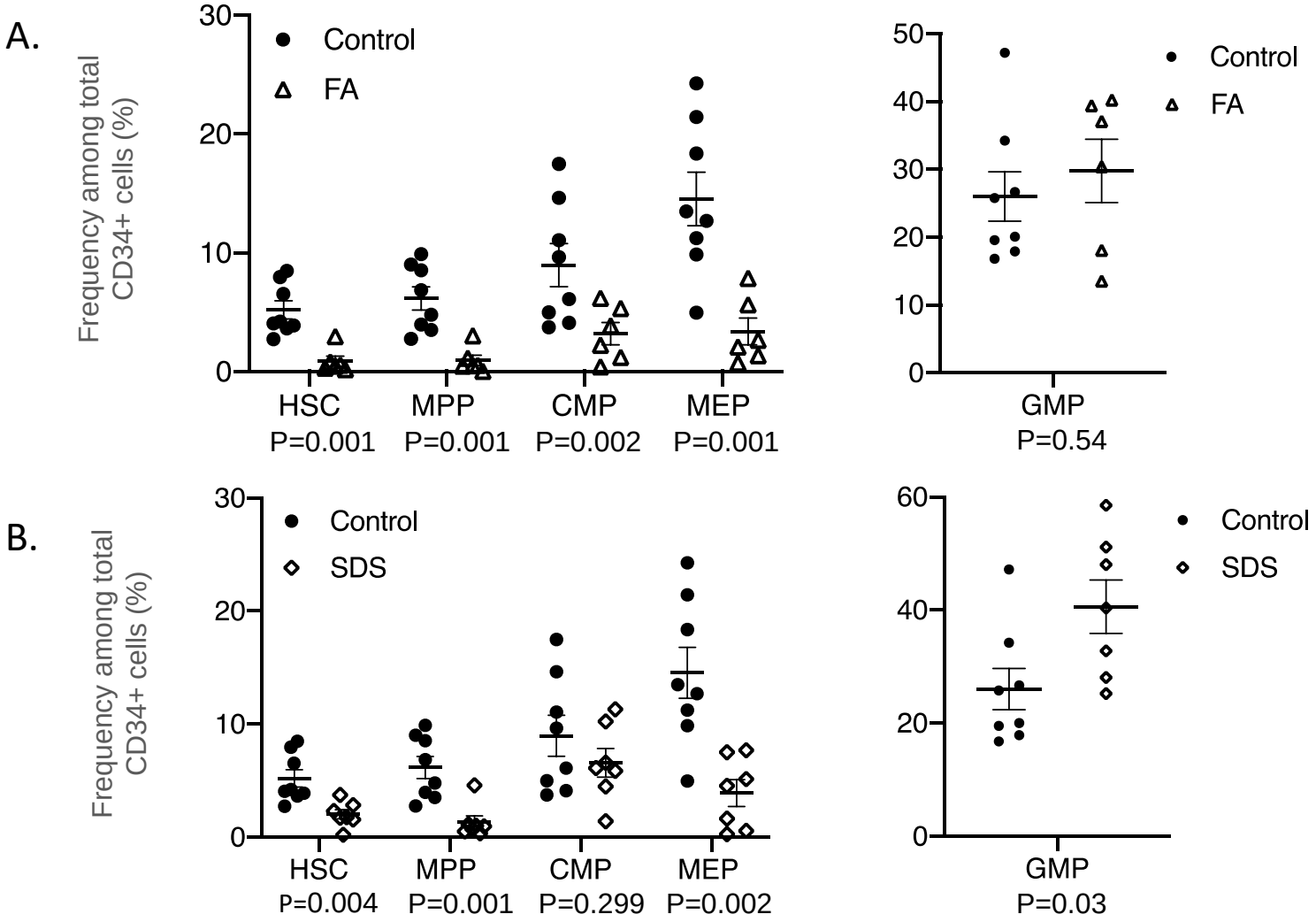
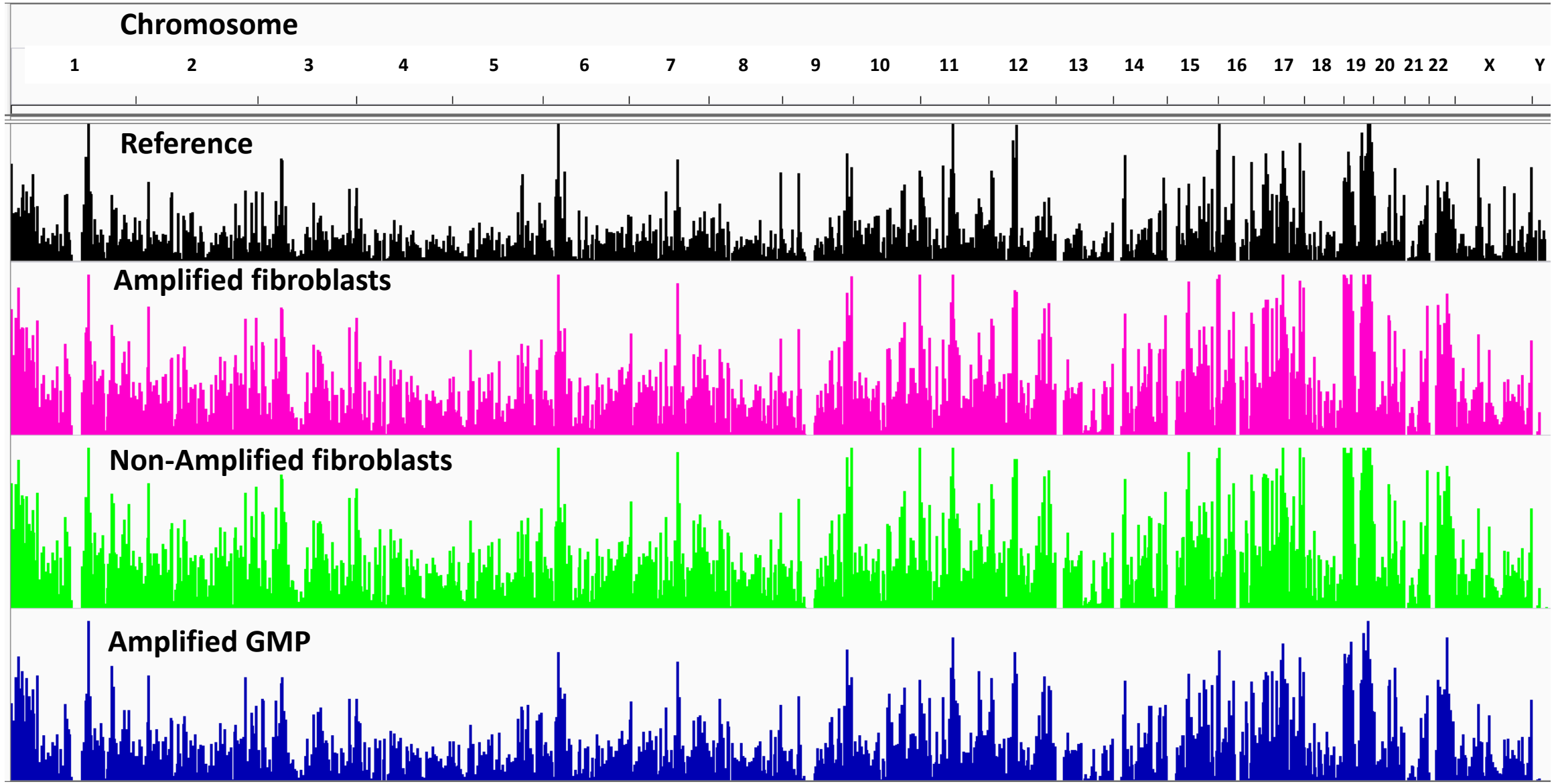
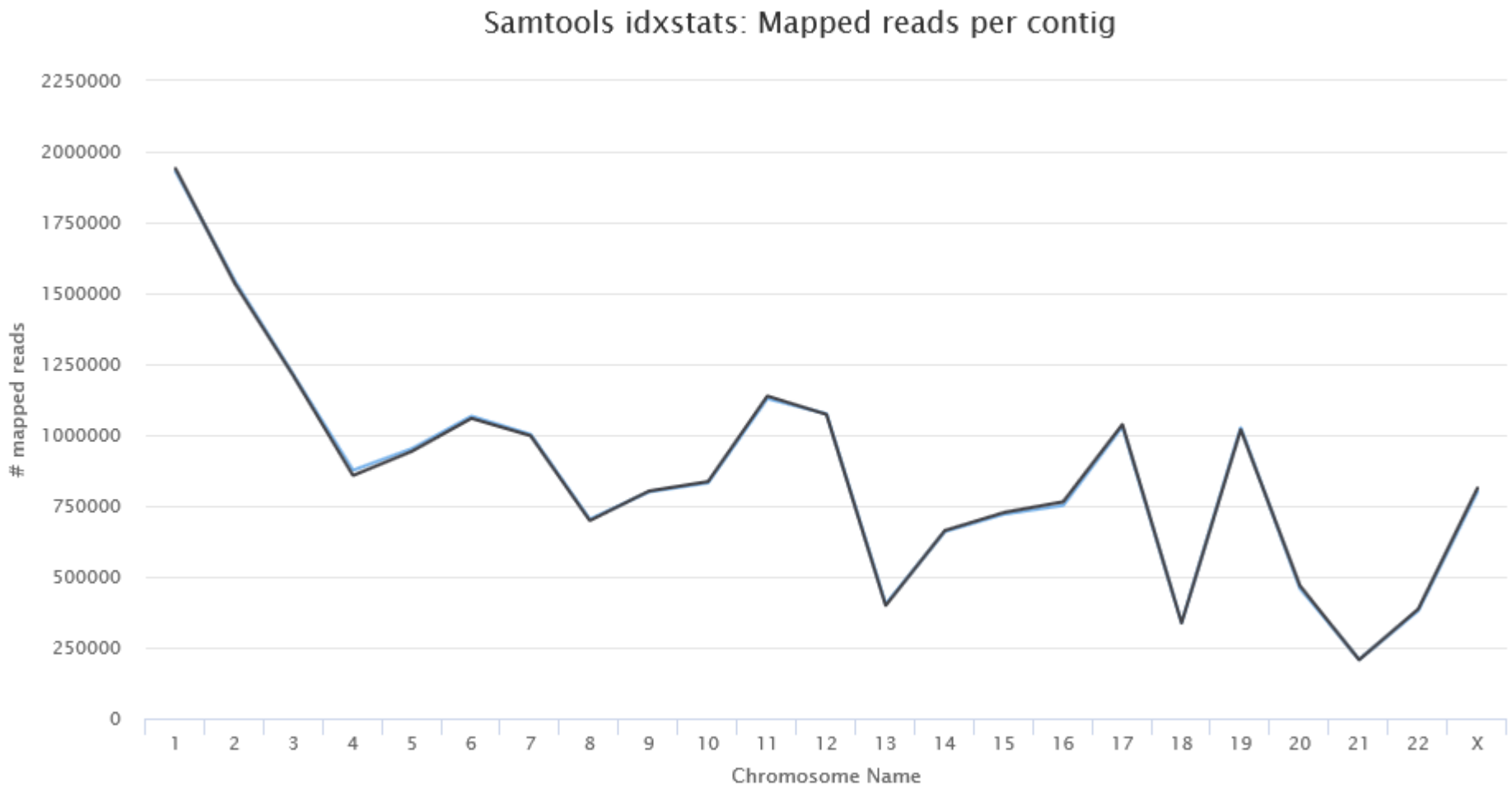
Supplemental Figure 1. Frequencies of hematopoietic stem cells and progenitors (HSCP) among bone marrow the CD34+ cell population. Bone marrow aspirate samples underwent deep immunophenotyping as described in the Methods. The proportions of cells among the total CD34+ cells are presented. (The frequency of HSCPs among live bone marrow mononuclear cells are presented in Figure 1 in the manuscript). **A.** Comparison of HSPCs between Fanconi anemia (FA) and controls. For better visualization of differences, GMPs are presented separately as their frequency among CD34+ cells was higher than other progenitors. **B.** Comparison of HSPCs between Shwachman-Diamond syndrome (SDS) and controls. For better visualization of differences, GMPs are presented separately as their frequency among CD34+ cells was higher than other progenitors. HSC, hematopoietic stem cell; MPP, multipotent progenitor cell; MEP, megakaryocyte erythroid progenitor; GMP, granulocyte monocyte progenitor. The mean percentage of a HSCP population among the total number of viable bone marrow cells is represented with standard error of mean (SEM). The same control data in Supplemental Figure 1B are also presented in Supplemental Figure 1A.



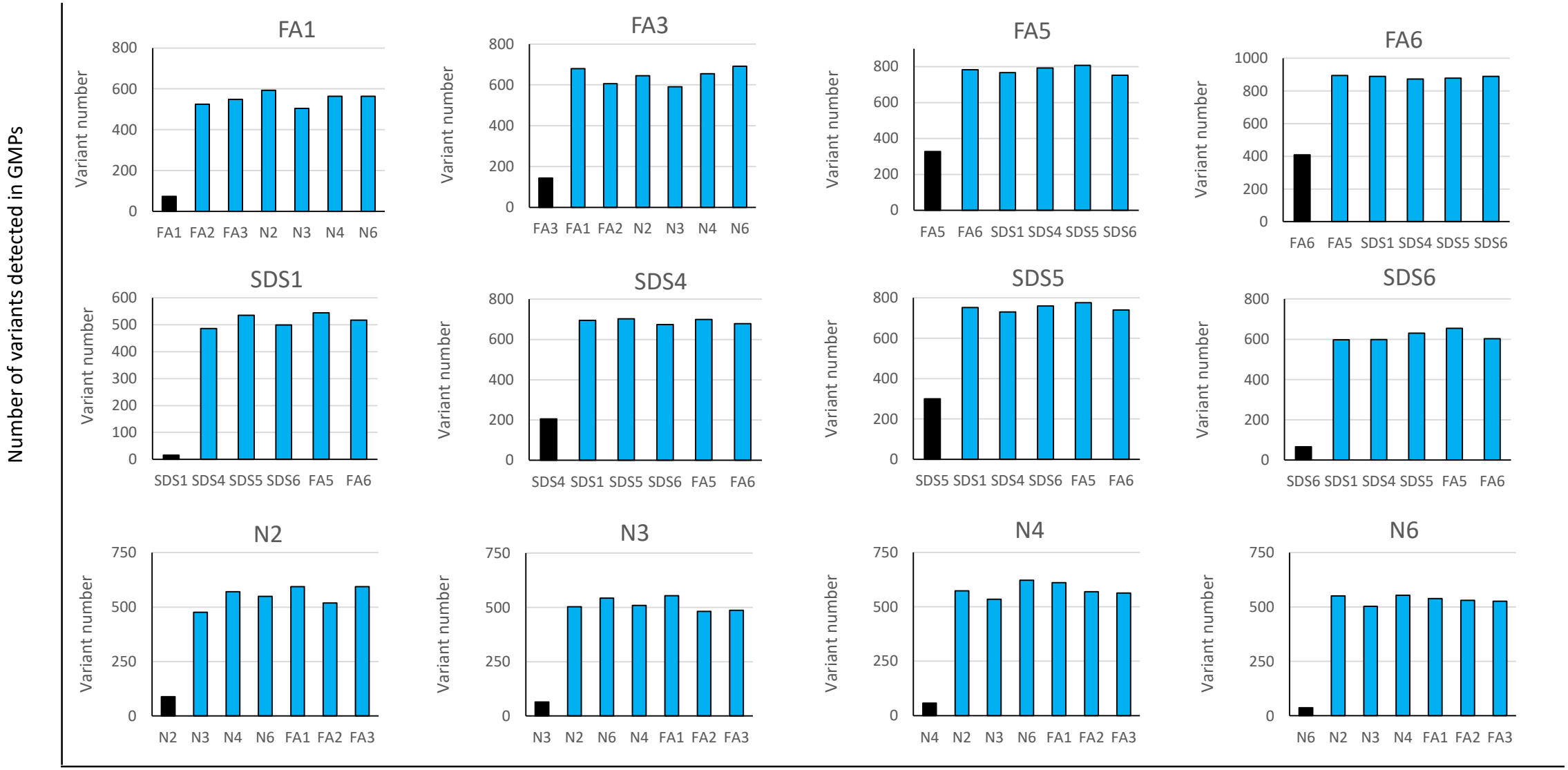
Supplemental Figure 2: Visualization of WES coverage across the human genome for amplified-genomic DNA from bone marrow fibroblasts (pink), non-amplified DNA from the same marrow fibroblast sample (green) and amplified-genomic DNA from GMPs of the same subject (blue). Coverage files (.bed file) were loaded onto the Integrated Genome Viewer (IGV). Human (b37) was used as a reference genome. Observed coverage of mapped reads between non-amplified and amplified samples indicates similar profile.



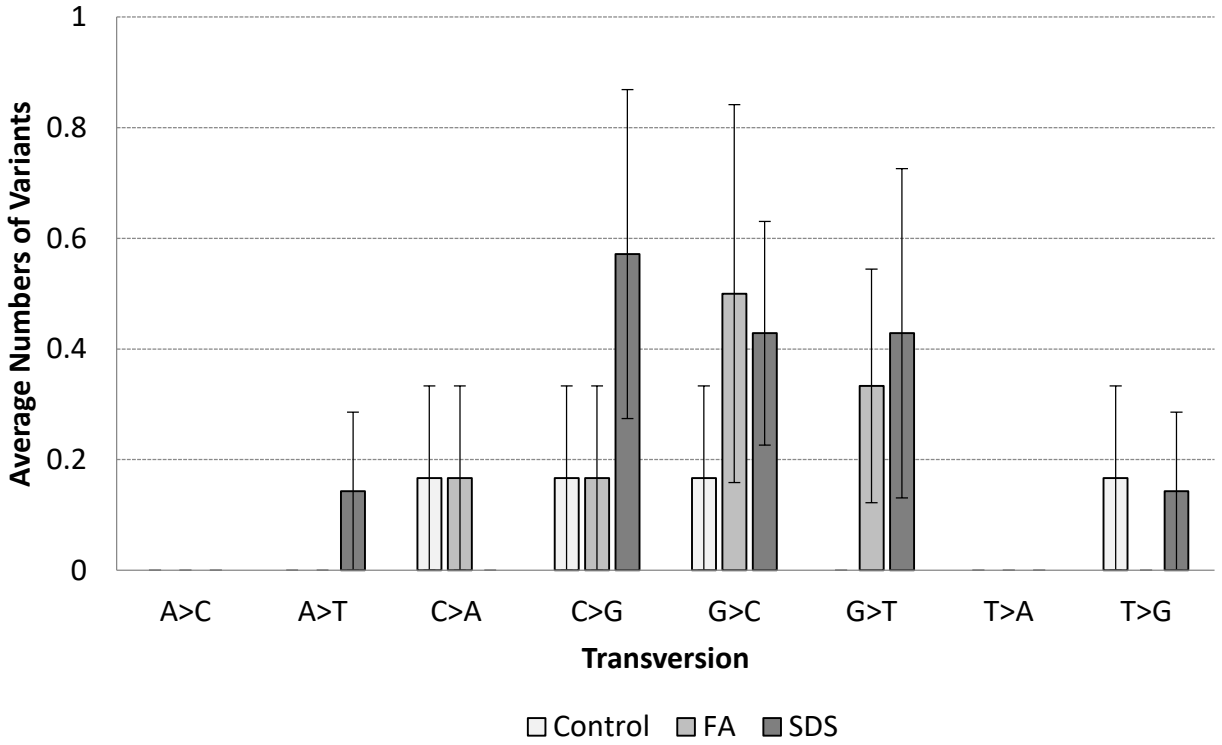
Supplemental Figure 3: Comparison of mapped reads from each chromosome using WES sequences from amplified-genomic DNA of marrow fibroblasts (Black line) and non-amplified genomic DNA of the same marrow fibroblast sample (blue line) in random 20M read pair subsets. Comparison was done by the Samtools idxstats software program. The results show almost identical numbers of variants detected in samples before and after DNA amplification.



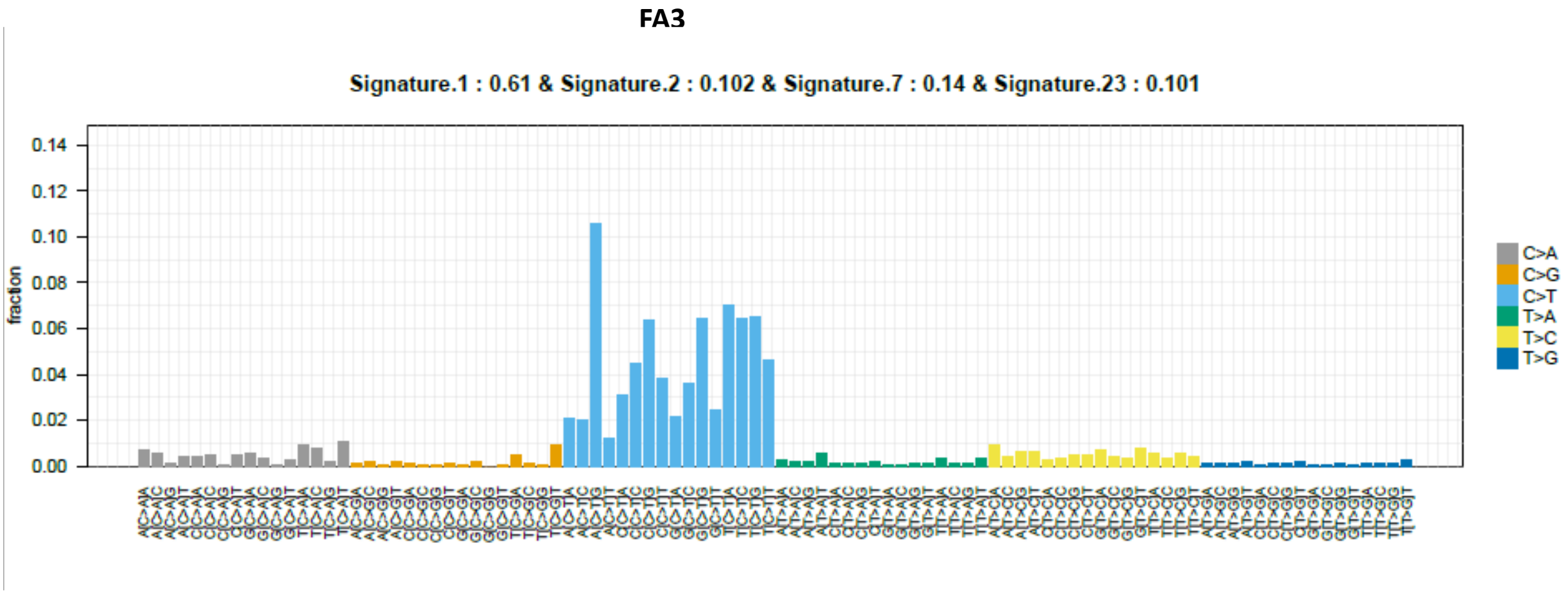
Supplemental Figure 4: Number of variants detected when a GMP data are compared to self marrow fibroblasts and to other subjects' marrow fibroblasts. Results of selected subjects from each group (FA, SDS and healthy controls) are shown. The variants used in this analysis were detected in both, WES and the cancer gene panel assays. They appeared in the cancer gene panel assay at >0.07% VAF in GMPs, did not appear in marrow fibroblasts and the read depth in both, GMPs and marrow fibroblasts, was at least 50.



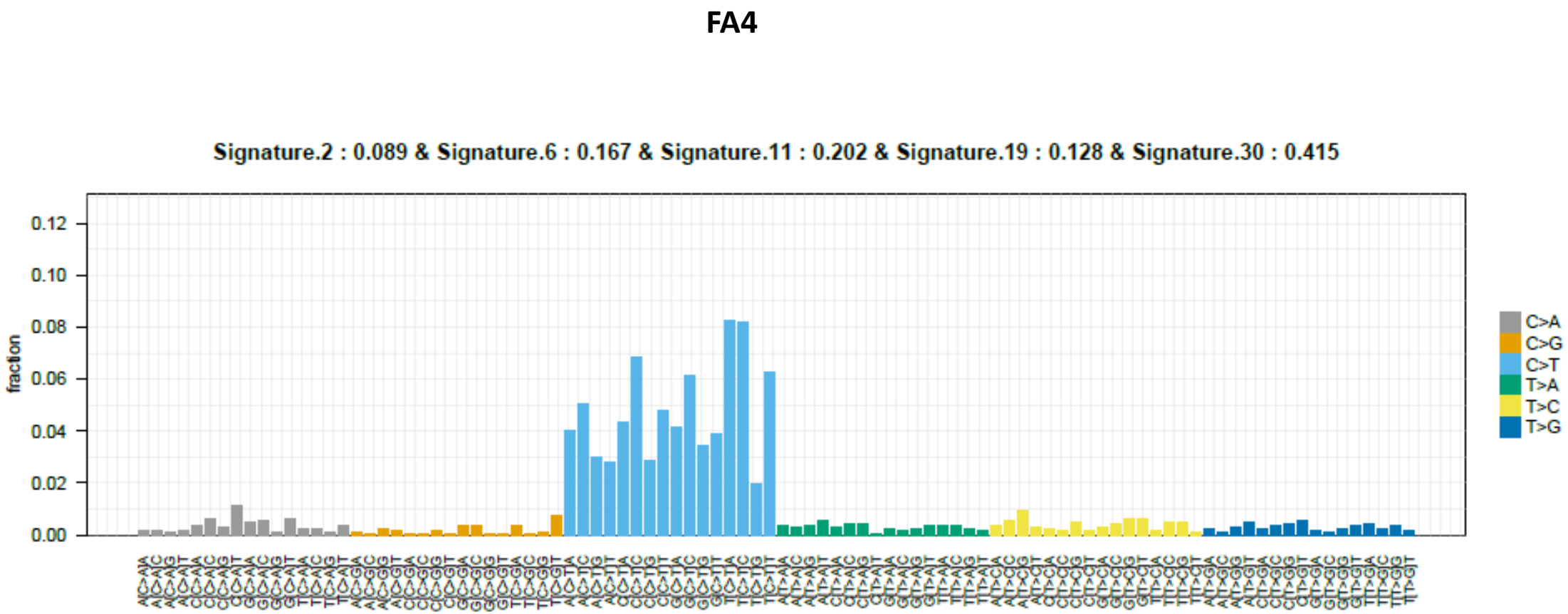
Supplemental Figure 5: Average number (+/- standard error of the mean) of each transversion variant per subject among the FA, SDS and healthy control groups.



Supplemental Figure 6: Trinucleotide signature generated from data of Fanconi anemia patient 1 (FA3) who did not develop bone marrow clones/MDS/AML. The signatures were generated by comparing the validated variants to the published signatures of cancer whole exome sequencing databases of the COSMIC published signatures. The trinucleotide signature is displayed by calculating the fraction of each trinuelotide variant relative to the total number of variants in this subject. The nucleotide change is displayed above each 16 trinucleotide block.

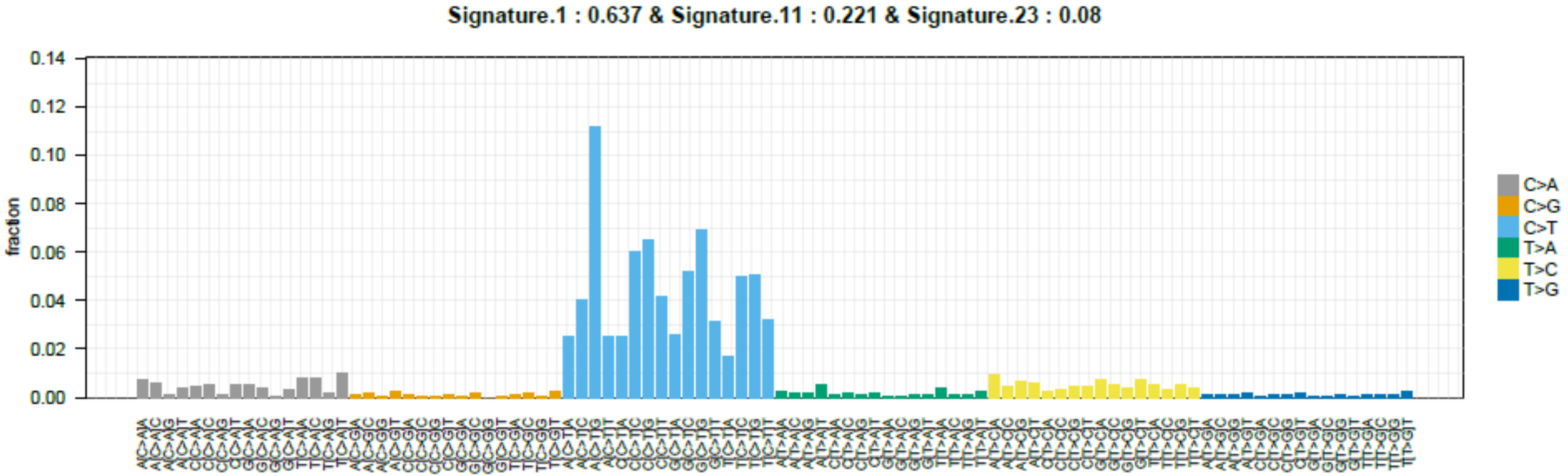


Supplemental Figure 7: Trinucleotide signature generated from data of Fanconi anemia patient 1 (FA4) who did not develop bone marrow clones/MDS/AML. The signatures were generated by comparing the validated variants to the published signatures of cancer whole exome sequencing databases of the COSMIC published signatures. The trinucleotide signature is displayed by calculating the fraction of each trinuelotide variant relative to the total number of variants in this subject. The nucleotide change is displayed above each 16 trinucleotide block.



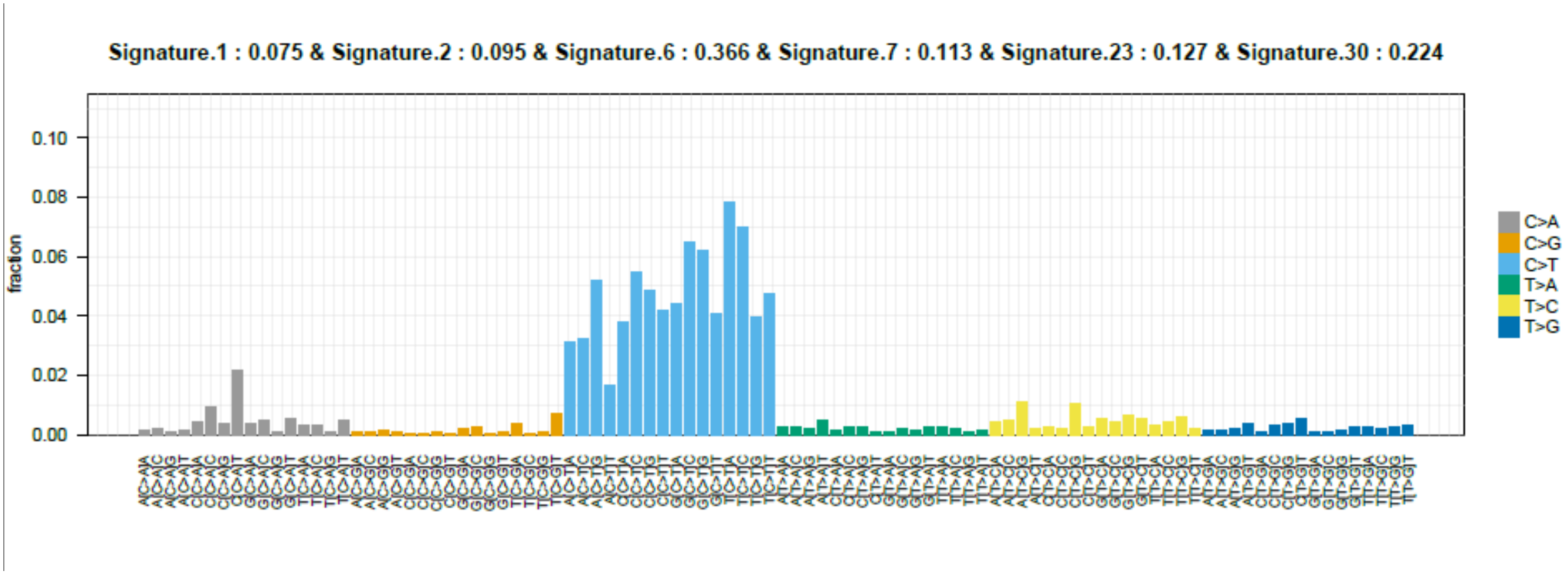
Supplemental Figure 8: Trinucleotide signature generated from data of Fanconi anemia patient 1 (FA5) who did not develop bone marrow clones/MDS/AML. The signatures were generated by comparing the validated variants to the published signatures of cancer whole exome sequencing databases of the COSMIC published signatures. The trinucleotide signature is displayed by calculating the fraction of each trinuelotide variant relative to the total number of variants in this subject. The nucleotide change is displayed above each 16 trinucleotide block.

FA5

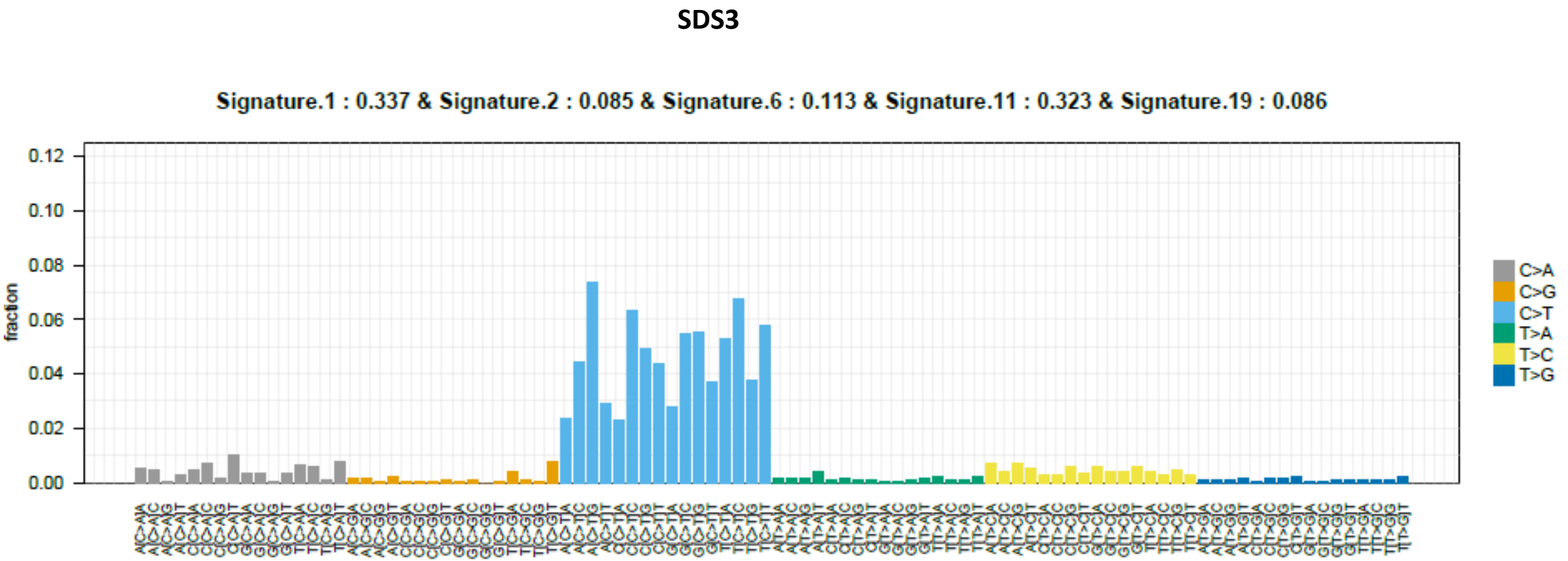


Supplemental Figure 9: Trinucleotide signature generated from data of Fanconi anemia patient 1 (FA6) who did not develop bone marrow clones/MDS/AML. The signatures were generated by comparing the validated variants to the published signatures of cancer whole exome sequencing databases of the COSMIC published signatures. The trinucleotide signature is displayed by calculating the fraction of each trinuelotide variant relative to the total number of variants in this subject. The nucleotide change is displayed above each 16 trinucleotide block.

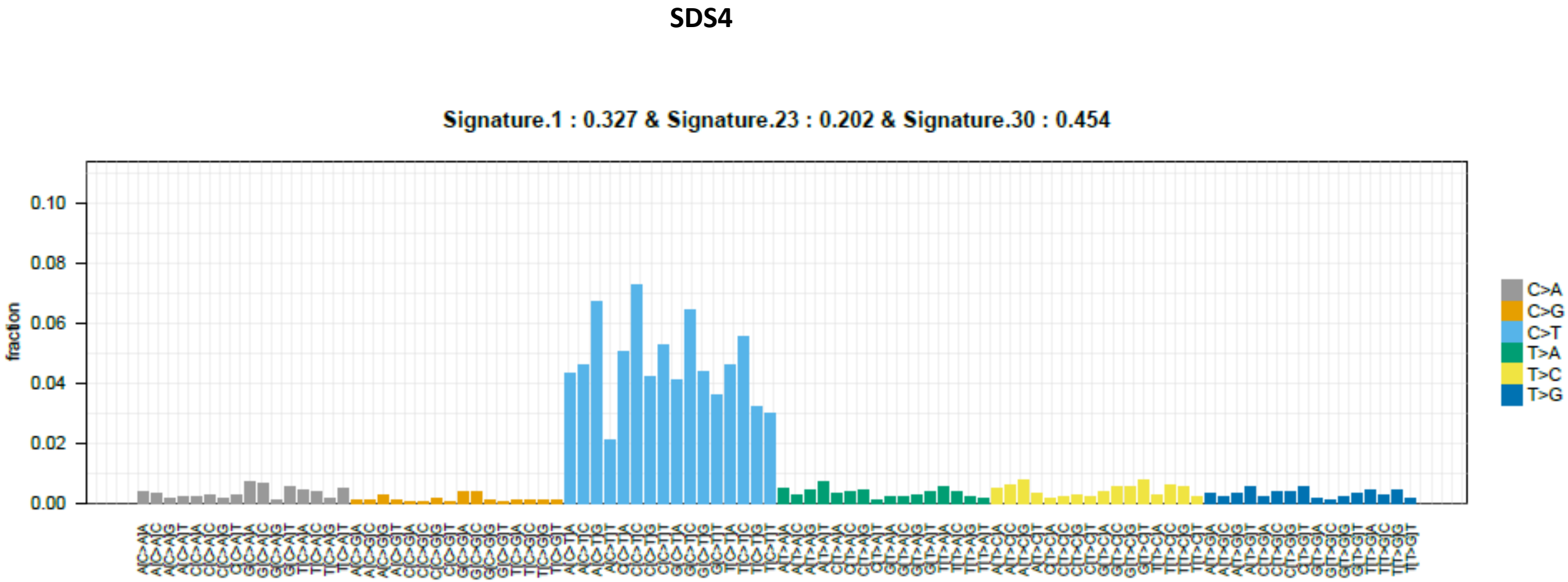
FA6



Supplemental Figure 10: Trinucleotide signature generated from data of Shwachman-Diamond syndrome patient 1 (SDS3) who did not develop bone marrow clones/MDS/AML. The signatures were generated by comparing the validated variants to the published signatures of cancer whole exome sequencing databases of the COSMIC published signatures. The trinucleotide signature is displayed by calculating the fraction of each trinucleotide variant relative to the total number of variants in this subject. The nucleotide change is displayed above each 16 trinucleotide block.

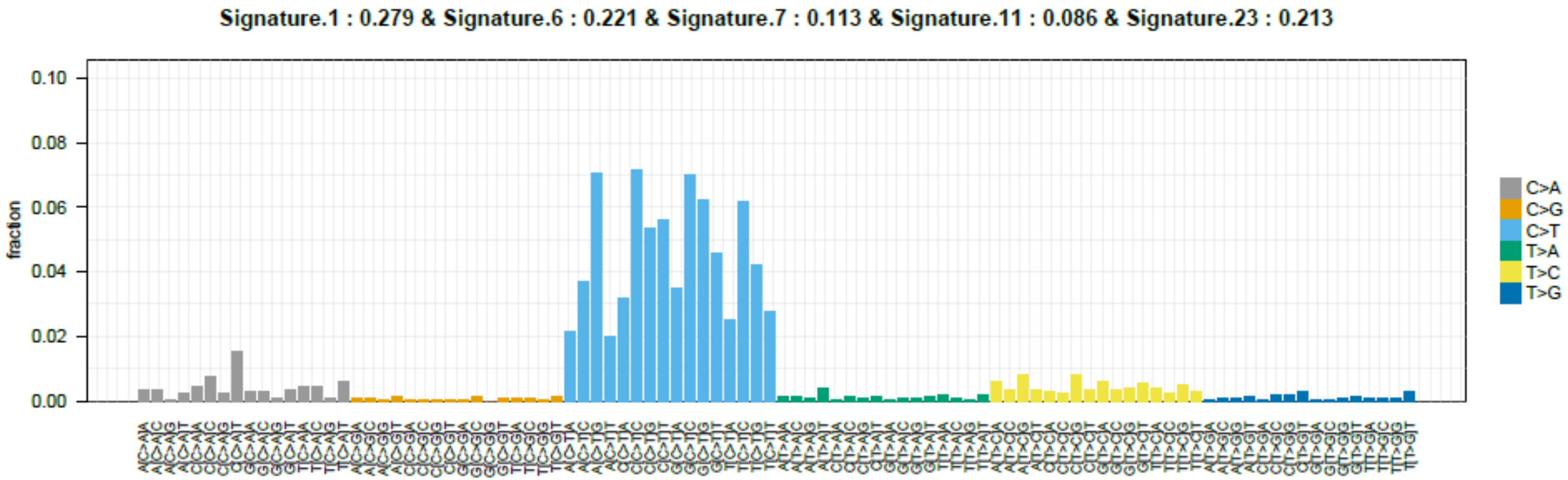


Supplemental Figure 11: Trinucleotide signature generated from data of Shwachman-Diamond syndrome patient 1 (SDS4) who did not develop bone marrow clones/MDS/AML. The signatures were generated by comparing the validated variants to the published signatures of cancer whole exome sequencing databases of the COSMIC published signatures. The trinucleotide signature is displayed by calculating the fraction of each trinucleotide variant relative to the total number of variants in this subject. The nucleotide change is displayed above each 16 trinucleotide block.



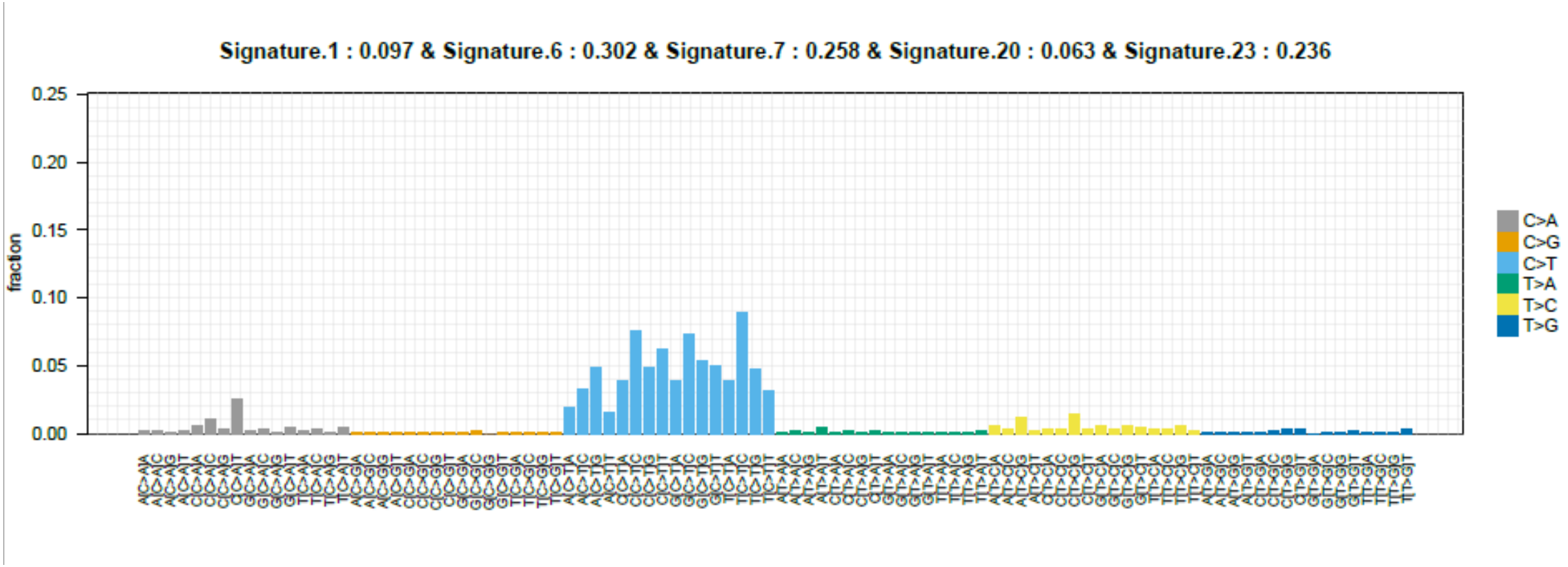
Supplemental Figure 12: Trinucleotide signature generated from data of Shwachman-Diamond syndrome patient 5 (SDS5) who did not develop bone marrow clones/MDS/AML. The signatures were generated by comparing the validated variants to the published signatures of cancer whole exome sequencing databases of the COSMIC published signatures. The trinucleotide signature is displayed by calculating the fraction of each trinucleotide variant relative to the total number of variants in this subject. The nucleotide change is displayed above each 16 trinucleotide block.

SDS5

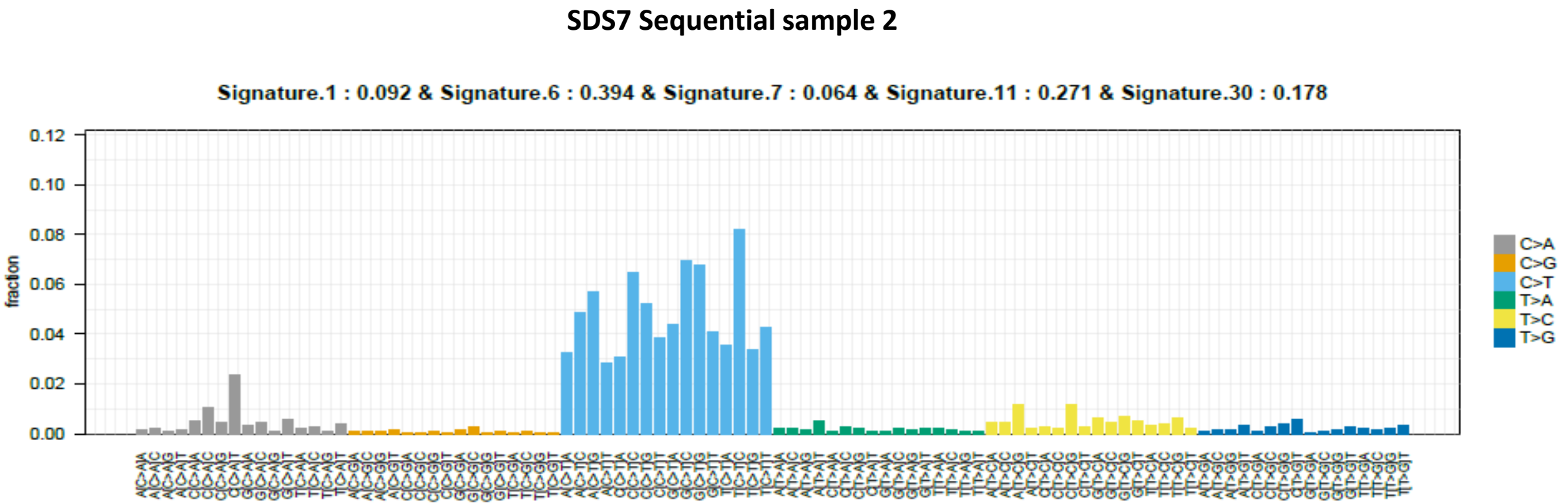


Supplemental Figure 13: Trinucleotide signature generated from data of Shwachman-Diamond syndrome patient 6 (SDS6) who did not develop bone marrow clones/MDS/AML. The signatures were generated by comparing the validated variants to the published signatures of cancer whole exome sequencing databases of the COSMIC published signatures. The trinucleotide signature is displayed by calculating the fraction of each trinucleotide variant relative to the total number of variants in this subject. The nucleotide change is displayed above each 16 trinucleotide block.

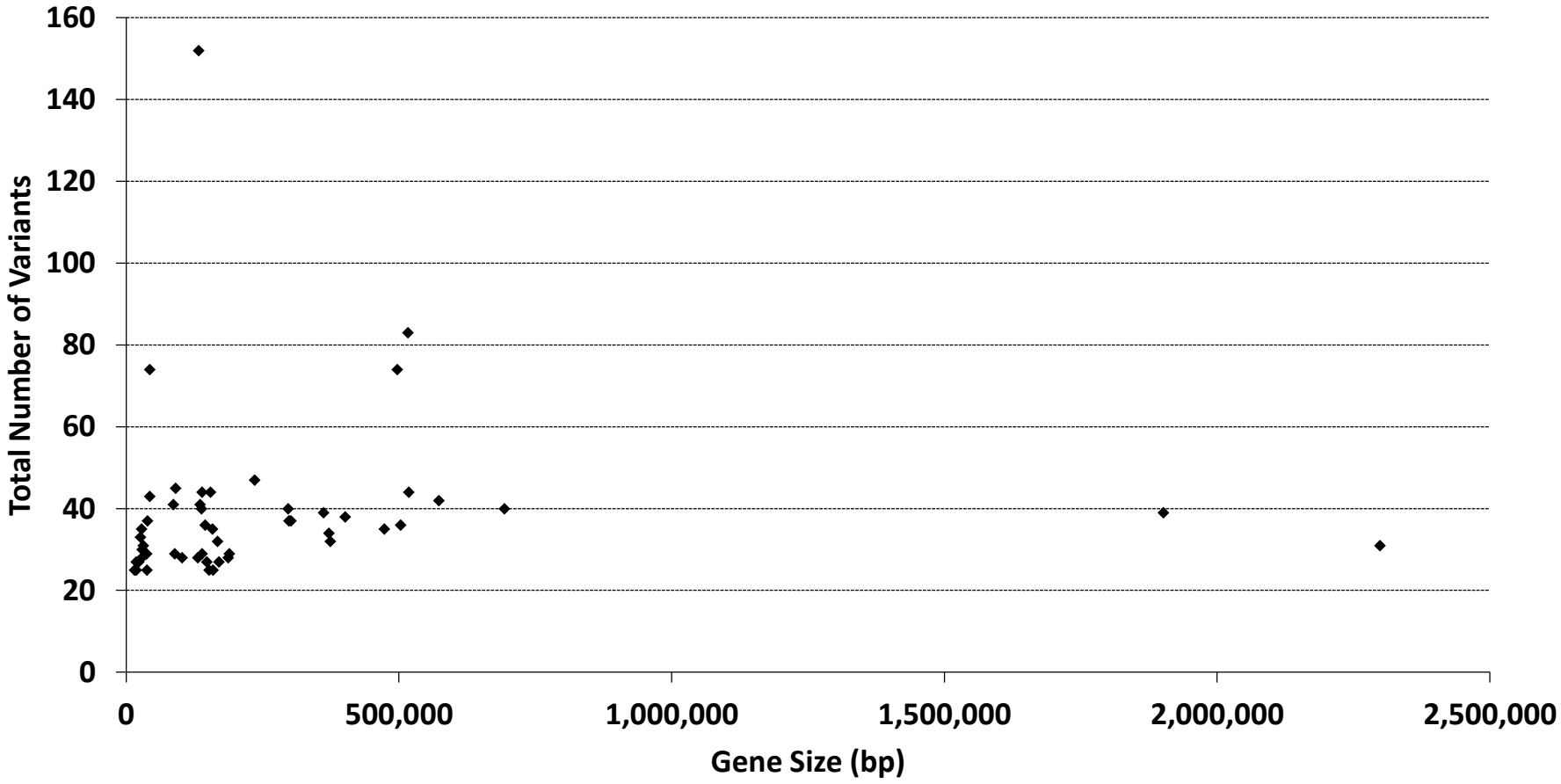
SDS6



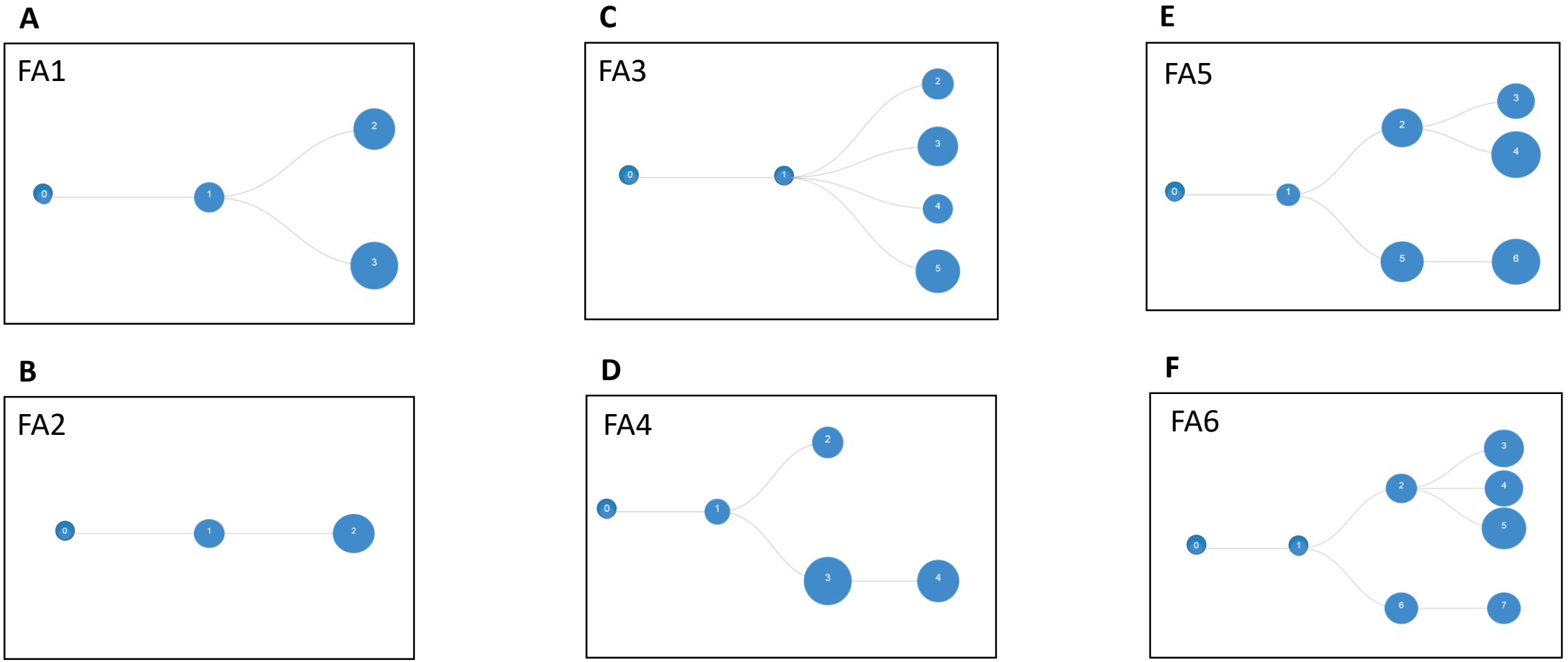
Supplemental Figure 14. Trinucleotide signatures generated from second sequential samples of Shwachman-Diamond syndrome patient 7 (SDS7) who did not develop bone marrow clones/MDS/AML at the time of the bone marrow sampling used in this figures, but subsequently develop AML (See Supplementary Figure 20). Cancer signature construction from sequential sample 1 was not possible due to low variant numbers. The signatures were generated by comparing the validated variants to the published signatures of cancer whole exome sequencing databases of the COSMIC published signatures. The trinucleotide signature is displayed by calculating the fraction of each trinucleotide variant relative to the total number of variants in this subject. The nucleotide change is displayed above each 16 trinucleotide block.



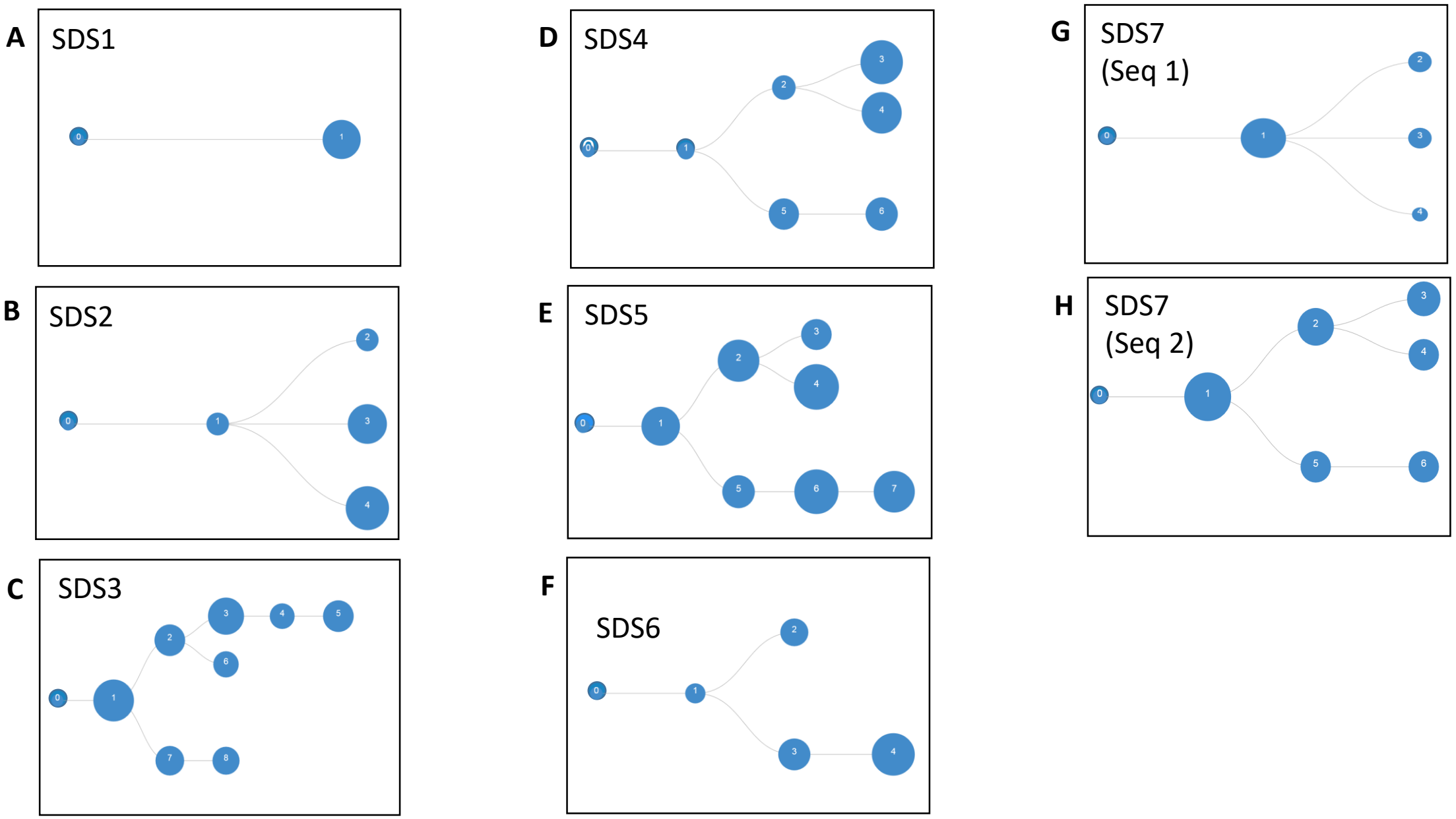
Supplemental Figure 15. Total number of variants per gene in each of the top 50 mutated genes ordered according to the size. There does was no linear correlation between the gene size and the number of variants



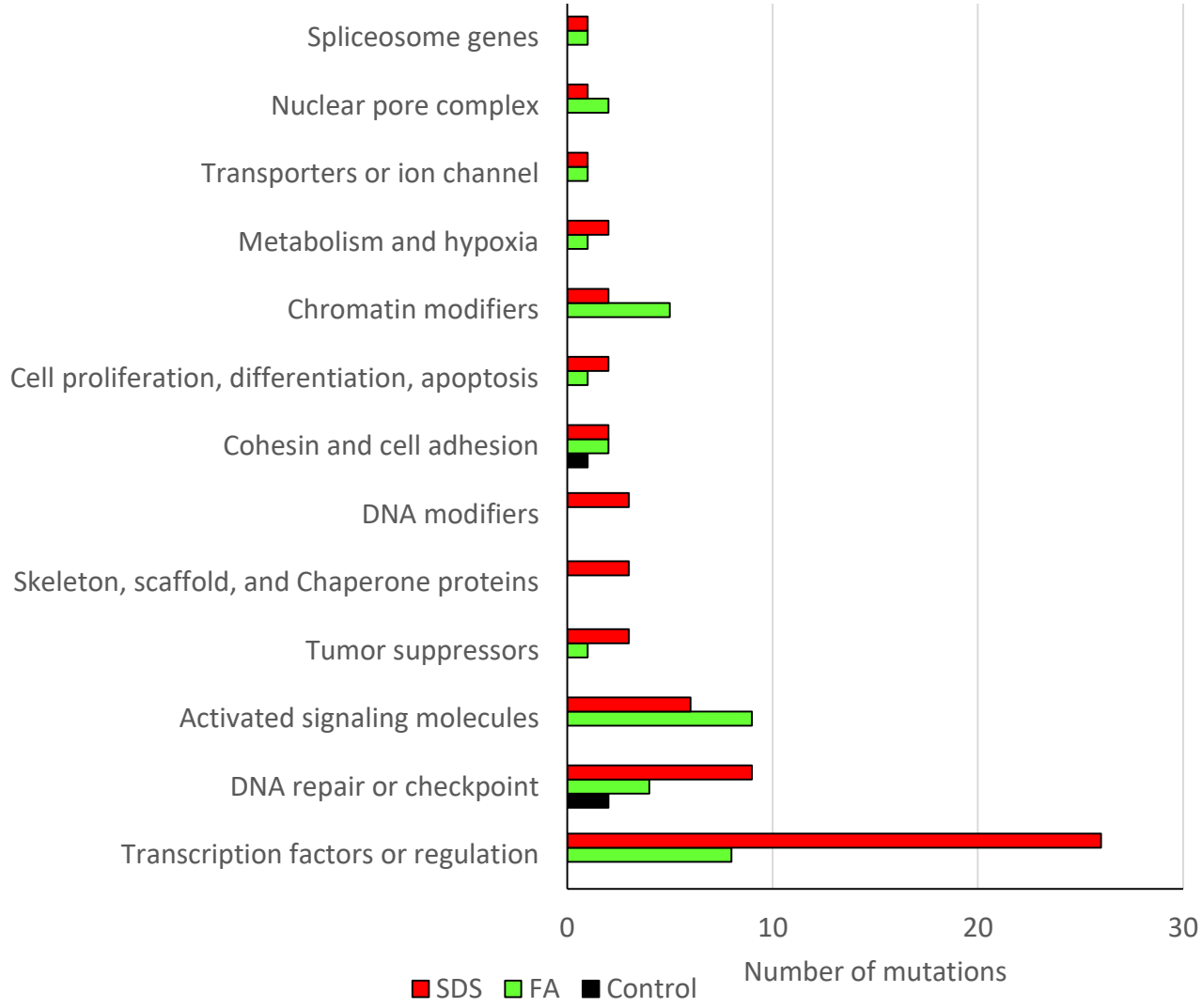
Supplemental Figure 16: Mutational tree of Fanconi anemia patient samples (without leukemia). The detailed list of mutated genes in each clone and sub-clones is in Supplementary Table 6.



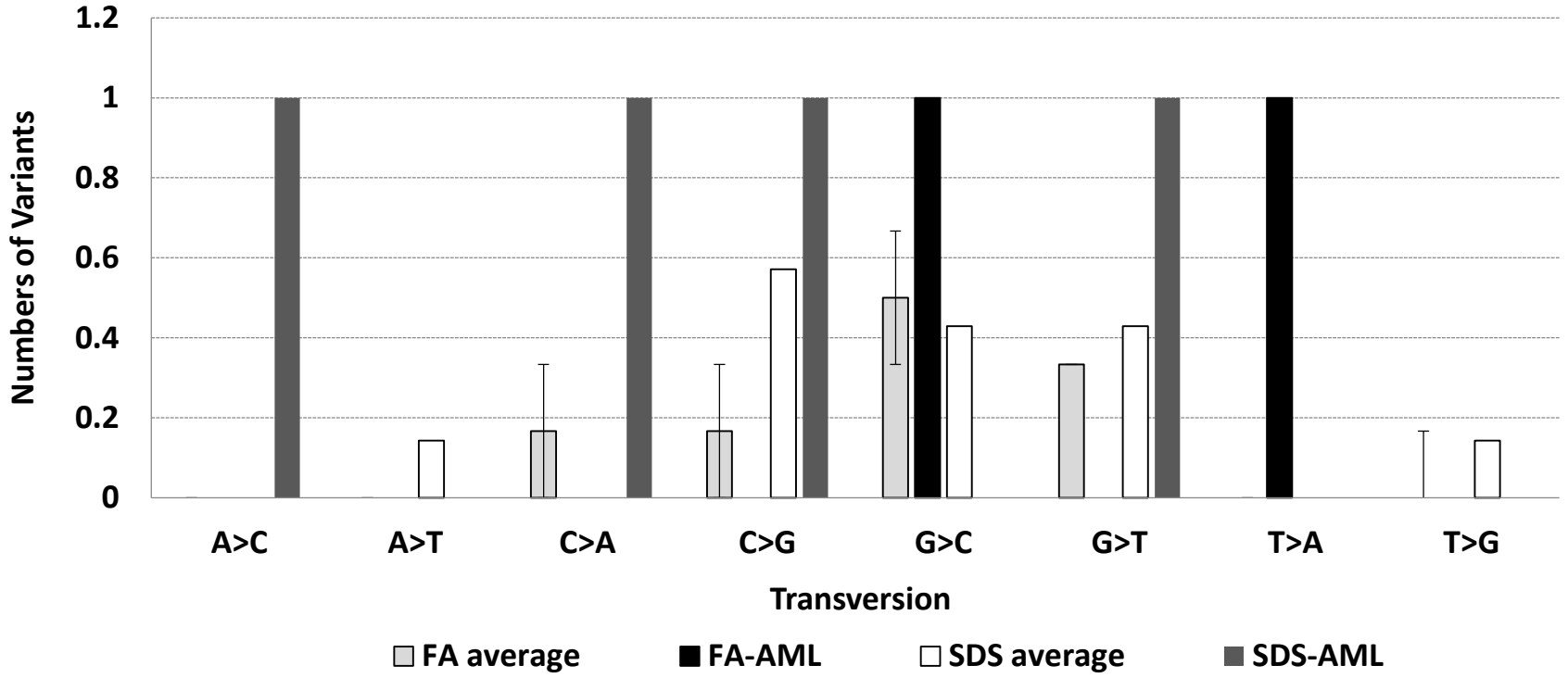
Supplemental Figure 17: Mutational tree of Shwachman-Diamond syndrome patients (without clones/MDS/leukemia) . The detailed list of mutated genes in each clone and sub-clone is in Supplementary Table 6.



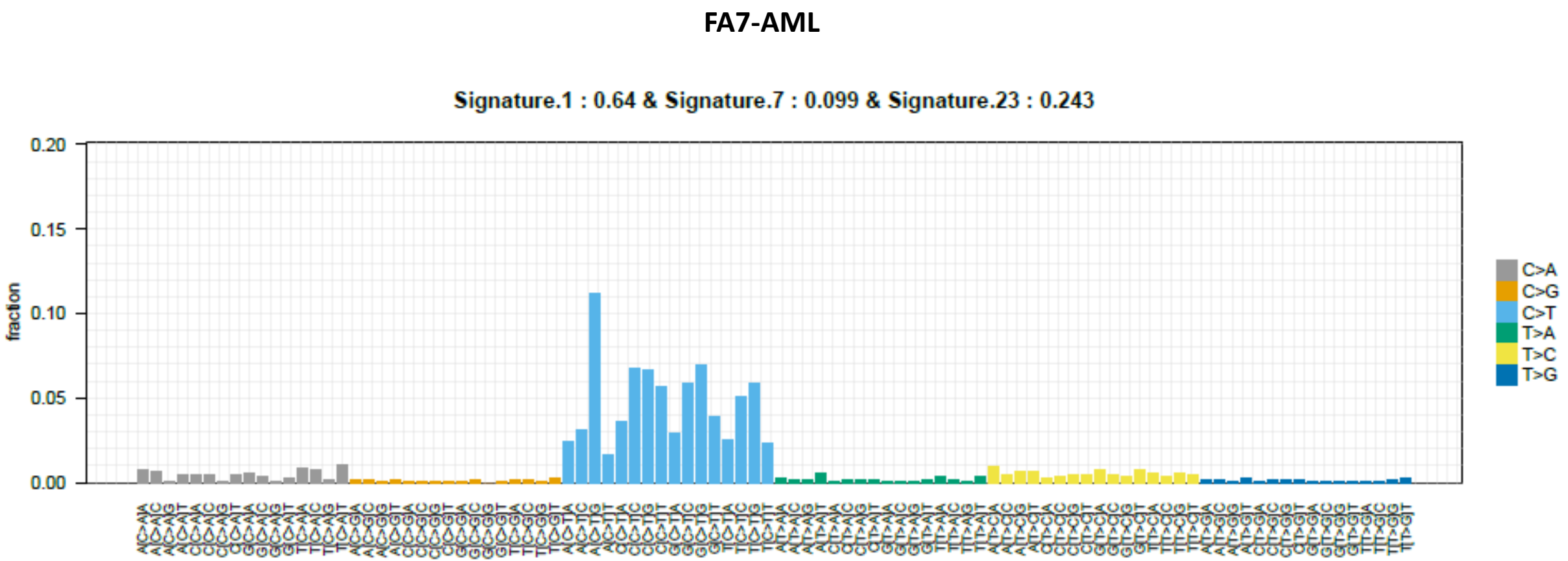
Supplemental Figure 18: List of pathways mutated in FA/SDS patients without transformation and in healthy control subjects.



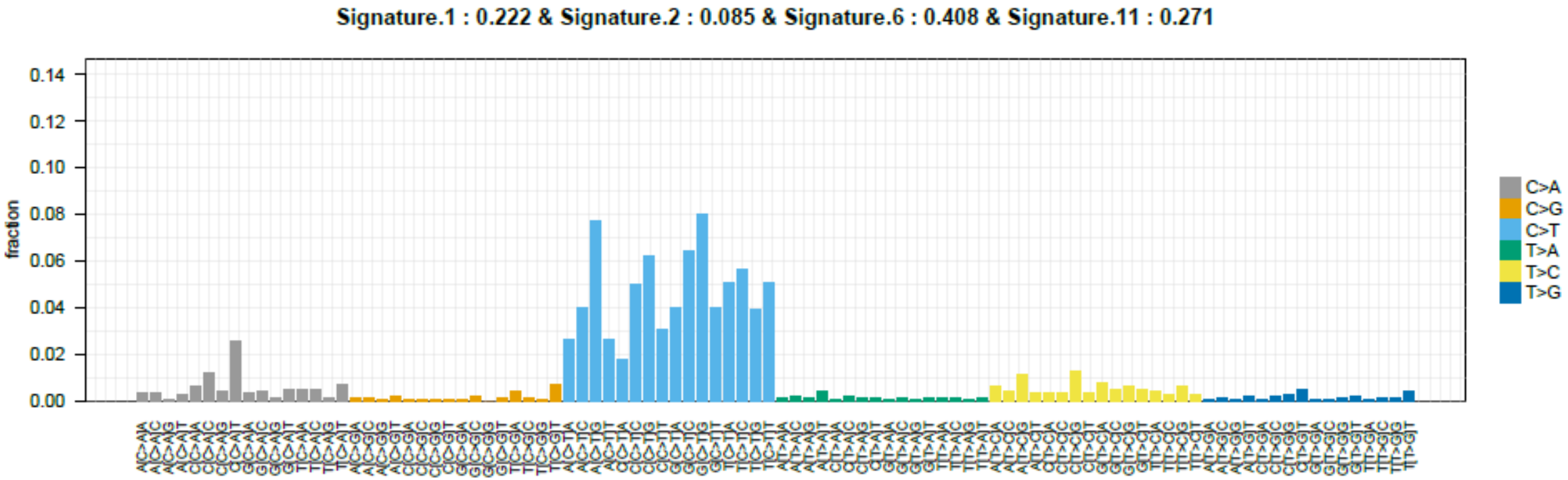
Supplemental Figure 19: Transversion mutations in leukemic and not leukemic samples



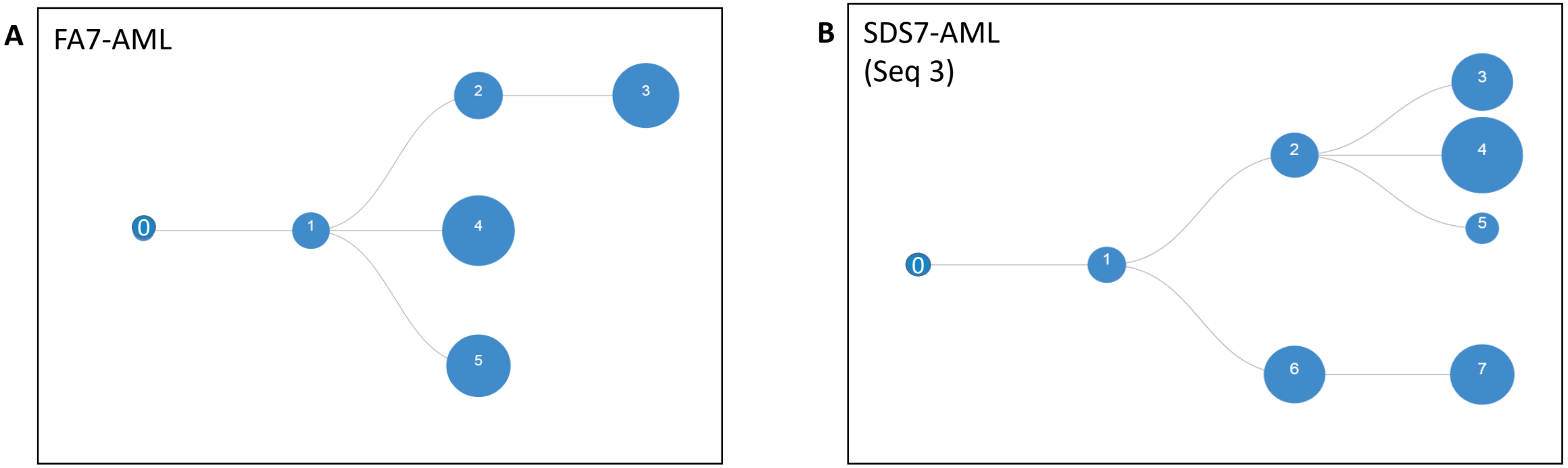
Supplemental Figure 20: Trinucleotide signature generated from data of Fanconi anemia patient 7 (FA7) who developed AML. The signatures were generated by comparing the validated variants to the published signatures of cancer whole exome sequencing databases of the COSMIC published signatures. The trinucleotide signature is displayed by calculating the fraction of each trinucleotide variant relative to the total number of variants in this subject. The nucleotide change is displayed above each 16 trinucleotide block.



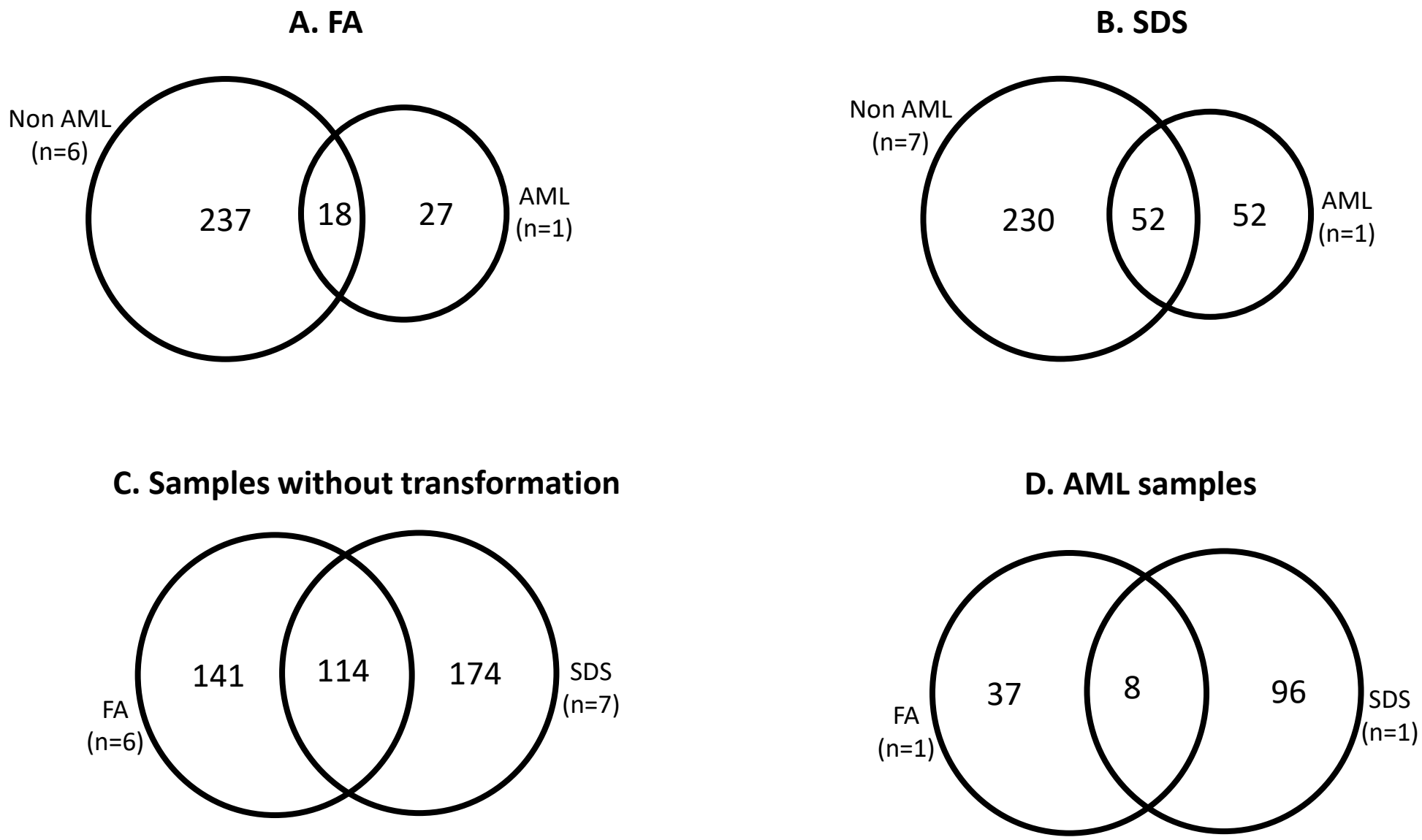
Supplemental Figure 21: Trinucleotide signature generated from data on one of three sequential samples from Shwachman-Diamond syndrome patient 7 (SDS7) who develop AML at the time of the bone marrow sampling used in this figure (See Supplementary Figure 13 for results from previous sampling) . The signatures were generated by comparing the validated variants to the published signatures of cancer whole exome sequencing databases of the COSMIC published signatures. The trinucleotide signature is displayed by calculating the fraction of each trinucleotide variant relative to the total number of variants in this subject. The nucleotide change is displayed above each 16 trinucleotide block.



Supplemental Figure 22 Mutational tree of FA7-AML and SDS-AML (sample with overt leukemia, which is also a sequential sample number 3 from patient SDS7). The detailed list of mutated genes in each clone and sub-clone is in Supplementary Table 6.



Supplemental Figure 23. A vane diagram showing the number of genes with somatic moderate to high impact variants in FA patients with or without AML and in SDS patients with or without AML.



Supplemental Figure 24. Percentage of total transversion mutations in each sequential sample.