Effects of Somatic Mutations Are Associated with SNP in the Progression of Individual Acute Myeloid Leukemia Patient: The Two-Hit Theory Explains Inherited Predisposition to Pathogenesis

Soyoung Park[1], Youngil Koh[1,2], Sung-Soo Yoon[1,2,3]*

[1]Cancer Research Institute, Seoul National University College of Medicine, Seoul 110-799, Korea,
[2]Department of Internal Medicine, Seoul National University Hospital,
Seoul National University College of Medicine, Seoul 110-799, Korea,
[3]Clinical Research Institute, Seoul National University Hospital, Seoul 110-799, Korea

This study evaluated the effects of somatic mutations and single nucleotide polymorphisms (SNPs) on disease progression and tried to verify the two-hit theory in cancer pathogenesis. To address this issue, SNP analysis was performed using the UCSC hg19 program in 10 acute myeloid leukemia patients (samples, G1 to G10), and somatic mutations were identified in the same tumor sample using SomaticSniper and VarScan2. SNPs in *KRAS* were detected in 4 out of 10 different individuals, and those of *DNMT3A* were detected in 5 of the same patient cohort. In 2 patients, both *KRAS* and *DNMT3A* were detected simultaneously. A somatic mutation in *IDH2* was detected in these 2 patients. One of the patients had an additional mutation in *FLT3*, while the other patient had an *NPM1* mutation. The patient with an *FLT3* mutation relapsed shortly after attaining remission, while the other patient with the *NPM1* mutation did not suffer a relapse. Our results indicate that SNPs with additional somatic mutations affect the prognosis of AML.

Keywords: acute myeloid leukemia, high-throughput nucleotide sequencing, point mutation, single nucleotide polymorphism

Introduction

Acute myeloid leukemia (AML) is the most common type of leukemia among hematologic malignances. This disease is a clonal hematopoietic disease that is caused by either inherited or acquired genetic mutations. AML is known to be a molecularly and clinically heterogeneous disease [1]. In order to cure this disease, cytosine-arabinoside (Ara-C)-based chemotherapy is widely used. Two thirds of AML patients who receive Ara-C based chemotherapy regimen, will achieve complete remission (CR). However, more than 50% of patients suffer a relapse after attaining CR [2]. Eventually, a great majority of AML patients die from progressive disease after relapse by various genetic changes that will endow tumor cells with resistance against most commonly used anticancer agents. Therefore, it is prerequisite to identify prognostic markers that will help in deciding the suitable treatment modalities. A lot of studies have focused on finding somatic mutations in cancer cells by whole-exome sequencing (WES). However, these studies did not fully explain the mechanisms of how disease relapses.

In the initial development of cancer in humans, Knudson's two-hit hypothesis has provided important insights into the pathogenesis of tumors in autosomal dominant tumor predisposition syndromes [3]. With an analogy, we focused on the additional genetic changes that will predispose susceptible individuals with a pre-existing genetic abnormality to develop leukemia. To address this issue, germline single nucleotide polymorphism (SNP) analysis was per-

formed using the UCSC hg19 program in 10 AML patients (samples, G1 to G10), and somatic mutations were identified in the same tumor sample using SomaticSniper and VarScan2. Our results suggest that SNPs with additional somatic mutations may directly affect the prognosis of AML and emphasize the need to validate in a larger patient cohort.

Methods

DNA sample preparation (normal and tumor samples)

Genomic DNA samples were prepared for WES. They were freshly obtained from AML patients' bone marrow at initial diagnosis. Subsequent bone marrow samples were aspirated if necessary according the disease status of the patients. AML blast tumor cells were isolated from bone marrow of each patient using Ficoll gradient separation. Matching patients' saliva samples were collected as germline controls. Each tumor and control DNA was extracted via standard protocols of DNA purification. DNA samples for validation were obtained from our frozen stock as needed.

Whole-exome sequencing

We prepared and sequenced exomes using the Solexa sequencing technology platform (HiSeq2000; Illumina, San Diego, CA, USA) following the manufacturer's instructions. We randomly sheared 3 μg of genomic DNA using the Covaris System to generate 150-bp inserts. The fragmented DNA was end-repaired using T4 DNA polymerase and Klenow polymerase, and Illumina paired-end adaptor oligonucleotides were ligated to the sticky ends. We analyzed the ligation mixture by electrophoresis on an agarose gel and sliced and purified fragments of 200–250 bp. The purified DNA library was hybridized with the SureSelect Human All Exons probes set (Agilent, Santa Clara, CA, USA) to capture 50 Mb of targeted exons following the manufacturer's

instruction. We prepared the HiSeq2000 paired-end flowcell with the manufacturer's protocol using the captured exome library. Clusters of PCR colonies were then sequenced on the HiSeq2000 platform using recommended protocols from the manufacturer.

Read mapping and alignment and variant analysis

We recently developed an in-house analysis pipeline for cancer genome sequencing data. This pipeline essentially includes 1) mapping and alignment, 2) discovery of sequence nucleotide variations (SNVs) and insertions and deletions (indels), and 3) filtering and annotation of SNVs and indels. Briefly, once the raw sequence data were created, the output short reads were aligned to a reference genome (NCBI human genome assembly build 36) using UCSC hg19. Each alignment was assigned a mapping quality score by UCSC hg19, which is the Phred-scaled probability that the alignment is incorrect. The PCR duplicates were detected and removed by SAMTOOLS (http://samtools.sourceforge.net/). After alignment, we used GATK v1.4 to call SNVs for each chromosomal position. Filtering and annotation of SNVs and indels were performed using ANNOVAR (http://www.openbioinformatics.org/annovar/). We defined somatic mutation SNVs as those detected by both SomaticSniper and VarScan2 with a somatic p-value < 0.05. We also set a minimum coverage of 5 in both normal and tumor samples. Conclusively, final somatic mutations were selected using the in-house filter.

Results and Discussion

Germline SNP analysis was done to detect germline heterozygous mutations that are carried by AML patients. To address this issue, germline SNP analysis was performed using the UCSC hg19 program in 10 AML patients (samples G1 to G10) (Table 1). We chose 18 major genes (*TET2*,

Table 1. Germline SNPs (*KRAS* and *DNMT3A*) in 10 AML patients (samples, G1 to G10)

Sample no.	Gene	Chrom	Type	Chrom start	Chrom end	Ref	Alt	Genotype	rs no.	Region	Gene position
G2	*KRAS*	chr12	SNV	25,362,776	25,362,777	A	G	AG	-	CDS	41,078
G6	*KRAS*	chr12	SNV	25,362,776	25,362,777	A	G	AG	-	CDS	41,078
G7	*KRAS*	chr12	SNV	25,362,776	25,362,777	A	G	GG	-	CDS	41,078
G8	*KRAS*	chr12	SNV	25,362,776	25,362,777	A	G	AG	-	CDS	41,078
G3	*DNMT3A*	chr2	SNV	25,469,501	25,469,502	C	T	CT	rs2276598	CDS	95,273
G4	*DNMT3A*	chr2	SNV	25,469,501	25,469,502	C	T	CT	rs2276598	CDS	95,273
G6	*DNMT3A*	chr2	SNV	25,469,501	25,469,502	C	T	CT	rs2276598	CDS	95,273
G8	*DNMT3A*	chr2	SNV	25,469,501	25,469,502	C	T	CT	rs2276598	CDS	95,273
G10	*DNMT3A*	chr2	SNV	25,469,501	25,469,502	C	T	TT	rs2276598	CDS	95,273

SNP, single nucleotide polymorphism; AML, acute myeloid leukemia; Chrom, chromosome; Ref, reference; Alt, alteration; SNV, single nucleotide variant.

Table 2. Distribution of germline SNPs (18 top genes)[a] in 10 AML patients (samples, G1 to G10)

Sample No.	Germline ID	TET2	ASXL1	DNMT3A	CEBPA	PHF6	WT1	TP53	EZH2	RUNX1	PTEN	FLT3	NPM1	HRAS	KRAS	NRAS	KIT	IDH1	IDH2
G1	201SD	-	o	-	-	-	o	o	-	-	-	o	-	o	-	-	-	-	-
G2	53SD	o	o	-	-	-	o	o	o	-	-	o	-	-	o	-	-	-	-
G3	218SD	o	o	o	-	-	o	-	-	-	-	o	-	-	-	-	-	-	-
G4	95SD	o	o	o	-	-	-	o	-	-	-	o	-	-	-	-	-	-	-
G5	219SD	-	o	-	-	-	o	o	-	-	-	o	-	-	-	-	-	-	-
G6	132SD	o	o	o	-	-	o	o	-	-	-	o	-	-	o	-	-	-	-
G7	206SD	o	o	-	-	-	o	o	-	-	-	o	-	-	o	-	o	-	-
G8	605SD	o	o	o	-	-	o	o	o	-	-	o	-	-	o	-	-	-	-
G9	234SD	o	o	-	-	-	o	o	-	-	-	o	-	-	-	-	o	-	-
G10	162SD	o	o	o	-	-	o	o	-	-	-	o	-	-	-	-	o	-	-

SNP, single nucleotide polymorphism; AML, acute myeloid leukemia.
[a]*TET2, ASXL1, DNMT3A, CEBPA, PHF6, WT1, TP53, EZH2, RUNX1, PTEN, FLT3, NPM1, HRAS, KRAS, NRAS, KIT, IDH1,* and *IDH2.*

Table 3. Somatic mutation calling results in 10 AML patients (samples, G1 to G10)

		G1_S1	G2_S1	G3_S1	G4_S1	G5_S1	G6_S1	G7_S1	G8_S1	G9_S1	G10_S1
Germline	*KRAS*	-	o	-	-	-	o	o	o	-	-
	KIT	-	-	-	-	-	-	o	-	o	o
	DNMT3A	-	-	o	o	-	o	-	o	-	o
		NFKBIZ	CXorf57	MEF2A	PPT1	FAT1	HIST1H3B	MUC17	MYH13	CACNB4	FER1L5
			ZNF84	POLR3B	VEGFC	THSD7B	KIAA0922	CECR2	IDH2	KCNH1	PSG11
			FKBP4	POLR3B	SCAMP1	ANKRD5	TEX30	TET2	NPM1	C1orf65	SSX1
			NRG3	CNNM3	TRMT12		DNAH17	TCF12	ALDH6A1	RPS6KA1	GPR126
			C20orf26	LYAR	DCAF8L1		C3	CLSTN1	UBE2A	RAD21	ANKHD1
			IFT140	PCDHB3			LHFPL1		NEB	NFX1	ZFHX4
			PRSS12	KRTAP4-6			IDH2		HSPA12B	KSR2	COBLL1
			ARHGAP10	MUC4			SPTBN5			FRG1	MUC16
			GABBR2	MUC5B			ITPR3			CYLD	FRAS1
			BLNK	MUC21			FLT3			KRTAP4-6	GPR98
			TPH2	ZNF516			RUNX1				PTPN13
			IDH2	NOS1			MYCBP2				DZANK1
			MUC4								ZSCAN2
			NPM1								CLMP

Germline SNPs of *KRAS, KIT,* and *DNMT3A* are indicated.
G2, 29 somatic mutations detected; G10, 129 somatic mutations detected, cutoff to fit the following table.
AML, acute myeloid leukemia; SNP, single nucleotide polymorphism.

ASXL1, DNMT3A, CEBPA, PHF6, WT1, TP53, EZH2, RUNX1, PTEN, FLT3, NPM1, HRAS, KRAS, NRAS, KIT, IDH1, and *IDH2*) to be analyzed in 10 samples (Table 2). Among these 18 genes, *KRAS* and *DNMT3A* seemed to play a critical role in the progression of AML, associated with the chemotherapy response. *KRAS* is a protein product of the normal *KRAS* gene, which performs an essential function in normal tissue signaling, and the mutation of a *KRAS* gene is an essential step in the development of many cancers [4]. *DNMT3A* is a key component of the epigenetic regulation of genes, as it catalyzes the addition of methyl groups to the cytosine residue of CpG dinucleotides [5]. However, variations only in *KRAS* and *DNMT3A* are not enough to explain the mechanism of AML progression, since additional somatic changes exist in AML patients.

SNPs of *KRAS* were detected in 4 different individuals (G2, G6, G7, and G8), as shown in Table 1, while those of *DNMT3A* were detected in 5 individuals (G3, G4, G6, G8, and G10) among 10 AML patients (Table 1). Patients G6 and G8 showed SNPs in both *KRAS* and *DNMT3A*. We then

Table 4. Clinical course of patients

Sample no.	CR after 1st ind	DFS	Relapse
G1	1	625	0
G2	1	109	0
G3	1	388	0
G4	0	205	1
G5	1	306	1
G6	1	390	1
G7	0	N/A	Persistence
G8	1	402	0
G9	1	356	1
G10	1	340	0

G7, DFS is undetermined (N/A) due to its persistent status (not in complete remission).
CR, complete remission; DFS, disease-free survival.

searched for somatic mutations of each individual in order to prove the 'two-hit' theory. Table 3 shows the results of the somatic mutation analysis. SNPs of *KRAS* and *DNMT3A* were detected in both samples G6 and G8 (Table 3). However, they had different clinical courses after the first induction. They both attained a CR after the first induction; however, G6 relapsed within 390 days after attaining remission, while G8 did not relapse for 402 days. We suggested that an additional somatic mutation may have affected the clinical courses above. A somatic mutation in *IDH2* was identified in both G6 and G8, while that of *FLT3* and *NPM1* was detected in G6 and G8, respectively. *FLT3* mutation is known as a bad prognostic factor in AML [6]. *NPM1* mutation, on the other hand, is usually known as a good prognostic factor in AML [7]. Therefore, we could speculate that SNPs (*KRAS* and *DNMT3A*) with additional somatic mutations (*FLT3* and *NPM1*) affect the prognosis of AML differentially.

In addition, G2 and G7 shared the same SNPs in *KRAS* but their clinical courses showed a striking difference (Table 4). G7 had an additional germline SNP in *KIT*. This patient never attained CR despite intensive chemotherapy. As we searched further for somatic mutations in G7, a somatic mutation in *TET2* was detected as a second hit. On the other hand, an *NPM1* mutation was observed in G2, showing a divergent outcome compared with G7. G2 attained CR after the first induction of chemotherapy and did not relapse.

We observed a divergent outcome in patients having the same SNPs but with different somatic mutations as a second hit. Somatic mutations that were differentially detected at the time of diagnosis strongly suggest clinical implications and warrant future study in a larger cohort to enhance our understanding of the mechanisms of disease progression in AML.

Acknowledgments

References

1. Grossmann V, Tiacci E, Holmes AB, Kohlmann A, Martelli MP, Kern W, *et al*. Whole-exome sequencing identifies somatic mutations of *BCOR* in acute myeloid leukemia with normal karyotype. *Blood* 2011;118:6153-6163.
2. Song JH, Kim SH, Kweon SH, Lee TH, Kim HJ, Kim TS, *et al*. Defective expression of deoxycytidine kinase in cytarabine-resistant acute myeloid leukemia cells. *Int J Oncol* 2009;34: 1165-1171.
3. Tucker T, Friedman JM. Pathogenesis of hereditary tumors: beyond the "two-hit" hypothesis. *Clin Genet* 2002;62:345-357.
4. Liang H, Zhang J, Shao C, Zhao L, Xu W, Sutherland LC, *et al*. Differential expression of RBM5, EGFR and KRAS mRNA and protein in non-small cell lung cancer tissues. *J Exp Clin Cancer Res* 2012;31:36.
5. Pezzi A, Moraes L, Valim V, Amorin B, Melchiades G, Oliveira F, *et al*. DNMT3A mutations in patients with acute myeloid leukemia in South Brazil. *Adv Hematol* 2012;2012:697691.
6. Smith CC, Wang Q, Chin CS, Salerno S, Damon LE, Levis MJ, *et al*. Validation of ITD mutations in *FLT3* as a therapeutic target in human acute myeloid leukaemia. *Nature* 2012;485: 260-263.
7. Dohner K, Schlenk RF, Habdank M, Scholl C, Rucker FG, Corbacioglu A, *et al*. Mutant nucleophosmin (*NPM1*) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations. *Blood* 2005;106:3740-3746.