Whole transcriptome sequencing reveals HOXD11-AGAP3, a novel fusion transcript in the Indian acute leukemia cohort

Sagar Sanjiv Desai[1,2†], Febina Ravindran[1†], Amey Panchal[3], Nishit Ojha[3], Sachin Jadhav[3] and Bibha Choudhary[1]*

[1]Department of Biotechnology and Bioinformatics, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India, [2]Graduate Student Registered Under Manipal Academy of Higher Education, Manipal, Karnataka, India, [3]Cancer Centre, Healthcare Global Enterprises Ltd., Bangalore, India

Introduction: Acute leukemia is a heterogeneous disease with distinct genotypes and complex karyotypes leading to abnormal proliferation of hematopoietic cells. According to GLOBOCAN reports, Asia accounts for 48.6% of leukemia cases, and India reports ~10.2% of all leukemia cases worldwide. Previous studies have shown that the genetic landscape of AML in India is significantly different from that in the western population by WES.

Methods: We have sequenced and analyzed 9 acute myeloid leukemia (AML) transcriptome samples in the present study. We performed fusion detection in all the samples and categorized the patients based on cytogenetic abnormalities, followed by a differential expression analysis and WGCNA analysis. Finally, Immune profiles were obtained using CIBERSORTx.

Results: We found a novel fusion HOXD11-AGAP3 in 3 patients, BCR-ABL1 in 4, and KMT2A-MLLT3 in one patient. Categorizing the patients based on their cytogenetic abnormalities and performing a differential expression analysis, followed by WGCNA analysis, we observed that in the HOXD11-AGAP3 group, correlated co-expression modules were enriched with genes from pathways like Neutrophil degranulation, Innate Immune system, ECM degradation, and GTP hydrolysis. Additionally, we obtained HOXD11-AGAP3-specific overexpression of chemokines CCL28 and DOCK2. Immune profiling using CIBRSORTx revealed differences in the immune profiles across all the samples. We also observed HOXD11-AGAP3-specific elevated expression of lincRNA HOTAIRM1 and its interacting partner HOXA2.

Discussion: The findings highlight population-specific HOXD11-AGAP3, a novel cytogenetic abnormality in AML. The fusion led to alterations in immune system represented by CCL28 and DOCK2 over-expression. Interestingly, in AML, CCL28 is known prognostic marker. Additionally, non-coding signatures (HOTAIRM1) were observed specific to the HOXD11-AGAP3 fusion transcript which are known to be implicated in AML.

1 Introduction

India ranks third in leukemia incidence rates worldwide, with a median survival rate of 35.5% (Bahl et al., 2015; Philip et al., 2015). A higher incidence of blood cancer is observed in men (9.5%) compared to women (5.5%) (Ahirwar et al., 2018). The median age of onset is 40 years, a significant departure from developed countries (Philip et al., 2015; Udupa et al., 2020). Of the two major classes of leukemias, Myeloid leukemias are more common in India and lymphoid in the west (Ahirwar et al., 2018).

Amongst the 7 subtypes of AML, M2 is the most common subtype of AML in the Indian population (Bhattacharyya et al., 2018; Udupa et al., 2020). Cytogenetic abnormalities, such as a change in chromosome number or chromosomal translocations, are considered hallmark genetic abnormalities of hematopoietic neoplasms (Küppers and Dalla-Favera, 2001; Nambiar et al., 2008; Falini et al., 2010). These chromosomal rearrangements lead to fusion transcripts frequently known to drive cancers. Philadelphia chromosome, t (9; 22): BCR-ABL1 was reported in chronic myeloid leukemia (CML), and is used as diagnostics for CML and also some uncertain cases of acute leukemia, and is associated with the worst prognosis in children and adults (Kang et al., 2016; Sampaio et al., 2021). The fusion transcripts resulting from the translocation lead to abnormal cell proliferation.

Detection of fusion transcripts has become robust after the emergence of NGS technologies like transcriptome sequencing, advanced sequencing platforms, and analysis tools (Kumar et al., 2016). The use of fast and accurate tools like STAR-Fusion, which works on RNA-seq data, has been reported to be among the best performers (Haas et al., 2019).

Targeted exome sequence analysis revealed the HOXD11-AGAP3 fusion gene in pediatric nervous system tumors and gastric tumors (Yuan et al., 2022) (Wu et al., 2019). AGAP3, an NMDA receptor signaling complex component, is suggested to be related to GTP binding activity (Oku and Huganir, 2013). GTP binding proteins have been shown to express in leukemic cell lines and could have a role in the cellular differentiation of megakaryoblasts (Nagata et al., 1995). Moreover, Rac, A GTP-binding protein, is known to play a pivotal role in BCR/ABL driven Leukemic conditions (Skorski et al., 1998).

Newman et al. developed a method to fractionate the immune cells using gene expression, CIBERSORT (Cell type Identification By Estimating Relative Subsets Of known RNA Transcripts) (Newman et al., 2015). It has been utilized in AML patients to predict inferior event-free survival (EFS) and overall survival (OS) and identifying novel prognostic markers (Wang et al., 2021). The relative presence of M0 vs. M2 macrophages predicted the relapse. Apart from the available risk stratification markers, the immune cell scoring system (ICSS) can be used as an independent prognostic predictor (Wang et al., 2021).

Co-expression analysis to understand disease pathogenesis has been widely used (Sundarrajan and Arumugam, 2016). Recent studies have utilized weighted gene co-expression network analysis (WGCNA) to identify co-expressed groups of genes from RNA-seq (Clarke et al., 2013), which is based on higher-order relationships and integrated function of gene modules (Zhao et al., 2010; Kadarmideen and Watson-Haigh, 2012).

There are limited NGS-based AML studies from India. Jain et al. (2020) performed NGS analysis on two AML patients and reported that NGS-guided therapy was better than chemotherapy, which was associated with poor survival (Jain et al., 2020). 34 AML patient samples mutation profiling was performed using WGS, and WES highlighting the differences between the genomic landscape of Indian AML and the published literature, and Patkar et al. used 34 gene panel targeted NGS to detect minimal residual disease (MRD) in AML (Dalal et al., 2019; Patkar et al., 2021).

In this study, we have performed transcriptome sequencing and analysis of 9 Indian patient samples from an HCG hospital in Bangalore. Primarily, we performed fusion detection in all the samples and detected a novel fusion HOXD11-AGAP3 in 3 patients, BCR-ABL1 in 4 patients, and KMT2A-MLLT3 in one patient. We performed differential expression of BCR-ABL1 patients and HOXD11-AGAP3 patients independently. The differentially expressed (DE) genes were used as input for WGCNA analysis, which revealed fusion transcript-specific enriched pathways. We also checked the samples' immune profiles and found distinct differences between the patients and immune cells correlated with HOXD11-AGAP3 fusion.

2 Methods

2.1 Subjects for the study

We obtained 9 peripheral blood samples diagnosed with AML at the Healthcare Global Enterprises Ltd., Bengaluru, Karnataka, India. The protocol was approved by the institutional review board of HCG and the Institute of Bioinformatics and Applied Biotechnology. Table 1 lists the clinical details of all the patients. Informed consent was obtained from all the participants.

2.2 Sample collection and culture of PBMCs

The Indian acute leukemia blood samples were collected in anticoagulant EDTA/HEPARIN (3–5 mL each) vacutainers. PBS (equal to that of the sample) was added to the samples, followed by layering the blood on lymphocyte separation media (LSM) (1:2 LSM: blood) along the walls of the tube and centrifuged at 500 RCF for 30 min. The buffy coat was collected and washed 3 times with PBS at 1,000 rpm for 10 min. The washed pellet (cells) was resuspended in 5–10 mL RPMI media (supplemented with 10% FBS) and cultured in T-25 flasks at 37°C with 5% CO_2. After the cells reached confluency, they were passaged. The cells were washed with PBS and stored at −80°C in 500 µL of Trizol Reagent.

2.3 Total RNA extraction

RNA extraction was carried out using Trizol Reagent (Ambion). 0.5×10^6 cells were resuspended in 500 µL of Trizol, 1/10 vol of 2 M sodium acetate pH4) was added to the tube and mixed, followed by 100 µL of chloroform and mixed vigorously and incubated on ice for 20 min. RNA precipitation was done using an equal volume of isopropanol, and the pellet was washed with 75% ethanol. The pellet

TABLE 1 Patient details and alignment statistics of the 10 Acute Leukemia Samples.

ID	Age	Gender	Type	Number of reads (millions)	Alignment percentage (%)
TL_1	NA	F	AML	120.9	86.23
TL_2	38	M	AML	107.7	93.92
TL_3	50	F	AML	107	93.97
TL_5	45	M	AML	223	87.68
TL_6	55	M	AML	93	93.26
TL_7	51	M	AML	111.8	82.27
TL_8	NA	F	AML	100.8	94.46
TL_9	55	M	AML	108.8	91.88
P3	22	M	AML	159	82.82

was dried and resuspended in DEPC-treated MilliQ water. RNA concentration and purity were checked using Qubit, and its integrity was examined by capillary electrophoresis (Tapestation 2200 Agilent) to ensure RNA integrity number >9 for RNA library preparation.

2.4 Illumina Truseq total RNA library preparation and sequencing

For transcriptome sequencing, total RNA libraries were prepared using Illumina TruSeq RNA Library Prep Kit v2. Briefly, the mRNA was isolated from total RNA using oligo-dT beads (NEB; New England Biolabs, Ipswich, Massachusetts, United States). The isolated mRNA was then fragmented to 200-250 bp, followed by double-stranded cDNA synthesis. Adapter ligation was performed after end-repair, and PCR amplification was performed. After constructing the libraries, their concentrations and insert sizes were confirmed using Qubit and Agilent Tapestation, respectively. Library sequencing was outsourced to Medgenome, Bengaluru, Karnataka, India, to obtain 150-bp paired-end reads.

2.5 Transcriptome analysis of acute leukemia samples

Raw reads obtained after sequencing were quality-checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed off sequencing adapters using Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The transcriptome raw data is available on NCBI-SRA (https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA915202). Quality-checked reads were aligned with the standard paired-end alignment options to the human hg38 genome using the STAR aligner (Dobin et al., 2013). The sorted alignment files were indexed using SAMTools (Li et al., 2009), and the read counts were generated using the multiBamCov utility of the BEDTools suite (Quinlan and Hall, 2010). Normalized expression values were calculated from the read counts (TPM-transcripts per million) using in-house python scripts. Fusion genes were identified in
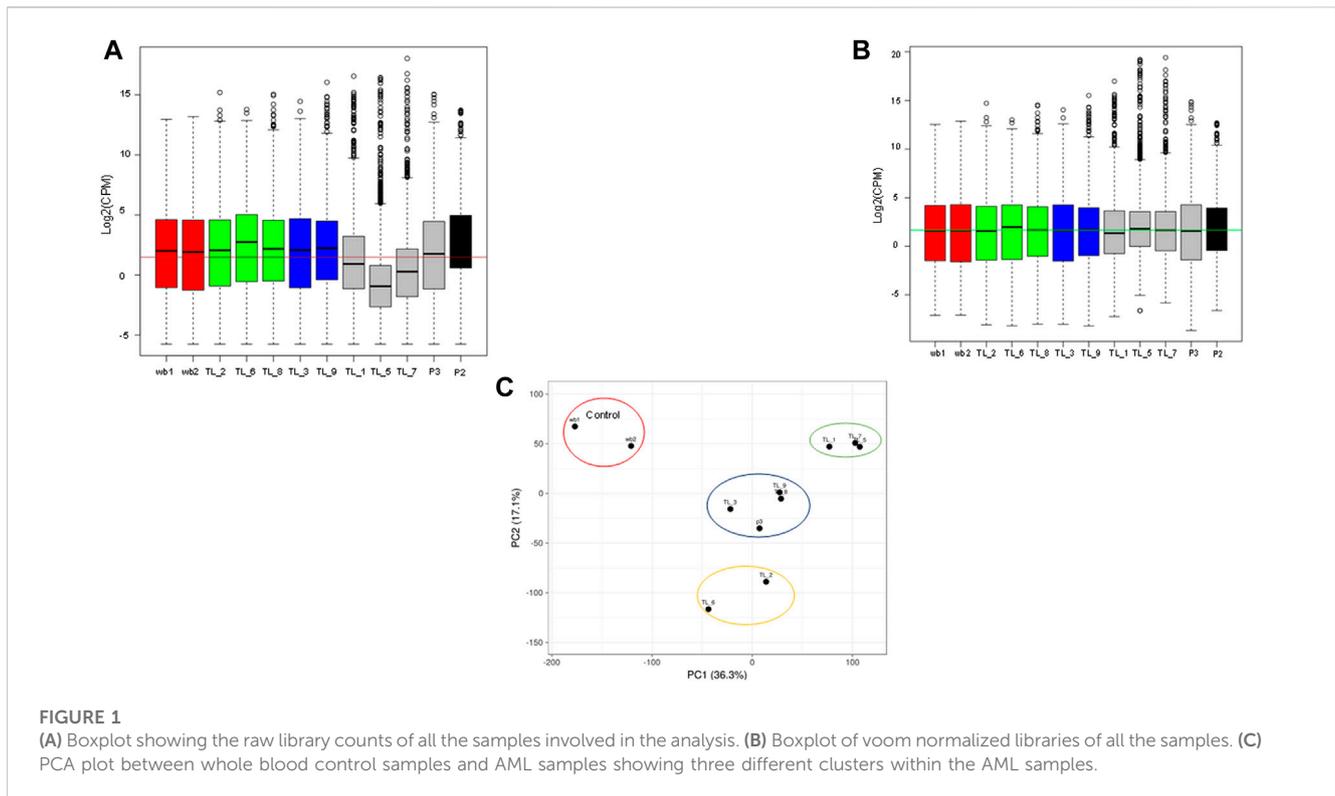
the samples using the STAR-Fusion aligner and the FusionInspector module incorporated to validate the identified fusion transcripts (Haas et al., 2019; Haas et al., 2022). Clusters were obtained from PCA analysis using the online ClustViz tool with normalized TPM values (Metsalu and Vilo, 2015). AGAP3 expression values were extracted for the Indian cohort, and a boxplot was generated using the boxplot package in R. The online database GEPIA2 (Tang et al., 2019) was used to obtain a boxplot of AGAP3 expression values in LAML vs. normal samples from the TCGA dataset, along with the Kaplan-Meier survival plot between high-expression and low-expression AML groups.

2.6 Immune cell fractions across fusion types

To determine the relative immune cell fractions in all the samples using their transcriptome profiles, the samples were compared against the two whole blood control samples. The immune cell fractions were subsequently compared between samples using the online tool CIBERSORTx. CIBERSORTx has a predetermined set of signature genes defined for every immune cell type and their standard expression values known as the Lm22 matrix. Uploading your transcriptome data into the software, it is compared to the Lm22 matrix, and the percentage of immune cell fraction in every sample is determined.

2.7 Differential expression and co-expression networks

The samples were divided into BCR-ABL1 samples and HOXD11-AGAP3 samples. Differential expression was performed for both categories with the two whole blood samples as control using the Bioconductor package DESeq2 in R (Love et al., 2014). Genes were considered upregulated with a log2FC cutoff of +1 and downregulated with −1. The entire transcriptome profile and the differentially expressed genes were subjected to a gene co-

FIGURE 1
(A) Boxplot showing the raw library counts of all the samples involved in the analysis. **(B)** Boxplot of voom normalized libraries of all the samples. **(C)** PCA plot between whole blood control samples and AML samples showing three different clusters within the AML samples.

expression network analysis using the WGCNA package in R. The most significantly correlated modules were subjected to pathway analysis using the REACTOME database.

2.8 Statistical analysis and plots

Samples were normalized using the voom method, and boxplots were generated using the raw and normalized libraries in R. PCA plot was generated online using the ClustVis tool. The fusion transcripts were visualized using the IGV genome browser. The stacked bar chart for immune profiles and the pathway charts were plotted using Microsoft excel. Results were considered to be significant with a *p*-value <0.05. GEPIA2 shows the difference in survival plots to be significant using the log-rank test and the difference in expression to be significant using the Student's t-test. The co-expression modules were correlated with the traits in WGCNA *via* the Pearson correlation coefficient.

3 Results

3.1 Patient details and clinical characteristics

We obtained 9 AML samples from the HCG hospital with a median diagnosis age of 50 years, with 3 of them being females and 6 males. Transcriptome library preparation and subsequent RNA-seq analysis was performed for all 9 transcriptome samples. The depth covered per sample was approximately 100X with an average of 116 million reads of length 151 each. After the
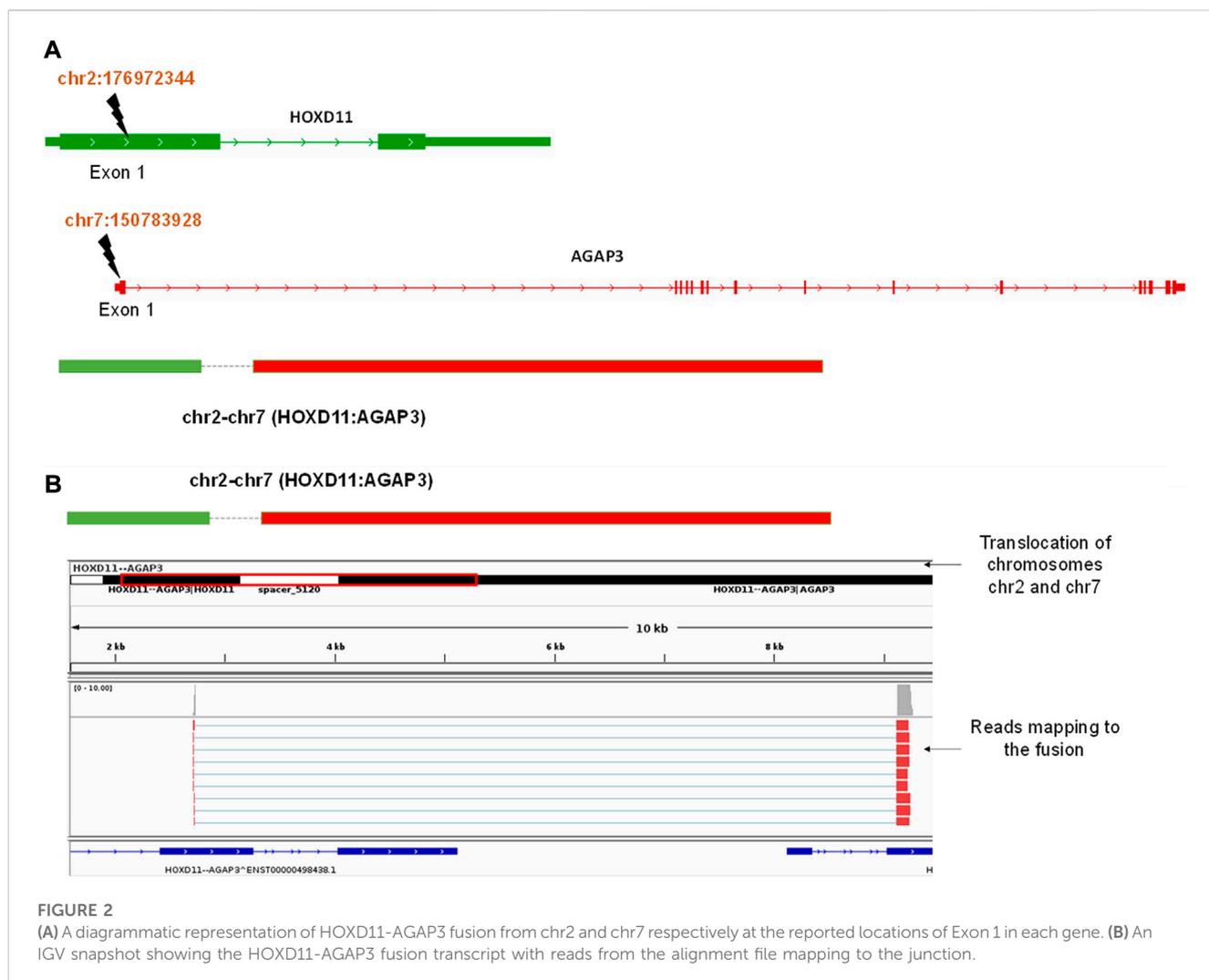
quality check, an average of 87% alignment was obtained for all the samples (Table 1).

3.2 PCA analysis revealed distinct clusters obtained using the normalized transcriptome profiles

To normalize sequencing variations between the samples, quantile normalization of the raw reads was performed using the voom normalization method (Figures 1A, B). Further, we performed PCA analysis, and the plot was generated using the online ClustViz tool. To obtain clusters, we got two normal whole blood samples as controls (SRR17278683 and SRR17278702) from NCBI SRA (Brennan et al., 2022). All the AML samples clustered together and away from the 2 control samples, indicating similarity among AML samples (Figure 1C). Further, on observing only the AML samples from the PCA analysis, three clusters were obtained (TL_1, TL_5, and TL_7), (TL_9, TL_8, TL_3, P3), and (TL_2 and TL_6), indicating differences among the three clustered groups in gene expression (Figure 1C). To account for the cluster-wise differences in the samples, we performed fusion transcript analysis to check whether the samples from different clusters harbored different fusion transcripts.

3.3 Transcriptome analysis reveals novel fusion gene HOXD11-AGAP3

We obtained three different fusions by performing fusion detection using the STAR-Fusion tool. The first was BCR-ABL1
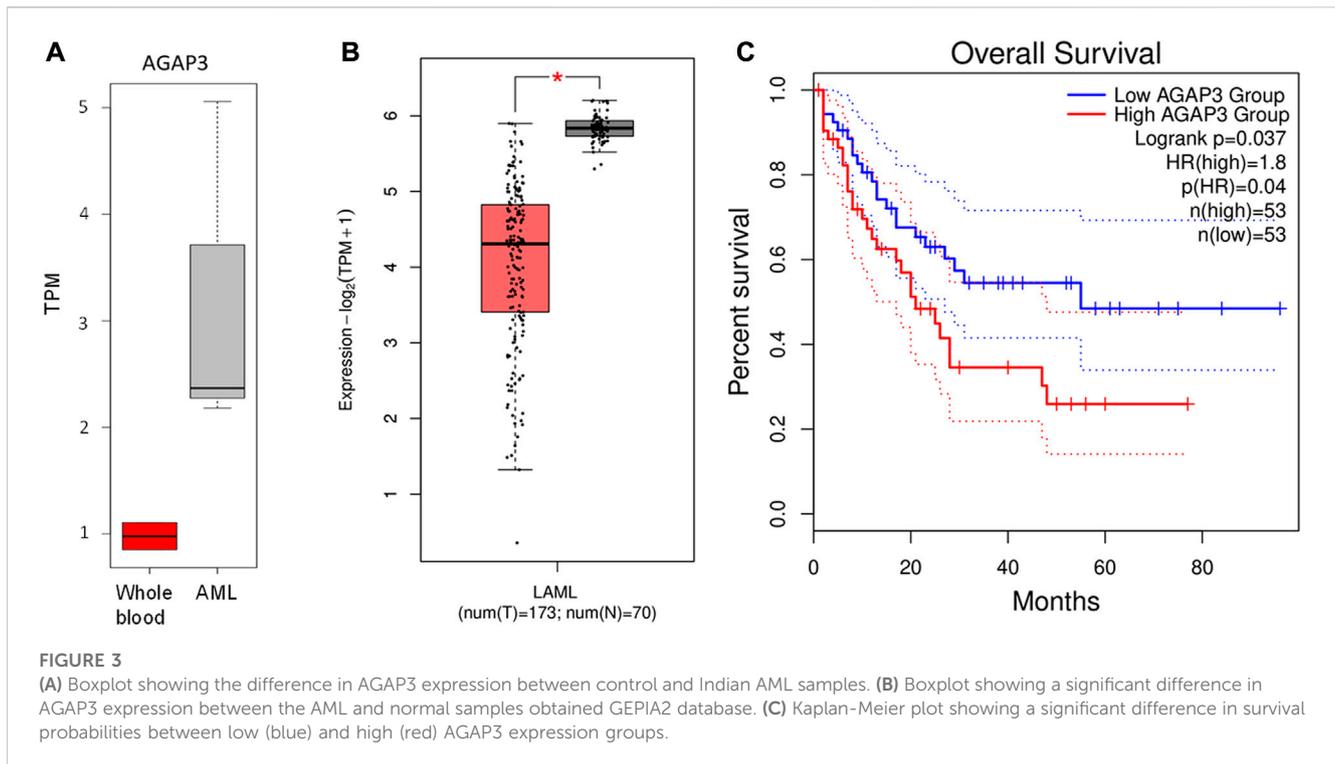
FIGURE 2
(A) A diagrammatic representation of HOXD11-AGAP3 fusion from chr2 and chr7 respectively at the reported locations of Exon 1 in each gene. **(B)** An IGV snapshot showing the HOXD11-AGAP3 fusion transcript with reads from the alignment file mapping to the junction.

fusion in 3 out of 9 samples, TL_2, TL_6, and TL_8 (Supplementary Figure S1, Table 2). TL_2, TL_6, and TL_8 have the same orientation of the BCR-ABL1 fusion with the same breakpoint locations in their respective exons across all three samples. But, TL_8 also has an ABL1-BCR fusion with different exon locations. This could explain the inconsistencies in the transcriptome profiles and the clustering of TL_2 and TL_6 away from TL_8. Secondly, we observed a novel fusion transcript formed between Exon 1 of HOXD11 on chr2 at position 176972344 and Exon1 of AGAP3 on chr7 at position 150783928, resulting in a chr2:chr7/HOXD11-AGAP3 fusion in samples TL-3, TL-6, and TL-9 (Figure 2A; Table 2). Additionally, TL-9 showed a KMT2A-MLLT3 fusion. To further validate the presence of these fusion transcripts, we ran the FusionInspector tool as a part of the STAR-Fusion output. We observed that all three fusions, BCR-ABL1, KMT2A-MLLT3, and HOXD11-AGAP3, had a significant number of reads spanning the respective junctions (Figure 2B, Supplementary Table S1). Interestingly, it was also seen that for all the fusion transcripts detected, the breakpoints in the participating genes occurred in the exonic regions, except HOXD11-AGAP3, which was due to a non-reference splice site involving intron 1 of HOXD11. The fusion sequences with the exact junctions have been depicted in the Supplementary Document.

TABLE 2 Fusion genes detected in Indian Acute Leukemia Samples.

Patient ID	Fusion gene detected
TL-1	--
TL-2	BCR; ABL1
TL-3	HOXD11; AGAP3
TL-5	--
TL-6	BCR; ABL1, HOXD11; AGAP3
TL-7	--
TL-8	BCR; ABL1
TL-9	KMT2A; MLLT3, HOXD11; AGAP3
P3	--

We performed ORF analysis on the fusion transcript harboring roughly 3 Kb long intron to check whether a functional AGAP3 protein or a non-functional truncated protein would result. Using ORF finder, we could observe that the HOXD11 homeobox domain fuses with the AGAP3 arfGAP

FIGURE 3
(A) Boxplot showing the difference in AGAP3 expression between control and Indian AML samples. **(B)** Boxplot showing a significant difference in AGAP3 expression between the AML and normal samples obtained GEPIA2 database. **(C)** Kaplan–Meier plot showing a significant difference in survival probabilities between low (blue) and high (red) AGAP3 expression groups.

GTPase domain (Supplementary Figure S2A) and from the fusion transcript, we found the longest ORF to be in the frame and had GTPase domains intact (Supplementary Figure S2B), although due to the presence of N nucleotide, it could not be confirmed whether the full-length protein was made. In the absence of a full-length transcript with no clarity on the truncated transcript or full length, we decided to check for the expression of AGAP3 in AML samples.

AGAP3 was upregulated in acute leukemia samples compared to the blood control samples (Figure 3A). In comparison, we saw that AGAP3 was downregulated in TCGA AML samples from the GEPIA2 database (Figure 3B), which suggests an Indian population-specific AGAP3 expression pattern. It was also seen that patients with higher AGAP3 expression have a significantly lower survival probability (*p*-value = 0.037) (Figure 3C) in the Caucasian population. We also obtained the interacting partners of AGAP3 from the STRING database, of which TTYH3, which has a significant effect on survival probabilities in AML samples (*p*-value = 0.03), was significantly upregulated in acute leukemia samples (Supplementary Figure S3). We extracted the expression profiles of AGAP3 interacting partners and HOXD11 interacting partners. It was seen that out of the 64 transcription factors binding to AGAP3, ARNT and EP300 were upregulated in TL_3, TL_6, and TL_9 and TTYH3 an interacting partner was the only one upregulated in all three samples with fusion. Amongst the 84 activators of HOXD11, ZNF223 and ARNT were expressed in the samples with HOXD11-AGAP3 fusion proteins.

It was interesting to note that HOXD11:AGAP3 signatures were evident in the expression profiles of lncRNAs associated with genes in their vicinity and the transcripts ZMYM4, ZMYM4_AS1, EP300, LIMS1, LMS1_AS1 SENCR, HOXA2, HOTAIRM1, ZBED3, NNT, NNT_AS1,

PWAK1, MACC1, MACC1_AS1, showed a HOXD11-AGAP3 fusion sample-specific expression profile (Supplementary Figure S4).

3.4 WGCNA reveals hub genes and networks specific to fusion genes

To identify co-expression networks specific to the three fusion gene categories (none, BCR and HOX), we performed a WGCNA analysis with TPM values of all the genes with the presence/absence of a fusion transcript (Table 2) as traits. WGCNA starts by constructing a gene co-expression network followed by identifying gene modules based on hierarchical clustering. It correlates these modules to the clinical data/parameters provided by the user and provides correlations values along with the significance of the correlation. The user can then select the most significantly correlated parameter-module pair and proceed with pathway analysis or network analysis of the genes in the module It was seen that the most significantly correlated module for the BCR-ABL1 patients was "midnight blue" (*p*-value = 0.002, Pearson correlation value, r = 0.85) with 91 genes that accounted for transcriptional regulation pathways (Figure 4A).The most significantly correlated module for HOXD11-AGAP3 samples was "brown" (*p*-value = 0.07, r = 0.6), with 4435 genes. The most enriched pathways were peptide chain elongation and termination, 40S ribosomal subunit formation, non-sense mediated decay, GTP hydrolysis, and downregulation of SMAD2/3 and SMAD4 (Figure 4B). To further investigate the effect of different fusion transcripts, we performed differential expression of BCR-ABL1 samples and HOXD11-AGAP3 samples, with whole blood samples as controls. In the case of BCR-ABL1 samples,
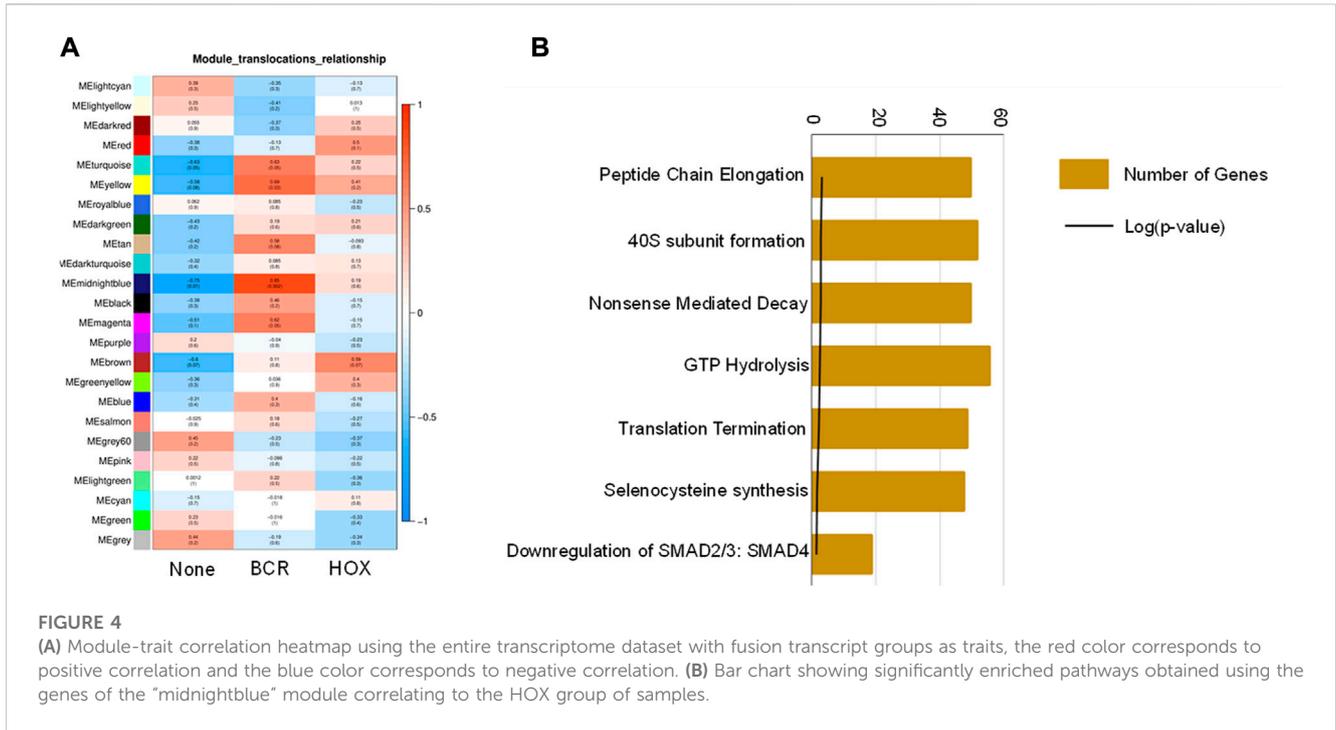
FIGURE 4
(A) Module-trait correlation heatmap using the entire transcriptome dataset with fusion transcript groups as traits, the red color corresponds to positive correlation and the blue color corresponds to negative correlation. **(B)** Bar chart showing significantly enriched pathways obtained using the genes of the "midnightblue" module correlating to the HOX group of samples.
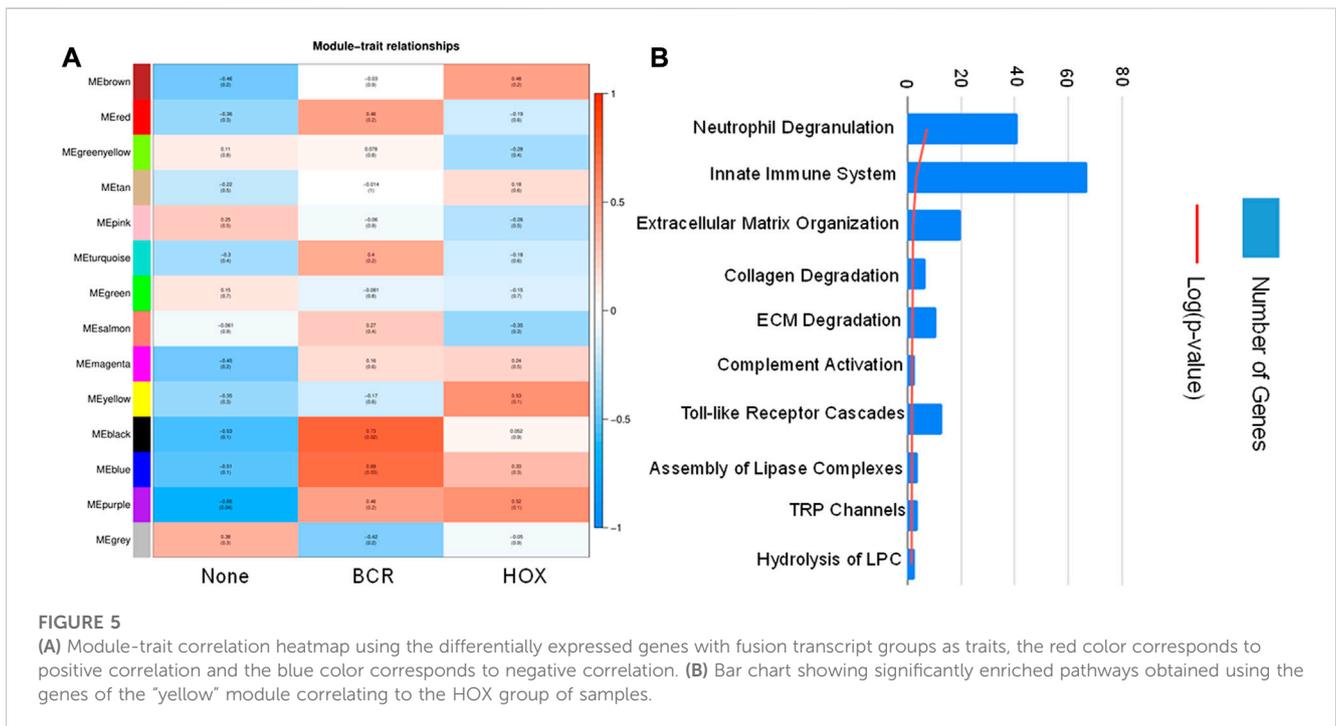

FIGURE 5
(A) Module-trait correlation heatmap using the differentially expressed genes with fusion transcript groups as traits, the red color corresponds to positive correlation and the blue color corresponds to negative correlation. **(B)** Bar chart showing significantly enriched pathways obtained using the genes of the "yellow" module correlating to the HOX group of samples.

1938 genes were differentially expressed (p-value <0.05), with 62.34% being upregulated (log2FC > 2) and 37.66% (log2FC < −2) being downregulated. We obtained a total of 3197 differentially expressed genes (p-value <0.05, log2FC > 2) in the HOXD11-AGAP3 case, of which 44.13% were upregulated, and 55.87% were downregulated (log2FC < −2). The differentially expressed genes (5135: 1938 from BCR-ABL1 samples +3197 from HOXD11-AGAP3 samples) were fed

into another iteration of WGCNA analysis. The module-trait correlation matrix showed that the "black" module is the most significantly correlated with the BCR category (p-value = 0.02, r = 0.73), and the "yellow" module shows the best correlation with HOX category (r = 0.53, p-value = 0.1) (Figure 5A). On performing pathway analysis, we saw that the black (BCR) module was enriched with gene expression regulation and collagen biosynthesis pathways
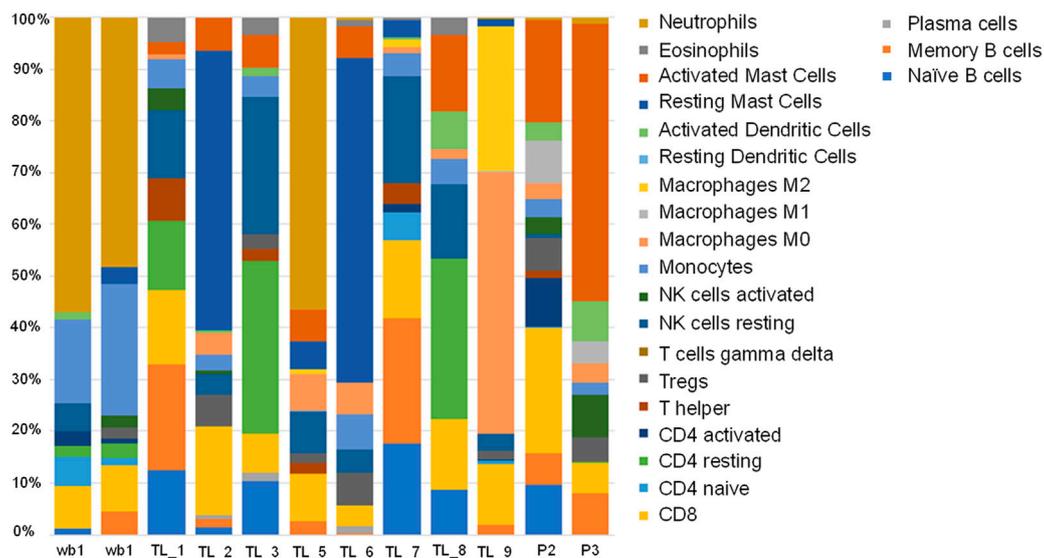
FIGURE 6
Stacked bar plot depicting 22 immune cell fractions across the control and Acute leukemia samples in terms of percentage.

(Supplementary Figure S5), and the yellow (HOX) module showed neutrophil degranulation, innate immune system, ECM organization, complement activation TLR-cascades, lipase complex assembly, to be the most altered (Figure 5B).

To get a better understanding of the immune system-related effect of HOXD11-AGAP3, we screened for additional signatures and amongst the chemokines, we obtained CCL28 and DOCK2 to be highly expressed in HOXD11-AGAP3 samples as compared to the rest. Additionally, we also observed that difference in CCL28 expression has a significant impact on patient survival ($p = 0.016$) (Supplementary Figure S6).

Since the fusion genes correlated with alterations in the immune system, we profiled immune cells to identify fusion-associated changes in the immune cell fractions.

3.5 Immune cell profiling of Indian patient samples

We evaluated the immune cell profiles with CIBERSORTx. It uses a set of genetic markers for 22 immune cells as reference, against which user provided expression values are compared to determine the percentage of immune cells present in the patient samples being analysed. All the acute leukemia samples had a very different immune profile compared to the whole blood controls. We checked for the immune profile in 3 different groups depending on the fusion transcripts. Among the groups with no fusion transcript (TL_1, TL_5, TL_7, and P3), TL_1 and TL_7 showed a very similar distribution of immune fractions with high mast cell and negligible neutrophil fractions (Figure 6). TL_5 was the only one with a significantly high fraction of neutrophils (45%) compared to the negligible levels in others (Figure 6 13). In the group with BCR-ABL1 fusion (TL_2, TL_6, and TL_8), TL_2 and TL_6 were similar with a higher fraction of Mast cells, Tregs, Macrophage M0, and CD8 cells, while TL_8 had a higher fraction of activated dendritic cells. In the subgroup HOXD11:

AGAP3 fusion, all three TL_3, TL_6, and TL_9 was dissimilar in immune fractions, which could be due to the presence/absence of additional fusion in those samples. TL_3, the only sample with just the HOXD11-AGAP3 fusion, had a high frequency of resting CD4 cells and NK cells, followed by CD4 naive and CD8 cells. TL_6's (that also showed BCR/ABL1 fusion) immune profile was closer to the TL_2 samples than as compared to TL_3 and TL_9. TL_9 showed the highest M0 and M2 macrophage fractions and a relatively high fraction of CD8 cells (Figure 6), indicating patient-specific immune cell fractions. We also observed that there was an exclusive representation of plasma cells in TL_3 and TL_6 samples. TL_3 was also enriched in NK cells, while TL_3 and TL_9 was enriched in CD4 T cells (Figure 6).

4 Discussion

One of the main objectives of this study was to identify and characterize karyotypes in the Indian acute leukemia population. Several studies have used transcriptome sequencing data to detect fusion transcripts, but no study has been conducted in India (Lilljebjörn et al., 2013; Stengel et al., 2018; Mata-Rocha et al., 2019). In India, there have been case study reports regarding treatment outcomes of chemotherapy, the genome landscape of two patients, and a study to detect MRD in AML using NGS, but so far, transcriptomics has not been used for any clinical applications (Dalal et al., 2019; Jain et al., 2020; Patkar et al., 2021). Our study reveals patient-specific complex karyotypes, including BCR-ABL1, KMT2A-MLLT3, and a novel fusion transcript HOXD11-AGAP3 in three out of the 10 patients. A differential expression followed by WGCNA analysis resulted in different gene-coexpression networks getting enriched according to the presence of the fusion transcripts.

For the longest time, BCR-ABL1 fusion was believed to be a hallmark of CML patient samples, but recent studies have described

a rare category of AML with BCR-ABL translocation (Kang et al., 2016; Neuendorff et al., 2016; Sampaio et al., 2021). Interestingly, we detected BCR-ABL fusion in TL_2, TL_6, and TL_8, which suggests that these samples could have a CML origin with primary blast crisis, which is the case with TL_2, or they could belong to the recently discovered rare category of AML, which are BCR-ABL positive (Neuendorff et al., 2016). In three samples, TL-3, TL-6, and TL-9, we detected a novel fusion gene pair, HOXD11-AGAP3. This fusion gene has been reported in neuroblastoma and gastric tumor cases but not in AML cases (Wu et al., 2019) (Yuan et al., 2022). Interestingly, TL_6 also had the BCR-ABL1 fusion transcript and the KMT2A-MLLT3 fusion gene in TL-9. Suggests that TL_6 and TL_9 are complex karyotypes.

Approximately 10%–12% of all AML samples are known to have complex karyotypes, yet they are not very well characterized at the molecular level (Mrózek et al., 2019). AGAP3 is an NMDA receptor signaling complex; its fusion gene is suspected to be involved in GTP-binding and GTPase signaling (Oku and Huganir, 2013; Wu et al., 2019). WGCNA analysis revealed a significant correlation between the HOXD11-AGAP3 fusion and GTP hydrolysis. It has been shown that BCR-ABL activates proteins of the GTP-binding network in the context of leukemogenesis (Skorski et al., 1998). Therefore, the fact that TL_6 has both BCR-ABL1 and HOXD11-AGAP3 could be of great genetic relevance for AML.

Long non-coding RNA in AML has been reviewed thoroughly (Ng et al., 2019). Long non-coding RNA regulates gene expression by binding to the DNA, RNA or protein and directing the localization of the protein. They have been used as diagnostic and prognostic indicators of disease progression, especially cancer. LincRNA is known to regulate cell proliferation and apoptosis of cancer cells (presented in the literature review). One of the lincRNA HOTAIRM1 (present in HOXA gene cluster) overexpression has a well-known impact on prognosis in a myeloid cell differentiation manner in AML (Zhang et al., 2014) (Gao et al., 2020). Interestingly, HOTAIRM1 and HOXA2 expression levels are elevated in HOXD11-AGAP3 samples.

CIBERSORTx analysis showed a distinct pattern of immune cells across the individuals and did not follow fusion gene patterns indicating the fusion products might not influence them. However, mild similarities between the patients were observed, indicating the heterogeneity of AML. TL_5 has a significantly high (45%) neutrophil content which was not observed in any of the samples of the Caucasian population. It has been observed that the neutrophil-to-lymphocyte ratio in the pretreatment group has a better prognosis than NLR high (Zhang et al., 2021). It is possible that the high neutrophil count in the TL_5 sample could be due to an infection and may indicate a bad prognosis. A recent study reported AGAP3 in Alzheimer's disease correlates with resting and naive CD4 cells, NK cells, CD cells, etc. (Zhao et al., 2022). TL_3, the only sample with just HOXD11-AGAP3, showed high frequencies of the same. WGCNA analysis of the full transcriptome profile enriched the innate immune system and neutrophil degranulation pathways. HOXD11-AGAP3 signatures also included elevated CCL28 and DOCK2 expression. Altered CCL28 expression, which shows a significant effect on survival in AML samples (GEPIA), has been associated with chemokine responsiveness in AML patients. DOCK2, which is expressed only in hematopoietic tissue, is known to be of prognostic value in AML

patients (Brenner et al., 2016) (Hu et al., 2019). These facts suggest an immune cell-related role of HOXD11-AGAP3 in AML.

The small sample size of patients in the study is a definite shortcoming, but our results reveal a novel fusion in 3 out of 10 samples, with two being complex Karyotypes. This and other results reveal Indian population-specific signatures for Acute Leukemia that need to be explored and validated further with a larger dataset.

Data availability statement

The data presented in the study are deposited in the NCBI-SRA repository, accession number PRJNA915202 (https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA915202).

Ethics statement

The study has been approved by the Institutional Ethics Committee of Institute of Bioinformatics and Applied Biotechnology (IBAB) and HealthCare Global Enterprises Ltd. (HCG); Study No: HIEC/48/21/08.

Author contributions

BC and SD conceived and coordinated the study; AP, NO, and SJ, provided the samples along with the clinical data; BC and SD collected the samples; FR performed the RNA isolation and transcriptome library preparation from the samples; BC and SD interpreted the data and wrote the manuscript. All the authors read, reviewed and approved the final manuscript.

Conflict of interest

Authors AP, NO, and SJ are employed by HealthCare Global Enterprises Ltd.

The remaining 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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1100587/full#supplementary-material

SUPPLEMENTARY FIGURE S1
An IGV snapshot showing the BCR-ABL1 fusion transcript with reads from the alignment file mapping to the junction.

SUPPLEMENTARY FIGURE S2
(A) ORF sequence predicted by the orf maker using the fusion transcript, along with the amino acid sequence of the orf. The amino acid sequence shows similarity with arf-GAP GTPase domain on performing a protein blast. **(B)** The orfs from the partner genes in the fusion transcript HOXD11 and AGAP3 along with their protein domains.

SUPPLEMENTARY FIGURE S3
Boxplot showing the difference in TTYH expression between control and AML samples and a Kaplan Meier plot showing significant difference in survival probabilities between low and high TTYH3 expression.

SUPPLEMENTARY FIGURE S4
A heatmap of lncRNA-mRNA pairs with upregulated expression profiles specific to the HOXD11-AGAP3 family of patients.

SUPPLEMENTARY FIGURE S5
Bar chart showing significantly enriched pathways obtained using the genes of the "black" module correlating to the BCR group of samples.

SUPPLEMENTARY FIGURE S6
CCL28 and DOCK2 expression patterns across all the AML patients with upregulation specific to the HOXD11-AGAP3 category of patients along with a Kapalan-Meier curve of CCL28 in AML from GEPIA between low and high CCL28 expression groups.

SUPPLEMENTARY TABLE S1
Number of reads mapping to the Fusion Junctions in respective samples.

References

Ahirwar, D. R., Kumar Nigam, D. R., and Parmar, D. D. (2018). A study of leukemias Profile in central India. *Trop. J. Pathology Microbiol.* 4, 181–187. doi:10.17511/jopm. 2018.i02.12

Bahl, A., Sharma, A., Raina, V., Kumar, L., Bakhshi, S., Gupta, R., et al. (2015). Long-term outcomes for patients with acute myeloid leukemia: A single-center experience from aiims, India. *Asia-Pacific J. Clin. Oncol.* 11, 242–252. doi:10.1111/ajco.12333

Bhattacharyya, J., Nath, S., Saikia, K. K., Saxena, R., Sazawal, S., Barman, M. P., et al. (2018). Prevalence and clinical significance of FLT3 and NPM1 mutations in acute myeloid leukaemia patients of Assam, India. *Indian J. Hematol. blood Transfus. official J. Indian Soc. Hematol. Blood Transfus.* 34 (1), 32–42. doi:10.1007/s12288-017-0821-0

Brennan, K., Zheng, H., Fahrner, J. A., Shin, J. H., Gentles, A. J., Schaefer, B., et al. (2022). NSD1 mutations deregulate transcription and DNA methylation of bivalent developmental genes in Sotos syndrome. *Hum. Mol. Genet.* 31 (13), 2164–2184. doi:10. 1093/hmg/ddac026

Brenner, A. K., Reikvam, H., and Bruserud, Ø. (2016). A subset of patients with acute myeloid leukemia has leukemia cells characterized by chemokine responsiveness and altered expression of transcriptional as well as angiogenic regulators. *Front. Immunol.* 7, 205. doi:10.3389/fimmu.2016.00205

Clarke, C., Madden, S. F., Doolan, P., Aherne, S. T., Joyce, H., O'Driscoll, L., et al. (2013). Correlating transcriptional networks to breast cancer survival: A large-scale coexpression analysis. *Carcinogenesis* 34 (10), 2300–2308. doi:10.1093/carcin/bgt208

Dalal, H. Y., Damodar, S., Veldore, V. H., K, C. M., Prabhu, S., Chawla, M., et al. (2019). A prospective analysis of the genomic landscape of patients with acute myeloid leukemia and its impact on clinical outcomes - data from a tertiary care center in India. *Blood* 134, 5163. doi:10.1182/blood-2019-130000

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). Star: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15–21. doi:10.1093/bioinformatics/bts635

Falini, B., Tiacci, E., Martelli, M. P., Ascani, S., and Pileri, S. A. (2010). New classification of acute myeloid leukemia and precursor-related neoplasms: Changes and unsolved issues. *Discov. Med.* 10 (53), 281–292.

Gao, J., Wang, F., Wu, P., Chen, Y., and Jia, Y. (2020). Aberrant LncRNA expression in leukemia. *J. Cancer* 11, 4284–4296. doi:10.7150/jca.42093

Haas, B. J., Dobin, A., Li, B., Stransky, N., Pochet, N., and Regev, A. (2019). Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol.* 20 (1), 213. doi:10.1186/s13059-019-1842-9

Haas, B. J., Dobin, A., Ghandi, M., Van Arsdale, A., Tickle, T., Robinson, J. T., et al. (2022). *Targeted* in silico *characterization of fusion transcripts in tumor and normal tissues via FusionInspector*. New York: Cold Spring Harbor Laboratory. doi:10.1101/2021.08.02.454639

Hu, N., Pang, Y., Zhao, H., Si, C., Ding, H., Chen, L., et al. (2019). High expression of DOCK2 indicates good prognosis in acute myeloid leukemia. *J. Cancer* 10 (24), 6088–6094. doi:10.7150/jca.33244

Jain, G., Thakral, D., Sahoo, R. K., Kumar, I., Vashishtha, S., Verma, P., et al. (2020). Next generation sequencing guided treatment modulation and prognosis in Acute myeloid leukemia: Case vignettes. *Am. J. blood Res.* 10 (4), 134–144.

Kadarmideen, H. N., and Watson-Haigh, N. S. (2012). Building gene co-expression networks using transcriptomics data for systems biology investigations: Comparison of methods using microarray data. *Bioinformation* 8 (18), 855–861. doi:10.6026/97320630008855

Kang, Z.-J., Liu, Y. F., Xu, L. Z., Long, Z. J., Huang, D., Yang, Y., et al. (2016). The Philadelphia chromosome in leukemogenesis. *Chin. J. Cancer* 35, 48. doi:10.1186/s40880-016-0108-0

Kumar, S., Razzaq, S. K., Duy Vo, A., Gautam, M., and Li, H. (2016). Identifying fusion transcripts using next generation sequencing. *Wiley Interdiscip. Rev. RNA* 7 (6), 811–823. doi:10.1002/wrna.1382

Küppers, R., and Dalla-Favera, R. (2001). Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* 20, 5580–5594. doi:10.1038/sj.onc.1204640

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Lilljebjörn, H., Agerstam, H., Orsmark-Pietras, C., Rissler, M., Ehrencrona, H., NiLsson, L., et al. (2013). RNA-seq identifies clinically relevant fusion genes in leukemia including a novel MEF2D/CSF1R fusion responsive to imatinib. *Leukemia* 28 (4), 977–979. doi:10.1038/leu.2013.324

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

Mata-Rocha, M., Rangel-Lopez, A., Jimenez-Hernandez, E., Morales-Castillo, B. A., Gonzalez-Torres, C., Gaytan-Cervantes, J., et al. (2019). Identification and characterization of novel fusion genes with potential clinical applications in Mexican children with acute lymphoblastic leukemia. *Int. J. Mol. Sci.* 20, 2394. doi:10.3390/ijms20102394

Metsalu, T., and Vilo, J. (2015). ClustVis: A web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic acids Res.* 43 (W1), W566–W570. doi:10.1093/nar/gkv468

Mrózek, K., Eisfeld, A. K., Kohlschmidt, J., Carroll, A. J., Walker, C. J., Nicolet, D., et al. (2019). Complex karyotype in de novo acute myeloid leukemia: Typical and atypical subtypes differ molecularly and clinically. *Leukemia* 33 (7), 1620–1634. doi:10.1038/s41375-019-0390-3

Nagata, K., Okano, Y., and Nozawa, Y. (1995). Identification of heterotrimeric GTP-binding proteins in human megakaryoblastic leukemia cell line, MEG-01, and their

alteration during cellular differentiation. *Life Sci.* 57 (18), 1675–1681. doi:10.1016/0024-3205(95)02147-b

Nambiar, M., Kari, V., and Raghavan, S. C. (2008). Chromosomal translocations in cancer. *Biochimica Biophysica Acta - Rev. Cancer* 1786, 139–152. doi:10.1016/j.bbcan.2008.07.005

Neuendorff, N. R., Burmeister, T., Dorken, B., and Westermann, J. (2016). BCR-ABL-positive acute myeloid leukemia: A new entity? Analysis of clinical and molecular features. *Ann. Hematol.* 95 (8), 1211–1221. doi:10.1007/s00277-016-2721-z

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Ng, M., Heckl, D., and Klusmann, J.-H. (2019). The regulatory roles of long noncoding RNAs in acute myeloid leukemia. *Front. Oncol.* 9, 570. doi:10.3389/fonc.2019.00570

Oku, Y., and Huganir, R. L. (2013). AGAP3 and Arf6 regulate trafficking of AMPA receptors and synaptic plasticity. *J. Neurosci. official J. Soc. Neurosci.* 33 (31), 12586–12598. doi:10.1523/JNEUROSCI.0341-13.2013

Patkar, N., Kakirde, C., Shaikh, A. F., Salve, R., Bhanshe, P., Chatterjee, G., et al. (2021). Clinical impact of panel-based error-corrected next generation sequencing versus flow cytometry to detect measurable residual disease (MRD) in acute myeloid leukemia (AML). *Leukemia* 35 (5), 1392–1404. doi:10.1038/s41375-021-01131-6

Philip, C., George, B., Ganapule, A., Korula, A., Jain, P., Alex, A. A., et al. (2015). Acute myeloid leukaemia: Challenges and real world data from India. *Br. J. Haematol.* 170, 110–117. doi:10.1111/bjh.13406

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi:10.1093/bioinformatics/btq033

Sampaio, M. M., Santos, M. L. C., Marques, H. S., Goncalves, V. L. d. S., Araujo, G. R. L., Lopes, L. W., et al. (2021). Chronic myeloid leukemia-from the philadelphia chromosome to specific target drugs: A literature review. *World J. Clin. Oncol.* 12, 69–94. doi:10.5306/wjco.v12.i2.69

Skorski, T., Wlodarski, P., Daheron, L., Salomoni, P., Nieborowska-Skorska, M., Majewski, M., et al. (1998). BCR/ABL-mediated leukemogenesis requires the activity of the small GTP-binding protein Rac. *Proc. Natl. Acad. Sci. U. S. A.* 95 (20), 11858–11862. doi:10.1073/pnas.95.20.11858

Stengel, A., Nadarajah, N., Haferlach, T., Dicker, F., Kern, W., Meggendorfer, M., et al. (2018). Detection of recurrent and of novel fusion transcripts in myeloid malignancies by targeted RNA sequencing. *Leukemia* 32 (5), 1229–1238. doi:10.1038/s41375-017-0002-z

Sundarrajan, S., and Arumugam, M. (2016). Weighted gene co-expression based biomarker discovery for psoriasis detection. *Gene* 593 (1), 225–234. doi:10.1016/j.gene.2016.08.021

Tang, Z., Kang, B., Li, C., Chen, T., and Zhang, Z. (2019). GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic acids Res.* 47, W556–W560. doi:10.1093/nar/gkz430

Udupa, M. S. N., Babu, K. G., Suresh Babu, M. C., Lakshmaiah, K. C., Lokanatha, D., Jacob, A. L., et al. (2020). Clinical profile, cytogenetics and treatment outcomes of adult acute myeloid leukemia. *J. Cancer Res. Ther.* 16, 18–22. doi:10.4103/jcrt.jcrt_1162_16

Wang, Y.-H., Hou, H. A., Lin, C. C., Kuo, Y. Y., Yao, C. Y., Hsu, C. L., et al. (2021). A CIBERSORTx-based immune cell scoring system could independently predict the prognosis of patients with myelodysplastic syndromes. *Blood Adv.* 5 (22), 4535–4548. doi:10.1182/bloodadvances.2021005141

Wu, W., Xu, W. J., Liu, J. B., Sun, J., Huang, Y. M., and Lv, Z. B. (2019). Exome sequencing identifies predisposing and fusion gene in ganglioneuroma, ganglioneuroblastoma and neuroblastoma. *Math. Biosci. Eng. MBE* 16 (6), 7217–7229. doi:10.3934/mbe.2019362

Yuan, L., Chen, S., Wang, Y., and Ma, Y. (2022). The molecular characteristics of gastric cancer patients living in Qinghai-Tibetan Plateau. *BMC Gastroenterol.* 22 (1), 244–248. doi:10.1186/s12876-022-02324-8

Zhang, X., Weissman, S. M., and Newburger, P. E. (2014). Long intergenic non-coding RNA HOTAIRM1 regulates cell cycle progression during myeloid maturation in NB4 human promyelocytic leukemia cells. *RNA Biol.* 11 (6), 777–787. doi:10.4161/rna.28828

Zhang, Q., Yang, Q., Weng, Y., Huang, Z., Chen, R., Zhu, Y., et al. (2021). Neutrophil-to-lymphocyte ratio correlates with prognosis and response to chemotherapy in patients with non-M3 de novo acute myeloid leukemia. *Transl. Cancer Res.* 10, 1013–1024. doi:10.21037/tcr-20-2179

Zhao, W., Langfelder, P., Fuller, T., Dong, J., Li, A., and Hovarth, S. (2010). Weighted gene coexpression network analysis: State of the art. *J. Biopharm. statistics* 20 (2), 281–300. doi:10.1080/10543400903572753

Zhao, K., Zhang, H., Wu, Y., Liu, J., Li, X., and Lin, J. (2022). Integrated analysis and identification of hub genes as novel biomarkers for Alzheimer's disease. *Front. aging Neurosci.* 14, 901972. doi:10.3389/fnagi.2022.901972