Identification of Mutations in GATA2 Gene in Adult Bengali Bangladeshi Acute Myeloid Leukemia (AML) Patients

Yesmin ZA[1*], Nishat L[2], Masum MM[3], Sharmin E[4], Abdullah-Al-Sayeef M[5]

[1,2,3]*Department of Anatomy Bangabandhu Sheikh Mujib Medical University, Bangladesh*
[4]*Department of Pharmacology, Bangabandhu Sheikh Mujib Medical University, Bangladesh*
[5]*Department of Anatomy, Magura Medical College, Magura, Bangladesh, Bangladesh*
**Corresponding Author*

Abstract: **AML is a malignant disease of the bone marrow in which hematopoietic precursors are arrested in an early stage of development due to genetic alterations in normal hematopoietic stem cells. Several genes are identified for the disease and *GATA2* gene is one of the important genes among them because mutations of *GATA2* gene suppress leukemogenesis. In a country like Bangladesh, genetic services are at an early stage of development or even not yet developed. There is no adequate reporting on *GATA2* gene mutations in adult Bengali Bangladeshi AML patients. The aim of the study was to determine the mutations of *GATA2* gene in AML patients in the adult Bengali Bangladeshi population. The present study was a cross-sectional type of descriptive study. The study was carried out in the Genetic Research Center (GRC) of the Department of Anatomy, Bangabandhu Sheikh Mujib Medical University (BSMMU), Bangladesh during the period of January 2018 to December 2018. The study includes 50 adult Bengali Bangladeshi AML patients. Informed written consent was obtained from the participants. For identifying the mutations peripheral venous blood was obtained from each patient and sequencing was done using ABI Sanger Sequencer 3730. Research shows the mean age is 25±14.2 and average age group of adult AML patients of Bengali Bangladeshis is 18-29 years (54%), among them 31 (62%) and 19 (38%) were females and males respectively. *GATA2* gene mutations were identified in 1 of the 50 AML patients and shows synonymous, frameshift and missense variant compared with the database of dbSNP. Identification of different variant with this small sample size emphasizes the importance of exploring the genetic makeup of adult Bengali Bangladeshi AML patients to develop a database for proper screening and genetic counseling of the disease.**

Keywords: ***GATA2*** **gene, Bengali Bangladeshi, AML**

I. INTRODUCTION

Acute myeloid leukemia (AML) is a malignant disease of the bone marrow. According to Dong et al. AML accounted for 18.0% of the total leukemia cases worldwide in 1990 [1]. Kirtane and Lee [2] reported that the incidence rate of AML varies among different races or geographic locations. Recently, with the development of methodologies of massive sequencing, different genes with new mutations associated with acute myeloid leukemia have been identified [3]. Among them, *GATA2* gene mutations are serving as powerful prognostic indicators and affect the treatment decision [4]. Bresnick et al. [5] revealed strong evidence that *GATA2* gene mutations cause hematologic disease. The human *GATA2* gene (14,579 bases, chr3:128,488,629-128,503,207) is located on 3q21.3 [6]. Hsu et al. [7] reported that the *GATA2* gene contains 7 exons. The two 5-prime exons are untranslated. Hahn et al. [8] analyzed *GATA2* gene in 5 families with acute myeloid leukemia and reported heterozygous missense and 3-bp deletion types of mutation in *GATA2* gene. So far there is no evidence of research on genetic analysis of *GATA2* gene among adult Bengali Bangladeshi AML patients. If any relationship with *GATA2* gene with AML, it will help to clarify the pathogenesis of the disease and provide new therapeutic and preventive strategies for the Bengali Bangladeshi population. For this reason, this research is undertaken to identify the mutations of *GATA2* gene in adult Bengali Bangladeshi AML patients.

II. METHODS

The study was carried out in the Genetic Research Center (GRC) of the Department of Anatomy, Bangabandhu Sheikh Mujib Medical University (BSMMU), Dhaka, Bangladesh during the period from January 2018 to December 2018. The study was conducted after getting the Institutional Review Board (IRB) clearance certificate from BSMMU (Appendix-I). Study patients were collected from four institutions of Bangladesh is the Department of Hematology, Bangabandhu Sheikh Mujib Medical University (BSMMU), Dhaka Medical College Hospital (DMCH), Sir Salimullah Medical College & Mitford Hospital (SSMC) and National Institute of Cancer Research & Hospital (NICRH).

According to sample size calculation (Appendix- II), samples were collected from 50 adult Bengali Bangladeshis AML patients using a convenient sampling technique. The patients were selected according to the inclusion and exclusion criteria followed by a checklist (Appendix- III). Informed written consent was obtained from all the patients to participate and after taking consent, a datasheet on AML was filled out based on the information from the patients. After that 3 ml of peripheral venous blood was obtained very carefully from each patient. The sample was then taken into a

tube containing ethylenediaminetetra acetic acid (EDTA). The blood was stored at -20^{0}C for preservation and further use.

Genomic DNA was isolated from peripheral blood samples using Promega ReliaPrep™ Blood gDNA isolation kit (Promega, USA) in a biological safety cabinet to prevent contamination of the DNA samples. At first, the frozen blood samples were thawed completely. Then the thawed blood samples were mixed thoroughly for 10 minutes in a rotisserie shaker at a speed of 600 rpm at room temperature. A 1.5 ml microcentrifuge tube was taken for each blood sample. After that 20 µl of proteinase K solution from the blood genomic DNA isolation kit was taken in the tube using micropipette (ranging 10-100 µl) with a yellow tip. This proteinase K enzyme was used to break the cellular proteins. Then 200 µl blood was taken in the tube using micropipette (ranging 100-1000 µl) with a barrier tip and was mixed with proteinase K solution briefly. The barrier tip was used to prevent accidental contact of blood with the micropipette that might cause contamination of the instrument. Later, 200 µl of cell lysis buffer was taken from the DNA isolation kit in the tube using micropipette (ranging 100 -1000 µl) with a blue tip. Cell lysis buffer was used to break down the cells. Then the DNA came out from the cells. Then the cap of the microcentrifuge tube was closed and the content was mixed by vortexing in a rotisserie shaker for 10 seconds at 2400 rpm. After mixing the reagents with the blood sample, the tube was incubated for 10 minutes in a heat block at 56° C temperature which is the optimum temperature for proteinase K. After removing the tube from the heat block, 250 µl of binding buffer (BBA) was added using a micropipette (ranging 100-1000 µl) with blue tip and again mixed by vortexing in a rotisserie shaker for 10 seconds at 2400 rpm. Then a binding column was placed into an empty collection tube. Later, the contents of the micro-centrifuge tube were transferred into the binding column using a micropipette (ranging 100-1000 µl) with barrier tip. The binding column cap was closed and placed in a microcentrifuge machine. Then it was centrifuged for 1 minute at 14,000 rpm speed. After checking the binding column to make sure that the lysate has completely passed through the membrane, the column was centrifuged for another one minute. When lysate was visible on top of the membrane then the collection tube containing flow-through was removed. Then the remnant liquid was discarded as bio-hazardous waste. The binding column was placed into a fresh collection tube. Then 500µl of column wash solution (CWD) from the DNA isolation kit was added using a micropipette (ranging 100-1000 µl) with a blue tip to the column. Then it was centrifuged for three minutes at a speed of 14,000 rpm and the flow through was discarded. This step was repeated two more time to have total of three washes consecutively. Then the column was placed in a clean 1.5ml microcentrifuge tube. Later 50 µl of TE (Tris EDTA) buffer were using a micropipette (ranging 10-100 µl) with the yellow tip in the binding column and centrifuged for one minute at a speed of 14,000 rpm. Then the binding column was discarded and micro-centrifuge tube containing genomic DNA was kept for further procedure.

Then the quality and the quantity of the extracted DNA were checked according to the SOP of GRC using NanoDrop 2000/2000c spectrophotometer (Thermo Fisher Scientific, USA) shown in Figure 1.

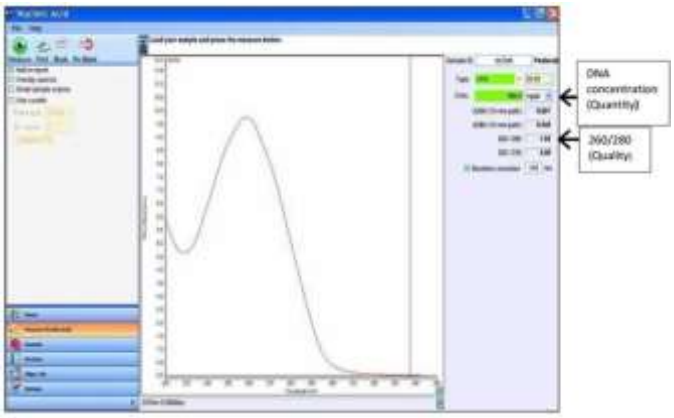

Fig. 1 Sample concentration reading in NanoDrop showing good quality and concentration of the gDNA

Amplification of *GATA2* exons 4 and 5 was performed using the following primers according to Fasan et al. [9]:

GATA2_Ex4-F: 5'-GGTTAAGCAGGCCCCCGTGTC-3'
GATA2_Ex4-R: 5'-ATTAACCGCCAGCTCCTGCC-3'
GATA2_Ex5-F: 5'-CAGCCTGCTGACGCTGCCTT-3'
GATA2_Ex5-R: 5'-GCCTCTTGCCTGGCAGCACA-3'

The desired portion of *GATA2* gene was amplified by short-range PCR and performed on Biometra thermal cycler (Biometra GmbH, Germany) shown in Table I.

Table I PCR Profile for GATA2 Gene

Step	Temperature °C	Time (m:s)	No of cycles
Pre-heating	95	5:00	1 (First)
Denaturation	95	0:30	35
Annealing	61.85 (ex4) 61.60 (ex5)	0:30	35
Initial Extension	72	0:30	30
Final Extension	72	5:00	1
Holding Temperature	4		Till removal of the amplicon

PCR amplified products were checked in ethidium bromide-stained 2% agarose gel. 50bp DNA step ladder (Promega, USA) was used as molecular weight marker for electrophoresis. The desired DNA fragments were visualized under UV light and subsequently photographed by gel documentation system shown in Figure 2.

Purification was done by PromegaWizard® DNA clean-up system of Promega, USA. The amplicons were run using ABI Sanger Sequencer 3730.

Fig. 2 Gel electrophoresis of the PCR product on 2% agarose gel (the desired product size is compared with standard 50 bp DNA ladder).

Data analysis

Data were analyzed by using data analyzing software. Percentage frequency of responses, mean and standard deviations were calculated using SPSS version 22. All the chromatogram results of *GATA2* exon 4 and 5 was first compared with the reference sequence of Ensemble GRCh 37 software, then the transcripts were analyzed. The chromatogram data was then blasted by UCSC BLAT system.

In case of *GATA2* exon 4, UCSC BLAT system shown 99.5%. matching. The matched ENST, transcript variant was 00000341105.2 and the exon number was 00001657978. All the variants in different parts were compared with the database of dbSNP. In the first case in sample no 3309630_16 F, the study found the SNP in cDNA no 1220. There in the alleles G was substituted by A. This was a synonymous variant and the amino acid sequence was V/V- this was a known variant of mutation. In the second case of that sample, the study found the Indel in cDNA no 1221. In these alleles T was substituted by TT. That was a frameshift variant and the amino acid sequence was N/X- this was a known variant of mutation. In the third case of that sample, the study found the Indel in cDNA no 1220. In here alleles were found TGTGGCCCCA. That was a known variant of mutation. In the fourth case, the study found the SNP in cDNA no 1259. There in the alleles G was substituted by A. this was a synonymous variant and the amino acid sequence was D/D- this was a known variant of mutation. In the fifth case, the study found the SNP in cDNA no 1281. There in the alleles T was substituted by G. this was a missense variant and the amino acid sequence was N/H- this was a known variant of mutation. In the sixth case, the study found the SNP in cDNA no 1283. There in the alleles A was substituted by G. this was a synonymous variant and the amino acid sequence was N/N - this was a known variant of mutation. In the seventh case of that sample, the study found the Indel in cDNA no 1285. In here the alleles were TTCATCTTGT. That was a frameshift variant and this was a known variant of mutation. In the eight cases in sample no 3309630_16 F, the study found the SNP in cDNA no 1292. There in the alleles G was substituted by A. This was a synonymous variant and the amino acid sequence was G/G- this was a known variant of mutation. In the ninth case, the study found the SNP in cDNA no 1298. There in the alleles G was substituted by A. This was a synonymous variant and the amino acid sequence was Y/Y- this was a known variant of mutation. In the tenth case of that sample, the study found the Indel in cDNA no 1302. In these alleles T was substituted by TT. That was a frameshift variant and the amino acid sequence was K/X- this was a known variant of mutation. In the eleventh case, the study found the SNP in cDNA no 1304. There in the alleles C was substituted by T. This was a synonymous variant and the amino acid sequence was K/K- this was a known variant of mutation. In the last case, the study found the SNP in cDNA no 1310. There in the alleles A was substituted by G. this was a synonymous variant and the amino acid sequence was N/N - this was a known variant of mutation.

III. RESULTS

In this research, the diagnosis of AML was found more frequent in between the age group of 18-29 years (54%) and the mean age is 25±14.2. The least frequent case was found in between the age group of 50-59 years (2%). Another important observation was found regarding the gender related to the case. The incidence of AML was more common in case of female in comparison to male shown in Figure 3.

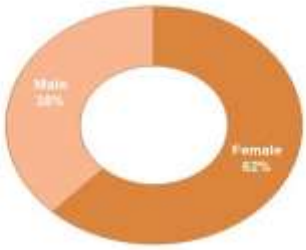

Fig.3 Percentage frequencies of adult Bengali Bangladeshi AML patients according to sex (n = 50)

All the chromatogram results of *GATA2* exon 4 and 5 were analyzed and *GATA2* gene mutations were identified in 1 of the 50 AML patients. In exon 4, synonymous, frameshift and missense type of known variant was found compared with the database of dbSNP. The results were presented in Table II and Figure 4.

Table II Mutation Reports of GATA2 Exon 4 in Adult Bengali Bangladeshi AML Patients (Source: dbSNP)

No	Class	Alleles	cDNA	Amino acid	Variant	Known/ New
01	SNP	G/A	1220	V/V	Synonymous	Known
02	Indel	TT/T	1221	N/X	Frameshift	
03	Indel	TGTGG CCCCA	1220		Frameshift	
04	SNP	G/A	1259	D/D	Synonymous	
05	SNP	T/G	1281	N/H	Missense	
06	SNP	A/G	1283	N/N	Synonymous	
07	Indel	TTCAT CTTGT	1285		Frameshift	
08	SNP	G/A	1292	G/G	Synonymous	
09	SNP	G/A	1298	Y/Y	Synonymous	
10	Indel	TT/T	1302	K/X	Frameshift	
11	SNP	C/T	1304	K/K	Synonymous	
12	SNP	A/G	1310	N/N	Synonymous	

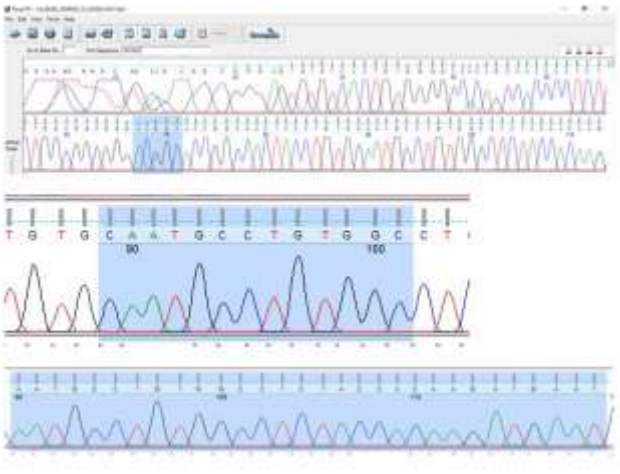

Fig. 4 Sanger sequence of different region of Sample 3309630_16 F exon 4 showing synonymous, frame shift and missense type of known variant

IV. DISCUSSIONS AND RECOMMENDATIONS

The present study reports that AML was more frequent in between the age group of 18-29 years and it's about 54%. Dong et al. [10] mentioned that the age group of AMLs is 15–49 years. According to American Cancer Society [11] AML is generally a disease of older people and is uncommon before the age of 45. The average age of people when they are first diagnosed with AML is about 68. But AML can occur in children as well. On the other hand, regarding sex, the present research shows a female predominance of AML but Juliusson et al. [12] and Aviv De-Morgan [13] reported that AML is significantly more common in males than in females.

The present study found synonymous, frameshift and missense type of known variant in 1 patient in *GATA2* exon 4 out of 50 adult Bengali Bangladeshi AML patients which is similar to the Gao et al. [14] where two heterozygous missense known mutations of *GATA2* were identified performing Sanger sequencing. Kozyra et al. [15] result was similar to the present research as they found synonymous and missense type of known variant in AML patient. Results of Hsu et al. [7] also in accord of the present research because they also found missense and in-frame deletion mutations in *GATA2* gene.

Despite trying hard to avoid bias and analyze the data flawlessly, the small sample size and the convenient sampling technique for choosing AML patients are the main limitations of the present study.

In Conclusion, age of onset and sex of AML patients varies among race and geographical distribution. Identification of known variant with this small sample size emphasizes the importance of exploring the genetic makeup of the Bangladeshi population to develop a database for proper screening and genetic counseling of the disease. Further studies are therefore required to develop a mutation database of adult Bengali Bangladeshi AML patients that could play an important role in the AML screening program of Bangladesh.

ACKNOWLEDGMENT

The authors of this article thank Professor Laila Anjuman Banu, and the Genetic Research Center of the Department of Anatomy, BSMMU for genetic analysis assistance.

This work was supported by the Research Grant for Teachers, BSMMU.

REFERENCES

[1]. Dong Y, Shi O, Zeng Q et al. Leukemia incidence trends at the global, regional, and national level between 1990 and 2017. Experimental hematology & oncology. 2020;9(1):1-11. Available from:https://doi.org/10.1186/s40164-020-00170-6.Accessed May 21, 2022.

[2]. Kirtane K, Lee SJ. Racial and ethnic disparities in hematologic malignancies. Blood. 2017;130(15):1699–1705.

[3]. Lagunas-Rangel FA, Chavez-Valencia V, Gomez-GuijosaMA, Cortes-PenagosC. Acute Myeloid Leukemia-Genetic Alterations and Their Clinical Prognosis. International journal of hematology-oncology and stem cell research. 2017;11(4):328–339.

[4]. WlodarskiMW, CollinM, HorwitzMS. GATA2 deficiency and related myeloid neoplasms. Seminars in hematology. 2017;54(2):81–86.

[5]. Bresnick EH, Jung MM, KatsumuraKR. Human GATA2 mutations and hematologic disease: how many paths to pathogenesis? Blood Advances. 2020;4(18):4584-4592.

[6]. GATA2-AS1, GATA2 antisense RNA 1 [Homo sapiens (human)] 2022 May 13]. Available from: https://www.ncbi.nlm.nih.gov/gene/101927167. Accessed May 21, 2022.

[7]. Hsu AP, Sampaio EP, Khan J, Calvo KR, Lemieux JE, Patel SY et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. Blood. 2011;118: 2653-2655.

[8]. Hahn CN, Chong CE, Carmichael CL, Wilkins EJ, Brautigan PJ, Li XC et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. Nature Genet. 2011;43: 1012-1017.

[9]. Fasan A, Eder C, Haferlach C, Kohlmann A, Dicker F et al.GATA2 mutations are frequent in intermediate-risk karyotype AML with biallelic CEBPA mutations and are associated with favorable prognosis. Leukemia. 2013; 27:482–485.

[10]. Dong Y, Shi O, Zeng Q, Xiaoqin Lu, Wei Wang, Yong Liet al. Leukemia incidence trends at the global, regional, and national level between 1990 and 2017. Exp Hematol Oncol. 2020; 9:14.

[11]. American Cancer Society January 12, 2022 [American Cancer Society's Cancer Statistics Center]. Key Statistics for Acute Myeloid Leukemia (AML) [updated 2022 Jan 12; cited 2019 June 24]. Available from:https://www.cancer.org/cancer/acute-myeloid-leukemia/about/key-statistics.html. Accessed May 22, 2022.

[12]. Juliusson G, Antunovic P, Derolf A, Lehmann S, Mollgard L, Stockelberg D, et al. Age and acute myeloid leukemia: real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry. Blood. 2009; 113:4179–87.

[13]. De-Morgan A, Meggendorfer M, Haferlach C, Shlush L. Male predominance in AML is associated with specific preleukemic mutations. Leukemia. 2021; 35:867–870.

[14]. GaoJ, Gentzler RD, Timms AE, Horwitz MS, Frankfurt O, Altman JK et al. Heritable GATA2 mutations associated with familial AML-MDS: a case report and review of literature. Journal of hematology & oncology. 2014; 7:36.

[15]. Kozyra EJ, Pastor VB, LefkopoulosS, Sahoo SS, Busch H, Voss RK et al. Synonymous GATA2 mutations result in selective loss of mutated RNA and are common in patients with GATA2 deficiency. Leukemia. 2020; 34:2673–2687.

[16]. Nanaa A, Viswanatha D, XieZ et al. Clinical and biological characteristics and prognostic impact of somatic GATA2 mutations in myeloid malignancies: a single institution experience. Blood cancer journal. 2021;11(6):1-4. Available from: https://doi.org/10.1038/s41408-021-00517-0. Accessed July 13, 2022.

Appendix- I

Institutional Review Board (IRB) clearance certificate from BSMMU

Appendix- II

Sample size calculation

Sample size (n)

$n = Z^2pq/d^2$

Here

n = Desired sample size

Z = Confidence limit (1.96 for 95% confidence level)

P = prevalence rate (mutations prevalence of GATA2 = 1-4% [16] = Mean 2.5 = 3)

Q = 1-p

D = Specified absolute precision

$Z^2 = 3.84$, P = 0.03, q = 0.97, $d^2 = 0.0025$

$$n = \frac{3.84 \times 0.03 \times 0.97}{0.0025}$$

= 44.69

= 45

Appendix- III

Selection criteria for the study patients

Inclusion criteria

Selection of patients was mainly based on the following criteria-

- Bengali Bangladeshi
- Both sex
- Age >18yrs
- Diagnosed AML patients

Exclusion criteria

- History of relapse episodes of AML
- History of chemotherapy and radiotherapy