Role of Paired Box 5 (PAX 5) Genetic Variant in the Development and Progression of Leukemia.

Amrita Bhat  
Shri Mata Vaishno Devi University

Monika Pandita  
Shri Mata Vaishno Devi University

Supriya Hakeem  
Shri Mata Vaishno Devi University

Gh. Rasool Bhat  
Shri Mata Vaishno Devi University

Sonali Verma  
Shri Mata Vaishno Devi University

Gresh Chander  
Shri Mata Vaishno Devi University

Shabab Angurana  
Government Medical College

Asif Amin  
University of Kashmir

Raies Ahmed Qadri  
University of Kashmir

Ruchi Shah  
University of Kashmir

Rakesh Kumar (✉ kumar.rakesh@smvdu.ac.in)  
Shri Mata Vaishno Devi University

Research Article

Keywords: Single Stranded Complimentary Polymorphism (SSCP), Gene Alignment tool Kit (GATK), Paired box (PAX), Single Nucleotide Polymorphism (SNP)

DOI: https://doi.org/10.21203/rs.3.rs-171897/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Leukemia is a heterogeneous disorder, characterized by a high proliferation of white blood cells. Various genetic studies have tried to reveal the role of contributory SNPs responsible for the development of leukemia. The expression of this factor is involved in several aspects of B-cell differentiation, including immunoglobulin gene rearrangement, BCR signal transduction, and B-cell survival so that the deletion or inactivating mutations of Pax5 causes cell arrest in the Pro-B-cell stage. The role of the present gene has previously been studied in various population groups; however, the role of this variant in Leukemia within Indian populations is unclear. In the present study, the role of genetic variant of rs3780135 of PAX5 was investigated within leukemia patients from Northern India. The variant was identified by exome sequencing and then genotyped using high throughput real-time based TaqMan assay technology and validated by SSCP and Sanger sequencing. The association of SNP with the disease was evaluated using logistic regression. It was observed that the variant rs3780135 of PAX5 showed significant association with Leukemia in the North Indian Population [OR of 1.94 (95% CI 1.51-2.48), \( p = 1.2 \times 10^{-6} \)] in real time based TaqMan assay and allelic OR of 1.54 (1.02-2.34), at 95% CI, \( p = 0.002 \) for sanger sequencing when corrected for age, gender, BMI, smoking, and alcohol. The present study concludes that the variant rs3780135 of PAX5 polymorphism act as a risk factor in the development of leukemia within Northern India population.

Introduction

Leukemia consists of a group of heterogeneous malignancies in which immature and dysfunctional hematopoietic progenitors proliferate and accumulate in the bone marrow\(^1\). In hematopoietic stem cells, a disruption of the cellular processes including proliferation, differentiation, and cell growth ultimately leads to Leukemia\(^2\). According to the Population-Based Cancer Registry of India, males are more affected than females by the ratio of 2:1\(^3\)\(^,\)\(^4\). The majority population groups of Jammu and Kashmir practice endogamy, thus preserving the gene pool. This factor has been increasingly documented as an inherited genetic factor that can contribute to the progression of leukemia\(^5\)\(^,\)\(^6\)\(^,\)\(^7\).

The variant rs3780135 of PAX5 is a fellow of the paired box domain gene family that translates for nuclear transcription factors which plays significant part in growth, variation, cell movement, propagation, and hematopoiesis\(^5\)\(^-\)\(^9\). The transcriptional movement of PAX5 (chr9:36840510-36840768 in genome browser) which regulates its target genes when determined by interaction of distinct partner proteins with the central and C-terminal protein interaction motifs of PAX5\(^8\). The fractional homeodomain of PAX5 associates with the TATA-binding protein of the basal transcription machinery and the C-terminal transactivation domain which regulates gene transcription most probably by interacting with histone acetyltransferases (HAT), such as the co-activator CBP or SAGA complex\(^9\) as shown in figure 1.

The appearance PAX5 in regular mature tissue is restricted to the hematopoietic system however it is abnormally articulated in a number of solid cancers as well as B-cell malignancies\(^10\)\(^,\)\(^11\). Damage of PAX5
expression hindered B cell development at an early pro-B-cell stage and relapses committed B-cell precursors (BCPs) to progenitors\textsuperscript{12,13}. \textit{PAX5} has regularly remained embattled for alterations and chromosomal translocations in childhood acute lymphoblastic leukemia (ALL)\textsuperscript{14-20}. It was observed that the genetic variant rs3780135 of \textit{PAX5} was associated with Acute lymphoblastic leukemia in various ethnic Chinese and Japanese populations\textsuperscript{21,22}. The aim of the present study is to discover the association of variant rs3780135 of \textit{PAX5} with the risk of Leukemia from the Jammu & Kashmir region. This study benefits in the detailed valuation of variant rs3780135 in the considered population and will highlight the pathways associated with it.

**Results**

The total exonic data generated was 45.63 GB with more than 100X sequencing coverage. A total of 62 shortlisted genes were screened and studied through this analysis. In the present study, it was observed that for homozygous 78,878 variants were reported in cases and 31,873 were present in controls while about 128,253 heterozygous variants were in cases and 46,938 were present in controls. A total of 15,024 missense variants were observed in all samples, out of which the \textit{PAX5} variant was frequently observed in all samples with a high percentage frequency. Gene details is given in Table 1. The exonic data has been uploaded in public domain UCSC Genome Browser and DBSNP with accession number NM_001280550.2.

**Table 1: Showing detail of variant rs3780135 of PAX5.**

| Gene | Zygosity       | SNP ID      | cDNA Change | Protein change | Exon No. | Function       | Type of Alteration  |
|------|---------------|-------------|-------------|----------------|----------|----------------|---------------------|
| \textit{PAX5} | Homozygosity   | rs3780135   | g.29416366 G>C | p.T264I        | 07       | Missense variant | Non-synonymous SNV |

As the variant of \textit{PAX5} was shortlisted used WES in a small subset of samples, and it was imperative to further validate in a larger sample size by using TaqMan assay and sanger sequencing. Furthermore, SSCP and was also done to confirm the mutation. The main aim of the study was to explore the association of variant rs3780135 of \textit{PAX5} with the risk of Leukemia in the present population.

**Table 2: Clinical Characteristic distribution in cases and controls.**
### Characteristics

| Characteristics | Cases     | Controls   | p-value |
|-----------------|-----------|------------|---------|
| Age *           | 40.7 (±14.5) | 49.1 (±15.30) | <0.01   |
| Gender (in %)   | F=40      | F=30       | -       |
|                 | M=60      | M=70       |         |
| BMI**           | 21.2 (±6.10) | 24.9 (±5.3) | <0.01   |

### Histopathology Types

- Acute Lymphoblastic Leukemia (ALL) 0.17
- Acute Myeloid Leukemia (AML) 0.23
- Chronic Myeloid Leukemia (CML) 0.50
- Chronic Lymphoid Leukemia (CLL) 0.10

### Smoker

- 0.63 vs 0.22

### Non-Smoker

- 0.37 vs 0.78

### Alcoholic

- 0.53 vs 0.17

### Non-Alcoholic

- 0.47 vs 0.83

*age in years and **BMI in kg/m²

Whole Exome Sequencing analysis showed PAX5 gene frequently mutated in respect to other shortlisted genes. A total of 600 subjects (200 leukemic cases and 400 healthy controls) were recruited for Real-Time TaqMan assay.

In Table 2, 60% of Leukemia patients were males and 40% were females, while in controls, 70% were males and 30% were females. The mean ages were 40.51 (±14.67) for leukemia patients and 50.76 (±13.30) years for controls. BMI in cases was 21.2 (±6.10) and 24.9 (±5.3) in Controls. The clinical characteristic distributions of the cases and the controls are given in Table 2. The allele frequency distribution of the variant rs3780135 of PAX5 between cases and controls is summarized in Table 3. In the current study, allele A is present in more cases (0.62) than in controls (0.54), hence suggesting that allele A is causing risk. It was observed that genetic allele A of variant rs3870135 of PAX5 is significantly associated with leukemia (p value =0.025), with H.W. E = 0.974.

Table 3: Allelic frequency distribution between cases and controls.
To observe the maximum effect of allele A, association of PAX5 was observed by using dominant model. The OR observed was 1.6 (0.94-2.4) at 95% CI in leukemia corrected for age, gender and BMI. Furthermore, evaluation was done on variant rs3780135 of PAX5 for other genetic models as per the risk allele and the results observed showed positive association of variant in all the three models as shown in Table 4.

Table 4: Showing the association of variant rs3780135 of PAX5 with leukemia in North Indian population using genetic models.

| SNP ID   | R  | Genetic Models | Genotype          | p-value | OR (95%CI)       |
|----------|----|----------------|-------------------|---------|-----------------|
| rs3780135| A  | Dominant model | TT +TC Vs CC      | 0.08    | 1.6 (1.01-2.16) |
|          |    | Recessive Model| TT vs TC + CC     | 0.001   | 1.7(0.8-3.54)   |
|          |    | Additive Model | TT vs TC vs CC    | 0.02    | 1.48 (1.01-2.16)|

*Corrected for Age, Gender and BMI

Validation of rs3780135 of PAX5 by SSCP and Sanger sequencing.

Single Strand Conformational Polymorphism:

For PCR amplified gene sequences, SSCP was done to analyze the mutational status in 200 cases and 400 controls. Based on the electrophoretic mobility, DNA strands separate into unique three-dimensional structures exhibiting different band patterns; heterozygous and homozygous mutant showing three to four bands. Its detection capacity was utilized to screen affected cases and controls to detect genetic variability in targeted populations as shown in figure 2.

Sanger Sequencing:

Sanger sequencing of the PAX5 (exon 7) with genomic position g:36840626 (AGT) to (AAT) G>A with the amino acid alteration from Threonine to Isoleucine with 391 amino acids. Allelic frequency distribution as shown in Figure 3 and the chromatogram of PAX5 sequence generated homozygous wild (GG), heterozygous (AG), and Homozygous mutant (AA) in as shown in Figure 4.

The allelic OR of variant rs3780135 of PAX5 was 1.54 (1.01-2.34) with p-value = 0.004 and H.W.E it was 0.838. The OR observed was 2.34(1.09-3.21, at 95% CI), p=0.002 as shown in Table 5 when corrected for
age, gender, and BMI. Thus, it was observed that the variant rs3780135 of PAX5 shows a significant association with the risk of leukemia for North Indian population.

Table 5: Showing Allelic frequency distribution of variant rs3780135 of PAX5.

| SNPs /Gene | Allele frequency Cases (N=150) | Allele frequency Controls (N=150) | Allelic OR* (95% C.I.) | p-value | Total HWE | Dominant OR* (95% C.I.) | p-value |
|-----------|-------------------------------|----------------------------------|------------------------|---------|-----------|------------------------|---------|
| rs3780135 PAX5 | G=51.1 | G=61.7 | 1.54 (1.0-2.34) | 0.004 | 0.838 | 2.34 (1.0-3.21) | 0.002 |

In-silico analysis of variant rs3780135 of PAX5

Secondary structure analysis of PAX5

The unstructured and uncharacterized region of the PAX5 (isoform 4) containing a mutation at T 264, from aa 211 till 295 was subjected to the secondary structure analysis in the Jpred 4, Protein Secondary Structure Prediction Server 23 as shown in figure 5.

It was observed from the Jpred secondary structure analysis that the region in question, (aa 211–295) containing the mutation at T 264, did not show any prominent structural characterization. However, the region from aa 231 till 251 shows a short beta-sheet and alpha-helix, as shown in Figure 5. Green represents the beta-sheet while red denotes the alpha helix. The mutation T 264 is marked in the box in the vicinity of the unstructured region.

The same region (aa 211 till 295) was also analyzed in the Raptor X server 24, for the characterization of the secondary structure elements. Figure 6 (a) also shows the absence of a structured region along with the mutation at T 264 which is marked in the box in the figure. The short alpha-helix is observed in the position from aa 240 to 250. Raptor X, also measures the protein disorder propensity, from the disorder propensity analysis it was observed that the chunk of 10 aa covering T 264 as shown in the box marked in the figure (Figure 6 b) indicates the maximum disorder in this region. The red block indicates the maximum disorder which is followed in the rest of the sequence as well.

Domain Identification of the Pax5 uncharacterized region.

The Pfam database at EMBL-EBI which is a large collection of protein families and analyses the protein sequence from Pfam matches and based on which it determines the family of the proteins 25. The region 211–295 aa of PAX5 which is largely unstructured as shown in the analysis above, was provided as an input to the Pfam search. The Pfam characterized this region in the Homeodomain family (PF00046) belonging to the clan CL0123. Homeodomain proteins are responsible for regulating gene expression and cell differentiation during the early stages of embryonic development and have a characteristic protein
fold that binds to the DNA and regulates the expression of target genes. The clan CL0123 has a diverse range of DNA-binding domains that predominantly contain a helix-turn-helix motif.

**Discussion**

PAX5 plays important role at time of checkpoints in B lymphoid maturation and leukemogenesis. Mutations and deletions of PAX5 have been well-thought-out so far as subordinate oncogenic measures because they were originated in several BCP-ALL subtypes and probably in slight subclones. In the present study, an attempt was made to explore the association of the variant rs3780135 of PAX5 with Leukemia among patients from North India. The same variant was found significantly associated with acute lymphoblastic leukemia in Turkish populations, Germany. However, the same variant show protective effect in B cell acute lymphoid leukemia in Pakistan population. The present study indicated that genetic variants rs3780135 of PAX5 pose as risk factors for leukemia.

PAX5 is a transcription factor that is required for B-cell development and its maintenance. PML is a tumor suppressor and a pro-apoptotic factor. PML has been found to be translocated to the PAX5 locus to generate a PAX5-PML fusion gene in childhood acute lymphoblastic leukemia which disrupts PAX5 function.

**PAX5 expression in Normal and Cancer State**

PAX5 mediated transcription of lymphoid cells requires the activation of PAX5 promoter and recruits RNA Pol II positively and negatively regulating variable gene expression while in myeloid lineage, Polycom complex-mediated transcription repression inhibits Pol II activity regulating gene expression for myeloid lineage as hypothesized in figure 7.

In the Cancer state, MAPK activity Polycomb complex is evicted from the PAX5 promoter resulting in ectopic expression leading to the development of AML. On the other hand, PAX5 adhered fusion gene PML represses the expression of PAX5 in lymphoid progenitors. This dominant-negative effect of PML-PAX5 confers negative regulation thereby impairing DNA binding and represses PAX5 expression on its target substrates. Also, inhibition of PML sumoylation induced by the fusion gene induces resistance towards programmed cell death as studied in HeLa cells and is thus one of the contributory factors in the development of ALL.

Therefore, as it indicates that PAX5 are regulators of cellular processes, differentiation, and haematopoiesis and slightly modification in these genes may lead to the extension of leukemic risk. Therefore, it is much desirable to discover the effect of these genetic variants in the molecular functioning of PAX5 and the linked downstream signalling pathways.

**Conclusions**
Our findings provide evidence that the variant rs3780135 of PAX5 is associated with the predisposition of leukemic risk in the Jammu and Kashmir region. A study on large cohort may help in understanding the effect of this variant in different ethnic populations and may act as a predictive or prognostic biomarker for leukemia.

**Material And Methods**

**Ethics statement**

An approval for the study under the notification number (SMVDU/IERB/16/41) was taken from the Institutional Ethics Review Board (IERB) of Shri Mata Vaishno Devi University (SMVDU). All the details were recorded in a predesigned proforma and a well written informed consent was obtained from both cases and controls. All experimental procedures were conducted according to the guidelines and regulations set by the IERB, SMVDU.

**Sample Collection and DNA isolation**

An overall of 600 subjects were drafted for the study, of which 200 were cases (leukemic patients) and 400 remained healthy controls. Using the FlexiGene® Qiagen DNA isolation Kit (Catalogue No.51206), the genomic DNA was isolated from the blood samples. The quality of the genomic DNA was checked by 0.8% agarose gel electrophoresis (Bio-Rad Gel Doc™ EZ imager) and quantification was done by using Bio-Spectrometer™ (Eppendorf India Pvt. Ltd.).

**Whole Exome Sequencing**

The exome sequencing for samples was done on the Illumina NGS platform. The alignment against the hg GrCh37 genome was carried out using Burrows -Wheeler Aligner (BWA) and the variant calling was done using the Gene Alignment Tool Kit (GATK) pipeline. The alignment against the hg19 GrCh37 genome was carried out by using BWA and the variant calling was done by GATK pipeline. The variants were identified by using the GATK tool and annotated using ANNOtate VARiation (ANNOVAR). The databases used for the study include the 1000 genome, (database for nonsynonymous SNP's functional predictions (DBNSFP), Single Nucleotide Database (dbSNP), Genome Wide Associated Studies (GWAS). After filtration of variants against various filters by the use of Exome Capture Kits targeted region.

**Genotyping**

Genotyping of variants rs3780135 of PAX5 was achieved by using the TaqMan allele discrimination assay MX3005p labeled with VIC and FAM dyes (Thermo Fisher Scientific) and UNG Master Mix (Applied Bio-systems, USA). The capacity of the total PCR reaction was 10µl, including of 2.5 µl of TaqMan UNG Master Mix, 0.25 µl of the probe, 3µl DNA (5ng/µl) and 4.25 µl nuclease-free water further together to make the final volume. The thermal conditions approved were 10 minutes at 95 °C, 40 cycles of 95°C for 15 seconds and 60°C for 1 min. Entirely the samples were run in a 96-well plate with three no template
controls (NTCs). The post PCR detection system was used to measure allele-specific fluorescence. A overall of 93 random samples respectively from cases and controls were selected and re-genotyped for cross-validation of the genotyping calls and the agreement rate was 100%.

**Single-Strand Conformational Polymorphism (SSCP):**

SSCP is a PCR based technique for the rapid detection of mutations in target gene fragments under non-denaturing conditions\(^{35}\). In such conditions, PCR amplified DNA fragments are denatured at 95°C and immediately snap chilled. Following this the samples are loaded and run through 10% native polyacrylamide (PAGE) gel (Bio-Rad- Mini-PROTEAN Tetra Cell)\(^{36}\) after that the gel was stained with Ethidium bromide (Etbr 1 microgram/ml) for 5-7 minutes and was checked (Bio-Rad Gel Doc™ EZ imager). If mutation at a specific nucleotide position persists in the DNA fragment, conformational alterations can be identified.

**Sanger Sequencing**

Genotyping of *PAX5* (rs3780135) was done by polymerase chain reaction (PCR) and the primers forward 5'-TCACCCTCAATAGGTGCCATC and reverse 3'-ACTGGGACAGAGATCTTGGTGA which were designed by primer 3 based software. Overall samples were amplified using the PCR program with 95°C for 1 minute, 95°C for 1 min, 60°C for 45 seconds with 36 cycles, 72°C for 1 min, 72°C for 2min, 4°C infinity (∞), and sequencing was performed by Aggrigenome Pvt. Limited using Sequence Scanner Software (ABI 3730 XL DNA Analyzer) and Chromas v2.6.6 (Technelysium Pty Ltd, South Brisbane, Australia).

**Secondary structure analysis of *PAX5***

The unstructured and uncharacterized region of the *PAX5* (isoform 5) containing a mutation at T 264, from amino acid (aa) 211 till 295 was subjected to the secondary structure analysis in the Jpred 4, Protein Secondary Structure Prediction Server\(^{23}\). The same region (aa 211 till 295) was also analyzed in the Raptor X server\(^{24}\), for the characterization of the secondary structure elements and the protein disorder propensity.

**Domain Characterisation in Pax5**

The domains present in the same uncharacterized region (from aa 211 till 295) of Pax5 were determined from the Pfam database at EMBL-EBI\(^{25}\).

**Statistical analysis:**

Statistical Analysis of the data was done by using Statistical Package for the Social Science (SPSS) software (v.20; Chicago, IL). Chi-square (\(\chi^2\)) was performed and genotype frequencies were tested for total Hardy–Weinberg equilibrium. Binary Logistic Regression was also used to estimate OR at 95% CI and the respective level of significance was estimated as p-value as the level of significance from confounding factors like age, gender, and BMI.
Declarations

Conflict of Interest: None

Acknowledgment:

RK and AB acknowledges Department of Science & Technology, Government of India SSTP/J&K/459.

Author information

Affiliations

Cancer Genetics Research Group, School of Biotechnology, Shri Mata Vaishno Devi University, Katra, J&K-182320.

Amrita Bhat, Gh. Rasool Bhat, Rakesh Kumar

Genomic Stability Lab, School of Biotechnology, Shri Mata Vaishno Devi University, Katra, J&K, India,182320.

Monika Pandita

Bioinformatics Infrastructure Facility, School of Biotechnology, Shri Mata Vaishno Devi University (SMVDU), Katra, J&K-182320.

Supriya Hakim

ICMR-CAR, Shri Mata Vaishno Devi University, Katra, J&K, India-182320.

Sonali Verma, Gresh Chander

Department of Radiotherapy, Government Medical College Kathua.

Shabab Angurana

ICMR-CAR, Department of Biotechnology, University of Kashmir-190006

Ruchi Shah, Asif Amin, Raies A. Qadri

Author Contribution

R.K, RS RAQ and A.B, planned the work. A.B carried out all the experimental work and wrote the manuscript MP and G.R.B helped in statistical analysis while S.V, GC, AA, SA helped in sampling processes. SH help in bioinformatic analysis. All authors finally revised and approved the manuscript.

Ethics declarations
Competing interests

No competing interests declared by Authors.

Additional information

Correspondence and request for materials should be addressed to RK and RS. No copyright material (figure and tables) is inserted in Manuscript and the data is totally based on Lab generated data and analysis.

References

1 National Cancer Institute. Cancer.gov en español2 (2018).

2 Leukemia. American Society of Hematology.1 (2017).

3 Asthana, S., Labani, S., Mehrana, S. & Bakhshi, S. Incidence of childhood leukemia and lymphoma in India. Pediatric Hematology Oncology Journal (2018).

4 Chokkalingam, A. P. & Buffler, P. A. Genetic susceptibility to childhood leukaemia. Radiation protection dosimetry132, 119-129 (2008).

5 Bhat, A. A. et al. Role of non-coding RNA networks in leukemia progression, metastasis and drug resistance. Molecular Cancer19, 1-21 (2020).

6 Bhat, G. R. et al. Association of newly identified genetic variant rs2853677 of TERT with non-small cell lung cancer and leukemia in population of Jammu and Kashmir, India. BMC cancer19, 1-8 (2019).

7 Bhat, A. et al. Association of ARID5B and IKZF1 variants with leukemia from Northern India. Genetic testing and molecular biomarkers23, 176-179 (2019).

8 Gu, J. et al. Identification of significant pathways induced by PAX5 haploinsufficiency based on protein-protein interaction networks and Cluster analysis in Raji Cell Line. BioMed research international2017 (2017).

9 Fazio, G., Biondi, A. & Cazzaniga, G. in Novel Aspects in Acute Lymphoblastic Leukemia (IntechOpen, 2011).

10 Yan, M. et al. Development of cellular immune responses against PAX5, a novel target for cancer immunotherapy. Cancer research68, 8058-8065 (2008).

11 Tiacci, E. et al. PAX5 expression in acute leukemias: higher B-lineage specificity than CD79a and selective association with t (8; 21)-acute myelogenous leukemia. Cancer Research64, 7399-7404 (2004).
12 Carotta, S., Holmes, M., Pridans, C. & Nutt, S. L. Pax5 maintains cellular identity by repressing gene expression throughout B cell differentiation. *Cell Cycle* **5**, 2452-2456 (2006).

13 Cobaleda, C., Jochum, W. & Busslinger, M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* **449**, 473 (2007).

14 Santoro, A. *et al.* Altered mRNA expression of PAX5 is a common event in acute lymphoblastic leukaemia. *British journal of haematology* **146**, 686-689 (2009).

15 Nebral, K. *et al.* Incidence and diversity of PAX5 fusion genes in childhood acute lymphoblastic leukemia. *Leukemia* **23**, 134 (2009).

16 Bousquet, M. *et al.* A novel PAX5-ELN fusion protein identified in B-cell acute lymphoblastic leukemia acts as a dominant negative on wild-type PAX5. *Blood* **109**, 3417-3423 (2007).

17 Fazio, G., Palmi, C., Rolink, A., Biondi, A. & Cazzaniga, G. PAX5/TEL acts as a transcriptional repressor causing down-modulation of CD19, enhances migration to CXCL12, and confers survival advantage in pre-BI cells. *Cancer research* **68**, 181-189 (2008).

18 Thomas-Tikhonenko, A. & Cozma, D. PAX5 and B-cell neoplasms: transformation through presentation. (2008).

19 Zwollo, P. *et al.* The Pax-5 gene is alternatively spliced during B-cell development. *Journal of Biological Chemistry* **272**, 10160-10168 (1997).

20 Arseneau, J. R., Laflamme, M., Lewis, S. M., Maïcas, E. & Ouellette, R. J. Multiple isoforms of PAX5 are expressed in both lymphomas and normal B-cells. *British journal of haematology* **147**, 328-338 (2009).

21 Moriyama, T., Relling, M. V. & Yang, J. J. Inherited genetic variation in childhood acute lymphoblastic leukemia. *Blood* **125**, 3988-3995 (2015).

22 Tasian, S. K. & Hunger, S. P. Genomic characterization of paediatric acute lymphoblastic leukaemia: an opportunity for precision medicine therapeutics. *British journal of haematology* **176**, 867-882 (2017).

23 Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. J. n. a. r. JPred4: a protein secondary structure prediction server. *43*, W389-W394 (2015).

24 Källberg, M., Margaryan, G., Wang, S., Ma, J. & Xu, J. in *Protein Structure Prediction* 17-27 (Springer, 2014).

25 Bateman, A. *et al.* The Pfam protein families database. *32*, D138-D141 (2004).

26 Gehring, W. J. J. t. i. b. s. The homeobox in perspective. *17*, 277-280 (1992).
27 Gu, Z. et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nature genetics* **51**, 296-307 (2019).

28 Passet, M. et al. PAX5 P80R mutation identifies a novel subtype of B-cell precursor acute lymphoblastic leukemia with favorable outcome. *blood* **133**, 280-284 (2019).

29 Firtina, S. et al. Evaluation of PAX5 gene in the early stages of leukemic B cells in the childhood B cell acute lymphoblastic leukemia. *Leukemia research* **36**, 87-92 (2012).

30 Auer, F. et al. Inherited susceptibility to pre B-ALL caused by germline transmission of PAX5 c. 547G>A. *Leukemia* **28**, 1136-1138 (2014).

31 Khalid, A., Aslam, S., Ahmed, M., Hasnain, S. & Aslam, A. Risk assessment of FLT3 and PAX5 variants in B-acute lymphoblastic leukemia: a case–control study in a Pakistani cohort. *PeerJ* **7**, e7195 (2019).

32 Delogu, A. et al. Gene repression by Pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells. *Immunity* **24**, 269-281 (2006).

33 Kurahashi, S. et al. PAX5–PML acts as a dual dominant-negative form of both PAX5 and PML. *Oncogene* **30**, 1822 (2011).

34 Desouki, M. M., Post, G. R., Cherry, D. & Lazarchick, J. PAX-5: a valuable immunohistochemical marker in the differential diagnosis of lymphoid neoplasms. *Clinical medicine & research* **8**, 84-88 (2010).

35 Gasser, R. B. et al. Single-strand conformation polymorphism (SSCP) for the analysis of genetic variation. *Nature protocols* **1**, 3121 (2006).

36 Spinardi, L., Mazars, R. & Theillet, C. Protocols for an improved detection of point mutations by SSCP. *Nucleic acids research* **19**, 4009 (1991).

**Figures**

![Figure 1](image-url)
PAX5 structure, showing functional domains and corresponding exons. DBD, DNA binding paired domain; OP, conserved octapeptide; HD, partial homeodomain; TAD, transactivation domain; ID, inhibitory domain.

Figure 2

Sample screening using Single Strand Conformational Polymorphism: (a) wild type (b) lane 1, lane 3 lane 5 and 6 shows wild type (2 bands) while as lane 2 shows heterozygosity (3 bands) and lane 4 shows homozygous mutant (4 bands).

Figure 3

Showing Allelic frequency distribution of variant rs3780135 of PAX5
Figure 4

Chromatogram of PAX5 showing missense mutation G>A a) homozygous wild (GG) b) heterozygous (AG) c) Homozygous mutant (AA).

Figure 5

Secondary structure analysis of PAX5 unstructured region containing the T 264 marked in the box
Figure 6

(a). Raptor X secondary structure Characterisation of the unstructured region showing lack of prominent secondary structure elements in the box marked with T264. (b). The protein disorder propensity showing the maximum disorder (red) in the marked box covering T 264.
Figure 7

Pathway of PAX5 in both normal and cancer state in myeloid and Lymphoid lineage.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TitlepagewithsupplimentaryfilePAX5manuscript.pdf