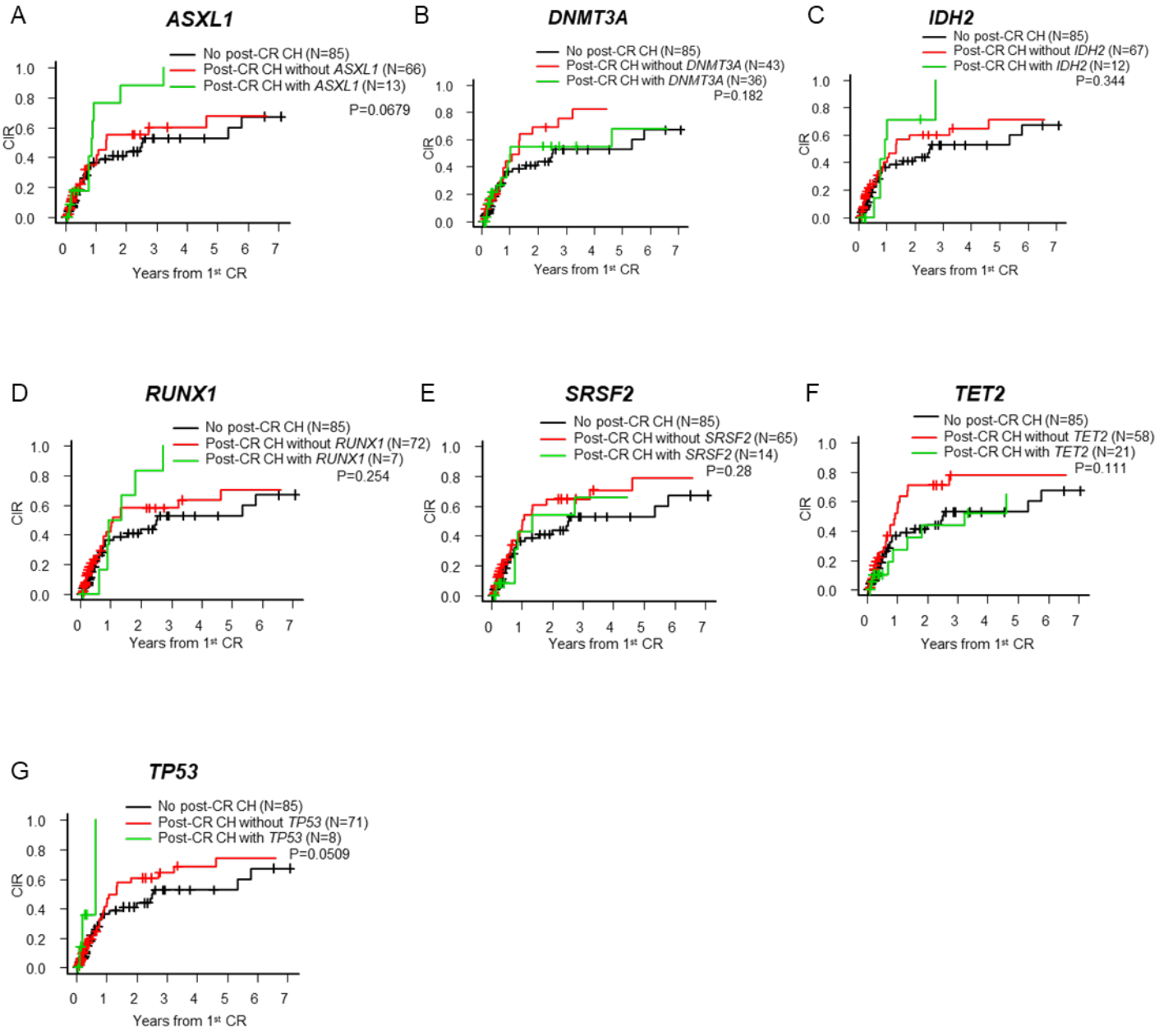


Figure S15
Cumulative incidence of relapse based on the mutated genes in post-CR CH. Only genes mutated in more than 5 patients are shown. Patients who proceed with allo-SCT were censored at the time of transplant.



No post-CR CH: black line, Post-CR CH positive without relevant gene: red line, Post-CR CH positive with relevant gene: green line. (**A**) *ASXL1*, (**B**) *DNMT3A*, (**C**) *IDH2*, (**D**) *RUNX1*, (**E**) *SRSF2*, (**F**) *TET2*, (**G**) *TP53*.

Figure S16

(**A**) Non-relapse mortality by the number of mutated genes as post-CR CH (0: black line, 1: red line, 2: green line). (**B**) Non-relapse mortality based on variant allele frequency (VAF) stratification (negative: blue line, low VAF (1st quartile): green line, intermediate VAF (2nd-3rd quartile): red line, high VAF (4th quartile): black line). Patients who proceed with allo-SCT were censored at the time of transplant.

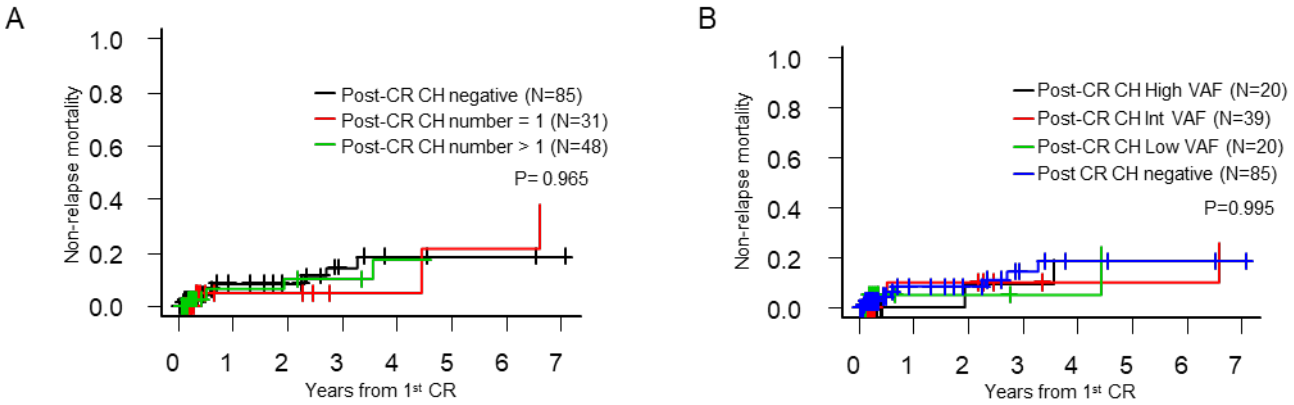
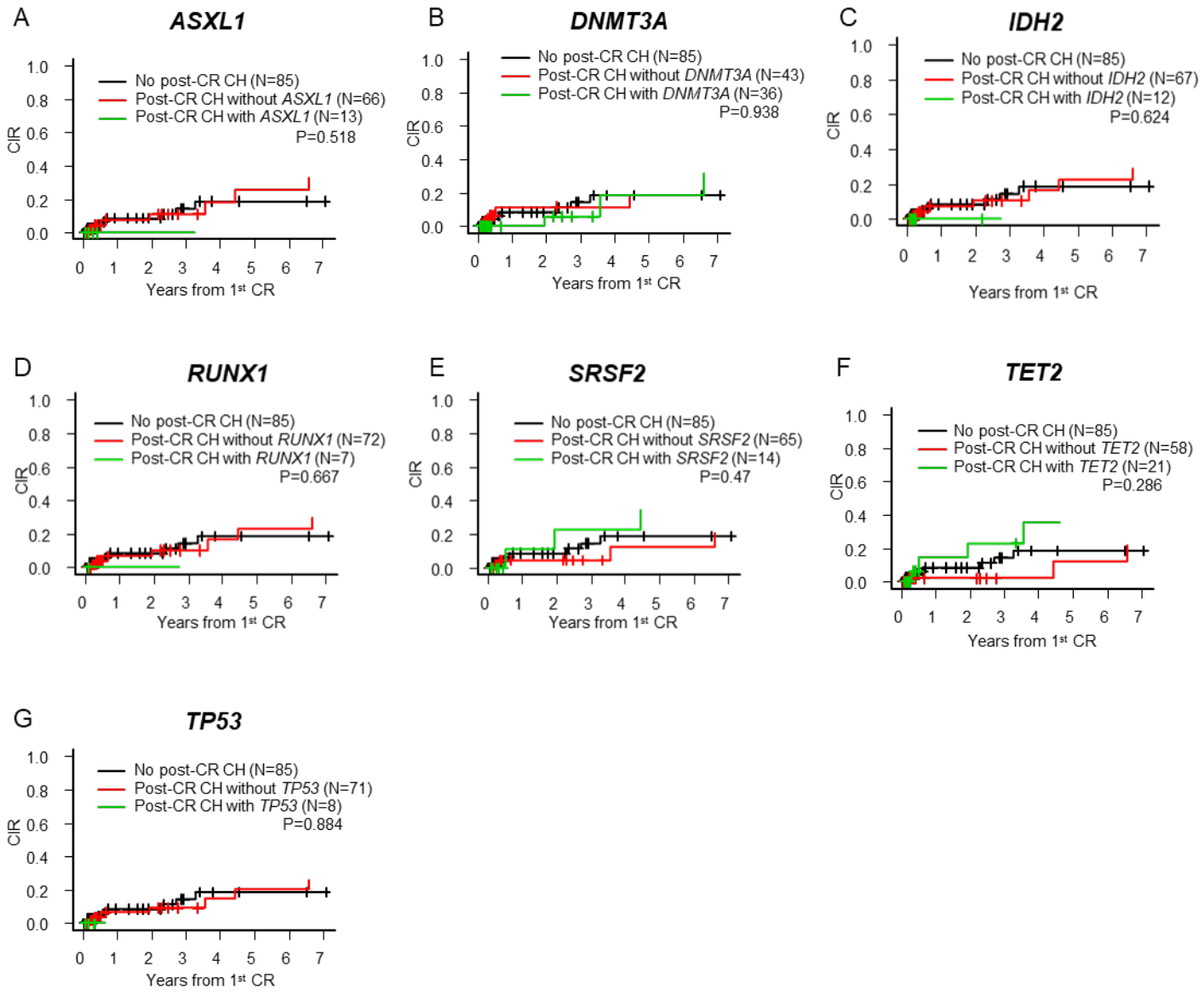


Figure S17
Non-relapse mortality based on the mutated genes in post-CR CH. Only genes mutated in more than 5 patients are shown. Patients who proceed with allo-SCT were censored at the time of transplant.



No post-CR CH: black line, Post-CR CH positive without relevant gene: red line, Post-CR CH positive with relevant gene: green line. (**A**) *ASXL1*, (**B**) *DNMT3A*, (**C**) *IDH2*, (**D**) *RUNX1*, (**E**) *SRSF2*, (**F**) *TET2*, (**G**) *TP53*.

FIgure S18
Cumulative incidence of atherosclerotic cardiovascular disease (ASCVD) in post-CR CH negative patients (black line) and post-CR CH positive patients (red line). Only one patient developed ASCVD in post-CR CH positive case.

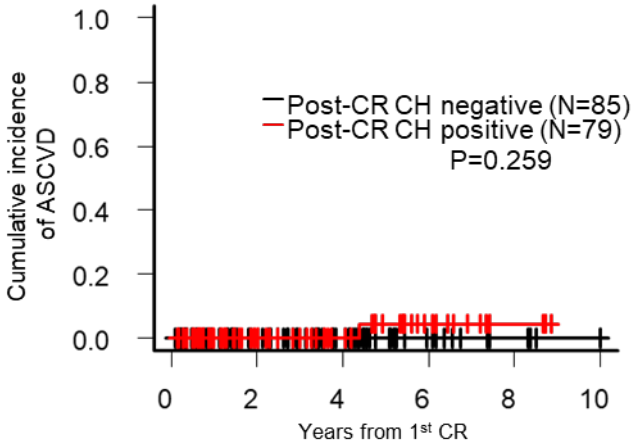
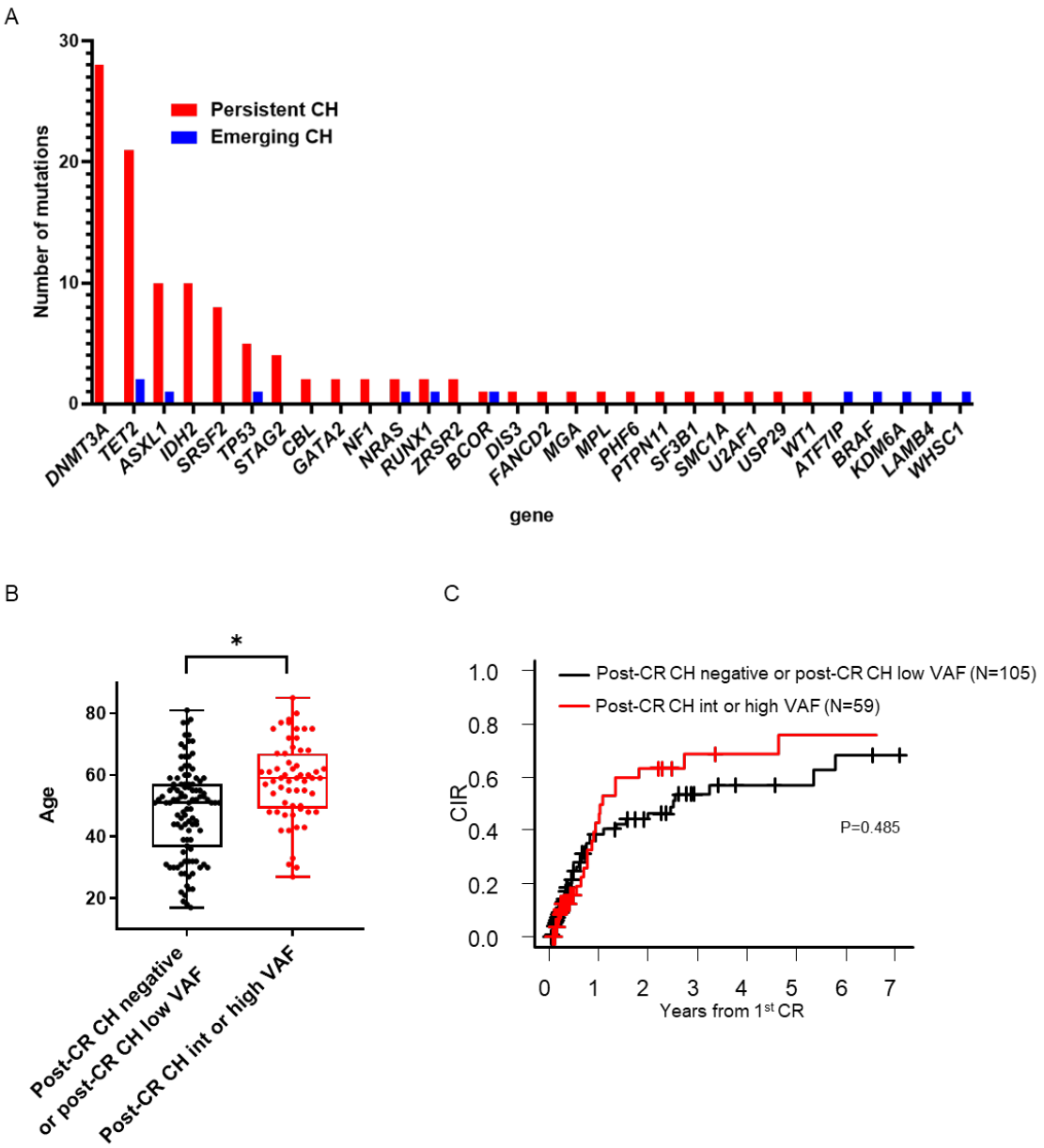


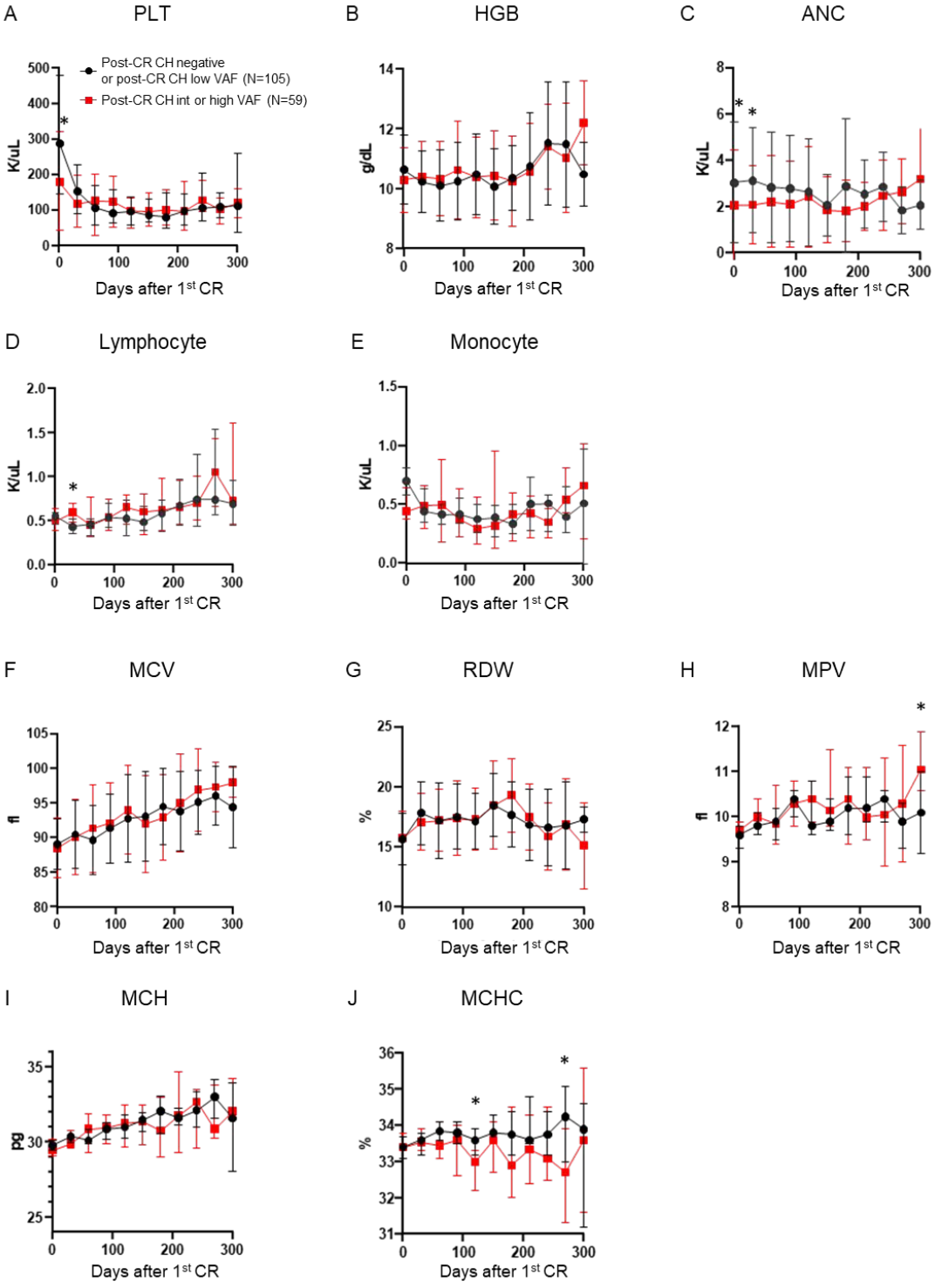
Figure S19
Frequency of somatic mutations, age, and cumulative incidence of relapse in post-CR CH intermediate (2nd-3rd quartile) or high (4th quartile) variant allele frequency (VAF) patients.



(**A**) Bar graph showing the number of mutations detected as persistent preleukemic CH (red bar) or emerging CH (blue bar) in intermediate or high VAF patients. (**B**) Box plots comparing the age of patients in post-CR CH negative or post-CR CH low (1st quartile) VAF patients, and in post-CR CH intermediate or high VAFpatients. (**C**) Cumulative incidence of relapse in post-CR CH negative or post-CR CH low VAF patients (black line), and in post-CR CH intermediate or high VAF patients (red line). Patients were censored when they underwent allo-SCT in 1st CR.

Figure S20

The long-term trend of peripheral blood counts in post-CR CH negative or post-CR CH low (1st quartile) variant allele frequency (VAF) patients (black line), and in post CR CH intermediate (2nd-3rd quartile) or high (4th quartile) VAF patients (red line).



Peripheral blood (PB) cell counts were checked every about 30 days at the recovery phase. PB data were censored when the patients had a relapse or allogeneic stem cell transplantation. (**A**) PLT, (**B**) HGB, (**C**) ANC, (**D**) lymphocyte, (**E**) Monocyte, (**F**) MCV, (**G**) RDW, (**H**) MPV, (**I**) MCH, (**J**) MCHC. * $P < 0.05$.

Abbreviation: PLT; platelets, HGB; Hemoglobin, ANC; absolute neutrophil counts, MCV; mean cell volume, RDW; red cell distribution width, MPV; mean platelet volume, MCH; mean corpuscular hemoglobin, MCHC; mean corpuscular hemoglobin concentration.

Supplemental Table

Table S1. List of 295 genes targeted by next-generation sequencing.
Table S2. Summary of high-confidence somatic mutations detected in 164 AML patients.
Table S3. Comparison of clinical characteristics between post-CR CH negative patients (N = 85) and post-CR CH positive patients (N = 79).
Table S4. Therapies in 43 AML patients who have multiple sequencing data after CR.
Table S5. Correlation between allo-SCT conditioning regimen and post-CR CH clearance.

Supplemental Reference

1. Zhang J, Fujimoto J, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science.* 2014;346(6206):256-259.
2. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-2079.
3. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* 2013;31(3):213-219.
4. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics.* 2009;25(21):2865-2871.
5. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood.* 2013;122(22):3616-3627; quiz 3699.
6. Morita K, Kantarjian HM, Wang F, et al. Clearance of Somatic Mutations at Remission and the Risk of Relapse in Acute Myeloid Leukemia. *J Clin Oncol.* 2018;36(18):1788-1797.
7. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplant.* 2013;48(3):452-458.
8. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med.* 2014;371(26):2488-2498.