

ORIGINAL ARTICLE**Identification of Prognostic and Susceptibility Markers in Chronic Myeloid Leukemia Using Next Generation Sequencing****Yogender Shokeen^{1*}, Neeta Raj Sharma³, Abhishek Vats², Vibha Taneja², Sachin Minhas¹, Mayank Jauhri¹, Satish Sankaran⁴, Shyam Aggarwal¹****OPEN ACCESS**

Citation: Yogender Shokeen, Neeta Raj Sharma, Abhishek Vats, Vibha Taneja, Sachin Minhas, Mayank Jauhri, Satish Sankaran, Shyam Aggarwal. Identification of Prognostic and Susceptibility Markers in Chronic Myeloid Leukemia Using Next Generation Sequencing. *Ethiop J Sci.* 2018;28(2):135.

doi:<http://dx.doi.org/10.4314/ejhs.v28i2.5>.

Received: August 3, 2017

Accepted: August 7, 2017

Published: March 1, 2018

Copyright: © 2018 Yogender S., et al .

This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was funded by Research Development Program, Sir Ganga Ram Hospital, Delhi, India (RDP Study Code: 4.9.34.001) and Novartis Healthcare Private Ltd, Mumbai, India.

The funding was solely for research purpose and funding agency had no role in subject enrollment, data generation, analysis or publication.

Competing Interests: The authors declare that this manuscript was approved by all authors in its form and that no competing interest exists.

Affiliation and Correspondence:

¹Department of Medical Oncology, Sir Ganga Ram Hospital, Rajinder Nagar, Delhi, India

²Department of Research, Sir Ganga Ram Hospital, Rajinder Nagar, Delhi, India

³Department of Pediatrics and Child Health and Pediatric Emergency Consultant, School of Medicine, Addis Ababa University, Ethiopia

³School of Bio-Engineering and Biosciences, Lovely Professional University, Jalandhar, Punjab, India

⁴Strand Center for Genomics and Personalized Medicine. UAS Alumini Building, Veterinary College Campus, Bellary Road, Hebbal, Bangalore, India

*Email: yogendershokeen@yahoo.co.in

ABSTRACT

BACKGROUND: Incidence of Chronic Myeloid Leukemia (CML) is continuously increasing and expected to reach 100,000 patients every year by 2030. Though the discovery of Imatinib Mesylate (IM) has brought a paradigm shift in CML treatment, 20% patients show resistance to this tyrosine kinase inhibitor (TKI). Therefore, it is important to identify markers, which can predict the occurrence and prognosis of CML. Clinical Exome Sequencing, panel of more than 4800 genes, was performed in CML patients to identify prognostic and susceptibility markers in CML.

METHODS: Enrolled CML patients (n=18) were segregated as IM responders (n=10) and IM failures (n=8) as per European Leukemia Net (ELN), 2013 guidelines. Healthy controls (n=5) were also enrolled. DNA from blood of subjects was subjected to Next Generation Sequencing. Rare mutations present in one patient group and absent in another group were considered as prognostic markers, whereas mutations present in more than 50% patients were considered as susceptibility markers.

RESULT: Mutations in genes associated with cancer related functions were found in different patient groups. Four variants: rs116201358, rs4014596, rs52897880 and rs2274329 in C8A, UNC93B1, APOH and CA6 genes, respectively, were present in IM responders; whereas rs4945 in MFG8 was present in IM failures. Mutations in HLA-DRB1 (rs17878951), HLA-DRB5 (rs137863146), RPHN2 (rs193179333), CYP2F1 (rs116958555), KCNJ12 (rs76684759) and FUT3 (rs151218854) were present as susceptibility markers.

CONCLUSION: The potential genetic markers discovered in this study can help in predicting response to IM as frontline therapy. Susceptibility markers may also be used as panel for individuals prone to have CML.

KEYWORDS: Chronic Myeloid Leukemia, Genetic Markers, Next Generation Sequencing (NGS)

INTRODUCTION

Incidence of Chronic Myeloid Leukemia in the United States is 1.6 per 100,000 of population, accounting approximately for 5,430 cases annually. The estimated prevalence in the US is around 25,000–30,000, which is expected to reach around 105,000 by the year 2030 (1). Its incidence in India is between 0.8–2.2 per 100,000 per annum (2). According to the current estimate, the world is expected to have 100,000 newly diagnosed CML patients per year, leading this chronic disorder into a serious health problem (3). Hence, the markers, with potential to detect individuals prone to have CML, are very important to control rising incidence.

According to the World Health Organization guidelines, CML is diagnosed with the presence of BCR-ABL gene, which is a reciprocal translocation between chromosome 9 and chromosome 22. Development of BCR-ABL inhibitor, Imatinib Mesylate (IM), spectacularly changed the prognosis with obstruction of its tyrosine kinase activity and impeding its further downstream signaling. The drug has high efficacy rates: the major molecular response (MMR) of IM as first line therapy at 12 months is 22%, which increases to 44% at 24 months (4). After 2 years of IM therapy, 40% of the patients achieve complete molecular response (CMR), which further increases to 81% at 10 years of treatment (5). Despite this commendable drug discovery, approximately 20% of the patients show resistance following treatment with this TKI therapy (6). European Leukemia Net (ELN) 2013 recommendations suggest that CML patients on first line TKI therapy can be segregated as responders or failures, where responders should continue with the regimen and failure are advised to change the treatment to either more potent TKI or dose escalation (7). Though prognostic scores (SOKAL and EUTOS) are available for estimation of outcomes, they have failed for several clinical conclusions (8,9). Therefore, genetic markers, with better reliability and accuracy, are required for estimation of prognosis and planning of personalized medicine in CML. In the present study, we performed clinical exome

sequencing in CML patients to search potential prognostic markers for IM response and susceptibility markers for CML.

MATERIALS AND METHODS

Subjects: The study was conducted after obtaining approval from the Ethics Committee of Sir Ganga Ram Hospital, a super specialty tertiary care centre in Delhi (India). Informed consent was taken from each subject before enrollment in the study. Adult patients (>18 years) diagnosed with CML were enrolled. Diagnosis of CML was confirmed by reverse transcription polymerase chain reaction (RT-PCR) for BCR-ABL fusion gene and fluorescent in situ hybridization (FISH) for translocation (9,22). Enrolled patients were classified as responders and resistant as per ELN, 2013 criteria. According to ELN 2013 criteria, CML patients with $\leq 10\%$ BCR-ABL or 35% Ph chromosome are considered as responders after 3 months of TKI treatment, whereas patients without complete hematological response or 95% Ph chromosome are considered as resistant. After 6 months of treatment, the criteria changes to presence of $< 1\%$ BCR-ABL or 0% Ph chromosome for response and for resistance $> 10\%$ BCR-ABL or 35% Ph chromosome. Patients harboring $< 0.1\%$ BCR-ABL or are referred as responders and patients with $> 1\%$ BCR-ABL as resistant after 12 months of treatment. Beyond 12 months of TKI treatment, loss of CHR, CMR or MMR is considered as resistance, whereas presence of $< 1\%$ BCR-ABL is considered as optimal response. MMR is defined as presence of $\leq 0.1\%$ BCR-ABL for the present study. Patients not achieving MMR at 18 months of IM treatment were regarded as suboptimal responders, whereas increase in BCR-ABL% at any time line compared to previous observation was considered as loss of molecular response and disease relapse (7). Healthy subjects with no known history of malignancy and above 18 years of age were also enrolled as controls.

Peripheral blood sample from both patients and controls was obtained in EDTA vials after taking informed consent. DNA was extracted (Nucleospin, Macheley-Nager, Germany) and was subjected to sequencing.

Clinical exome sequencing

Next generation sequencing: A targeted panel (TruSight One, Illumina, USA) with probes covering all coding exons and essential splice

sites of >4800 clinically relevant genes was used for sequencing the DNA samples using Illumina's TruSight technology (Illumina, USA). The target size for this is panel 12 Mb and designed based on the information in the Human Gene Mutation Database (HGMD), the Online Mendelian Inheritance in Man (OMIM) catalogue, GeneTests.org, Illumina TuSight sequencing panels, etc.

DNA (50 ng) is taken for library preparation. TruSight One library preparation uses transposon based shearing of the genomic DNA. A limited cycle PCR step allows the incorporation of adaptors, platform-specific tags and barcodes to prepare the libraries. The tagged and amplified sample libraries were analyzed for quality using BioAnalyzer (Agilent, USA) and quantified using Qubit. And, 500 ng of each library was pooled and hybridized to biotinylated probes. The hybridized target DNA fragments were pulled down using streptavidin beads and two successive enrichment steps were performed to optimize the regions of interest. Target libraries were amplified using loaded for sequencing on the MiSeq or HiSeq (Illumina, USA) to obtain ~3 GB per sample.

Sequence analysis: The trimmed FASTQ files were generated using MiSeq Reporter from Illumina for MiSeq data and Casava software for HiSeq data. The reads were aligned against the whole genome build hg19 using STRAND® NGS V2.1.6 (Strand LifeSciences Pvt. Ltd., Bangalore). Five base pairs from the 3' end were trimmed, as 3' end were bases with quality below 10. Reads which had lengths less than 25 bp after trimming were not considered for alignment. A maximum of 5 matches of alignment score at least 90% were computed. Reads that failed QC (quality control), reads with average quality less than 20, reads with ambiguous characters were all filtered out. The STRAND® NGS variant caller was used to detect variants at locations in the target regions covered by a minimum of 10 reads with at least 2 variant reads. Variants with a decibel score of at least 50 were reported.

Interpretation: Interpretation of the variant data was done using the StrandOmics software, V1.9. The variant annotation engine includes different algorithms to identify variant impact on gene using public content (HGMD, ClinVar, OMIM, HPO, links to dbSNP, 1000 Genomes, Exome Variant Server) and proprietary content (curated variant records).

Statistical analysis: Statistical analyses were performed using SPSS for Windows 16 (SPSS, Chicago IL, USA). Non-parametric data is presented as median with standard deviation (SD). Statistical significance of differences of quantitative data was determined using Mann-Whitney U test. P-value of less than 0.05 was considered to be significant.

Ethics statement: The study and informed consent forms were reviewed and approved by the Ethics Committee of Sir Ganga Ram Hospital, Delhi, India (EC No.: EC/11/12/439). All procedures performed in the this study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration

RESULTS

Subjects: Eighteen patients and five healthy controls were enrolled in the study. Mean (SD) age of patients was 42.11 (12.76) years, whereas that of controls was 38.40 (12.76) years ($p = 0.570$). All the patients were in chronic phase at the time of diagnosis. The IM responder group had 10 patients (55.55%), while 8 patients (44.45%) were in IM failure group. Patient no.4 and patient no.5 were included in responder group despite having >1% of BCR-ABL as the patients were responding well to IM therapy. Reduction in BCR-ABL percentage (compared to time of diagnosis) from 9.8% to 1.3% in patient no.4 and from 8.6% to 1% in patient no.5 was observed in span of 4 months and 5.5 months respectively. The mean (SD) age of the responder group was 35.1 (6.74) years, whereas failure was 50.88(13.36) years ($p = 0.05$).

Genetic markers: Our main objective was to detect prognostic and susceptibility markers. The criterion for genetic variants to be considered as marker was presence in at least 50% patients of one group and absence in all patients of the opposite group, whereas mutations present in more than 50% (≥ 9) of CML patients were considered as susceptibility markers. Only rare mutations, which were not present in healthy population of Asia, Europe and

Africa as per HapMap project, were considered. The shortlisted mutations were annotated manually with GeneCards and Kegg pathways for their functions (Table 3). Important genetic information about discovered mutations is given in Table 1.

Table 1: Genetic information of potential markers discovered in CML patients

| Mutation | Gene | Allele | Accession No. | Chromosomal Position ^ε | Codon | Amino Acid change | CpG islands [▲] |
|----------------------------------|----------|--------|----------------|-----------------------------------|---------|-------------------|--------------------------|
| Markers for IM Responder | | | | | | | |
| c.107C>A (rs116201358) | C8A | A/C | NM_000562.2 | 1; 57333311 | GCA-GAA | p.36Ala>Glu | None |
| c.1494G>A (rs4014596) | UNC93B1 | C/T | NM_030930.2 | 11; 67759316 | GTG-ATG | p.499Val>Met | 67758155-67759445 |
| c.422T>C (rs52797880) | APOH | A/G | NM_000042.2 | 17; 64216854 | ATC-ACC | p.141Ile>Thr | None |
| c.209G>C (rs2274329) | CA6 | C/G | NM_001215.3 | 1; 9009451 | GGC-GCC | p.70Gly>Ala | None |
| Marker for IM Failures | | | | | | | |
| c.7C>A (rs4945) | MFGE8 | G/T | NM_001114614.1 | 15; 89456544 | CGC-AGC | p.3Arg>Ser | 89456225-89456830 |
| CML Susceptibility Marker | | | | | | | |
| c.181T>C (rs137863146) | HLA-DRB5 | G/G | NM_002125.3 | 6; 32489871 | TAT-CAT | p.61Tyr>His | 32489743-32490128 |
| c.200T>C (rs17878951) | HLA-DRB1 | A/G | NM_002124.3 | 6; 32552056 | GTG-GCG | p.67Val>Ala | 32551852-32552331 |
| c.1070A>T (rs193179333) | RHPN2 | A/T | NM_033103.4 | 19; 33493188 | CAC-CTC | p.357His>Leu | None |
| c.798G>C (rs117958555) | CYP2F1 | C/G | NM_000774.3 | 19; 41628014 | CAG-CAC | p.266Gln>His | None |
| c.785T>G (rs76684759) | KCNJ12 | G/T | NM_021012.4 | 17; 21319439 | ATC-AGC | p.262Ile>Ser | 21318651-21320041 |
| c.1007A>C (rs151218854) | FUT3 | G/T | NM_000149.3 | 19; 5843844 | GAT-GCT | p.336Asp>Ala | None |

^εChromosomal Position = Chromosome no.; nucleotide position, [▲]CpG = Cytosine phosphate Guanine

Clinical exome analysis: Data was uploaded to SeattleSeq Annotation 141 (<http://snp.gs.washington.edu/SeattleSeqAnnotation141>) for annotation. Analysis of patients and controls was done separately. All the variants present in the control group were aligned in one table and all variants of patients group were aligned in another table. In total, 97,462 variants in patients and 65,546 variants in controls were found. After filtration of common variants, 45,950 variants were selected exclusively in the patient group. The distribution of genetic variants in the patient group included 30,525 intronic, 2,837 variants were in untranslated regions (3'UTR = 1,737 and 5' UTR = 1,100), 5,132 variations were synonymous and 7,494 variants were non-synonymous. Non-synonymous mutations were further segregated as 6,524 missense mutations,

766 frame-shift and 166 stop gained/stop lost mutations (Figure 1). We targeted only missense mutations, as these mutations play a central role in altering phenotype of the gene at protein level.

Variant analysis of significant genes: Total 2,033 missense variants were common in at least 2 patients, and only 273 variants were left after considering the variants present in more than 25% (≥ 5) patients. These variants were further analyzed for molecular pathway and functional correlation with CML. All the mutations observed were subjected to Kyoto Encyclopedia for Genes and Genomics (Kegg) pathways and GeneCards to identify the cellular pathway and functions associated with the mutated genes. We focused on the genes which were most frequently mutated in this group.

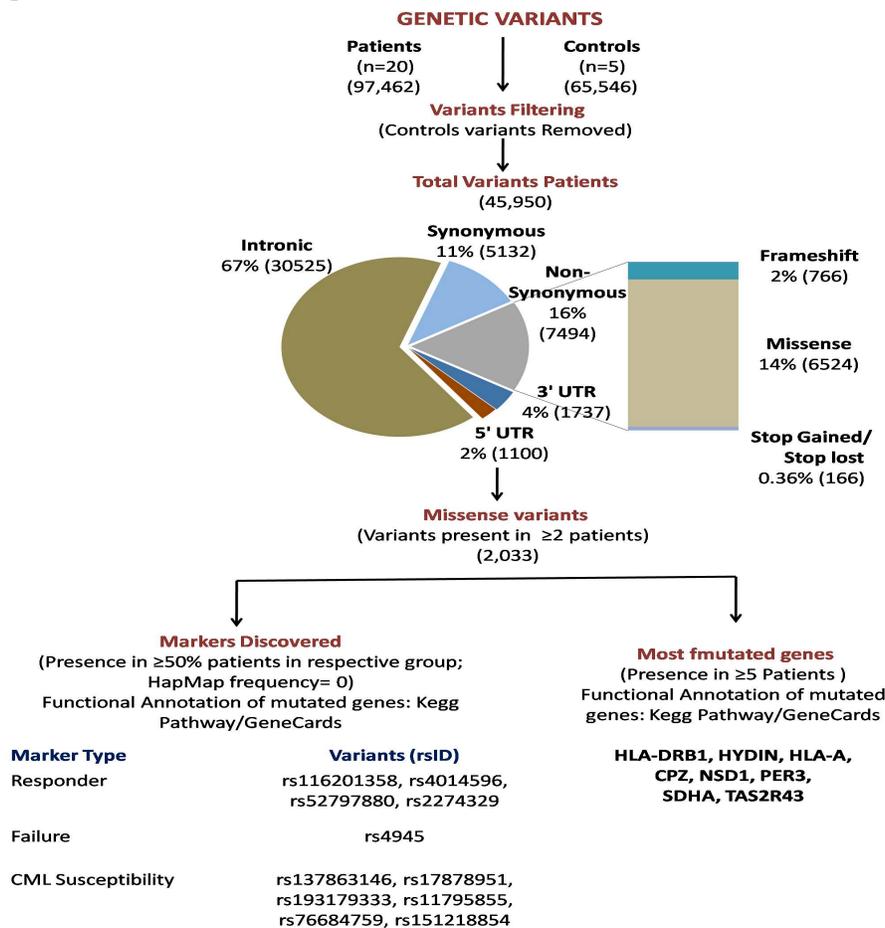


Figure 1: Work flow of clinical exome data interpretation to define potential prognostic and susceptibility markers in CML cohort.

Prognostic marker variants: We discovered rs116201358 (c.107C>A) in C8A, rs4014596 (c.149G>A) in UNC93B1, rs52797880 (c.422T>C) in APOH and rs2274329 (c.209G>C) in CA6 as markers for good response as mutations were present in responding patients only.

Amongst these mutations, rs116201358 (c.107C>A) and rs2274329 (c.209G>C) are expected to be most damaging (PolyPhan score: 0.994 and 1 respectively), while rs4014596 (c.149G>A) was moderately damaging (Table 2).

In the treatment failure group, rs4945 (c.7C>A) in MFGE8 was discovered. The mutation lies in conserved region and expected to be moderately damaging (SIFT: 0.00, PolyPhan score: 0.65) (Table 2).

HLA-DRB1 and HYDIN had 5 variants each, whereas HLA-A had 4 variants (Supplementary Table 2). These genes demonstrated maximum number of variants among all the genes. Apart from these candidate genes, CPZ, NSD1, PER3, SDHA and TAS2R43 had 3 mutations each (Supplementary Table 2). All the remaining genes had either 1 or 2 variation.

Table 2: Damaging potential of candidate variants related to CML prognosis and susceptibility.

| Mutation | Gene | Allele | MAF [▲] (dbsnp) | Mutation scoring | | | | Patients (n [#]) | | |
|----------------------------------|----------|--------|--------------------------|-------------------|----------|----------|-------------------|----------------------------|----------------|--------------|
| | | | | SIFT [€] | PolyPhan | Grantham | GERP [¶] | Responders (n=10) | Failures (n=8) | Total (n=18) |
| Markers for IM Responder | | | | | | | | | | |
| c.107C>A (rs116201358) | C8A | A/C | 0.021 | 0.00 | 0.994 | 107 | 5.09 | 5 | 0 | 5 |
| c.1494G>A (rs4014596) | UNC93B1 | C/T | 0.048 | 0.00 | 0.744 | 21 | 2.8 | 5 | 0 | 5 |
| c.422T>C (rs52797880) | APOH | A/G | 0.066 | 0.00 | 0 | 89 | 3.37 | 5 | 0 | 5 |
| c.209G>C (rs2274329) | CA6 | C/G | 0.0004 | 0.00 | 1 | 60 | 3.39 | 5 | 0 | 5 |
| Marker for IM Failures | | | | | | | | | | |
| c.7C>A (rs4945) | MFGE8 | G/T | 0.435 | 0.00 | 0.065 | 110 | -0.479 | 0 | 4 | 4 |
| CML Susceptibility Marker | | | | | | | | | | |
| c.181T>C (rs137863146) | HLA-DRB5 | G/G | 0.011 | 0.00 | 0.021 | 83 | 1.02 | 7 | 5 | 12 |
| c.200T>C (rs17878951) | HLA-DRB1 | A/G | NA | 0.00 | 0 | 64 | -0.433 | 9 | 6 | 15 |
| c.1070A>T (rs193179333) | RHPN2 | A/T | 0.031 | 0.00 | 0.989 | 99 | 4.61 | 6 | 4 | 10 |
| c.798G>C (rs117958555) | CYP2F1 | C/G | 0.008 | 0.00 | 0.994 | 24 | -4.65 | 7 | 2 | 9 |
| c.785T>G (rs76684759) | KCNJ12 | G/T | 0.197 | 0.00 | 0.999 | 142 | 5.43 | 7 | 5 | 12 |
| c.1007A>C (rs151218854) | FUT3 | G/T | 0.002 | 0.00 | 0 | 126 | -4.57 | 6 | 5 | 11 |

[▲]MAF = Minor allele frequency of variant according to dbsnp (NCBI), [€]SIFT = Sorting Intolerant From Tolerant, [¶]GERP = Genomic Evolutionary Rate Profiling, [#]n = total number of patients in respective

Susceptibility marker: HLA family genes HLA-DRB1 had rs17878951 (c.200T>C) and HLA-DRB5 had rs137863146. Other mutations observed in our cohort as CML susceptibility markers were rs193179333 in RHPN2, rs117958555 in CYP2F1, rs76684759 in KCNJ12 and rs151218854 in FUT3. Among these mutations, rs193179333, rs117958555 and rs76684759 were most deleterious as per SIFT, PolyPhan, Grantham and GERP scoring (Table 2).

Supplementary Table 2: Summary of most mutated genes (in ≥ 5 CML patients) revealing effected pathway and functions in CML.

| Genes | Mutations (rsID [▲]) | No. of Patients (n [€] =18) | Associated Pathways/ Functions |
|----------|--------------------------------|--------------------------------------|---|
| HLA-DRB1 | rs17885382 | 6 | hsa04514:Cell Adhesion Molecules(CAMs); hsa04612:Antigen Processing and Presentation; hsa04640:Hematopoietic Cell Lineage; hsa04672:Intestinal Immune Network For IgA Production; hsa04940:Type I Diabetes Mellitus; hsa05140:Leishmaniasis; hsa05310:Asthma; hsa05320:AutoimmuneThyroid Disease; hsa05322:Systemic Lupus Erythematosus; hsa05330:Allograft Rejection; hsa05332:Graft-versus-host Disease; hsa05416:Viral Myocarditis (Kegg Pathways) |
| | rs17878951 | 15 | |
| | rs9269951 | 8 | |
| | rs11554462 | 8 | |
| | c.239C>G | 8 | |
| HYDIN | rs147394921 | 8 | Involved in cilia motility. Mutations in this gene cause of autosomal recessive primary ciliary dyskinesia-5. (GeneCards) |
| | rs201356436 | 7 | |
| | rs202145987 | 5 | |
| | rs117922041 | 6 | |
| | rs148896798 | 7 | |
| HLA-A | rs1059455 | 6 | hsa04144:Endocytosis; hsa04514:Cell Adhesion Molecules (CAMs); hsa04612:Antigen Processing And Presentation; hsa04650:Natural Killer Cell Mediated Cytotoxicity; hsa04940:Type I Diabetes Mellitus; hsa05320:Autoimmune Thyroid Disease; hsa05330:Allograft Rejection; hsa05332: Graft-versus-host Disease; hsa05416:Viral Myocarditis (Kegg Pathways) |
| | rs199474430 | 8 | |
| | rs1136949 | 8 | |
| | rs1137160 | 8 | |
| CPZ | rs79736750 | 6 | Encodes Carboxypeptidase Z., has carboxypeptidase activity towards substrates with basic C-terminal residues (GeneCards). |
| | rs35993494 | 6 | |
| | rs34964084 | 5 | |
| NSD1 | rs35848863 | 7 | hsa00310:Lysine Degradation (Kegg Pathways) |
| | rs34165241 | 7 | |
| | 78247455 | 5 | |
| PER3 | rs10462020 | 7 | hsa04710:Circadian Rhythm-Mammal (Kegg Pathways) |
| | rs10462021 | 7 | |
| | rs2640909 | 10 | |
| SDHA | rs1042052 | 5 | hsa00020:Citrate Cycle (TCACycle); hsa00190:Oxidative Phosphorylation; hsa01100:Metabolic Pathways; hsa05010:Alzheimer's Disease; hsa05012: Parkinson's Disease; hsa05016: Huntington's Disease (Kegg Pathways) |
| | rs10426960 | 6 | |
| | rs10426962 | 6 | |
| TAS2R43 | rs111846092 | 10 | hsa04742:Taste Transduction (Kegg Pathways) |
| | rs200533679 | 7 | |
| | rs201618803 | 9 | |

[▲]rsID = rsiD provided in DBSNP (NCBI), [€]n = total number of patients enrolled

Table 3: Functional annotation of genes having potential markers using Kegg pathways and Gene Cards.

| Gene | Patients [▲] | Associated Pathways/Functions |
|--|---|--|
| Genes of Responder Markers | | |
| C8A | P2, P3, P4, P5, P8 | hsa04610:Complement And Coagulation Cascades; hsa05020:Prion Diseases; hsa05322:Systemic Lupus Erythematosus (Kegg Pathways) |
| UNC93B1 | P2, P3, P4, P5, P7 | Regulates toll-like receptor Signaling. Protein deficiency associated with herpes simplex encephalitis. (GeneCard) |
| APOH | P4, P5, P6, P8, P10 | Associated with lipoprotein metabolism, coagulation, and the production of antiphospholipid autoantibodies. (GeneCard) |
| CA6 | P3, P4, P5, P7, 10 | hsa00910:Nitrogen Metabolism. (Kegg Pathways) |
| Gene of Failure Marker | | |
| MFGE8 | P11, P13, P16, P18 | Encodes Lectadherin, involved in phagocytosis of apoptotic cells. Implicated in healing, autoimmune disease and cancer. (GeneCard) |
| Genes of Susceptibility Markers | | |
| HLA-DRB5 | hsa04514: Cell Adhesion Molecules (CAMs); hsa04612:Antigen Processing And Presentation; hsa04640:Hematopoietic Cell Lineage; hsa04672:Intestinal Immune Network For IgA Production; hsa04940:Type I Diabetes Mellitus; hsa05140:Leishmaniasis; hsa05310:Asthma; hsa05320:Autoimmune Thyroid Disease; hsa05322:Systemic Lupus Erythematosus; hsa05330:Allograft Rejection; hsa05332:Graft-versus-host Disease; hsa05416:Viral Myocarditis. (Kegg Pathways) | |
| HLA-DRB1 | hsa04514:CellAdhesionMolecules(CAMs); hsa04612:Antigen Processing And Presentation; hsa04640:Hematopoietic Cell Lineage; hsa04672:Intestinal Immune Network For IgA Production; hsa04940:Type I Diabetes Mellitus; hsa05140:Leishmaniasis; hsa05310:Asthma; hsa05320:Autoimmune Thyroid Disease; hsa05322:Systemic Lupus Erythematosus; hsa05330:Allograft Rejection; hsa05332:Graft-versus-host Disease; hsa05416:Viral Myocarditis (Kegg Pathways) | |
| RHPN2 | Encodes Rho-GTPase binding protein. Involved in actin cytoskeleton. (GeneCard). | |
| CYP2F1 | hsa00980:Metabolism of Xenobiotics By Cytochrome P450 (Kegg Pathways) | |
| KCNJ12 | Encodes K ⁺ channels. (GeneCard) | |
| FUT3 | hsa00601:Glycosphingolipid Biosynthesis Lacto And Neolacto Series; hsa01100:Metabolic Pathways (Kegg Pathways) | |

▲ Patients= patients having mutation of respective gene

DISCUSSION

The discovery of Imatinib Mesylate in drastically changed the scenario of CML treatment. Despite the paradigm shift in prognosis, around 20% of patients fail to achieve expected remission (6). Increasing incidence of CML will soon convert it into a serious medical problem. Therefore, discovery of more prognostic markers which will help in proper outcome estimation and planning for personalized CML treatment is required. Susceptibility markers are also needed to control increasing incidence of CML. The susceptibility markers identified have been reported for the first

time in the present study. Therefore, these findings are required to be validated in larger cohort of CML patients. Once validated, these markers can be tested in high risk population, which is more prone to develop CML (10,11). In search of the potential prognostic and susceptibility markers, we performed NGS in 18 CML patients, classified as responders and failures. In the responder group, recommended time line for analysis (3 months and 6 months) was customized in case of 2 patients to detect BCR-ABL% for research purpose. Change in the time line of BCR-ABL analysis of patients for research purpose was one of the limitations of the study.

The key objective of the study was to find missense variants, capable of predicting the outcome of patients treated with Imatinib Mesylate as frontline therapy and potential markers to detect healthy individuals prone to have CML. As per the criterion for prognostic markers, we demonstrated mutations in 5 genes as markers. Out of these 5 variations observed, mutation in MFG-E8 was in the failure group, whereas 4 mutations, discovered in the responders, were in genes C8A, UNC93B1, APOH and CA6.

Prognostic markers: MFG-E8 (Milk Fat Globule–Epidermal growth factor-8), is primarily responsible for production of membrane glycoprotein lactadherin, which promotes phagocytosis of apoptotic cells. In melanoma cells, MEG-E8 promotes progression by triggering epithelial to mesenchymal transition, stimulated invasion and immune suppression, whereas the high expression is linked with disease progression in oral cancer, breast cancer, prostate cancer and colon cancer (12-16). As this gene was found to be involved in the progression of many cancers, rs4945, discovered in our study was also expected to be associated with IM treatment failure in CML.

Complement component 8-alpha (C8A), encodes alpha subunit of C8. C8A participates in the formation of membrane attack complex (MAC). Elevated RNA levels of C8A gene were associated with better prognosis of Hepatocellular carcinoma (17) and in breast cancer patients, treated with trastuzumab (18). We found a deleterious mutation, rs116201358 in the responders, which seems to be related with better treatment outcome similar to elevated expression in earlier studies.

UNC93B1 (Unc-93 homolog B1) is involved in regulation of toll like receptor signaling. The encoded protein traffics nucleotide sensing receptors to endolysosomes from endoplasmic receptor. It is an IFN1 signature gene, induced by chemotherapy cyclophosphamide. IFN1 has recently been found involved in inducing anti-tumor immunity (19). Impaired UNC93B1 dependent immunity is also associated herpes simplex virus 1 encephalitis (HSE) pathogenesis

in children (20). We observed that the presence of mutation rs4014596 in the majority of the responder patients associates it with better prognosis, due to potential role in anti-tumor immunity.

APOH (Apolipoprotein H) is linked with several pathways including lipoprotein metabolism, coagulation and the production of antiphospholipid antibodies. Its elevated serum levels were found to be associated with incidence of colorectal cancer (21), better prognosis in Estrogen Receptor negative (ER-ve) breast cancer (22) and acute myeloid leukemia (23). We observed mutation rs52797880 in IM the responding cohort, which corroborated with the previous findings correlating elevated serum levels with better prognosis in breast cancer and AML.

Carbonic Anhydrase 6 (CA6) is a protein involved in reversible hydration of carbon-dioxide in saliva. Strong expression of this gene is associated with serous adenocarcinomas of ovary and endometrium (24,25), in carcinomas of breast, pancreas, urothelium and cancers of renal and pulmonary origin (26). In our cohort, 60% of the responding patients had mutation rs2274329, which makes it a potential candidate as marker for better prognosis

Co-occurrence of different genetic variants in an individual or a group of patients is considered as haplotype marker (27). In this study, we observed the co-existence of rs116201358 with rs4014596 and rs4014596 with rs2274329 in 4 and 3 IM responding patients respectively (Table 3). This observation suggests that coexistence of mentioned variants can be a potential haplotypes for good prognosis in CML patients.

Susceptibility markers: In our search for CML susceptibility markers, we observed rs17878951 and rs137863146 in HLA-DRB1 (Major Histocompatibility Complex Class II, DR Beta1) and HLA-DRB5 in our cohort. Both of the genes are participants of Interferon-gamma and CXCR4 signaling and play a key role in antigen presentation. Variants of HLA-DRB1 were found to be associated with incidence of CML in different populations throughout the world (28-32), and overexpression of HLA-DRB5 is a

potential susceptibility marker of Lung cancer (33), whereas the genetic variants are markers for CLL (34) and breast cancer (35). Another variant (rs193179333) of RHPN2 (Rhopilin Rho GTPase Binding Protein 2) gene, which codes for Ras Homologous (Rho)-GTPase binding protein, was observed in 10 CML patients of our cohort. The RHPN2 binds to GDP and GTP bound RhoA and RhoB, involved in organization of actin cytoskeleton. Mutations of RHPN2 are linked with suspicion of colorectal cancer incidence (36), prognosis of lung adenocarcinoma and malignant glioma (37,38). The polymorphism of CYP2F1 (Cytochrome P450 family 2 subfamily F member 1) gene is linked to lung cancer in different ethnicities (39), whereas differential expression are associated with prognosis of ovarian and breast cancer (40-42). In this study, rs117958555 was found in the majority of CML patients, which hints at being a susceptibility marker. In normal circumstances, it is involved in drug metabolism, hence genetic change in the gene can be possibly related to pathogenesis of CML. However, FUT3 (Fucosyltransferase 3) is involved in embryogenesis, tissue differentiation, tumor metastasis and inflammation. Incidences of gastro-intestinal and breast cancer are linked to FUT3 genetic variations (43,44). Similarly, we also discovered rs151218854 in FUT3 related to incidence of CML. We also observed another mutation rs76684759 in KCNJ12 (Potassium Voltage-Gated Channel Subfamily J Member 12), Normal KCNJ12 gene is involved in cardiac inward rectifier current (IK1). Its association with any type of cancer is not reported to the best of our knowledge, but we observed rs76684759 in more than 70% of our patients. The mutation is present in conserved region (SIFT Score: 0) and highly damaging (PolyPhen : 0.999).

In conclusion, we discovered 11 missense mutations, as potential prognostic and susceptibility markers. Variants in C8A, UNC93B1, APOH and C6A genes were found in IM responders whereas variant of MFGE8 was present in IM failures. Susceptibility markers were discovered in HLA-DRB1, HLA-DRB5, RHPN2, CYP2F1, KCNJ12 and FUT3. All the

genes were associated with cancer. The study is a step forward to discovering reliable markers for selecting individuals prone to CML and estimate the outcome of treatment. The findings of the study can be useful in individualizing cancer treatment after validation of discovered mutations in larger cohort at genetic and functional level.

ACKNOWLEDGEMENT

We are heartily thankful to our colleagues Mirza Adil Beg for assistance in DNA extraction of subjects. We acknowledge the assistance of Parul Thakar Chugh for her valuable inputs in statistical analysis.

REFERENCES

1. Huang X, Cortes J, Kantarjian H. Estimation of increasing prevalence and plateau prevalence of Chronic Myeloid Leukemia in the era of Tyrosine Kinase Inhibitors therapy. *Cancer* 2012; 118:3123-27.
2. Prasad RR, Singh P. Report of chronic myeloid leukemia from Indira Gandhi Institute of Medical Sciences, Regional Cancer Center, 2002-2009. *Indian J Med Paediatr Oncol.* 2013; 34:172-4.
3. Leitner AA, Hochhaus A, Müller MC. Current treatment concepts of CML. *Curr Cancer Drug Targets.* 2011; 11:31-43.
4. Saglio G, Kim DW, Issaragrisil S, *et al.* Nilotinib versus imatinib for newly diagnosed chronic myeloid leukemia. *N Engl J Med.* 2010; 362:2251-19.
5. Kalmanti L, Saussele S, Lauseker M, *et al.* Safety and efficacy of Imatinib in CML over a period of 10 years: Data from the randomized CML-study IV. *Leukemia.* 2015; 29:1123-32.
6. Quintás-Cardama A, Kantarjian HM, Cortes JE. Mechanisms of primary and secondary resistance to Imatinib in chronic myeloid leukemia. *Cancer Control* 2009; 16:122-31.
7. Baccarani M, Deininger MW, Rosti G, *et al.* European LeukemiaNet recommendations for the management of chronic myeloid leukemia: 2013. *Blood* 2013; 122:872-84.

8. Mahon FX. Discontinuation of tyrosine kinase therapy in CML. *Ann Hematol.* 2015; 94:S187–93.
9. Takahashi N, Kyo T, Maeda Y, *et al.* Discontinuation of imatinib in Japanese patients with chronic myeloid leukemia. *Haematologica* 2012; 97:903–06.
10. Rodriguez-Abreu D, Bordoni A, Zucca E. Epidemiology of hematological malignancies. *Ann Oncol.* 2007; 18 (Suppl 1):i3-i8.
11. Ugai T, Matsuo K, Sawada N, *et al.* Smoking and subsequent risk of leukemia in Japan: The Japan Public Health Center-based Prospective Study. *J Epidemiol.* 2017; 27(7):305-10.
12. Jinushi M, Nakazaki Y, Carrasco DR, *et al.* Milk Fat Globule EGF-8 Promotes Melanoma Progression through Coordinated Akt and Twist Signaling in the Tumor Microenvironment. *Cancer Res.* 2008; 68:8889-98.
13. Yamazaki M, Maruyama S, Abé T, *et al.* MFG-E8 expression for progression of oral squamous cell carcinoma and for self-clearance of apoptotic cells. *Lab Invest.* 2014; 94:1260–72.
14. Carrascosa C, Obula RG, Missiaglia E, *et al.* MFG8/lactadherin regulates cyclins D1/D3 expression and enhances the tumorigenic potential of mammary epithelial cells. *Oncogene.* 2012; 31:1521-32.
15. Soki FN, Koh AJ, Jones JD, *et al.* Polarization of prostate cancer-associated macrophages is induced by milk fat globule-EGF factor 8 (MFG-E8)-mediated efferocytosis. *J Biol Chem.* 2014; 289:24560-72.
16. Kusunoki R, Ishihara S, Tada Y, *et al.* Role of milk fat globule epidermal growth factor 8 in colonic inflammation and carcinogenesis. *J Gastroenterol.* 2015; 50:862-75.
17. Awan FM, Naz A, Obaid A, *et al.* Identification of Circulating Biomarker Candidates for Hepatocellular Carcinoma (HCC): An Integrated Prioritization Approach. *PLoS ONE* 2015; 10: e0138913
18. Willis S, Young BM, Williams C, Leyland-Jones B. C8A, a potential predictive marker of trastuzumab benefit, is associated with immunological signatures in MSigDB. *Annals of Oncology.* 2015; 26 (suppl3):1524.
19. Moschella F, Torelli GF, Valentini M, *et al.* Cyclophosphamide induces a type I interferon-associated sterile inflammatory response signature in cancer patients' blood cells: implications for cancer chemoimmunotherapy. *Clin Cancer Res.* 2013; 19:4249-61.
20. Lafaille FG, Pessach IM, Zhang SY, *et al.* Impaired intrinsic immunity to HSV-1 in human iPSC-derived TLR3-deficient CNS cells. *Nature.* 2012; 491:769-73.
21. Ma Y, Zhang P, Wang F, Liu W, Yang J, Qin H. An integrated proteomics and metabolomics approach for defining oncofetal biomarkers in the colorectal cancer. *Ann Surg.* 2012 ; 255:720-30.
22. Chung L, Moore K, Phillips L, Boyle FM, Marsh DJ, Baxter RC. Novel serum protein biomarker panel revealed by mass spectrometry and its prognostic value in breast cancer. *Breast Cancer Res.* 2014; 16:R63.
23. Lee SW, Kim IJ, Jeong BY, *et al.* Use of MDLCDIGE and LCMS/MS to identify serum biomarkers for complete remission in patients with acute myeloid leukemia. *Electrophoresis* 2012 Jul; 33:1863-72
24. Smith NL, Halliday BE, Finley JL, Wennerberg AE. Immuno histochemical distribution of tumor associated antigen CA6 in gynecological neoplasms as detected by monoclonal antibody DS6. *Int J Gynecol Pathol.* 2001; 20:260-6.
25. Kearse KP, Smith NL, Semer DA, *et al.* Monoclonal antibody DS6 detects a tumor-associated sialoglycotope expressed on human serous ovarian carcinomas. *Int J Cancer.* 2000; 88:866-72.
26. Smith NL, Halliday BE, Finley JL, Wennerberg AE. The spectrum of immunohistochemical reactivity of monoclonal antibody DS6 in non-gynecologic neoplasms. *Appl Immunohistochem Mol Morphol.* 2002; 10:1528.

27. The International HapMap Consortium. A haplotype map of human genome. *Nature* 2005; 437:1299-20.
28. Barion LA, Tsuneto LT, Testa GV, *et al.* Association between HLA and leukemia in a mixed Brazilian population. *Rev Assoc Med Bras.(1992)*. 2007; 53:252-56.
29. Yasukawa M, Ohminami H, Kojima K, *et al.* Analysis of HLA-DRB1 alleles in Japanese patients with chronic myelogenous leukemia. *Am J Hematol*. 2000; 63:99-01.
30. Naugler C, Lewisky R. HLA risk markers for chronic myelogenous leukemia in Eastern Canada. *Leuk Lymphoma*. 2009; 50:254-59.
31. Rivera-Pirela SE, EcheverríaM, Salcedo P, *et al.* HLA DRB1*, DQB1*, DPA1*, and DPB1* and their association with the pathogenesis of leukemia in the population of Venezuela. *Rev Alerg Mex*. 2016; 63:237-51.
32. Wang XJ, Zhang YZ, Sun HY , Li QH , Ru K. Associations of HLA gene with leukemia in 1186 cases. *Zhongguo Shi Yan Xue Ye Xue Za Zhi*. 2014; 22:263-68.
33. Li ZH, Wang YZ, Liu J *et al.* DNA damage associated genetic variants contribute to lung cancer susceptibility in a Han Chinese population. *Zhonghua Yu Fang Yi Xue Za Zhi*. 2016; 50:721-27.
34. Hojjat-Farsangi M, Jeddi-Tehrani M, Amirzargar AA, *et al.* Human leukocyte antigen class II allele association to disease progression in Iranian patients with chronic lymphocytic leukemia. *Hum Immunol*. 2008; 69:666-74.
35. Yang XX, Pan HZ, Li PY, *et al.* HLA class II variants in Chinese breast cancer patients. *Asian Pac J Cancer Prev*. 2011; 12:3075-79.
36. Carvajal-Carmona LG, Cazier JB, Jones AM, *et al.* Fine-mapping of colorectal cancer susceptibility loci at 8q23.3, 16q22.1 and 19q13.11: refinement of association signals and use of in silico analysis to suggest functional variation and unexpected candidate target genes. *Hum Mol Genetics*. 2011; 20:2879–88.
37. Wu K, Zhang X, Li F, *et al.* Frequent alterations in cytoskeleton remodeling genes in primary and metastatic lung adenocarcinomas. *Nat Commun*. 2015; 6:101-31.
38. Danussi C, Akavia UD, Niola F, *et al.* RHPN2 Drives Mesenchymal Transformation in Malignant Glioma by Triggering RhoA Activation. *Cancer Res*. 2013; 73:5140-50.
39. Tournel G, Cauffiez C, Leclerc J, *et al.* CYP2F1 genetic polymorphism: identification of interethnic variations. *Xenobiotica*. 2007; 37:1433-38.
40. Kumarakulasingham M, Rooney PH, Dundas SR, *et al.* Cytochrome p450 profile of colorectal cancer: identification of markers of prognosis. *Clin Cancer Res*. 2005; 11:3758.
41. Downie D, McFadyen MC, Rooney PH, *et al.* Profiling Cytochrome P450 Expression in Ovarian Cancer: Identification of Prognostic Markers. *Clin Cancer Res*. 2005; 11:7369–75.
42. Calaf GM, Roy D. Human drug metabolism genes in parathion and estrogen treated breast cells. *Int J Mol Med*. 2007; 20:875-81.
43. Duell EJ, Bonet C, Muñoz X, *et al.* Variation at ABO histo-blood group and FUT loci and diffuse and intestinal gastric cancer risk in a European population. *Int J Cancer*. 2015; 136: 880-93.
44. doNascimento JC, Ferreira Sde A, Vasconcelos JL, *et al.* Fut3 role in breast invasive ductal carcinoma: Investigating its gene promoter and protein expression. *Exp Mol Pathol*. 2015; 99:409-15.