JCI insight

Cellular and molecular architecture of hematopoietic stem cells and progenitors in genetic models of bone marrow failure

Stephanie Claudia Heidemann, … , John E. Dick, Yigal Dror

JCI Insight. 2020. https://doi.org/10.1172/jci.insight.131018.

Research In-Press Preview Hematology Oncology

Inherited bone marrow failure syndromes (IBMFSs) such as Fanconi Anemia (FA) and Shwachman-Diamond syndrome (SDS) feature progressive cytopenia and a risk of acute myeloid leukemia (AML). Using deep phenotypic analysis of early progenitors in FA/SDS bone marrow samples we revealed selective survival of progenitors that phenotypically resembled granulocyte-monocyte progenitors (GMP). Whole exome and targeted sequencing of GMP-like cells in leukemia-free patients revealed a higher mutation load than in healthy controls and molecular changes that are characteristic of AML: increased G>A/C>T variants, decreased A>G/T>C variants, increased trinucleotide mutations at Xp(C>T)pT and decreased mutation rates at Xp(C>T)pG sites compared to other Xp(C>T)pX sites and enrichment for Cancer signature 1 (X indicates any nucleotide). Potential pre-leukemic targets in the GMP-like cells from FA/SDS patients included SYNE1, DST, HUWE1, LRP2, NOTCH2 and TP53. Serial analysis of GMPs from a SDS patient, who progressed to leukemia revealed a gradual increase in mutational burden, enrichment of G>A/C>T signature and emergence of new clones. Interestingly, the molecular signature of marrow cells from two FA/SDS patients with leukemia was similar to that of FA/SDS patients without transformation. The predicted founding clones in SDS-AML harbored mutations in several genes including TP53, while in FA-AML the mutated genes included ARID1B and SFPQ. We described an architectural change in the hematopoietic hierarchy of FA/SDS with remarkable preservation of GMP-like populations harboring […]

Cellular and Molecular Architecture of Hematopoietic Stem Cells and Progenitors in Genetic Models of Bone Marrow Failure

*Stephanie Heidemann, MD[1,2], *Brian Bursic[1], Sasan Zandi, MD/PhD[3,4], Hongbing Li PhD[1], Sagi Abelson PhD[4], Robert J Klaassen, MD, FRCP[5], Sharon Abish, MD[6], Meera Rayar MD, FRCP(C)[7], Vicky R. Breakey, BSc, MD, MEd, FRCP(C)[8], Houtan Moshiri[1], Santhosh Dhanraj MSc[1,9], Richard de Borja, PhD[1], Adam Shlien PhD[1] John E. Dick, PhD[4,10], Yigal Dror, MD, FRCP(C)[1,2,9]

[*]Both authors contributed equally to the manuscript and should be considered primary authors.

Affiliations: [1]Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, ON, Canada; [2]Marrow Failure and Myelodysplasia Program, Division of Hematology/Oncology, Department of Pediatrics, The Hospital for Sick Children, Toronto, ON, Canada; [3]currently Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada; [4]Princess Margaret Cancer Centre, Toronto, ON, Canada; [5]Department of Pediatrics, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; [6]Pediatric Hematology Oncology, Montreal Children's Hospital, Montreal, QC, Canada; [7]Division of Hematology/Oncology, UBC & B.C. Children's Hospital, Vancouver, BC, Canada; [8]Department of Pediatrics, McMaster University, Hamilton, ON, Canada; [9]Institute of Medical Science, University of Toronto, Toronto, ON, Canada, [10]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada.

Correspondence to: Dr. Yigal Dror, Division of Hematology/Oncology and the Genetic and Genome Biology Program, the Hospital for Sick Children, 555 University Avenue, Toronto Ontario M5G 1X8, Canada. Telephone: 416 813-5630; fax: 416 813-8327; e-mail: yigal.dror@sickkids.ca.

Abstracts word count	233
Text word count (Manuscript body)	6467
Total words (title page, abstract page, manuscript, figure legends, tables, and references)	11,074
Figures	6
Tables	3
Supplemental Figures	24
Supplemental Tables	8

ABSTRACT

Inherited bone marrow failure syndromes (IBMFSs) such as Fanconi Anemia (FA) and Shwachman-Diamond syndrome (SDS) feature progressive cytopenia and a risk of acute myeloid leukemia (AML). Using deep phenotypic analysis of early progenitors in FA/SDS bone marrow samples we revealed selective survival of progenitors that phenotypically resembled granulocyte-monocyte progenitors (GMP). Whole exome and targeted sequencing of GMP-like cells in leukemia-free patients revealed a higher mutation load than in healthy controls and molecular changes that are characteristic of AML: increased G>A/C>T variants, decreased A>G/T>C variants, increased trinucleotide mutations at Xp(C>T)pT and decreased mutation rates at Xp(C>T)pG sites compared to other Xp(C>T)pX sites and enrichment for Cancer signature 1 (X indicates any nucleotide). Potential pre-leukemic targets in the GMP-like cells from FA/SDS patients included *SYNE1*, *DST*, *HUWE1*, *LRP2*, *NOTCH2* and *TP53*. Serial analysis of GMPs from a SDS patient, who progressed to leukemia revealed a gradual increase in mutational burden, enrichment of G>A/C>T signature and emergence of new clones. Interestingly, the molecular signature of marrow cells from two FA/SDS patients with leukemia was similar to that of FA/SDS patients without transformation. The predicted founding clones in SDS-AML harbored mutations in several genes including *TP53*, while in FA-AML the mutated genes included *ARID1B* and *SFPQ*. We described an architectural change in the hematopoietic hierarchy of FA/SDS with remarkable preservation of GMP-like populations harboring unique mutation signatures. GMP-like cells might represent a cellular reservoir for clonal evolution.

INTRODUCTION

Myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) comprise a spectrum of hematopoietic disorders. Despite intensive chemotherapy and hematopoietic stem cell (HSC) transplantation the overall survival of advanced MDS/AML remains low; ~60% in children and ~30% in adults (1). The outcome is further compromised by treatment-related, long-term adverse events (2).

Hematopoiesis is a complex developmental system that is organized as a hierarchy sustained by multipotent HSCs. Although typically depicted with increasingly restricted oligopotent and unipotent progenitors downstream of the HSC, recent studies demonstrated a reshaping of the architecture of human hematopoietic hierarchy between in-utero fetal liver and adulthood timepoints (3-5). Transcriptional and functional analysis suggests that by adulthood, there is predominantly a two tier hierarchy of multipotent and unipotent HSPCs (5).

AML is a heterogeneous disorder that derives from early HSCP, which undergoes malignant transformation to leukemic blast and clonal expansion. Deep sequencing of leukemic samples extrapolated the existence of founding clones and derived subclones (6). AML is sometimes preceded by MDS. MDS is a clonal pre-leukemic disease state with cytopenia due to underproduction, abnormal differentiation, increased apoptosis, varying degrees of leukemic blasts and a high risk of progression to leukemia. The incidence of both, MDS and AML, increases with age (7), but both can present in early childhood (8, 9).

Several cytogenetic abnormalities have been identified in bone marrow samples from patients with de novo MDS/AML, including -7, +8, del(12q). Genes that are mutated and might be involved in MDS/AML evolution have been recently discovered; for example, RNA-splicing machinery (e.g. *SRSF2, SF3B1, U2AF1*), DNA methylation (e.g. *IDH1*, *IDH2, TET2, DNMT3A*), transcription factors (e.g. *RUNX1*), chromatin modification (e.g. *EZH2, ASXL1*), signal transduction (e.g. *FLT3*), RAS pathway (e.g. *KRAS*), cohesin complex (e.g. *STAG2*) and DNA repair (e.g. *FANCL*) genes (reviewed in(10)). These data advanced our knowledge about MDS/AML pathology; however, the mechanisms underlying clonal initiation and progression are largely unknown.

Although rare, inherited bone marrow failure syndromes (IBMFSs) provide an opportunity to study AML evolution and progression due to a high risk of MDS/AML (11, 12) and stepwise progression from non-malignant hematopoietic phase, to MDS (13) and on to AML (14-16). We previously showed that by the age of 18 years, patients with the more common IBMFSs, Fanconi anemia (FA) and Shwachman Diamond syndrome (SDS), have 75% and 25% risk respectively of developing marrow cytogenetic abnormalities/MDS/AML (11). AML secondary to MDS has particularly poor outcome. Only few studies that focused on clonal hematopoiesis in IBMFSs have been published. *TP53* mutations were identified in some SDS patients with (17) or without MDS/AML (18). *RUNX1* mutations have been detected in whole marrow cells from several patients with FA patients without transformation (19). *CSFR3*(18, 20, 21) and *RUNX1* (22) mutations have been detected in whole marrow cells from severe congenital neutropenia patients with and without MDS/AML. Further studies are necessary to decipher the cells that found transformation and why they abnormally accumulate mutations.

In this study, we aimed to discover cellular and molecular signatures underlying early clonal evolution when no clinical signs of MDS/AML are detected in two relatively prevalent IBMFSs that feature an initial marrow failure phase and frequently progress to MDS/AML: FA and SDS. FA is caused by germline mutations in one of 23 DNA repair genes collectively referred to as the FA pathway (23), and SDS is caused by germline mutations in genes that are involved in late stage of 60S ribosome subunit maturation, *SBDS* (24), *DNAJC21*(25) and *EFL1 (26)*, but also in *SRP54* (27), which is involved in the cotranslation protein-targeting pathway. We found that granulocyte-monocyte progenitor (GMP)-like population are relatively preserved compared to marked exhaustion of other cell populations, they carry a high mutation load and a unique trinucleotide mutation signature; suggesting that GMP-like cells are a reservoir for clonal evolution.

RESULTS

Hematopoietic stem cells and multipotent progenitors are markedly reduced in FA/SDS

We and others showed global reduction in hematopoietic cells and in CD34+ cells in bone marrows from patients with FA (28) and SDS (29). We hypothesized that in both disorders defects begin within the most early hematopoietic cells, and applied 12-paramater deep

immunophenotyping profiling methodology based on recently developed approaches (5, 30)(Figure 1A). Cell numbers were normalized to the viable (propidium iodide-negative) cells in the sample. Within the CD34+CD38-primitive progenitor compartment and compared to healthy controls, the relative numbers of CD90+CD45RA- hematopoietic stem cells (HSC) were reduced 14.1 and 4.6-fold in FA and SDS, respectively, and the CD90-/CD45RA- multipotent progenitors (MPP) were reduced 17.7 and 7.8-fold in FA and SDS, respectively (Figure 1B-C). Since most FA/SDS patients included in this study had hypocellular bone marrow specimens (Supplemental Table 1), we suggest that the average fold decrease in absolute numbers of patients' HSCPs compared to healthy controls is probably higher than that of the above relative numbers.

FA/SDS are characterized by variable levels of oligopotent hematopoietic progenitor loss

CD34+CD38+ progenitors include the common myeloid progenitors (CMP), megakaryocyte erythroid progenitor (MEP) and GMPs. CMPs and MEPs were markedly and significantly reduced in the patients. CMPs were reduced 8.1 and 3.5-fold in FA and SDS, respectively. MEPs were reduced 12.3 and 15.5-fold in FA and SDS, respectively (Figure 1D-E).

Unexpectedly, the reduction of HSCs did not result in universal reduction of all their downstream progenies. In SDS, MEPs represented the most affected population compared to CMPs or GMPs. In FA, MEPs and CMPs were markedly reduced compared to GMPs. Furthermore, in both SDS and FA, GMPs (CD34+/CD38+/FLT3+/CD45RA+) were least affected and relatively preserved, with only 1.5-fold reduction in SDS and 2.3-fold reduction in FA. In SDS, the percentages of GMPs were not significantly different from controls (Figure 1D-E). Remarkable, when HSCP frequencies were normalized to the total number of CD34+ cells in the respective sample, the average percentage of SDS GMPs was a modest 1.56 fold higher than the average percentage of healthy controls' GMPs ($p=0.03$). In FA, the average GMPs was 1.15 time higher than that of controls, but the difference did not reach statistical significance (Supplemental Figure 1). These data about FA/SDS GMPs were surprising for both disorders, but particularly in SDS, as granulopoiesis is the most affected series in SDS (29, 31).

FA/SDS feature abnormally high frequency of somatic variants in GMPs

The IBMFS are difficult to study genetically since there is a paucity of cells to work with. Therefore, we undertook genetic analysis to gain insight into the mutations present within the GMP population that seemed to be persisting more extensively than other progenitors. In addition, because of the relative abundance of GMP-like cells we reasoned that they are more likely to carry mutations that confer growth advantage than other progenitors that were markedly reduced.

We analyzed somatic tier 1-2 variants in GMP-like cells, as described in the Methods. Bone marrow fibroblasts were used as a surrogate germline tissue. The cogency of variant detection was supported by a high congruence of mapped reads across the genome (Supplemental Figure 2) and per chromosome (Supplemental Figure 3). Analysis of a marrow fibroblast sample demonstrated that this congruence was seen between amplified and non-amplified DNA before whole exome sequencing (WES). Importantly, we consistently saw lower variant numbers when GMPs were compared to self fibroblasts versus fibroblasts from other subjects; which is expected given normal genomic variations between individuals (Supplemental Figure 4). There was a consistently higher number variants in patients versus controls who were processed and analyzed in an identical fashion (see below). Detection of calls by MuTect2 and by other mutation caller software programs (Sterlka and VarScan) was also highly congruent (Data not shown). In addition, there was no correlation between gene size and number of variants detected, which would be expected from random mutations along the genome. Also, we found no aberrantly high rates of C>T (G>A) errors in analysis of a GMP DNA sample compared to a blood DNA sample amplified by single cell REPli-G whole-genome amplification kit and by VarScan mutation caller software (Data not shown). Last, detecting variants by WES and the Cancer gene panel showed high congruence (Supplemental Table 2).

The numbers of somatic variants among FA patients (mean 111) and SDS patients (mean 108) were remarkably higher than that among control subjects (mean 25), whose samples were processed in the same way (P values of 0.04 and 0.02, respectively)(Figure 2A). All variants were rare (minor allele frequency of $\leq$1%) or absent in the general population's databases (Data not shown). There was no significant age difference between FA/SDS patients and controls

(p=0.34 and p=0.41, respectively). Interestingly, the frequency of variants in FA was not statistically different from SDS (Figure 2A).

The total number of variants in each subject according to age at sampling are in Figure 2B-D. Statistically significant correlation between mutation burden and age could not be accurately determined since a larger number of subjects in each group are required for this analysis. Importantly, the variants in SDS/FA appeared in significantly higher allele frequencies compared to those of controls ($p<0.0001$)(Figure 2E).

Types of nucleotide change across patients

Due to their AML-predisposition, we reasoned that mutations in FA/SDS GMP-like cells are characterized by previously published AML mutational patterns. Therefore, we employed multiple analytical techniques to understand the mutational process and patterns underlying the high mutational load in FA/SDS. First, we determined the variants underlying transition changes (interchanges between purines bases or between pyrimidine bases, Figure 3A) and transversion changes (interchanges between purine and pyrimidine bases, Supplemental Figure 5). We found that the most abundant single nucleotide variants (SNVs) in all groups (FA, SDS and controls) was as seen in AML (32); namely G>A/C>T transitions, followed by A>G/T>C transitions and G>T/C>A, C>G/G>C transversions. Nevertheless, the proportions of G>A/C>T transitions in FA/SDS were significantly higher than those of control subjects ($p<0.05$).

To gain further insight into the mutational processes in FA/SDS, we analyzed variants in the context of a trinucleotide change: the six options of nucleotide substitutions and the 16 combinations of bases immediately 3' and 5' to this variant. Overall, this results in a mutational signature that comprises 96 different trinucleotide frames for each subject that are displayed in a heatmap in Figure 3B. All the subject groups showed a high C>T mutation rate regardless of the flanking 5' and 3'-nucleotide (Xp(C>T)pX sites). However, but this propensity was much more prominent in FA (p=0.04) and SDS (p=0.02) patients than in controls. The visualization of vertical rows on the heatmap suggests that the 3' base has a greater influence on the mutational pattern. The vertical rows seen within the C>T region indicate that most patients have greater mutation rates at Xp(C>T)pT and lesser mutation rates at Xp(C>T)pG sites (Arrows in Figure

3B) compared to other Xp(C>T)pX sites. This pattern was less prominent in healthy control subjects. The low amount of mutations seen at Xp(C>T)pG sites may be attributed to the relatively low amount of CpG sites in the genome and could be the result of the deamination of methylated cytosines (33). Last, there was a modestly increased mutational load at T>C sites in FA/SDS.

Different cancers generate mutations through distinct processes and leave their mark on the genome through a unique mutational signatures (34). To identify the specific cancer trinuceotide signature of GMP-like cells from each subject, we first normalized variants to the relative contribution of each trinucleotide in the exome region using the DeconstructSigs R package and then compared our results to those in the Catalogue of Somatic Mutation in Cancer (COSMIC) database. Normalization entails determining the amount of a certain trinucleotide variant relative to the amount of native trinucelotides occurring within the respective genome. De novo AML has previously been characterized by the COSMIC database to have a trinucleotide pattern contributed by Signatures 1 (spontaneous deamination of 5-methylcytosine) and 5 (transcriptional strand bias for T>C substitutions at ApTpX context). Due to a minimum 50 variant criterion for analysis, cancer signatures could be constructed from 9 of the 14 FA/SDS GMP-like cell samples, but from none of the control subjects (Supplemental figure 6-14). Importantly, the AML signature 1 was more frequent (8 of the 9 patients), and more often the dominant signature (4 of the 9 patients) than other signatures (Supplemental Table 3).

The analysis of tier 1-2 SNVs and indels predicted varying degrees of damage to the encoded protein from stopgain, frameshift, start lost, splicing and missense alterations to potentially less severe effect of 3’ UTR, 5’ UTR and synonymous changes (Figure 3C). The distribution of mutation types for patients was similar to controls although the rates of mutations was higher.

Mutated genes and mutational trees

To identify genes that might be involved in malignant myeloid transformation and to construct mutational trees, we selected genes with mutations that fulfilled the criteria described in the Methods and had moderate to high software-predicted impact on the protein; namely, nonsense,

splicing, frameshift, indel/inframe, start site loss and missense. The average number of mutated genes per subject was significantly higher in FA (61) and SDS (58) compared to controls (14) (Figure 3D); but was not statistically different when FA and SDS were compared.

Importantly, there were a substantial number of genes with moderate to high impact mutations in more than one FA/SDS patient (Table 1). Commonly mutated cancer-related genes in both, FA and SDS, included the nuclear membrane gene *SYNE1* and the ubiquitin E3 ligase gene *HUWE1*. Genetic mutations or dysregulation of these genes have previously been implicated in several solid tumor such colon and gastric cancer, though not in leukemia. Several known MDS/AML-driver genes were recurrently mutated in SDS (e.g. *ASXL1*, *TP53* and *CUX1*) and FA (e.g. *BCOR*)(Supplemental Table 4). It is noteworthy that mutations in the TP53 binding protein 1 gene, *TP53BP1* (p.Asp11Asn and p.Val687Ile) were seen in two SDS patients. The number of variants in mutated genes was not related to the gene size (Supplemental Figure 15), indicating a non-random distribution of mutations.

Compilation of a dominant mutational tree in samples without clinical evidence of transformation was performed as described (35) in all FA (Supplemental Figure 16A-F) and SDS samples (Supplemental Figure 17A-G). The specific genes and variants in each clone are listed in Supplemental Table 5. In all samples there were mutations in known MDS/AML genes and in other cancer-related genes that have not previously been reported in MDS/AML. Interestingly, in two FA samples the founding clones harbored somatic mutations in MDS/AML-related genes (*KDM6A* in FA3 and *FANCE* in FA5); while in the rest of the FA samples the founding clones harbored cancer-related genes that have not been previously associated with MDS/AML. *TP53* mutations were part of the founding clones in two SDS patients (SDS1 and SDS5) (Supplemental Table 5), but in none of the samples of FA patients without leukemia. Other MDS/AML related genes were identified in the founding clones in three other patients with SDS (Supplemental Table 5).

Analysis of MDS/AML-related gene pathways showed high rates of mutations in the transcription factors/regulation pathway, DNA repair/checkpoint gene pathway, and activated signaling molecules pathway in FA/SDS (Supplemental Figure 18).

Clonal landscape of AML samples in FA/SDS

To gain insight into the relevance of variants and mutated genes detected in samples without transformation, we analyzed leukemic cells from one FA (FA7) patient with AML and one SDS patient (SDS7) with AML. Although only two AML cases from these rare disorders were available for the study, these anecdotes provide a unique opportunity to observe processes that appeared at two stage: before any clinical and standard laboratory evidence of transformation and at an ultimate catastrophic phase of leukemia. Blast cell samples were paired with marrow fibroblasts or T-cells from the same subject, and somatic variants in blasts were analysed as described in the Methods. The mutations rate in SDS-AML blasts was slightly higher than the rates in all other SDS samples without transformation, but the number of variants in FA-AML blasts was within the range of those in non-transformed FA samples (Figure 4A).

Similar to our findings in non-AML samples, both FA/SDS samples showed higher G>A/C>T transition rates than controls; the predominant mutation type in de-novo AML (Figure 4B)(34). The number of transversion was low (Supplemental Figure 19), and meaningful comparison between transformed to non-transformed samples was impossible.

The trinucleotide heatmap depicting the variant change and adjacent 5’ and 3’ bases in non-AML and AML patients is in Figure 4C. All samples, including AML-blasts and GMPs from subjects at no transformation, featured high mutation rates at Xp(T>C)pX sites. Importantly, the AML signature 1 was the predominant trinucleotide signature in FA-AML blasts (64%), and comprised a substantial faction also in SDS-AML blasts (22%) (Supplemental Figure 20-21).

Analysis of the potential impact of mutations on the protein showed a generally similar pattern in FA/SDS-AML samples compared to those without AML (Figure 4D).

Cancer-associated genes with moderate to high impact mutations in AML samples are listed in Table 2. The genes with the highest variant allele frequency are in Figure 5A-B. Several genes harbored variants with high frequency in FA-AML and were predicted to be part of the founding clone by mutational tree analysis (Supplemental Figure 22A and Supplemental Table 5). These genes were *ARID1B*, *SFPQ*, *PCDH15*, *EPPK1* and *MAP2K1*. The founding clone gave rise to three sub-clones that included mutations in the MDS/AML genes *NUP98*, *PML/BRCA1* and *TP53/BRCA2*, respectively. The first clone gave rise to an additional clone with mutations in the *CREBBP* MDS/AML associated gene.

The genes that appeared in highest allele frequency in SDS-AML included *MYH1*, *TP53*, *FLT4*, *LPHN3* and *DICER1* (Table 2). These genes were predicted to be part of the founding clone, which gave rise to two sub-clones (Supplemental Figure 22B and Supplemental Table 5). The mutated genes in one of the sub-clones included the MDS/AML gene *PTPRD*, and other cancer genes (e.g. *JAK1* and an additional mutation in *DICER1*). This sub-clone gave rise to additional clones harbouring mutations in MDS/AML genes such as *BRAF* and *SETD2*. The second sub-clone featured a mutation in *SFPQ* and subsequent clones included mutations in *NCOR1*, *SMAD4*, *NF1* and *BRCA1*. Similar to samples without AML (Supplemental Figure 18), in both, FA-AML and SDS-AML, commonly mutated pathways included the transcription factor or transcription factor regulation and DNA repair (Figure 5C).

Last, we evaluated whether genes with high/moderate impact mutations that appeared in patients without transformations were also mutated in the AML phase. In FA, 18 of the 255 genes that were part of clonal hematopoiesis in patients without MDS/AML appeared in the AML blasts (Supplemental Figure 23 and Supplemental Table 6). In SDS, 52 of the 282 genes that were part of clonal hematopoiesis in patients without transformation appeared also in the AML blasts (Supplemental Figure 23 and Supplemental Table 6).

Clonal evolution and progression observed in sequential samples

From the SDS patient who developed leukemia, two additional samples 36 months and 25 months prior to the development of AML were available. The number of mutations grew prominently from stage to stage (Pearson R value 0.99)(Figure 6A). The growth was more

prominent than the age-related mutation increment we found in our SDS patient cohort (Figure 2C). Interestingly, there was a gradual increase in G>A (R= 0.99) and C>T transitions (R= 0.99938), but not in A>G or T>C transitions (Figure 6B). There was also a gradual accentuation of the trinucleotide signature (Heatmap in Figure 6C). The number of transversions was low (Supplemental Figure 24), and did not show a conclusive pattern.

Construction of trinucleotide cancer signatures using the COSMIC database was feasible for the last 2 sequential samples. Interestingly, signature 1 accounted for 9.2% of the mutational signature in the second sequential sample (Supplemental Figure 14) and increased to 22.2% at the stage of AML (Supplemental Figure 21).

Similar to the variant numbers, there was also a gradual increase in the number of genes with moderate/high impact mutations in each sequential sample; 15, 65 and 103 respectively (Table 3) (R=0.945). Of the 15 genes with mutations in first sequential sample, two were mutated in the second and third samples. Of the 65 mutated genes in second sequential sample, 10 were mutated in the third sample.

In each of the sequential samples a dominant mutational tree could be constructed. However, as seen with bone marrow cytogenetic abnormalities in FA (36) and SDS (37), the dominant tree may arise and regress, and in each sequential sample a different dominant tree was apparent. The founding clone in sequential sample 1 harbored 13 genes with high/moderate impact mutations including *ARHGEF12* and *NOTCH2*; in sequential sample 2 there were 28 such genes, including *IDH2* and *MYH2*; in the third sample (AML) there were 6 such genes including *TP53*. The known pathogenic mutation in *TP53* (c.742C>T; p.Arg248Trp) was dominant in the AML stage (52%). It is noteworthy that with progression from sequential sample 1 to 3, the proportion of mutations in transcription factors, transcription factor regulation, activated signaling molecules and DNA repair and checkpoint molecules increased (Figure 6D).

DISCUSSION

The present study focused on evaluating the cellular and molecular events before overt leukemia develops and their potential impact on malignant transformation. We report for the first time,

detailed analysis of the very early hematopoietic cells (HSC, MPP) and subsequent progenitors (CMPs, MEPs, GMPs) in FA and SDS. Most HSCPs were markedly reduced except for GMPs, which were much more preserved. Molecular analysis of phenotypically GMP cells, revealed a high number somatic mutations compared to control subjects and genetic signatures that resembled those seen in AML. Using sequential SDS samples before and at AML stage we were able to show that somatic nucleotide-level mutations develop and disappear very rapidly in this disorder; resembling observations related to some large clonal marrow cytogenetic abnormalities (36, 37). The reconstructed founding clone at the AML stage harbored mutations in several genes including in *TP53.*

The overrepresentation of immunophenotypic GMPs versus other myeloid progenitors in FA/SDS patients suggests that these cells feature higher survival or growth properties and possibly harbour some of the initial transformational events that lead to MDS/AML. We cannot rule out the possibility that relative preservation of GMP-like cells reflects a general compensatory mechanism for the bone marrow failure unrelated to leukemia risk. Although possibly, it would be surprising that a compensatory mechanism targets GMPs regardless of whether the mostly affected lineage is granulocytic (SDS) or megakaryocytic/erythrocytic (FA). It is noteworthy that the initiating events may occur in an earlier HSPC, which then acquire the immunophenotype of GMPs. The markedly elevated somatic variants in FA/SDS GMP-like cells is in keeping with this hypothesis. It is possible that some of these mutations enhance proliferation or inhibit cell death; thereby confer grow advantage to these progenitors. For example, the *TP53* mutation p.Arg248Trp seen in SDS patients inactivate the protein and its proliferation regulatory properties. Future studies are necessary to decipher the mechanism underlying the relative preservation of GMP-like cells in FA/SDS bone marrows and whether it is related to increased proliferation, decreased apoptosis or self-renewal.

Interestingly, despite different functions of FA genes from SDS genes, in both conditions GMPs were relatively more preserved and there were no significant differences in the average number of somatic mutations. This raises the possibility that at least in part clonal evolution in bone marrow failure disorders does not depend on the direct biochemical sequela of the germline

mutation, and might be related to the consequent growth disadvantage of bone marrow cells, mitotic stress and a drive for survival through growth-promoting somatic mutations.

The cause of an increased propensity for MDS/AML in IBMFSs and the mechanisms of leukemogenesis are unclear, and several hypotheses have been proposed (38, 39). Our findings of an increased mutation rate in GMP-like cells and their relative preservation provide a groundwork for research focusing on these questions. Several pathological processes have been identified in FA/SDS, and may be considered while trying to explain an increased risk of somatic mutations. FA proteins are involved in correction of interstrand DNA crosslinks (40) and telomere length maintenance (41) leading to chromosomal instability. There is also evidence for short telomeres and genomic instability in SDS (42, 43). These pathologies may lead to somatic structural chromosomal abnormalities that are commonly seen in SDS (37, 43, 44) and in FA (45, 46); however, they may not directly explain the increased numbers of SNVs seen in our study. In FA, DNA interstrand crosslinks may lead to DNA double-strand breaks due to prolonged stalling of the replication fork or collapse. This may eventually lead to errors during repair or replication. Oxidative stress has been implicated in causing DNA damage and cancer development (47, 48), and is increased in both, FA (49-51) and SDS (52, 53). In addition, the accelerated cell death and slow-growing cells in FA (49, 50, 54) and SDS (55-57) may lead to replicative stress, which can consequently increase the rate of randomly occurring mutations. Interestingly, it has been suggested that the slow-growing HSCPs in bone marrow failure disorders are under selective pressure for mutations that reverse their growth defect and ameliorate the restraints on proliferation (58, 59). Last, similar to AML (60) and MDS (61), SDS bone marrow stroma features increased angiogenesis (62). SDS bone marrow stroma has also been shown to be functionally impaired in humans (29) and in mice (63). In the latter study, deletion of *sbds* in mouse mesenchymal stem cells resulted in DNA damage in HSCPs and in pro-inflammatory response that was shown to contribute to leukemic transformation (63).

To our knowledge, there are no published data about the rate, type and signature of somatic variants in GMPs from inherited leukemia-predisposition syndromes, and only little information is available about somatic mutations in bone marrow samples from patients with FA (19) and SDS (17, 18). An explicit comparison between results from the present work to those from

previously published studies on IBMFSs is challenging due to different methodologies and analytic approaches. Nonetheless, the number of variants in our study might be different from that reported in few published papers on FA/SDS, and there are several possible explanations for that. First, mutation rates in GMP-like cells have not previously been published. GMP-like cells were relatively preserved in FA/SDS, which might be attributed to higher rate of somatic mutations that confer growth advantage. Second, published studies focused on mutations with high allele frequency; for example, in the study on somatic mutations in FA patients (19), mainly Sanger sequencing was utilized; the technique typically detect variant with allele frequency of over 10-20%. In the published WES data on two patients with SDS (18), few variants were reported; however, the authors focused on variants at the expected binomial distribution around 50%. Due to the analysis of highly purified progenitors and limited number of progenitors in FA/SDS, we used amplified DNA. Quality assessment of the data, paired analysis of amplified and non-amplified DNA from control marrow fibroblasts across the genome (described in the Results section) and our internal robust methodology suggest that the trends seen herein are real and that significant bias by the DNA amplification is unlikely.

The molecular changes found herein in FA/SDS GMP-like cells are reminiscent of those seen in AML; for example, abundance of G>A/C>T and G>T.(32) G>A/C>T hypermutations have been attributed to the endogenous process of deamination at methylated cytosine sites (32). Importantly, this pattern was also dominant in FA/SDS-AML samples and steadily increased in sequential samples from a SDS patient who eventually developed AML. Studies of sequential samples from additional cases are needed to determine whether gradual acquirement of this pattern is indeed part of the transformational process in FA/SDS.

The characterization of mutational signatures unveils a new hypothesized mutational aetiology that could give insight into the mutational processes underlying leukemic predisposition in FA/SDS. By the COSMIC, AML features a trinucleotide pattern contributed by Signatures 1 (spontaneous deamination of 5-methylcytosine and increased mutations at CpG sites) and signature 5 (transcriptional strand bias for T>C substitutions at ApTpX sites). To our knowledge, the COSMIC trinucleotide signature database has not been previously applied to FA/SDS bone marrow samples. Out results suggest that GMP-like cells are prone to develop AML-type

trinucleotide signature in FA/SDS. This hypothesis is solidified by finding Signature 1 in AML cells derived from FA/SDS patients, and by observing an increment in the proportion of Signature 1 in sequential samples from a patient with SDS who eventually developed AML. The predominance of signature 1 indicates that deamination of methylated cytosines plays a role in the mutations seen in FA/SDS; however, mutational processes related to the other concomitant signatures may also be in play.

In most samples we were able to reconstruct a dominant mutational tree. However, most mutations were not part of the dominant mutational tree, suggesting that FA/SDS marrows contain multiple non-related clones. Further, we cannot rule out a possibility that at the stage of AML, additional smaller non-related AML clones co-existed. Importantly, using sequential samples, we found that similar to large cytogenetic abnormalities that may appear and disappear with time in FA (36) and SDS (37) (including del(20q10-11) and i(7q)), SNVs may also appear and disappear; as described in one patient with severe congenital neutropenia (21). Our study further shows that most clones do not culminate in leukemia evolution, and despite a burst of evolving clones, most of them disappear and become outnumbered by new clones. This process probably continues until a combination of critical mutations appear in the same clone and drive progression towards MDS/AML.

It is noteworthy that the frequency of mutations in genes, which are commonly mutated in de-novo MDS/AML (e.g. *DNMT3A, TET2* and *SF3B1*), was low in FA/SDS patients, particularly in the cases who developed AML, suggesting that transformation in FA/SDS may utilize novel mechanisms. *PCDH15* was mutated in FA-AML with high VAF (69%) and was predicted to be part of the founding clone of the dominant mutational tree. *PCDH15* is a member of the cadherin superfamily, which encode integral membrane proteins that mediate calcium-dependent cell-cell adhesion. It is mutated in several solid cancers including breast cancer, glioma and lymphoma (64-66). The findings of mutations in this gene also in two SDS patient without AML (one of them in the founding clone) suggest a potential pathogenic role.

It is noteworthy that *SFPQ* was mutated in both our AML patients; in the founding clone in FA and in a sub-clone in SDS. To our knowledge, *SFPQ* was previously reported to be mutated only

in one AML subject (67). A recent study suggested downregulation of *SFPQ* by MicroRNA-1296 in colorectal cancer as a mechanism for cell proliferation (68). The published mutation in an AML patient was described as non-synonymous without further details. The mutation in our FA-AML patient was a missense variant in the N-terminal domain (p.Gly14Ser). The mutation in the SDS-AML patient was a missense variant in the C-terminal domain (p.Glu699Lys). It is possible that loss or aberrant SFPQ alters splicosome function and drives MDS/AML. Further studies are necessary to determine whether *SPFQ* mutations are more common in IBMFS-associated MDS/AMLthan in de novo MDS/AML, and whether there is synergism between hematopoietic stem cell loss and SFPQ in developing leukemia.

It is important to note that *TP53* was mutated herein mainly in SDS patients. It was mutated in the SDS-AML founding clone and in two SDS patients without transformation, indicating that it is indeed an early transformational event. It is interesting that in sequential samples, the *TP53* mutation p.Arg248Trp (previously reported as pathogenic) was detected in the founding clone of SDS-AML but not in the founding clone in previous samples. This information supports the notion that early hematopoietic cells in IBMFSs have heightened tendency for clonal evolution, but most clones eventually subside and do not progress.

In summary, FA/SDS are characterized by a burst of clonal evolution. Although the molecular changes largely follow AML features, most hematopoietic clones do not progress, and at a leukemic stage only a few clones become predominant. The differences between clones that progress to leukemia and those that do not need to be further elucidated. Future studies should also evaluate the prognostic value of the identified molecular changes in this study and their potential use for early detection of irreversible transformation or therapeutic targets in FA and SDS. Last, since AML blasts from only two FA/SDS patients were available for this study, the molecular data at the AML stage are anecdotal and multicenter collaborative efforts are required to collect larger number of AML samples from these rare disorders to validate our observations.

METHODS

Flow cytometry

Bone marrow HSCP population sizes were evaluated by multi-parametric immunophenotyping (Figure 1A) as previously described (5). Cell frequencies were normalized as previously described to the total bone marrow mononuclear cells (5, 69) and to total bone marrow CD34+ cells (70, 71).

DNA preparation for genomic studies

To identify the spectrum of somatic mutations and affected genes, we analyzed DNA from phenotypical sorted GMP cells. DNA samples from 200-965 sorted GMPs were amplified by whole genome amplification (REPli-G Mini Kit) for 16 hours with adjustment of reagents to cell number as per the manufacturer instructions and as previously described (72-75).

To eliminate germline variants we paired each subject's data with their marrow fibroblast genome as a source of non-hematopoietic DNA. We enriched marrow fibroblasts by culturing marrow cells, removing floating hematopoietic cells and passaging 3-5 times. Due to poor growth of passaged patient cells, DNA of marrow fibroblasts from close to half of the patients (and one healthy subject for quality control) was amplified, with no apparent effect on the number of filtered somatic variants (Supplemental Table 7) and no apparent bias toward specific nucleotide change (Supplemental Table 8). Furthermore, matched amplified and non-amplified DNA from fibroblasts showed a high congruence of mapped reads across the genome and per chromosome (Supplemental Figure 2-3).

To study molecular events in AML samples, we sorted blast cells. In a case of SDS patient with AML, amplified DNA from marrow myeloblasts underwent paired analysis with DNA from marrow fibroblasts. In a FA patient with AML, peripheral blood sample was available and amplified DNA from myeloblasts underwent paired analysis with amplified DNA from T-cells.

Whole exome sequencing

DNA underwent exome enrichment by the Sure Select 50 Mb Human All Exon Capture Kit (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instructions, and sequenced on the Illumina HigSeq2500 at The Center for Applied Genomics (The Hospital for Sick Children, Toronto) as previously described (25). The average reads per nucleotide among the analyzed subjects was 146 (range 116-189).

Next generation sequencing cancer gene panel assay

To augment mutation discovery by deep variant analysis and validate variants in cancer-related genes found by WES, we used a deep sequencing panel of 877 genes, which were either known cancer-related genes from the COSMIC or are hypothesized to play a role in cancer (based on published expression in tumors, known function or constitutive mutation in cancer-predisposition syndromes). The total number of bases for non-overlapping exons covered by the panel +/- 10 bp is 3,012,823 base pairs. The panel was developed by our group (AS) as previously described (76). The average reads per nucleotide among the analyzed subjects was 1216 (range 775-2098).

Variant calling

Somatic variant calling followed the bcbio pipeline (https://github.com/bcbio/bcbio-nextgen/). The pipeline is used to identify somatic variants by comparing them to normal human genome alignments and annotating the mutations for subsequent analysis. The pipeline includes alignment of FASTQ files to the reference genome (GRCh37) using Burrows-Wheeler Alighner (BWA) mem v0.7.17 (http://bio-bwa.sourceforge.net), duplication of marking using biobambam v2.0.87 (https://github.com/gt1/biobambam), and removal of low complexity regions by bedtools v2.27.1 (https://github.com/arq5x/bedtools2).

GMP and marrow fibroblast FASTQ files were aligned and mapped separately to the reference genome to create binary alignment map (BAM) files, and both BAM files were then processed using MuTect v1.1.5 (https://www.broadinstitute.org/cancer/cga/mutect) for somatic point mutations and indels.

Variants from GMP WES and Cancer Panel sequencing were selected as true somatic variants if 1) they appeared in GMPs from both, WES and the Cancer Panel, 2) the variant frequency in marrow fibroblasts was 0, 3) the variant comprised >7% of the total reads for the respective nucleotides in GMPs; using this threshold, >90% of the variants fulfilled all criteria in both, WES and Cancer Panel (Supplemental Table 2), and 4) the read depths by the Cancer Panel in GMP and in the marrow fibroblasts was >50.

Analysis of somatic variants

Somatic variants were classified into tiers as described (77). As conventionally done in cancer genomics analysis, we used only tier 1 and 2 variants, which are more likely to have a pathogenic effect than tier 3 and 4 variants.

The R package, deconstructSigs (https://github.com/raerose01/deconstructSigs) was used to construct tumor signatures from somatic variants, to normalize signatures according to variant frequencies and compare them to known tumor signatures in COSMIC. A mutation signature is determined by comparing the total variant profile of a patient to the known variant profile of different cancers. For this analysis a minimum of 50 somatic variants per sample is required to construct a signature. ComplexHeatmap (https://bioconductor.org) was employed to create sample heatmap of somatic variants. Variant Effect Predictor (http://grch37.ensembl.org/info/docs/tools/vep/index.html) was used to annotate the mutations for functional consequence.

Mutational trees were reconstructed by the PhyloWGS software program as developed by Dr. Quaid Morris' group (35)(https://github.com/morrislab/phylowgs). The program can reconstruct related clonal subpopulations in tumor samples from WGS/WES data. It is based on variant allele frequencies of the mutations and uses the Markov chain Monte Carlo procedure. It can construct mutational trees with or without data about CNV (https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0602-8). Using this software we designated marrow fibroblasts cells as molecular group 0. Subsequent clones were ordered and numbered by the software program.

Statistical analysis

Descriptive analysis was used to characterize groups. Two tail Student's t-test was used to determine the statistical significance of differences between two means. To determine significant differences between multiple means, the non-parametric Kruskal-Wallis test was performed followed by Dunn's post hoc test. Wilcoxon signed-rank test was used for testing whether 3 samples have different VAF distributions. P value of <0.05 was considered significant. The statistical analyses were performed using Excel or XLSTAT Version 2019.1.2 (Addinsoft, New

York, NY) or GraphPad v8. The bioinformatics software programs used in this study are described with the respective analyses in the Methods and Results sections.

Study approval

SDS patients were eligible for the study if they fulfilled the international consensus diagnostic criteria (78) and had biallelic *SBDS* mutations. FA patients were eligible if they had clinical diagnosis of FA and positive chromosome fragility testing. At the time of testing, most patients without leukemia had cytopenia and hypocellular bone marrow (Supplemental Table 1); no patient had clonal marrow cytogenetic abnormalities. Healthy control subjects were donors for bone marrow transplantation. The study was approved by the Research Ethics Board and informed written consent was obtained from all enrolled subjects. Usage of a sample that had been cryopreserved in the Tissue Bank at The Hospital for Sick Children was according to the Research Ethics Board's regulations and approval. A total of 7 FA, 8 SDS, and 8 healthy control subjects were studied. The list of subjects and samples is in Supplemental Table 7.

AUTHOR CONTRIBUTIONS

SH contributed to design, acquisition of data, analysis and interpretation of data, and assisted in writing the manuscript. BB contributed to design, acquisition of data, analysis and interpretation of data, drafted the article and revised it for important intellectual content. SZ contributed to design, acquisition of data, analysis and interpretation of data and assisted in writing the manuscript. HL contributed to design, acquisition of data, analysis and interpretation of data and assisted in writing the manuscript. SA contributed to analysis and interpretation of data, reviewed/revised the manuscript for important intellectual content. RJK heads one of the Canadian Inherited Marrow Failure Registry site research team that contributed acquisition of vital data and interpretation of data; reviewed/revised the manuscript for important intellectual content. SA heads one of the Canadian Inherited Marrow Failure Registry site research team that contributed acquisition of vital data and interpretation of data; reviewed/revised the manuscript for important intellectual content. MR heads one of the Canadian Inherited Marrow Failure Registry site research team that contributed acquisition of vital data and interpretation of data; reviewed/revised the manuscript for important intellectual content. VRB heads one of the Canadian Inherited Marrow Failure Registry site research team that contributed acquisition of vital data and interpretation of data; reviewed/revised the manuscript for important intellectual content. RDB contributed to design, analysis and interpretation of data. HM contributed to analysis, generation of figures and interpretation of data. SD contributed to design, interpretation of data, and review of the manuscript. AS contributed to design, analysis and interpretation of data and assisted in writing the manuscript. JED contributed to design, analysis and interpretation of data and assisted in writing the manuscript. YD contributed to conception, design, acquisition of data, analysis and interpretation of data; drafted and revised the article.

CONFLICT OF INTERST

The authors have no conflict of interest to declare.

ACKNOWLEDGMENTS

The work in the manuscript was support by grants from the Canadian Institute of Health Research application number 377600, from the Canadian Institute of Health Research application number 378100 and from Fanconi Anemia Canada.

FIGURES

Figure 1 **Deep immunophenotyping revealed striking loss of most, but not all, hematopoietic stem cells and progenitors in bone marrows from FA/SDS patients. A**. Analytic strategy of bone marrow aspirate cells by immunophenotyping. **B-C.** Comparison of multipotent cells between FA (n=6), SDS (n=7) and control (n=8). The mean percentage of each HSCPs among the viable bone marrow mononuclear cells is presented with standard error of mean (SEM). **D-E.** Comparison of oligopotent progenitors between FA, SDS and control patients. The mean percentage of each HSCPs among the viable bone marrow mononuclear cells is presented with standard error of mean (SEM). HSC, hematopoietic stem cell; MPP, multipotent progenitor cell; MEP, megakaryocyte erythroid progenitor; GMP, granulocyte monocyte progenitor. Student's t-test was used to compare between patients and controls. The same control data in Figures 1C and 1E are also presented in Figure 1B and 1D, respectively.

Figure 2 **Frequency of somatic variants in bone marrow samples from patients with SDS, FA and healthy control subjects. A**. Comparison of average (+/-standard error of the mean) variant rate between FA (n=6), SDS (n=7) and healthy control subjects (n=6). Results by student T-test are shown. P value by Kruskal-Wallis test with Dunn's post hoc test was 0.0695. **B**, **C** and **D** show the variant rate among controls, FA and SDS subjects organized according to ages. **E**. Allele frequency of the various variants in controls, FA and SDS. The group were compared using the Wilcoxon signed-rank test. Asterisk (*) indicates statistical significance compared to controls with $p<0.0001$. The Y-axis represent the variant frequency and the X-axis represent the variants arranged from those with the highest allele frequency to the lowest. In each group, each number may represent a different variant.

Figure 3 **Patterns of single and tri-nucleotide alterations among FA, SDS and healthy control subjects. A**. Average number (+/- standard error of the mean) of each transition (inside the CT purine group or inside the GA pyrimidine group) variant

per subject among the FA, SDS and healthy control groups. *indicate statistically significant differences (P<0.05). P value by Kruskal-Wallis test with Dunn's post hoc test was 0.9313 for A>G, 0.0735 for G>A, 0.086 for C>T, and 0.2586 for T>C. The average numbers of transversions (change from pyrine to pyrimidine or vise versa) are in Supplemental Figure 5. **B.** Heatmap depicting trinucleotide SNV patterns. The heatmap depicts specific triculeotide variants (SNV including the base immediately 3' and 5' to the SNV site). The 5' base is shows on the y-axis and the 3' base on the x-axis. Z-score of the log transformed values from 0-2 was used. To generate the heatmap, number of each variant plus 1 was converted to log. **C.** Percentage of SNVs and indels according to their damaging effects on the protein in each of the study subject group. **D.** Mean number of mutated genes in FA, SDS and controls with standard error of the means. Results by student T-test are shown. P value by Kruskal-Wallis test was 0.069.

Figure 4 **Patterns of single and tri-nucleotide alterations in FA/SDS-associated AML** **A.** Mutation rate in FA/SDS patients with AML or without AML. **B**. Percentage of each type of transition mutations across FA/SDS patients with or without AML samples. Percentages of transversions are in Supplemental Figure 19. **C**. Triculeotide heatmap of FA, FA-AML, SDS, SDS-AML patients. The tricucelotide mutations are shown with the 5' base on the y-axis and the 3'base on the x-axis. **D.** Type of mutations in AML versus non-AML samples.

Figure 5 **Genes mutated in FA/SDS-associated AML. A.** Top 30 genes mutated in AML cells from a patient with FA. **B.** Top 30 genes mutated in AML cells from a patient with SDS. **C.** Pathways that are disrupted in AML blasts from a patient with FA and in a patient with SDS.

Figure 6 **Transformational alterations in sequential samples from a non-malignant to malignant state.** The figure displays results from three sequential samples from a patient with SDS. **A**. Total number of mutations in each sequential sample. **B**. Percentage of total transition mutations in each sequential sample. Percentages of

transversions are in Supplemental Figure 24. **C.** Changes in trinucleotide signature heatmap in each sequential sample. **D.** Pathways affected in each sequential sample.

Table 1: Recurrently mutated genes in each study group according to the number of patients with mutations in the gene.

	Recurrence in 5 subjects	**Recurrence in 4 subjects**	**Recurrence in 3 subjects**	**Recurrence in 2 subjects**
FA			*ARID1A, CHD4, HUWE1, INTS1, ITPKB, SYNE1, THBS1*	*APC, ATF7IP, ATP2B3, ATRX, BCOR, BCR, BRCA2, CSMD3, CYLD, DST, EPHA7, FBN2, FES, FLCN, FLT4, GFI1B, GPC3, HIP1, KIAA1549, KMT2C, KMT2D, LRP1B, LRP2, LRRC7, LRRK1, MYH1, NAV1, NCAPD3, PDGFRB, PER1, PRDM1, PRKDC, PTPN13, PTPRT, RELN, SETBP1, SETDB1, SPTAN1, SRCAP, SRGAP3, STIL, TET1, TNR, TRIM24, TRIP11, UBR5, WAS, WDFY3, WDFY4, XPO1*
SDS	*SYNE1*	*RNF213*	*ASXL1, CAMTA1, COL1A1, COL7A1, EP400, EPPK1, HUWE1, KDM5A, LRP1B, NCOR2, PRKDC*	*ADAMTS20, ALK, AMER1, ARHGEF12, ATM, BUB1B, CDK12, CIITA, CNTN5, CNTRL, COL5A1, COLEC12, CREBBP, CUX1, DDX60, DNAH14, DSCAM, PB41L3, EPHA2, FAT1, FGFR1, FGFR3, FLG, FN1, FOXO1, IGF2R, KDM5C, KMT2D, LRIG3, MARK4, MGA, MLLT6, MN1, MPO, MTOR, MYC, MYH11, NOTCH2, NOTCH4, ODC1, PCDH15, PCM1, RABEP1, RAP1GDS1, RBL1, RPS6KA2, RUNX1, SETD2, SLC26A3 SMARCB1, SPTAN1, TP53, TP53BP1, TPR, TRIM33, UBR5, USP6, WDFY3, WDFY4, ZMYM3*
Healthy				*HUWE1, PIK3CB, SRCAP*

Table 2: Genes that were somatically mutated in leukemic blasts from a FA patient and a SDS patient

	VAF >0.07 to 0.25	**VAF >0.25 to 0.75**	**VAF >0.75-1**
FA-AML	*AFF1, AKT2, BCL9L, BRCA1, BRCA2, CENPF, CHD8, CLSTN2, COL11A1, CREBBP, DAB2IP, DDX60, DICER1, EPHA7, ERBB3, ERC1, KAT6B, KMT2C, LCP1, LRP2, MLLT1, MLLT10, NBN, NTRK1, NUP98, PIK3CB, PML, POLQ, PRCC, PTPN13, RAD50, ROS1, SOS1, STK19, SUFU, TP53, UBR5, TRIP11*	*PCDH15, ARID1B, SFPQ, EPPK1, MAP2K1, IL21R, HMGA1*	
SDS-AML	*AKAP9, AKT3, AMER1, ARID5B, ASTN1, ATF1, ATF7IP, BAI3, BAP1, BRAF, BRCA1, CACNA1D, CARD11, CDC6, CDH1, CDK12, CHD1, CHD7, COL5A1, CSMD3, DCC, DNM2, DYNC1H1, EGR3, ELF4, EP400, EPCAM, ERCC6, FGFR2, FLT1, GALNT15, GNAQ, GOLGA5, GRM3, HOXD11, HUWE1, KALRN, KMT2A, KMT2C, KMT2D, KTN1, LIFR, LRRC7, LRRK1, MBD1, MDC1 MED13, MGA, MKL1, MTCP1, MYH9, NAV3, NCOA2, NCOR1, NF1, NFE2, NUP214, OLIG2, PARK2, PBRM1, PDGFB, PHF20, PIK3CA, POT1, PRCC, PREX2, PTGS2, RSPO2, SERPINE1, SETD2, SETDB1, SFPQ, SMAD4, SOX2, SUZ12, SYNE1, TAF1, TBX18, TFE3, THBS1, TP53BP1, TRIM24, UBR5, WDFY3, WHSC1, XIRP2, ZNF91*	*DICER1, FLG, FLT4, HNF1A, IKZF1, JAK1, LPHN3, LRP2, MAST4, NCKIPSD, PTPRD, STK4, TCEB1, TNR, TP53*	*MYH1*

Table 3: Genes mutated in sequential samples from an SDS patient who eventually develop leukemia (Sequential sample 3)

	VAF >0.07 to 25	**VAF >0.25 to 75**	**VAF >0.75-1**
Seq-1	*BTK, AKT1, AXIN2, TET2*	*ARHGEF12*	*CACNA1D, CDH1, CHD3, DST, JAK3, KDM3B, NOTCH2, SDHC, TRRAP, WRN*
Seq -2	*ACVR1B, ARHGEF12, ARID1B, ATM, BUB1B, CHD6, COL1A1, COL7A1, DEK, DYNC1H1, EPHA6, EPHB4, FANCA, FGFR1, GABRG1, HDAC9, HUWE1, JMJD1C, KALRN, KDM5A, KDR, LRP1B, LRP2, MET, MN1, PRKDC, PRRC2A, SYNGAP1, TET2, TPR, ZMYND8*	*ALK, BCL11B, CAMTA1, CDK6, CNTN5, CREBBP, FLI1, HNF1A, IDH2, KDM5C MARK4, MAST4, MDC1, MLLT6, MYH11, NCOR2, NOTCH4, OLIG2, PCDH15, POT1, RARA, RBM10, SYNE1*	*AMER1, CHN1, FLT1, IL7R, MYH2, NCAPD3, NF2, ODC1, PRDM2, RUNX1T1, TSC2*
Seq-3 (AML)	*AKAP9, AKT3, AMER1, ARID5B, ASTN1, ATF1, ATF7IP, BAI3, BAP1, BRAF, BRCA1, CACNA1D, CARD11, CDC6, CDH1, CDK12, CHD1, CHD7, COL5A1, CSMD3, DCC, DNM2, DYNC1H1, EGR3, ELF4, EP400, EPCAM, ERCC6, FGFR2, FLT1, GALNT15, GNAQ, GOLGA5, GRM3, HOXD11, HUWE1, KALRN, KMT2A, KMT2C, KMT2D, KTN1, LIFR, LRRC7, LRRK1, MBD1, MDC1 MED13, MGA, MKL1, MTCP1, MYH9, NAV3, NCOA2, NCOR1, NF1, NFE2, NUP214, OLIG2, PARK2, PBRM1, PDGFB, PHF20, PIK3CA, POT1, PRCC, PREX2, PTGS2, RSPO2, SERPINE1, SETD2, SETDB1, SFPQ, SMAD4, SOX2, SUZ12, SYNE1, TAF1, TBX18, TFE3, THBS1, TP53BP1, TRIM24, UBR5, WDFY3, WHSC1, XIRP2, ZNF91*	*DICER1, FLG, FLT4, HNF1A, IKZF1, JAK1, LPHN3, LRP2, MAST4, NCKIPSD, PTPRD, STK4, TCEB1, TNR, TP53*	*MYH1*

Seq, sequential sample from the same subject; VAF, variant allele frequency

REFERENCES

1. DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, Alteri R, Robbins AS, and Jemal A. Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin.* 2014;64(4):252-71.
2. Schultz KA, Chen L, Chen Z, Kawashima T, Oeffinger KC, Woods WG, Nicholson HS, and Neglia JP. Health conditions and quality of life in survivors of childhood acute myeloid leukemia comparing post remission chemotherapy to BMT: a report from the children's oncology group. *Pediatric blood & cancer.* 2014;61(4):729-36.
3. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak D, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol.* 2017;19(4):271-81.
4. Laurenti E, and Gottgens B. From haematopoietic stem cells to complex differentiation landscapes. *Nature.* 2018;553(7689):418-26.
5. Notta F, Zandi S, Takayama N, Dobson S, Gan OI, Wilson G, Kaufmann KB, McLeod J, Laurenti E, Dunant CF, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science.* 2016;351(6269):aab2116.
6. Saultz JN, and Garzon R. Acute Myeloid Leukemia: A Concise Review. *Journal of clinical medicine.* 2016;5(3).
7. Cogle CR, Craig BM, Rollison DE, and List AF. Incidence of the myelodysplastic syndromes using a novel claims-based algorithm: high number of uncaptured cases by cancer registries. *Blood.* 2011;117(26):7121-5.
8. Hasle H, Wadsworth LD, Massing BG, McBride M, and Schultz KR. A population-based study of childhood myelodysplastic syndrome in British Columbia, Canada. *British journal of haematology.* 1999;106(4):1027-32.
9. Das U, Appaji L, Kumari BS, Lakshmaiah KC, Padma M, Kavitha S, and Sathyanarayanan V. A single center experience in 266 patients of infantile malignancies. *Pediatr Hematol Oncol.* 2014;31(6):489-97.
10. Ganguly BB, and Kadam NN. Mutations of myelodysplastic syndromes (MDS): An update. *Mutat Res Rev Mutat Res.* 2016;769(47-62.
11. Cada M, Segbefia C, Klaassen R, Fernandez C, Yanofsky R, Wu J, Pastore Y, Silva M, Lipton J, Brossard J, et al. The impact of the category, cytopathology and cytogenetics on development and progression of clonal and malignant myeloid transformation in inherited bone marrow failure syndromes. *Haematologia.* 2015;100(5):633-42.
12. Mandel K, Dror Y, Poon A, and Freedman M. A practical, comprehensive classification for pediatric myelodysplastic syndromes: the CCC system. *J Pediatr Hematol Oncol.* 2002;24(7):596-605.
13. Hasle H, Niemeyer CM, Chessells JM, Baumann I, Bennett JM, Kerndrup G, and Head DR. A pediatric approach to the WHO classification of myelodysplastic and myeloproliferative diseases. *Leukemia.* 2003;17(2):277-82.
14. Dror Y. In: Ikehara K ed. *Advances in the Study of Genetic Disorders:*. Rijeka: InTech Open Access Publisher, Croatia; 2011:1-37.
15. Smith OP, Hann IM, Chessells JM, Reeves BR, and Milla P. Haematological abnormalities in Shwachman-Diamond syndrome. *British journal of haematology.* 1996;94(2):279-84.
16. Butturini A, Gale RP, Verlander PC, Adler-Brecher B, Gillio AP, and Auerbach AD. Hematologic abnormalities in Fanconi anemia: an International Fanconi Anemia Registry study. *Blood.* 1994;84(5):1650-5.

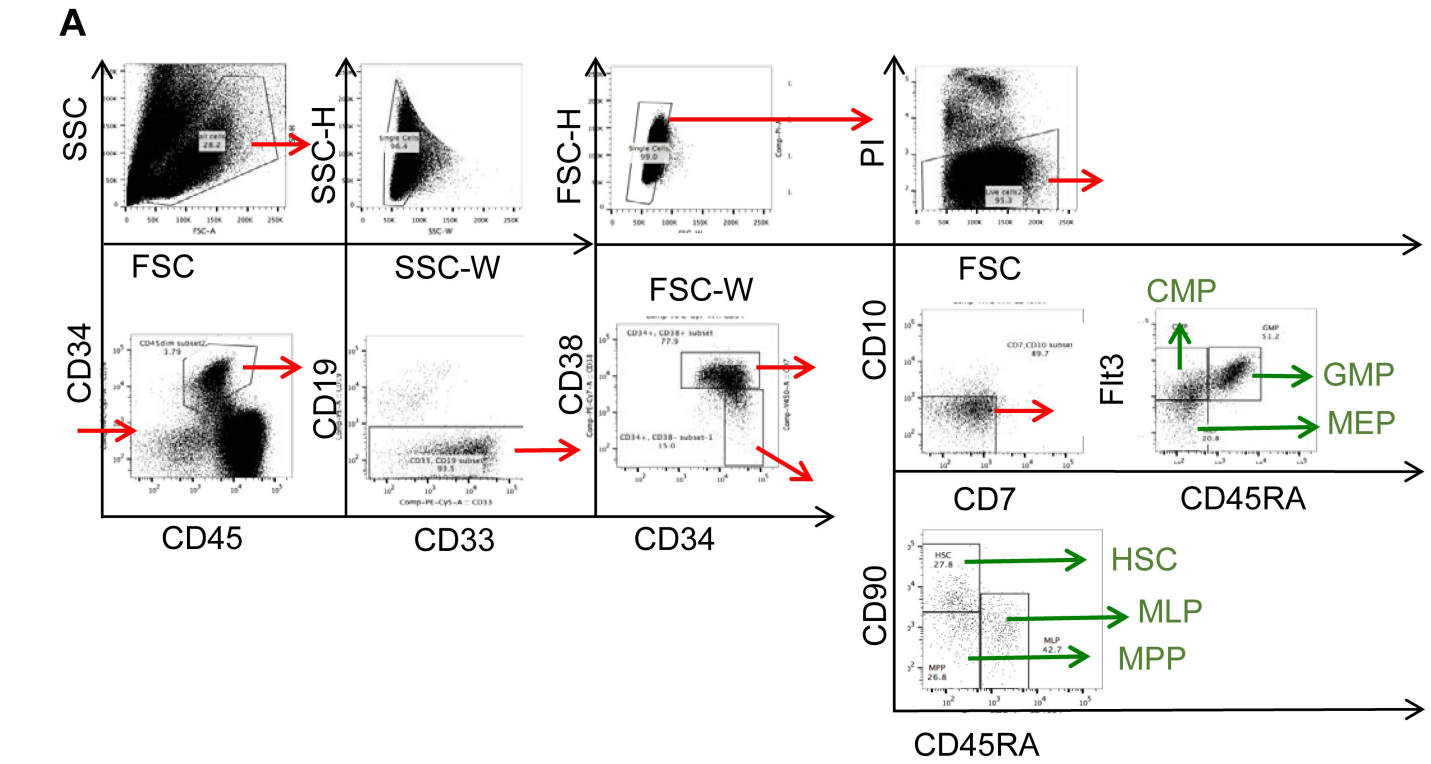
17. Lindsley RC, Saber W, Mar BG, Redd R, Wang T, Haagenson MD, Grauman PV, Hu ZH, Spellman SR, Lee SJ, et al. Prognostic Mutations in Myelodysplastic Syndrome after Stem-Cell Transplantation. *The New England journal of medicine.* 2017;376(6):536-47.
18. Xia J, Miller CA, Baty J, Ramesh A, Jotte MRM, Fulton RS, Vogel TP, Cooper MA, Walkovich KJ, Makaryan V, et al. Somatic mutations and clonal hematopoiesis in congenital neutropenia. *Blood.* 2018;131(4):408-16.
19. Quentin S, Cuccuini W, Ceccaldi R, Nibourel O, Pondarre C, Pages MP, Vasquez N, Dubois d'Enghien C, Larghero J, Peffault de Latour R, et al. Myelodysplasia and leukemia of Fanconi anemia are associated with a specific pattern of genomic abnormalities that includes cryptic RUNX1/AML1 lesions. *Blood.*117(15):e161-70.
20. Germeshausen M, Skokowa J, Ballmaier M, Zeidler C, and Welte K. G-CSF receptor mutations in patients with congenital neutropenia. *Curr Opin Hematol.* 2008;15(4):332-7.
21. Beekman R, Valkhof MG, Sanders MA, van Strien PM, Haanstra JR, Broeders L, Geertsma-Kleinekoort WM, Veerman AJ, Valk PJ, Verhaak RG, et al. Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia. *Blood.* 2012;119(22):5071-7.
22. Skokowa J, Steinemann D, Katsman-Kuipers JE, Zeidler C, Klimenkova O, Klimiankou M, Unalan M, Kandabarau S, Makaryan V, Beekman R, et al. Cooperativity of RUNX1 and CSF3R mutations in severe congenital neutropenia: a unique pathway in myeloid leukemogenesis. *Blood.* 2014;123(14):2229-37.
23. Ceccaldi R, Sarangi P, and D'Andrea AD. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol.* 2016;17(6):337-49.
24. Boocock GR, Morrison JA, Popovic M, Richards N, Ellis L, Durie PR, and Rommens JM. Mutations in SBDS are associated with Shwachman-Diamond syndrome. *Nat Genet.* 2003;33(1):97-101.
25. Dhanraj S, Matveev A, Li H, Lauhasurayotin S, Jardine L, Cada M, Zlateska B, Tailor CS, Zhou J, Mendoza-Londono R, et al. Biallelic mutations in DNAJC21 cause Shwachman-Diamond syndrome. *Blood.* 2017;129(11):1557-62.
26. Stepensky P, Chacon-Flores M, Kim KH, Abuzaitoun O, Bautista-Santos A, Simanovsky N, Siliqi D, Altamura D, Mendez-Godoy A, Gijsbers A, et al. Mutations in EFL1, an SBDS partner, are associated with infantile pancytopenia, exocrine pancreatic insufficiency and skeletal anomalies in aShwachman-Diamond like syndrome. *J Med Genet.* 2017;54(8):558-66.
27. Carapito R, Konantz M, Paillard C, Miao Z, Pichot A, Leduc MS, Yang Y, Bergstrom KL, Mahoney DH, Shardy DL, et al. Mutations in signal recognition particle SRP54 cause syndromic neutropenia with Shwachman-Diamond-like features. *J Clin Invest.* 2017;127(11):4090-103.
28. Rackoff WR, Orazi A, Robinson CA, Cooper RJ, Alter BP, Freedman MH, Harris RE, and Williams DA. Prolonged administration of granulocyte colony-stimulating factor (filgrastim) to patients with Fanconi anemia: a pilot study. *Blood.* 1996;88(5):1588-93.
29. Dror Y, and Freedman MH. Shwachman-Diamond syndrome: An inherited preleukemic bone marrow failure disorder with aberrant hematopoietic progenitors and faulty marrow microenvironment. *Blood.* 1999;94(9):3048-54.
30. Doulatov S, Notta F, Eppert K, Nguyen LT, Ohashi PS, and Dick JE. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol.* 2010;11(7):585-93.
31. Hashmi S, Allen C, Klaassen R, Fernandez C, Yanofsky R, Wu J, Champagne J, Silva M, Lipton J, Brossard J, et al. Comparative analysis of Shwachman-Diamond syndrome to other inherited bone marrow failure syndromes and genotype-phenotype correlation. *Clinical genetics.* 2011;79(5):448-58.

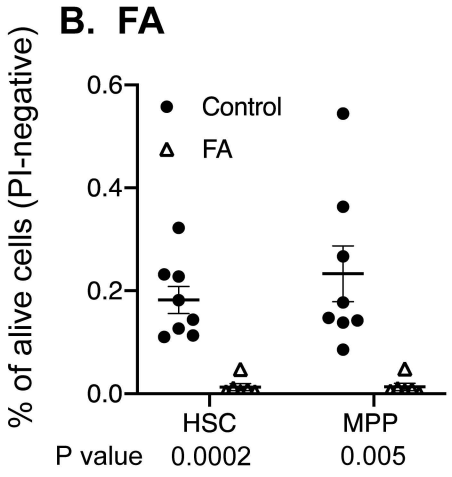
32. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, Wartman LD, Lamprecht TL, Liu F, Xia J, et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell.* 2012;150(2):264-78.
33. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell.* 2012;149(5):979-93.
34. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, et al. Signatures of mutational processes in human cancer. *Nature.* 2013;500(7463):415-21.
35. Deshwar AG, Vembu S, Yung CK, Jang GH, Stein L, and Morris Q. PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol.* 2015;16(35.
36. Alter BP, Caruso JP, Drachtman RA, Uchida T, Velagaleti GV, and Elghetany MT. Fanconi anemia: myelodysplasia as a predictor of outcome. *Cancer Genet Cytogenet.* 2000;117(2):125-31.
37. Dror Y, Durie P, Ginzberg H, Herman R, Banerjee A, Champagne M, Shannon K, Malkin D, and Freedman MH. Clonal evolution in marrows of patients with Shwachman-Diamond syndrome: a prospective 5-year follow-up study. *Experimental Hematology.* 2002;30(7):659-69.
38. Dror Y. Shwachman-Diamond syndrome: implications for understanding the molecular basis of leukaemia. *Expert Rev Mol Med.* 2008;10(e38.
39. Cooper JN, and Young NS. Clonality in context: hematopoietic clones in their marrow environment. *Blood.* 2017;130(22):2363-72.
40. Walden H, and Deans AJ. The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. *Annual review of biophysics.* 2014;43(257-78.
41. Sarkar J, and Liu Y. Fanconi anemia proteins in telomere maintenance. *DNA repair.* 2016;43(107-12.
42. Thornley I, Dror Y, Sung L, Wynn RF, and Freedman MH. Abnormal telomere shortening in leucocytes of children with Shwachman-Diamond syndrome. *British journal of haematology.* 2002;117(1):189-92.
43. Valli R, Minelli A, Galbiati M, D'Amico G, Frattini A, Montalbano G, Khan AW, Porta G, Millefanti G, Olivieri C, et al. Shwachman-Diamond syndrome with clonal interstitial deletion of the long arm of chromosome 20 in bone marrow: haematological features, prognosis and genomic instability. *British journal of haematology.* 2019;184(6):974-81.
44. Valli R, De Paoli E, Nacci L, Frattini A, Pasquali F, and Maserati E. Novel recurrent chromosome anomalies in Shwachman-Diamond syndrome. *Pediatric blood & cancer.* 2017;64(8).
45. Cioc AM, Wagner JE, MacMillan ML, DeFor T, and Hirsch B. Diagnosis of myelodysplastic syndrome among a cohort of 119 patients with fanconi anemia: morphologic and cytogenetic characteristics. *American journal of clinical pathology.* 2010;133(1):92-100.
46. Tonnies H, Huber S, Kuhl JS, Gerlach A, Ebell W, and Neitzel H. Clonal chromosomal aberrations in bone marrow cells of Fanconi anemia patients: gains of the chromosomal segment 3q26q29 as an adverse risk factor. *Blood.* 2003;101(10):3872-4.
47. Bavarva JH, Tae H, McIver L, and Garner HR. Nicotine and oxidative stress induced exomic variations are concordant and overrepresented in cancer-associated genes. *Oncotarget.* 2014;5(13):4788-98.
48. Dizdaroglu M. Oxidatively induced DNA damage: mechanisms, repair and disease. *Cancer letters.* 2012;327(1-2):26-47.
49. Cumming RC, Lightfoot J, Beard K, Youssoufian H, O'Brien PJ, and Buchwald M. Fanconi anemia group C protein prevents apoptosis in hematopoietic cells through redox regulation of GSTP1. *Nature medicine.* 2001;7(7):814-20.

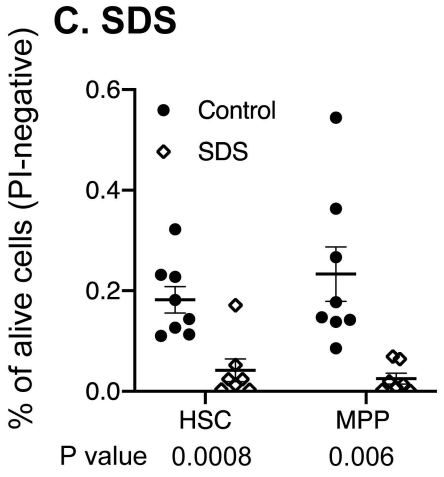
50. Li J, Sejas DP, Zhang X, Qiu Y, Nattamai KJ, Rani R, Rathbun KR, Geiger H, Williams DA, Bagby GC, et al. TNF-alpha induces leukemic clonal evolution ex vivo in Fanconi anemia group C murine stem cells. *J Clin Invest.* 2007;117(11):3283-95.
51. Kumari U, Ya Jun W, Huat Bay B, and Lyakhovich A. Evidence of mitochondrial dysfunction and impaired ROS detoxifying machinery in Fanconi anemia cells. *Oncogene.* 2014;33(2):165-72.
52. Ambekar C, Das B, Yeger H, and Dror Y. SBDS-deficiency results in deregulation of reactive oxygen species leading to increased cell death and decreased cell growth. *Pediatric blood & cancer.* 2010;55(6):1138-44.
53. Sen S, Wang H, Nghiem CL, Zhou K, Yau J, Tailor C, Irwin MS, and Dror Y. The ribosome-related protein, SBDS, is critical for normal erythropoiesis. *Blood.* 2011;118(24):6407-17.
54. Huang F, Ben Aissa M, Magron A, Huard CC, Godin C, Levesque G, and Carreau M. The Fanconi anemia group C protein interacts with uncoordinated 5A and delays apoptosis. *PloS one.* 2014;9(3):e92811.
55. Dror Y, and Freedman MH. Shwachman-Diamond syndrome marrow cells show abnormally increased apoptosis mediated through the Fas pathway. *Blood.* 2001;97(10):3011-6.
56. Rujkijyanont P, Watanabe K, Ambekar C, Wang H, Schimmer A, Beyene J, and Dror Y. SBDS-deficient cells undergo accelerated apoptosis through the Fas-pathway. *Haematologica.* 2008;93(3):363-71.
57. Watanabe K, Ambekar C, Wang H, Ciccolini A, Schimmer A, and Dror Y. SBDS-deficiency results in specific hypersensitivity to Fas stimulation and accumulation of Fas at the plasma membrane. *Apoptosis.* 2009;14(1):77-89.
58. Warren AJ. Molecular basis of the human ribosomopathy Shwachman-Diamond syndrome. *Advances in biological regulation.* 2018;67(109-27.
59. Stanley N, Olson TS, and Babushok DV. Recent advances in understanding clonal haematopoiesis in aplastic anaemia. *British journal of haematology.* 2017;177(4):509-25.
60. Padro T, Ruiz S, Bieker R, Burger H, Steins M, Kienast J, Buchner T, Berdel WE, and Mesters RM. Increased angiogenesis in the bone marrow of patients with acute myeloid leukemia. *Blood.* 2000;95(8):2637-44.
61. Pruneri G, Bertolini F, Soligo D, Carboni N, Cortelezzi A, Ferrucci PF, Buffa R, Lambertenghi-Deliliers G, and Pezzella F. Angiogenesis in myelodysplastic syndromes. *British journal of cancer.* 1999;81(8):1398-401.
62. Leung EW, Rujkijyanont P, Beyene J, Wei K, Abdelhaleem M, Freedman MH, and Dror Y. Shwachman-Diamond syndrome: an inherited model of aplastic anaemia with accelerated angiogenesis. *British journal of haematology.* 2006;133(5):558-61.
63. Zambetti NA, Ping Z, Chen S, Kenswil KJG, Mylona MA, Sanders MA, Hoogenboezem RM, Bindels EMJ, Adisty MN, Van Strien PMH, et al. Mesenchymal Inflammation Drives Genotoxic Stress in Hematopoietic Stem Cells and Predicts Disease Evolution in Human Pre-leukemia. *Cell stem cell.* 2016;19(5):613-27.
64. Yap YS, Singh AP, Lim JHC, Ahn JH, Jung KH, Kim J, Dent RA, Ng RCH, Kim SB, and Chiang DY. Elucidating therapeutic molecular targets in premenopausal Asian women with recurrent breast cancers. *NPJ Breast Cancer.* 2018;4(19.
65. Nikas JB. Independent validation of a mathematical genomic model for survival of glioma patients. *Am J Cancer Res.* 2016;6(6):1408-19.
66. Rouget-Quermalet V, Giustiniani J, Marie-Cardine A, Beaud G, Besnard F, Loyaux D, Ferrara P, Leroy K, Shimizu N, Gaulard P, et al. Protocadherin 15 (PCDH15): a new secreted isoform and a potential marker for NK/T cell lymphomas. *Oncogene.* 2006;25(19):2807-11.
67. Dolnik A, Engelmann JC, Scharfenberger-Schmeer M, Mauch J, Kelkenberg-Schade S, Haldemann B, Fries T, Kronke J, Kuhn MW, Paschka P, et al. Commonly altered genomic regions in acute

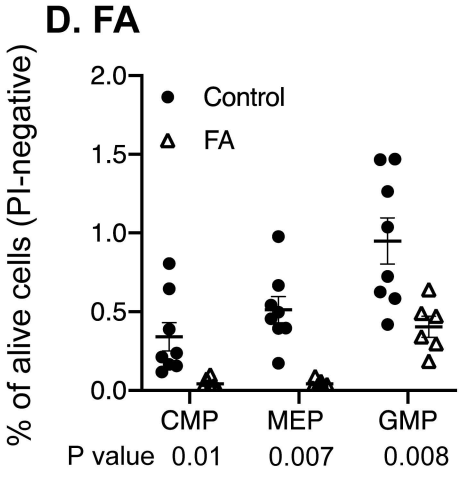
myeloid leukemia are enriched for somatic mutations involved in chromatin remodeling and splicing. *Blood.* 2012;120(18):e83-92.
68. Tao Y, Ma C, Fan Q, Wang Y, Han T, and Sun C. MicroRNA-1296 Facilitates Proliferation, Migration And Invasion Of Colorectal Cancer Cells By Targeting SFPQ. *J Cancer.* 2018;9(13):2317-26.
69. Manso BA, Zhang H, Mikkelson MG, Gwin KA, Secreto CR, Ding W, Parikh SA, Kay NE, and Medina KL. Bone marrow hematopoietic dysfunction in untreated chronic lymphocytic leukemia patients. *Leukemia.* 2019;33(3):638-52.
70. Townsley DM, Scheinberg P, Winkler T, Desmond R, Dumitriu B, Rios O, Weinstein B, Valdez J, Lotter J, Feng X, et al. Eltrombopag Added to Standard Immunosuppression for Aplastic Anemia. *The New England journal of medicine.* 2017;376(16):1540-50.
71. Wong WM, Dolinska M, Sigvardsson M, Ekblom M, and Qian H. A novel Lin-CD34+CD38- integrin alpha2- bipotential megakaryocyte-erythrocyte progenitor population in the human bone marrow. *Leukemia.* 2016;30(6):1399-402.
72. Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, Kennedy JA, Schimmer AD, Schuh AC, Yee KW, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature.* 2014;506(7488):328-33.
73. Chen J, Kao YR, Sun D, Todorova TI, Reynolds D, Narayanagari SR, Montagna C, Will B, Verma A, and Steidl U. Myelodysplastic syndrome progression to acute myeloid leukemia at the stem cell level. *Nature medicine.* 2019;25(1):103-10.
74. Schmidt M, Rinke J, Schafer V, Schnittger S, Kohlmann A, Obstfelder E, Kunert C, Ziermann J, Winkelmann N, Eigendorff E, et al. Molecular-defined clonal evolution in patients with chronic myeloid leukemia independent of the BCR-ABL status. *Leukemia.* 2014;28(12):2292-9.
75. Aubry MC, Roden A, Murphy SJ, Vasmatzis G, Johnson SH, Harris FR, Halling G, Knudson RA, Ketterling RP, and Feldman AL. Chromosomal rearrangements and copy number abnormalities of TP63 correlate with p63 protein expression in lung adenocarcinoma. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* 2015;28(3):359-66.
76. Shlien A, Campbell BB, de Borja R, Alexandrov LB, Merico D, Wedge D, Van Loo P, Tarpey PS, Coupland P, Behjati S, et al. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat Genet.* 2015;47(3):257-62.
77. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. *The New England journal of medicine.* 2009;361(11):1058-66.
78. Dror Y, Donadieu J, Koglmeier J, Dodge J, Toiviainen-Salo S, Makitie O, Kerr E, Zeidler C, Shimamura A, Shah N, et al. Draft consensus guidelines for diagnosis and treatment of Shwachman-Diamond syndrome. *Ann N Y Acad Sci.* 2011;1242(40-55.

Figure 1 **Deep immunophenotyping revealed striking loss of most, but not all, hematopoietic stem cells and progenitors in bone marrows from FA/SDS patients. A**. Analytic strategy of bone marrow aspirate cells by immunophenotyping. **B-C.** Comparison of multipotent cells between FA (n=6), SDS (n=7) and control (n=8). The mean percentage of each HSCPs among the viable bone marrow mononuclear cells is presented with standard error of mean (SEM). **D-E.** Comparison of oligopotent progenitors between FA, SDS and control patients. The mean percentage of each HSCPs among the viable bone marrow mononuclear cells is presented with standard error of mean (SEM). HSC, hematopoietic stem cell; MPP, multipotent progenitor cell; MEP, megakaryocyte erythroid progenitor; GMP, granulocyte monocyte progenitor. Student's t-test was used to compare between patients and controls. The same control data in Figures 1C and 1E are also presented in Figure 1B and 1D, respectively.


A
SSC
FSC
SSC-H
SSC-W
FSC-H
FSC-W
PI
FSC
CD34
CD45
CD19
CD33
CD38
CD34
CD10
CD7
Flt3
CMP
GMP
MEP
CD45RA
CD90
HSC
MLP
MPP
CD45RA


B. FA
% of alive cells (PI-negative)
Control
FA
HSC
MPP
P value 0.0002 0.005


C. SDS
% of alive cells (PI-negative)
Control
SDS
HSC
MPP
P value 0.0008 0.006


D. FA
% of alive cells (PI-negative)
Control
FA
CMP
MEP
GMP
P value 0.01 0.007 0.008

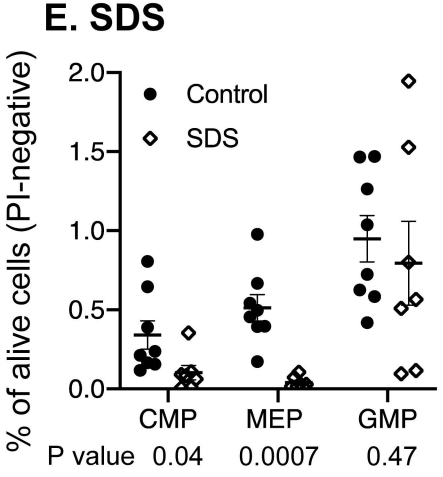

E. SDS
% of alive cells (PI-negative)
Control
SDS
CMP
MEP
GMP
P value 0.04 0.0007 0.47

Figure 1

Figure 2 **Frequency of somatic variants in bone marrow samples from patients with SDS, FA and healthy control subjects. A**. Comparison of average (+/-standard error of the mean) variant rate between FA (n=6), SDS (n=7) and healthy control subjects (n=6). Results by student T-test are shown. P value by Kruskal-Wallis test with Dunn's post hoc test was 0.0695. **B**, **C** and **D** show the variant rate among controls, FA and SDS subjects organized according to ages. **E**. Allele frequency of the various variants in controls, FA and SDS. The group were compared using the Wilcoxon signed-rank test. Asterisk (*) indicates statistical significance compared to controls with $p<0.0001$. The Y-axis represent the variant frequency and the X-axis represent the variants arranged from those with the highest allele frequency to the lowest. In each group, each number may represent a different variant.

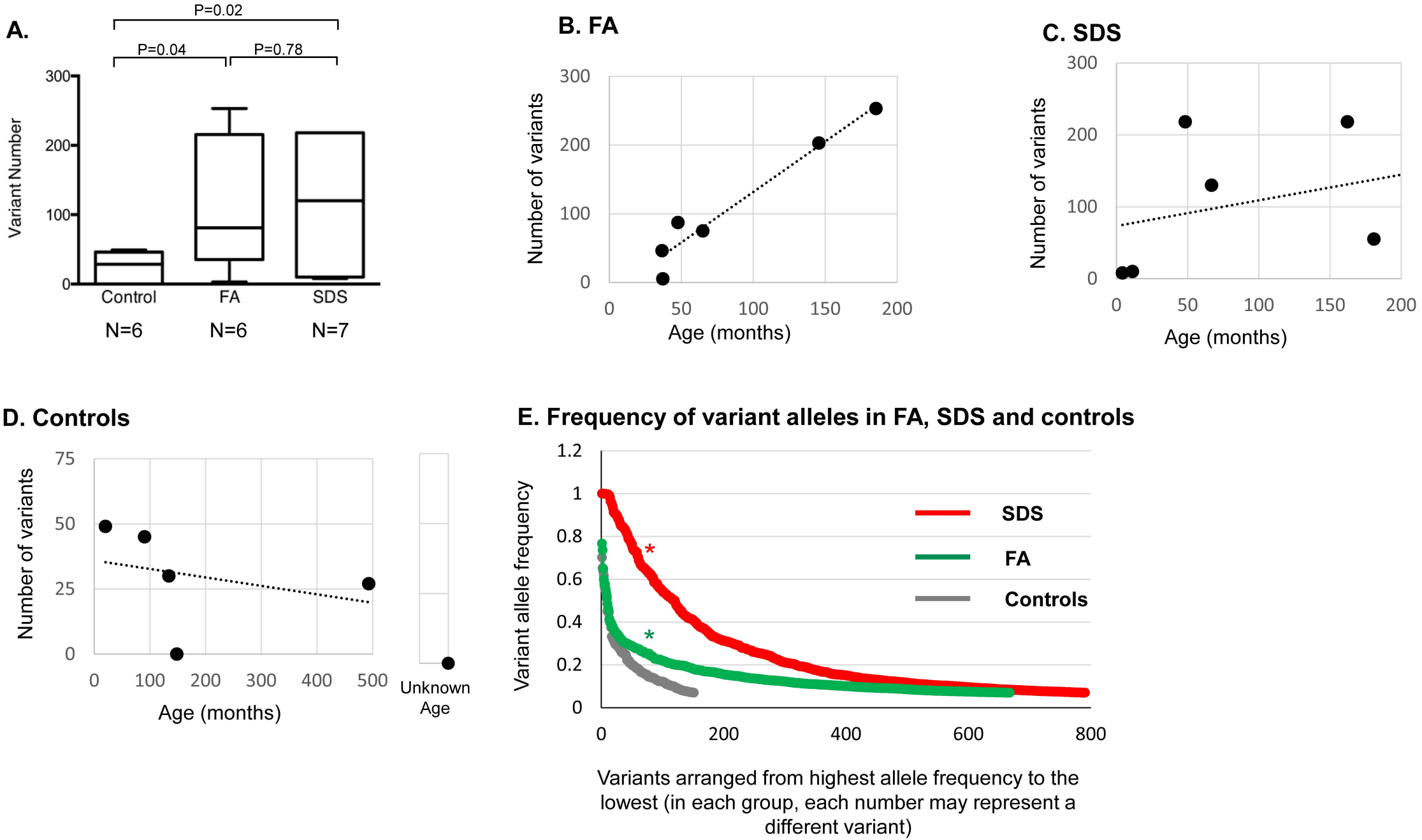

A.
P=0.02
P=0.04
P=0.78
Variant Number
300
200
100
0
Control
FA
SDS
N=6
N=6
N=7
B. FA
Number of variants
Age (months)
C. SDS
Number of variants
Age (months)
D. Controls
Number of variants
75
50
25
0
0
100
200
300
400
500
Age (months)
Unknown Age
E. Frequency of variant alleles in FA, SDS and controls
Variant allele frequency
1.2
1
0.8
0.6
0.4
0.2
0
0
200
400
600
800
SDS
FA
Controls
*
*
Variants arranged from highest allele frequency to the lowest (in each group, each number may represent a different variant)

Figure 2

Figure 3 **Patterns of single and tri-nucleotide alterations among FA, SDS and healthy control subjects. A**. Average number (+/- standard error of the mean) of each transition (inside the CT purine group or inside the GA pyrimidine group) variant per subject among the FA, SDS and healthy control groups. *indicate statistically significant differences ($P<0.05$). P value by Kruskal-Wallis test with Dunn's post hoc test was 0.9313 for A>G, 0.0735 for G>A, 0.086 for C>T, and 0.2586 for T>C. The average numbers of transversions (change from pyrine to pyrimidine or vise versa) are in Supplemental Figure 5. **B.** Heatmap depicting trinucleotide SNV patterns. The heatmap depicts specific triculeotide variants (SNV including the base immediately 3' and 5' to the SNV site). The 5' base is shows on the y-axis and the 3' base on the x-axis. Z-score of the log transformed values from 0-2 was used. To generate the heatmap, number of each variant plus 1 was converted to log. **C.** Percentage of SNVs and indels according to their damaging effects on the protein in each of the study subject group. **D.** Mean number of mutated genes in FA, SDS and controls with standard error of the means. Results by student T-test are shown. P value by Kruskal-Wallis test was 0.069.

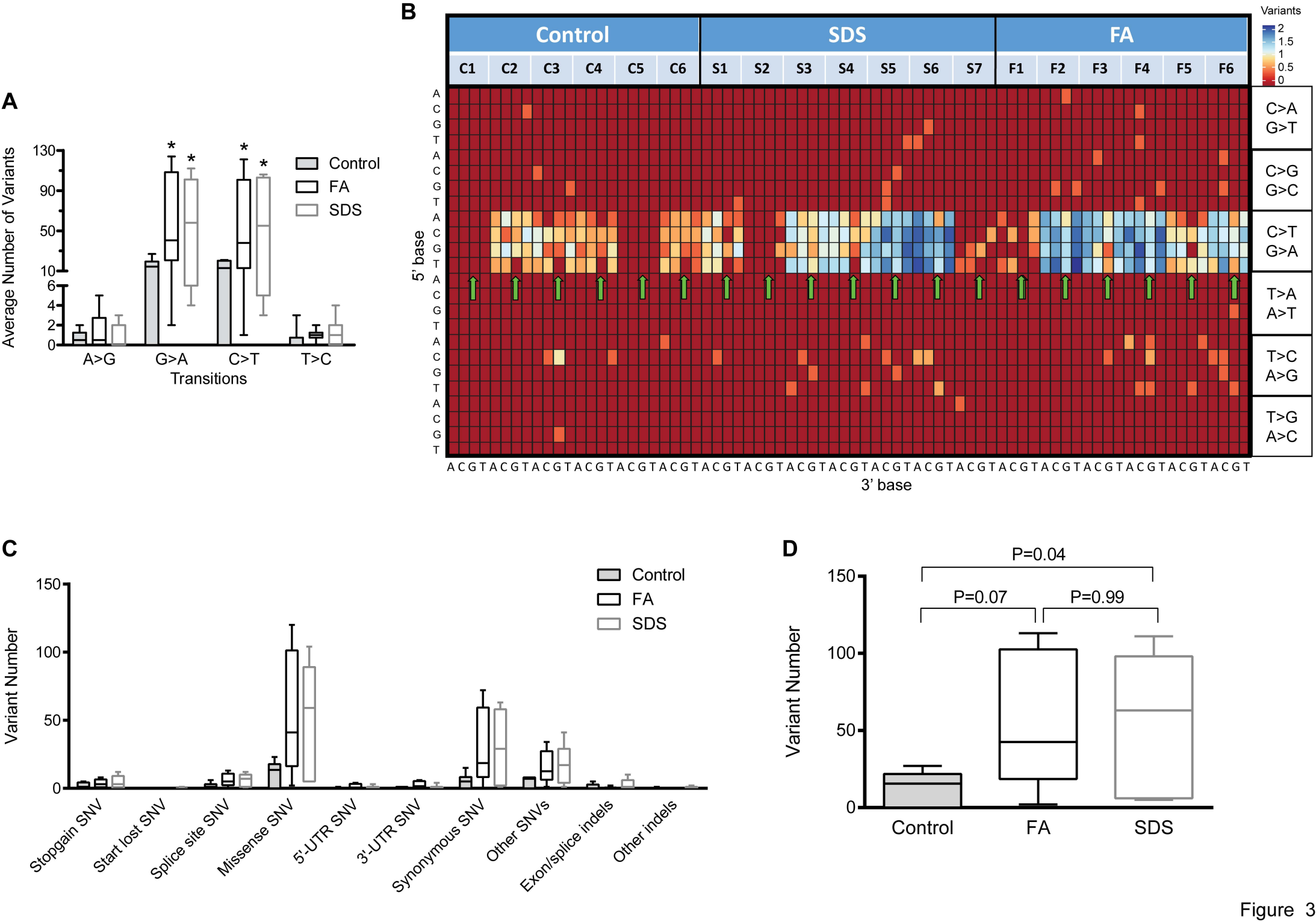

A
Average Number of Variants
130
90
50
10
6
4
2
0
*
*
*
*
Control
FA
SDS
A>G
G>A
C>T
T>C
Transitions
B
Control
SDS
FA
C1
C2
C3
C4
C5
C6
S1
S2
S3
S4
S5
S6
S7
F1
F2
F3
F4
F5
F6
Variants
2
1.5
1
0.5
0
5' base
3' base
C>A
G>T
C>G
G>C
C>T
G>A
T>A
A>T
T>C
A>G
T>G
A>C
C
Variant Number
150
100
50
0
Control
FA
SDS
Stopgain SNV
Start lost SNV
Splice site SNV
Missense SNV
5'-UTR SNV
3'-UTR SNV
Synonymous SNV
Other SNVs
Exon/splice indels
Other indels
D
Variant Number
P=0.04
P=0.07
P=0.99
Control
FA
SDS

Figure 3

Figure 4 **Patterns of single and tri-nucleotide alterations in FA/SDS-associated AML**

A. Mutation rate in FA/SDS patients with AML or without AML. **B**. Percentage of each type of transition mutations across FA/SDS patients with or without AML samples. Percentages of transversions are in Supplemental Figure 19. **C**. Triculeotide heatmap of FA, FA-AML, SDS, SDS-AML patients. The tricucelotide mutations are shown with the 5' base on the y-axis and the 3'base on the x-axis. **D.** Type of mutations in AML versus non-AML samples.

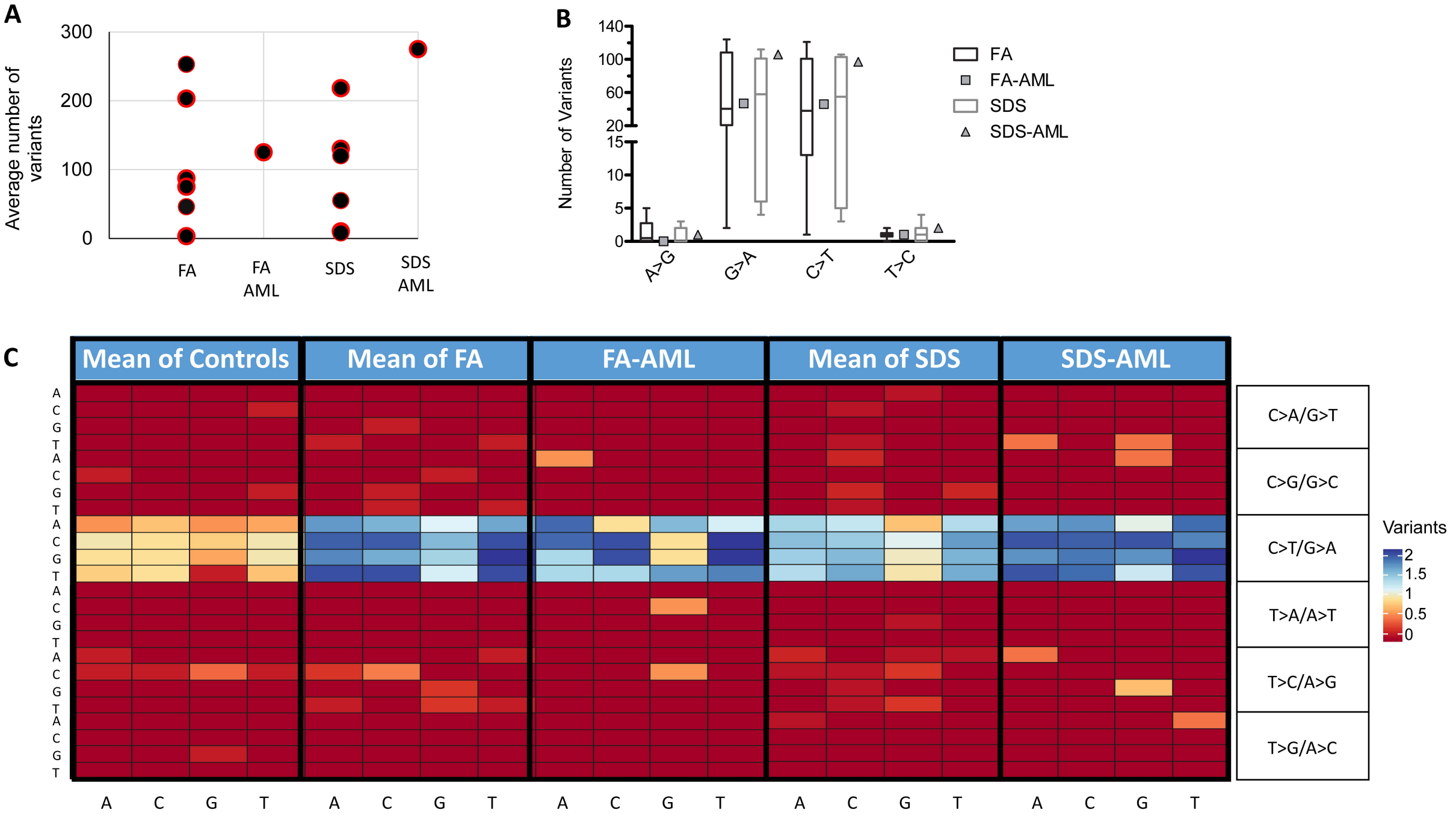

A
Average number of variants
300
200
100
0
FA
FA AML
SDS
SDS AML
B
Number of Variants
140
100
60
20
15
10
5
0
A>G
G>A
C>T
T>C
FA
FA-AML
SDS
SDS-AML
C
Mean of Controls
Mean of FA
FA-AML
Mean of SDS
SDS-AML
A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T
A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T
C>A/G>T
C>G/G>C
C>T/G>A
T>A/A>T
T>C/A>G
T>G/A>C
Variants
2
1.5
1
0.5
0

Figure 4A-C

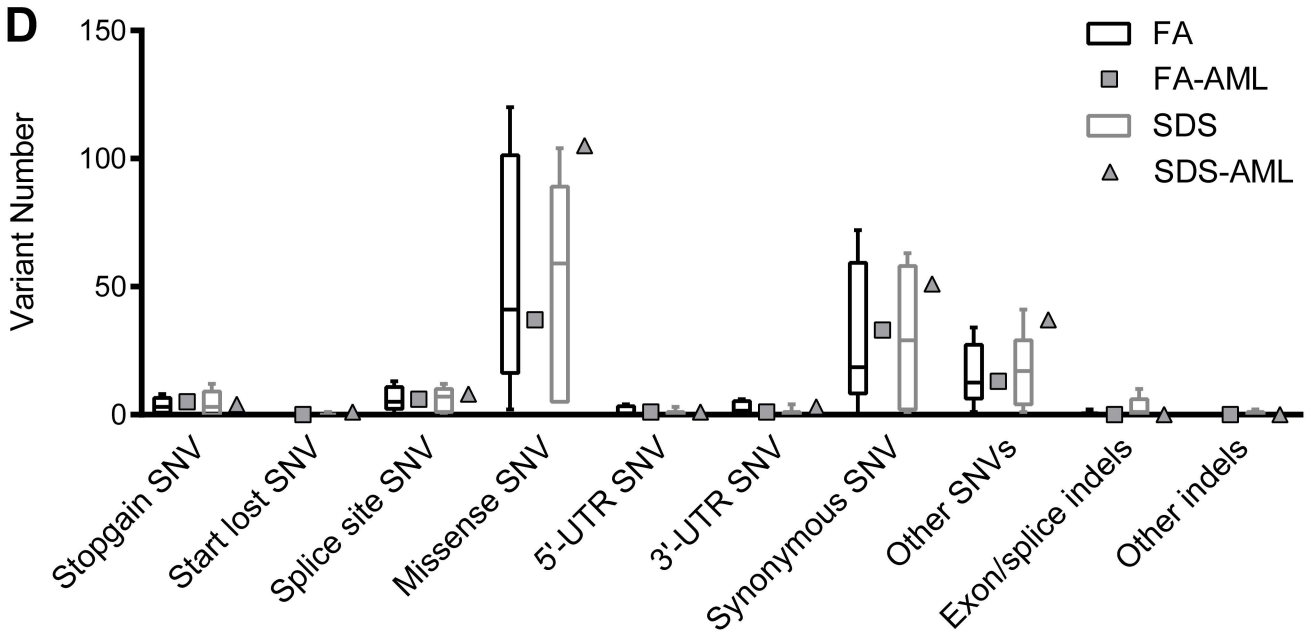

D
150
100
50
0
Variant Number
FA
FA-AML
SDS
SDS-AML
Stopgain SNV
Start lost SNV
Splice site SNV
Missense SNV
5'-UTR SNV
3'-UTR SNV
Synonymous SNV
Other SNVs
Exon/splice indels
Other indels

Figure 4D

Figure 5 **Genes mutated in FA/SDS-associated AML. A.** Top 30 genes mutated in AML cells from a patient with FA. **B.** Top 30 genes mutated in AML cells from a patient with SDS. **C.** Pathways that are disrupted in AML blasts from a patient with FA and in a patient with SDS.

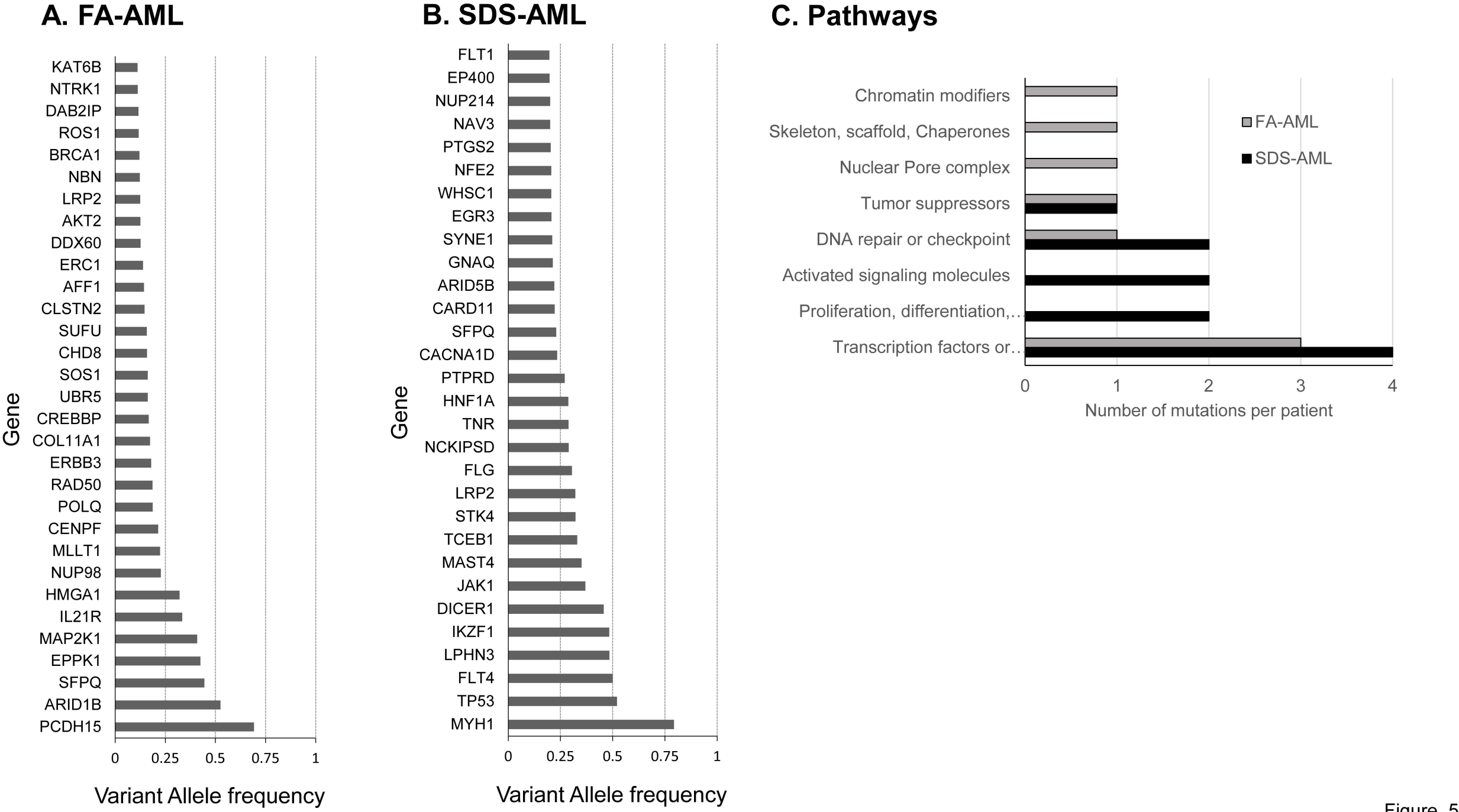

A. FA-AML
KAT6B
NTRK1
DAB2IP
ROS1
BRCA1
NBN
LRP2
AKT2
DDX60
ERC1
AFF1
CLSTN2
SUFU
CHD8
SOS1
UBR5
CREBBP
COL11A1
ERBB3
RAD50
POLQ
CENPF
MLLT1
NUP98
HMGA1
IL21R
MAP2K1
EPPK1
SFPQ
ARID1B
PCDH15
Gene
0
0.25
0.5
0.75
1
Variant Allele frequency
B. SDS-AML
FLT1
EP400
NUP214
NAV3
PTGS2
NFE2
WHSC1
EGR3
SYNE1
GNAQ
ARID5B
CARD11
SFPQ
CACNA1D
PTPRD
HNF1A
TNR
NCKIPSD
FLG
LRP2
STK4
TCEB1
MAST4
JAK1
DICER1
IKZF1
LPHN3
FLT4
TP53
MYH1
Gene
0
0.25
0.5
0.75
1
Variant Allele frequency
C. Pathways
Chromatin modifiers
Skeleton, scaffold, Chaperones
Nuclear Pore complex
Tumor suppressors
DNA repair or checkpoint
Activated signaling molecules
Proliferation, differentiation,..
Transcription factors or..
FA-AML
SDS-AML
0
1
2
3
4
Number of mutations per patient

Figure 5

Figure 6 **Transformational alterations in sequential samples from a non-malignant to malignant state.** The figure displays results from three sequential samples from a patient with SDS. **A**. Total number of mutations in each sequential sample. **B**. Percentage of total transition mutations in each sequential sample. Percentages of transversions are in Supplemental Figure 24. **C.** Changes in trinucleotide signature heatmap in each sequential sample. **D.** Pathways affected in each sequential sample.

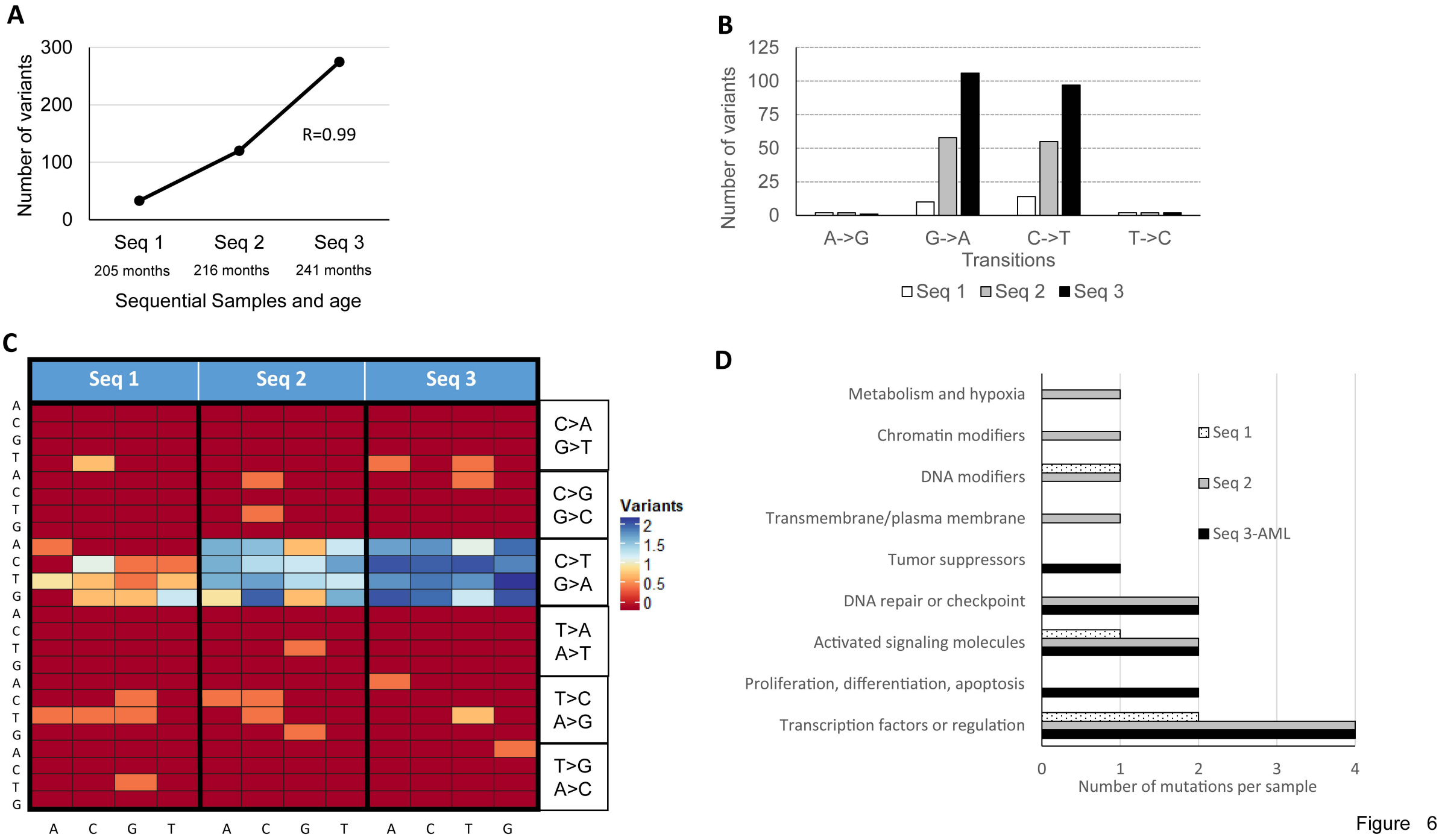

A
Number of variants
300
200
100
0
R=0.99
Seq 1
205 months
Seq 2
216 months
Seq 3
241 months
Sequential Samples and age
B
Number of variants
125
100
75
50
25
0
A->G
G->A
C->T
T->C
Transitions
Seq 1
Seq 2
Seq 3
C
Seq 1
Seq 2
Seq 3
C>A
G>T
C>G
G>C
C>T
G>A
T>A
A>T
T>C
A>G
T>G
A>C
Variants
2
1.5
1
0.5
0
D
Metabolism and hypoxia
Chromatin modifiers
DNA modifiers
Transmembrane/plasma membrane
Tumor suppressors
DNA repair or checkpoint
Activated signaling molecules
Proliferation, differentiation, apoptosis
Transcription factors or regulation
Seq 1
Seq 2
Seq 3-AML
0
1
2
3
4
Number of mutations per sample

Figure 6