

ORIGINAL ARTICLE

The clonal origins of leukemic progression of myelodysplasia

T Kim^{1,2}, MS Tyndel^{2,3}, HJ Kim^{4,5}, J-S Ahn^{4,5}, SH Choi⁴, HJ Park⁴, Y-k Kim^{4,5}, D-H Yang⁵, J-J Lee⁵, S-H Jung⁵, SY Kim^{4,5}, YH Min⁶, J-W Cheong⁶, SK Sohn⁷, JH Moon⁷, M Choi⁸, M Lee⁸, Z Zhang^{1,2,9} and D(DH) Kim¹⁰

The genetics behind the progression of myelodysplasia to secondary acute myeloid leukemia (sAML) is poorly understood. In this study, we profiled somatic mutations and their dynamics using next generation sequencing on serial samples from a total of 124 patients, consisting of a 31 patient discovery cohort and 93 patients from two validation cohorts. Whole-exome analysis on the discovery cohort revealed that 29 of 31 patients carry mutations related to at least one of eight commonly mutated pathways in AML. Mutations in genes related to DNA methylation and splicing machinery were found in T-cell samples, which expand at the initial diagnosis of the myelodysplasia, suggesting their importance as early disease events. On the other hand, somatic variants associated with signaling pathways arise or their allelic burdens expand significantly during progression. Our results indicate a strong association between mutations in activated signaling pathways and sAML progression. Overall, we demonstrate that distinct categories of genetic lesions play roles at different stages of sAML in a generally fixed order.

Leukemia (2017) 31, 1928–1935; doi:10.1038/leu.2017.17

INTRODUCTION

Acute myeloid leukemia (AML) that develops from pre-existing hematologic diseases, rather than developing *de novo*, is known as secondary AML (sAML). A number of hematologic malignancies can progress to sAML, including but not limited to myelodysplastic syndromes (MDSs), chronic myelomonocytic leukemia and atypical chronic myeloid leukemia.^{1–3} Roughly one-third of patients diagnosed with MDSs or other related hematologic malignancies (henceforth collectively referred to as MDS) progress to sAML.¹ MDS are heterogeneous clonal hematopoietic disorders characterized by dysplastic changes in one or more cellular lineages, causing impaired bone marrow function.¹ The transformation from a normal stem cell into a preleukemic or leukemic state involves the accumulation of genetic abnormalities.

High throughput sequencing technology has led to a number of discoveries in the field of AML biology, including the existence of recurrently mutated genes, their prognostic values and their genomic classifications.^{4,5} For example, commonly mutated genes and their relevant biological pathways have been revealed in a number of different tumors, including AML.^{6,7} The Cancer Genome Atlas (TCGA) consortium has compiled a list of eight most commonly mutated pathways in *de novo* AML, which accounts for over 99% of the adult AML cases in their cohort.⁶ In line with this, a number of studies in recent years have improved our understanding on the pathophysiology of MDS and its prognostic factors as well.^{8–12} In particular, studies have used next generation sequencing to identify genetic alterations that are commonly mutated and have prognostic value as well as great genetic heterogeneity that accounts for much of the clinical heterogeneity of MDS.^{10–18} Mutations in genes affecting splicing machinery and

chromatin modifiers have been shown to be overrepresented in sAML when compared with *de novo* AML or therapy-related AML.¹⁹

Although a number of studies have investigated the patterns of somatic mutations in various myeloid neoplasms, the complete molecular and genetic characteristics of sAML still remain largely unclear. It has been demonstrated that mutations in splicing machinery, DNA methylation and chromatin modifications are commonly mutated in MDS.^{10,11,19–23} On the other hand, mutations in activated signaling pathway such as RAS mutations have been associated with the leukemic progression of MDS.^{19,20,24–30} Although relative timing and pattern of mutation acquisition in sAML can be computationally inferred based on variant allele frequency (VAF), the inferred model of clonal evolution would still require validation. Furthermore, the highly heterogeneous mutational profiles of the MDS and sAML complicate the search for general patterns of sAML progression. Previous studies have used serial sequencing to examine that sAML progression is associated with acquisition of new mutations.^{19,30,31} In particular, Lindsley *et al.*¹⁹ demonstrated that sAML progression is associated with the acquisition of mutations in genes associated with activated signaling pathways and myeloid transcription factors. Similar observation was recurrently shown by a recent study by Makishima *et al.* as well.³⁰

To address these questions and dissect the order of mutation acquisition throughout the course of the disease, we performed whole-exome sequencing and/or targeted deep sequencing on serial samples from 31 sAML patients as well as targeted deep sequencing on an additional 93 patients, the latter consisting of a progressed and a non-progressed cohort. Our approach of analyzing genes grouped by the pathways they affect helps

¹Department of Computer Science, University of Toronto, Toronto, ON, Canada; ²The Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada; ³The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada; ⁴Genome Research Center for Hematopoietic Disease, Chonnam National University, Hwasun, Korea; ⁵Department of Hematology-Oncology, Chonnam National University Hwasun Hospital, Hwasun, Korea; ⁶Department of Internal Medicine, Yonsei University, Seoul, Korea; ⁷Department of Hematology-Oncology, Kyungpook National University, Daegu, Korea; ⁸Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea; ⁹Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada and ¹⁰Department of Medical Oncology, Princess Margaret Cancer Center, University Health Network, Faculty of Medicine, University of Toronto, Toronto, ON, Canada. Correspondence: Dr HJ Kim, Genome Research Center for Hematopoietic Disease, College of Medicine, Chonnam National University, 322 Seoyang-ro, Hwasun-eup, Hwasun-gun 519-763, Jeollanam-do, Republic of Korea or Dr Z Zhang, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College St., Rm 608, Toronto, ON, Canada M5S 3E1. E-mail: hjoonk@chonnam.ac.kr or zhaolei.zhang@utoronto.ca

Received 8 September 2016; revised 23 December 2016; accepted 30 December 2016; accepted article preview online 16 January 2017; advance online publication, 3 February 2017

de-convolute the heterogeneity and reveal generalized patterns of the disease model. In the following study, we present the dynamics of the VAFs of somatic variants across different disease stages and demonstrate correlations between associated biological pathways and leukemic progression using serial samples.

SUBJECTS AND METHODS

Cohorts

This study examined several cohorts with a combined total of 124 patients (Table 1, Supplementary Table S1 and Supplementary Methods, Section A.1). The discovery cohort (C1) consists of 31 patients diagnosed with antecedent hematologic malignancies who all progressed to sAML. Whole-exome sequencing was performed for each case in C1 on fractionated T-cell samples as controls (CD3+ fraction) and bone marrow samples taken at the diagnosis of the antecedent malignancy and after sAML progression. The other cohorts include 72 non-progressed MDS patients (C2a, median follow-up 3.5 years) and an additional (non-overlapping) 21 sAML patients progressed to sAML from MDS (C2b), for whom samples from the MDS stage were not available. Targeted deep sequencing was performed on all cohorts.

Sample preparation, sequencing, variant calling and computational analysis

The study was approved by the Research Ethics Board at Chonnam National University (Hwasun, Korea), Kyungpook National University (Daegu), Korea and Yonsei University (Seoul), Korea. All sequencing data are deposited at the European Nucleotide Archive (Study Accession: PRJEB18698). T-cell (CD3+ fraction) fractionation was performed using the MACS separation column (25 MS Columns; Milteniy Biotec, Bergisch

Gladbach, Germany). DNA for all 258 bone marrow mononuclear cells was extracted using the QIAamp DNA Blood Mini Kit (Qiagen, Venlo, Netherlands). Whole-exome sequencing (Agilent SureSelect v4) was performed on the 93 samples from the discovery cohort as per the manufacturer's protocol using an Illumina HiSeq 2000 sequencer (Santa Clara, CA, USA). Exome sequencing reads were processed for each stage using the methodology described in our previous case study³² and a list of significant variants was generated for each case in the discovery cohort. We selected 92 genes (Supplementary Table S2) for validation using targeted deep sequencing on all 258 samples from all of the cohorts. Targeted sequencing was performed using an Agilent custom probe set of the selected 92 genes. We multiplexed and sequenced the samples using an Illumina HiSeq 2000. The reads were processed in the same way as the whole-exome sequencing data. T-cell samples were used as controls for disease samples when calling variants for all cases in whole-exome sequencing data and all cases except for C2b in targeted sequencing data, and were also used to call CHIP (clonal hematopoiesis of indeterminate potential) variants (Supplementary Methods A.3 and Supplementary Figure S1). Sequencing metrics are detailed in Supplementary Tables S3 and S4. Our in-house variant calling pipeline and algorithms, pathway analysis, and subclonal analysis are described in Supplementary Method, Sections A.2–5.

RESULTS

Exome sequencing reveals somatic variants in MDS and secondary AML

Whole-exome sequencing was performed on a trio of T-cell, MDS and sAML samples for each patient in a discovery cohort of 31 patients (C1, median time until sAML progression of 1.2 years) to profile the origin and landscape of somatic variants (Figure 1 and

Table 1. Summary and cohort division of 124 patients involved in this study

Variable	sAML, discovery cohort	sAML, validation cohort	non-progressed MDS, validation cohort
Number of patients	31	21	72
Gender, male (%)	19 (61)	11 (52)	47 (65)
Age, median (range, years)	63 (37–74)	67 (52–79)	67 (31–84)
<i>Classification (ancestral disease or MDS) (%)</i>			
RA	4 (13)	1 (5)	5 (7)
RARS	0	1 (5)	0
RCMD	5 (16)	2 (10)	40 (56)
RAEB-1	6 (19)	7 (33)	10 (14)
RAEB-2	8 (26)	7 (33)	15 (21)
5q–	0	0	1 (1)
MDS/MPN unclassified	1 (3)	0	0
MDS-U	2 (6)	1 (5)	2 (3)
CMML	3 (10)	1 (5)	0
aCML	1 (1)	1 (5)	5 (7)
<i>Blood count at ancestral disease or non-progressed (at diagnosis)</i>			
WBC, median (range, × 10 ⁹ /l)	3.3 (0.6–157.1.0)	3.5 (1.5–394.0)	2.5 (1.1–9.6)
Hemoglobin, median (range, g/dl)	8.6 (3.5–13.1)	7.8 (4.6–11.9)	8.8 (5.5–16.5)
Platelet, median (range, × 10 ⁹ /l)	57.0 (5.0–380.0)	77.0 (25.0–599.0)	60.0 (5.0–566.0)
<i>Bone marrow feature at ancestral disease or non-progressed</i>			
BM blasts, < 5%	10 (32)	7 (33)	60 (83)
5 to < 10%	10 (32)	7 (33)	5 (7)
≥ 10%	11 (36)	7 (33)	7 (10)
karyotype			
Good ^a	18 (58)	13 (62)	44 (61)
Intermediate ^b	4 (13)	4 (19)	12 (17)
Poor ^c	3 (10)	4 (19)	7 (10)
No mitosis/not available	6 (19)	0	9 (12)
Duration from ancestral disease to sAML, median (range, months)	11.1 (1.6–87.4)	9.1 (1.6–57.9)	NA

Abbreviations: aCML, typical chronic myeloid leukemia; BM, bone marrow; CMML, chronic myelomonocytic leukemia; MDS, myelodysplastic syndrome; MDS/MPN, myelodysplastic/myeloproliferative neoplasms, unclassifiable; RA, refractory anemia; RAEB, refractory anemia with excess blasts; RARS, refractory anemia with ringed sideroblasts; RCMD, refractory anemia with multi-lineage dysplasia; sAML, secondary acute myeloid leukemia; WBC, white blood cell; NA, not available. ^aGood: normal, -Y, del(5q) alone, del(20q) alone. ^bPoor: complex (≥ abnormalities), chromosome 7 anomaly. ^cIntermediate: others.

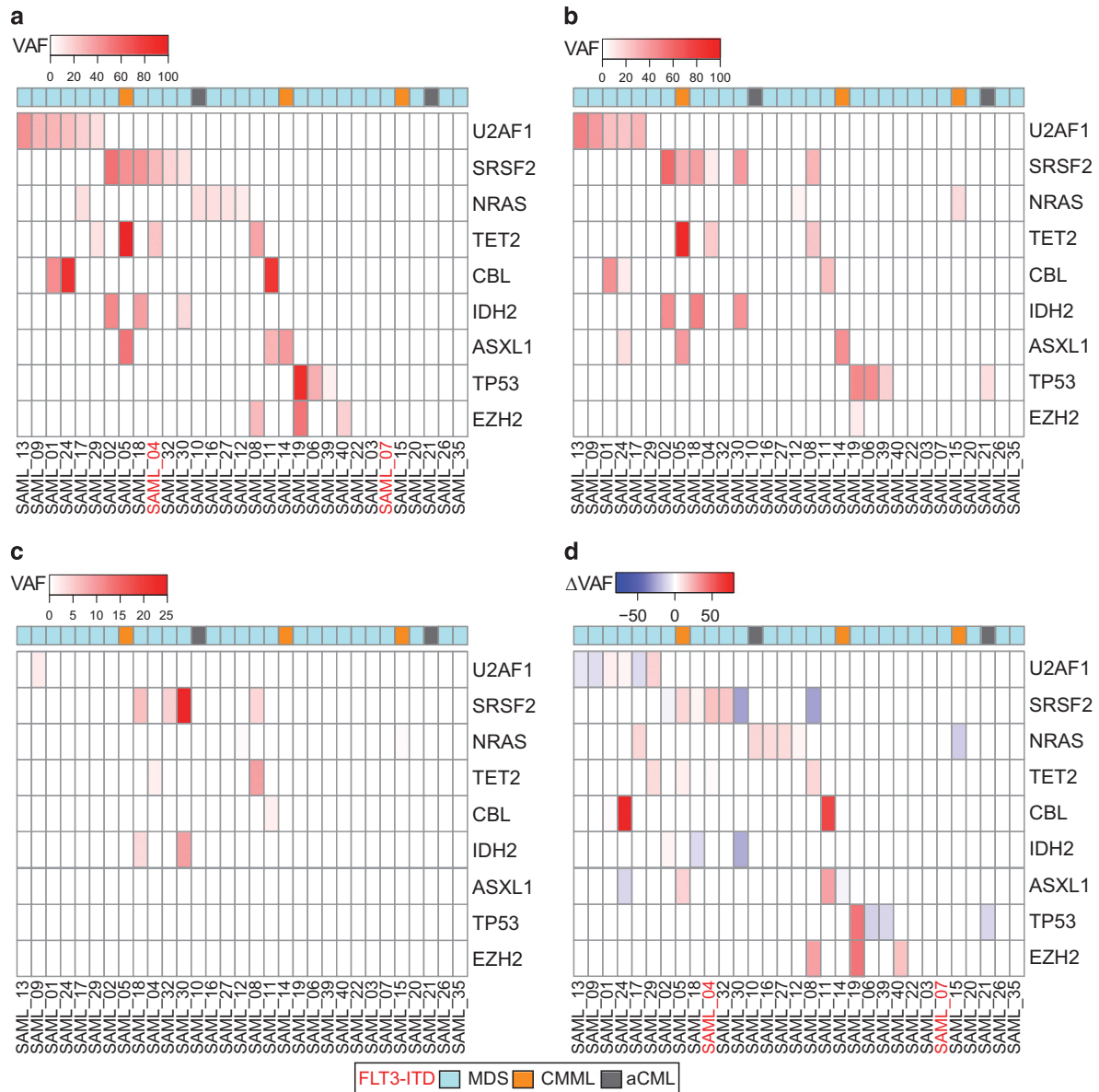


Figure 1. Gene-wise mutational landscape. Clustered by gene using the whole-exome data of (a) samples taken after sAML, (b) samples taken at MDS and (c) T-cell samples. (d) The change in VAF from MDS to sAML. The color scale corresponds to VAFs in (a–c), and the Δ VAFs in (d) Red ID labels indicate the patients that gained *FLT3-ITD* at sAML. Only genes that occur in at least two patients are shown for visual clarity.

Supplementary Table S5). The average read depth retrieved for target regions was 73 \times and at least 80% of the captured regions had a read coverage of 20 \times or greater for all 93 samples (Supplementary Table S4). After calling and prioritizing variants as described in Subjects and methods, we found a mean and median of 7.7 and 6 significant variants per patient at the time of initial diagnosis, and 12.4 and 10 variants after sAML progression, respectively (the mean depth of all significant variants was 109 \times). The mean VAF of the variants at the MDS stage expanded from 18.8 to 31.2% at the sAML stage (Supplementary Figure S2). In total, we detected 399 variants (see Supplementary Methods, Section A.2 for variant calling details). The 399 somatic mutations consist of 261 non-synonymous single nucleotide variants, 87 synonymous single nucleotide variants, 25 stop-gain mutations, 9 frameshift deletions, 5 frameshift insertions, 2 non-frameshift deletions and 10 splicing variants from 340 genes. Among these, 31 genes had variants in multiple patients. A total of 89 variants in

48 commonly mutated genes in myeloid disorders were further validated with a true positive rate of 100% using targeted deep sequencing in all 93 C1 samples (mean depth of 1224.24 \times and VAF correlation between two platforms measured by Pearson's Rho \sim 0.96, Supplementary Table S5 and Supplementary Figure S3). We also detected using Genomon-ITD that *FLT3-ITD* emerged in two patients (SAML-04 and SAML-07) after sAML progression.³³

Pathway analyses dissect the order of stepwise somatic variant acquisition in relevant biological pathways

We compared the proportion of the 31 cases having variants in genes affecting each pathway (see Table 2 for the gene lists per pathway) at each stage (Figure 2a). Consistent with previous studies, there are a significant number of cases with variants in genes involved in DNA methylation and/or splicing machinery (35.5% and 48.3%, respectively).^{10,11} However, the portion of cases with variants affecting these pathways increases significantly at

Table 2. Eight commonly mutated pathways in AML and genes involved in each of them

Pathway	Genes
NPM1	NPM1
Tumor suppressors	TP53, WT1, PHF6, DDX41
DNA methylation	IDH1, IDH2, TET2, TET1, DNMT1, DNMT3A, DNMT3B
Activated signaling	FLT3, KIT, ABL1, DYRK4, EPHA2, EPHA3, JAK3, MST1R, PDGFRB, WEE1, ACVR2B, ADRBK1, AKAP13, BUB1, CPNE3, DCLK1, MAPK1, YLK2, MYO3A, NRK, PRKCG, RPS6KA6, SMG1, STK32A, STK33, STK36, TRIO, TTBK1, WNK3, WNK4, PTPN11, PTPRT, PTPN14, KRAS, NRAS, MPL, JAK2, CALR, CBL, CSF3R, RYR2, MTOR, SH2B3, SPRED2, GNA12
Spliceosome	CSTF2T, DDX1, DDX23, DHX32, HNRNPK, METTL3, PLRG1, PRPF3, PRPF8, RBMX, SF3B1, SNRNP200, SRRM2, SRSF6, SUPT5H, TRA2B, U2AF1, U2AF1L4, U2AF2, SRSF2, ZRSR2
Cohesin complex	SMC1A, SMC3, SMC5, STAG2, RAD21
Myeloid TFs	RUNX1, CEBPA, GATA2, CBFβ, ETV6, ETV3, GLI1, IKZF1, MYB, MYC, MLLT10, CEP164
Chromatin modifiers	ASXL1, EZH2, KMD6A, ARID4B, ASXL2, ASXL3, BRPF1, CBX5, CBX7, EED, HDAC2, HDAC3, JMJD1C, KAT6B, KDM2B, KDM3B, MLL2, MLL3, MTA2, PRDM9, PRDM16, RBBP4, SAP130, SCML2, SUDS3, SUZ12, ZBTB33, ZBTB7B, CREBBP, RPN1, MECOM, KAT6A, BCOR

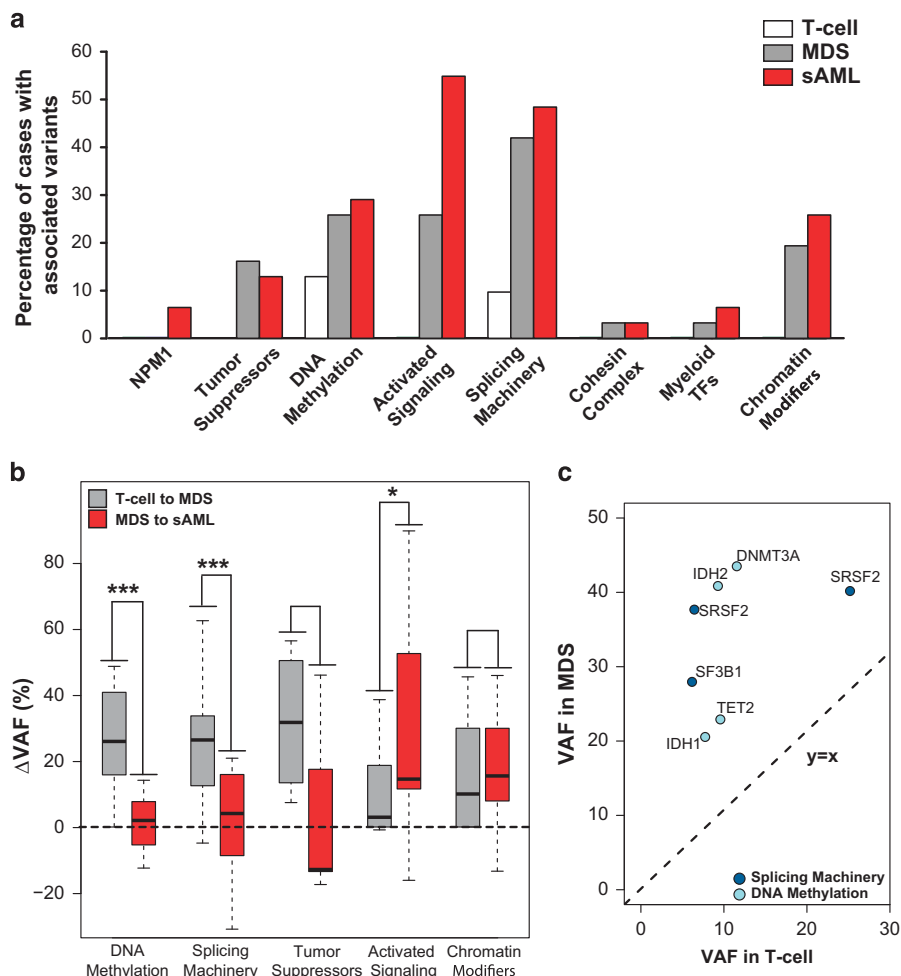


Figure 2. Pathway hierarchy in the discovery cohort. Each case was considered to have affected pathway if it had at least one variant with a minimum VAF of 5% in a gene associated with that pathway. **(a)** Portion of cohort with mutated pathways in each stage. **(b)** Inter time point change of VAFs in mutated pathways. **(c)** Expansion of CHIP variants by MDS.

the MDS stage, but does not change much by the sAML stage. This suggests that variants within DNA methylation and splicing machinery contribute to the development of MDS but not to sAML progression. On the other hand, variants in genes involved in activated signaling pathways show a distinctive pattern. The portion of cases with variants affecting activated signaling

pathways noticeably increases at the sAML step (25.8 to 54.8%). The changes in VAF between stages within cases of these variants reveal a similar pattern (Figure 2b). Indeed, the changes from T-cells to MDS VAFs are significantly distinct compared with the changes from antecedent malignancy to sAML VAFs in DNA methylation, splicing machinery and signaling pathway variants

(two-sample Kolmogorov–Smirnov test, $P < 0.002$, $P < 0.004$ and $P < 0.013$, respectively), though the change between time points in the case of signaling pathways was in the opposite direction of the other two biological pathways. Mutations affecting DNA methylation or splicing machinery were also detected in the T-cell samples of five patients with VAFs that all expanded by the diagnosis of the MDS (Figure 2c). The distinct behaviors of these pathways allow us to infer that there exists a general trend in the multi-step mutation acquisition process leading to sAML progression.

These findings are consistent with previous studies, which have shown that somatic variants in genes involved in activated signaling pathway are enriched in *de novo* AML.²⁰ Changes in VAFs of splicing machinery variants (Figure 3a) and DNA methylation variants (Figure 3b) over disease stages have very similar patterns (general linear model with repeated measures insignificant, P -value ~ 0.357), while both have a distinct pattern compared with the change of VAFs in activated signaling variants (general linear model with repeated measures, P -values ~ 0.044 and ~ 0.012 ; Figure 3c). The VAF changes of variants associated with chromatin modifiers show a mixed pattern, suggesting that they can be either early or late events (Figure 3d).

Tracing the origin of somatic variants using serial samples

In our pathway analysis, we used a threshold of a 5% VAF when marking a pathway as 'mutated' in a given sample at a given stage. We also investigated the origin of somatic variants in the three pathways with distinct inter-stage patterns. As was done in Klco *et al*,³⁴ we applied a threshold of 2.5% (which means 5% of cells carry the variant for a heterozygous mutation) to search for

evidence of early development of mutations in cases where they were insignificant in one or two stages but significant in the other one or two. For each variant, we examined all disease stages prior to the time of clear detection level (VAF > 5%). Importantly, none of the 20 mutations related to activated signaling pathways had detectable origins in T-cell samples (measured VAFs range from 0 to 0.82%) and only one variant not already deemed significant by the MDS stage had evidence of originating at that time (*NRAS-G13D*, 4.81% in whole-exome sequencing and 3.84% in targeted sequencing). Among the eight total mutations related to DNA methylation (there were four additional variants already significant in the T-cell samples), three of them were detected at more than 2.5% (3.16%, 4.76% and 3.5% in *TET2*, *DNMT3A* and *IDH2*, respectively). VAFs of all three were also measured at higher than 2.5% in the targeted deep sequencing data (4.00, 3.55 and 7.25%). Similarly, 2 of the 16 mutations related to splicing machinery were also detected with VAFs higher than 2.5% (4.35 and 4.55% in *SRSF2*) in T-cell samples (three additional variants had already been deemed significant in the T-cell samples in our pathway analysis). Overall, these tendencies show that the mutations related to activated signaling pathways occur at a later stage than the other two commonly mutated pathways. sAML progression is therefore not likely to directly originate from genetic characteristics of preleukemic clones.

Mutation profiles of progressed and non-progressed MDS patients support the disease model

We further validated our observations of stage-specific genetic traits using two independent non-overlapping cohorts consisting

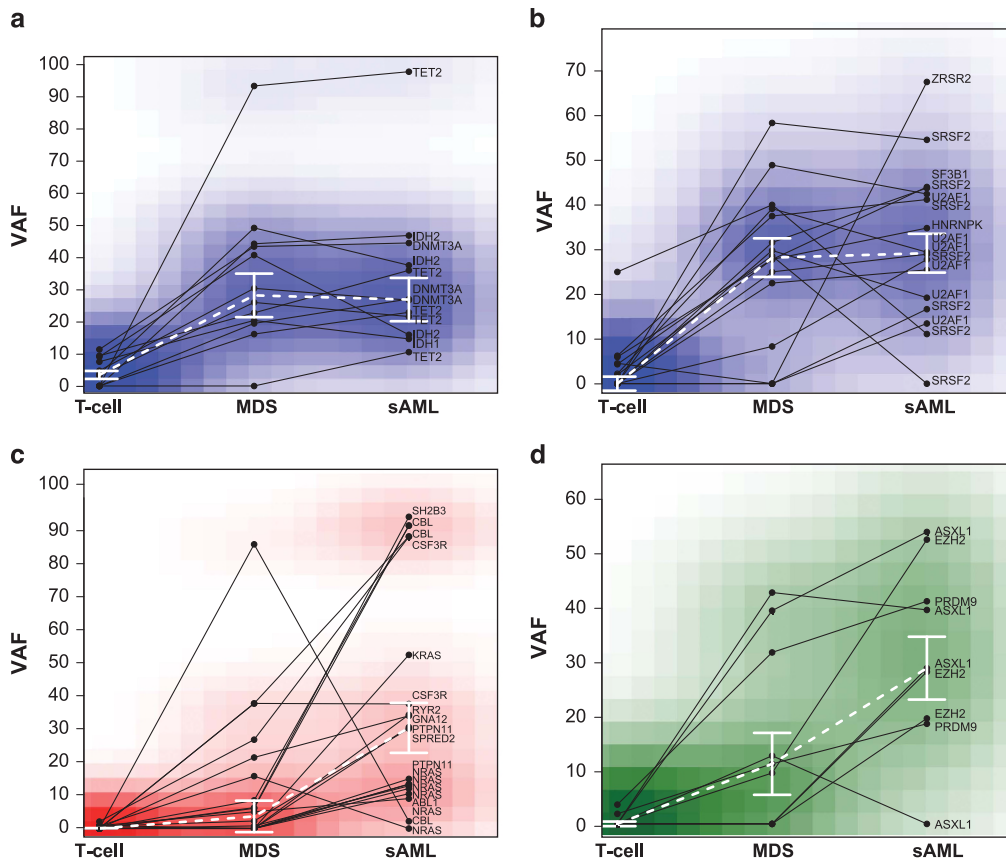


Figure 3. VAF dynamics of pathways across stages. Dynamics of VAFs at three time points of variants associated with (a) splicing machinery, (b) DNA methylation, (c) activated signaling and (d) chromatin modifier. The white points mark the medians per stage point, and spread marks indicate standard error.

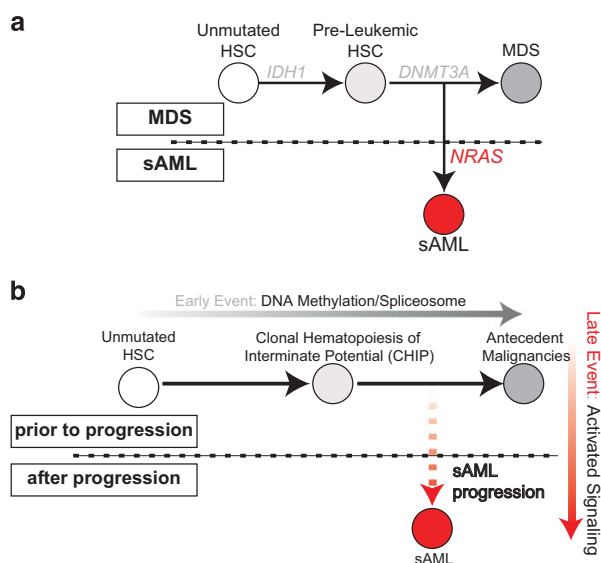


Figure 4. Clonal evolution of sAML. (a) Clonal evolution of a single sAML patients (SAML-12). (b) Generalized pattern of clonal evolution observed in this study.

of progressed and non-progressed MDS patients (C2b and C2a). As such, 72 additional MDS patients who did not progress to sAML (C2a) and 21 sAML patients progressed from MDS (C2b) were included for targeted deep sequencing (Supplementary Figures S4 and S5 and Supplementary Tables S4 and 6, mean on-target coverage of 808x). Using Fisher's exact test, no significant differences in proportion were noted between the two groups in DNA methylation (23.8% vs 20.8% in C2b vs C2a, not significant) and in splicing machinery (23.8% vs 20.8% in C2b vs C2a, not significant); however, there was a significantly higher frequency of variants in activated signaling pathway genes in C2b compared with those in C2a (19.0% vs 2.8%, $P < 0.03$). As in the discovery cohort where 16% of patients harbored variants in T-cell samples, we found 20 out of 72 patients in the non-progressed cohort (C2a) with variants in the T-cell samples, 14 of which included variants in DNA methylation and splicing machinery (7/20 in DNA methylation, 7/20 in splicing machinery exclusively), and one of which included a variant in activated signaling pathways (*KRAS*). Eighteen of these 20 patients only carried one detected variant at this stage (Supplementary Figure S6).

DISCUSSION

Our study describes the general pattern of sAML progression of hematologic malignancies using mutational profiles of serial samples. It is clear that sAML progression is associated with an increased mutation burden in terms of the number of variants and/or the VAFs (Figure 1). In particular, the mutation burden affecting activated signaling pathways increases significantly as MDS progress to sAML. On the other hand, mutations associated with DNA methylation and splicing machinery including pre-leukemic mutations increase significantly with the development of MDS but not during sAML progression.

Clonal origin of sAML from a pathway perspective

Many of the clonal analyses results of the cases in the discovery cohort had similar patterns to each other. We can posit two general clonal categories based on their dynamics across disease stages. The first contains pathways correlated with the development of the MDS, and includes clones driven by mutations in

splicing machinery and DNA methylation. The second is characterized by the patterns shown in activated signaling variants, which undergo expansion during sAML progression. These two clear patterns are consistent with previous studies.^{19,30} Pathways such as chromatin modification, which show mixed patterns in different cases with different particular variants, may have subdivisions that belong to either category.

Ten distinct clonal evolution patterns can be postulated based on the acquisition/expansion of mutations related to the three signature pathways (Figure 4 and Supplementary Figure S7 and Supplementary Methods A.5). Sixteen percent of patients develop the MDS from preleukemic mutations associated with DNA methylation or splicing machinery, and 26% are first seen to develop clones of this category at the MDS stage (a total of 42%). Forty-eight percent of patients show growth or development of clones containing activated signaling pathway variants at the sAML stage. A further 10% had clones with activated signaling pathways that did not significantly grow at the sAML stage, and these notably lacked observed variants from any of the other seven pathways. Twenty-two percent of patients carry only mutations associated with other commonly mutated pathways in AML or no mutations, but they were not observed frequently enough to generalize their patterns.

De novo AML vs secondary AML

Previous studies on de novo AML have already shown notably frequent occurrence of mutations in activated signaling pathways.^{6,11} Recent work by Lindsley *et al.*¹⁹ concluded that variants in splicing machinery and chromatin modifiers are sAML-specific when compared with de novo AML and therapy-related AML. In our study, we also noted a frequent occurrence of splicing machinery mutations at the MDS clone, which indeed carries over with a similar rate in the sAML clone. Notably, however, the VAFs of DNA methylation pathway variants and splicing machinery variants plateaued during the progression, implying their relatively weak association with sAML progression (Figures 3a and b).

Patients with preleukemic mutations

The presence of variants in T-cell samples in 5 C1 and 20 C2a patients provides evidence on the relative timing of early events for a subset of patients. The T-cell variants with pathway associations are all involved in DNA methylation (*DNMT3A*, *IDH1/2* and *TET2*) or splicing machinery (*SRSF2* and *SF3B1*). Both cohorts notably lack activated signaling pathway variants at this stage (with one exception in C2a). These patients show evidence of having CHIP, a preleukemic condition recently proposed by Steensma *et al.*³⁵ CHIP refers to the condition where a somatic variant is present but failed to meet any diagnostic criteria of hematologic malignancies. A recent large-scale study demonstrated that variants associated with CHIP significantly increase the risk of hematologic cancer (hazard ratio = 11.1).³⁶ Most variants we found in T-cell samples are consistent with this study (*TET2*, *DNMT3A*, *IDH1/2*, *U2AF1/2*, *SRSF2*, *SF3B1*, *ASXL1*, *BCOR*, *JAK2* and *TP53*).³⁵ Interestingly, 24% of the patients in our cohorts (25/103) develop the MDS from such variants, suggesting that there could be a link between preleukemic mutations and hematologic malignancies. However, VAFs of these variants only expand significantly at the time of initial diagnosis of the myeloid disorder, suggesting that they may be involved in the development of MDS but might not play a direct role in the sAML progression.

Overview and future direction

Comprehensive analyses in this work present a novel overview of sAML progression by investigating groups of genes associated with biological pathways in serial samples, rather than focusing only on individual variants at single disease stages from unrelated

individuals. Our analyses show that antecedent malignancies can be thought of as a subtype of preleukemic clones that are also themselves diseases. Our work shows clear patterns of clonal evolution from a pathway perspective and distinct roles of different biological pathways at different disease stages.

Multi-omic profiling of other genetic or epigenetic makeups such as RNA and bisulfite sequencing could provide further insight into progression, in particular for patients who do not acquire new mutations between stages. In addition, further validation in independent cohorts is an important next step in expanding on these discoveries. Our findings also call for functional studies on how somatic variants in the genes that affect activated signaling pathway are related to sAML progression. Since sAML ultimately appears to be distinguished from de novo AML by the variants acquired during the development of the MDS, functional studies might also offer insight into how sAML clones containing variants from MDS differ from de novo AML clones that lack them, as well as insight into their poor prognosis.

CONCLUSION

Our study dissects the distinct genetic characteristics of the development of MDS and their sAML progression. Progression to sAML is associated with additional mutation burden in genes affecting activated signaling pathways regardless of the subtype of the antecedent hematologic malignancies. This is consistent with the known biology behind de novo AML, thus sAML progression of antecedent malignancies is likely to be driven by altered activated signaling pathways.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by a grant from the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT and Future Planning (NRF-2015R1A2A1A10054579) and the Leading Foreign Research Institute Recruitment Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT and Future Planning (NRF-2011-0030034). The biospecimens and data used for this study were provided by the Biobank of Chonnam National University Hwasun Hospital, a member of the Korea Biobank Network. TK is supported by the Queen Elizabeth II Graduate Scholarship for Science and Technology (University of Toronto) and the scholarship from Natural Science and Engineering Research Council of Canada (NSERC-PGS D2). We thank Dr Aaron Schimmer and Dr Meong Hi Son for their insightful discussions and advice on the manuscript. In addition, they would like to thank the reviewers for their valuable comments and suggestions.

AUTHOR CONTRIBUTIONS

TK, HK, ZZ and DK designed the study and its concept. HP, SHC, JA, YK, DY, JL, SJ, SYK, YHM, JC, SKS and JHM collected the samples and performed experiments. YK, MST, MC, ML and ZZ analyzed the sequencing data. TK, MST, ZZ and DK performed all downstream and statistical analyses. TK, MST, HK, ZZ and DK wrote the manuscript.

REFERENCES

- Greenberg P, Cox C, LeBeau MM, Fenaux P, Morel P, Sanz G et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood* 1997; **89**: 2079–2088.
- Tefferi A, Vardiman JW. Classification and diagnosis of myeloproliferative neoplasms: The 2008 World Health Organization criteria and point-of-care diagnostic algorithms. *Leukemia* 2008; **22**: 14–22.
- Jaffe ES, Harris NL, Stein H, Vardiman JW. *Tumours of haematopoietic and lymphoid tissues*. WHO Classification of Tumours. WHO, 2001.
- Grimwade D, Ivey A, Huntly BJP. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. *Blood* 2016; **127**: 29–41.

- Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med* 2016; **374**: 2209–2221.
- Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013; **368**: 2059–2074.
- Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science* 2013; **339**: 1546–1558.
- Greenberg PL, Stone RM, Bejar R, Bennett JM, Bloomfield CD, Borate U et al. Myelodysplastic syndromes, version 2.2015. *J Natl Compr Canc Netw* 2015; **13**: 261–272.
- Bejar R. Myelodysplastic syndromes diagnosis: what is the role of molecular testing? *Curr Hematol Malig Rep* 2015; **10**: 282–291.
- Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 2011; **478**: 64–69.
- Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 2013; **122**: 3616–3627.
- Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 2014; **28**: 241–247.
- Delhommeau F, Dupont S, Valle Della V, James C, Trannoy S, Massé A et al. Mutation in TET2 in myeloid cancers. *N Engl J Med* 2009; **360**: 2289–2301.
- Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G et al. Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med* 2011; **364**: 2496–2506.
- Walter MJ, Shen D, Shao J, Ding L, White BS, Kandoth C et al. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia* 2013; **27**: 1275–1282.
- Malcovati L, Papaemmanuil E, Bowen DT, Boulwood J, Porta Della MG, Pascutto C et al. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood* 2011; **118**: 6239–6246.
- Bejar R, Stevenson KE, Caughey B, Lindsley RC, Mar BG, Stojanov P et al. Somatic mutations predict poor outcome in patients with myelodysplastic syndrome after hematopoietic stem-cell transplantation. *J Clin Oncol* 2014; **32**: 2691–2698.
- Papaemmanuil E, Cazzola M, Boulwood J, Malcovati L, Vyas P, Bowen D et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med* 2011; **365**: 1384–1395.
- Lindsley RC, Mar BG, Mazzola E, Grauman PV, Shareef S, Allen SL et al. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood* 2015; **125**: 1367–1376.
- Corces-Zimmerman MR, Hong W-J, Weissman IL, Medeiros BC, Majeti R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc Natl Acad Sci USA* 2014; **111**: 2548–2553.
- Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* 2014; **506**: 328–333.
- Bejar R. Clinical and genetic predictors of prognosis in myelodysplastic syndromes. *Haematologica* 2014; **99**: 956–964.
- Cazzola M, Porta Della MG, Malcovati L. The genetic basis of myelodysplasia and its clinical relevance. *Blood* 2013; **122**: 4021–4034.
- Cargo CA, Rowbotham N, Evans PA, Barrans SL, Bowen DT, Crouch S et al. Targeted sequencing identifies patients with preclinical MDS at high risk of disease progression. *Blood* 2015; **126**: 2362–2365.
- Meggendorfer M, De Albuquerque A, Nadarajah N, Alpermann T, Kern W, Steuer K et al. Karyotype evolution and acquisition of FLT3 or RAS pathway alterations drive progression of myelodysplastic syndrome to acute myeloid leukemia. *Haematologica* 2015; **100**: e487–e490.
- Shih L-Y, Huang C-F, Wang P-N, Wu J-H, Lin TL, Dunn P et al. Acquisition of FLT3 or N-ras mutations is frequently associated with progression of myelodysplastic syndrome to acute myeloid leukemia. *Leukemia* 2004; **18**: 466–475.
- van Kamp H, de Pijper C, Verlaan-de Vries M, Bos JL, Leeksa CH, Kerkhofs H et al. Longitudinal analysis of point mutations of the N-ras proto-oncogene in patients with myelodysplasia using archived blood smears. *Blood* 1992; **79**: 1266–1270.
- Horiike S, Misawa S, Nakai H, Kaneko H, Yokota S, Taniwaki M et al. N-ras mutation and karyotypic evolution are closely associated with leukemic transformation in myelodysplastic syndrome. *Leukemia* 1994; **8**: 1331–1336.
- Dicker F, Haferlach C, Sundermann J, Wendland N, Weiss T, Kern W et al. Mutation analysis for RUNX1, MLL-PTD, FLT3-ITD, NPM1 and NRAS in 269 patients with MDS or secondary AML. *Leukemia* 2010; **24**: 1528–1532.
- Makishima H, Yoshizato T, Yoshida K, Sekeres MA, Radivoyevitch T, Suzuki H et al. Dynamics of clonal evolution in myelodysplastic syndromes. *Nat Genet* 2016; e-pub ahead of print 19 December 2016; doi:10.1038/ng.3742.
- Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K et al. Clonal architecture of secondary acute myeloid leukemia. *N Engl J Med* 2012; **366**: 1090–1098.
- Kim TH, Yoshida K, Kim YK, Tyndel MS, Park HJ, Choi SH et al. Clonal dynamics in a single AML case tracked for 9 years reveals the complexity of leukemia progression. *Leukemia* 2015; **30**: 295–302.

- 33 Chiba K, Shiraishi Y, Nagata Y, Yoshida K, Imoto S, Ogawa S *et al*. Genomon ITDetector: a tool for somatic internal tandem duplication detection from cancer genome sequencing data. *Bioinformatics* 2015; **31**: 116–118.
- 34 Klco JM, Miller CA, Griffith M, Petti A, Spencer DH, Ketkar-Kulkarni S *et al*. Association between mutation clearance after induction therapy and outcomes in acute myeloid leukemia. *JAMA* 2015; **314**: 811–822.
- 35 Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP *et al*. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* 2015; **126**: 9–16.
- 36 Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG *et al*. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* 2014; **371**: 2488–2498.

Supplementary Information accompanies this paper on the Leukemia website (<http://www.nature.com/leu>)