Article

Integrated Genomic Analysis Identifies ANKRD36 Gene as a Novel and Common Biomarker of Disease Progression in Chronic Myeloid Leukemia

Zafar Iqbal [1,2,3], Muhammad Absar[2], Tanveer Akhtar[2], Aamer Aleem[4], Abid Jameel[5], Sulman Basit [6], Anhar Ullah[7], Sibtain Afzal[8], Khushnooda Ramzan[9], Mahmood Rasool[10], Sajjad Karim[10], Zeenat Mirza[10], Mudassar Iqbal[11], Maryam AlMajed[1], Buthinah AlShehab[1], Sarah AlMukhaylid[1], Nouf AlMutairi[1], Nawaf Al-anazi[1, 12], Muhammad Farooq Sabar[13], Muhammad Arshad[14], Muhammad Asif[15], Masood Shammas[16] and Amer Mehmood[15]

[1]: Clinical Laboratory Sciences, College of Applied Medical Sciences (CoAMS-A), King Saud Bin Abdulaziz University for Health Sciences, King Abdulaziz Medical City/Kind Abdullah International Medical Research Centre (KAIMRC)/Saudi Society for Blood and Marrow Transplantation (SSBMT), King Abdulaziz Medical City, National Guard Health Affairs, Al-Ahsa, Saudi Arabia.
[2]:Hematology, Oncology and Pharmaco-genetic Engineering Sciences (HOPES) Group, Health Sciences Laboratories (HaSiL), Department of Zoology, University of the Punjab (ZPU), Lahore 54590, Pakistan.
[3]: Next-Generation Medical Biotechnology & Genomic Medicine Division, Department of Biotechnology, Qarshi University, Lahore, Pakistan.
[4]Division of Hematology/Oncology, Department of Medicine, KKUH and College of Medicine, King Saud University, Riyadh, Saudi Arabia.
[5] Khyber Teaching Hospital and Hayatabad Medical Complex, Peshawar, Pakistan.
[6]Taibah University Madinah-Center for Genetics and, Inherited Diseases Center for Genetics and Inherited Diseases,, Taibah University Madinah, 30001, Saudi Arabia, Almadinah Almunawarah 30001, Saudi Arabia.
[7] National Heart and Lung Institute, Imperial College London, UK.
[8]Biomedical Research Laboratory, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia.
[9]Research Centre, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia.
[10] Centre of Excellence in Genomic Medicine Research (CEGMR) & King Fahd Medical Research Center, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.
CEGMR, King Abdulaziz University, Jeddah, Saudi Arabia.
[11]Asian Medical Institute, Kant, Kyrgyzstan.
[12]: Division of Hematology/Oncology, Department of Pediatrics, King Abdulaziz Hospital, Al-Ahsa, Saudi Arabia.
[13]Centre for Applied Molecular Biology (CAMB), University of the Punjab, Lahore, Pakistan.
[14]:Department of Biological Sciences, International Islamic University, Islamabad, Pakistan.
[15]Research Operations and Development, BUITEMS, Quetta, Pakistan.
[16]Department of Medical Oncology, Harvard (Dana Farber) Cancer Institute, Boston, MA 02115, Harvard University, Research Health Scientist, VABHCS, Harvard Medical School at VAMC, West Roxbury, MA 02132.
[17]Cancer Stem Cell Unit, Department of Anatomy, King Saud University, Riyadh, Saudi Arabia.
Name: Dr Zafar Iqbal
Address: Clinical Laboratory Sciences, College of Applied Medical Sciences (CoAMS-A), King Saud Bin Abdulaziz University for Health Sciences, King Abdulaziz Medical City/KAIMRC, National Guard Health Affairs, Al-Ahsa, Saudi Arabia
Phone numbers: +966-53-86-355-86
Facsimile numbers: NA
E-mail address: drzafar.medgen@yahoo.com, iqbalz@ksau-hs.edu.sa

Simple Summary: Chronic Myeloid Leukemia is a type of blood cancer that is regarded as a success story in finding out exact biological origin, pathogenesis, and development of a molecularly targeted (mutation-specific) therapy, that has led successful treatment of this fatal cancer. It is resulted due to BCR-ABL oncogene due to translocation between chromosomes 9 and 22. Anti—BCR-ABL drugs, known as tyrosine kinase inhibitors (TKIs) have led longterm remissions in more than 80% of CML patients and even cure in about one third patients. Nevertheless, many patients face drug resistance and disease progression in about 30% of CML patients, leading to morbidities and mortality. Unfortunately, no biomarkers of CML progression are available due to poor understanding of mechanism of progression. Therefore, finding reliable molecular biomarkers of CML progression is one of the most attractive research area in 21st century cancer research. In this study, we report

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 August 2021 doi:10.20944/preprints202108.0498.v1

novel genomic variants exclusively found in all our advanced phase CML patients. This study will help in identifying CML patients at risk of disease progression and timely therapeutic interventions to avoid or at least delay fatal disease progression in this cancer.

Abstract: Background: Chronic Myeloid Leukemia (CML) is initiated in bone marrow due to chromosomal translocation t(22;9) leading to fusion oncogene BCR-ABL. Targeting BCR-ABL by tyrosine kinase inhibitors (TKI) have changed fatal CML into an almost curable disease. Despite that, TKIs lose their effectiveness due to disease progression. Unfortunately, mechanism of CML progression is poorly understood and common biomarkers for CML progression are unavailable. This study was conducted to find out novel biomarkers of CML progression by employing whole exome sequencing (WES).

Materials and Methods: WES of accelerated phase (AP-) and blast crisis (BC-) CML patients was carried out, with chronic phase CML (CP-CML) patients as control. After DNA library preparation and exome enrichment, clustering and sequencing was carried out using Illumina platforms. Statistical analysis was carried out using [SAS/STAT] software version 9.4 and R package employed to find mutations shared exclusively by all AP-/BC-CML. Confirmation of mutations was carried out using Sanger sequencing and protein structure modelling using I-Tasser followed by mutant generation and visualization using PyMOL.

Results: Three novel genes (ANKRD36, ANKRD36B and PRSS3) were mutated exclusively in all AP-/BC-CML patients. Only ANKRD36 gene mutations (c.1183_1184 delGC and c.1187_1185 dupTT) were confirmed by Sanger sequencing. Protein modelling studies showed that mutations induce structural changes in ANKRD36 protein.

Conclusions: Our studies show that ANKRD36 is a potential common biomarker and drug target of early CML progression. ANKRD36 is yet uncharacterized in human. It has the highest expression in bone marrow, specifically myeloid cells. We recommend carrying out further studies to explore the role of ANKRD36 in biology and progression of CML.

Keywords: CML; Disease progression; common biomarker; drug target; ANRD36.

1. Introduction

Chronic Myeloid Leukemia (CML) is a neoplasm of hematopoietic cells, which is characterized by a deregulated high production of immature granulocytes and their progenitors [1]. Since these cells are immature, they are not fully functional [2]. The excessive proliferation of progenitor cells and blasts results in a change in the balance between regeneration and differentiation [3]. Approximately 15% of all leukemias are CML, which means that 2 out of 100,000 individuals develop CML yearly. Out of those patients, 5-10% have exposure to excessive radiations [4].

CML was the first neoplasm to be linked to a chromosomal abnormality, and it is also one of the most intensely investigated malignancies [3]. CML is instigated by a reciprocal chromosomal translocation t(9:22) giving rise to Philadelphia chromosome [5]. The translocation ensues between ABL proto-oncogene on the long arm of chromosome 9 and breakpoint cluster region (BCR) on chromosome 22, giving rise to BCR-ABL fusion oncogene [6]. This fusion oncogene encodes a new oncoprotein called bcr-abl [7, 8]. The bcr-abl oncoprotein has enhanced tyrosine kinase activity that hinders apoptosis, alters cell cycles, and deregulates cell division, leading to leukemogenesis [9, 10]. In last two decades, tyrosine-kinase inhibitors (TKIs) have revolutionized CML treatment and recently overall survival of CML patients equals to general public due to introduction of TKIs [11].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 August 2021 doi:10.20944/preprints202108.0498.v1

There are three main disease phases of CML, which include chronic phase (CP), accelerated phase (AP), and blast-crisis phase (BC) [12]. Most of the CML patients are diagnosed in chronic phase and hence overall survival of CML is excellent [13]. Nevertheless, about 20% of CML patients progress to advanced phases of the disease that results in drug resistance, intolerance, morbidities, and mortality [14]. Unfortunately, mechanism of CML progression is poorly comprehended [15]. Moreover, universal biomarkers to early diagnose disease progression are not available. Discovery of common biomarkers for CML progression can help in early determination of CML patients at risk of progression and clinically manage these patients to avoid or delay disease progression [16, 17]. Thereby, this study was intended to determine common gene variants associated with CML progression using highthroughput DNA sequencing methods like whole-exome sequencing.

2. Materials and Methods.

2.1. Patient Inclusion and Exclusion Criteria

The study was carried out from January 2012 until Dec 2019. One hundred forty-one (141) CML patients were enrolled in the study from Hayatabad Medical Complex (HMC) Peshawar, Khyber Pakhtunkhawa (KP) Pakistan. Peripheral blood samples were collected from all CML patients along with clinical data. Out of 141 patients, 123 were CP-CML, 12 AP-CML and 6 BC-CML. AP- & BC-CML patients were experimental group while CP-CML patients were included as controls. Additionally, 10 age/gender matched healthy controls were included in the study.

Regarding treatment, imatinib mesylate (IM) was the first line of therapy for all patients. However, nilotinib (NI) was prescribed in case of IM resistance. The criteria of all responses were per European Leukemia Net guidelines 2013 [18].

2.2. Definitions of Clinical Phases of Chronic Myeloid Leukemia (CML) for Staging

Chronic phase (CP) was identified by the presence of three main parameters in the circulation, which are 15%-19% of basophils, less than 30% of blasts and promyelocytes, and less than 5% of blast cells. Moreover, evidence of blast cells in extramedullary sites were not available [19]. Accelerated phase (AP) was described by an increase of blasts up to 15%-29%, or 30% of promyelocytes in bone morrow or blood. Furthermore, ≥20% of basophils with constant low platelet counts of less than 100 × 109/L were detected, and chromosomal abnormalities in Philadelphia cells were discovered [20]. Blast crisis (BC) was defined by the presence of blasts equal or greater than 30% in bone morrow or blood. In BC, blasts were present in the spleen and in other extramedullary sites [18].

2.3. Criteria for Assessment of Treatment Response in Chronic Myeloid Leukemia

Patient blood count and physical examination were performed every 4-8 weeks to monitor treatment response. The listed response tools were applied to evaluate the effectiveness of CML medication in all patients [21, 22].

2.3.1 Complete Hematological Response (CHR):

CHR was defined as the absence of immature cells, normal platelet count of less than 450 × 10^9/L, and normal basophils count of less than 5%. Also, impalpable spleen was documented [18, 23].

2.3.2. Cytogenetic Response (CyR):

Cytogenetics and differential morphology or FISH bone marrow aspirates were evaluated for diagnosis every 6 months and 12 months, and then yearly. Cytogenetic testing results of Ph+; Complete cytogenetic response (CCyR): 0% or < less than 1% BCR–ABL nuclei by FISH/≥200 cells, partial (PCyR): 1-35% Ph+, minor cytogenetic response (mCyR):

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 August 2021 doi:10.20944/preprints202108.0498.v1

Ph+= 36-65%, minimal cytogenetic response (miCyR): Ph+= 66–95% and no cytogenetic response (nCyR) as Ph+= > 95% [18, 22, 23].

2.3.3. Criteria for Calculation of Molecular Response (MR):

Major molecular response (MMR) was described as a BCR-ABL/ABL ratio cut-off of ≤0.1%. Moreover, a ratio of ≤ 0.0032% was termed as $MR^{4.5}$. Responses seen in molecular and cytogenetic level were outlined as shown in (Figure 1) [18, 22, 23].

2.4. Criteria for Calculation of European LeukemiaNet (ELN) Responses & survival:

Following criteria was used for calculating of European LeukemiaNet (ELN) treatment responses and survivals:

Optimal response: at 3 months if Ph+ =≤35%, at 6 months if Ph+ =0, at 12 months if BCR-ABL1 by PCR was ≤0.1% then and at any time as BCR-ABL1 ≤0.1% [18].

Warning: It was defined at baseline as High risk or CCA/Ph+ major route, at 3 months Ph+ 36-95%, at 6 months Ph+ 1-35%, at 12 months as BCR-ABL1 by PCR = >0.1-1% [18].

Failure: It was defined at 3 months as Non-CHR and/or Ph+ >95%, at 6 months as Ph+ >35%, at 12 months as Ph+ >0 and then at any time as Loss of CHR, Loss of CCyR [18].

Overall Survival (OS): The overall survival was taken as the the beginning of the IM therapy to the patient expired date or last follow up [24].

Progression-Free Survival (PFS): PFS was measured from the day IM began till the development of CML to AP or BC or to death. Any patient who survived as per last day of study was censored at the last follow-up date. The confirmation of the survival status of patients who were absent from the last follow-up was conducted by contacting patients based on the registered contact information. The survival analysis was determined as per kaplan-Meier Method [25].

2.5. Criteria for Documenting Adverse Events

According to the standard terminologies (version 4.03), hematological undesirable effects were categorized [19].

2.6. Ethical Approval:

The protocols of this study were approved by King Abdullah International Medical Research Center (KAIMRC), King Saud bin Abdulaziz University for Health Sciences (KSAU-HS) Saudi Arabia, and Hayatabad Medical Complex (HMC) Peshawar Pakistan and University of the Punjab Lahore Pakistan. A written informed consent was obtained from every enrolled patient in this study. The study was carried out per regulations of the Declaration of Helsinki [26, 27].

2.7. Sample Collection & DNA Extraction:

10ml peripheral blood was were collected in EDTA tubes (BD Vacutainer Systems, Franklin Lakes, and N.J.). QIAamp DNA Blood Mini Kit (QIAGEN) was used to extract DNA from all patients [28]. DNA quantitation was performed by utilizing NanoDrop Spectrophotometer (NanoDrop Technologies, Inc., USA). After that, it was diluted into aliquots of 70-80 ng/µl for mutation detection by Whole Exome Sequencing (WES). The excess amount of DNA was diluted to 40 ng/µl for Sanger Sequencing. DNA was stored in refrigerator at -80°C [29].

2.8. Whole Exome Sequencing

In this study, the SureSelect^{XT} V6-Post Capture Exome kit (Agilent Technologies Inc, Santa Rosa, CA, USA) was utilized for the formulation of libraries and target enrichment. For exonic and intron flanking regions, exome enrichment was done by SureSelect^{XT2} Target Enrichment System for Illumina Paired-End Multiplexed Sequencing (Illumina, San Diego, CA, USA) based on the manufacturer protocol (Agilent Technologies Inc, Santa

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 August 2021 doi:10.20944/preprints202108.0498.v1

Rosa, CA, USA). DNA fragmentation and tagmentation were performed per manufacturer's protocols. Following that, purification and amplification of the DNA were conducted. Magnetic beads were used to purify the amplified DNA fragments. The whole exome was used to capture target regions. Subsequently, PCR amplified the enriched DNA fragments. To enumerate the augmented fragments, the Qubit fluorometer was operated on the enriched libraries. Moreover, using Agilent Bioanalyzer (Agilent Technologies Inc, Santa Rosa, CA, USA), the library size distribution was quantified. Last of all, for cluster generation and whole exome sequencing, the amplified DNA fragments were loaded on flow cell on an Illumina NextSeq500 instrument (Illumina, San Diego, CA, USA) [30].

2.9. Exome Sequencing Data Analysis

The WES output BCL records were transformed to FASTQ files with the aid of BCL2FASTQ software. The BWA-MEM algorithm aided in the alignment of the FASTQ records to the human genome (GRCh37/hg19), using the BWA aligner. Whole exome sequencing data statistics including sequencing depth and some summary statistics are presented in table 1 and 2. For variants analysis, the Genome analysis tool kit (GATK) was utilized. Illumina Variant Studio was used for genomic variants annotation and filtration [31]. The resulting annotated files on average have approximately 90,000 variants. This includes synonymous, coding, intergenic, intronic, splice-site as well as 5′ and 3′ UTR variants Variant statistics are presented in table 3.

2.10. Primary Analysis:

In order to identify a shared biomarker for CML growth, mutated genes were analyzed in all advanced phase CML patients. An Excel file presenting the WES was modified using the filtration strategy, which excluded all synonymous and intron variants while rare variants were called. Moreover, all recognized Tolerant (T) and Benign (B) Variants (with known prediction) were eliminated. For multiple B or T, we considered it B, if the frequency of B was ≥70%. On the other hand, it is thought out to be T if the frequency of T was ≥70% [32]. In summary, synonymous, intergenic and deep intronic variant were removed from the annotated file. Only those variants having high and intermediate protein effects as well as splice variants were retained. Moreover, variants with a population frequency of more than 0.005 in the dbSNP and ESP (Exome Sequencing Project) database were also removed. A total of approximately 124 on average rare variants were obtained as a result of this analysis. Finally, further data analysis was performed to find driver mutations in novel genes, i.e., mutations that are shared by all advanced phase CML patients, but absent in chronic phase CML or healthy controls. Hence, these variants might have a significant role in disease progression [16, 17]. Data generated from next generation has been submitted to NCBI and can be accessed through link https://www.ncbi.nlm.nih.gov/sra/PRJNA734750 (SRA accession number PRJNA734750).

2.11. Validation of Mutation by Sanger Sequencing:

In order to validate the WES detected variants, Sanger sequencing was carried out in all samples under investigation. The recovery of distinct Genomic primers of the variants in identified genes was obtained from the University of California Santa Cruz genome database browser (Table 3), and primers were obtained from Applied Biosystems, California, USA. Template DNA amplification was performed by PCR. DNA sequencing reactions were prepared using ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction kits (Applied Biosystems, California, USA) [33]. Then forward and reverse DNA templates were sequenced by Sanger sequencing using ABI Prism 3730 Genetic Analyzer (Applied Biosystems, California, USA) [34, 35].

2.12. Statistical Analysis of Patient Clinical Data

Based on the normality test, absolute numbers and percentages for categorical variables were demonstrated, as well as mean with an appropriate measure of variation for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 August 2021 doi:10.20944/preprints202108.0498.v1

continuous variables. For categorical data, Chi-Square or Fisher's exact tests were used to compare two groups while a two-sample independent test or Mann Whitney U test was used for the continuous data. ANOVA or Kruskal-Wallis test was used to analyze variance for groups of ≥3. To assess the survival outcome, Kaplan Meier survival analysis curves were carried out [25]. The group comparison were performed by log rank test. SAS/STAT software version 9.4 were used for data analysis (SAS institute Inc., Cary, NC, USA.). For statistical computing, the R foundation was operated (Vienna, Austria) [36]. The Eutos risk score, Euro risk score, and the Sokal risk score were measured [18, 37- 40].

2.12. Protein Modelling Studies

One of the most significant issues in computational structural biology is the prediction of 3-dimensional protein structures from amino acid sequences. The protein structure of ANKRD36, which is yet uncharacterized in human, was done using I-Tasser web-server[41] . It resulted in computational prediction of its structure and an assessment of these mutations [41]. Mutagenesis was specifically done on residues 395 and 396 using PyMOL Wizard. Further, the wild-type and mutated structures were superimposed using Schrodinger's PyMOL Molecular Graphics System,Version 2.5[42].

3. Results

A total of 141 CML patients were included in this study. Mean age of the patients was 34.6 (Table 1), and male to female ratio was 1.6:1. Gender statistics revealed that females were 60.2%, and males were 39.8%. The mean hemoglobin was 10.1, and the mean WBC count was 317.9. In addition, platelet count in CML patient was 400.2. Overall, females were more commonly affected by CML.

During course of study, 12.8% (n=18) of patients progressed to advanced phases (AP=2, BC=12). CP-, AP- and BC-CML patients had mean age of 33.5, 35.6, and 38.1 years, respectively. In addition, there was male dominance found in all the CML phases. Furthermore, the male to female ratio was calculated to be 2:1, 2:1, and 1.5:1 in BC, AP, and CP, accordingly. Moreover, anemia was common among two-thirds of the patients. Of all CML patients, 56% of them had a leukocyte count 50 × 109/L or higher (n=79). Imatinib was first-line TKI, and it was administered to 66.7%, 66.7%, and 58.36% of CP, AP & BC CML patients, respectively. The chemotherapy was given to 8.1%, 66.7%, and 75% of CP, AP & BC CML patients, accordingly. Overall, 12.7% of CML patients (n=18), developed to AP-CML (n=6), or they progressed to BC-CML (n=12) (table 2).

There was significant difference between chronic and advanced phase patients with respect to male to female ratio, hemoglobin level, WBC count, platelet count, type of treatment received, hepatomegaly, splenomegaly and survival status (Table 2).

Table 1. Comparisons between our finding and other studies.

Characteristics	Japan	Iraq [43]	US [48]	EU [47]	India [44]	Our Study
# of Patients	506	100	1106	210	90	141
Age, yrs Mean	56	51.7	41.1	55	38.6	36.4
Male	349	58%	59%	54%	57%	60.2%
Female	157	42%	41%	46%	42.2%	39.8%
Ratio: Male: Female	2.2:1	1.4:1	1.4:1	1:1	1.4:1	1.6:1
Hemoglobin (g/dL) Mean	4	12.28	10	12.6	9.41	10.1
WBC count (× 10^9/L) Mean		45.26	19	80.2	182	317.9
Platelets (× 10^9/L) Mean	47.2	341.5	77	373	328	400.2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 August 2021 doi:10.20944/preprints202108.0498.v1

Table 2. Comparison of demographics, clinical data and laboratory parameters between three phases of CML.

Characteristics	Patients Group			
	CP-CML, n (%)	AP-CML, n (%)	BC-CML, n (%)	P-VALUE
# of Patients	123 (87.2)	6 (4.3)	12 (8.5)	
Age, yrs				
Mean (range)	35.5 (9-67)	35.6 (27-43)	38.1 (29-50)	
Gender				
Male	74 (60.2)	4 (66.67)	8 (66.7)	P = 0.6004
Female	49 (39.8)	2 (33.33)	4 (33.3)	P = 0.5987
P-VALUE	P = 0.0272	P = 0.3980	P = 0.2933	
Ratio: Male: Female	1.5:1	2:1	2:1	
Hemoglobin (g/dL) Mean	10.1			
<12g/dl	69 (56.1)	5 (83.3)	9 (75)	P = 0.0642
>12g/dl	14 (11.4)	1 (16.7)	3 (25)	P = 0.2609
P-VALUE	P = 0.0024	P = 0.2154	P = 0.1380	
WBC count ($\times 10^9$/L) Mean	313.7	315	325	
<50	20 (16.3)	1 (20)	2 (16.7)	P = 0.8276
>/=50	64 (52)	5 (80)	10 (83.3)	P = 0.0184
P-VALUE	P = 0.0052	P = 0.2752	P = 0.0661	
Platelets ($\times 10^9$/L) Mean	400.2			
<450	75 (61)	4 (66.7)	10(83.3)	P = 0.2528
>/=450	33 (26.8)	2 (33.3)	2 (16.7)	P = 0.8722
P-VALUE	P = 0.0011	P = 0.4786	P = 0.0661	
Imatinib				
Yes	82 (66.7)	4 (66.7)	7 (58.3)	P = 0.7260
Nilotinib as 2nd Line				
Yes	41 (33.3)	4 (66.7)	8 (66.7)	P = 0.0065
Hydroxyurea				
Yes	82 (66.7)	3 (50)	10 (83.3)	P = 0.9967
Interferon				
Yes	41 (33.3)	0 (0)	0 (0)	P = 0.0038
Chemotherapy				
Yes	10 (8.1)	4 (66.7)	9 (75)	P < 0.0001
Splenomegaly				
<5cm	4 (3.3)	0 (0)	0 (0)	P = 0.4358
5-8cm	9 (7.3)	1 (16.7)	3 (25)	P = 0.0619
>8cm	70 (56.9)	5 (83.3)	9 (75)	P = 0.0732
No splenomegaly	40 (32.5)	0 (0)	0	P = 0.0044
Hepatomegaly				
Yes	35 (28.5)	4 (66.7)	8 (66.7)	P = 0.0014
Anemia				
Yes	97 (78.9)	5 (83.3)	9 (75)	P = 0.9807
Pregnant				
Yes	4 (8.2)	0 (0)	0 (0)	P = 0.2090
Survival Status				
Confirmed deaths	10 (8.1)	0 (0)	9 (75)	P = 0.0003
Alive at last follow up (overall survival)	113(91.9)	6 (100)	3(25)	P = 0.0003

Table legend: WES: Whole Exome Sequencing, WBC; White blood Cells, CP; Chronic Phase, AP; Accelerated Phase, BC; Blast Phase, CP-CML; Chronic Phase-Chronic Myeloid Leukemia, AP-CML; Accelerated Phase-Chronic Myeloid Leukemia, BC-CML; Blast Phase-Chronic Myeloid Leukemia.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 August 2021 doi:10.20944/preprints202108.0498.v1

Table 3: Statistics obtained before alignment of reads with the reference genome

Statistics	ID 1	ID 2	ID 3	ID 4	ID 5
Total number of reads	70,508,170	75,173,754	75,622,396	71,328,320	76,940,162
Q30 (%)	96.6	97.0	96.8	96.9	97.1
Average Read Length (bp)	101.0	101.0	101.0	101.0	101.0
Total Yield (Mbp)	7,121	7,592	7,637	7,204	7,770
Target region (bp)	60,456,963	60,456,963	60,456,963	60,456,963	60,456,963
Average depth (X)	117.7	125.5	126.3	119.11	128.5

Total yield = total number of reads x Average read length, Average depth is the throughput depth of the target regions (X) = Total yield/Target regions.

Table 4: Statistics obtained after alignment of reads with the reference genome

Statistics	ID 1	ID 2	ID 3	ID 4	ID 5
Initial Mappable Reads	70,471,133	75,143,023	75,592,332	71,300,726	76,912,816
%Non-Redundant Reads	88.1	86.0	86.9	86.3	87.1
%On target reads	75.2	77.9	77.7	78.0	77.7
Depth of target region (X)	69.1	74.4	75.5	70,9	76.9
Coverage (%>10X)	97.0	97.3	97.3	96.9	97.1
Coverage (%>30X)	82.1	84.0	84.6	82.9	84.2

Initial Mappable Reads; Number of mapped reads to human genome, % Non-Redundant Reads = 100 x Non-redundant reads/Initial mappable reads, % On-Target Reads = 100 x On-target reads/Non-redundant reads, On-Target Yield (bp) = The sum of the bases in the final alignment to the target regions, Mean Depth of Target Regions (X) = On-target yield/Target regions. Coverage statistics. The percentage of bases in target regions with a depth of coverage are mentioned.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 August 2021 doi:10.20944/preprints202108.0498.v1

Table 5: Different types of variants identified in each exome sequenced sample

Variant Type	ID 1	ID 2	ID 3	ID 4	TD 5
Number of SNPs	88,892	90,562	88,725	90,441	86,484
Synonymous variants	11,945	12,268	11,810	12,053	11,444
Missense Variant	11,139	11,467	11,116	11,408	10,776
Stop Gained	88	111	107	109	107
Stop Lost	41	40	48	44	41
Number of INDELs	9,911	10,000	10,126	10,003	9,637
Frameshift Variant	312	310	322	322	296
Inframe Insertion	178	169	165	175	166
Inframe Deletion	200	195	208	186	184
% found in dbSNP142	97.1	97.0	96.9	96.9	96.9
Het/Hom Ratio	1.4	1.7	1.3	1.6	1.1
Ts/Tv Ratio	2.3	2.3	2.3	2.3	2.3

Het/Hom Ratio; Ratio of number of heterozygous variants to number of homozygous variants, Ts/Tv Ratio; Ratio of transition rate of SNVs that pass the quality filters divided by transversion rate of SNVs that pass the quality filters. Transition rate of SNVs that pass the quality filters divided by transversion rate of SNVs that pass the quality filters.

*3.1. Exome Sequencing: Initial Screening for Novel Genes_*

Rare variants, as well as those variants which were absent in the population variation databases, were prioritized for further analysis. Initially 55 candidate variants in 22 genes were prioritized based on filteration criteria described in materials and methods section. Variants in advanced phase CML patients were filtered. Three novel genes (ANKRD36, ANKRD36B and PRSS3) mutated in all advanced phase CML patients but not in CP-CML and healthy controls were found. Data generated from next generation has been submitted to NCBI and can be accessed through link https://www.ncbi.nlm.nih.gov/sra/PRJNA734750 (SRA accession number PRJNA734750).

3.2. Mutation Validation by Sanger Sequencing

ANKRD36B (c.2758A>G) and PRSS3 (c.473_474insCC and c.478_479delAC) variants were not confirmed using Sanger sequencing. However, ANKRD36 gene mutations (c.1183_1184 delGC and c.1187_1185 dupTT) were confirmed by Sanger sequencing in BC-CML patients (figure 1), demonstrating the association between ANKRD36 variants and CML progression. ANKRD36 mutations were confirmed in AP-CML as well, showing that these mutations are an early indicator of CML progression. It also shows that ANKRD36 mutations are a potential early biomarker of CML progression.

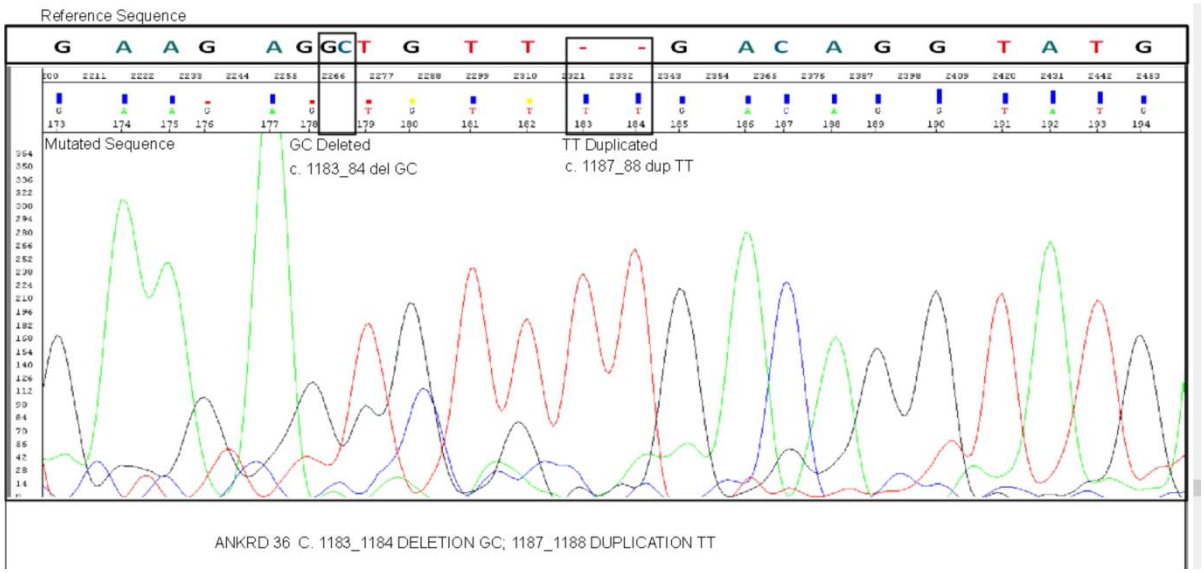

Figure 1. Confirmation of the presence of c.1183_1184 delGC and c.1187_1185 dup TT mutation in ANKRD36 gene by Sanger sequencing in unknown gene variants common in Accelerate/Blast phase (AP/BC) CML patients (AP, n=5, BC, n=7).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 August 2021 doi:10.20944/preprints202108.0498.v1

3.3. Protein Modelling studies

As structure of protein encoded by ANKRD36 was unknown and no prior PDB deposit was available. Therefore, ANKRD36 modelling studies were carried out using ANKRD36 protein sequence retrieved from UniProt (https://www.uniprot.org/uniprot/A6QL64). Computational prediction of the protein structure was done using I-Tasser webserver. The mutation was manually evaluated and the wild and mutated structures were superimposed using PyMOL to shed light on structural changes induced.

Effect of nonsynonymous missense mutation has been shown in figure 2 wherein we zoomed into the region harboring the two nonsynonymous missense mutation.

Our analysis shows that these mutations induced structural changes in ANKRD36 protein due to incorporation of bigger cysteine (Cys) and phenylalanine (Phe) residues instead of the comparatively smaller alanine (ala) and valine (val) on residues 395 and 396, respectively (Figure 2). The RMSD lied in range of 0.025-0.043 [45, 46]. A395C mutation has not been previously reported and could be of significance. Functional changes and possible pathogenesis associated with ANKRD36 gene may have been due to these mutations that lead to structural changes in protein encoded by ANKRD36. It also indicates that mutated ANKRD36 protein may have important role in CML progression and may be a potential new drug target in CML progression.

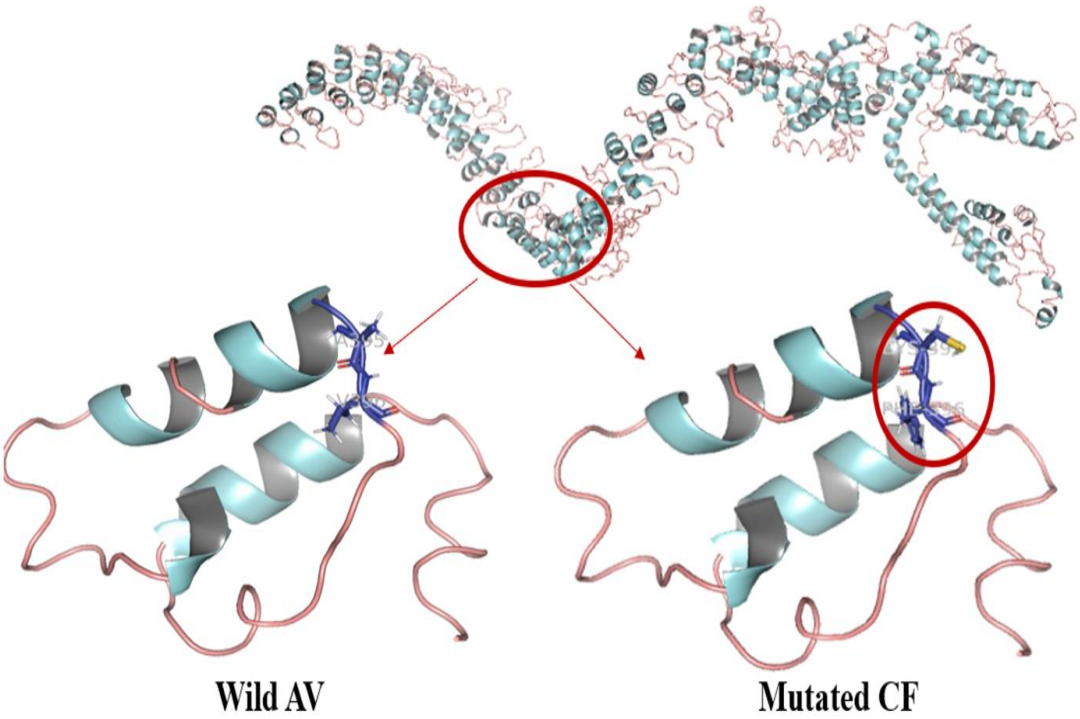

Figure 2. Protein modelling studies of normal wild-type and mutated ANKRD36.

4. Discussion:

This study included overall 141 patients from different phases of CML. In our study mean age of the patients was 36.4 ± 5.2 years. This is important to mention that mean age of our CML patients is significantly different from western populations. In Europe, the mean age of CML patients was 55 years [47]. A study reported the mean age of CML patients in the USA to be 41.1±13.3 years [48]. In Japan, the mean age of patients diagnosed with CML was 56 years [49]. Due to this factor, life expectancy of CML patients is not comparable to general population in developing countries as it is for developed countries like US, Europe, and Japan [47–49]. Furthermore, there was a significant difference between

 doi:10.20944/preprints202108.0498.v1

chronic and advanced phase patients with respect to male to female ratio, hemoglobin level, WBC count, platelet count, splenomegaly, and survival status that is in accordance with previous reports [47–49].

During course of study, 12.8% (n=18) of patients progressed to advanced phases (AP=6, BC=12). A European study reported 9 (4.2%) out 210 enrolled CML patients developed to advanced phases (AP=5, BC=4) [50]. Based on the results of the pivotal International Randomized Study of Interferon and STI571 (IRIS) trial involving 1,106 randomized patients newly diagnosed with CML, rate of progression of imatinib –treated patients was 3.3% [51]. A study carried out in Japan reported 7.5% (n=16) of CML patients progressed to advanced phases [49]. Variation in WBC and platelet counts in our subjects and patients from other populations shows biological differences in AP-/BC-CML patients from different geographic regions that might be due to ethnic variations, different genetic basis of CML progression in different ethnic groups and approach to clinical management of CML [47, 49-51]. A higher frequency of CML progression in our patients can be attributed to unavailability of all FDA-approved drugs, very few bone marrow transplantation centers and non-compliance of CML patients. It necessitated finding out some early biomarkers of disease progression for our CML patients.

As there are no common molecular biomarkers available for early detection of CML progression [52], we subjected our advanced phase CML patients to exome sequencing and compared them with CP-CML as well as healthy controls. We found ANKRD36 gene exclusively mutated in all BC-and AP-CML patients but in none of the CP-CML patients and healthy controls. ANKRD36 is a novel gene which is still uncharacterized in human. Nevertheless, maximum expression of ANKRD36 is reported to be in myeloid cells of the bone marrow [53]. It is located on chromosome 2q11.2.

ANKRD36 main function and its exact role in CML or any other cancer is still unknown. However, various studies found an association between specific health conditions and ANKRD36. In type 2 diabetes mellitus patients (T2DM), ANKRD36 expression was found to be significantly upregulated as compared to normal controls [54]. CircANKRD36 (circular RNA transcribed by ANKRD36) level was positively correlated with glucose, glycosylated hemoglobin, and IL-6. Furthermore, leucocytes expressed high levels of circANKRD36 in T2DM patients. Therefore, circANKRD36 may be used as a biomarker for screening chronic inflammation in patients with T2DM [54]. Another study showed an association between pneumonia pathogenesis and circANKRD36 [55]. Irritated MRC-5 cell injury by lipopolysaccharide (LPS) evoked circANKRD36 to activate the NF-κB signaling pathway and caused inflammation in MRC-5 cells. When circANKRD36 was silenced, NF-κB pathway was inactivated, and it significantly increased the viability of LPS-aroused MRC-5 and decreased cell apoptosis [55]. Moreover, a similar study revealed the association between circANKRD36 and NF-κB pathway activation in H9c2 cells treated with LPS [56]. These studies show that ANKRD36 mutations can be categorized as "likely to be pathogenic" and this gene may have a role in CML biology and progression.

Our protein biomodelling studies also indicate that ANKRD36 mutations reported by us fall under the category of "likely to be pathogenic" genetic alterations. This gene contains "Ankyrin repeat" which is a 33-residue motif in proteins consisting of two alpha helices separated by loops. Typically, Ankyrin repeats fold together to form a single, linear solenoid structure called ankyrin repeat domains, one of the most common protein–protein interaction platforms especially in eukaryotes. They participate in diverse functions such as transcriptional initiators, cell cycle regulators, cytoskeletal, ion transporters, and signal transducers. Clinical significance of natural variations in several of ankyrin proteins have been previously reported to affect the specificity of protein interactions [57, 58]. Mutation effect due to simultaneous "deletion of GC and Insertion of TT" results in 2 amino acid change: Ala to Cys (395) and Val to Phe (396). Both of Val and Phe are hydrophobic, positionally interchangeable and resonate the same overall protein function because protein function is preserved due to retention of specific nucleotides in the DNA codon that encode amino acids with similar polarity or hydrophobicity substitution [59]. Nevertheless,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 August 2021 doi:10.20944/preprints202108.0498.v1

A395C mutation hasn't been previously reported and might be of more importance as rare mutations are more pathogenic than the frequent one. The mutation location is on surface exterior linking the two alpha helices and might alter the flexibility of the protein. This might hamper the potential interaction with other interacting proteins [57-59]. Possible predictions of functional annotation of partially characterized proteins and their functional domains surely needs further validation.

We searched "The Cancer Genome Atlas (TCGA)" of the National Cancer Institute of the National Institute of Health and "cBioPortal for Cancer Genomics" to find out any leukemia specific ANKRD36 mutations. Nevertheless, we could not find ANKRD36 mutations related to any type of leukemia. However, various studies have found role of ANKRD36 in different cancers. A study analyzing the antitumor role of miR-144-5p on renal cell carcinoma (RCC) showed that ANKRD36 gene is targeted by miR-144-5p [60]. In this study, poor survival was associated with high expression of miR-144-5p-regulated-ANKRD36. Data from miRTarBase database of micro-RNAases shows that ANKRD36 is also regulated by miR-182, which is a miRNA expressed in early stages of tumor growth [61]. A study showed that the silencing of miR-182 enhanced apoptosis. Moreover, reduction in tumor growth was observed in-vivo when anti-miR-182 treated cells were transplanted in immunodeficient mice. From these studies, it can be inferred that ANRD36 has role in carcinogenesis and in in regulation of apoptosis. Moreover, it also indicates that silencing of ANKRD36 *miR-182* and *miR-144-5p* can suppress the tumor growth and increase apoptotic activity of the cancer cells. Thus, inhibition *of miR-182* and *miR-144-5p* might be important drug targets to find new treatment for advanced phases of cancers where ANKRD36 has some role including CML [61]. In another study, mutational status of ANKRD36 genes was found to be correlated with proximal gastric cancer [62]. ANKRD36 has been reported to be co-expressing and interacting with other genes on locus 2q11.2 including ANKRD36C, ITPRIPL1, FAHD2B, FAM178B, and CNNM3 that shows that ANKRD36 is involved in some important biological network associated with cancers [63] Studies have also found that ANKRD36 is upregulated by PIM1 inhibitors [64]. All these studies highlight significance of ANKRD36 in important biological functions and its association with cancer, as well as show that this gene is targetable and druggable if found mutated. As this gene has been found to have highest expression in myeloid cells of the bone marrow, it may serve as a novel biomarker and drug target for CML patients with advanced phases of the disease [53]. Further studies are recommended for biological characterization of this gene in human and to find its possible role in CML progression and pathogenesis of other diseases.

In our studies, Two out of three variants were not confirmed using Sanger sequencing. These variants may arise as a result of inevitable technical artifacts that are not uncommon in NGS-based studies and might be resulted due to a number of reasons. Next-generation sequencing techniques generate low-interest variants in the form of genotype false positives. Biases in the library construction mat lead to errors [65-69]. Moreover, we used NextSeq for WES and this technology generates short reads. It is difficult to call genotypes at the end of short reads [70]. False-positive in NGS data may also arise as a result of misalignment of sequencing reads to the RefSeq and inaccuracies or biases of the RefSeq compared to a specific local population [71]. Therefore, these factors should also be kep in mind during NGS-based investigations to avoid false positive results.

5. Conclusions:

We hereby report mutations in a novel gene ANKRD36, associated with disease progression in CML and hence can serve as important biomarker to identify CML patients at risk of disease progression. Our protein biomodelling studies show that these mutations change the structure of ANKRD36 protein that might affect its biological functions. Although this gene is yet to be characterized in human, various studies indicate its involvement in different biological functions and pathogenesis of diseases, including cancer. As this gene has found to have maximum expression in bone marrow, specifically myeloid cells, this gene may important role in hematopoiesis and a therefore a potential role in hematopoietic diseases, specifically in CML progression. Accordingly, we recommend

further studies to find out exact biological functions of this gene, specifically its role in apoptosis and cancer carcinogenesis.

Acknowledgments/Source(s) of support: This work was funded by the National Plan for Science, Technology and Innovation (MAARIFAH), King Abdul-Aziz City for Science and Technology, Kingdom of Saudi Arabia, grant Number 14-Med2817-02. The study was approved by King Abdullah International Medical Research Centre (KAIMRC), National Guard Health Affairs, Saudi Arabia although no research funding was provided (project # RA17/002/A). This study was partially supported by the College of Medicine, Research Centre, Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia.

Informed consent statement:
Formal informed consent was obtained from all study subjects.

Institutional review board statement:
The study was approved by institutional review boards of participating centers.

Data Availability Statement:
Data generated from next generation has been submitted to NCBI and can be accessed through link https://www.ncbi.nlm.nih.gov/sra/PRJNA734750 (SRA accession number PRJNA734750).

Reference

1- Ning L, Hu C, Lu P, Que Y, Zhu X, Li D. Trends in disease burden of chronic myeloid leukemia at the global, regional, and national levels: a population-based epidemiologic study. Exp Hematol Oncol. 2020 Nov 3;9(1):29. doi: 10.1186/s40164-020-00185-z. PMID: 33292593; PMCID: PMC7607878.

2- Flis S, Chojnacki T. Chronic myelogenous leukemia, a still unsolved problem: pitfalls and new therapeutic possibilities. Drug Des Devel Ther. 2019 Mar 8;13:825-843. doi: 10.2147/DDDT.S191303. PMID: 30880916; PMCID: PMC6415732.

3- Sampaio MM, Santos MLC, Marques HS, Gonçalves VLS, Araújo GRL, Lopes LW, Apolonio JS, Silva CS, Santos LKS, Cuzzuol BR, Guimarães QES, Santos MN, de Brito BB, da Silva FAF, Oliveira MV, Souza CL, de Melo FF. Chronic myeloid leukemia-from the Philadelphia chromosome to specific target drugs: A literature review. World J Clin Oncol. 2021 Feb 24;12(2):69-94. doi: 10.5306/wjco.v12.i2.69. PMID: 33680875; PMCID: PMC7918527.

4- Turgeon ML. Clinical hematology: theory and procedures. 5th ed. Wolters Kluwer Health/Lippincott Williams & Wilkins; 2012. p. 363-7.

5- Cortes J, Lang F. Third-line therapy for chronic myeloid leukemia: current status and future directions. J Hematol Oncol. 2021 Mar 18;14(1):44. doi: 10.1186/s13045-021-01055-9. PMID: 33736651; PMCID: PMC7976694.

6- Nasser A, Hussein A, Chamba C, Yonazi M, Mushi R, Schuh A, Luzzatto L. Molecular response to imatinib in patients with chronic myeloid leukemia in Tanzania. Blood Adv. 2021 Mar 9;5(5):1403-1411. doi: 10.1182/bloodadvances.2020002973. PMID: 33666650; PMCID: PMC7948290.

7- Kang ZJ, Liu YF, Xu LZ, Long ZJ, Huang D, Yang Y, Liu B, Feng JX, Pan YJ, Yan JS, Liu Q. The Philadelphia chromosome in leukemogenesis. Chin J Cancer. 2016 May 27;35:48. doi: 10.1186/s40880-016-0108-0. PMID: 27233483; PMCID: PMC4896164.

8- Jabbour E, Kantarjian H. Chronic myeloid leukemia: 2018 update on diagnosis, therapy and monitoring. Am J Hematol. 2018 Mar;93(3):442-459. doi: 10.1002/ajh.25011. PMID: 29411417.

9- Quintás-Cardama A, Cortes J. Molecular biology of bcr-abl1-positive chronic myeloid leukemia. Blood. 2009 Feb 19;113(8):1619-30. doi: 10.1182/blood-2008-03-144790. Epub 2008 Sep 30. PMID: 18827185; PMCID: PMC3952549.

10- Ali MA. Chronic Myeloid Leukemia in the Era of Tyrosine Kinase Inhibitors: An Evolving Paradigm of Molecularly Targeted Therapy. Mol Diagn Ther. 2016 Aug;20(4):315-33. doi: 10.1007/s40291-016-0208-1. PMID: 27220498.

11- Soverini S, Abruzzese E, Bocchia M, Bonifacio M, Galimberti S, Gozzini A, Iurlo A, Luciano L, Pregno P, Rosti G, Saglio G, Stagno F, Tiribelli M, Vigneri P, Barosi G, Breccia M. Next-generation sequencing for BCR-ABL1 kinase domain mutation testing in patients with chronic myeloid leukemia: a position paper. J Hematol Oncol. 2019 Dec 5;12(1):131. doi: 10.1186/s13045-019-0815-5. PMID: 31801582; PMCID: PMC6894351.

12- Sessions J. Chronic myeloid leukemia in 2007. Am J Health Syst Pharm. 2007 Dec 15;64(24 Suppl 15):S4-9. doi: 10.2146/ajhp070484. PMID: 18056931.

13- Valent P, Herndlhofer S, Schneeweiß M, Boidol B, Ringler A, Kubicek S, Gleixner KV, Hoermann G, Hadzijusufovic E, Müllauer L, Sperr WR, Superti-Furga G, Mannhalter C. TKI rotation-induced persistent deep molecular response in multi-resistant blast crisis of Ph+ CML. Oncotarget. 2017 Apr 4;8(14):23061-23072. doi: 10.18632/oncotarget.15481. PMID: 28416739; PMCID: PMC5410285.

14- Kotagama K, Chang Y, Mangone M. miRNAs as Biomarkers in Chronic Myelogenous Leukemia. Drug Dev Res. 2015 Sep;76(6):278-85. doi: 10.1002/ddr.21266. Epub 2015 Aug 18. PMID: 26284455; PMCID: PMC4758911.

15- Shet AS, Jahagirdar BN, Verfaillie CM. Chronic myelogenous leukemia: mechanisms underlying disease progression. Leukemia. 2002 Aug;16(8):1402-11. doi: 10.1038/sj.leu.2402577. PMID: 12145676.

16- Branford S, Wang P, Yeung DT, Thomson D, Purins A, Wadham C, Shahrin NH, Marum JE, Nataren N, Parker WT, Geoghegan J, Feng J, Shanmuganathan N, Mueller MC, Dietz C, Stangl D, Donaldson Z, Altamura H, Georgievski J, Braley J, Brown A, Hahn C, Walker I, Kim SH, Choi SY, Park SH, Kim DW, White DL, Yong ASM, Ross DM, Scott HS, Schreiber AW, Hughes TP. Integrative genomic analysis reveals cancer-associated mutations at diagnosis of CML in patients with high-risk disease. Blood. 2018 Aug 30;132(9):948-961. doi: 10.1182/blood-2018-02-832253. Epub 2018 Jul 2. PMID: 29967129.

17- Xu J, Wu M, Sun Y, Zhao H, Wang Y, Gao J. Identifying Dysregulated lncRNA-Associated ceRNA Network Biomarkers in CML Based on Dynamical Network Biomarkers. Biomed Res Int. 2020 Feb 18;2020:5189549. doi: 10.1155/2020/5189549. PMID: 32149112; PMCID: PMC7049421.

18- Baccarani M, Deininger MW, Rosti G, Hochhaus A, Soverini S, Apperley JF, Cervantes F, Clark RE, Cortes JE, Guilhot F, Hjorth-Hansen H, Hughes TP, Kantarjian HM, Kim DW, Larson RA, Lipton JH, Mahon FX, Martinelli G, Mayer J, Müller MC, Niederwieser D, Pane F, Radich JP, Rousselot P, Saglio G, Saußele S, Schiffer C, Silver R, Simonsson B, Steegmann JL, Goldman JM, Hehlmann R. European LeukemiaNet recommendations for the management of chronic myeloid leukemia: 2013. Blood. 2013 Aug 8;122(6):872-84. doi: 10.1182/blood-2013-05-501569. Epub 2013 Jun 26. PMID: 23803709; PMCID: PMC4915804.

19- Cortes JE, Talpaz M, O'Brien S, Faderl S, Garcia-Manero G, Ferrajoli A, Verstovsek S, Rios MB, Shan J, Kantarjian HM. Staging of chronic myeloid leukemia in the imatinib era: an evaluation of the World Health Organization proposal. Cancer. 2006 Mar 15;106(6):1306-15. doi: 10.1002/cncr.21756. PMID: 16463391.

20- Suttorp M, Millot F, Sembill S, Deutsch H, Metzler M. Definition, Epidemiology, Pathophysiology, and Essential Criteria for Diagnosis of Pediatric Chronic Myeloid Leukemia. Cancers (Basel). 2021 Feb 14;13(4):798. doi: 10.3390/cancers13040798. PMID: 33672937; PMCID: PMC7917817.

21- Cagnetta A, Garuti A, Marani C, Cea M, Miglino M, Rocco I, Palermo C, Fugazza G, Cirmena G, Colombo N, Grasso R, Nencioni A, Gobbi M, Patrone F. Evaluating treatment response of chronic myeloid leukemia: emerging science and technology. Curr Cancer Drug Targets. 2013 Sep;13(7):779-90. doi: 10.2174/15680096113139990084. PMID: 23906053.

22- Cross NC, White HE, Müller MC, Saglio G, Hochhaus A. Standardized definitions of molecular response in chronic myeloid leukemia. Leukemia. 2012 Oct;26(10):2172-5. doi: 10.1038/leu.2012.104. Epub 2012 Apr 16. PMID: 22504141.

23- Mahon FX, Etienne G. Deep molecular response in chronic myeloid leukemia: the new goal of therapy? Clin Cancer Res. 2014 Jan 15;20(2):310-22. doi: 10.1158/1078-0432.CCR-13-1988. Epub 2013 Oct 28. PMID: 24166905.

24- Driscoll JJ, Rixe O. Overall survival: still the gold standard: why overall survival remains the definitive end point in cancer clinical trials. Cancer J. 2009 Sep-Oct;15(5):401-5. doi: 10.1097/PPO.0b013e3181bdc2e0. PMID: 19826360.

25- Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. Int J Ayurveda Res. 2010 Oct;1(4):274-8. doi: 10.4103/0974-7788.76794. PMID: 21455458; PMCID: PMC3059453.

26- World Medical Association. Ethics Unit. *Declaration of Helsinki* 2007. www.wma.net/e/ethicsunit/helsinki.htm.

27- Goodyear MD, Krleza-Jeric K, Lemmens T. The Declaration of Helsinki. BMJ. 2007 Sep 29;335(7621):624-5. doi: 10.1136/bmj.39339.610000.BE. PMID: 17901471; PMCID: PMC1995496.

28- Ip SC, Lin SW, Lai KM. An evaluation of the performance of five extraction methods: Chelex® 100, QIAamp® DNA Blood Mini Kit, QIAamp® DNA Investigator Kit, QIAsymphony® DNA Investigator® Kit and DNA IQ™. Sci Justice. 2015 May;55(3):200-8. doi: 10.1016/j.scijus.2015.01.005. Epub 2015 Jan 28. PMID: 25934373.

29- Absar M, Mahmood A, Akhtar T, Basit S, Ramzan K, Jameel A, Afzal S, Ullah A, Qureshi K, Alanazi N, Iqbal Z. Whole exome sequencing identifies a novel FANCD2 gene splice site mutation associated with disease progression in chronic myeloid leukemia: Implication in targeted therapy of advanced phase CML. Pak J Pharm Sci. 2020 May;33(3(Special)):1419-1426. PMID: 33361032.

30- Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, Vertino-Bell A, Smaoui N, Neidich J, Monaghan KG, McKnight D, Bai R, Suchy S, Friedman B, Tahiliani J, Pineda-Alvarez D, Richard G, Brandt T, Haverfield E, Chung WK, Bale S. Clinical application of whole-exome sequencing across clinical indications. Genet Med. 2016 Jul;18(7):696-704. doi: 10.1038/gim.2015.148. Epub 2015 Dec 3. PMID: 26633542.

31- AlAsiri S, Basit S, Wood-Trageser MA, Yatsenko SA, Jeffries EP, Surti U, Ketterer DM, Afzal S, Ramzan K, Faiyaz-Ul Haque M, Jiang H, Trakselis MA, Rajkovic A. Exome sequencing reveals MCM8 mutation underlies ovarian failure and chromosomal instability. J Clin Invest. 2015 Jan;125(1):258-62. doi: 10.1172/JCI78473. Epub 2014 Dec 1. PMID: 25437880; PMCID: PMC4382257.

32- Carson AR, Smith EN, Matsui H, Brækkan SK, Jepsen K, Hansen JB, Frazer KA. Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. BMC Bioinformatics. 2014 May 2;15:125. doi: 10.1186/1471-2105-15-125. PMID: 24884706; PMCID: PMC4098776.

33- Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, Murphy KM. Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. J Mol Diagn. 2010 Jul;12(4):425-32. doi: 10.2353/jmoldx.2010.090188. Epub 2010 Apr 29. PMID: 20431034; PMCID: PMC2893626.

34- Beck TF, Mullikin JC; NISC Comparative Sequencing Program, Biesecker LG. Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. Clin Chem. 2016 Apr;62(4):647-54. doi: 10.1373/clinchem.2015.249623. Epub 2016 Feb 4. PMID: 26847218; PMCID: PMC4878677.

35- Diaz-Horta O, Duman D, Foster J 2nd, Sırmacı A, Gonzalez M, Mahdieh N, Fotouhi N, Bonyadi M, Cengiz FB, Menendez I, Ulloa RH, Edwards YJ, Züchner S, Blanton S, Tekin M. Whole-exome sequencing efficiently detects rare mutations in autosomal recessive nonsyndromic hearing loss. PLoS One. 2012;7(11):e50628. doi: 10.1371/journal.pone.0050628. Epub 2012 Nov 30. PMID: 23226338; PMCID: PMC3511533.

36- *R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/*

37- Sokal JE, Cox EB, Baccarani M, Tura S, Gomez GA, Robertson JE, Tso CY, Braun TJ, Clarkson BD, Cervantes F, et al. Prognostic discrimination in "good-risk" chronic granulocytic leukemia. Blood. 1984 Apr;63(4):789-99. PMID: 6584184.

38- Hasford J, Pfirrmann M, Hehlmann R, Allan NC, Baccarani M, Kluin-Nelemans JC, Alimena G, Steegmann JL, Ansari H. A new prognostic score for survival of patients with chronic myeloid leukemia treated with interferon alfa. Writing Committee for the Collaborative CML Prognostic Factors Project Group. J Natl Cancer Inst. 1998 Jun 3;90(11):850-8. doi: 10.1093/jnci/90.11.850. PMID: 9625174.

39- Hasford J, Baccarani M, Hoffmann V, Guilhot J, Saussele S, Rosti G, Guilhot F, Porkka K, Ossenkoppele G, Lindoerfer D, Simonsson B, Pfirrmann M, Hehlmann R. Predicting complete cytogenetic response and subsequent progression-free survival in 2060 patients with CML on imatinib treatment: the EUTOS score. Blood. 2011 Jul 21;118(3):686-92. doi: 10.1182/blood-2010-12-319038. Epub 2011 May 2. PMID: 21536864.

40- Kuntegowdanahalli LC, Kanakasetty GB, Thanky AH, Dasappa L, Jacob LA, Mallekavu SB, Lakkavalli RK, Kadabur LN, Haleshappa RA. Prognostic and predictive implications of Sokal, Euro and EUTOS scores in chronic myeloid leukaemia in the imatinib era-experience from a tertiary oncology centre in Southern India. Ecancermedicalscience. 2016 Oct 6;10:679. doi: 10.3332/ecancer.2016.679. PMID: 27899952; PMCID: PMC5102687.

41- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods. 2015 Jan;12(1):7-8. doi: 10.1038/nmeth.3213. PMID: 25549265; PMCID: PMC4428668.

42- *DeLano, W. L. CCP4 Newsletter on Protein Crystallography. 2002; 40:82-92.*

43- Khazaal MS, Hamdan FB, Al-Mayah QS. Association of BCR/ABL transcript variants with different blood parameters and demographic features in Iraqi chronic myeloid leukemia patients. Mol Genet Genomic Med. 2019 Aug;7(8):e809. doi: 10.1002/mgg3.809. Epub 2019 Jun 17. PMID: 31206255; PMCID: PMC6687619.

44- Kumar S, Gupta VK, Bharti A, Meena LP, Gupta V, Shukla J. A study to determine the clinical, hematological, cytogenetic, and molecular profile in CML patient in and around Eastern UP, India. J Family Med Prim Care. 2019 Jul;8(7):2450-2455. doi: 10.4103/jfmpc.jfmpc_307_19. PMID: 31463275; PMCID: PMC6691447.

45- Koene HR, Kleijer M, Algra J, Roos D, von dem Borne AE, de Haas M. Fc gammaRIIIa-158V/F polymorphism influences the binding of IgG by natural killer cell Fc gammaRIIIa, independently of the Fc gammaRIIIa-48L/R/H phenotype. Blood. 1997 Aug 1;90(3):1109-14. PMID: 9242542.

46- Castro-Chavez F. The rules of variation: amino acid exchange according to the rotating circular genetic code. J Theor Biol. 2010 Jun 7;264(3):711-21. doi: 10.1016/j.jtbi.2010.03.046. Epub 2010 Apr 3. PMID: 20371250; PMCID: PMC3130497.

47- Hoffmann VS, Baccarani M, Hasford J, Lindoerfer D, Burgstaller S, Sertic D, Costeas P, Mayer J, Indrak K, Everaus H, Koskenvesa P, Guilhot J, Schubert-Fritschle G, Castagnetti F, Di Raimondo F, Lejniece S, Griskevicius L, Thielen N, Sacha T, Hellmann A, Turkina AG, Zaritskey A, Bogdanovic A, Sninska Z, Zupan I, Steegmann JL, Simonsson B, Clark RE, Covelli A, Guidi G, Hehlmann R. The EUTOS population-based registry: incidence and clinical characteristics of 2904 CML patients in 20 European Countries. Leukemia. 2015 Jun;29(6):1336-43. doi: 10.1038/leu.2015.73. Epub 2015 Mar 18. PMID: 25783795.

48- Szabo SM, Levy AR, Davis C, Holyoake TL, Cortes J. A multinational study of health state preference values associated with chronic myelogenous leukemia. Value Health. 2010 Jan-Feb;13(1):103-11. doi: 10.1111/j.1524-4733.2009.00573.x. Epub 2009 Jul 29. PMID: 19659707.

49- Kizaki M, Takahashi N, Iriyama N, Okamoto S, Ono T, Usui N, Inokuchi K, Nakaseko C, Kurokawa M, Sumi M, Nakamura F, Kawaguchi T, Suzuki R, Yamamoto K, Ohnishi K, Matsumura I, Naoe T; New TARGET investigators. Efficacy and safety of tyrosine kinase inhibitors for newly diagnosed chronic-phase chronic myeloid leukemia over a 5-year period: results from the Japanese registry obtained by the New TARGET system. Int J Hematol. 2019 Apr;109(4):426-439. doi: 10.1007/s12185-019-02613-1. Epub 2019 Feb 14. PMID: 30762219.

50- Cervantes F, López-Garrido P, Montero MI, Jonte F, Martínez J, Hernández-Boluda JC, Calbacho M, Sureda A, Pérez-Rus G, Nieto JB, Pérez-López C, Román-Gómez J, González M, Pereira A, Colomer D. Early intervention during imatinib therapy in patients with newly diagnosed chronic-phase chronic myeloid leukemia: a study of the Spanish PETHEMA group. Haematologica. 2010 Aug;95(8):1317-24. doi: 10.3324/haematol.2009.021154. Epub 2010 Mar 10. PMID: 20220063; PMCID: PMC2913080.

51- O'Brien SG, Guilhot F, Larson RA, Gathmann I, Baccarani M, Cervantes F, Cornelissen JJ, Fischer T, Hochhaus A, Hughes T, Lechner K, Nielsen JL, Rousselot P, Reiffers J, Saglio G, Shepherd J, Simonsson B, Gratwohl A, Goldman JM, Kantarjian H, Taylor K, Verhoef G, Bolton AE, Capdeville R, Druker BJ; IRIS Investigators. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. N Engl J Med. 2003 Mar 13;348(11):994-1004. doi: 10.1056/NEJMoa022457. PMID: 12637609.

52- Ahmad R, Tripathi AK, Tripathi P, Singh S, Singh R, Singh RK. Malondialdehyde and protein carbonyl as biomarkers for oxidative stress and disease progression in patients with chronic myeloid leukemia. In Vivo. 2008 Jul-Aug;22(4):525-8. PMID: 18712183.

53- Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K, Asplund A, Sjöstedt E, Lundberg E, Szigyarto CA, Skogs M, Takanen JO,

Berling H, Tegel H, Mulder J, Nilsson P, Schwenk JM, Lindskog C, Danielsson F, Mardinoglu A, Sivertsson Å, von Feilitzen K, Forsberg M, Zwahlen M, Olsson I, Navani S, Huss M, Nielsen J, Ponten F, Uhlén M. Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics. Mol Cell Proteomics. 2014 Feb;13(2):397-406. doi: 10.1074/mcp.M113.035600. Epub 2020 Oct 1. PMID: 33498127.

54- Fang Y, Wang X, Li W, Han J, Jin J, Su F, Zhang J, Huang W, Xiao F, Pan Q, Zou L. Screening of circular RNAs and validation of circANKRD36 associated with inflammation in patients with type 2 diabetes mellitus. Int J Mol Med. 2018 Oct;42(4):1865-1874. doi: 10.3892/ijmm.2018.3783. Epub 2018 Jul 18. PMID: 30066828; PMCID: PMC6108858.

55- Guo R, Zhang L, Meng J. Circular RNA ANKRD36 attends to lipopolysaccharide-aroused MRC-5 cell injury via regulating microRNA-31-3p. Biofactors. 2020 May;46(3):391-401. doi: 10.1002/biof.1592. Epub 2019 Dec 2. PMID: 31793082.

56- Shi S, Zhang S, Zhang H, Jin Q, Wu D. Silencing circANKRD36 protects H9c2 cells against lipopolysaccharide-induced injury via up-regulating miR-138. Exp Mol Pathol. 2019 Dec;111:104300. doi: 10.1016/j.yexmp.2019.104300. Epub 2019 Aug 20. PMID: 31442445.

57- Neville MJ, Johnstone EC, Walton RT. Identification and characterization of ANKK1: a novel kinase gene closely linked to DRD2 on chromosome band 11q23.1. Hum Mutat. 2004 Jun;23(6):540-5. doi: 10.1002/humu.20039. PMID: 15146457.

58- Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY. The ankyrin repeat as molecular architecture for protein recognition. Protein Sci. 2004 Jun;13(6):1435-48. doi: 10.1110/ps.03554604. PMID: 15152081; PMCID: PMC2279977.

59- Castro-Chavez F. The rules of variation: amino acid exchange according to the rotating circular genetic code. J Theor Biol. 2010 Jun 7;264(3):711-21. doi: 10.1016/j.jtbi.2010.03.046. Epub 2010 Apr 3. PMID: 20371250; PMCID: PMC3130497.

60- Yamada Y, Arai T, Kojima S, Sugawara S, Kato M, Okato A, Yamazaki K, Naya Y, Ichikawa T, Seki N. Regulation of antitumor miR-144-5p targets oncogenes: Direct regulation of syndecan-3 and its clinical significance. Cancer Sci. 2018 Sep;109(9):2919-2936. doi: 10.1111/cas.13722. Epub 2018 Jul 28. PMID: 29968393; PMCID: PMC6125479.

61- Perilli L, Tessarollo S, Albertoni L, Curtarello M, Pastò A, Brunetti E, Fassan M, Rugge M, Indraccolo S, Amadori A, Bortoluzzi S, Zanovello P. Silencing of miR-182 is associated with modulation of tumorigenesis through apoptosis induction in an experimental model of colorectal cancer. BMC Cancer. 2019 Aug 20;19(1):821. doi: 10.1186/s12885-019-5982-9. PMID: 31429725; PMCID: PMC6700772.

62- Cong T, Liu GX, Cui JX, Zhang KC, Chen ZD, Chen L, Wei B, Huang XH. [Exome sequencing of gastric cancers screened the differences of clinicopathological phenotypes between the mutant and the wide-type of frequently mutated genes]. Zhonghua Yi Xue Za Zhi. 2018 Jul 24;98(28):2242-2245. Chinese. doi: 10.3760/cma.j.issn.0376-2491.2018.28.006. PMID: 30078278.

63- Jacquet L, Wood V, Kadeva N, Cornwell G, Codognotto S, Stephenson E, Ilic D. Generation of KCL040 clinical grade human embryonic stem cell line. Stem Cell Res. 2016 Jan;16(1):173-6. doi: 10.1016/j.scr.2015.12.035. Epub 2016 Jan 3. PMID: 27345807; PMCID: PMC4757728.

64- Garcia PG. The role of the PIM1 kinase in T-cell acute lymphoblastic leukemia/lymphoma. 2017. Available from: https://lib.ugent.be/fulltxt/RUG01/002/349/919/RUG01-002349919_2017_0001_AC.pdf

65- Fuentes Fajardo KV, Adams D; NISC Comparative Sequencing Program, Mason CE, Sincan M, Tifft C, Toro C, Boerkoel CF, Gahl W, Markello T. Detecting false-positive signals in exome sequencing. Hum Mutat. 2012 Apr;33(4):609-13. doi: 10.1002/humu.22033. Epub 2012 Mar 5. PMID: 22294350; PMCID: PMC3302978.

66- Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome biology*. 2011;12:R18.

67- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456:53–59.

68- Koboldt DC, Ding L, Mardis ER, Wilson RK. Challenges of sequencing human genomes. *Briefings in bioinformatics.* 2010;11:484–498.

69- Teer JK, Bonnycastle LL, Chines PS, Hansen NF, Aoyama N, Swift AJ, Abaan HO, Albert TJ, Margulies EH, Green ED, et al. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. *Genome research.* 2010;20:1420–1431.

70- Ledergerber C, Dessimoz C. Base-calling for next-generation sequencing platforms. Brief Bioinform. 2011 Sep;12(5):489-97. doi: 10.1093/bib/bbq077. Epub 2011 Jan 18. PMID: 21245079; PMCID: PMC3178052.

71- Church DM, Schneider VA, Graves T, Auger K, Cunningham F, Bouk N, Chen HC, Agarwala R, McLaren WM, Ritchie GR, et al. Modernizing reference genome assemblies. *PLoS Biol.* 2011;9:e1001091.