Novel Genetic Variations in Acute Myeloid Leukemia in Pakistani Population

Saba Shahid1*, Muhammad Shakeel1*, Saima Siddiqui2, Shariq Ahmed1, Misha Sohail1, Ishtiaq Ahmad Khan2, Aiysha Abid4 and Tahir Shamsi3

1 Department of Genomics, National Institute of Blood Diseases and Bone Marrow Transplantation, Karachi, Pakistan, 2 Jamil-ur-Rahman Center for Genome Research, Dr. Panywani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan, 3 Department of Hematology, National Institute of Blood Diseases and Bone Marrow Transplantation, Karachi, Karachi, Pakistan, 4 Centre for Human Genetics and Molecular Medicine, Sindh Institute of Urology and Transplantation (SIUT), Karachi, Pakistan

Acute myeloid leukemia (AML) is a hematological malignancy characterized by clonal expansion of blast cells that exhibit great genetic heterogeneity. In this study, we describe the mutational landscape and its clinico-pathological significance in 26 myeloid neoplasm patients from a South Asian population (Pakistan) by using ultra-deep targeted next-generation DNA sequencing of 54 genes (~5000 ×) and its subsequent bioinformatics analysis. The data analysis indicated novel non-silent somatic mutational events previously not reported in AML, including nine non-synonymous and one stop-gain mutations. Notably, two recurrent somatic non-synonymous mutations, i.e., STAG2 (causing p.L526F) and BCORL1 (p.A400V), were observed in three unrelated cases each. The BCOR was found to have three independent non-synonymous somatic mutations in three cases. Further, the SRSF2 with a protein truncating somatic mutation (p.Q88X) was observed for the first time in AML in this study. The prioritization of germline mutations with ClinVar, SIFT, Polyphen2, and Combined Annotation Dependent Depletion (CADD) highlighted 18 predicted deleterious/pathogenic mutations, including two recurrent deleterious mutations, i.e., a novel heterozygous non-synonymous SNV in GATA2 (p.T358P) and BCORL1 (p.A400V), were observed in three unrelated cases each. The BCOR was found to have three independent non-synonymous somatic mutations in three cases. Further, the SRSF2 with a protein truncating somatic mutation (p.Q88X) was observed for the first time in AML in this study. The prioritization of germline mutations with ClinVar, SIFT, Polyphen2, and Combined Annotation Dependent Depletion (CADD) highlighted 18 predicted deleterious/pathogenic mutations, including two recurrent deleterious mutations, i.e., a novel heterozygous non-synonymous SNV in GATA2 (p.T358P) and a frameshift insertion in NPM1 (p.L258fs), found in two unrelated cases each. The WT1 was observed with three independent potential detrimental germline mutations in three different cases. Collectively, non-silent somatic and/or germline mutations were observed in 23 (88.46%) of the cases (0.92 mutation per case). Furthermore, the pharmGKB database exploration showed a missense SNV rs1042522 in TP53, exhibiting decreased response to anti-cancer drugs, in 19 (73%) of the cases. This genomic profiling of AML provides deep insight into the disease pathophysiology. Identification of pharmacogenomics markers will help to adopt personalized approach for the management of AML patients in Pakistan.

Keywords: genomic screening, AML, next generation sequencing, myeloid sequencing panel, novel no-silent somatic mutation
INTRODUCTION

Acute myeloid leukemia (AML) is the most frequent form of acute leukemia in adults with a poor survival rate of about 5 years only (Horton and Huntly, 2012; Cancer Genome Atlas Research Network et al., 2013). It is caused by pathogenic variations in normal progenitor myeloid hematopoietic cells, leading to altered differentiation, proliferation, and self-renewal capability of the cells (Papaemmanuil et al., 2016). In the last decade, there has been significant increase in understanding of underlying mutational landscape of AML (Arber et al., 2016; Papaemmanuil et al., 2016). Consequently, the prognosis, diagnosis, and treatment have been transformed from histological findings to cytogenetic and genomic testing (Grimwade et al., 2010). Analyzing the genetic alterations in AML can be helpful to reduce ambiguities in further characterization of the molecular heterogeneity of normal karyotype AML (Renneville et al., 2008).

Recent studies on the molecular pathogenesis have identified prognostic significance of genetic variation and their contribution in the pathogenesis of AML (Papaemmanuil et al., 2016). The improved AML prognosis associated with mutated NPM1 and biallelic mutations in the CEBPA have resulted in a change in the disease definition (May Green et al., 2010; Hollink et al., 2011). These recent advances have changed the classification and introduced molecular subtypes of the AML with gene mutations (NPM1 and CEBPA) by the recommendation of WHO classification of hematopoietic tumors in 2008 (Campo et al., 2008). In the revised version of 2016, WHO classification introduced additional germline predisposition associated with genetic alterations in the genes CEBPA, DDX41, RUNX1, ANKR26, ETV6, and GATA2 (Cazzola, 2016; Swerdlow et al., 2016). Further studies on genetic landscape of AML have expanded the mutational spectrum where TET2, DNMT3A, NPM1, SRSF2, and ASXL1 genes are mutated frequently in elderly people (Prassek et al., 2018).

Mutational profiling plays an important role in the diagnosis of AML and is now routinely available as a part of the diagnostic workup. It provides diagnostic accuracy, which increases the precision in risk stratification and helps in adopting therapeutic options (Papaemmanuil et al., 2013; Kuo and Dong, 2015). The development of FLT3 and IDH2 inhibitors (Lee et al., 2017; Stein et al., 2017) is achieved only by extensive genomic studies. With the advent of next-generation DNA sequencing (NGS), the cost of genome sequencing has decreased significantly. Amplicon-based targeted sequencing represents an attractive mutation detection method in selected gene panels (Harissendy et al., 2011; Jünemann et al., 2013). This strategy needs less amount of DNA and provides large data of multiple genes in a short turnaround time. Therefore, the genomic tractability of AML makes it a feasible option for targeted NGS testing clinically. The aim of this study was to assess the frequency and clinico-pathological significance of frequently mutated genes by targeting sequencing in AML cases. The targeted sequencing panel comprises of genes involved in various biological functions such as epigenetic regulator genes, the cohesin complex protein encoding genes, genes of activated signaling, tumor repressor genes, and spliceosome genes. This is the first study on molecular characterization of AML patients from South Asia using myeloid sequencing panel, which will be helpful in early diagnosis as well as risk management.

MATERIALS AND METHODS

Ethical and Consent Statement

For this study, 26 AML patients were recruited and sequenced for TruSight myeloid sequencing panel between December 2015 and 2018. These patients included 15 males and 11 females with a median age of 35 years (range: 7–51 years). The clinical presentation of the cases, chromosomal abnormalities, and percentage of circulating blast cells are given in Supplementary Table S1. The study design was approved by the Research Ethics Committee and Review Board of NIBD, and in accordance with the tenets of the Declaration of Helsinki. A written informed consent was obtained from patients and their legal guardians for participation in this study and publication of the findings. Peripheral venous blood specimens of all the recruited patients were collected in EDTA tubes, and stored at 4°C till DNA isolation and subsequent analysis.

DNA Extraction

Genomic DNA was isolated from peripheral blood by using QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer’s protocol. The quality of the extracted DNA was assessed by 2% agarose gel and quantified by Qubit DNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific, United States).

Myeloid Sequencing Panel

TruSight myeloid sequencing panel (Illumina, San Diego, CA, United States) is designed to sequence targeted regions of 54 genes frequently reported for somatic mutations (complete coding exons of 15 genes and exonic hotspots of 39 genes). The genes whose complete coding exons were sequenced include BCOR, BCORL1, CDKN2A, CEBPA, CUX1, DNMT3A, ETV6/TEL, EZH2, KDM6A, IKZF1, PHF6, RAD21, RUNX1/AML1, STAG2, and ZRS2, and exonic hotspots of 39 genes include ABL1, ASXL1, ATRX, BRAF, CALR, CBL, CBLB, CBLC, CSF3R, FBXW7, FLT3, GATA1, GATA2, GNAS, HRAS, IDH1, IDH2, JAK2, JAK3, KIT, KRAS, KMT2A/MLL, MPL, MYD88, NOTCH1, NPM1, Nras, PDE4A, PTK2B, PTEN, PTEN11, SETBP1, SF3B1, SMIC1, SMIC3, SRSF2, TET2, TP53, U2AF1, and WT1. The panel consists of 568 amplicons (length range: 225–275 bp) and covers ~141 kb of genomic region of ~250-bp fragment lengths.

DNA Libraries Preparation

The sequencing libraries were prepared from 50 ng of genomic DNA per sample using TruSight myeloid sequencing panel according to the manufacturer’s protocol. Briefly, the libraries were prepared by annealing uniquely targeted specific oligos at upstream and downstream to the region of interest (ROI), followed by the removal of unbound oligos in subsequent washing steps by using a filter plate. In extension and
ligation step, DNA polymerase was used to connect the hybridized upstream and downstream oligos resulting in the formation of products containing the targeted regions of interest flanked by sequences required for amplification. Next, the amplification step added indexes adapters and prepared for cluster generation. Then, the libraries were cleaned up by using AMPure XP beads to purify PCR products. After the purification, libraries were quantified by Qubit DNA HS Assay kit (Life Technologies, United States).

Libraries were normalized to attain equal library representation that pooled in batches of four samples as per the given guideline. A Pooled Amplicon Library (PAL) was prepared by mixing 5 µl of each of uniquely indexed library. Then, libraries were diluted by taking 6 µl of PAL and 594 µl of ice-cold HT1 incorporation buffer and heat-denatured at 92°C for 2 min. The diluted amplicon libraries were placed on the ice water bath for 5 min, and then 600 µl of the final sample was loaded into the sequencing reagent cartridge kit V2 (MS-102-2002). Workflow for DNA library preparation using Illumina TruSight myeloid sequencing panel is given in Supplementary Figure S1. The DNA sequencing was performed on a MiSeq instrument with standard V2 flow cells with paired end sequencing (150 bp × 2), as per the manufacturer’s instruction.

Sanger Sequencing

Sanger Sequencing was performed to confirm the variants that were identified as pathogenic through standard protocol (BigDye® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems®). The status of known mutations in NPM1 and FLT3 genes were checked by Sanger sequencing and later by allele-specific polymerase chain reaction (PCR) and PCR-restriction fragment length polymorphism (PCR-RFLP) analysis. Electropherogram of identified mutation in AML cases are given in Supplementary Figure S2.

Data Analysis

For data analysis, variants calling was performed using the standard pipeline, as described elsewhere (Lek et al., 2016). The alignment of short DNA sequences with human reference genome hg19 (UCSC) was performed by using Burrows–Wheeler Aligner (BWA-MEM) algorithm (Li and Durbin, 2009). The sequence alignment files (SAM) were converted into binary format (BAM) files using SAMtools (Li et al., 2009); and the removal of duplicates (PCR artifacts) was performed using the PICARD tool1. The base quality score recalibration (BQSR), realignment around small insertions and deletions, and variants calling were carried out by using on-instrument pipeline with Genome Analysis Tool Kit (GATK) best practices (DePristo et al., 2011). The variants with QUAL < 50, GQ < 20, and population variant allele frequency ≥1% in either gnomAD genome or 1000 Genomes Project were filtered out as recommended previously (Tyner et al., 2018). Given that no matching normal tissue samples were sequenced, a bit stringent criterion was applied for somatic variants; i.e., the variants with allelic fraction (VAF) less than half of the percent circulating blast cells in each patient (VAF < 1/2 × %circulating blasts) were considered as somatic. To find possibly pathogenic and/or deleterious somatic associated with AML, a multi-tool prioritization approach was adopted, as recommended by American College of Medical Genetics and Genomics (Richards et al., 2015; Figure 1). The identified variants were annotated with ANNOVAR (Yang and Wang, 2015) and Variants Effect Predictor (McLaren et al., 2016) tools to determine their functional consequences. The deleterious impact of non-synonymous variants was assessed with SIFT, Polyphen2, and Combined Annotation Dependent Depletion (CADD), as described previously (Shakeel et al., 2018).

To prioritize biologically active driver mutations over inactive passengers, the parsimony-guided unsupervised functional impact predictor ParsSNP tool was used. This tool uses an expectation maximization framework to find mutations that explain tumor incidence broadly, without using predefined training labels that can introduce biases (Kumar et al., 2016). The identified variants were also searched in ClinVar database (Landrum et al., 2014) for pathogenic/likely pathogenic association with myeloid malignancies. The interaction between the proteins with deleterious variants in the same samples was determined using STRING database (von Mering et al., 2005). The curation from pharmGKB database (Hewett et al., 2002) was performed to determine variants that likely have a role in leukemic chemotherapy.

1http://picard.sourceforge.net
This panel was designed to identify somatic mutations in myeloid (Pakistan) through Illumina TruSight myeloid sequencing panel. variations in 26 AML cases of a South Asian population.

RESULTS

Non-frameshift deletion 3
Splicing 2
Synonymous 61
Non-synonymous 71

Functional Impact
UTR3 9
UTR5 2
Downstream 3
Exonic 146

The Landscape of Somatic Mutations

Considering the variants with allelic fraction less than half of the %circulating blasts in each case, there were 38 somatic mutations as a whole, including 31 SNVs and 7 insertions/deletions (1.46 mutation/mutation including 18 non-synonymous SNVs, 2 stop-gain SNVs, 1 splicing SNV, and 2 frameshift deletions, and 17 silent mutations including 2 synonymous SNVs, 2 downstream SNVs, 1 3’untranslated region SNV, 7 intronic SNVs, 2 non-frameshift insertions, and 1 non-frameshift deletion (Figure 2 and Supplementary Table S2). Further, it was observed that some cases contained higher number of somatic mutations in different genes. A Kruskal-Wallis test and post hoc Dunn test of multiple comparisons among all the cases showed a significantly higher number of somatic mutations in two cases, AM01 and AM03 (p < 0.01 after multiple corrections).

Strikingly, three recurrent non-silent, and five recurrent silent somatic mutations were observed in more than one case. The non-silent recurrent somatic mutations included a non-synonymous SNV in CDKN2A (p.R90C) in four cases, a non-synonymous SNV in STAG2 (p.L526F) in three cases, and a non-synonymous SNV in BCO RL1 (p.A400V) in three cases. Notably, p.L526F(STAG2) and p.A400V(BCORL1) affected all the transcripts of respective genes, whereas p.R90C(CDKN2A) affected only one out of six transcripts. The VAF of p.R90C (CDKN2A) was double in one case AM21 (VAF 0.214) than in the other three cases who carried similar burden of this variant (VAF 0.084–0.105). The mutational burden of two other recurrent mutations, i.e., p.L526F(STAG2) and p.A400V(BCORL1), was similar among the cases, i.e., 0.102–0.111 and 0.08–0.106, respectively. Further, it was noted that five genes, CSF3R, NOTCH1, CBL, RUNX1, and BCO R, were found to have independent non-silent mutational events in two cases each, with VAF 0.098–0.108, 0.158–0.278, 0.1448–0.203, 0.159–0.25, and 0.135–0.167, respectively. Two genes, KIT and EZH2, had two coexisting mutations each (VAF 0.418–0.42 and 0.043–0.045, respectively) in the same cases (AM03 and AM01, respectively), affecting all the transcripts of their genes.

Curation of somatic mutations in COSMIC database highlighted 10 mutational events not observed in the database, whereas four mutations had been cataloged with a different variation type at the sites than observed in this study (detailed in Supplementary Table S2). The non-cataloged 10 mutations also included the stop-gain SNV (p.Q88X) in SRSF2, affecting all two transcripts of this gene, and the two recurrent non-synonymous SNVs (p.L526F of STAG2 and p.A400V of BCO RL1). Filtration of somatic mutations with ClinVar database highlighted three pathogenic SNVs already associated with hematological disorders. We also assessed conservation status of the non-silent variants sites using PhyloP conservation scores of non-neutral substitution rates based on alignment with 100 vertebrates (Pollard et al., 2010). This analysis revealed 13 mutations in comparatively high conserved regions (PhyloP score >4), 5 mutations in moderately conserved regions (4 ≤ PhyloP score ≥1), and 5 mutations in non-conserved regions (PhyloP < 1) of proteins. Among the recurrent non-silent mutations, p.L526F was observed in the highly conserved region of STAG2, suggesting its more profound deleterious effect, p.A400V occurred in moderately conserved region of BCO RL1, whereas p.R90C occurred in the non-conserved region of CDKN2A. The assessment of somatic mutations for potentially driver role through ParsSNP tool highlighted a

| TABLE 1 | Total variants pertaining to various genomic regions and their functional distribution. |
|----------|-------------------------------------------------|
| Genomic region | No. of variants |
| Exonic | 146 |
| Intronic | 133 |
| Upstream | 0 |
| Downstream | 3 |
| UTR5 | 2 |
| UTR3 | 9 |
| Non-synonymous | 71 |
| Synonymous | 61 |
| Stop-gain | 3 |
| Splicing | 2 |
| Frameshift insertion | 3 |
| Frameshift deletion | 2 |
| Non-frameshift insertion | 2 |
| Non-frameshift deletion | 3 |
driver mutation p.G12S in NRAS (rs121913250), which is well characterized and already reported recurrently in COSMIC database (COSV54736621 and COSM563).

**Co-existence of Somatic Mutations**

It was also noted that seven cases carried more than one non-silent variants; however, the genes harboring co-existing mutations were different among all the cases. The difference in VAF of co-existing mutations gave a clue to define the clonal composition, i.e., a founding clone (the clone with the highest VAF values) and the subclone (Figure 3). In AM01, the novel p.R80H mutation in the conserved region of Runt-related transcription factor 1 (RUNX1), might be the somatic event (VAF 0.25) followed by disrupting splice site (c.1096-2) in Cbl proto-oncogene (CBL) (VAF 0.203) in founding clones leading to the abnormal proliferation of hematopoietic stem cells. Analysis of protein interaction between RUNX1 and CBL through STRING database revealed no interaction between these two proteins, indicating independent mutational events. Previously, a different mutation, p.R80A, at same position of RUNX1, has been shown to strongly reduce its binding with DNA (Bravo et al., 2001). In AM03, the co-existing p.D419fs and p.R420fs deletions in KIT (VAF 0.418 and 0.42) originated more probably in founding clone, prior to the p.V1649I of BCOR (0.135) in subclone. In AM19, the novel p.L509V mutation in CUX1 (VAF 0.366) might be originated in founding clone, followed by p.L526F of STAG2 (0.126) in subclone; in AM26, the p.Q88X in SRSF2 (VAF 0.214) in founding clone and p.A400V in BCORL1 (VAF 0.08) in subclone. The co-existing somatic mutations in three other cases, AM16 [p.R730C of DNMT3A (VAF 0.112) and p.L526F of STAG2 (VAF 0.102)], AM23 [ p.R90C of CDKN2A (VAF 0.214) and p.N434K of RUNX1 (VAF 0.159)], and AM25 [p.T618I of CSF3R (VAF 0.108) and p.A400V of BCORL1 (VAF 0.097)], more probably originated in the same clones. It was noteworthy that all the coexisting mutational events happened in genes belonging to different biological functional categories previously described in myeloid leukemias (Cancer Genome Atlas Research Network et al., 2013), indicating that different underlying processes were involved in the pathophysiology of AML in this cohort.

**Germline Mutation Predisposition**

In addition to the somatic mutations, the predisposition due to germline mutations was also assessed. For this, the germline variants with ClinVar pathogenic/likely pathogenic significance, producing a stop-gain or stop-loss site, disrupting splicing sites, frameshift insertions, or deletions, and non-synonymous alterations predicted as deleterious by SIFT and Polyphen2 tools, were brought into subsequent analysis, as described previously (Bertelsen et al., 2019). This analysis prioritized 18 germline variants pertaining to 15 genes including 13 non-synonymous SNVs, a stop-gain SNV, a splice-site SNV, and 3 frameshift insertions (Supplementary Table S3). Two recurrent mutations, p.T358P in GATA2 affecting all three transcripts, and p.L258fs insertion in NPM1 affecting two out of seven transcripts, were observed in two non-related cases each. Further, WT1 was observed recruiting three independent germline mutations in three different unrelated cases. Filtration with ClinVar database highlighted six non-synonymous pathogenic SNVs.
Shahid et al. NGS Based Screening of AML Patients

FIGURE 3 | Comparison of variants allelic fraction (VAF) of co-existing non-silent somatic mutations in various cases. The mutations in RUNX1 and CBL in AM01, the KIT mutation in AM03, the CUX1 mutation in AM19, and SRSF2 mutation in AM26 originated more likely in founding clones, followed by mutations in other genes in subclones.

and a frameshift pathogenic insertion already associated with hematological neoplasms. The germline variants were filtered with COSMIC database, which revealed six novel variants not cataloged in this database. The assessment of PhyloP scores revealed 11 mutations in comparatively high conserved regions, 2 mutations in moderately conserved regions, and 5 mutations in non-conserved regions of respective proteins. The variants affecting highly conserved regions included the recurrent p.T358P in GATA2, protein truncating p.R441X in WT1, and splicing c.418+1 in PHF6. Exploration of protein truncating p.R441X (rs121907909) in Ensembl genome browser revealed that it affects eight protein coding transcripts introducing a premature stop codon, whereas two protein coding transcripts are protected through NMD pathway (Supplementary Table S4).

To explore the role of identified genetic variants in drug response, pharmGKB database was searched. This analysis showed a missense SNV rs1042522 (G > C) in TP53, with GG and GC genotypes in 19 (73%) cases. These genotypes have been found to show decreased response to cisplatin, paclitaxel, capecitabine, and oxaliplatin anti-cancer drugs as compared to the CC genotype.

DISCUSSION

Next-generation sequencing analysis of myeloid neoplasm including AML and other related disorders has yielded several significant advances in the identification of diagnostic, prognostic, and therapeutic markers for these disorders (Arber et al., 2016; Papaemmanuil et al., 2016). This study was designed to screen AML patients in a clinical diagnostic setup by using a specifically designed myeloid sequencing panel and provides clinico-pathological significance of identified deleterious/non-silent mutations in the Pakistani population. We identified 293 variants including single-nucleotide variants, and small insertions and deletions in coding as well as in non-coding regions in a small cohort of 26 AML patients. Sequence variants not observed in ClinVar, dbSNP, and gnomAD were considered as novel variants. The pathogenicity of sequence variants with a global minor allele frequency (GMAF) of <0.1 was assessed by using several in silico bioinformatics tools and the variants were classified according to the ACMG criteria (Richards et al., 2015). The deleterious impact of non-synonymous variants was assessed with SIFT, Polyphen2, and CADD, as described previously (Shakeel et al., 2018). Although variants were not functionally validated using any in vitro system, in silico analyses have generated strong and convincing scores that suggest the possible pathogenicity of the identified variants in respective cases. To the best of our knowledge, this is the first study to report genetic variations in myeloid malignancies from this South Asian population using NGS technology.

The higher nonsyn/syn ratio in AML cohort represents higher mutation rate and/or positive selection on non-synonymous sites, as indicated in various cancers previously (Greenman et al., 2007; Pengyuan et al., 2012). By applying the multi-tool prioritization approach, we were able to find at least one pathogenic/deleterious non-silent somatic or predisposing germline mutation in 23 of the 26 cases (88.46%), where 9 cases had both the somatic and germline mutations, 8 cases had somatic mutation only, and 6 cases had germline mutation only. In order to explore possible biological relationship between a predisposing germline mutation and somatic mutational events in the cases having coexisting germline and somatic mutations, a circos plot was constructed, which revealed that the two cases with germline non-synonymous mutation p.T358P in GATA2 (AM01 and

2www.ensembl.org/index.html
AM03) had higher number of non-silent somatic mutations (Figure 4). GATA2 encodes an endothelial transcription factor GATA-2 that plays an essential role in gene regulation during vascular development and hematopoietic differentiation. The observed mutation is within the highly conserved region of GATA2 (a zinc finger domain). Although this mutation is novel and not cataloged in dbSNP or COSMIC databases, a p.R361P change near the observed p.T358P mutation in the novel and not cataloged in dbSNP or COSMIC databases, a GATA2 (a zinc finger domain). Although this mutation is observed mutation is within the highly conserved region of vascular development and hematopoietic differentiation. The GATA-2 that plays an essential role in gene regulation during

*Frontiers in Genetics | www.frontiersin.org

June 2020 | Volume 11 | Article 560

Shahid et al. NGS Based Screening of AML Patients

FIGURE 4 | Co-existence of prioritized/pathogenic germline variants with the non-silent somatic mutations in cases with both germline and somatic events. The names of cases are mentioned above the genes with germline mutations.
rich (SR) class of splicing factors involved in both constitutive and alternative mRNA splicing. Previously, dysfunctional SRSF due to sequence variations at p.P95H position have been found to activate aberrant alternative splicing in hematopoietic cells, whereby, having its role in onset of myelodysplastic syndromes (MDS) and AML (Liang et al., 2018; Masaki et al., 2019). In this context, the truncating mutation observed in this study would more likely result in non-functional SRSF, which would lead to malignancy due to hampered alternative splicing in hematopoietic cells. This study reports a first truncating mutation in the SRSF2 RNA-binding domain in an AML case. The other stop-gain somatic mutation in RAD21, causing a protein truncation at p.R478X, affects two protein coding transcripts. This variation has been cataloged in the COSMIC database with four recurrences (COSM1735718) associated with AML. The third and germline stop-gain mutation in WT1 causes p.R441X truncation. Previously, the protein truncating variations in WT1 have been shown to attenuate the TP53-induced DNA damage response in T-cell acute lymphoblastic leukemia (Bordin et al., 2018).

CONCLUSION

In conclusion, this is the first report of a comprehensive analysis of somatic as well as germline mutations in AML from Pakistan using next-generation DNA sequencing technology. Our data strongly support and extend the spectrum of detrimental mutations identified in previous studies employing targeted resequencing approach for the diagnosis of AML. This study also highlights the usefulness of panel sequencing in cases where prognosis becomes challenging. The small cohort size and retrospective nature, i.e., sample collection from a single medical center, are the limiting factors of the study. Furthermore, the novel findings of this preliminary study require validation in a larger cohort with different time scales. Nevertheless, the findings provide an assessment of predisposing detrimental mutations of AML in this region and its utility in clinical settings.

DATA AVAILABILITY STATEMENT

The raw datasets presented in this article have been deposited in BioProject – accession PRJNA627793.

REFERENCES

Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., and Le Beau, M. M. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 20, 2391–2405. doi: 10.1182/ blood-2016-03-643344

Bertelsen, B., Tuxen, I. V., Yde, C. W., Gabrielaite, M., Torp, M. H., Kinalis, S., et al. (2019). High frequency of pathogenic germline variants within homologous recombination repair in patients with advanced cancer. Genom. Med. 4:13. doi: 10.1038/s41525-019-0087-6

Bordin, F., Piovan, E., Masiero, E., Ambesi-Impiombato, A., Minuzzo, S., Bertorelle, R., et al. (2018). WT1 loss attenuates the TP53-induced DNA damage response in T-cell acute lymphoblastic leukemia. Haematologica 103, 266–277. doi: 10.3324/haematol.2017.170431

Bravo, J., Li, Z., Speck, N. A., and Warren, A. J. (2001). The leukemia-associated AML1 (Runx1)–CBF beta complex functions as a DNA-induced molecular clamp. Nat. Struct. Biol. 8, 371–378. doi: 10.1038/86264

Campo, E., Sverdlov, S. H., Harris, N. L., Pileri, S., Stein, H., and Jaffe, E. S. (2008). The 2008 WHO classification of lymphoid neoplasms and beyond: evolving

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee and Review Board of NIBD. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

SSH: study design and execution, and manuscript writing. MSh: data analysis, result interpretation, and manuscript writing. SSI: clinical examination and evaluation of AML patients. SA: library preparation for myeloid sequencing panel. MSo: DNA extraction and quantification. AA: review of manuscript. IK and TS: involving in study design, patient recruitment, review of manuscript, and supervision throughout the study.

ACKNOWLEDGMENTS

We appreciate the patients who participated in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00560/full#supplementary-material

FIGURE S1 | Workflow for DNA library preparation using Illumina TruSight myeloid sequencing panel.

FIGURE S2 | Electropherograms of Sanger sequencing of identified mutations in AML cases.

FIGURE S3 | Density plot to show the distribution/spectrum of variants allelic fraction (VAF) of all variants across the 26 AML patients of this study. The perpendicular at 0.51 allelic fraction represents the median percentage of circulating blasts cells across all cases.

FIGURE S4 | Protein-protein interaction between GATA2 and the proteins with somatic mutations in AM01.

TABLE S1 | Clinical features of 26 AML cases.

TABLE S2 | SOMATIC Mutations.

TABLE S3 | Germline Mutations in AML.

TABLE S4 | Effect of stopgain and frameshift variants (germline and somatic) on transcripts.
Cancer Genome Atlas Research Network, Ley, T. J., Miller, C., Ding, L., Raphael, B. J., Mungall, A. J., et al. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N. Engl. J. Med. 368, 2059–2074. doi: 10.1056/NEJMoa1301689

Cao, Q., Gearhart, M., Gery, S., Shojaei, S., Yang, H., Sun, H., et al. (2016). BCR regulates myeloid cell proliferation and differentiation. Leukemia 30, 1155–1165. doi: 10.1038/leu.2016.2

Cazzola, M. (2016). Introduction to a review series: the 2016 revision of the WHO classification of tumors of hematopoietic and lymphoid tissues. Blood 127, 2361–2364. doi: 10.1182/blood-2016-03-657379

DePristo, M., Banks, E., Poplin, R., Garimella, K., Maguire, J., and Hartl, C. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.

Estey, E. H. (2014). Acute myeloid leukemia: update on risk stratification and management. Am. J. Hematol. 89, 1063–1081. doi: 10.1002/ajh.23834

Fan, Z., Yamaza, T., Lee, J. S., Wang, S., Fan, G., et al. (2009). BCOR regulates myeloid cell proliferation and differentiation. Blood 114, 384–392. doi: 10.1182/blood-2009-11-254441

Grossmann, V., Tiacci, E., Holmes, A. B., Dalglish, G. L., Hunter, C., Bignell, G., et al. (2007). Patterns of somatic mutation in human cancer genomes. Nature 446, 153–158.

Grimeade, D., Hills, R. K., Moorman, A. V., Walker, H., Chatters, S., Goldstone, A. H., et al. (2010). Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood 116, 354–365. doi: 10.1182/blood-2009-11-254441

Harismendy, O., Schwab, R. B., Kools, C., Demmers, J. A., Rodriguez, P., Strouboulis, J., and Vidal, M. (2011). Updating bench top sequencing performance comparison. Nucleic Acids Res. 39, 491–498.

Horton, S. J., and Huntly, B. J. (2012). Recent advances in acute myeloid leukemia. Carcinogenesis 33, 1270–1276. doi: 10.1093/carcin/bgr148

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V. I., Paschka, P., Roberts, N. D., et al. (2016). Genomic classification and prognosis in acute myeloid leukemia. N. Engl. J. Med. 374, 2209–2221. doi: 10.1056/NEJMoa1516192

Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., et al. (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 122, 3616–3627. doi: 10.1182/blood-2013-08-51888

Peng, Y., Jang, W., Dong, H., Xiong, D., Vedell, P., Cui, P., et al. (2012). Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. Carcinogenesis 33, 1270–1276. doi: 10.1093/carcin/bgr148

Sanchez, I., Demmers, J. A., Rodriguera, P., Strouboulis, J., and Vidal, M. (2007). Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbl10/Idm1B histone demethylase and the Bcl6 interacting corepressor. Mol. Cell. Proteomics 6, 820–834. doi: 10.1074/mcp.m600275-mcp200
Shakeel, M., Irfan, M., and Khan, I. A. (2018). Estimating the mutational load for cardiovascular diseases in Pakistani population. *PLoS One* 13:e0192446. doi: 10.1371/journal.pone.0192446

Stein, E. M., DiNardo, C. D., Pollyea, D. A., Fathi, A. T., Roboz, G. J., Altman, J. K., et al. (2017). Enasidenib in mutant-IDH2 relapsed or refractory acute myeloid leukemia. *Blood* 130, 722–731. doi: 10.1182/blood-2017-04-779405

Swerdlow, S. H., Campo, E., Pileri, S. A., Harris, N. L., Stein, H., and Siebert, R. (2016). The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* 127, 2375–2390. doi: 10.1182/blood-2016-01-643569

Thol, F., Bollin, R., Gehlhaar, M., Walter, C., Dugas, M., Suchanek, K. J., et al. (2014). Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. *Blood* 123, 914–920. doi: 10.1182/blood-2013-07-518746

Tijssen, M. R., Cvejic, A., Joshi, A., Hannah, R. L., Ferreira, R., Forrai, A., et al. (2011). Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev. Cell* 20, 597–609. doi: 10.1016/j.devcel.2011.04.008

Timchenko, N., Wilson, D. R., Taylor, L. R., Abdel-Sayed, S., Wilde, M., Sawadogo, M., et al. (1995). Autoregulation of the human C/EBP alpha gene by stimulation of upstream stimulatory factor binding. *Mol. Cell. Biol.* 15, 1192–1202. doi: 10.1128/mcb.15.3.1192

Tyner, J. W., Tognon, C. E., Bottomly, D., Wilmot, B., Kurtz, S. E., Savage, S. L., et al. (2018). Functional genomic landscape of acute myeloid leukaemia. *Nature* 7728, 526–531. doi: 10.1038/s41586-018-0623-z

von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., et al. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 33(Database issue), D433–D437. doi: 10.1093/nar/gki905

Wamstad, J. A., Corcoran, C. M., Keating, A. M., and Bardwell, V. J. (2008). Role of the transcriptional corepressor Bcor in embryonic stem cell differentiation and early embryonic development. *PLoS One* 3:e2814. doi: 10.1371/journal.pone.0002814

Yang, H., and Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and WANNOVAR. *Nat. Protoc.* 10, 1556–1566. doi: 10.1038/nprot.2015.105

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Shakeel, Siddiqui, Ahmed, Sohail, Khan, Abid and Shamsi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.