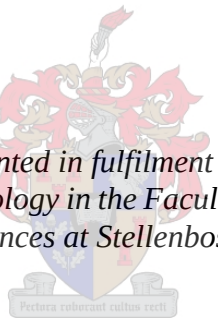
Screening and Characterisation of *BCR::ABL1* Kinase Domain Mutations in Chronic Myeloid Leukaemia Participants at Tygerberg Hospital, South Africa

by
Shareefa Isaacs



Thesis presented in fulfilment for the degree of
Master of Pathology in the Faculty of Medicine and Health Sciences at Stellenbosch University

Supervisor: Dr Carmen Swanepoel
Co-supervisor: Prof Zivanai Chapanduka

March 2023

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2023

Abstract

To date, an increased number of drug resistant cancers have been observed. The underlying mechanisms are not well understood; therefore, we use Chronic Myeloid Leukaemia (CML) as a disease model to explore the type and frequencies of potential drug resistant causing mutations and the effect on survival at Tygerberg Hospital (TBH). The underlying genetic abnormality is a t(9::22) chromosomal translocation, and consequent fusion between the Breakpoint Cluster Region and the Non-Receptor Tyrosine Kinase Abelson genes (*BCR::ABL1*). Although Tyrosine Kinase Inhibitors (TKD), such as Imatinib (Gleevec), effectively decrease *BCR::ABL1* mRNA transcript levels whereby the drug acts on the protein and inhibits its mode of action, point mutations within the kinase domain (KD) have shown to confer resistance to treatments.

A retrospective audit was conducted on a CML cohort from 2013 to 2020, and 20 of these participants were recruited for detection of KD mutations within the *ABL1* oncogene. Peripheral blood from routine diagnostic screening by the NHLS Molecular Haematology diagnostic laboratory were obtained for the mutation detection assay via bi-directional Sanger Sequencing. Identified sequence variants were then subjected to various bioinformatics tools to investigate variant, protein, and consequent effect on pathways.

For the audit, a total of 165 patients with confirmed CML treated at TBH was captured. Descriptive statistics showed 46.1% of CML patients were female, and 53.9% were male. The patients were from 69 different areas in the Northern part of the Cape Metropole and surrounding Boland areas in Western Cape, South Africa. Quantitative analysis showed the youngest patient to be two years old and the oldest 88 years old, $\bar{x}$ = 45-46 years. We found a weak linear correlation between age and *BCR::ABL1* mRNA transcript levels. A significant difference in *BCR::ABL1* mRNA transcript levels ($p < 0.000$) was seen in CML patients on treatment after 1 year. Of interest's sake, 104 of the initial 165 patients who were diagnosed were still undergoing treatment. Moreover, 57 out of the 165 patients presented with possible resistance; 75% were categorized as failure and 25% at warning stage according to European Leukaemia Net guidelines. In the second part of the study, 20 participants' samples were screened for variants. The screening sample group had a mortality rate of 20% who had a relative survival rate of less than five years. Sanger sequence analysis showed potential variants of interest in Exon 4 and of the *ABL1* gene, although bioinformatics analysis alludes to the fact that these variants are not likely to play a role in resistance.

In conclusion, the TBH CML cohort has a younger age at diagnosis which puts a greater strain on the public sector, and a significant number of patients are lost to follow up. More so, the age relative to CML deaths, as well as relative survival greatly differs from literature. Additionally, although synonymous mutations have recently shown to be pathological, the two variants found within our cohort has not been published to do so. CML could be a poster child for personalised medicine as the survival rate of CML patients is nearing that of the general population and treatment-free remission is attainable. Unfortunately, the paucity of CML data, drug availability and monitoring limitations are some of the obstacles that hinders South Africa from achieving these goals.

Opsomming

Tot op datum is 'n verhoogde aantal middel weerstandige kankers waargeneem. Die onderliggende meganismes word nie goed verstaan nie; daarom gebruik ons Chroniese Myeloïede Leukemie (CML) as 'n siektemodel om die tipe en frekwensies van potensiële geneesmiddel weerstandige veroorsakende mutasies en die effek op oorlewing te ondersoek. Die onderliggende genetiese abnormaliteit is 'n translokasie, en gevolglike samesmelting tussen die Breakpoint Cluster Region geen en die Nie-reseptor Tyrosien Kinase Abelson geen (BCR::ABL1), t(9::22). Alhoewel Tyrosien kinase-inhibeerders, soos Imatinib (Glivec), transkripsievlakke effektief verlaag duur inhibisie, het punt mutasies binne die kinase domein (KD) getoon dat dit weerstand teen behandelings verleen.

'n Terugwerkende oudit is uitgevoer op 'n CML-kohort vanaf 2013 tot 2020, en 20 van hierdie deelnemers is gewerf vir die opsporing van KD-mutasies binne *ABL1*-onkogeen. Perifere bloed van roetine-diagnostiese toets deur die NHLS Molekulêre Hematologie-diagnostiese laboratorium is verkry vir die mutasie-opsporingstoets via tweerigting Sanger-volgordebepaling. Geïdentifiseerde variante is aan verskeie bioinformatika-instrumente onderwerp om proteïen en gevolglike effekte te ondersoek.

Vir die oudit is 'n totaal van 165 pasiënte met bevestigde CML wat by TBH behandel is, opgeneem. Beskrywende statistieke het getoon 46,1% van CML-pasiënte was vroulik, en 53,9% was manlik. Die pasiënte was van 69 verskillende gebiede in die Noord-distrik van Wes-Kaap. Kwantitatiewe analise het getoon die jongste pasiënt was twee jaar oud en die oudste was 88 jaar oud, $\bar{x}$ = 45-46 jaar. Ons het 'n swak lineêre korrelasie tussen ouderdom en *BCR::ABL1*-transkripsievlakke gevind. 'n Beduidende verskil in *BCR::ABL1*-transkripsievlakke ($p < .000$) is gesien by pasiënte op behandeling na 1 jaar. Interessant is dat 104 van die aanvanklike 164 pasiënte wat gediagnoseer is, steeds behandeling ondergaan. Altesaam 57 uit die 165 pasiënte het moontlike weerstand getoon; 75% is gekategoriseer as mislukking en 25% op waarskuwing stadium volgens ELN-riglyne. In die tweede deel van die studie is 20 deelnemers monsters vir variante gekeur en die steekproefgroep het 'n sterftesyfer van 20% gehad, met 'n relatiewe oorlewingsyfer van minder as vyf jaar. Volgorde-analise het potensiële interessante variante in Ekson 4 en 9 van die *ABL*1-geen getoon. Bioinformatiese analise sinspeel egter op die feit dat hierdie variante waarskynlik nie 'n rol in weerstand sal speel nie.

Ten slotte, die TBH CML-kohort kom voor by jonger ouderdom groepe, wat 'n groter druk op die openbare sektor plaas, en 'n groot aantal pasiënte is verlies om op te volg. Die ouderdom tot CML-sterftes, sowel as relatiewe oorlewing verskil grootliks van literatuur. Alhoewel synonymous mutasies

onlangs as patologies getoon het, is die twee variante wat binne ons kohort gevind word nie gepubliseer om dit te doen nie. Selfs al kan CML 'n plakkaatkind vir persoonlike medisyne wees met die oorlewingsyfers van CML-pasiënte wat dié van die algemene bevolking nader en behandelingsvrye remissie wat nou bereikbaar is, is min CML-data, dwelmbeskikbaarheid en moniterings beperkings van die dinge wat Suid-Afrika verhinder om hierdie doelwitte te bereik.

Dedication

To my parents

Acknowledgements

This work could not have materialised without my remarkable supervisor, Dr Carmen Swanepoel. Her tireless efforts, patience, guidance, and mentorship is immeasurable. I have not only gained tremendous tools and knowledge, but character and confidence as a scientist and as a person.

Additionally, I like to thank my co-supervisor, Prof Zivanai Chapanduka and amazing collaborators; Mr Faghri February, Dr Fatima Bassa and Dr Dominique Anderson as well as the entire division of Haematological Pathology at Stellenbosch University and the National Health Laboratory Services team. A special vote of thanks to the National Research Foundation and Harry Crossley Foundation for funding this research.

Finally, I would like to thank my family and friends for their endless support, encouragement, and love, not forgetting the ones we lost along the way.

Table of Contents

List of Tables

List of Figures

List of abbreviations

AA	Amino Acid
ABL	Abelson Gene
ACMG	American College of Medical Genetics Genomics
ADP	Adenosine Diphosphate
AE	Elution Buffer
AFR	Africa
AlloSCT	Allogeneic Stem Cell Transplantation
A-loop	Activation Loop
AMP	Association for Molecular Pathology
AMR	America
AP	Accelerated Phase
ATP	Adenosine Triphosphate
CBL	Cbl E3 ubiquitin protein ligase
BC	Blast Crisis Phase
BCR	BreakPoint Cluster Region Gene
CCyR	Complete Cytogenetic Response
CDS	Coding sequence
CHR	Complete Haematologic Response
CML	Chronic Myeloid Leukaemia
CP	Chronic Phase
CTNNB1	Catenin (cadherin-associated protein) beta 1
DNA	Deoxyribonucleic Acid
Dr.	Doctor

EAS	East Asian
EDTA	Ethylenediaminetetraacetic Acid
ELN	European Leukaemia Net
EUR	European
Exac	Exome Aggregation Consortium
FDA	Food and Drug Administration
FISH	Fluorescence in situ Hybridization
HIV	Human Immunodeficiency Virus
HREC	Health Research Ethics Council
ICPerMed	International Consortium for Personalised Medicine
IS	International Scale
kDa	Kilodalton
LOF	Loss Of Function
MAF	Minor Allele Frequency
MAX	Maximum
M-bcr	Major Break Point Region
m-bcr	Minor Breakpoint Cluster Region
mCyR	Minor Cytogenetic Response
minCyR	Minimal Cytogenetic Response
MD	Molecular Dynamics
MIN	Minimum
MMR	Major Molecular Response

mRNA	Messenger Ribonucleic Acid
NA	Not Applicable
NCBI	National Center for Biotechnology Information
NCCN	National Comprehensive Cancer Network
NHGRI	National Human Genome Research Institute
NHLS	National Health Laboratory Services
noCyR	No Cytogenetic Response
P	Phosphate
PCR	Polymerase Chain Reaction
PCyR	Partial Cytogenetic Response
PDB	Protein Data Bank
Ph	Philadelphia Chromosome
P-loop	Phosphate Binding Loop
POPIA	Protection of Personal Information Act
qRT-PCR	Quantitative Real-Time Polymerase Chain Reaction
RCSB	Research Collaboratory for Structural Bioinformatics
RT-PCR	Real-Time Polymerase Chain Reaction
RMSD	Root Mean Square Deviation
SA	South Africa
SAS	South East Asian
SEER	Surveillance Epidemiology and End Results Program
SH3	src Homolog-3

SH2	src Homolog-2
SNV	Single Nucleotide Variant
SOP	Standard Operating Procedures
SU	Stellenbosch University
TBH	Tygerberg Hospital
TFR	Treatment-Free Remission
TK	Tyrosine Kinase
TKI	Tyrosine Kinase Inhibitor
USA	Unites States of America
UTR	Untranslated Region
WBC	White Blood Cell
WHO	World Health Organisation
2D	Two Dimensional
3D	Three Dimensional

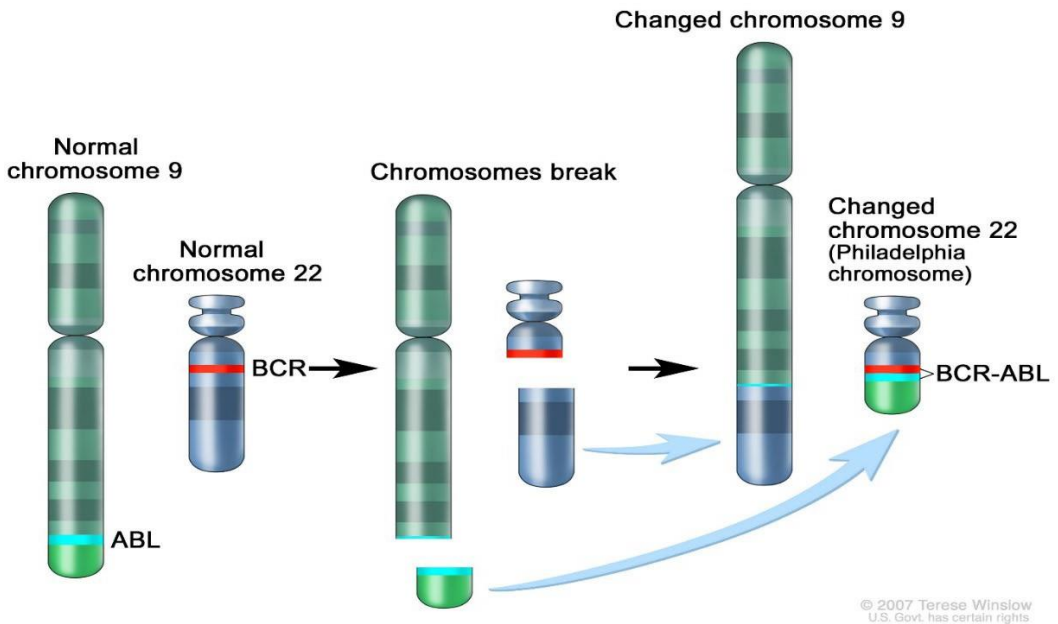
Chapter 1

Introduction

1.1 Literature review

1.1.1 Chronic Myeloid Leukaemia background

Chronic Myeloid Leukaemia (CML) is a malignancy of the bone marrow which results in an increased and unregulated production of myeloid cells in the blood. According to Bauer et al., 2012, the "Philadelphia (Ph)" chromosome, which is a consequence of the fusion between the breakpoint cluster (*BCR*) gene at chromosome 22 and the Abelson (*ABL1*) tyrosine kinase gene at chromosome 9 (t(9:22)) **(Figure 1.1)**, is the underlying genetic abnormality of CML (Bauer & Romvari, 2012). This is also known as the *BCR::ABL1* proto-oncogene, which encodes the chimeric BCR::ABL tyrosine kinase (TK) protein.



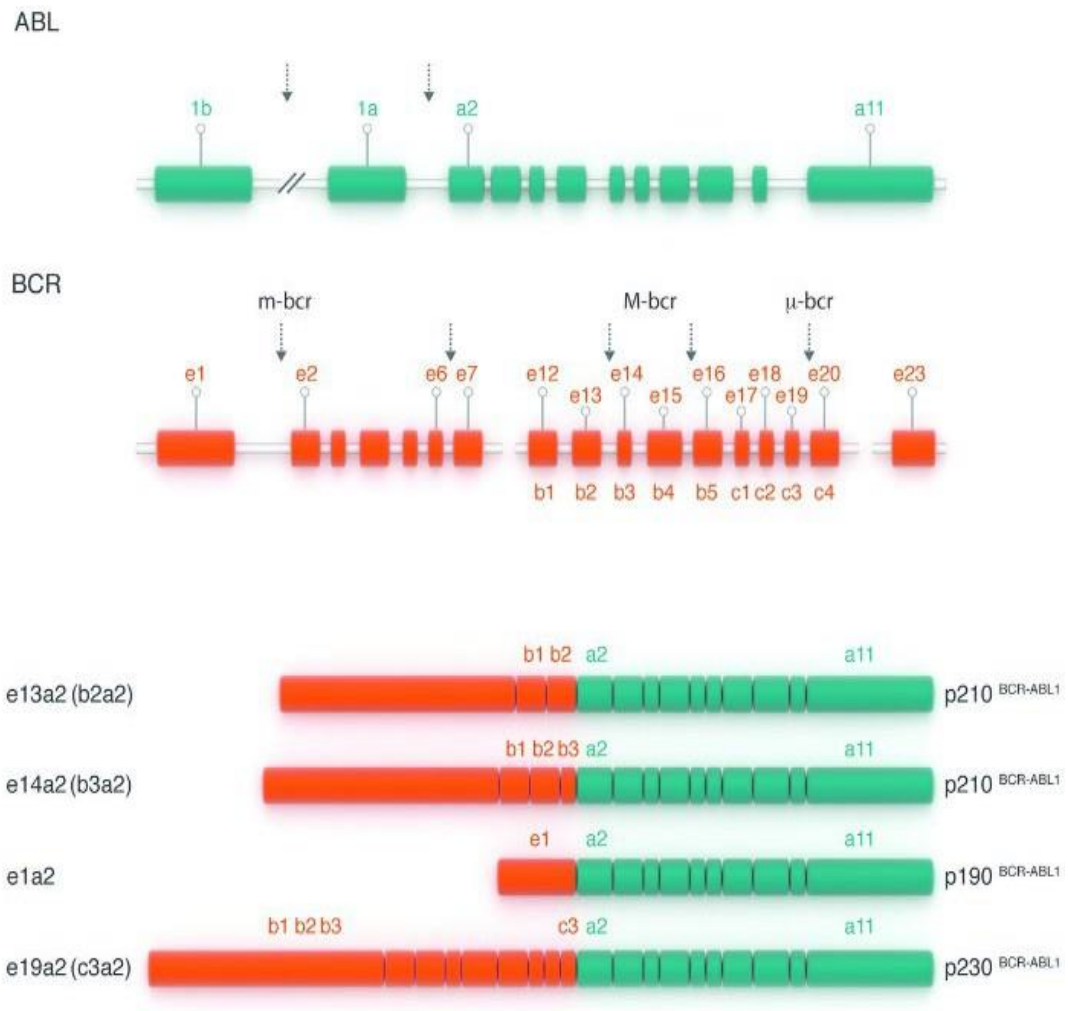
Extracted from the National Cancer Institute - https://www.cancer.gov/publications/dictionaries/cancer-terms/def/bcr-abl-fusion-gene

Figure 1.1 Formation of the *BCR::ABL1* oncogene fusion

To govern cellular functions such as proliferation, differentiation and survival, TK proteins transfer phosphates from Adenosine Triphosphate (ATP) to tyrosine residues on substrate proteins (Druker & Mauro, 2001). Therefore, the fusion BCR::ABL1 protein has upregulated kinase activity that alters cellular adhesion, activates mitogenic signalling and inhibits apoptosis, resulting in an excessive proliferation of hematopoietic stem cells (malignant transformation) (Mauro & Druker,2001). Additionally, it is linked to poor DNA repair, which leads to more chromosomal alterations and mutations (Tiwari & Wilson, 2019). This may partially explain the aggressive nature of CML.

Depending on the location of the fusion, various breakpoints occur resulting in variants with different molecular weight isoforms. The majority of CML patients have the p210 BCR::ABL1 fusion protein (Van Dongen et al., 1999). Breakpoints are found in the major break point region (M-bcr), located in intron 13 or 14 of the *BCR* gene on chromosome 22, also called b2 and 3 (Sandwith, 1859) which fuses to the large intron between *ABL1* exons 1b and 2 **(Figure 1.2)**. The hybrid *BCR::ABL1* mRNA transcript would therefore contain either *BCR* exon b2 or b3 and *ABL1* exon 2 (also called exon a2) which encodes the 210 kDa BCR- ABL protein **(Figure 1.2)**. Fifty-five percent (55%) of the b3-a2 transcript and 40% of the b2- a2 transcript junctions are associated with the 210 kDa BCR::ABL protein. In 5% of the cases, both b3-a2 and b2-a2 transcripts can be formed due to alternative splicing. However, no clinical outcome difference has been reported between the two p210 transcripts (Kang et al., 2016).

Some Ph+ patients have a larger *BCR::ABL1* fusion transcript that results from a fusion between exon 19 of the *BCR* gene (c3) and exon 2 of the *ABL* gene **(Figure 1.2)**. This occurs by breakpoints in the minor breakpoint cluster region (m-bcr) in the *BCR* gene between exon 19 and 20. This fusion encodes a large 230 kDa *BCR::ABL1* protein and disrupts granulocyte differentiation (Ayatollahi et al., 2018). It is interesting to note that intermittent Ph+ and Ph- CML with breakpoints in the *BCR* gene outside of the p210, p190 and p230 reported, have been shown to involve several other introns (Ayatollahi et al., 2018).



Extracted from - Chronic myeloid leukemia: reminiscences and dreams. Haematologica 2016;101(5):541-558; https://doi.org/10.3324/haematol.2015.139337.

Figure 1.2 *BCR* and *ABL1* breakpoints and fusion transcripts

The World Health Organization (WHO) book on Classification for Tumours of Haematopoeitic and Lymphoid tissue (4th Edition) states that bone marrow aspiration is mandatory for all patients who present with CML symptoms. Upon the presentation of persistent leucocytosis, routine cytogenetics; Fluorescence *in situ* hybridization (FISH) and molecular studies are performed. FISH and cytogenetics will detect the presence of the chromosomal abnormality, the t(9;22) (q34;q11) translocation whereas reverse transcriptase polymerase chain reaction (RT-PCR) is performed, which can either be qualitative to provide information about the presence of the *BCR::ABL1* mRNA transcript or quantitative to assess the amount of *BCR::ABL1* mRNA transcripts. The former is used for diagnosis and essentially identifies the specific type of rearrangement, and the latter for monitoring of treatment efficacy, residual disease, relapse, or possible development of resistance (WHO- 4th Edition - © International

Agency for Research on Cancer; (Classification for Tumours of Haematopoeitic and Lymphoid tissue); Page 32, 4th Edition, 2017)

1.1.2 CML prevalence and incidence

CML accounts for about 15% of adult leukaemia (Deininger et al., 2020) with an annual incidence of 1-2 cases per 100,000 adults with a slight male predominance (Redaelli et al., 2004). The median age at diagnosis is approximately 60 years (Hogland et al., 2015), however CML occurs in all age groups according to the Surveillance, Epidemiology, and End Results Program (SEER) statistics (https://seer.cancer.gov/statfacts/html/cmyl.html). Incidence data for Africa, and particularly for South Africa, is lacking. Data on prevalence is unsubstantial due to being categorised with leukaemia's generally. Unpublished studies as mentioned by Louw, 2012 showed that the median age of presentation for South African patients is lower, with a slightly lower male:female ratio (Louw, 2012). Similar findings have also been reported from other African countries where median ages of 38, 35, and 39 have been reported for newly diagnosed CML patients in Nigeria, Kenya, and Malawi, respectively (Boma et al., 2006, Othieno-Abinya et al., 2002, and Mukiibi et al., 2003).

1.1.3 Treatment response profiles and molecular monitoring

CML has been described to occur in three phases; chronic phase (CP), accelerated phase (AP) and the blast crisis phase (BC). When untreated, the patients would progress to the next phase, respectively, in 3-5 years on average (Deininger et al., 2020). The efficacy of the treatment is measured by response profiles which are used for the management of CML patients. The ELN recommendations has defined three types of responses, namely, Haematologic, Cytogenetic and Molecular Responses **(Table1.1)**.

Table 1.1. ELN guidelines definitions of Haematologic, Cytogenetic and Molecular Response

Response by Type	Definitions
a) Haematologic • Complete (CHR)	• WBC < 10×10^9/L • Basophils < 5% • No myelocytes, promyelocytes, myeloblasts in the differential • Platelet count < 450×10^9/L • Spleen nonpalpable
b) Cytogenetic • Complete (CCyR) • Partial (PCyR) • Minor (mCyR) • Minimal (minCyR) • None (noCyR)	• No Ph+ metaphases • 1% to 35% Ph+ metaphases • 36% to 65% Ph+ metaphases • 66% to 95% Ph+ metaphases • > 95% Ph+ metaphases
c) Molecular • Complete • Major (MMR)	• Undetectable *BCR::ABL1* mRNA transcripts by real time quantitative and/or nested PCR in two consecutive blood samples of adequate quality (sensitivity > 10^4) • Ratio of *BCR::ABL1* to *ABL1* (or other housekeeping genes) ≤ 0.1% on the international scale

Table adapted from ELN 2020 (Hochhaus et al., 2020)
Abbreviations: CHR, complete hematologic response; WBC White Blood Cell, CCyR, complete cytogenic response; PCyR, partial cytogenetic response; Ph+, Philadelphia chromosome positive; MCyR, major cytogenetic response; minCyR, minor cytogenetic response; noCR, no cytogenetic response; CMR, complete molecular response; PCR, polymerase chain reaction;. MMR, major molecular response.

To date molecular monitoring of the *BCR::ABL1* transcript has become standard of care as it aids in prognosis, help with monitoring treatment response and predict relapse. Guidelines by the Europe against Cancer and National Comprehensive Cancer Network (NCCN) and European LeukemiaNet (ELN) state that qRT-PCR should be performed at diagnosis prior to treatment in order to determine baseline levels of the *BCR::ABL1* transcript (Bauer & Romvari, 2012). Thereafter, molecular monitoring should be continued at 3 months intervals until a complete cytogenetic response (CCyR) or major molecular response (MMR) has been achieved **(Table 1.1)**.

Molecular responses are assessed according to the International Scale (IS) as the ratio of *BCR::ABL1* mRNA transcripts to *ABL1* transcripts. These are recorded on a graph as the log percentage decrease in *BCR::ABL1* transcript, during therapy, where 1% corresponds to a decrease of 2 logs; 0.1% to decrease of 3 logs; 0.01% to a decrease of 4 logs, 0.0032% to a decrease of 4.5 logs and 0.001% corresponds to a decrease of 5 logs **(Table 1.2)** (Hochhaus et al., 2020). A 3 log reduction in *BCR::ABL1* transcript from a standardized baseline value needs to be observed to define a MMR. The chances of transforming to advance phases of the disease are thus minimal and the patient has a 100% chance to remain progression free (Jabbour & Kantarjian, 2020).

Table 1.2 ELN recommended *BCR::ABL1* transcript levels for scoring molecular responses

	MMR	**MR^4**	**MR$^{4.5}$**	**MR^5**
BCR::ABL 1 transcript level on the IS	≤0.1%	≤0.01%	≤0.0032%	≤0.001%

Abbreviations: BCR – Breakpoint cluster region, ABL- Abelson, IS – International Standard, MMR – Major Molecular Response, MM – Molecular response
Adapted from (Hochhaus et al., 2020)

The molecular test results obtained could also help patients understand their response to treatment and whether treatment will be continued as is or considered to be changed especially since treatment for CML patients is lifelong. An increase in *BCR::ABL1* transcript after therapy might predict loss of treatment response, relapse or the occurrence of secondary mutations in the *BCR::ABL1* fusion gene. However, it may also be indicative of non-adherence, drug-drug or drug-food interactions, different laboratory, or procedures and/or assay errors (Braun, Eide, & Druker, 2020).

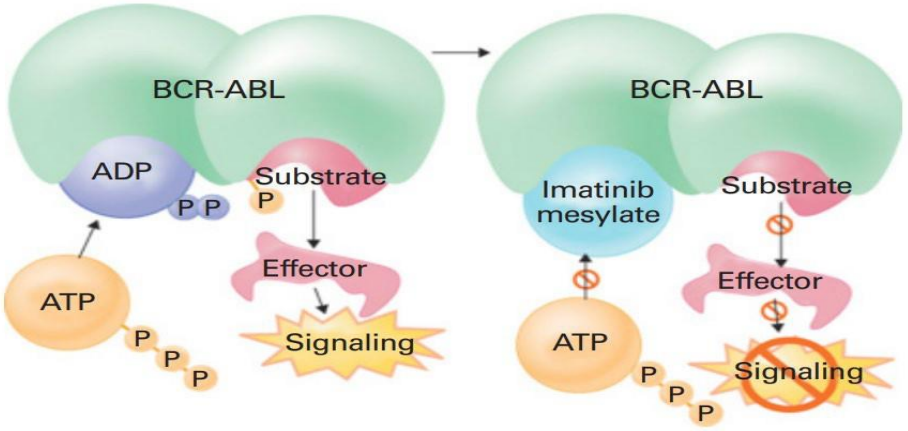
Patients are considered to have reached CCyR when no Philadelphia chromosomes are observed on the bone marrow. A complete cytogenetic response (CCyR; 0% Ph-positive metaphases by cytogenetics) is equal to a negative FISH test (+/- 2%) and *BCR::ABL1* transcripts IS <1%. A partial cytogenetic response (PCyR; ≤35% Ph-positive methaphases) is equal to *BCR::ABL1* transcripts ≤10% IS (Jabbour & Kantarjian, 2020).

1.1.4 CML management and treatment

The major goal for CML clinical management is to prevent progression from chronic to accelerated and blast crisis phases. Since the introduction of the first tyrosine kinase inhibitor (TKI), Imatinib (Glivec), in 1998, this TKI revolutionised CML treatment decreasing the mortality rate from 10%-20% to 1%-2% (Jabbour & Kantarjian, 2020). Imatinib mesylate (Glivec) was the first TKI approved by the Food and Drug Administration (FDA) until the market exclusive usage of Imatinib lapsed in most countries in 2016. Since then, Imatinib generics entered the market and due to its readily available option, many countries were encouraged to use it to lower health care expenses. Switching from branded Imatinib to generic imatinib is substantially cost saving. However, the efficacy, safety, and quality of generic imatinib as well as the impact on patient care were an issue of concern. A study done at M.D Anderson Cancer Centre in Houston Texas reported that 38 patients, who were switched from the branded Imatinib to generics demonstrated a stable molecular response in 89% of the patients for a median of 19.4 months, 8% showed improvements and 3% of the patients worsened while no patient lost MMR (Abou Dalle et al., 2019).

The FDA approval of Imatinib (Glivec) was followed by Dasatinib (Sprycel) and Nilotinib (Tasigna). The latter TKI's were developed for use when patients exhibit a slow response to Imatinib (Glivec), relapse and/or development of *ABL1* tyrosine kinase domain (TKD) mutations which lead to Imatinib resistance (Yang & Fu, 2015). Furthermore, they are also more potent, and their efficacy is more specific. These TKI's act by blocking the binding of ATP to the BCR::ABL1 tyrosine kinase thereby inhibiting its activity **(Figure 1.3)**.

Dasatinib is a second generation TKI and inhibits the BCR::ABL fusion protein through competitive inhibition of the ATP binding sites. Dasatinib as a first line therapy for newly diagnosed CML-CP patients compare to Imatinib has been demonstrated (Yang & Fu, 2015). Nilotinib targets the BCR::ABL1 kinase domain with greater potency and efficacy, and does so by binding to the inactive configuration of the *ABL* KD , with P-loop folding over the ATP binding site, then the activation-loop block the substrate binding site to disrupt the ATP-phosphate-binding site and inhibit the catalytic activity of the enzyme (Yang & Fu, 2015). Ponatinib works by its ability to make multiple points of contact so that individual mutations have less effect on the overall binding affinity of the drug (Yang & Fu, 2015). Bosutinib inhibits BCR::ABL1 mediated signalling at nanomolar concentrations as there seems to be a direct correlation between *BCR::ABL1* over expression and the generation of mutations.



Abbreviations: ADP- Adenosine diphosphate, P – phosphate groups, ATP – adenosine triphosphate,

Extracted from (Lee et al., 2013).

Figure 1.3. Mode of Action of Imatinib on BCR::ABL1

Both the ELN and NCCN guidelines and recommendations are widely used for the treatment and management of CML. In SA, local experts from the private and public sector formed a CML expert panel that meets twice a year to discuss issues regarding CML (V. J. Louw, 2012). The panel had also produced a set of recommendations for the management of CML in 2011, where they consider monitoring, stem cell transplantation, TKI intolerance, management of HIV coinfection, pregnancy, and breastfeeding.

Interestingly, it has been reported that in the South African population responses to treatment is delayed (V. J. Louw, 2012). Contributing factors to the delayed response could be limitations of CML management and treatment which are often socio-economic and political. Louw (2012), reported on the various challenges which are faced regarding drug accessibility, co-infections, and lack of local recommendations suitable for the SA setting (Louw, 2012). The burden of HIV, a high prevalence of other communicable and non-communicable disease, and the focus on preventative and primary healthcare, all contributed to resource limitations to expensive cancer drugs such as the TKIs (Louw, 2012). In a similar study performed in India in 2019, they found that availability and efficacy of drugs; environmental factors; nutritional factors; and unknown genetic factors all contributed to differences in haematological cancers in developing countries to that of the developed world (Malhotra et al., 2019). In addition, the availability of *BCR::ABL1* testing, secondary and tertiary TKI availability, compliance, toxicity, ensuring standard of care treatment and monitoring are all challenges faced.

The South African study also highlighted that only a few patients who had medical insurance could afford the treatment, monitoring and associated costs while some insurers refused to cover the high costs. Moreover, Black South Africans, who were the predominant ethnic group, simply did not have access to treatment due to minimum household incomes (Louw, 2012). In 2002, the Glivec International Patient Assistance Program (GIPAP) had been set up as a partnership between Novartis and the MAX Foundation where TKIs are supplied to poor patients at no cost. Five hundred patients in SA were active on the program but the project had later ceased (Dr. F. Bassa, personal communication, 19/03/2021).

Interestingly, (Sissolak et al., 2015) investigated treatment outcomes in patients treated with Imatinib at Tygerberg Hospital in Cape Town, SA and found that while delays in initial responses of the patients on treatment was observed, the overall treatment was similar to those of developed countries (Sissolak et al., 2015). This would indicate that CML patients, who did not show resistance, stood a fair chance at survival if there was access to treatment and monitoring. The research and medical advancement regarding CML management recommendations has led to the life expectancy of CML patients matching CML free individuals. Strategies are being identified which focuses on the quality of life and the possibilities of treatment free remission (TFR).

1.1.5 CML and drug resistance

Cancers can develop resistance to therapies which in turn result in an increased number of drug resistant cancers. Mechanisms that underlie drug resistance in cancers are not well understood. Various categories of mechanism that can promote or enable drug resistance includes drug inactivation (interactions that modify, partly degrade the drug), drug target alteration, drug influx/efflux, DNA damage repair, cell death inhibition and inherent tumour cell heterogeneity as well as epigenetic modifications that may contribute to development of cancer progenitor cells that are not killed by normal cancer therapies (Mansoori et al., 2017). These mechanisms can either act independently or in combination through various signal transduction pathway.

In the case of CML, Imatinib specifically targets the BCR::ABL1 protein and induces remission. Unfortunately, a proportion of CML patients (20-30%) treated with Imatinib develop resistance at some point during therapy (Cortes et al., 2012). Some patients may fail to respond to initial treatment with Imatinib (primary resistance), while others stop responding with prolonged therapy after an initial response (acquired resistance).

The ELN defines primary resistance as, not achieving complete haematological response (CHR) by 3 months, no cytogenic response by 6 months and no major molecular response by 12 months. Secondary resistance is defined as the patient losing the previously achieved haematologic or cytogenetic response and patients exhibiting a consecutive 1 log increase in transcript level after achieving MMR **(Table 1.3)**.

Table 1.3 ELN guidelines recommended resistance criteria in accordance with IS%

	Optimal	**Warning**	**Failure**
Baseline	NA	High Risk	NA
3 months	≤10%	>10%	>10% if confirmed within 1-3 months
6 months	≤1%	>1-10%	>10%
12 months	≤0.1%	>0.1-1%	>1%
Any time	≤0.1%	>0.1-1%	>1% resistance mutations

Abbreviations: NA – Not applicable

Adapted from (*Hochhaus et al., 2020)*

The ELN guidelines for CML treatment provide resistance criteria according to IS% relative to time **(Table 1.3).** An optimal response is having achieved less than or equal to 10% IS at 3 months; less than or equal to 1% at 6 months and less than or equal to 0.1% at 12 months on treatment. A warning to treatment resistance is defined as having more than 10% IS at 3 months; more than 1-10% IS at 6 months and more than 0.1-1% IS at 12 months. Moreover, treatment failure is defined as having more than 10% IS at 3 months; more than 10% at 6 months and more than 1% at 12 months. The guidelines allures to the possibility of resistance mutations if IS% is more than 1% at any time, after being on treatment for 12 months (Hochhaus et al., 2020). According to the patients characteristics, comorbidities and tolerance, the monitoring milestones of *BCR::ABL1* transcript levels by IS% at 3,6 and 12 months, determine whether the current treatment should be continued, changed or carefully considered for continuation or change (Deininger et al., 2020)

It is important to note that these recommendations are based upon a European setting and although they are followed, many discrepancies arise within a setting such as at TBH in SA (Dr. F. Bassa, personal communications, 19/03/2021).

1.1.6 Secondary mutations

Several mechanisms of Imatinib resistance have been proposed that account for loss of Imatinib efficacy in patients with CML. The most common mechanism for Imatinib resistance (60%) are due to mutations in the *BCR::ABL1* kinase domain that compromise inhibitor binding, by interference with critical hydrogen bonds and subsequently lead to increased kinase activity or impairing the ability of the BCR::ABL protein to bind to the ATP protein site of the inhibitor molecule (Kotagama et al., 2016, Quintás-Cardama, Kantarjian & Cortes, 2009). Over the years several mutations have been identified within the kinase domain of *BCR::ABL1* and their significance on phenotype have been evaluated with different mutations conferring different levels of resistance.

Mutations within this domain **(Figure 1.4)** is observed in 30-50% of chronic phase patients who had developed secondary resistance to Imatinib. An increase in frequency of the mutations is observed in the accelerated phase (AP) and usually in blast crisis (BC) patients. Studies suggest that CML-BC is associated with dramatic changes in the leukaemia cell phenotype but molecular mechanisms for this change are still unknown (Yang & Fu, 2015). More than 100 different mutations involving different amino acids have been reported (Jabbour & Kantarjian, 2020), but less than 20 of these mutations are frequently identified in clinical practice (Tanaka et al., 2011).

Reports show that the T315I confers complete resistance to Imatinib, Dasatini, Nilotinib, and Bosutinib (Yang & Fu, 2015). The T315A, F317L/I/V/C, F359V/C and Y253H mutations confers to Nilotinib. E255K/V, F359C/V, Y253H, and T315I mutants are mostly associated with disease progression and relapse. Bosutinib is effective in patients who are resistant to Dasatinib (F317L) and Nilotinib (Y253H, E255k/V and F359C/I/V). It did, however, have limited activity against the F318L mutant. T315I, G250E, and V299L mutants are resistant to Bosutinib. Ponatinib would work on mutants that are resistant to Dasatinib or Nilotinib, including E255V, Y253H, and F359V and the infamous T315I. **Table 1.4** shows a list of the most reported mutations and the TKIs they confer resistance to.

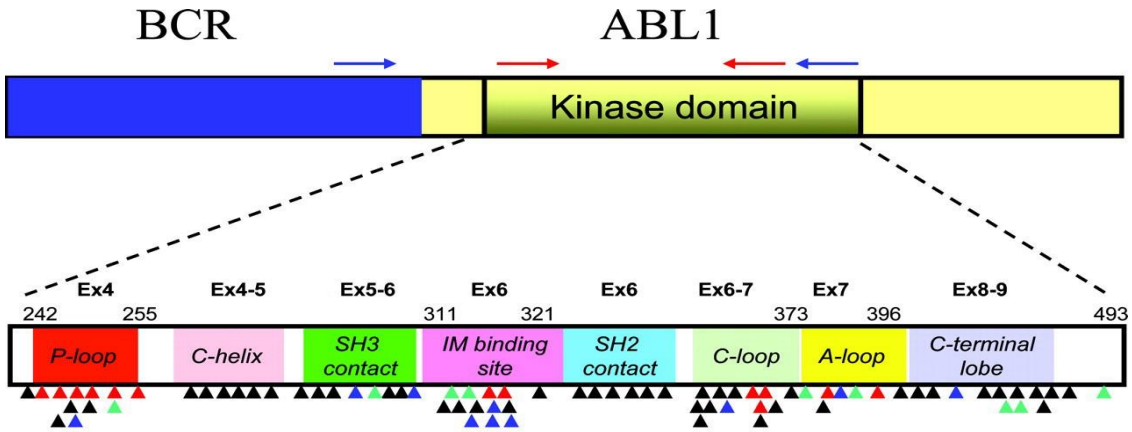


Figure 1.4. Schematic of *BCR::ABL* fusion transcript and the location of reported KD mutations and alternate transcripts. Common areas for location of BCR (blue box) and ABL1 (yellow box including KD) primers for nested PCR (blue arrows) and BCR::ABL1 KD PCR (red arrows) used for direct sequencing assays are indicated. Coloured triangles indicate the location of KD mutations reported in TKI-resistant samples (black for Imatinib, green for Nilotinib/Imatinib, blue for Dasatinib/Imatinib, and red for all three TKIs). The KD subdomains, exons, and amino acid numbers are shown.

Abbreviations: P-loop: phosphate binding loop; IM binding site: imatinib binding region; C-loop: kinase catalytic domain; and A-loop: activation loop.

Extracted from: (Jones et al., 2009)

Table 1.4 Most reported mutations and the TKI therapy they confer resistant to

Amino acid variation	Drug resistance consequence
M244V	Imatinib
L248V	Imatinib
G250E	Imatinib
Q252H	Imatinib
Y253H	Imatinib, Nilotinib
E255V	Imatinib, Nilotinib, Bosutinib
L273M	Imatinib
E255K	Imatinib, Nilotinib, Bosutinib
D276G	Imatinib
E279K	Imatinib
T277A	Imatinib
F311L	Imatinib
V299L	Dasatinib, Bosutinib
T315I	Imatinib, Nilotinib, Dasatinib, Bosutinib
T315A	Dasatinib
T315M	Ponatinib
T315L	Ponatinib
F317L	Imatinib, Dasatinib
F317I	Imatinib, Dasatinib
F317V	Imatinib, Dasatinib
F317C	Imatinib, Dasatinib
M351T	Imatinib
E355G	Imatinib

F359C	Imatinib, Nilotinib
F359I	Imatinib, Nilotinib
F359V	Imatinib, Nilotinib
E379K	Imatinib
L384M	Imatinib
L387M	Imatinib
L387F	Imatinib
H396R	Imatinib
H396P	Imatinib
E459K	Imatinib
F486S	Imatinib

Adapted from (Abruzzese et al., 2019)

Fortunately, science has progressed to the extent where bioinformatics allows the structural modifications in the BCR::ABL1 mutant protein to be visually analysed in 3D and investigated (Iqbal et al., 2020). Structural genomics and the contributions thereof is deposited to the Protein Data Bank (PDB), which is an electronic repository of 3D structures of proteins and nucleic acids (Bishop et al., 2008).(Chandrasekhar et al., 2019). This is performed in *silico* to correlate structural and functional analysis of the *BCR::ABL1* gene. Four main mutational regions for TKI resistance have been reported: the P-loop, SH3 contact, SH2 contact and A loop (Iqbal et al., 2020 & Hughes et al., 2019). One example of the type of analysis that is possible as a result is the Molecular dynamic (MD) simulation of wild-type and mutated ABL1 structures. (Chandrasekhar et al., 2019) subjected ABL structures to MD simulations to investigate the energy variations, and found that the T315I showed reduced energy values, which is suggestive of non-functional proteins in the patient (Chandrasekhar et al., 2019). They also measured the PDB Sum conformational fluctuations in ABL mutated structures compared with normal ABL structures and found that conformation changes had occurred because of the T315I mutations, where they observed fewer helices and β turns.

Recently, Iqbal et al. 2020 performed studies on the molecular structure of mutations specifically found in the kinase domain of the *BCR::ABL1* gene conferring drug resistance (Iqbal et al., 2020). Briefly, they crystallographically determined human ABL kinase protein structure in complex with Nilotinib was retrieved from the worldwide archive of structural data of biological macromolecules, (Research Collaboratory for Structural Bioinformatics RCSB's PDB (ID: 3cs9)) and was viewed using Delano Scientifics PyMOL. The 3D structure of the protein-ligand complex is given as an input in PDB format, and the software displays the interacting residues and bonds (Iqbal et al., 2020).

LigPlot+ is another program that can be used to check and schematically plot the polar and hydrophobic interactions between the BCR::ABL kinase protein and the inhibitor molecule (Cortes et al., 2012). They performed a schematic 2D binding mode representation with the ligand, Nilotinib generated by Ligplot, and they found that the infamous T315I to be one of the primarily interacting residues. The use of bioinformatics structural analysis is therefore exceptional.

The remaining 40% of Imatinib resistant cases are suggestively independent of *BCR::ABL1* kinase domain mutation mechanisms (Eide et al., 2016). Thus, there are various other mechanisms that promote or enable drug resistance independent of BCR::ABL such drug influx/efflux, epigenetic modifications and methylation (Braun et al., 2020).

1.2 Present study

The need to identify these mutations within our South African CML cohort is very relevant and important to clinical treatment as it would aid in the treatment decision making processes. While most diagnostic laboratories implement and screen for secondary mutations within the BCR::ABL1 kinase domain, no incidence data related to these mutations is available in the South African context. Thus, we aim to implement and validate the diagnostic screening of the ABL1 kinase domain and explore known and novel secondary mutations and characterise them within our unit at Tygerberg Hospital, with hopes of extending this to other parts of SA as well. Variants identified will be further subjected to *in silico* studies (computational simulation) to aid in identifying potential targets for therapeutic strategies and understanding the underlying basis of the disease.

Chapter 2
Methodology

2.1 Study Cohort

The study was conducted within the Division of Haematology, at the Molecular Haematology laboratory of the National Health Laboratory Services (NHLS) and Stellenbosch University's Faculty of Medicine and Health Sciences, Tygerberg. For Objective 1, a retrospective audit of CML diagnosed and treated cases at Tygerberg Hospital for a period of 8 years (2013-2020) was conducted. Objective 2, which is the laboratory-based portion of this project focused on the mutational screening of the *BCR::ABL1* kinase domain for future diagnostic implementation. This study was approved by the Health Research Ethics Committee (HREC) of the faculty of Medicine and Health Sciences of the University of Stellenbosch (HREC: S19/10/272) **(Appendix VII)**

2.1.1 Target and Study population

From the diagnosed and treated CML patients reviewed, cases characterized for resistance were grouped based on the recommendations by European LeukemiaNet 2009 guidelines. Patients not achieving complete haematological response (CHR) by 3 months, no cytogenetic response by 6 months or Major Molecular Response (MMR) by 12 months was categorised as primary resistant and patients losing the previously achieved hematologic or cytogenetic response or patients exhibiting a consecutive 1 log increase in *BCR::ABL1* mRNA transcript level after achieving an MMR was categorized as secondary resistant patients. The resistant patients were further categorised as Failure or Warning. Patients were categorised as warning if they exhibited more than 10% IS at 3 months; more than 1-10% IS at 6 months; more than 0.1-1% at 12 months or more than 0.1-1% IS at any time. Patients were categorised as failure if they exhibited more than 10% if confirmed within 1-3 months; if they showed more than 10% at 6 months on treatment; more than 1% at 12 months on treatment.

2.1.2 Inclusion Criteria

Patients were included if:

- Cytogenetics confirmed the translocation between chromosomes 9 and 22;
- are in the accelerated and blast crisis phase at diagnosis.

- they had the typical *BCR::ABL1* transcripts b3a2 and b2a2 ;
- they either failed to respond to 1st line Imatinib therapy or have an increase in *BCR::ABL1*mRNA transcript levels which would lead to MMR loss or those that respond sub optimally on 1st line Imatinib therapy;
- Those that fail haematological or cytogenetic response during 2nd line TKI therapy.

2.2 Objective 1: Audit of CML cases

One hundred and sixty-five cases of CML were diagnosed in the reviewed period between 2013 and 2020. Data was reviewed with the aid and input from a team of haematologist, clinicians, and scientists at the Tygerberg Hospital Haematology Clinic in association with the Division of Haematopathology and the National Health Laboratory Services (NHLS). A CML cohort database was created to be used in further sub studies which forms part of a larger CML based study focusing on determining alternative contributions that underlie resistance mechanisms in CML. Descriptive statistics on newly diagnosed CML and follow up cases and Imatinib resistance cases was generated.

2.2.1 Minimal clinical data collection and management

The Trak-care laboratory information management system (v04.16.04.814, Laboratory System Technologies, Johannesburg, SA, L6.10, InterSystems Corporation, Cambridge, MA, USA) was used to obtain the clinical and diagnostic data.

Response profiles were obtained from routine monitoring of transcripts levels using quantitative real-time PCR, via the Cepheid based p210*BCR::ABL1* Ultra test used on the Cepheid GeneXpert® platform (GeneXpert®, Cepheid, California, USA). Treatment response profiles (MMR, CMR, MCyR, CCyR) were generated and analysed with the help of a haematological pathologist and classified in accordance with the European LeukemiaNet guidelines. Time-to- respond profiles (MMR, CMR, MCyR) were also determined. The patients who progressed to the accelerated phase or blast crisis phase of CML were classified as non- responders.

Relevant permissions were obtained to access and use the clinical and laboratory data and unique participant numbers was assigned to ensure that the patient's individual information remains de-identified and protected in accordance with POPIA, NHLS, TBH and SU privacy regulatory policies **(Appendices VII, VIII and IX).**

2.3 Objective 2: Mutational screening of *BCR::ABL1* kinase domain and evaluation and implementation of diagnostic sequencing assay

2.3.1 Study population:

A total of 20 voluntary consented participants who attends the haematology/Oncology (X-Block) clinic at Tygerberg Hospital for routine diagnostics were recruited for the mutational screening. Therefore, no additional blood was requested as the remaining blood from routine diagnostic testing was used for mutation screening. It should be noted for internal diagnostic evaluations and validations within the NHLS, a minimum of 20 samples is required according to GPL1530 Verification/Validation of Quantitative and Qualitative methods SOP. **(Appendix II).**

2.3.2 Sample Collection

One EDTA tube (Vacutainer, USA), of 3-4ml (per tube) of peripheral blood for routine diagnostic testing was recruited for the mutation screening when patients visited the Haematology Oncology clinics (X-Block, Tygerberg Hospital). This was received by the NHLS Molecular Haematology diagnostics laboratory for processing. Part of the protocol is to check White Cell Counts (WCC) to determine the volume of blood to be used for further processing **(Appendix I).** Of the 20 patients, only five patients had baseline samples, for whom two samples were screened (at baseline and after 12 months on treatment).

A buffy coat was prepared from the left-over blood by centrifugation (MSE MISTRAL 1000, MSE Scientific, Leicestershire, England). Centrifugation was done for 15 minutes at room temperature, using a speed of 1200 rpm. The white buffy coat, found in the centre of the centrifuged sample containing concentrated leukocytes was stored in a pre-labelled 1.5ml Eppendorf tube for downstream DNA extraction purposes.

2.3.3 DNA extraction

Genomic DNA was extracted using the QIAamp DNA Mini Kit as per manufacturer's instruction (QIAGEN) and NHLS diagnostics SOP **(Appendix IV).** In brief, a volume of 20 µl QIAGEN Protease K was transferred into the bottom of a 1.5 ml microcentrifuge tube (Whitehead Scientific, USA) along with a 200 µl aliquot of buffy coat (fresh or thawed). A volume of 200 µl Buffer AL was added to the sample and the mixture was mixed by pulse-vortexing (Rotamixer Deluxe, Hook and Tucker instruments Ltd, UK) for 15 seconds. The mixture was then incubated at 56°C on a heating block (Accublock, Labnet International, USA)

for 10 minutes. Thereafter, the 1.5 ml microcentrifuge tube containing the lysed mixture was quickly centrifuged to remove condensation droplets from the inside of the lid. This was followed by the addition of 200 µl of ethanol (100%) to the sample, which was again mixed by pulse vortexing for 15 seconds. Thereafter, this mixture was added to the QIAamp Mini spin column placed within a 2 ml collection tube and subsequently centrifuged (Prism Microcentrifuge, Labnet, USA) at 8000rpm for 1 minute. Following that, the QIAamp Mini spin column was placed in a clean 2 ml collection tube and the tube containing the filtrate was discarded. Subsequently, 500 µl of Buffer AW2 was added to the QIAamp Mini spin column and was centrifuged at 8000 rpm for 1 minute. The old collection tube containing the filtrate was discarded while the QIAamp mini column was once again placed in a clean collection tube and centrifuged at full speed for 1 minute. Thereafter the QIAamp Mini spin column was placed in a clean 1.5 ml microcentrifuge tube and the collection tube with the filtrate was discarded. Lastly, 200 µl of Buffer AE was added to the QIAamp Mini spin column which was then incubated at room temperature (15 -25°C) for 1 minute, and then centrifuged at 8000rpm for 1 minute. DNA quantification was performed on isolated DNA using the BioDrop µLITE spectrophotometer (ISOGEN Life Science, Netherlands) to measure the purity and concentration of the DNA prior to use in PCR.

2.3.4 Polymerase Chain Reaction (PCR) Amplification

A standard PCR reaction was performed using previously described primer sets, as shown in Table 2.1. Prior to synthesis, the primers' locations were verified using published reference sequence (NCBI reference sequence: NM_005157.6) obtained from the NCBI database (https://www.ncbi.nlm.nih.gov/) and NCBI Blast programs. Additionally, each primer sequence was subjected to self-complementarity, primer-primer complementarity tests and melting temperature compatibility checks using Primer 3 Plus primer design (https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/) and DNAMAN X (Lynnon Biosoft, Australia. Version 10.3.3.126) software before being synthesized by Intergrated DNA Technologie, Inc. (IDT), supplied by Whitehead Scientific (Pty) Ltd.

For the amplification of each DNA sample, 200ng of extracted DNA was used as a template in a 25µL reaction mix, which contained the following reagents: 0.5µl (0.5ng/ul) of each primer (Table 2.1.), 0.125µl Taq Polymerase (Promega), 1.5uL MgCL2. 5µl buffer, 0.5µL dNTP mix and ddH2O (varies) to a total volume of 25µL. Amplification was done in a thermal cycler (Bio-Rad T100 Thermal Cycler, California, USA) using the following thermal cycling profile: the holding step was at 95°C for 10 minutes for 1 cycle, followed by denaturation, annealing and

extension steps at 95°C for 30 seconds, annealing temperature for respective primers for 30 seconds and 72°C for 45 seconds, respectively. The final holding step was at 22°C for 1 minute.

Table 2.1 PCR primers, annealing temperatures and template product size

Exon		Forward and Reverse Primers	Annealing Temp (C)	Product size (bp)
4	F R	GCGAGTAACTTAGAGCACACG CAGCAGGCAGGAGGTAGACT	59.8	559
5	F R	ACCTGTCTGCAGCAATGTGG ATTCCAACGAGGTTTTGTGC	56.1	339
6	F R	CGGAGCCACGTGTTGAAGT TAGGCTGGGGCTTTTTGTAA	56.1	298
7	F R	CCATTGTCAGCATTGCACCT AAGGGAAGAGCAAGAAAGAGG	57.3	511
8	F R	CCATCCCTTCTGAGGTCTGC TGGAATGGAGAGAGCTCATGG	60	367
9	F R	TCCTGCCAGCATCTAACGTCT ACCAGGAAGTGAACCGAATG	57.4	380
10	F R	CACACCAAGCCAACACCAGT TTCCGAAGCTGAACCATTGAC	60	609

Abbreviations: F, forward; R, Reverse, Temp, Temperature; bp, base pairs, C, Celsius

2.3.5 Verification of PCR product – Agarose Gel Electrophoresis

PCR amplified products were separated on a 2% agarose gel for verification of successful amplification. The gel was made up by a 2g Agarose powder (SeaKem LE Agarose, Lonza Rockland, ME USA) and 1X TBE buffer (Appendix V), which was mixed and microwaved until dissolved and poured into electrophoresis tray (BioRad USA). The combs were inserted, and gel was allowed to set. The agarose gel was submerged in 1X TBE running buffer in an electrophoresis chamber as wells were subsequently loaded with a mixture of 2µl Novel Juice (Genedirex, USA) and 8µl of amplified PCR product along with a size marker. Each product was visualised for the correct size using the gel documentation system (Gel Doc XR+, BioRad, USA)

2.3.6 Bidirectional Sanger sequencing

Bidirectional sequencing by the Sanger method remains the most widely used and recommended technique. PCR amplification of samples are cleaned and subjected to bi-directional sequencing for analysis.

The PCR products were cleaned using Exonuclease 1 and Calf Intestinal Alkaline phosphate (New England Biolabs) (Mix 1). The sequencing reaction was done in accordance with the NHLS Molecular Haematology diagnostic laboratory protocol **(Appendix VI)** using the AB Big Dye Sequencing Kit (Thermo Fisher Scientific, USA) (Mix 2). In brief, 1µl mix of the two enzymes each was added to 3µl of PCR product and subsequently incubated at 37ºC for 15 minutes and 80ºC in a thermal cycler (Bio-Rad T100 Thermal cycler, California, USA). The incubated mix 1 was then centrifuged at 8000rpm for one minute. Thereafter, 3 µl of respected primers (diluted) was added to the 4µl template followed by 4µl of Mix 2 **(Appendix VI)**.

Bi-directional sequencing was completed on the ABI Prism 3100 automated sequencer using the PCR specific cycle sequencing primers at the Central Analytic Facility (CAF), University of Stellenbosch, South Africa. The BioEdit sequencing software (Applied biosystems, USA) was used for analysis of chromatograms and sequence alignments. Sequences were extracted in FASTA format which were compared against BLAST database at the National Centre for Biotechnology Information (NCBI). The obtained nucleotide sequences were translated into protein sequence and compared with the BCR:;ABL1 protein sequence from the NCBI database (NP_005148.2) and the reference sequence of the *ABL1* gene, (ensemble ID ENSG00000097007). All nucleotide substitutions were confirmed by bi-directional sequencing. To identify mutations, multiple pairwise sequence alignments were carried out using DNAMAN, which allowed for the observation of any abnormalities.

2.4 Structural analysis

All variations identified were subjected to suitable and available informatics tools. The type of computational tool used was dependent on the variations observed. These tools allow the investigation of structural changes on the function. These include:

- Ensembl Genome Browser (https://www.ensembl.org/index.html)
- ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/);
- SynMICdb (http://synmicdb.dkfz.de/rsynmicdb/)
- synVEp v 1.0 (https://services.bromberglab.org/synvep/home) and
- MutationTaser (https://www.genecascade.org/MutationTaster2021/#transcript).
- Population frequencies was obtained from
    - the 1000 Genomes project (https://www.internationalgenome.org/)
    - and gnomAD (https://gnomad.broadinstitute.org/).

The ABL1 structure was retrieved from the Protein Data Bank (PDB: 1OPL) (http://www.rcsb.org/pdb/explore/explore.do?structureId=1OPL) with the resolution of 3.42 Å and was loaded into the Molecular Operating Environment software tool (MOE 2011.10 Chemical Computing Group Inc), ignoring all water molecules and heteroatoms. The structure was subjected to protonation followed by energy minimization under MMFF94x force field (Merck Molecular Force Field) 36,37 to an RMSD of 0.05 where the implicit solvated environment was specified, and the stabilized conformation was saved in PDB format.

2.5 Statistical analysis

The statistical analysis was performed using Microsoft Excel and the statistical software package SPSS v20 (SPSS Inc, Chicago, Illinois) and summarised and analysed in conjunction with the Department of Statistics, University of Stellenbosch. Categorical data were reported as percentages and frequencies. The Wilcoxon Rank Sum Test, Mann-Whitney and Chi-square tests as well as the Spearman's and Pearson's correlation coefficients. Statistical significance was accepted as a two-sided *p*-value of 0.05.

Chapter 3

Results

The present study was composed of three parts. Firstly, an audit was conducted on a CML cohort to identify trends based on age, gender, and prognosis. Secondly, mutational/variant screening of the *ABL1* gene, for kinase domain variants was performed on 20 selected CML patients with resistance profiles. Lastly, *in silico* structural analysis was performed on identified sequence variants to determine whether amino acid changes have any effect on the structure or function of a protein without conducting functional studies.

3.1 Incidence of CML cohort at TBH

Part one of the present study was a retrospective audit which showed a total of 165 CML patients were diagnosed and treated at Tygerberg Hospital over a period of 8 years, spanning from January 2013 until December 2020. Data was collected from CML patients at diagnosis and/or during TKI treatment. From the 165 CML patients, 34.5% (n=57) CML patients showed possible resistance and fell within either the failure or warning stages according to ELN recommendations. As per NHLS new test evaluation guidelines, and pending on biospecimen availability, a minimum of 20 of these CML patients were selected for sequence variant/mutational screening. As backup buffy coat storage has only been implemented since 2018, the availability of biospecimens associated with these resistant cases were limited.

3.1.1 Demographical data

3.1.1.1 Gender

Descriptive statistics related to gender showed that 46.1% (n=76) of the patients were found to be female and 53.9% (n=89) were male **(Table 3.1).**

Table 3.1. Gender proportions of CML cohort

Gender	Frequency (N)	Percent (%)
Female	76	46.1
Male	89	53.9
Total	165	100

3.1.1.2 Patient per district distribution

As Tygerberg Hospital is an academic hospital in the Northern Suburban area of the Western Cape, the patients were from 69 different locations in and around the Northern district. The largest number of patients per districts were from Khayelitsha (19), Delft (10) and Strand (9), respectively. The rest of the locations had 8 or less patients.

3.1.1.3 Age

Quantitative analyses were performed to investigate the age of the cohort. The youngest patient was two years old, and the oldest patient was 88 years old, with a range (max-min) of 86. The average age of the patients was 45-46 years old **(Table 3.2)**.

Table 3.2. Age distribution of CML cohort

Min age	2
Max age	88
Range	86
Average	45.59

Abbreviations: Min, minimum; Max, maximum

Furthermore, viewing the data graphically, the highest age peaks between 40-60 years **(Figure 3.1**)

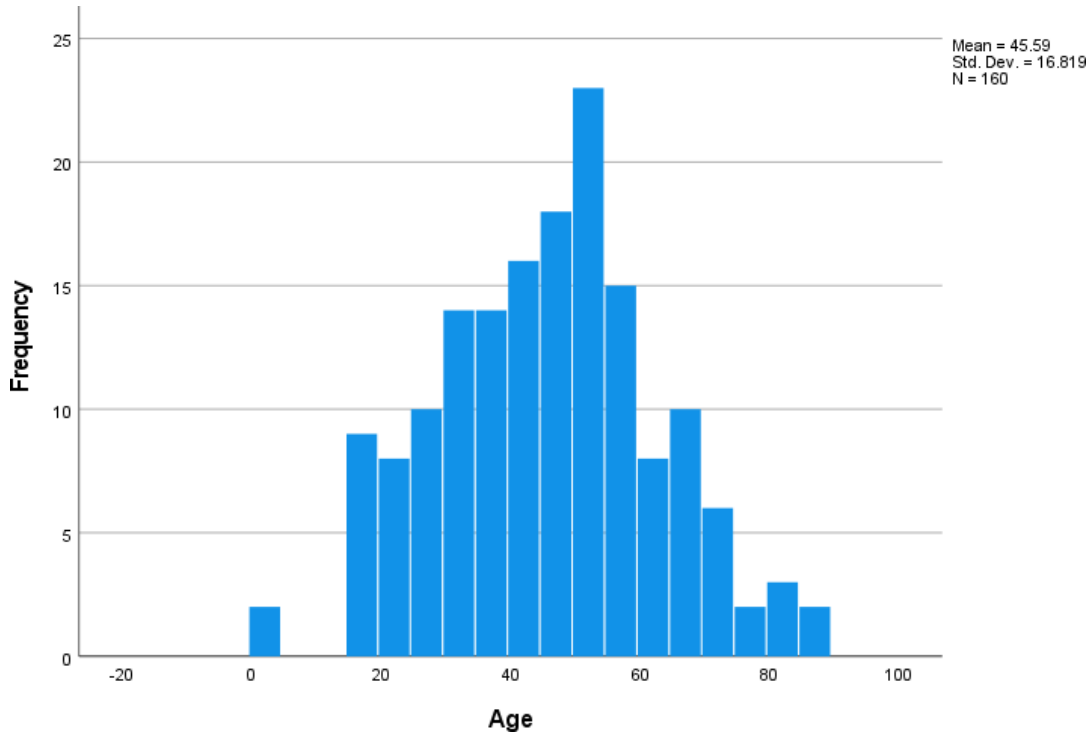


Figure 3.1. Graphical distribution of age groups of CML cohort

3.1.1.4 Patient admissions per year

The frequency of patients that were admitted per year from 2013 till 2020 were also calculated. Data revealed that highest number of patients was admitted during 2014 at 22.4% and thereafter in 2013 at 20.6%. The lowest year of admissions were in 2018 at 10.9% **(Table 3.3)**.

Table 3.3. Admission frequency of CML patients per year

Admission Year	Frequency (N)	Percent (%)
2013	35	21.2
2014	**37**	**22.4**
2015	16	9.7
2016	13	7.9
2017	20	12.1
2018	12	7.3
2019	18	10.9
2020	14	8.5
Total	165	100

3.1.2 BCR::ABL1 transcript levels as interpreted per International Scale (IS%)

3.1.2.1 Differences in IS% related to active treatment undergoing cases from diagnosis to 1-year follow-ups.

The ultimate aim of CML treatments (Gleevec®-Imatinib mesylate - a ABL1 tyrosine kinase inhibitor) is to reduce the transcript number of the genetic abnormality, which is *BCR::ABL1* gene expression. The efficacy of these CML treatment responses are gauged via monitoring of the haematological (return of blood counts to normal), cytological (reducing and elimination the Ph+ chromosome) and molecular (reducing and elimination of *BCR::ABL1* transcript) responses.

Interestingly, after a 1 year of follow-up, only 104 patients were still actively undergoing treatment, compared to the 165 who were initially diagnosed and started treatment 12 months prior. For molecular monitoring, *BCR::ABL1* levels are monitored by RQ-PCR which measures BCR::ABL1 % according to an international scale. Treatment responses are based on IS %, therefore IS% values of the 104 patients actively undergoing CML treatment were compared at diagnosis (IS1), and then at 1 year follow-up (IS5) on CML treatment. To

determine whether the population mean differs between diagnosis starting point and 1 year after treatment, Wilcoxon signed rank testing was used to test paired difference of repeated measurements of each participant. An average of 54.65% (n=89) cases showed a lesser IS% after being on treatment for one year and 39.77% 9(n=15) of the cases showed a higher transcript at 12 months on treatment compared to IS% at diagnosis **(Table 3.4)**. Following this analysis, a significant difference of 0.000 between IS1 (at diagnoses) and IS 5 (1 year on treatment) were observed.

Table 3.4. IS% at diagnosis vs at 12 months (IS5) on treatment

Ranks			
		N	Mean Rank
is5 - is1*	Negative Ranks	89[a]	54.65
	Positive Ranks	15[b]	39.77
	Total	104	

a – Is5> IS1; b – IS5<IS1

**IS1-at diagnosis; IS5 at 1 year on treatment*

3.1.2.2 Gender distribution frequency related to clinical attendance

The gender distribution related to clinic attendance after a 1-year period was also analysed. From this analysis, 46% (n=48) female and 54%(n=56) male CML patients still attended the clinic after 1 year for CML monitoring and treatments **(Table 3.5)**. We compared the gender frequencies of follow-up patients to those who were lost to follow up. Data indicated that out of a total of 76 females who started treatment and routine testing, 28 (36.8%) were lost to follow up. Additionally, out of a total of 89 male patients who started treatment, 33 (37.07%) were lost to follow up. No statistical significance was found between the gender ratio of the group lost to follow up, which corresponds with the insignificant difference found in the follow up group as well when looking at the ratios of the entire cohort. It can therefore be concluded that the patient's gender does not impact whether they continue treatment and maintain routine testing.

Table 3.5. Gender proportions of CML resistant and Lost To Follow Up (LTFU) groups

	At Diagnosis		At 1 year Follow-up		Lost to follow up (LTFU)	
Sex	**Frequency (N)**	**Percent (%)**	**Frequency (N)**	**Percent (%)**	**Frequency (N)**	**Percent (%)**
Female	76	46.1	48	46,2	28	45,9
Male	89	53.9	56	53,8	33	54,1
Total	**165**	100	**104**	100	**61**	100
	P = 0,992345		**P = 0,982418**		**P = 0,990307**	

.

3.1.2.3 Treatment response efficacy (IS%) in relation to age

IS% in relation to patient age at time of diagnosis was also analysed. Spearman's Correlation statistics was used to measure how strong the relationship/association is between the variables of interests. This test measures the strength and direction of a monotonic association between ranked variables (i.e., IS% and Age). This essentially means that the variables tend to move in the same relative direction, but not necessarily the same rate. This differs from Pearson corelations as it measures the direction of linear associations between variables which are not needed to be ranked. The correlation coefficient is a value between -1 and +1. If the value is negative, it indicates negative relationship in the sense that as the one value increases, the other decreases. Contrastingly, a positive value would indicate a positive relationship in the sense that if one value increases, the other increases as well. A value +1 value would indicate a perfectly positive correlation, whereas a value near -1 would indicate a perfectly negative correlation and a value at 0 would indicate no correlation. The monotonic association between Age and IS% indicate a correlation coefficient of -0.077. The Sig (2-tailed) essentially is the two-tailed p-value which evaluates the null against an alternative. When looking at the P values, $p > 0.05$ would indicate the null hypothesis, i.e.., that there is no correlation between age and IS%. A $p < 0.05$ indicates the alternative hypothesis, i.e., that there is a correlation between variables, **(Table 3.6)**. The results therefore indicate weak negative correlation between age and IS% (-0.077) and the null hypothesis (0.335), where there is no statistically significant correlation between age and IS% within the cohort.

Table 3.6. IS% in relation to age at diagnosis

Correlations				
			age	is1
Spearman's rho	age	Correlation Coefficient	1.000	-0.077
		Sig. (2-tailed)	.	0.335
	is1*	Correlation Coefficient	-0.077	1.000
		Sig. (2-tailed)	0.335	.

Is1 - at diagnosis*

When the IS% in relation to age is graphically compared, a random dispersion is observed. No clear pattern or trend is noticed, further indicating that age is not associated with increase in IS% in our cohort **(Figure 3.2)**.

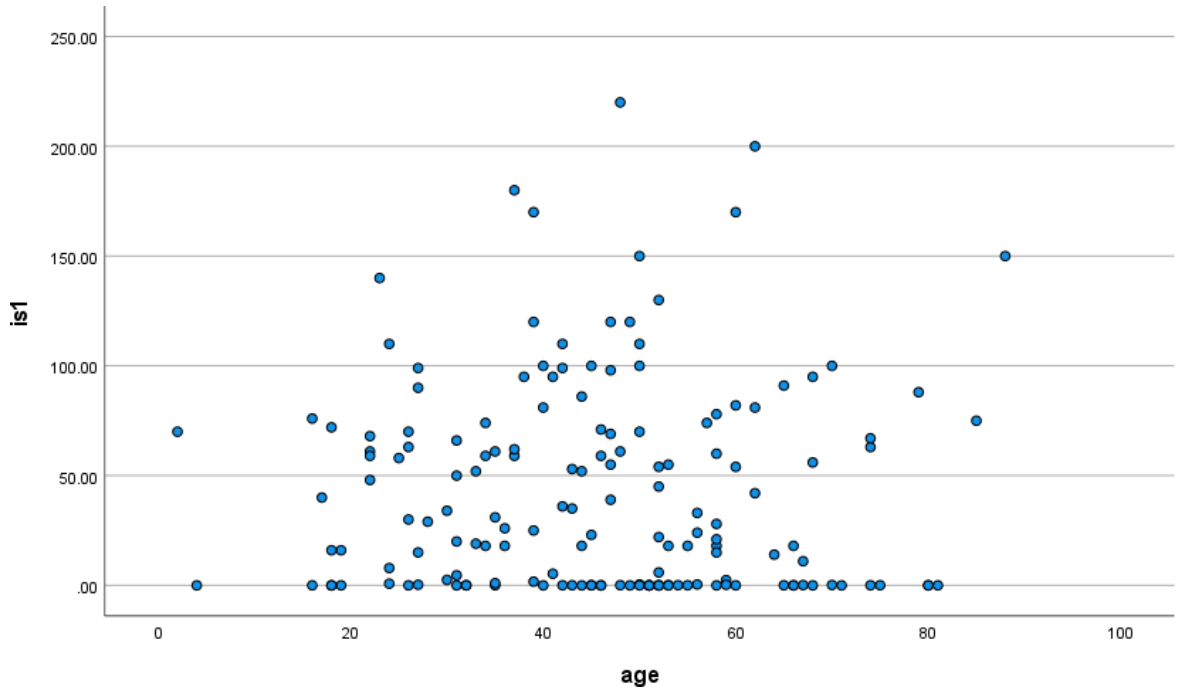


Figure 3.2. Graphical representation of participants age at diagnosis and corresponding IS%

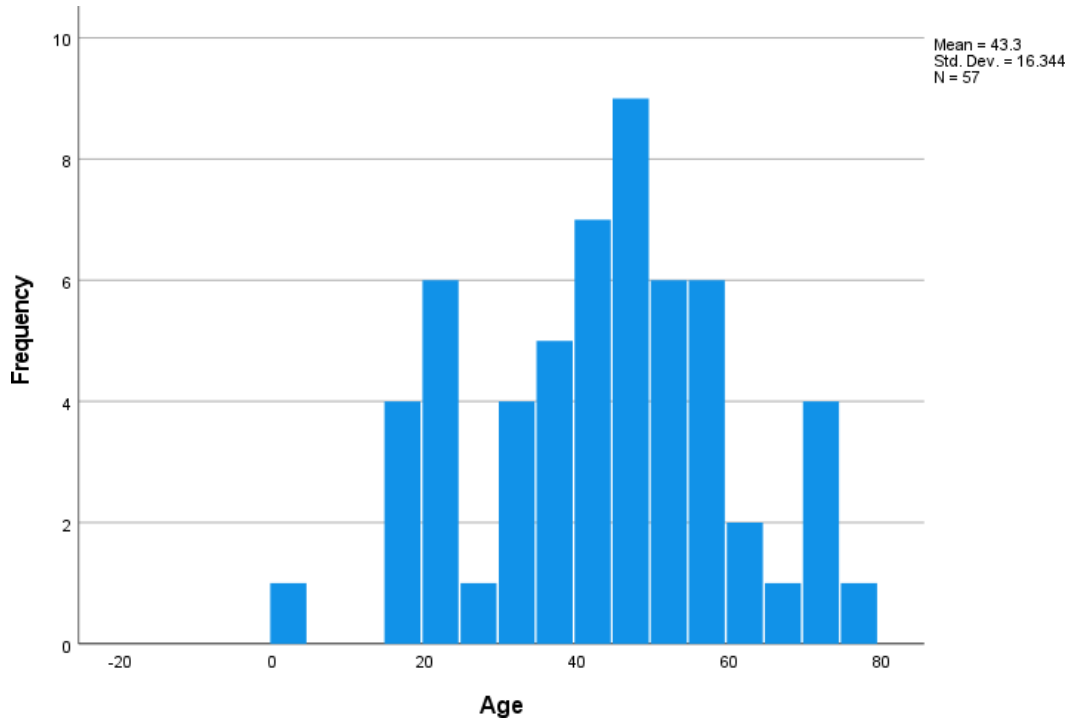
**is1-at diagnosis*

3.2 Population description of the CML resistant participants

During the review period, a total of 165 CML had been diagnosed and/or treated. Out of these 165 CML patients, 57 were classified as resistant to first line therapy, according to ELN guidelines, and were either stated to be in the warning or failure stage. **Table 3.7** summarises the demographics of the resistant patients. 2020 was the year of national lockdowns due the COVID-19 pandemic, however, the results did not deviate significantly from the pre-pandemic admissions rates, as rates throughout where rather inconsistent – similar to rates observed during 2015, 2016 and 2018. We then compared the rate of resistant cases to the admissions for the respective years and found that 2016 observed the highest rate of resistant cases, at 69%

Table 3.7. Demographics and ELN response for CML resistant cohort

Year of admission	Total admissions (n=165)	Resistant cases (n=57)	Resistance Frequency (%)
2013	35	12	34%
2014	37	11	30%
2015	16	4	25%
2016	13	9	69%
2017	20	10	50%
2018	12	3	25%
2019	18	7	39%
2020	14	1	7%
Total	165	57	
Gender			
	Frequency	**Percent**	
F	26	45.6	
M	31	54.4	
	57	100.0	
ELN response			
	Frequency (n)		
Unknown	1		
Failure	44		
Warning	12		
Total	57		



Abbreviations: Std.Dev, Standard Deviation of the Mean; N, Sample size

Figure 3.3. CML resistant cohort age groups at diagnosis

When looking at the age groups of the resistant classified participants, the average age was found to be 43.3 years old, with the youngest resistant patient being two years old and the eldest being 75 years old **(Table 3.7 and Figure 3.3).** Additionally, results showed that 45.6% (n=26) of the resistant patients were females, with n=21 being classified to be at the failure stage of resistance and 5 being categorised in the warning stage of resistance according to ELN guidelines. Moreover, a total of 54.4% (n=31) of resistant patients were male, with 23 of these patients being classified as failure to respond to treatment, and 7 being classified as being within the warning stage of resistance. One of the male's resistance statuses according to ELN recommendations was still under investigation **(Table 3.8)**

Table 3.8. Cross-tabulation of CML resistant cohort participants gender and ELN response

Gender * ELN response Crosstabulation					
		ELN response			Total
		Unknown	Failure	Warning	
Gender	F	0	21	5	26
	M	1	23	7	31
Total		1	44	12	57

3.3 Descriptive Statistics of selected CML patients for sequence variant/mutation screening

Out of the 57 resistant patients, 20 participants were selected for the subsequent sequence variant/mutation screening portion of the sub-study. The participants were randomly selected based on biospecimen availability at diagnosis and at one year follow ups, as well as biospecimen integrity and availability of all relevant clinical data This includes, but is not limited to; disease phase, adherence, choice of treatment, and/or any change in treatments. Determining whether any sequence variants/mutations were present at diagnosis and at the one-year follow-up for each of the selected CML patients was also of importance. Results show a gender ratio of 55% (n=11) female participants compared to 45% (n=9) male participants **(Table 3.9**). The ELN response categorises 75% (n=15/20) of patients as failure, whereas 25% (n=/20) were in the warning stage of failure to treatment **(Table 3.9)**. The year 2019 had the biggest intake of 35% (n=7/20) of the 20 selected CML resistant participants **(Table 3.9)**.

Table 3.9. Demographics of CML participants selected for variant/mutation screening

Admission year			
		Frequency	**Percent**
	2014	1	5.0
	2016	4	20.0
	2017	4	20.0
	2018	3	15.0
	2019	7	35.0
	2020	1	5.0
	Total	20	100.0
Gender			
		Frequency	**Percent**
	F	11	55.0
	M	9	45.0
	Total	20	100.0
ELN response			
		Frequency	**Percent**
	Failure	15	75.0
	Warning	5	25.0
	Total	20	100.0

Furthermore, the sample group selected for screening showed a 20% (n=4) mortality rate. The *BCR::ABL1* IS level log charts, supported by clinical data, indicate that two of the participants (378 and 356) had an overall survival of less 12 months **(Figure 3.3.1).** Clinical data indicates that participant 378 was found to be in the Accelerated phase of disease, and participant 356 had transformed from CML to B-Cell Acute Lymphoblastic Leukaemia (B-ALL) **(Table 3.9.1).** Participant 311 was found to have progressed to the Accelerated phase as well, while participant 322 was found to have progressed to the Blast Crisis phase of disease and had an overall survival of four years and two years, respectively.

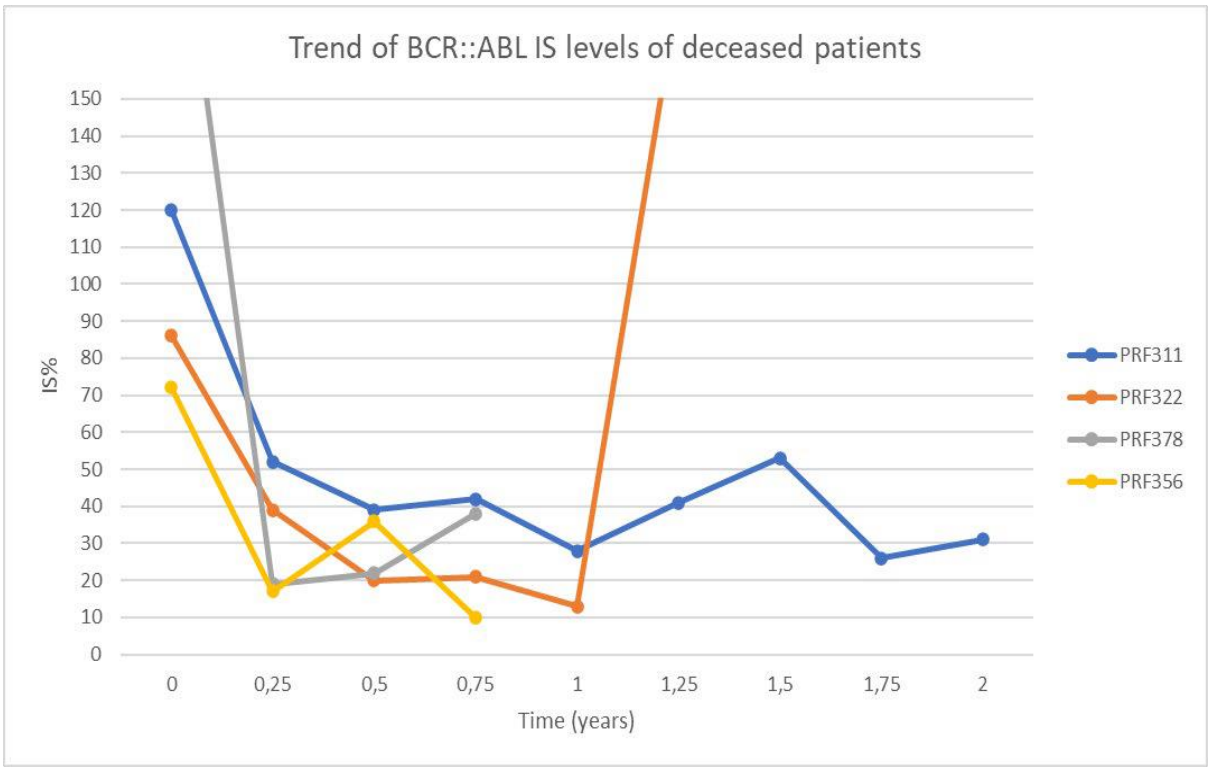


Figure 3.3.1. Trend of BCR::ABL IS% levels in deceased patients in screening cohort

Moreover, the average age of the CML deaths within the cohort selected for screening was found to be 39.2 years of age, with a median of 44.5 years. The youngest deceased patient being only 19 years old **(Table 3.9.1).** The exact cause of death of these patients were inconclusive from reports, and cause of death was merely attributed to CML.

Table 3.9.1 Age, gender and CML phase of participants selected for screening who are deceased

Participant	Initial Diagnosis	Age at demise	Gender	CML Phase
311	March 2018	43	Male	AP
322	June 2018	46	Female	BC
378	October 2019	49	Male	B-ALL transformation from CML
356	March 2019	19	Male	AP

Abbreviations: AP, accelerated phase, BC, Blast crisis phase, B-ALL, B-Cell Acute Lymphoblastic Leukaemia

3.4 Screening of the *BCR::ABL 1* Kinase domain using Sanger sequencing

This portion of the study aimed to identify documented and novel clinically relevant sequence variants within the kinase domain (KD) of the *ABL1* portion of the *BCR::ABL1* gene. The end goal was to evaluate a mutation screening assay with possible implementation at the NHLS, Molecular Haematology Laboratory at Tygerberg Hospital. Globally, direct sequencing is the gold standard for *ABL1* sequence variant/mutation screening. With more than 100 amino acid substitutions detected in the KD portion of *ABL1*, exon 4 to exon 10 of the *ABL1* gene was screened.

3.4.1 Primer design and verification

The *BCR::ABL1* gene's reference sequence (NM_005157.6) was retrieved from the NCBI data base (https://www.ncbi.nlm.nih.gov/). The Primer3Plus primer design software (https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) was used to design forward and reverse primers for the 593bp (Exon 4-10) region that code for the KD of the ABL1 gene of the BCR::ABL1 oncogene. Primers were designed according to PCR primer design guidelines taking into considerations several criteria such as length, GC content, melting temperature, etc for optimum performance. Each primer sequence was subjected to self-complementarity and primer-primer complementarity tests using DNAMAN software. The primers where then verified using the NCBI BLAST program from the NCBI database. Once passing all the quality checks, the primers where then synthesised by Whitehead Scientific (Pty) Ltd Integrated DNA Technologies. The primers were received in a lyophilised state and were diluted according to manufacturer's instructions. Aliquots were made for downstream applications to prevent contamination and degradation due to thawing and usage.

3.4.2 Sample selection and concentrations

The 20 selected CML participants' DNA was extracted from archived buffy coats which were stored at -20°C. Only those with a confirmed CML diagnosis and a positive p210 BCR-ABL1 oncogene identification was selected for participation. A total of 5/20 CML patients had an archived baseline biospecimen (sample collected at diagnosis/before commencement of treatment). For these five CML participants, an additional sample taken 12 months after being on treatment was also screened. Following the DNA extraction from frozen buffy coats, all 25

samples DNA concentration (A260/280) and purity were determined using the Biodrop instrument. Concentrations yielded as well as the A260/280 ratio is shown in **Appendix I**.

3.4.3 Polymerase Chain Reaction (PCR)

Following extensive optimization at various temperatures, and using a magnesium gradient, each of the seven exons of the KD domain of the *ABL1* gene of the *BCR::ABL1* gene was successfully amplified, using a standard PCR procedure as detailed in the methodology in Chapter 2. Each PCR product generated fragment sizes according to what was expected **(Figure 3.4)**

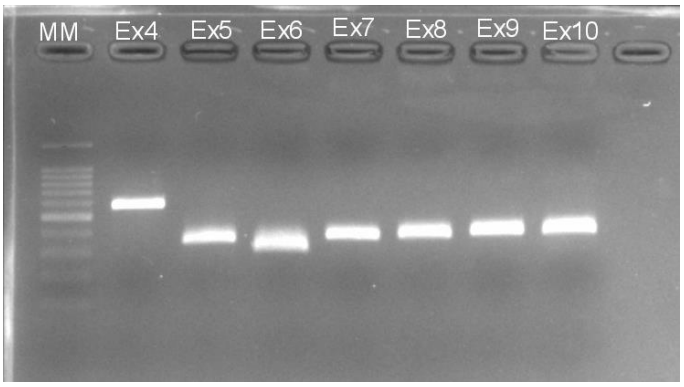
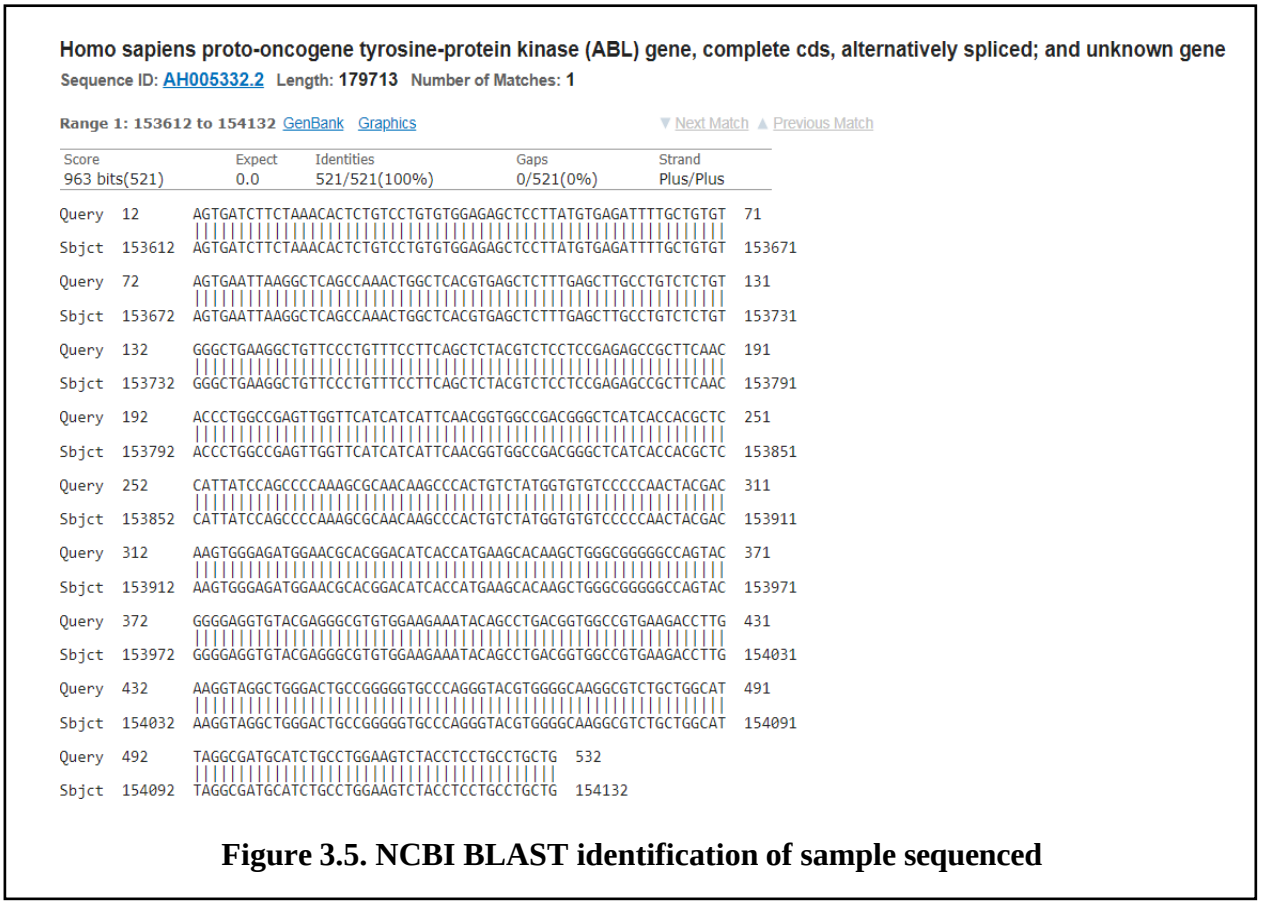


Figure 3.4. Representative example of amplified products of exon 4 to exon 10 amplicons visualised on 2% agarose gel.

3.4.4 Mutational screening results

Following successful PCR amplification and confirmative visualisation by gel electrophoresis, PCR products were cleaned for subsequent Sanger Sequencing. The seven exons of the 20 participants' samples, as well as the additional five baseline samples were all successfully amplified. Therefore, a total of 175 amplicons were bidirectionally Sanger sequenced.

Figure 3.5 is a representative example for the BLAST results for one of the biospecimens, which indicates 100% identity to the reference *ABL1* gene. The 175 products sequenced produced the same quality and accuracy depicted here.



Homo sapiens proto-oncogene tyrosine-protein kinase (ABL) gene, complete cds, alternatively spliced; and unknown gene

Sequence ID: AH005332.2 Length: 179713 Number of Matches: 1

Range 1: 153612 to 154132 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
963 bits(521)	0.0	521/521(100%)	0/521(0%)	Plus/Plus

```
Query  12      AGTGATCTTCTAAACACTCTGTCCTGTGTGGAGAGCTCCTTATGTGAGATTTTGCTGTGT  71
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  153612  AGTGATCTTCTAAACACTCTGTCCTGTGTGGAGAGCTCCTTATGTGAGATTTTGCTGTGT  153671

Query  72      AGTGAATTAAGGCTCAGCCAAACTGGCTCACGTGAGCTCTTTGAGCTTGCCTGTCTCTGT  131
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  153672  AGTGAATTAAGGCTCAGCCAAACTGGCTCACGTGAGCTCTTTGAGCTTGCCTGTCTCTGT  153731

Query  132     GGGCTGAAGGCTGTTCCCTGTTTCCTTCAGCTCTACGTCTCCTCCGAGAGCCGCTTCAAC  191
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  153732  GGGCTGAAGGCTGTTCCCTGTTTCCTTCAGCTCTACGTCTCCTCCGAGAGCCGCTTCAAC  153791

Query  192     ACCCTGGCCGAGTTGGTTCATCATCATTCAACGGTGGCCGACGGGCTCATCACCACGCTC  251
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  153792  ACCCTGGCCGAGTTGGTTCATCATCATTCAACGGTGGCCGACGGGCTCATCACCACGCTC  153851

Query  252     CATTATCCAGCCCCAAAGCGCAACAAGCCCACTGTCTATGGTGTGTCCCCCAACTACGAC  311
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  153852  CATTATCCAGCCCCAAAGCGCAACAAGCCCACTGTCTATGGTGTGTCCCCCAACTACGAC  153911

Query  312     AAGTGGGAGATGGAACGCACGGACATCACCATGAAGCACAAGCTGGGCGGGGGCCAGTAC  371
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  153912  AAGTGGGAGATGGAACGCACGGACATCACCATGAAGCACAAGCTGGGCGGGGGCCAGTAC  153971

Query  372     GGGGAGGTGTACGAGGGCGTGTGGAAGAAATACAGCCTGACGGTGGCCGTGAAGACCTTG  431
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  153972  GGGGAGGTGTACGAGGGCGTGTGGAAGAAATACAGCCTGACGGTGGCCGTGAAGACCTTG  154031

Query  432     AAGGTAGGCTGGGACTGCCGGGGGTGCCCAGGGTACGTGGGGCAAGGCGTCTGCTGGCAT  491
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  154032  AAGGTAGGCTGGGACTGCCGGGGGTGCCCAGGGTACGTGGGGCAAGGCGTCTGCTGGCAT  154091

Query  492     TAGGCGATGCATCTGCCTGGAAGTCTACCTCCTGCCTGCTG  532
               |||||||||||||||||||||||||||||||||||||||||
Sbjct  154092  TAGGCGATGCATCTGCCTGGAAGTCTACCTCCTGCCTGCTG  154132
```

Figure 3.5. NCBI BLAST identification of sample sequenced

Of the 175 products that were screened, only 10 showed sequence variant changes for further investigation. Two amplicons displayed a sequence base pair change at coding sequence base pair position 720, in exon 4, resulting in a G>A heterogenous sequence change which can be seen in the representative chromatogram file in Figure 3.6. Eight amplicons displayed a heterozygous A> G sequence change at coding sequence base pair position 1497, in exon 9, which can be seen in the representative chromatogram file in Figure 3.7. When a double peak is observed in a chromatograph, it is an indication that two different alleles are present **(Figure 3.6 and Figure 3.7).**

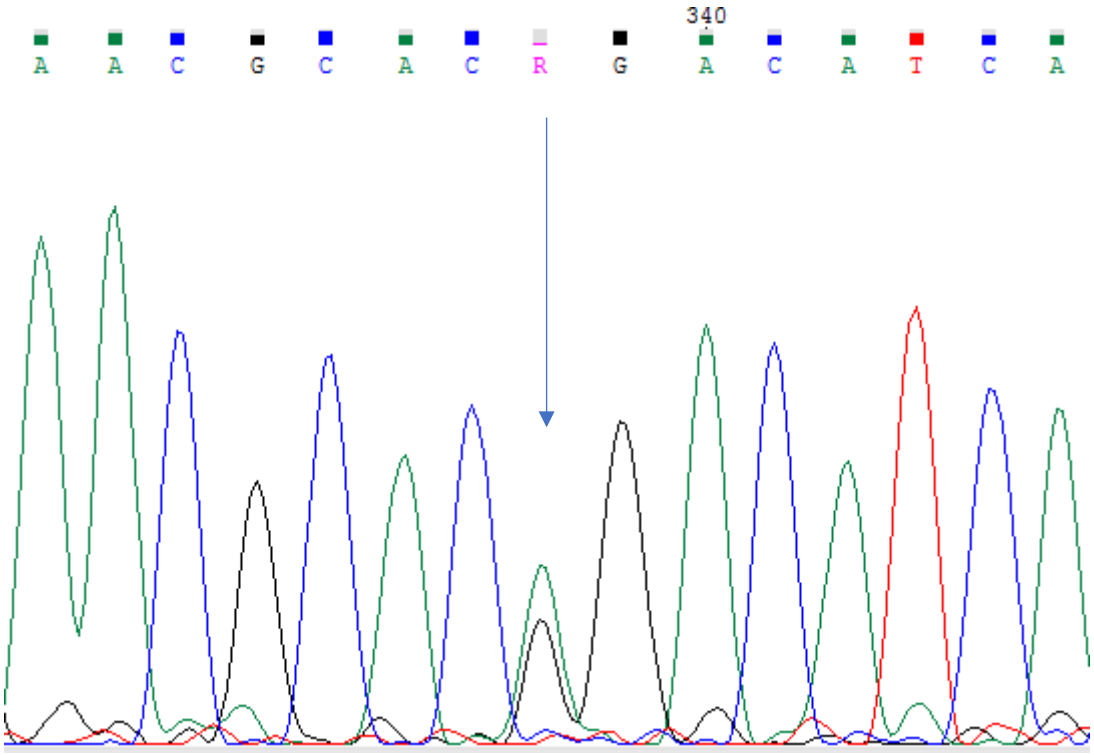


Figure 3.6. Representative chromatograph of area of interest, Exon 4. A double peak can be observed, indicating a heterozygous variant be as indicated by arrow.

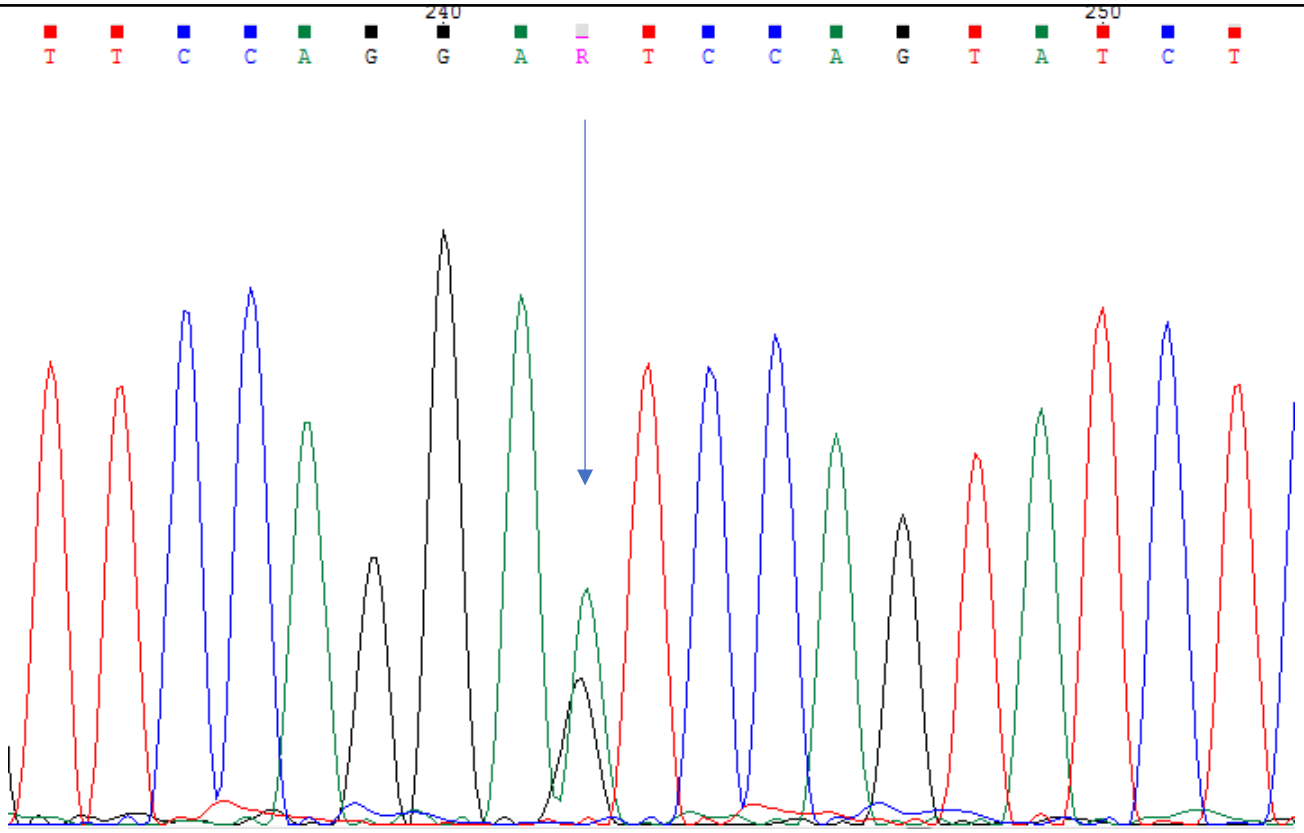


Figure 3.7. Representative chromatograph of area of interest, Exon 9. A double peak can be observed, indicated by the arrow, indicating a heterozygous variant as well.

Multiple sequence alignment using the DNAMAN software was carried out for all sequenced amplicons and both the forward as well as the reverse compliment sequences was used for these alignments. The observed sequence changes as stated above and observed in the chromatogram file are confirmed in the multiple sequence alignments for the variant identified in Exon 4 and Exon 9, respectively, when compared to the reference sequence. **(Figure 3.8 and 3.9)**

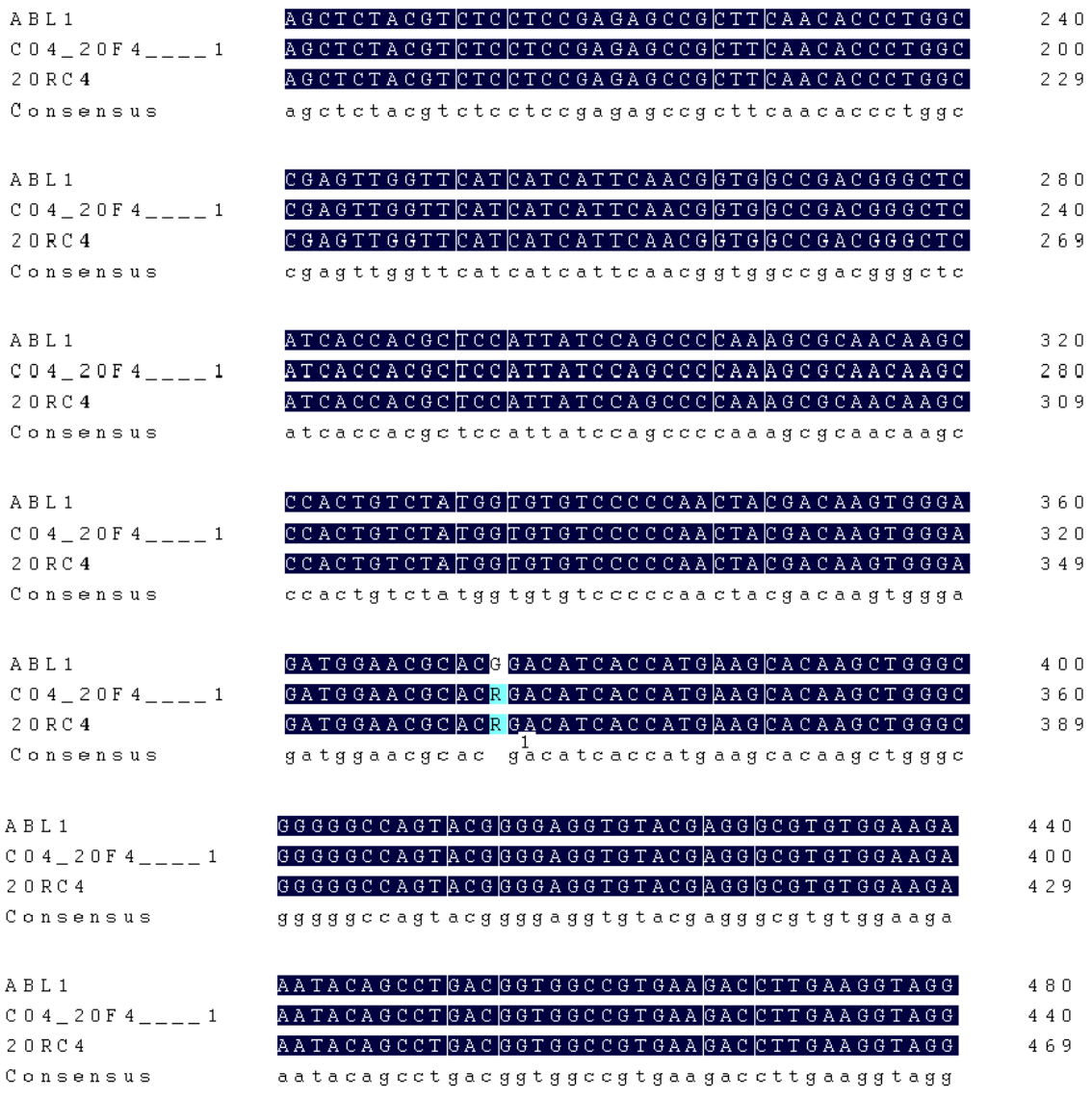


Figure 3.8. Representative Multiple Sequence Alignment of Exon 4 to confirm the variant in the forward strand and reverse strand, as compared to the reference sequence.

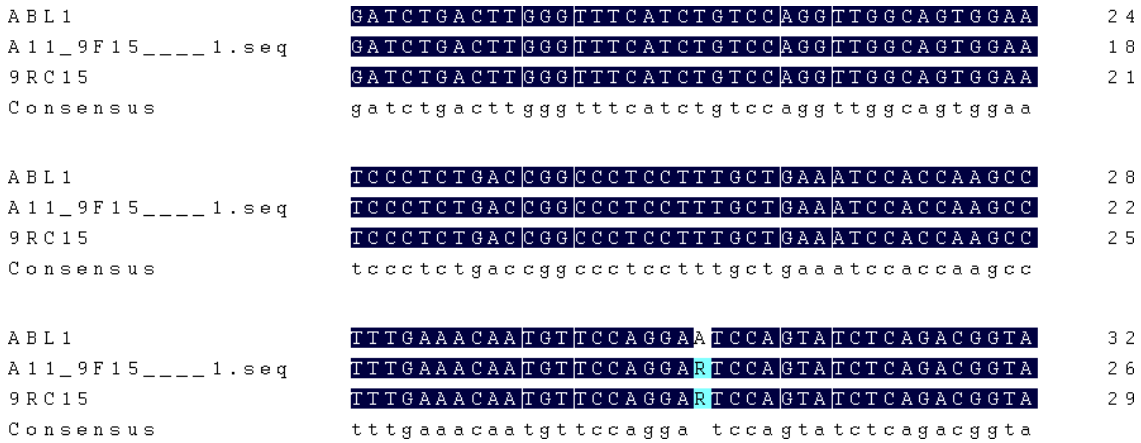


Figure 3.9. Representative Multiple Sequence Alignment of Exon 9

Thus, in summary, two heterozygous sequence variants that might be clinically relevant, was found and confirmed in exon 4 and exon 9 in the region that spans a portion of the kinase domain. Since the *ABL1* proto-oncogene fuses with various translocation genes in various leukaemia's, we were specifically interested in the t(9;22) translocation that results in a fusion with the 5' end of the breakpoint cluster region gene (BCR; MIM:151410). Considering that there are also two transcript variants of the *ABL1* gene due to alternative splicing, it was of utmost importance to use the correct mRNA transcript for downstream analysis. Publicly available databases, Ensembl Genome Data Browser, National Centre for Biotechnology Information (NCBI), UniProt as well as OMIM and ClinVar (part of NCBI) were used for further investigation of these sequence variants as well as their clinical relevance. It should be noted that variant (a) of the *ABL1* includes an alternate 5'terminal exon (exon 1a instead of exon 1b), and differs in the 5' UTR and 5' coding region, compared to variant b. In addition, the encoded isoform (a) of ABL1, has a distinct N-terminus and is shorter than isoform b and are localised to the nucleus.

Furthermore, the transcript ID ENST00000318560.6 (https://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000097007;r=9:130713043-130887675;t=ENST00000318560) which also matches to the NCBI reference sequences, mRNA transcript variant isoform a, NM_005157.6 (https://www.ncbi.nlm.nih.gov/nuccore/NM_005157.6) and NP_005148.2 (https://www.ncbi.nlm.nih.gov/protein/62362414) was used for further analysis . The *ABL1* transcript 'a' spans 11 exons and coding regions and is composed of 5578 base pairs and encodes a protein composed of 1130 amino acids in length. The coding sequence retrieved fromNCBI composes of 3393 nucleotides and are shown in **Figure 3.10**

(https://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi?REQUEST=CCDS&DATA=CCDS35166.1) , displaying the alternating exons in blue and highlighting the position of where the sequence variants occur (Red font and bold) in the respective exons.

Pairwise sequence alignments were done using the FASTA nucleotide sequence of the cases which harbours the respective identified sequence variants and the reference *ABL1* transcript 'a'. In depth review of the literature as well as databases searches revealed that both sequence variant changes had been previously identified and were found to be synonymous. This means that while a change in the DNA sequence was observed, the encoded amino acid remains unchanged.

CCDS35166.1

Nucleotide Sequence (3393 nt):

ATGTTGGAGATCTGCCTGAAGCTGGTGGGCTGCAAATCCAAGAAGGGGCTGTCCTCGTCCTCCAGCTGTT
ATCTGGAAGAAGCCCTTCAGCGGCCAGTAGCATCTGACTTTGAGCCTCAGGGTCTGAGTGAAGCCGCTCG
TTGGAACTCCAAGGAAAACCTTCTCGCTGGACCCAGTGAAAATGACCCCAACCTTTTCGTTGCACTGTAT
GATTTTGTGGCCAGTGGAGATAACACTCTAAGCATAACTAAAGGTGAAAAGCTCCGGGTCTTAGGCTATA
ATCACAATGGGGAATGGTGTGAAGCCCAAACCAAAAATGGCCAAGGCTGGGTCCCAAGCAACTACATCAC
GCCAGTCAACAGTCTGGAGAAACACTCCTGGTACCATGGGCCTGTGTCCCGCAATGCCGCTGAGTATCTG
CTGAGCAGCGGGATCAATGGCAGCTTCTTGGTGCGTGAGAGTGAGAGCAGTCCTGGCCAGAGGTCCATCT
CGCTGAGATACGAAGGGAGGGTGTACCATTACAGGATCAACACTGCTTCTGATGGCAAGCTCTACGTCTC
CTCCGAGAGCCGCTTCAACACCCTGGCCGAGTTGGTTCATCATCATTCAACGGTGGCCGACGGGCTCATC
ACCACGCTCCATTATCCAGCCCCAAAGCGCAACAAGCCCACTGTCTATGGTGTGTCCCCCAACTACGACA
AGTGGGAGATGGAACGCACGGACATCACCATGAAGCACAAGCTGGGCGGGGGCCAGTACGGGGAGGTGTA
CGAGGGCGTGTGGAAGAAATACAGCCTGACGGTGGCCGTGAAGACCTTGAAGGAGGACACCATGGAGGTG
GAAGAGTTCTTGAAAGAAGCTGCAGTCATGAAAGAGATCAAACACCCTAACCTGGTGCAGCTCCTTGGGG
TCTGCACCCGGGAGCCCCCGTTCTATATCATCACTGAGTTCATGACCTACGGGAACCTCCTGGACTACCT
GAGGGAGTGCAACCGGCAGGAGGTGAACGCCGTGGTGCTGCTGTACATGGCCACTCAGATCTCGTCAGCC
ATGGAGTACCTGGAGAAGAAAAACTTCATCCACAGAGATCTTGCTGCCCGAAACTGCCTGGTAGGGGAGA
ACCACTTGGTGAAGGTAGCTGATTTTGGCCTGAGCAGGTTGATGACAGGGGACACCTACACAGCCCATGC
TGGAGCCAAGTTCCCCATCAAATGGACTGCACCCGAGAGCCTGGCCTACAACAAGTTCTCCATCAAGTCC
GACGTCTGGGCATTTGGAGTATTGCTTTGGGAAATTGCTACCTATGGCATGTCCCCTTACCCGGGAATTG
ACCTGTCCCAGGTGTATGAGCTGCTAGAGAAGGACTACCGCATGGAGCGCCCAGAAGGCTGCCCAGAGAA
GGTCTATGAACTCATGCGAGCATGTTGGCAGTGGAATCCCTCTGACCGGCCCTCCTTTGCTGAAATCCAC
CAAGCCTTTGAAACAATGTTCCAGGAATCCAGTATCTCAGACGAAGTGGAAAAGGAGCTGGGGAAACAAG
GCGTCCGTGGGGCTGTGAGTACCTTGCTGCAGGCCCCAGAGCTGCCCACCAAGACGAGGACCTCCAGGAG
AGCTGCAGAGCACAGAGACACCACTGACGTGCCTGAGATGCCTCACTCCAAGGGCCAGGGAGAGAGCGAT
CCTCTGGACCATGAGCCTGCCGTGTCTCCATTGCTCCCTCGAAAAGAGCGAGGTCCCCCGGAGGGCGGCC
TGAATGAAGATGAGCGCCTTCTCCCCAAAGACAAAAAGACCAACTTGTTCAGCGCCTTGATCAAGAAGAA
GAAGAAGACAGCCCCAACCCCTCCCAAACGCAGCAGCTCCTTCCGGGAGATGGACGGCCAGCCGGAGCGC
AGAGGGGCCGGCGAGGAAGAGGGCCGAGACATCAGCAACGGGGCACTGGCTTTCACCCCCTTGGACACAG
CTGACCCAGCCAAGTCCCCAAAGCCCAGCAATGGGGCTGGGGTCCCCAATGGAGCCCTCCGGGAGTCCGG
GGGCTCAGGCTTCCGGTCTCCCCACCTGTGGAAGAAGTCCAGCACGCTGACCAGCAGCCGCCTAGCCACC
GGCGAGGAGGAGGGCGGTGGCAGCTCCAGCAAGCGCTTCCTGCGCTCTTGCTCCGCCTCCTGCGTTCCCC
ATGGGGCCAAGGACACGGAGTGGAGGTCAGTCACGCTGCCTCGGGACTTGCAGTCCACGGGAAGACAGTT
TGACTCGTCCACATTTGGAGGGCACAAAAGTGAGAAGCCGGCTCTGCCTCGGAAGAGGGCAGGGGAGAAC
AGGTCTGACCAGGTGACCCGAGGCACAGTAACGCCTCCCCCCAGGCTGGTGAAAAAGAATGAGGAAGCTG
CTGATGAGGTCTTCAAAGACATCATGGAGTCCAGCCCGGGCTCCAGCCCGCCCAACCTGACTCCAAAACC
CCTCCGGCGGCAGGTCACCGTGGCCCCTGCCTCGGGCCTCCCCCACAAGGAAGAAGCTGGAAAGGGCAGT
GCCTTAGGGACCCCTGCTGCAGCTGAGCCAGTGACCCCCACCAGCAAAGCAGGCTCAGGTGCACCAGGGG
GCACCAGCAAGGGCCCCGCCGAGGAGTCCAGAGTGAGGAGGCACAAGCACTCCTCTGAGTCGCCAGGGAG
GGACAAGGGGAAATTGTCCAGGCTCAAACCTGCCCCGCCGCCCCCACCAGCAGCCTCTGCAGGGAAGGCT
GGAGGAAAGCCCTCGCAGAGCCCGAGCCAGGAGGCGGCCGGGGAGGCAGTCCTGGGCGCAAAGACAAAAG
CCACGAGTCTGGTTGATGCTGTGAACAGTGACGCTGCCAAGCCCAGCCAGCCGGGAGAGGGCCTCAAAAA
GCCCGTGCTCCCGGCCACTCCAAAGCCACAGTCCGCCAAGCCGTCGGGGACCCCCATCAGCCCAGCCCCC
GTTCCCTCCACGTTGCCATCAGCATCCTCGGCCCTGGCAGGGGACCAGCCGTCTTCCACCGCCTTCATCC
CTCTCATATCAACCCGAGTGTCTCTTCGGAAAACCCGCCAGCCTCCAGAGCGGATCGCCAGCGGCGCCAT
CACCAAGGGCGTGGTCCTGGACAGCACCGAGGCGCTGTGCCTCGCCATCTCTAGGAACTCCGAGCAGATG
GCCAGCCACAGCGCAGTGCTGGAGGCCGGCAAAAACCTCTACACGTTCTGCGTGAGCTATGTGGATTCCA
TCCAGCAAATGAGGAACAAGTTTGCCTTCCGAGAGGCCATCAACAAACTGGAGAATAATCTCCGGGAGCT
TCAGATCTGCCCGGCGACAGCAGGCAGTGGTCCAGCGGCCACTCAGGACTTCAGCAAGCTCCTCAGTTCGGTGAA
GGAAATCAGTGACATAGTGCAGAGGTAG

Figure 3.10. NCBI coding sequence, Exon4 (273bp) – c.720G>A and Exon 9 (90bp) – c.1497. The blue and black sequences indicate alternating exons on the coding sequence. The nucleotides where sequences changes of interest occur within the respective exons are highlighted in red.

3.5 In silico analysis

Bioinformatics tools were used to conduct *in silico* analysis to predict the potential impacts of the detected variants. Predictive programs were used concurrently, as individual programs would have limitations. The 2015 American College of Medical Genetics and Genomics (ACMG)/ Association for Molecular Pathology (AMP) framework were used for variant assessment and interpretations. Therefore, knowledge regarding the variant spectrum for the gene and disease as well as population frequency was used to assess for and against pathogenicity.

Two heterozygous variants were confirmed using Sanger sequencing. The *ABL1* mRNA transcript variant isoform a spans 11 exons and coding regions and is composed of 5578 nucleotides and encodes a protein composed of 1130 amino acids (AA). The evaluation of the possible functional impacts of the structural changes of these variants were investigated.

rs2229069 - (NM_005157.6(ABL1):c.720G>A (p.Thr240Thr)

The rs2229069 heterozygous variant has an ancestral G allele and a mutational A allele (ACG> ACA). This variant has been classed as a coding single nucleotide variant (SNV) and does not alter the amino acid sequence (T (Thr) > T (Thr), and consequently the protein it codes for (NP_005148.2:p.Thr240Thr). According to ClinVar, the variation has been found to have a benign clinical effect (https://www.ncbi.nlm.nih.gov/clinvar/RCV000968160.2/). The last clinical evaluation was December 2019 and submitted via Invitae.

rs2227985 - NM_005157.6 (ABL1):c.1497A>G (p. Glu499Glu)

The rs222795 heterozygous variant has an ancestral A allele, and a mutational G allele (GAA > GAG). This variant has also been classified as a coding SNV and does not alter the amino acid sequence (E (Glu) > E (Glu) and consequently the protein is unaffected as well (NP_005148.2:p.Glu499). According to ClinVar, the clinical significance of this variant has not been reported on (https://www.ncbi.nlm.nih.gov/snp/rs2227985#clinical_significance).

3.5.1 Ensembl Genome Browser

The Ensemble Genome Browser provides the identification and analysis of each variant/mutation found within a specific gene. **Table 3.10** serves as a summary of results retrieved for the variants under investigation. The results indicate both ancestral and mutational alleles and well as the clinical significance, which indicates that rs2229069 has been found to be benign and rs2227985's clinical significance is yet to be reported.

Table 3.10. Ensembl Genome Browser results for each variant

	rs2229069 – Exon 4	**rs2227985 – Exon 9**
Alleles	G/A, ancestral G	A/G, ancestral A
MAF	A=0.003108/781 (GnomAD_exomes) A=0.012979/65 (1000Genomes) A=0.03012/30 (HapMap)	G=0.063068 (15860/251474, GnomAD_exome) G=0.0857 (429/5008, 1000G) G=0.1206 (197/1634, HapMap)
Clinical significance	Benign	None reported
Variant type	SNV	SNV
Chromosomal Location	chr9:130862933 (GRCh38.p13)	chr9:130880141 (GRCh38.p13)
Ensembl HGVS	NC_000009.11:g.133738320G>A)	NC_000009.11:g.133755528A>G
cDNA position	720	1497
Protein	240	499
Amino acids	Thr/Thr	Glu/Glu
Codons	ac**G**/ac**A**	ga**A**/ga**G**

MAF-minor allele frequency, SNV-single nucleotide variant, HGVS- Human Genome Variation, Thr-Threonine, Glu – Glutamic acid

3.5.2 NCBI CDS sequences (NM_005157.6; NC_000009.12) indicating positions of variant alleles

The NCBI sequence is a reference standard in the RefSeqGene project, which defines genomic sequences to be used as reference standards for well-characterised genes. This allows sequenced data to be compared against globally accepted, published standards.

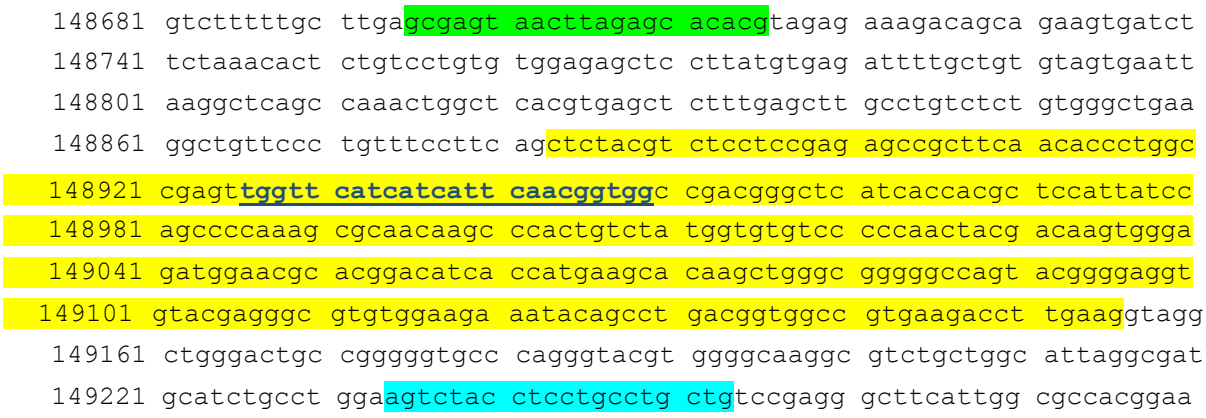


Figure 3.11. Representative nucleotide sequence of Exon 4 of the *ABL1* genomic area. The area in yellow represents the exonic area, with the forward primer highlighted in green and the reverse primer highlighted in blue. The position where the mutational allele occurs is highlighted in red.

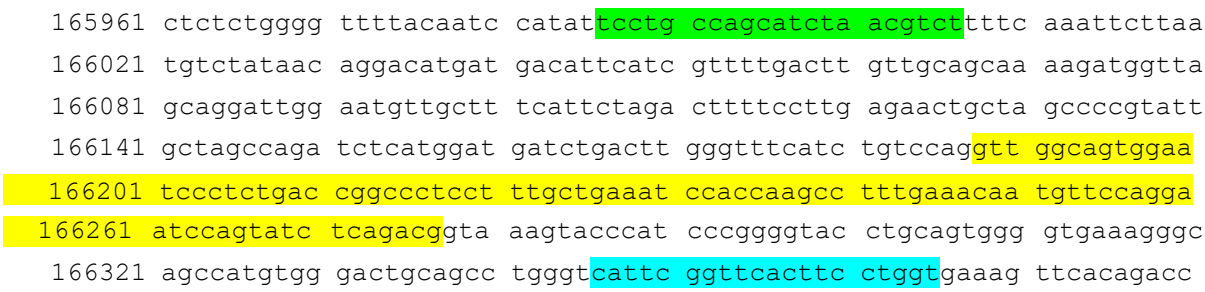


Figure 3.12. Representative nucleotide sequence of Exon 9 of the *ABL1* gene. The area in yellow represents the exonic area, with the forward primer highlighted in green and the reverse primer highlighted in blue. The position where the mutational allele occurs is highlighted in red.

3.5.3 In depth bioinformatics analysis of synonymous sequence variants

It was always thought that all synonymous mutations were silent and their effects on molecular function of the corresponding proteins were assumed to be minimal (Zeng & Bromberg, 2019); resultantly our ability to predict effects were limited. Variant rs2229069 has been published as having a benign clinical significance, with ClinVar Accession number RCV000968160.2, as a single nucleotide variant in 2019. The clinical significance of variant rs2227985 has not been

published according to the ClinVar database. However, further research into synonymous SNV has demonstrated that silent mutations may alter protein expression, conformation as well as function. Therefore, we explored the potential impact of these synonymous SNVs on splicing and transcription using the following databases and bioinformatics tools.

3.5.3.1 SynMICdb

SynMICdb is a comprehensive resource that was developed for cancer related synonymous mutations. The site displays 56 entries in the *ABL1* oncogene for synonymous sequence variants. However, the two variants identified within our population was not listed (http://synmicdb.dkfz.de/rsynmicdb/) .

3.5.3.2 synVep v 1.0

synVep is a machine learning-based predictor that was used for evaluating the effects of human synonymous variants and it uses observed and generated variants. The prediction output is the probability of whether a human synonymous variant have an clinical effect.

For both the rs2229069 and rs2227985 variant, the following results was observed. Within Table 3.12, Class are labels used in model development. On the other hand, the synVep output indicates the syn-Vep predicted score, and "effect' is the synVep binary prediction. SNV, rs2229069 was classified as 'observed', with a predicted score of 0.0055 however no clinical effect. On the other hand, SNV rs2227985 was classified as 'observed' with a predicted score of 0.848 and with a clinical effect. However, the software does not allude to what the effect is, more so that there is a possible effect.

Table 3.11. SynVep results for variants of interest

Gene	Ensembl	SNV	N variants queried	N effect	N no-effect	Fraction effect (%)
ABL1	ENST00000318560	rs2229069	1	0	1	0.00
		rs2227985	1	1	0	100.00

Table 3.12. SynVep Effect results for variants of interest

Gene	Ensembl	dbSNP	Ref	Alt	Class	synVep	Effect	Transcript Position	Strand	Codon	Position _hg19	Position _hg38
ABL1	ENST00000318560	rs2229069	G	A	observed	0.0055	no effect	720	+	AC**G**->AC**A**	133738320	130862933
		rs2227985	A	G	observed	0.8480	effect	1497	+	GA**A**->GA**G**	133755528	130880141

3.5.3.3 MutationTaster

MutationTaster, another *in silico* tool,was used to help predict the disease causing potential of DNA variants (Steinhaus et al., 2021). This tool subjects each variant to an array of *in silico* tests, where the software predicts the impact of variants and provides various analyses.

A "Known variant "output is given if a variant has been found in large scale sequencing projects or listed as disease mutations in ClinVar. Variantsfrom the 1000 Genomes Project (1000G), ExAC and gnomAD are also incorporated and filtered via MutationTaser and determines whether the alleles are homozygous (-/-) or heterozygous (-/+). On the other hand, the PhyloP and phastCons phylogenetics methods that are build into the MutationTaster *in silico* tool are used to determine the grade of conservation of a given nucleotide and produces a concervation score. phastCons values vary between 0 and 1 and reflect the probability that each nucleotide belongs to a conservant element, based on the multiple sequence alignment of 46 different species. Thus, the closer the value is to 1, the probability of conservation is high and vice versa.. It considers each individual alignment column and its flanking column. On the other hand, PhyloP values between -14 and +6 separately measures conservation at individual columns. It can measure conservation and acceleration (faster than expected). Thus, sites predicted to be conserved are assigned positive scores, while sites predicted to be fast-evolving are assigned negative scores. Another score output is the ExAC pLi scores (gene constraints metrics) which are retrieved from the gnomAD database and assess the tolerance of a gene against loss-of function variants. The Konzak consensus sequence which are a nucleotide motif that starts upstream of the codon (AUG) and plays a major role in the initiation of translation were also analysed to determine whether the SNV's have any effect on it.

For the rs2229069 variant, a benign effect was predicted, without any amino acid changes similar to results outputs from previous analysis. For the rs222785 variant the software predicted a benign effect, without any amino acid changes as well **(Table 3.13).**

Table 3.13. MutationTaster outputs for variants of interest

Analysis Indices	rs2229069				rs2227985			
Prediction	benign				benign			
ExAC LOF metrics	LOF: 1.00, misssense: 2.66, synonymous: 0.83				LOF: 1.00, misssense: 2.66, synonymous: 0.83			
Ensembl transcript ID	ENST00000318560.5				ENST00000318560.5			
Genbank transcript ID	NM_005157 (exact from MANE)				NM_005157 (exact from MANE)			
Variant type	Single base exchange				Single base exchange			
Gene region	CDS				CDS			
DNA changes	c.720G>A				c.1497A>G			
	g.148988G>A				g.166196A>G			
AA changes	no AA changes				no AA changes			
Frameshift	No				No			
Length of protein	Normal				Normal			
Allelle Frequency	database	homozygous (A/A)	heterozygous	allele carriers	database	homozygous (G/G)	heterozygous	allele carriers
	1000G	0	65	**65**	1000G	**33**	363	396
	ExAC	10	441	**451**	ExAC	**379**	7285	7664
	gnomAD	20	1103	**1123**	gnomAD	**872**	17027	17899
Phylogenetic conservation		*PhyloP*	*PhastCons*			*PhyloP*	*PhastCons*	
	(flanking)	5.716	1		(flanking)	4.998	1	
		-2.466	0.006			-0.119	0.976	
	(flanking)	5.716	1		(flanking)	4.998	1	
Splice sites	No abrogation of potential splice sites				No abrogation of potential splice sites			
Distance from splice site	N/A				17			
Kozak consensus sequence altered?	No				No			
AA sequence altered	No				No			
Chromosome	9				9			
Position of stopcodon in wt / mu CDS	3393 / 3393				3393 / 3393			
Position (AA) of stopcodon in wt / mu AA sequence	1131 / 1131				1131 / 1131			
Position of stopcodon in wt / mu cDNA	3774 / 3774				3774 / 3774			
Position of start ATG in wt / mu cDNA	382 / 382				382 / 382			
Last intron/exon boundary	2059				2059			
Theoretical NMD boundary in CDS	1627				1627			
Length of CDS	3393				3393			
Coding sequence (CDS) position	720				1497			
cDNA position	1101				1878			
gDNA position	148988				166196			
Chromosomal position	133738320				133755528			

Abbreviations: Exac – Exome Aggregation Consortium; LOF – Loss of Function

3.5.4 Population Allele frequency - The 1000 Genomes Project

The allele frequency for the respective SNV analysed were also obtained from the 1000 Genomes project, which houses a catalogue of human variations from different populations. The graphs below (Figure 3.13) portray the rs2229069:c.720G>A and rs222785:c.1497A>G allele frequencies, which was established in phase 3 of the 1000 Genomes Project. The population frequency that was included and compared is from Africa (AFR), America (AMR), East Asian (EAS), European (EUR) and South Asian (SAS). In most of the populations, the ancestral G allele for SNV, rs2229069: c.720G>A occur at a total frequency of 99-100% compared to 95% in the African population. The alternative A allele occurs at a frequency of 5% in the African population which is the highest frequency compared to either very low frequency or no-existence in the other populations **(Figure 3.13)**. Interestingly the ancestral A allele for SNV, rs222785: c.1497A>G occurs at a lower frequency ranging between 3-9% for most of the populations, while it is 0% for the East Asian population. On the other hand, the highest frequency occurrence for the A allele is 21% for the African population. The alternative G-allele occurs at a frequency of 100% in the east Asian population and the lowest frequency at 79% are observed in the African population **(Figure 3.13)**.

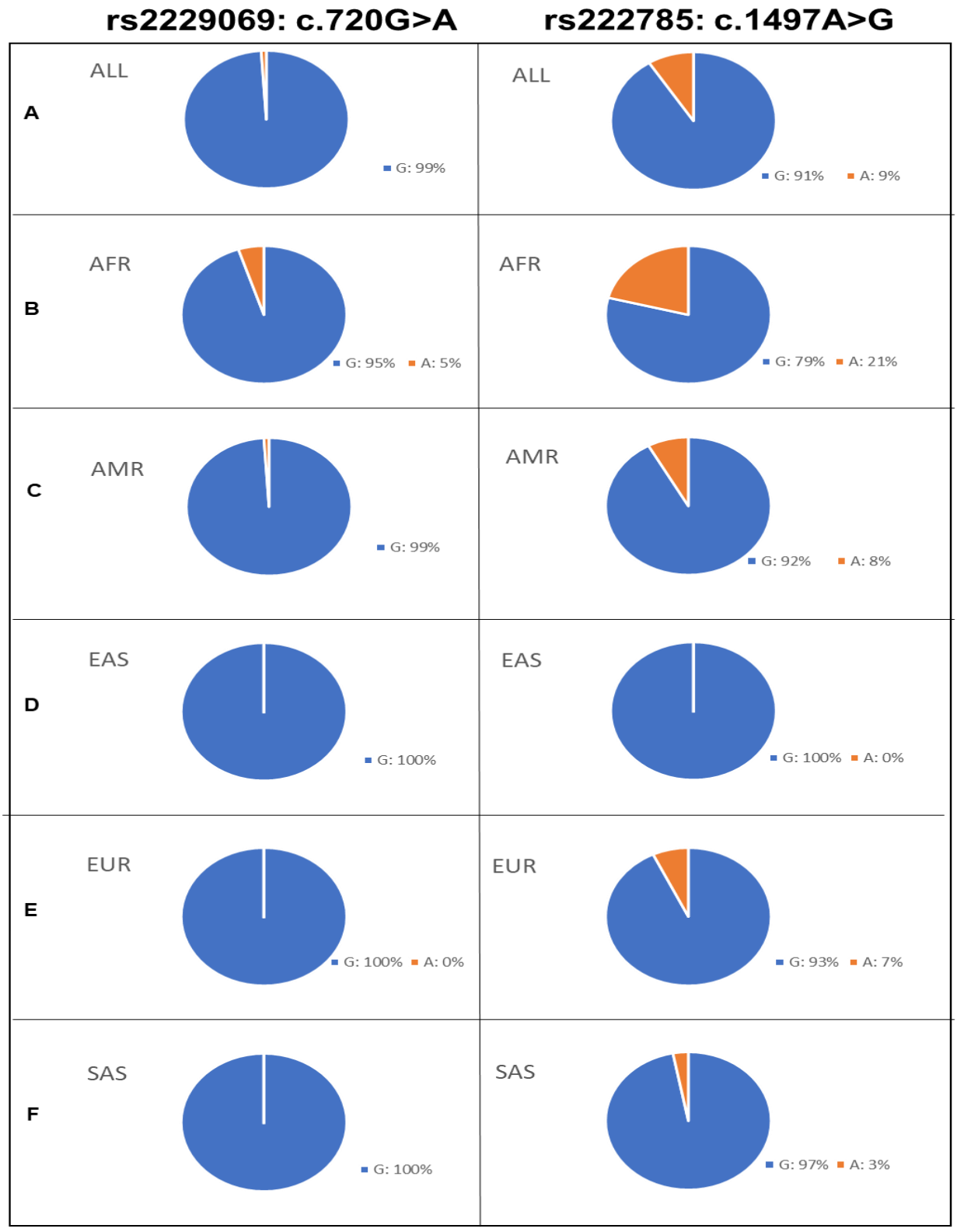


Figure 3.13. 1000Genomes Project's population allele frequencies for variants of interest

Abbreviations: ALL – total overall, AFR- Africa, AMR- AmericaEAS – East Asian, EUR – European, SAS – South Asian

3.5.5 Participants samples expressing variants

In summary, our study revealed heterozygous coding sequence variants in 11 out of the 25 samples investigated **(Table 3.14).** Additionally, the infamous *T315I* mutation which is the most frequently identified mutation in *CML* TKI-resistant patients has not been identified in our cohort. The table below also shows the SNV's detected in each sample. Samples 1-9 presented with the rs2227985 variant in exon 9, and 10 and 11 presented with the rs2229069 variant in exon 4. Interestingly, out of the five baseline samples included in the study, only one harboured the SNV, rs2227985 which interestingly was already present at baseline prior treatment. The same SNV was detected as expected in the sample which was collected at least 12 months after the commencement of the treatment. For the rest of the samples where variants were observed, no baseline samples were available.

Table 3.14. Samples screened expressing variants

	Sample ID	Exon/Region	SNV
1	**280-6**	**9**	**A/G – rs2227985**
2	**322-2**		
3	**329-2**		
4	**360-2**		
5	**361-6**		
6	**378-4**		
7	**381-1**		
8	**381-4**		
9	**245-6**		
10	**248-9**	**4**	**G/A – rs2229069**
11	**143-20**		

Abbreviations: SNV – single nucleotide variant

Additionally, we looked at the IS% data in relation to these patients as an indication of how they fare on treatment. **Figure 3.14** shows the graphical representation of each participant IS%. The graph has been formatting to be viewed more closely, and we can see that participants' 378 and 322 exceeds the graphs maximum value of 130%, and they reach 220% and 180%, respectively **(Table 3.15.)**. The dotted line indicates the point of IS% of participants at one year on treatment. A 3-log reduction in transcript from the standardised baseline value needs to be observed to define a Major Molecular Response **(Table 1.1)**. It can be observed that five participants have IS% <10% at 12 months on treatment. And four have IS%>10%. One of the patients had not been on treatment for an entire year at the time of review.

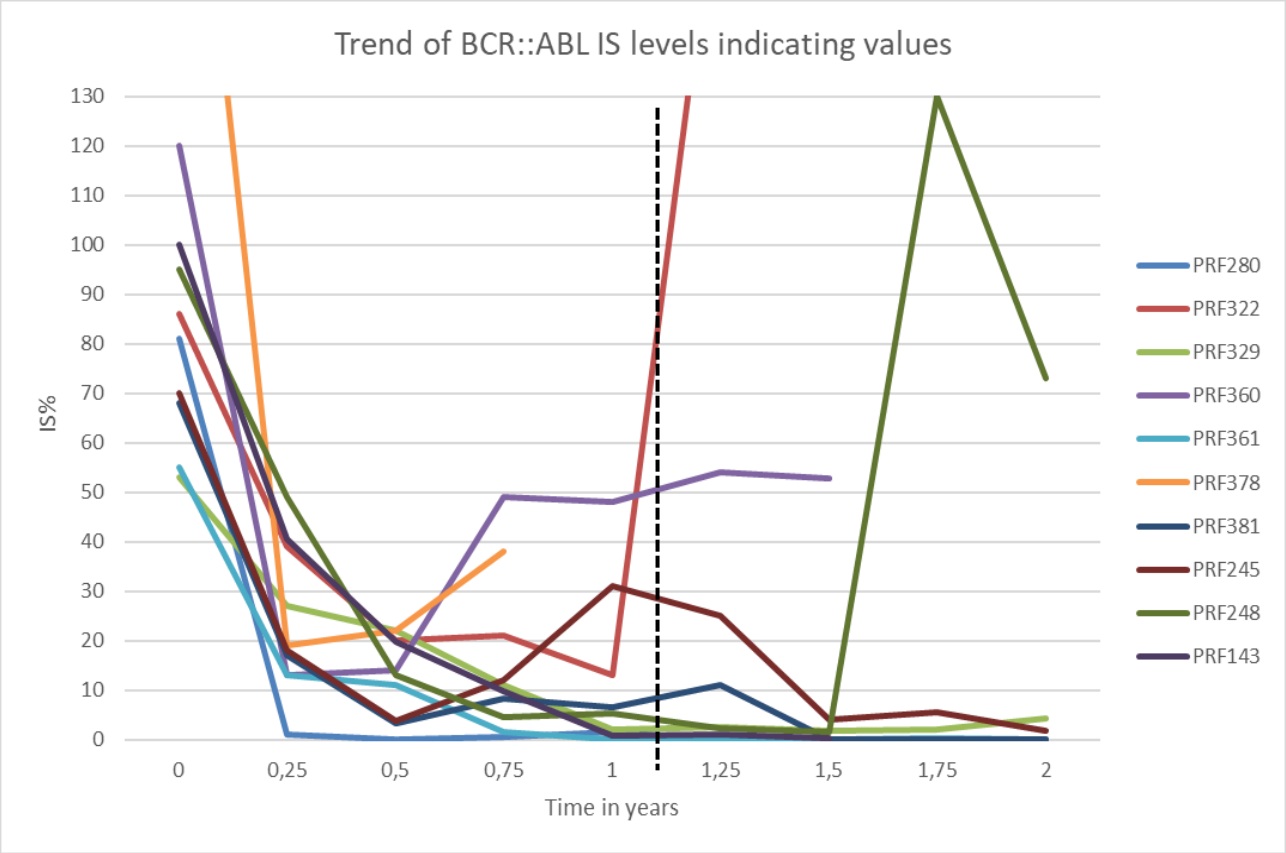


Figure 3.14. IS% response trend of patients with single nucleotide variants A/G – rs2227985 and G/A – rs2229069. The graph depicts changes in the BCR::ABL mRNA transcripts over the course of therapy, reported as an international standardised percentage.

Table 3.15. IS% percentages in relation to time in years of participants presenting with variants

Time (years)	IS% of respective participants									
	P280	**P322**	**P329**	**P360**	**P361**	**P378**	**P381**	**P245**	**P248**	**P143**
0	81	86	53	120	55	220	68	70	95	100
0,25	0,91	39	27	13	13	19	17	18	49	40,55
0,5	0,083	20	22	14	11	22	3,2	3,9	13	19,66
0,75	0,44	21	11	49	1,6	38	8,4	12	4,5	9,84
1	1,6	13	2,1	48	0,11		6,6	31	5,2	0,73
1,25	0,034	180	2,6	54	0,087		11	25	2,4	1,11
1,5	0,032		1,9	52,69	0,077		0,041	4	1,6	0,23
1,75	0,013		2		0,27		0,023	5,6	130	
2	0,0024		4,4		0,0012		0,0044	1,9	73	

Time in years is given in 3-month intervals, i.e., 0 years would indicate diagnosis, 0.25 years would be 3 months on treatment, 0.5 years would be 6 months on treatment, etc.

Participant 381 **(no. 7 and 8 in Table 3.14)** is of particular interest, as this participant's samples at both baseline and at 12 months follow up contained the variant of interest. Baseline samples, as stated in the methods were provide before commencement of treatment. To reiterate, the ELN defines primary resistance as not achieving complete haematological response (CHR) by 3 months, no cytogenic response by 6 months and no major molecular response by 12 months. Secondary resistance is defined as the patient losing the previously achieved haematologic or cytogenetic response and patients exhibiting a consecutive 1 log increase in transcript level after achieving MMR. In **Figure 3.15** it can be observed the participant 381 had an IS% of > 10% at three months (0.25 years), which classifies this participant within the 'warning' group of treatment failure. At 12 months, the participant 381 had an IS% > 1%, which would classify the participant within the 'failure' group according to ELN treatment resistance recommendations **(Figure 3.15)**.

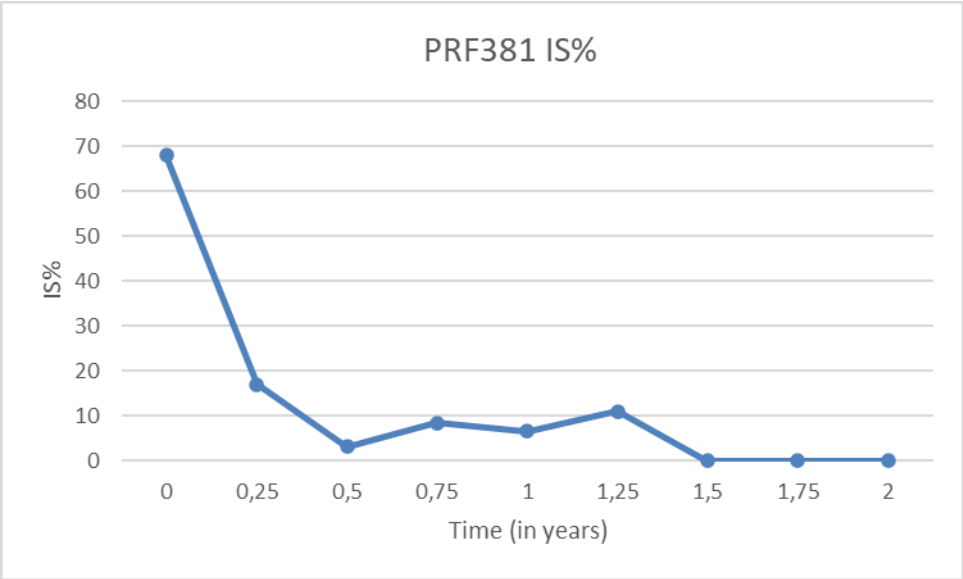


Figure 3.15. *BCR::ABL1* IS% Log graph of participant 381, displaying variant of interest at baseline and at 12 months follow up

However, the patient achieves an IS%< 0.1 at 1.5 years **(Table 3.16)**, and this drop is consistent up until two years on treatment. Clinical data obtained with relevant permissions allures to non-compliance **(Table 3.16.1)**. No change in treatment to second-line therapy was needed and interventions included discussions about the importance of compliance with treatment, which would effectively alleviate symptoms and the patient's treatment.

Table 3.16. IS% of participant 381 relative to years on treatment

Time (Years)	IS%
0	68
0.25	17
0.5	3.2
0.75	8.4
1	6.6
1.25	11
1.5	0.041
1.75	0.023
2	0.0044

In summary, the variant present within the kinase domain of participant 381 had no effect on transcript levels, nor on response to treatment. Once the participant had resumed compliance, required molecular targets were reached. This emphasises the need for an individualistic approach into the investigations of treatment responses.

The clinical data of these rest of the participants who displayed sequence variants was also looked at **(Table 3.16.1).** Six of the participants wherein the chronic phase of disease, one in the accelerated phase, one participant had progressed to blast crisis phase and one participant had transformed from CML to B-ALL.

Table 3.16.1. Clinical findings of CML resistant cohort selected for KD screening

ID	HIV status	Comorbidities/symptoms	CML disease phase	Treatment compliance
381	Neg	Schizophrenia, Asymptomatic leucocytosis	CP	Compliance issues
329	Neg	Splenomegaly	CP	Compliance issues
322	Pos	TB, thrombocytosis	BC	Adherent
360	-	Iron deficiency		Compliance issues
280	Neg	Diabetes mellites	CP	-
378	-	Splenomegaly, additionally Ph+ chromosome in 50% on FISH,	B-ALL	-
245	-	Paraplegia (spinal cord injury); blind, type 2 diabetes, HP, Cholesterolemia, chronic kidney disease, pancytopenia	AP	Compliance issues
248	-	Peptic ulcer disease, splenomegaly, foci of fibrosis	CP	Compliance issues
143	-	Pedal oedema, splenomegaly, hepatomegaly, leucocytosis	CP	-
361	-	Splenomegaly	CP	-

Abbreviations: Neg, negative; Pos, positive, TB, Tuberculosis, (-) indicates no result found

Furthermore, compliance issues and comorbidities seem to be regular phenomena in this cohort, with splenomegaly the most common observation. No HIV status was available for six of the 10 patients and testing thereof was not indicated. The patients highlighted in red were found to be deceased; both presented with the exon 9, rs2227985 variant. This is of interest, and although the relation might be spurious, it will be further investigated.

3.5.6 Protein Data Bank

Bioinformatic *in silico* tools analysis also included protein modelling. Even though synonymous sequence variants with no protein change were detected, we continued simulations on the ABL1 structure to confirm that no effects but more importantly showcase competence with these *in silico* tools. Molecular docking of the wild type ABL1 structure with Imatinib was also carried out along with STRING analysis to determine direct interactions with the ABL1 wild type structure.

The ABL structures was retrieved from the Protein Data Bank (https://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000097007;r=9:130713016-130887675;t=ENST00000318560;v=rs2229069;vdb=variation;vf=729530037). **Figure 3.16** gives a representation of the 3D tyrosine protein model and each colour represent an exon with colour as well as location given in **Table 3.17.**

.

Figure 3.16. PDB model of ABL structure

Table 3.17 Protein Data bank ABL structure displaying each exon

Exon	Representative colour	PDB model location
Exon 4	Purple	248-293
Exon 5	Pink	294-322
Exon 6	Orange	322-381
Exon 7	Yellow	381-443
Exon 8	Neon green	443-494
Exon 9	Green	494-524
Exon 10	Blue	524-534

The PDB model (6XR7) of the first variant observed in exon 4 (rs2229069;c. 720G>A) can be seen in **Figure 3.17**. The purple highlighted area represents Exon 4, and the variant location is indicated by the arrow. **Figure 3.18** provides a posterior view of the model.

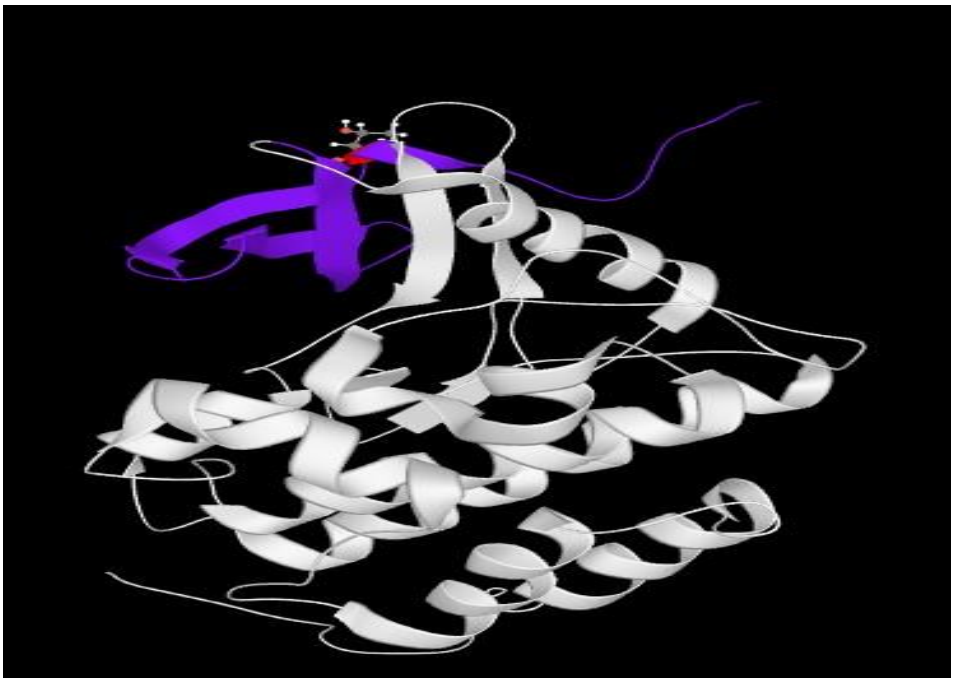


Figure 3.17. Anterior view of PDB model with location of variant rs2229069

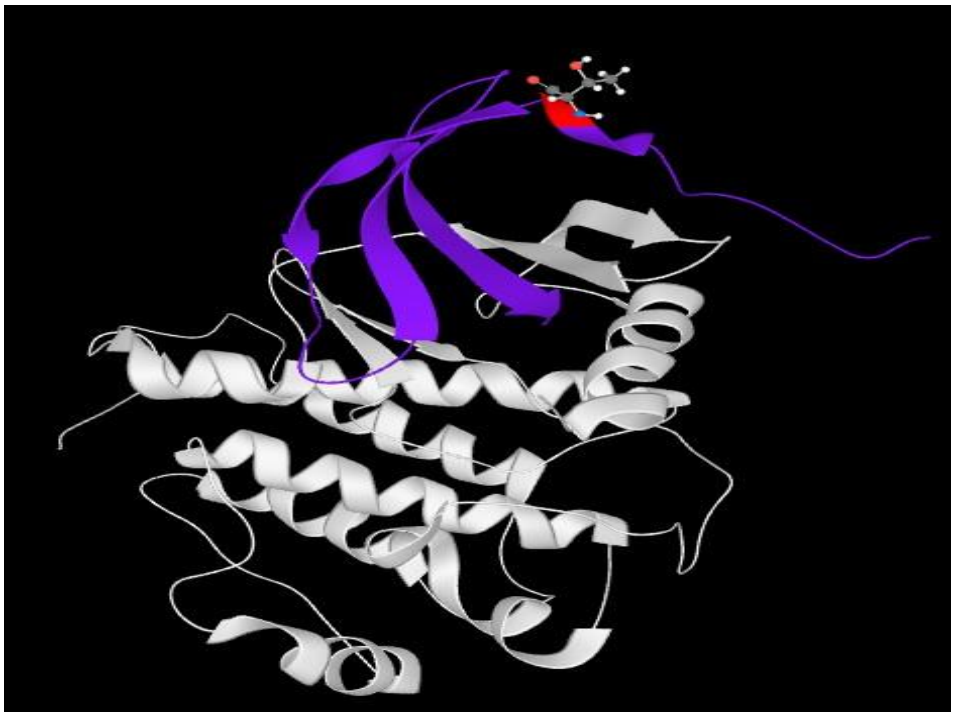


Figure 3.18. Posterior view of PDB model with location of variant rs2229069

The PDB model of the second variant which was identified in exon 9 (rs2227985:c. 1479A>G) can be seen in **Figure 3.19.** The green highlighted area represents Exon 9, and the variant location is indicated by the arrow. **Figure 3.20** provides a posterior view of the variant.

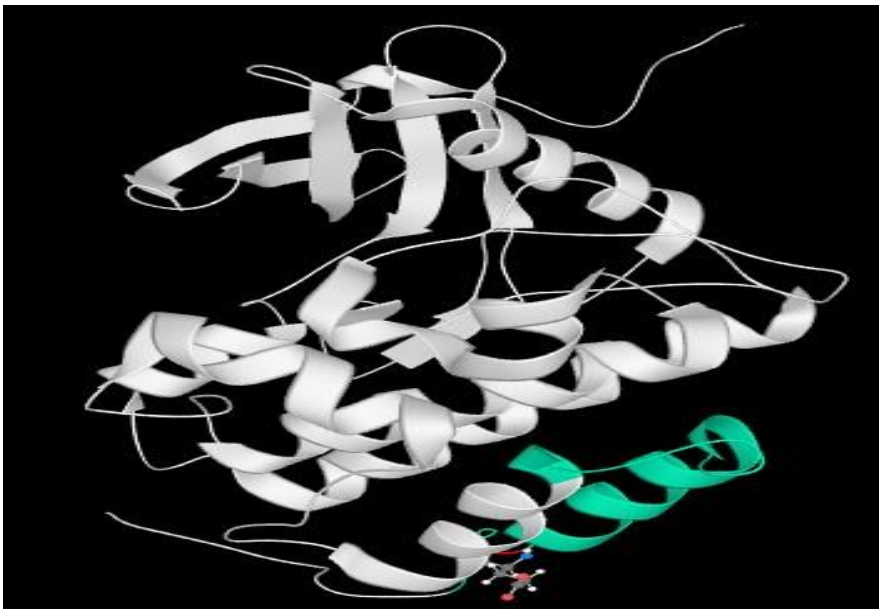


Figure 3.19. Anterior view of PDB model with location of variant rs2227985

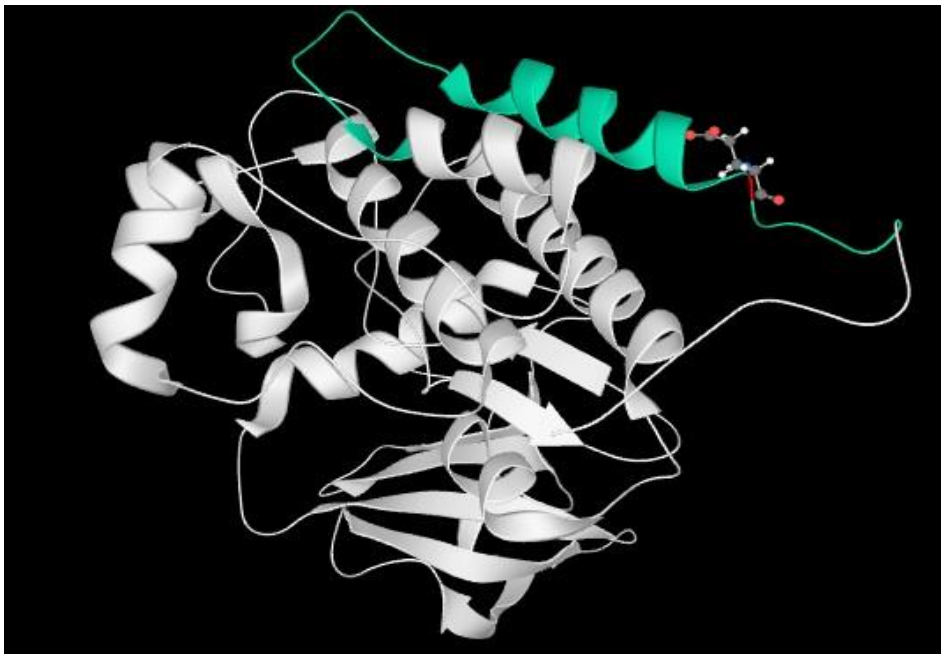


Figure 3.20. Posterior view of PDB model with location of variant rs2227985

Being able to view the exact location of variants of gene structures allows a gateway into investigating which structures and consequently functions could possibly be implicated. The variant of interest is highlighted in red on the structure. If a variant resulted in protein truncation or a frameshift, the peptide chain after the truncation will be displayed in dark red. However, both the variants in question are displayed in bright red. However, Protein truncations and frameshifts, if present, can be explored with a variety of tools. They also provide various protein information, as well as SIFT and PolyPhen scores associated with the variants. Unfortunately, these were not applicable to the variants observed in this study.

3.5.7 Molecular Docking with Molecular Operating Environment (MOE)

We also performed molecular docking of the wild-type structure with Imatinib using MOE (MOE 2011.10 Chemical Computing Group Inc) is an integrated computer-aided molecular design platform. It has various interfaces for various applications. However, this software is not free resource and requires a purchased license for usage. The ABL structure for this application was retrieved from the Protein Data Bank (PDB: 1OPL) (http://www.rcsb.org/pdb/explore/explore.do?structureId=1OPL) with the resolution of 3.42 Å and was loaded into the MOE working environment ignoring all water molecules and heteroatoms. The optimised conformation of Imatinib was docked into the active site of the ABL structure. The conformations are ranked according to London dG scoring function, which essentially estimates the binding energy of a ligand (Galli et al., 2014). Results indicated imatinib having a docking score of -9.1593 kcal/mol by interacting with the ABL1 structure. **Figure 3.21** is a representation of the ligand with the binding residues.

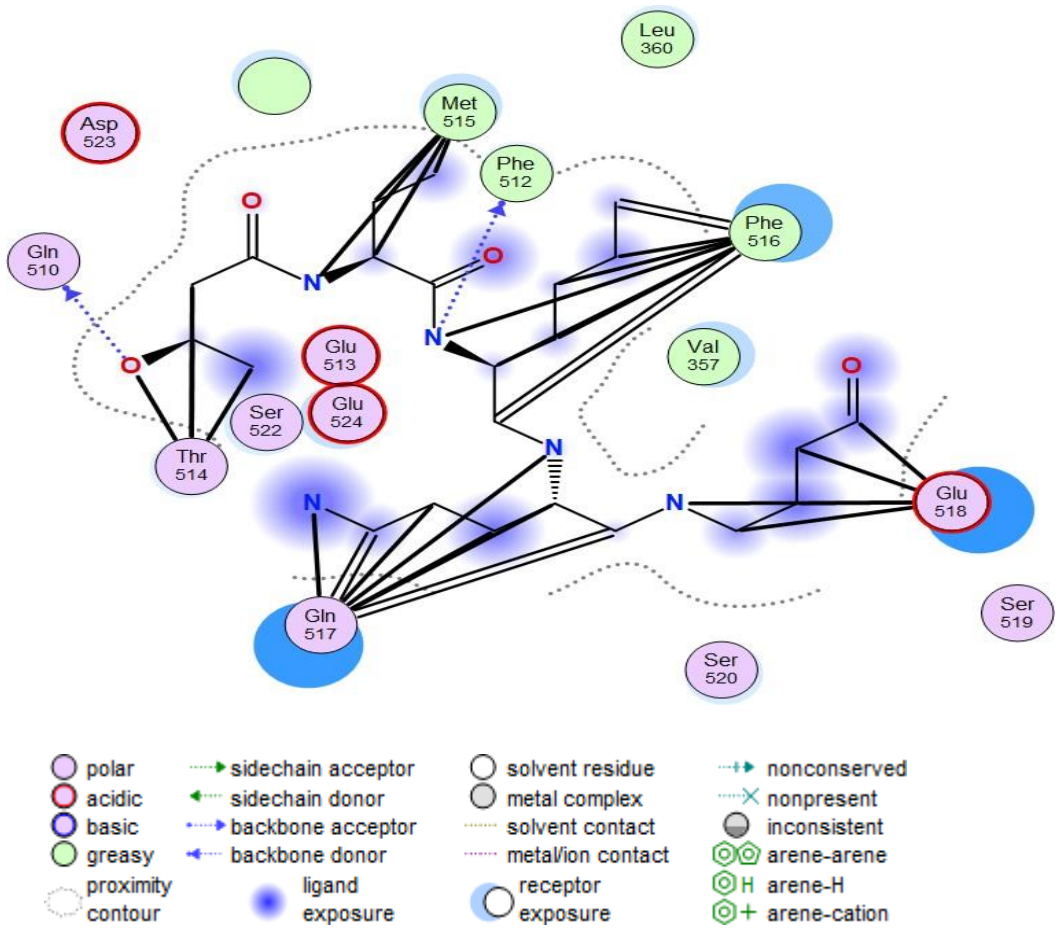
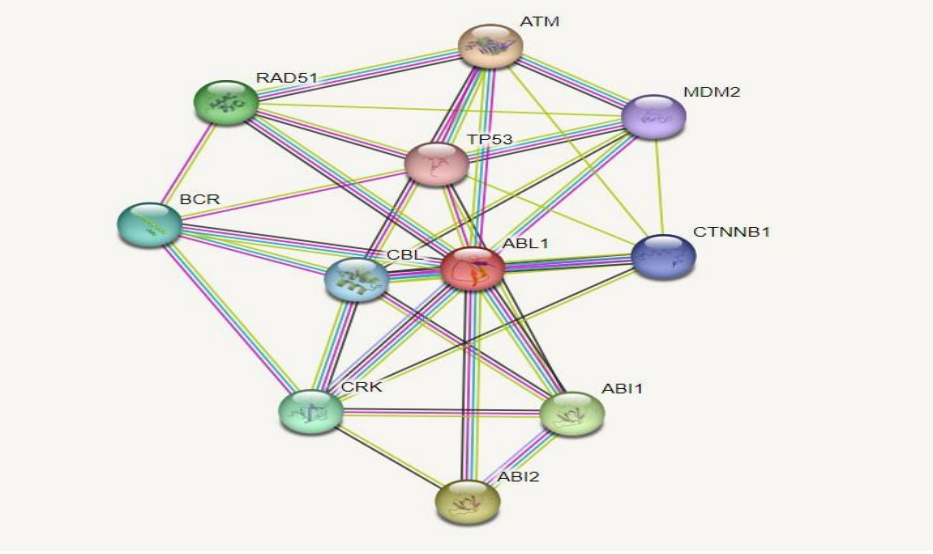


Figure 3.21 Docking of ABL wild type structure with Imatinib

3.5.8 STRING

We also looked at the STRING database to identify whether there are any interactions with BCR::ABL structure. The result indicates that there are direct interactions between the ABL1 and CTNNB1 and CBL **(Figure 3.22)**. The interactions include co-expression (black line), experimentally determined interactions (pink line), interactions from curated databases (blue line); and text mining (yellow line). CTNNB1, Catenin Beta-1, has been suggested to play a role in chronic phase progression (Chen et al., 2013). This is a multifunctional protein that contributes to cell development under normal physiological conditions and is a crucial transcriptional factor as well as plays a role in stem cell renewal and organ regeneration (Tomasello et al., 2020). In 2013, an American study identified a ubiquitin-binding domain in the NH2 terminal of the p210 transcript and found that it co-localises with the binding site for beta-catenin. They allude to the fact that p210 BCR::ABL may influence lineage-specific leukaemic expansion by directly binding and phosphorylating beta-catenin and altering its transcriptional activity (Chen et al., 2013). CBL, Cbl proto-oncogene, E3 ubiquitin protein ligase, also displays ubiquitination biochemical activity. This essentially is an interaction inferred from the biochemical effect of on protein on another. In 2010, a Japanese study found that CBL induced ubiquitin-dependent degradation of mature and phosphorylated BCR::ABL1 proteins, and efficiently supressed BCR::ABL-dependent leukaemic growth (Tsukahara & Maru, 2010).



Abbreviations: BCR- break point cluster, ABL – Abelson murine, CTNNB1- catenin (cadherin-associated protein) beta 1, CBL-Cbl E3 ubiquitin protein ligase

Figure 3.22. Protein interactions with BCR::ABL1 proto-oncogene

Chapter 4

Discussion

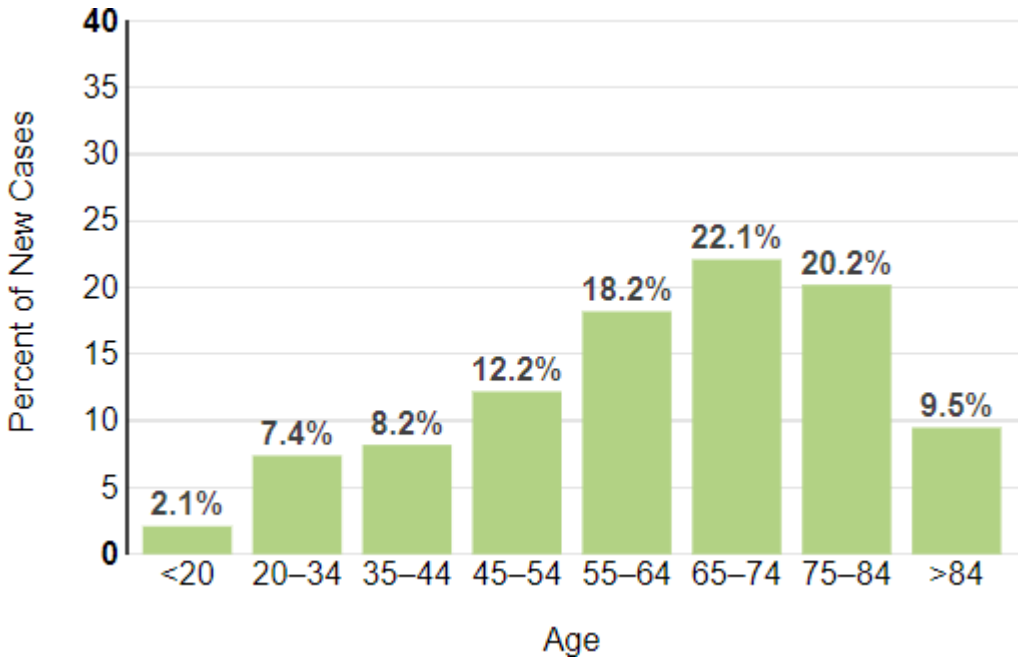
Over the years there has been an increase in drug resistance cancers while the mechanisms that underlie these drug resistances are not well understood. Various categories of mechanisms that promote drug resistance, independently or in combination through various signal transduction pathway has been proposed. This study specifically focuses on the most common mechanism (60%) for Imatinib resistance which are due to mutations in the BCR::ABL kinase domain that spans over exon 4-9 of the *ABL1* portion of the *BCR::ABL1* fusion gene. These mutations compromise inhibitor binding and subsequently lead to increased kinase activity or the inability of Imatinib to bind to its target due to alterations of the target protein (Kotagama *et al.*, 2016, Quintás-Cardama, Kantarjian and Cortes, 2009). More than 100 different mutations have been reported (Jabbour & Kantarjian, 2020), but less than 20 of these mutations are frequently identified in clinical practice (Tanaka et al., 2011). The most frequent mutation identified in CML drug resistance patients is the T315I mutation which confers complete resistance to Imatinib, Dasatini, Nilotinib, and Bosutinib(Yang & Fu, 2015).

Thus, the need to identify these mutations within our South African CML cohort was very relevant as knowledge regarding our patient cohort mutational landscape have an impact on clinical treatment decision making processes. While most diagnostic labs implement and screen for 2nd mutations within the BCR::ABL1 kinase domain, no incidence data related to these mutations is available in the South African context. Thus, we aimed to implement and validate the diagnostic screening of the ABL1 kinase domain and p-loop region with the hope to identify known and novel 2nd sequence nucleotide variants. Further *in silico* analysis was also done to identify potential diagnostic or prognostic markers for therapeutic strategies.

4.1. Incidence of TBH CML cohort

From the audit, 165 cases of CML were diagnosed in the review period of January 2013 to December 2020. The median age of the TBH CML cohort was 45-46 years old, with the youngest patient being 2 years old, and the oldest being 88 years old. There was also a slight predominance of males, 53.9% (n=89) compared to females at 46.1% (Table 3.1, page 27). Furthermore, patient geographical distributions show that most of the patients were found to be from Khayelitsha, Delf and Strand. Our results regarding age deviates slightly from global statistics **(Figure 4.1**), where the graph leans more

to the right, indicating an older age range at presentation, as opposed to our graph where the peak is centred **(Figure 3.1, Page 29),** indicating a younger age at presentation.



Extracted from SEER Cancer Stat Facts: Chronic Myeloid Leukemia. National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/statfacts/html/cmyl.html

Figure 4.1. SEER CML statistics relative to Age at presentation

Moreover, Statistics from the National Cancer Institute's Surveillance, Epidemiology, and End Results Program (SEER) indicate that the median age of CML patients is 65 years, with CML being diagnosed most frequently among people aged 65-74 and is more common amongst men. Patients admissions per year were also investigated and since data collection for our cohort spanned from 2013 to 2020, two years fell in the COVID pandemic period. As the COVID pandemic warranted national lockdowns which started on Thursday 26 March 2019 this massively impacted patient's admissions numbers for 2020-2021.

The world is currently undergoing a Pandemic which started in 2019 and warranted national lockdowns across the globe due to the rapid spread of disease caused by SARS-CoV-2. The South African President instated the lockdown to begin at mid-night on Thursday 26 March 2020. Looking at the effects on admission of patients which lead to the consequent diagnosis of CML during that period, results for

patients' admissions showed that 10.9% (n=18) CML patients were diagnosed during 2019 and 8.5% (n=14) during 2020. Interestingly, these results did not deviate significantly from the pre-pandemic admissions rates, which was found to be 7.3% (n=12) during 2018, 12.1% (n=20) during 2017, 7.9% (n=130 during 2016 and 7.9% (n=16) during 2015. The years 2013 and 2014 accounted for the highest rates of admissions, at 21.2% (n=35) and 22.4% (n=37), respectively (Table 3.3, Page 30)

CML patients are treated with Gleevec®-Imatinib mesylate which is a ABL1 tyrosine kinase inhibitor that reduces *BCR::ABL1* gene expression or eliminate it. Thus, the efficacy of these CML treatment responses are gauged via monitoring of the haematological (return of blood counts to normal), cytological (reducing and elimination the Ph chromosome) and molecular (reducing *BCR::ABL1* gene expression) responses per international guidelines. Thus, it is important for patients to continuously receive these treatments until treatment free remission have been achieved. Sadly, patient loss to follow-up (LTFU) is quite common in a public hospital setting, thus it was not that surprising that only 63% (n=104) of the patients were still actively attending the Haematology Clinic for routine diagnostics and molecular monitoring after 1 year follow-up. An astounding 37% (n=61) were lost to follow-up, and information on the cessation of routine monitoring and treatment of these patients are limited. Possible reasons for LTFU especially in a resource-limited setting may be financial difficulty, lack of patient knowledge that treatment needs to be lifelong, hospitalisation or illness, medication toxicity, and death. The frequency of gender distribution related to clinic attendance after a 1-year period was also analysed and there was no statistical significance found based on gender (p=0.676) of the 104 patients.

For molecular monitoring, *BCR::ABL1* levels are monitored by RQ-PCR which measures *BCR::ABL1* mRNA transcript levels according to an international scale, thus treatment efficacy responses are based on IS %. As we were interested to measure how strong the relationship/association is between various variables, the Pearson's Correlation Coefficient test was used to identify associations between age and *BCR::ABL1* mRNA transcript levels. A value between -1 and 1 are calculated, where 1 indicates a strong positive relationship and -1 indicates a strong negative relationship, and zero indicates no relationship at all. Thus, the association between age and IS% values of the 104 patients, actively undergoing CML treatment were compared at diagnosis, and then at 1 year follow-up of CML treatment Results showed a p value of -.077, indicating a weak negative relationship between age and *BCR::ABL1* mRNA transcript level, thus age was not associated with an higher IS% within our cohort. This could further be seen when comparing the data graphically, where a random dispersion was observed, with no clear patterns or trends (Figure 3.2, page 32). This also corresponds to a 2020 study based on an Algerian population (Nachi et al., n.d.). Another study conducted in India, shows that the median age of CML at

diagnosis was astonishingly 35 years old and investigated the clinical profiles and outcomes of elderly versus younger patients. They categorised patients older than 60 years as the elderly, and found that although this group presented with more comorbidities than the latter, there was no statistical significant difference between the groups regarding presentation in the blast phase nor in achieving haematological or major molecular responses (Nagendrappa et al., 2020).

The impact of Gleevec®-Imatinib mesylate therapy on *BCR::ABL1* mRNA transcript from the diagnosis starting point and 12 months after treatment was also investigated. The Wilcoxon Signed Ranks Test, which is a non-parametric test, was used to identify any differences between data that does not have a normal distribution. A statistical significance of .000 was observed between *BCR::ABL1* mRNA transcripts levels at the start of therapy compared to at 12 months.

4.2. TBH Resistant CML cohort

Fifty-seven (35%) of the 165 CML patients were characterized as resistant to first-line therapy, according to ELN guidelines. Of the 57 resistant CML patients, 77.2% (n=44) were found to be within the failure stage of resistance, and 21.1% (n=12) were in the warning stage. One patient, who was deemed/classified to be in failure stage, response to treatment was unclear. Descriptive statistics of the resistant cohort was also done to identify any trends. The results indicated that 45.6% (n=26) of these patients were female, with a slight male predominance of 54.4% (n=31), with the median age being 43.3 years, with the youngest resistant patient being two years old and the eldest being 75 years old (Figure 3.3, Page 35).

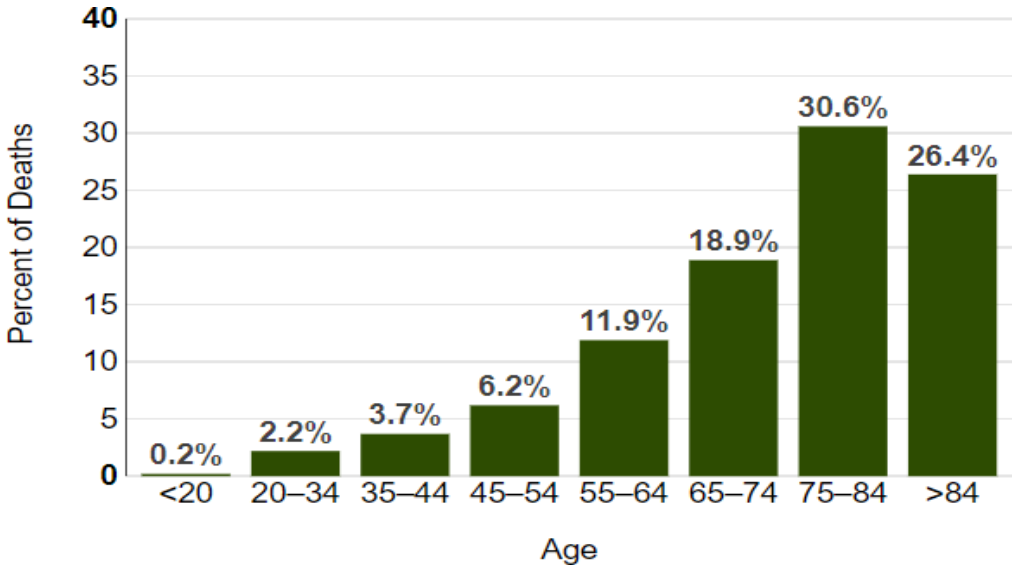
4.3. TBH CML Screening cohort

Of the 57 resistant patients, 20 patients were selected for the subsequent sequence variant/mutation screening portion of the sub-study. Screening cohort selection was based on biospecimen availability at diagnosis and 1 year follow ups as well as biospecimen integrity and availability of all relevant clinical data. Determining whether any sequence variants/mutations were present or absent at diagnosis and the 1 -year follow-up period for each of the selected CML patients were also of importance. Within the cohort of 20 screened patients, 55% (n=11) of the patient's samples were female while 45% (9) were male. Seventy five percent (n=15) of these patients were categorized within the failure stage, as compared to 25% (n=5) who fell in the warning stage of treatment resistance according to ELN guidelines. Interestingly, most of the resistant patients were diagnosed with CML during 2019, accounting for 35% (n=7) of the screening cohort.

Five of the participants who displayed sequence variants achieved and IS% of less than 10% at 12 months, and four had more than 10% IS. The BCR::ABL log charts enable treatment response, and alludes to whether the participant is indeed non responsive (consistent increase in transcript) or whether other factors, such as non-compliance could be contributing factors. This was further emphasised by taking a closer look at one participant, 381, where the log chart clearly indicates compliance issues as there is a steady, but delayed response to remain. This renders issues as the participant would be flagged as being resistant according to ELN guidelines due to not meeting response goals timeously. This reinstates that careful monitoring of CML patients is required. Interestingly, participant 381 had variant rs2227985 present in the sample screened at diagnosis, as well as at 12-month follow up. It is therefore clear that the variant did not have an impact on the participants response to treatment.

Additionally, there was a 20% (n=4) mortality rate amongst the patients selected for screening and the age at death was found to be an average of 39.2 years These participants had progressed to advanced phased of disease (AP and BC), with one transformation into B-ALL and all (100%) of these patients had survived less than five years. Data based on SEER statistics indicate that the 5-Year relative survival is 70.4%, where the median age at death was found to be 77 years old, with death being highest amongst people aged 75-84 years (Figure 4.2).

Although progression to advanced phases of disease is associated with worse prognosis, our data on a small number of participants is alarming and warrants further investigation, as it contradicts literature. A European study conducted by (Lauseker et al., 2019) found the median survival of patients with advanced disease states to be 8.2 years, while our participants had a relative survival rate of less than 5 years.



Extracted from Extracted from SEER Cancer Stat Facts: Chronic Myeloid Leukemia. National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/statfacts/html/cmyl.html

Figure 4.2. SEER statistics percent of deaths relative to age

4.4. TBH CML BCR::ABL1 kinase domain variant screening outcome

Five of the 20 patients had baseline samples (taken at diagnosis), which allowed for comparative analysis. We therefore screened 25 samples for sequence nucleotide variants (SNV) via bidirectional Sanger Sequencing.

Two heterozygous synonymous SNV were identified during the mutational screening of our CML patient cohort. Namely, a G/A variant at codon 720 in exon 4 (rs2229069 - NM_005157.6(ABL1): c.720G>A (p.Thr240Thr) and a A/G variant at codon 1497 on exon 9 (rs2227985 - NM_005157.6 (ABL1):c.1497A>G (p. Glu499Glu). Additionally, rs2227985 (c.1497A>G) was found in nine of the samples screened of which one was a baseline sample. This indicates that the variant was found in two samples screened for the same patient from bloods drawn at different times. The variant was present in the baseline sample, which was drawn before the commencement of treatment, and found in a sample collected at 12 months of treatment. This highlights that the variant was present prior treatment and that nucleotide sequence changes observed at this locus was not due to the CML treatment. On the other hand, the second variant rs2229069 (c.720G>A) identified was found only in two of the samples screened however no identification in a baseline sample.

The variants were verified using the Ensembl Genome Data Browser using the transcript ID ENTS00000318560.6. The Ensembl (https://www.ensembl.org/) bioinformatics tool was employed to gain biological information that are integration from biological data derived from the genomic sequences and other linked sequence databases. To assess variants for and against pathogenicity and interpret its clinical impact, a standard framework was developed in 2015 by the American College of Medical Genetics and Genomics (ACMG)/ Association for Molecular Pathology (AMP). This framework has to be applied during variant assessment. Both variants observed has been reported, with rs2229069 having a benign clinical significance and no reports on the clinical significance of the rs2227985 variant and both have been identified as "synonymous mutations". In the past, it was always thought that all synonymous mutations were silent and their effects on molecular function of the corresponding proteins were assumed to be minimal, however recent data says otherwise (Zeng & Bromberg, 2019).

4.5. Variation functionality predictions.

The National Institute of Health's' National Human Genome Research Institute (NHGRI) defines a mutation as a change in the DNA sequence of an organism. Single nucleotide variants (SNV'S) are the most common variants in the human genome with regulatory, non-synonymous and synonymous variants being of particular interest. Many computational tools have been developed to investigate the functional effects of regulatory and non-synonymous variants, but synonymous variants have been disregarded as functionally irrelevant. Shen *et al*., 2013, provides a comprehensive characterization of human genome variation by high coverage whole genome sequencing, where they found that about 10,000 variants in the coding region of every human have no effect on the resulting protein sequence, termed synonymous variants (SNVs) (Shen et al., 2013). Provided that there is no consequential amino acid change, the effects on functionality have been described as minimal (Zeng & Bromberg, 2019)., therefore we have found that experimental data and research tools into predicting the effects of these effects are limited, and interpretation is severely lacking.

Synonymous variants, in exons, do not change the amino acid sequence and consequent protein structure, and only affects the DNA and mRNA. However, recent studies suggest they can confer disruption in transcription, splicing, co-translational folding, mRNA stability and can affect regulatory factors within protein coding regions by altering protein expression, conformation and function (Zeng & Bromberg, 2019;Sharma et al., n.d.).

For example, Kimchi-Sarfaty et al (2007), demonstrated that synonymous mutations in the *ABCB1 or MDR1* gene that encodes the transporter ATP binding cassette subfamily B member 1 affect protein conformation and have functional and clinical consequences and may be implicated in drug resistance to chemotherapeutic agents (Kimchi-Sarfaty et al., 2007). Likewise, Ernest et al. (2008) demonstrated that synonymous SNV's rs2229069 and rs2227985 (our variant of interest) of the gene *ABL1* in the fusion protein BCR::ABL1 may contribute to primary, although not secondary, resistance to tyrosine kinase inhibitors (Ernst et al., 2008).

Efforts have been made to curate data by performing *in silico* tests to investigate functionality of these synonymous mutations. We therefore continued to explore the available resources in pursuit of the effect of the variants found in our cohort. Each software interface consists of input data fields, and usually integrates various sources of data to provide as much information as possible. The Synonymous Mutations In Cancer database (SynMICdb), is one of the first resources we explored. This database provides five different search options, the first being by Gene whereby the user can search synonymous mutations present in the gene of interest and they accept the HGNG gene symbol, gene name or the Ensembl ID. Other options include search by position in the CDS, search by region, search by organ and the present and advanced search option which allows a combination of different search options. When searching the *ABL1* oncogene, a total of 56 entries were extracted. Unfortunately, our variants could not be found amongst them.

We further investigated Synonymous Variant effect Predictor (SynVep) software created by US BrombergLab company, which essentially is a predictor for evaluating the effects of human synonymous variants. It uses a series of three models using observed (gnomAD) and generated (possible, but not seen in gnomAD) variants, which is a positive-unlabeled learning (PUL) model to identify 'unobservable' variants from the generated dataset; an intermediate model, which differentiates the observed and unobserved; and a final model to classify assumed effect and no-effect variants. They used 'functional effect' rather than 'pathogenicity'. The results showed that variant rs2229069 has no predicted effects and was classified as 'observed' with a predicted score of 0.0055. For the rs2227985 variant, one effect was observed. This variant has a SynVep effect prediction score of 0.8480. Interestingly, the latter SNV was the same SNV that had no reports on the clinical significance due to lack of data according to ClinVar on NCBI's dbSNP.

MutationTaster was another computation tool that was explored, which helps predict the disease-causing potential of DNA variants. ‘Known variant’ indicates if this variant has been found in large scale sequencing projects or listed as disease mutations in ClinVar. The database contains all variants from the 1000 Genomes Project (abbreviated as 1000G in tale x), ExAC and gnomAD.

Results indicated no abrogation of potential splice sites, no frameshift or length of protein changes, no Konzak consensus sequence alterations and no amino acid changes for either of the variants.

Furthermore, we investigated the allele frequencies using the 1000 Genomes Project. For the rs2229069 variant, Europe, South Asia, and East Asia showed 100% ancestral G allele frequency, whereas American data indicated a 1% alternate A allele frequency and Africa having an A allele frequency of 5%. For the rs2227985 variant, East Asia had a 100% ancestral A allele frequency, South Asia, Europe, and America had an alternate G alle frequency of 3%, 7% and 8%, respectively. Africa had an G alle frequency of 9%. When looking at the genotype frequency results showed that Africa had a 0.091% A|G frequency for the rs2229069 variant and 0.324% A|G frequency for the rs2227985 variant. Unfortunately, we were unable to generate our own baseline minor allele frequency for these variants in our cohort due to a lack of sample size. However, this may be further explored in a larger study on a more national scale especially for rs2227985 as one functional effect was observed with an effect prediction score of 0.8480 via the SynVep software.

4.6. CML diagnosis, monitoring and treatment

Early detection of cancer leads to the most favourable outcomes. The American Cancer Society therefore recommends screening tests for certain cancers in people who have no symptoms. Currently, no screening tests are routinely recommended to find CML early. Other than presenting with symptoms which would further be investigated, CML can sometimes be found when routine blood tests are done for other reasons such as routine physicals. At TBH the diagnostic work-up consists of a bone marrow aspirate to determine the proportion of blast cells and basophils to distinguish the phase of CML. Cytogenetics on marrow cells, and qRT-PCR on peripheral blood are performed where the translocation, and type of *BCR::ABL1* mRNA transcript is identified, with qRT-PCR aiding in the monitoring of CML patients' response to treatment. Combination therapy with TKIs and intensive chemotherapy, consolidated with allogeneic haematopoietic stem cell transplantation (AlloSCT) is described to offer patients the most favourable outcomes (Jain et al., 2017), with Treatment-Free Remission being the goal. However, in South Africa, and Africa access to second or third generation TKIs and AlloSCT is limited.

At TBH, Imatinib mesylate is the first-line therapy and Nilotinib serves as a second-line therapy when there is progression to more advanced disease and clinicians characterise the patient's treatment response profiles using the ELN guidelines with some alterations. Generally, if a patient does not reach 10% or less IS at 3 months, they will continue treatment as is, whereas the ELN guidelines warrants this as a warning of possible resistance. However, if patients' treatment do not reach 0.1% after 12 months, then it is concerning. Ultimately, each patient presenting with sub-optimal or inadequate responses to treatment is assessed individually. Clinical data indicates that non-adherence is a major issue amongst the cohort. This, along with the full blood count results, presentation of symptoms and kinase domain mutation screening for secondary mutations are amongst the variables investigated when trying to identify contributions or causes for non-responsiveness to therapy. However, due to the socio-economic background of patients as well as many patients presenting with additional, CML unrelated comorbidities, the process of investigation can be lengthy (communications with Dr. Fatima Bassa, Head of Clinical Haematology).

Published guidelines are based on European populations, but in a setting as the South African public health-sector, options are limited. Consequently, a combination of guidelines is used to determine patient progression, as well as monitoring trends and clinical features to aid in classification and decision-making regarding treatment.

4.7. Drugs and associated costs

TBH is situated in a poverty-stricken region, where patients are not by the means to procure medication privately. The strain of costs of treatment is therefore placed almost entirely on the public health sector. Literature dictates that SA had been part of the GIPAP program run by the Max Foundation and Novartis. However, they had later retracted sponsorship. At Tygerberg specifically, a program was initiated where patients were put on the branded Imatinib drug, Gleevac. In consultation with one of TBH consultants, it has been indicated that patients will remain on the branded treatment until the end of the program in 2022 and patients who were not part of the initial program are placed on the generic drug which was initiated in 2015. Adverse side effects, after changing to other brands, have been reported. However, the brand changes often to whatever is more readily available and cost effective (Communications, Dr Fatima Bassa). The South African government would have to decide on how to procure future treatment post 2022.

4.8. Genomic Medicine, Precision Medicine, Personalised Medicine

CML is considered the poster child for personalised medicine. The establishment of TKIs that target the molecular abnormality that causes the disease has resulted in CML patients achieving life- expectancies like healthy populations. Furthermore, genetic changes contributing towards resistance to first-line therapy can now be targeted. Several mutations cause resistance, and identifying the mutations relevant to individual patients and consequently selecting the right treatment is a rapidly developing part of personalised medicine for CML.

Genomic Medicine, Precision medicine and Personalised Medicine are three terms often interchangeably used to define the use of a patients' individual genotypic information in order to adapt their clinical care to their specific needs. (Roden et al., 2014) suggest that the term 'genomic medicine' comes with more implications for discovery and implementation; precision medicine carries an expectation of perfect outcomes; and that "personalised medicine" more accurately alludes to the ultimate goal (Manuscript et al., 2014). Precision medicine has drastically improved clinical care across the globe. Moreover, *in silico* detection and analysis of mutations are fast-tracked by the availability of high throughput informatics. The American Cancer Society and Precision Medicine Initiative (established by Barak Obama in 2015) defines precision medicine as, "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person" (Collins et al., 2016).

In Europe, an initiative called the International Consortium for Personalised Medicine, ICPerMed, was launched in 2016 by over 30 European and international members together with the European Commission as an observer (http://idwf.de/-CiGzBA). They define personalised medicine as a reference to "a medical model using characterization of individuals phenotypes and genotypes (e.g., molecular profiling, medical imaging, lifestyle data) for tailoring the right therapeutic strategy for the right person at the right time, and/or to determine the predisposition to disease and/or to deliver timely and targeted prevention". Precision medicine therefore is the approach of allowing the investigation into the specific molecular and genetic make-up of the patient's cancer, which ultimately results in more specific treatments, with fewer side-effects, which are also more cost effective to the patient's overall care.

Due to advances in cancer experimentation, the possibility to explore genomes are at the forefront of precision medicine (Zeng & Bromberg, 2019). However, the variability in these prediction methods and categorization can pose challenges to some populations. Africa harbours the highest genetic diversity which are underrepresented in databases, reference panels and bioinformatic platforms and this may be due to lack of infrastructure, funds, and resources in public health sector. A study from the University of Cape Town suggests that *in silico* prediction tools could lead to high rates of false positive or false negative results, especially in African genomes (Bope et al., 2019). They confer that *in silico* prediction in the context of African populations poses additional challenges; the biggest challenge being that these prediction tools are based off non-African benchmarks.

4.9. Limitations

The limitations of this study included the use of a relatively small cohort. However, CML is rare and consequently sample availability was limited. Further limitations include the screening methodology and lack of sensitivity regarding Sanger sequencing. Another limitation included the paucity of clinical data of the screened samples. The data that was available was notably difficult to access, and for the most part illegible. The CML risk scores, SOKAL and EUTOS, were also unavailable for this study. Moreso, our ability to compare our findings to those of the rest of SA was severely hampered by the lack of entity specific and current SA cancer registries. Finally, there was an impeding impact of COVID on data retrieval and sample processing.

4.10. Future Prospects and Conclusions

It would be of interest to investigate which variants other sites in SA had been identified through screening. In addition, functional studies, such as animal or cellular models, may assist in a better understanding of how some of the variants and their interactions contribute to the development of CML drug resistance, such as the novel variant identified, in this study, in Exon 9 of the ABL kinase gene. Furthermore, drug transporters involved in the uptake of the drug or related to the pathways where the BCL-ABL fusion protein participates in. An in-depth investigation into survival rates and causes of death could shed light on possible contributing factors for progression to accelerated stages of disease. We also suggest that investigation into additional chromosomal abnormalities be performed, as this was outside the scope of this study, but could contribute to a better understanding of disease progression.
The ELN guidelines, along with other international guidelines, provides optimal care for CML patients, where it is generally no longer considered a life-threatening disease. However, the reality for developing countries such as South Africa is significantly different. Drug availability, monitoring limitations, patient's non-adherence, non-commitment and the younger age at presentation are some of the realities faced. Lack of intervention at required times lead to more progressive disease. Moreover, the financial strain that a younger presenting population poses, where consequent longer durations of treatments is needed is paramount. South Africa has therefore adopted a triage mind-set when it comes to CML patient care.

Our study emphasises the need for specific entities in cancer registries in SA. Our data shows that our cohort deviates significantly from global CML cohorts which warrants population specific investigation.

References

Abou Dalle, I., Kantarjian, H., Burger, J., Estrov, Z., Ohanian, M., Verstovsek, S., Ravandi, F., Borthakur, G., Garcia-Manero, G., Jabbour, E., & Cortes, J. (2019). Efficacy and safety of generic imatinib after switching from original imatinib in patients treated for chronic myeloid leukemia in the United States. Cancer Medicine, 8(15), 6559–6565. https://doi.org/10.1002/cam4.2545

Abruzzese, E., Aldo, B., Baccarani, M., Abruzzese, E., Accurso, V., Albano, F., Annunziata, M., Barulli, S., Beltrami, G., Bergamaschi, M., Binotto, G., Bocchia, M., Caocci, G., Capodanno, I., Cavazzini, F., Cedrone, M., Cerrano, M., Crugnola, M., Adda, M. D., … Fozza, C. (2019). Managing chronic myeloid leukemia for treatment-free remission : A proposal from the GIMEMA CML WP Managing chronic myeloid leukemia for treatment-free remission : a proposal from the GIMEMA CML WP. December. https://doi.org/10.1182/bloodadvances.2019000865

Ayatollahi H, Keramati M, Shirdel A, Kooshyar M, Raeiszadeh M, Shakeri S, et al . BCR-ABL fusion genes and laboratory findings in patients with chronic myeloid leukemia in northeast Iran. Caspian J Intern Med 2018; 9 (1) :65-70.

Bauer, S., & Romvari, E. (2012). Interpreting Molecular Monitoring Results and International Standardization in Chronic Myeloid Leukemia. Journal of the Advanced Practitioner in Oncology, 3(3), 151–160.http://www.ncbi.nlm.nih.gov/pubmed/25031941%0Ahttp://www.pubmedcentral.nih.gov/articlere nder.fcgi?artid=PMC4093320

Bishop, A. Ö. T., De Beer, T. A. P., & Joubert, F. (2008). Protein homology modelling and its use in South Africa. South African Journal of Science, 104(1–2), 2–6.

Braun, T. P., Eide, C. A., Druker, B. J., Insitute, K. C., & Health, O. (2020). HHS Public Access. 37(4), 530–542. https://doi.org/10.1016/j.ccell.2020.03.006.Response

Chandrasekhar, C., Kumar, P. S., & Sarma, P. V. G. K. (2019). Novel mutations in the kinase domain of BCR-ABL gene causing imatinib resistance in chronic myeloid leukemia patients. Scientific Reports, 9(1), 1–17. https://doi.org/10.1038/s41598-019-38672-x

Collins, H., Calvo, S., Greenberg, K., Forman Neall, L., & Morrison, S. (2016). Information Needs in the Precision Medicine Era: How Genetics Home Reference Can Help. Interactive journal of medical research, 5(2), e13. https://doi.org/10.2196/ijmr.5199

Shen H, Li J, Zhang J, Xu C, Jiang Y, et al. (2013) Comprehensive Characterization of Human Genome Variation by High Coverage Whole-Genome Sequencing of Forty Four Caucasians. PLOS ONE 8(4): e59494. https://doi.org/10.1371/journal.pone.0059494

Cortes, J., Goldman, J. M., & Hughes, T. (2012). Current Issues in Chronic Myeloid Leukemia : Monitoring , Resistance , and Functional Cure Current Issues in Chronic Myeloid Leukemia : Monitoring , Resistance , and Functional Cure. Journal of the National Comprehensive Cancer Network, 10(October).

Deininger, M. W., Shah, N. P., Altman, J. K., Berman, E., Bhatia, R., Bhatnagar, B., DeAngelo, D. J., Gotlib, J., Hobbs, G., Maness, L., Mead, M., Metheny, L., Mohan, S., Moore, J. O., Naqvi, K., Oehler, V., Pallera, A. M., Patnaik, M., Pratz, K., … Sundar, H. (2020). Chronic myeloid leukemia, version 2.2021. JNCCN Journal of the National Comprehensive Cancer Network, 18(10), 1385–1415. https://doi.org/10.6004/jnccn.2020.0047

Eide, C. A., Bottomly, D., Savage, S. L., White, L., Wilmot, B., Schultz, A. M. R., Uchida, K. A., Agarwal, A., Traer, E., & Beppu, L. (2016). Characterization of the genomic landscape of BCR-ABL1 kinase-independent mechanisms of resistance to ABL1 tyrosine kinase inhibitors in chronic myeloid leukemia. Am Soc Hematology.

Ernst, T., Hoffmann, J., Erben, P., Hanfstein, B., Leitner, A., Hehlmann, R., Hochhaus, A., & Müller,

M. C. (2008). ABL single nucleotide polymorphisms may masquerade as BCR-ABL mutations associated with resistance to tyrosine kinase inhibitors in patients with chronic myeloid leukemia. Haematologica, 93(9), 1389–1393. https://doi.org/10.3324/haematol.12964

Galli, C. L., Sensi, C., Fumagalli, A., Parravicini, C., Marinovich, M., & Eberini, I. (2014). A computational approach to evaluate the androgenic affinity of iprodione, procymidone, vinclozolin and their metabolites. PLoS ONE, 9(8). https://doi.org/10.1371/journal.pone.0104822

Hochhaus, A., Baccarani, M., Silver, R. T., Schiffer, C., Apperley, J. F., Cervantes, F., Clark, R. E., Cortes, J. E., Deininger, M. W., Guilhot, F., Hjorth-Hansen, H., Hughes, T. P., Janssen, J. J. W. M., Kantarjian, H. M., Kim, D. W., Larson, R. A., Lipton, J. H., Mahon, F. X., Mayer, J., … Hehlmann, R. (2020). European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. Leukemia, 34(4), 966–984. https://doi.org/10.1038/s41375-020-0776-2

Hughes, T. P., Mauro, M. J., Cortes, J. E., Minami, H., Rea, D., DeAngelo, D. J., Breccia, M., Goh, Y.-T., Talpaz, M., Hochhaus, A., le Coutre, P., Ottmann, O., Heinrich, M. C., Steegmann, J. L., Deininger, M. W. N., Janssen, J. J. W. M., Mahon, F.-X., Minami, Y., Yeung, D., … Kim, D.-W. (2019). Asciminib in Chronic Myeloid Leukemia after ABL Kinase Inhibitor Failure. New England Journal of Medicine, 381(24), 2315–2326. https://doi.org/10.1056/nejmoa1902328

Iqbal, Z., Absar, M., Mahmood, A., Aleem, A., Iqbal, M., Jameel, A., Akhtar, T., Karim, S., Rasool,

M., Mirza, Z., Khalid, M., Akram, A., Sabar, M. F., Khalid, A. M., Aljarrah, K., Iqbal, J., Khalid, M., Shah, I. H., & Alanazi, N. (2020). Discovery and Protein Modeling Studies of Novel Compound Mutations Causing Resistance to Multiple Tyrosine Kinase Inhibitors in Chronic Myeloid Leukemia. Asian Pacific Journal of Cancer Prevention, 21(12), 3517–3526. https://doi.org/10.31557/APJCP.2020.21.12.3517

Jabbour, E., & Kantarjian, H. (2020). Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring. American Journal of Hematology, 95(6), 691–709. https://doi.org/10.1002/ajh.25792

Jones, D., Kamel-Reid, S., Bahler, D., Dong, H., Elenitoba-Johnson, K., Press, R., Quigley, N., Rothberg, P., Sabath, D., Viswanatha, D., Week, K., & Zehnder, J. (2009). Laboratory practice guidelines for detecting and reporting BCR-ABL drug resistance mutations in chronic myelogenous leukemia and acute lymphoblastic leukemia. Journal of Molecular Diagnostics, 11(1), 4–11. https://doi.org/10.2353/jmoldx.2009.080095

Kang, Z. J., Liu, Y. F., Xu, L. Z., Long, Z. J., Huang, D., Yang, Y., Liu, B., Feng, J. X., Pan, Y. J., Yan,

J. S., & Liu, Q. (2016). The philadelphia chromosome in leukemogenesis. Chinese Journal of Cancer, 35(1), 1–15. https://doi.org/10.1186/s40880-016-0108-0

Kotagama, K., Chang, Y., Mangone, M., State, A., & State, A. (2016). miRNAs as Biomarkers in Chronic Myelogenous Leukemia. Drug Development Research, 76(6), 278–285. https://doi.org/10.1002/ddr.21266.miRNAs

Lauseker, M., Bachl, K., Turkina, A., Faber, E., Prejzner, W., Olsson-Strömberg, U., Baccarani, M.Lomaia, E., Zackova, D., Ossenkoppele, G., Griskevicius, L., Schubert-Fritschle, G., Sacha, T., Heibl, S., Koskenvesa, P., Bogdanovic, A., Clark, R. E., Guilhot, J., Hoffmann, V. S., … Pfirrmann, M. (2019). Prognosis of patients with chronic myeloid leukemia presenting in advanced phase is defined mainly by blast count, but also by age, chromosomal aberrations and hemoglobin. American Journal of Hematology, 94(11), 1236–1243. https://doi.org/10.1002/ajh.25628

Lee, N., Kim, K. H., & Lee, S. Oral chemotherapeutic agents in current use. Journal of the Korean Medical Association. 2011; 54(11): 1191-1198. DOI:10.5124/jkma.2011.54.11.1191

Louw, V. (2012). Chronic myeloid leukaemia in South Africa. In Hematology (Amsterdam, Netherlands): Vol. 17 Suppl 1. https://doi.org/10.1179/102453312X13336169155817

Louw, V. J. (2012). Chronic myeloid leukaemia in South Africa. Hematology, 17(SUPPL. 1). https://doi.org/10.1179/102453312X13336169155817

Malhotra, H., Radich, J., & Garcia-Gonzalez, P. (2019). Meeting the needs of CML patients in resource-poor countries. Hematology (United States), 2019(1), 433–442. https://doi.org/10.1182/hematology.2019000050

Mansoori, B., Mohammadi, A., Davudian, S., Shirjang, S., & Baradaran, B. (2017). The different mechanisms of cancer drug resistance: A brief review. Advanced Pharmaceutical Bulletin, 7(3), 339–348. https://doi.org/10.15171/apb.2017.041

Roden, D.M. and Tyndale, R.F. (2013), Genomic Medicine, Precision Medicine, Personalized

Medicine: What's in a Name?. Clinical Pharmacology & Therapeutics, 94: 169-172. https://doi.org/10.1038/clpt.2013.101

Nachi, M., Kihel, I., & Entasoltane, B. (n.d.). Impact of the major BCR-ABL1 transcript type on clinical and biological parameters and molecular response in patients with chronic myeloid leukemia. Hematology/Oncology and Stem Cell Therapy, xxxx, 8–12. https://doi.org/10.1016/j.hemonc.2020.08.003

Nagendrappa, K., Jitendra, L., Pehalajani, K., Loknatha, D., Khandare, S. C. S., & Ashok, P. (2020). CML in Elderly : Does Age Matter ? Indian Journal of Hematology and Blood Transfusion, 36(1), 47–50. https://doi.org/10.1007/s12288-019-01143-4

Quintás-Cardama, A., Kantarjian, H. M., & Cortes, J. E. (2009). Mechanisms of primary and secondary resistance to imatinib in chronic myeloid leukemia. Cancer Control, 16(2), 122–131. https://doi.org/10.1177/107327480901600204

Sandwith, H. (1859). The journal. British Medical Journal, s4-1(148), 879–880. https://doi.org/10.1136/bmj.s4-1.148.879-a

Sharma, Y., Miladi, M., Dukare, S., Boulay, K., Caudron-herger, M., Groß, M., Backofen, R., & Diederichs, S. (n.d.). A pan-cancer analysis of synonymous mutations. Nature Communications, 2019. https://doi.org/10.1038/s41467-019-10489-2

Sissolak, G., Badenhorst, J., Steenkamp, J., Heaney, M., Louw, V., Schnugh, D., & Willem, P. (2015). Treatment Outcomes in CML Patients Treated with Tyrosine Kinase Inhibitors at a Tertiary Teaching Hospital in South Africa. Clinical Lymphoma, Myeloma and Leukemia, 15(12), 803–810. https://doi.org/10.1016/j.clml.2015.09.006

Steinhaus, R., Proft, S., Schuelke, M., Schwarz, J. M., Seelow, D., & Cooper, D. N. (2021). MutationTaster2021. 49(April), 446–451.

Tanaka, R., Kimura, S., Ashihara, E., Yoshimura, M., Takahashi, N., Wakita, H., Itoh, K., Nishiwaki, K., Suzuki, K., Nagao, R., Yao, H., Hayashi, Y., Satake, S., Hirai, H., Sawada, K. ichi, Ottmann, O. G., Melo, J. V., & Maekawa, T. (2011). Rapid automated detection of ABL kinase domain mutations in imatinib-resistant patients. Cancer Letters, 312(2), 228–234. https://doi.org/10.1016/j.canlet.2011.08.009

Tomasello, L., Vezzalini, M., Boni, C., Bonifacio, M., Scaffidi, L., Yassin, M., Al-Dewik, N., Kamga, P. T., Krampera, M., & Sorio, C. (2020). Regulative loop between β-catenin and protein tyrosine receptor type γ in chronic myeloid leukemia. International Journal of Molecular Sciences, 21(7), 1–19. https://doi.org/10.3390/ijms21072298

Tsukahara, F., & Maru, Y. (2010). Bag1 directly routes immature BCR-ABL for proteasomal degradation. Blood, 116(18), 3582–3592. https://doi.org/10.1182/blood-2009-10-249623

Van Dongen, J. J. M., Macintyre, E. A., Gabert, J. A., Delabesse, E., Rossi, V., Saglio, G., Gottardi, E., Rambaldi, A., Dotti, G., Griesinger, F., Parreira, A., Gameiro, P., Diáz, M. G., Malec, M., Langerak, A. W., San Miguel, J. F., & Biondi, A. (1999). Standardized RT-PCR analysis of fusion gene transcripts from chromosome aberrations in acute leukemia for detection of minimal residual disease. Report of the BIOMED-1 Concerted Action: Investigation of minimal residual disease in acute leukemia. Leukemia, 13(12), 1901–1928. https://doi.org/10.1038/sj.leu.2401592

Yang, K., & Fu, L. wu. (2015). Mechanisms of resistance to BCR-ABL TKIs and the therapeutic strategies: A review. Critical Reviews in Oncology/Hematology, 93(3), 277–292. https://doi.org/10.1016/j.critrevonc.2014.11.001

Zeng, Z., & Bromberg, Y. (2019). Predicting Functional Effects of Synonymous Variants: A SystematicReview and Perspectives. 10(October), 1–15. https://doi.org/10.3389/fgene.2019.00914

Fiona Cunningham , James E. Allen, Jamie Allen, Jorge Alvarez-Jarreta, M. Ridwan Amode, Irina M. Armean , Olanrewaju Austine-Orimoloye, Andrey G. Azov, If Barnes, Ruth Bennett, Andrew Berry, Jyothish Bhai, Alexandra Bignell, Konstantinos Billis , Sanjay Boddu, Lucy Brooks, Mehrnaz Charkhchi, Carla Cummins , Luca Da Rin Fioretto, Claire Davidson, Kamalkumar Dodiya, Sarah Donaldson, Bilal El Houdaigui, Tamara El Naboulsi, Reham Fatima, Carlos Garcia Giron , Thiago Genez, Jose Gonzalez Martinez, Cristina Guijarro-Clarke, Arthur Gymer, Matthew Hardy, Zoe Hollis, Thibaut Hourlier , Toby Hunt, Thomas Juettemann , Vinay Kaikala, Mike Kay, Ilias Lavidas, Tuan Le, Diana Lemos, José Carlos Marugán, Shamika Mohanan, Aleena Mushtaq, Marc Naven, Denye N. Ogeh, Anne Parker, Andrew Parton, Malcolm Perry, Ivana Pili˘zota, Irina Prosovetskaia, Manoj Pandian Sakthivel, Ahamed Imran Abdul Salam, Bianca M. Schmitt, Helen Schuilenburg, Dan Sheppard, José G. Pérez-Silva, William Stark, Emily Steed, Kyösti Sutinen, Ranjit Sukumaran, Dulika Sumathipala, Marie-Marthe Suner, Michal Szpak, Anja Thormann, Francesca Floriana Tricomi, David Urbina-G´omez, Andres Veidenberg, Thomas A. Walsh, Brandon Walts, Natalie Willhoft, Andrea Winterbottom, Elizabeth Wass, Marc Chakiachvili, Bethany Flint, Adam Frankish , Stefano Giorgetti, Leanne Haggerty, Sarah E. Hunt , Garth R. IIsley, Jane E. Loveland , Fergal J. Martin , Benjamin Moore, Jonathan M. Mudge, Matthieu Muffato, Emily Perry , Magali Ruffier , John Tate, David Thybert, Stephen J. Trevanion, Sarah Dyer, Peter W. Harrison , Kevin L. Howe , Andrew D. Yates , Daniel R.Zerbino and Paul Flicek Ensembl (2022)Nucleic Acids Res., vol. 50(1):D988-D995 PubMed PMID:34791404. doi:10.1093/nar/gkab1049

Appendices

Appendix I: DNA Concentrations of isolated samples

Appendix II: Gene Xpert

Appendix III: Genomic DNA was extracted using the QIAamp DNA Mini Kit

Appendix IV: TBE Gel preparation

Appendix V: Sequencing reaction preparation

Appendix VI: SU FMHS HREC Ethics Approval

Appendix VII: NHLS Ethics approval

Appendix VII: Department of Health TBH Ethics approval

Appendix I: Biodrop readings for Concentrations of isolated samples

	Sample	DNA Concentration (ng/ul)	260/280
1	40-1	116.7	1.89
2	401-4	13.4	1.69
3	374-1	382.4	1.40
4	374-5	14.7	1.85
5	361-6	44.6	1.92
6	371-1	345.9	1.22
7	371-3	20	1.91
8	381-1	20	2.11
9	381-4	83.7	1.88
10	356-3	20.8	1.92
11	311-1	104	1.85
12	311-10	17.8	2.43
13	329-2	59.5	1.46
14	300-2	48.2	1.39
15	322-2	24.6	1.80
16	294-2	24.3	2.04
17	360-2	18.6	2.19
18	280-6	67.3	1.59
19	250-8	282.8	1.56
20	378-4	140.4	1.79
21	245-6	21.9	1.92
22	248-9	112.6	1.89
23	264-7	129.4	1.86
24	252-14	82.7	1.80
25	143-20	63.5	1.97

Appendix II GPL1530 Verification/Validation of Quantitative and Qualitative methods SOP extract

SECTION D: EXPERIMENTS

Characteristic to be evaluated	Quantitative Method	Qualitative Method	Method	SOP
Precision - Repeatability	√		EP15 Precision	GPL2140
Precision – Within laboratory	√		EP 15 Precision	GPL2140
Trueness – Bias	√		Recovery of known values OR Alternate Comparison	GPL2140 Or GPL1530
Accuracy – Total Error	√		Calculation Total error vs Total allowable error TE= Systematic error +random error TE= |bias| + 1.65 *SD	
Agreement		√	EP12	GPQ0006
Symmetry		√	EP12	GPQ0006
Specificity and Sensitivity*		√	EP12	GPQ0006

Sample types and size

Alternate method comparison:
A minimum of 20 samples up to 40 samples are to be run in parallel on both the current and new analyser.
These samples should be distributed more or less evenly across the methods reportable range and include data points that cover the full range of medical decision points (see CHE1896 or consult pathologist). It is essential to consider sample stability carefully.
Approximately 10 samples should be run per day after the precision samples.
CSF and urine samples matrix are usually difficult to obtain, therefore a minimum of 20 samples will be statistically acceptable.

Appendix III: Gene Xpert Protocol

PRF

Document: Factor V

Version 1

Date: 13 August 2018

Author: S. Isaacs

Factor V assay

1. Check that the patient WB-EDTA tube details and assigned number **(e.g. FVL 18-39)** match the details on the NHLS request form.
2. Switch on the GeneXpert machine and computer.
3. Log onto the Cepheid profile using the password: cphd
4. Log on again using user: johanna and password: Dinsday52.
5. GeneXpert will run a self- test say **NO** to all the screens that pop up.
6. Click on create test.
7. Enter the assigned number then enter.
8. Scan the cartridge.
9. Decontaminate your working area with **RNase AWAY™** as well as pipette.
10. Remove cartridge from the package, which is kept at room temperature.
11. Label the cartridge with the assigned number.
12. Mix WB-EDTA tube by inverting the tube multiple times, until homogenous.
13. Open the cartridge lid, using a pipette with an aerosol resistant tip transfer **50ul** of EDTA anticoagulated blood to the bottom wall of the "S" opening of the cartridge.

14. Close the cartridge lid and load onto module assigned by GeneXpert machine.

15. Close module and hold close.

16. Click **Start Test**, once the test panel indicates green release module.

Appendix IV: QIAamp DNA Mini Kit

Protocol: DNA Purification from Blood or Body Fluids (Spin Protocol)

This protocol is for purification of total (genomic, mitochondrial, and viral) DNA from whole blood, plasma, serum, buffy coat, lymphocytes, and body fluids using a microcentrifuge. For total DNA purification using a vacuum manifold, see "Protocol: DNA Purification from Blood or Body Fluids (Vacuum Protocol)" on page 29.

Important points before starting

- All centrifugation steps are carried out at room temperature (15–25°C).
- Use carrier DNA if the sample contains <10,000 genome equivalents (see page 17).
- 200 µl of whole blood yields 3–12 µg of DNA. Preparation of buffy coat (see page 18) is recommended if a higher yield is required.

Things to do before starting

- Equilibrate samples to room temperature (15–25°C).
- Heat a water bath or heating block to 56°C for use in step 4.
- Equilibrate Buffer AE or distilled water to room temperature for elution in step 11.
- Ensure that Buffer AW1, Buffer AW2, and QIAGEN Protease have been prepared according to the instructions on page 16.
- If a precipitate has formed in Buffer AL, dissolve by incubating at 56°C.

Procedure

1. **Pipet 20 µl QIAGEN Protease (or proteinase K) into the bottom of a 1.5 ml microcentrifuge tube.**
2. **Add 200 µl sample to the microcentrifuge tube. Use up to 200 µl whole blood, plasma, serum, buffy coat, or body fluids, or up to 5 x 10^6 lymphocytes in 200 µl PBS.**

   If the sample volume is less than 200 µl, add the appropriate volume of PBS.

   QIAamp Mini spin columns copurify RNA and DNA when both are present in the sample. RNA may inhibit some downstream enzymatic reactions, but not PCR. If RNA-free genomic DNA is required, 4 µl of an RNase A stock solution (100 mg/ml) should be added to the sample before addition of Buffer AL.

   Note: It is possible to add QIAGEN Protease (or proteinase K) to samples that have already been dispensed into microcentrifuge tubes. In this case, it is important to ensure proper mixing after adding the enzyme.

3. **Add 200 µl Buffer AL to the sample. Mix by pulse-vortexing for 15 s.**

   To ensure efficient lysis, it is essential that the sample and Buffer AL are mixed thoroughly to yield a homogeneous solution.

   If the sample volume is larger than 200 µl, increase the amount of QIAGEN Protease (or proteinase K) and Buffer AL proportionally; for example, a 400 µl sample will require 40 µl QIAGEN Protease (or proteinase K) and 400 µl Buffer AL. If sample volumes larger than 400 µl are required, use of QIAamp DNA Blood Midi or Maxi Kits is recommended; these can process up to 2 ml or up to 10 ml of sample, respectively.

   Note: Do not add QIAGEN Protease or proteinase K directly to Buffer AL.

4. **Incubate at 56°C for 10 min.**

   DNA yield reaches a maximum after lysis for 10 min at 56°C. Longer incubation times have no effect on yield or quality of the purified DNA.

5. **Briefly centrifuge the 1.5 ml microcentrifuge tube to remove drops from the inside of the lid.**

6. **Add 200 µl ethanol (96–100%) to the sample, and mix again by pulse-vortexing for 15 s. After mixing, briefly centrifuge the 1.5 ml microcentrifuge tube to remove drops from the inside of the lid.**

   If the sample volume is greater than 200 µl, increase the amount of ethanol proportionally; for example, a 400 µl sample will require 400 µl of ethanol.

7. **Carefully apply the mixture from step 6 to the QIAamp Mini spin column (in a 2 ml collection tube) without wetting the rim. Close the cap, and centrifuge at 6000 x *g* (8000 rpm) for 1 min. Place the QIAamp Mini spin column in a clean 2 ml collection tube (provided), and discard the tube containing the filtrate.***

   Close each spin column to avoid aerosol formation during centrifugation.

   Centrifugation is performed at 6000 x *g* (8000 rpm) to reduce noise. Centrifugation at full speed will not affect the yield or purity of the DNA. If the lysate has not completely passed through the column after centrifugation, centrifuge again at higher speed until the QIAamp Mini spin column is empty.

   Note: When preparing DNA from buffy coat or lymphocytes, centrifugation at full speed is recommended to avoid clogging.

8. **Carefully open the QIAamp Mini spin column and add 500 µl Buffer AW1 without wetting the rim. Close the cap and centrifuge at 6000 x *g* (8000 rpm) for 1 min. Place the QIAamp Mini spin column in a clean 2 ml collection tube (provided), and discard the collection tube containing the filtrate.***

   It is not necessary to increase the volume of Buffer AW1 if the original sample volume is larger than 200 µl.

9. **Carefully open the QIAamp Mini spin column and add 500 µl Buffer AW2 without wetting the rim. Close the cap and centrifuge at full speed (20,000 x *g*; 14,000 rpm) for 3 min.**

10. **Recommended: Place the QIAamp Mini spin column in a new 2 ml collection tube (not provided) and discard the old collection tube with the filtrate. Centrifuge at full speed for 1 min.**

    This step helps to eliminate the chance of possible Buffer AW2 carryover.

11. **Place the QIAamp Mini spin column in a clean 1.5 ml microcentrifuge tube (not provided), and discard the collection tube containing the filtrate. Carefully open the QIAamp Mini spin column and add 200 µl Buffer AE or distilled water. Incubate at room temperature (15–25°C) for 1 min, and then centrifuge at 6000 x *g* (8000 rpm) for 1 min.**

    Incubating the QIAamp Mini spin column loaded with Buffer AE or water for 5 min at room temperature before centrifugation generally increases DNA yield.

    A second elution step with a further 200 µl Buffer AE will increase yields by up to 15%.

    Volumes of more than 200 µl should not be eluted into a 1.5 ml microcentrifuge tube because the spin column will come into contact with the eluate, leading to possible aerosol formation during centrifugation.

    Elution with volumes of less than 200 µl increases the final DNA concentration in the eluate significantly, but slightly reduces the overall DNA yield (see Table 5, page 25). For samples containing less than 1 µg of DNA, elution in 50 µl Buffer AE or water is recommended. Eluting with 2 x 100 µl instead of 1 x 200 µl does not increase elution efficiency.

    For long-term storage of DNA, eluting in Buffer AE and storing at –30 to –15°C is recommended, since DNA stored in water is subject to acid hydrolysis.

    A 200 µl sample of whole human blood (approximately 5 x 10^6 leukocytes/ml) typically yields 6 µg of DNA in 200 µl water (30 ng/µl) with an A_{260}/A_{280} ratio of 1.7–1.9.

    For more information about elution and how to determine DNA yield, purity, and length, refer to pages 24–25 and Appendix A, page 50.

Appendix V: TBE (Tris-borate-EDTA) Gel preparation

Lonza Rockland, Inc.
www.lonza.com
biotechserv@lonza.com
Tech Service: 800-521-0390
Customer Service: 800-638-8174
Document # 18100-0807-8
Rockland, ME 04841 USA

SeaKem® LE Agarose

The first and finest name in agarose.

Introduction

SeaKem® LE Agarose is an all purpose agarose for routine nucleic acid electrophoresis of fragments between 500bp-23,000 bp. SeaKem® LE Agarose has no detectable DNase or RNase activity.

Analytical Specifications

Gelling temperature (1.5%)	36°C ±1.5°C
Melting temperature (1.5%)	≥90°C
Gel strength (1%)	≥1,200 g/cm²

Applications

- Analytical electrophoresis of DNA and RNA ≥1,000 bp
- Blotting of DNA and RNA

Suggested Agarose Concentrations

Size Range (Base Pairs)	Final Agarose Concentration (%) 1X TAE Buffer	1X TBE Buffer
1,000-23,000	0.60	0.50
800-10,000	0.80	0.70
400-8,000	1.00	0.85
300-7,000	1.20	1.00
200-4,000	1.50	1.25
100-3,000	2.00	1.75

Dye Mobility Table

Migration of double-stranded DNA in relation to Bromophenol Blue (BPB) and Xylene Cyanol (XC) in SeaKem® LE Agarose Gels.

1X TAE Buffer XC	BPB	% Agarose	1X TBE Buffer XC	BPB
24,800	2,900	0.30	19,400	2,850
11,000	1,650	0.50	12,000	1,350
10,200	1,000	0.75	9,200	720
6,100	500	1.00	4,100	400
3,560	370	1.25	2,500	260
2,800	300	1.50	1,800	200
1,800	200	1.75	1,100	110
1,300	150	2.00	850	70

Precautions

Always wear eye protection when dissolving agarose and guard yourself and others against scalding solutions. Refer to Material Safety Data Sheet for additional safety and handling information.

Microwave Instructions for Agarose Preparation

1. Choose a beaker that is 2-4 times the volume of the solution.
2. Add room temperature 1X or 0.5X electrophoresis buffer and a stir bar to the beaker.
3. Slowly sprinkle in the agarose powder while the solution is rapidly stirred.
4. **Remove the stir bar if not Teflon® coated.**
5. Weigh the beaker and solution before heating.
6. Cover the beaker with plastic wrap.
7. Pierce a small hole in the plastic wrap for ventilation.
8. Heat the beaker in the microwave oven on **High** power until bubbles appear.
9. Remove the beaker from the microwave oven. **Caution: Any microwaved solution may become superheated and foam over when agitated.**
10. **GENTLY** swirl the beaker to resuspend any settled powder and gel pieces.
11. Reheat the beaker on **HIGH** power until the solution comes to a boil.
12. **Hold at boiling point for 1 minute** or until all of the particles are dissolved.
13. Remove the beaker from the microwave oven.
14. **GENTLY** swirl the beaker to thoroughly mix the agarose solution.
15. After dissolution, add sufficient hot distilled water to obtain the initial weight.
16. Mix thoroughly.
17. Cool the solution to 50°C-60°C prior to casting.

Appendix VI: Sequencing reaction preparation

PCR Clean-up and Sequencing Reaction preparation workflow

PCR Clean-up Enzyme – Mix 1

(Can make up a working solution and store for future use)

Mix 1	Brand/Cat nr	Volumes	Storage
Exonuclease 1 & Buffer	New England Biolabs (M0293S)	0.5ul	-20°C
CIP or FASTAP	New England Biolabs (M0290S)	0.5ul	

Applied Biosystems BigDye® Terminator v3.1 Cycle Sequencing Kits – Mix 2

Mix 2	Brand/Cat nr	Volumes	Storage
AB Big Dye® Terminator v3.1	?	60ul	-20°C
Sequencing Buffer	?	140ul	

x2

1. Identify # of PCR products tube and rearrange tubes according to sequencing primers and sample nr conferring to list on PRF sequencing form (Appendix#)

2. Assign CAF labelled sequencing tubes to respective PCR product tube.

3. Complete PRF sequencing form (Appendix #) (Double check that tubes arrangement matches those on form). E.g. CAF sequencing tube nr – PRF sample nr, PRF765 – 150/16. NB: note name and date.

4. Add **3 ul** of PCR product to assigned CAF labelled sequencing tubes

5. Use **Mix 1** which is stored in freezer @ -20°C. Take **1 ul** of Mix and add to CAF sequencing tube containing the PCR product.

6. Take CAF sequencing tube containing **4ul** of Mix 1 and PCR product combination and incubate @ **37°C** for **15min** and **80°C** for **15 min** in PCR machine (AB GeneAmp PCR System 2700)

7. Centrifuge CAF sequencing tubes @ **8000rpm** for **1min**

8. Add **3ul** of respective **primers** (~3pmol–[1ul~1pmol]) to the **4ul** template in CAF sequencing tube.

9. Add **4ul** of **Mix 2** to CAF sequencing tube containing primer and Clean PCR template mix.

10. PCR cycle for sequencing reaction:

98°C	98°C	TA	60°C	25
1min	30sec	15sec	4min	1min
	25 cycles			

Appendix VII: SU FMHS HREC Ethics Approval

26/08/2022

Project ID: 11509

Ethics Reference No: S19/10/272

Project Title: Screening and characterization of BCR/ABL kinase domain mutations in Chronic Myeloid Leukaemia patients at Tygerberg Hospital, South Africa

Dear Miss S Isaacs

We refer to your request for an extension/annual renewal of ethics approval received 01/08/2022.

The Health Research Ethics Committee reviewed and approved the annual progress report through an expedited

review process. The approval of this project is extended for a further year.

Approval date: 26 August 2022
Expiry date: 25 August 2023

Kindly be reminded to submit progress reports two (2) months before expiry date.

Kindly note that the HREC uses an electronic ethics review management system, *Infonetica*, to manage ethics applications and ethics review process. To submit any documentation to HREC, please click on the following link: https://applyethics.sun.ac.za.

Please remember to use your Project Id 11509 and ethics reference number S19/10/272 on any documents or correspondence with the HREC concerning your research protocol.

Please note that for studies involving the use of questionnaires, the final copy should be

uploaded on Infonetica. Yours sincerely,

Melody Shana
Coordinator: Health Research Ethics Committee1

National Health Research Ethics Council (NHREC)
Registration Number: REC-130408-012
(HREC1) •REC-230208-010 (HREC2)

Federal Wide Assurance Number: 00001372
Office of Human Research Protections (OHRP) Institutional Review
Board (IRB) Number: IRB0005240 (HREC1) •IRB0005239 (HREC2)

The Health Research Ethics Committee (HREC) complies with the SA National Health Act No. 61 of 2003 as it pertains to health research. The HREC abides by the ethical norms and principles for research, established by the World Medical Association (2013). Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects; the South African Department of Health (2006).Guidelines for Good Practice in the Conduct of Clinical Trials with Human Participants in South Africa (2nd edition); as well as the Department of Health (2015). Ethics in Health Research: Principles, Processes and Structures (2nd edition).
The Health Research Ethics Committee reviews research involving human subjects conducted or supported by the Department of Health and Human Services, or other federal departments or agencies that apply the Federal Policy for the Protection of Human Subjects to such research (United States Code of Federal Regulations Title 45 Part 46); and/or clinical investigations regulated by the Food and Drug Administration (FDA) of the Department of Health and Human Services.

Appendix VIII: NHLS Ethics approval

09 July 2021

Applicant: Shareefa Isaacs **Institution:** Stellenbosch University **Department:** Haematological Pathology
Email: shareefa@sun.ac.za
Tel: 021 938 4204 **Cell:** 081 270 7016

CC: Carmen Swanepoel

Academic Affairs and Research
Modderfontein Road, Sandringham, 2031
Tel: +27 (0)11 386 6142
Fax: +27 (0)11 386 6296
Email: babatyi.kgokong@nhls.ac.za
Web: www.nhls.ac.za

Re: Approval to access National Health Laboratory Service (NHLS) Data

Your application to undertake a research project **"Screening and characterization of BCR/ABL kinase domain mutations in Chronic Myeloid Leukaemia Patients at Tygerberg Hospital, South Africa - Secondary mutations, drug resistance and associated problems, Ref No: PR2116803"** using data from the NHLS database has been reviewed. This letter serves to advise that the application has been approved and the required data will be made available to you ***without patient names*** to conduct the proposed study as outlined in the submitted application. Submissions should be made annually on the AARMS system – https://aarms.nhls.ac.za.

Please note that approval is granted on your compliance with the NHLS conditions of service and that the study can only be undertaken provided that the following conditions have been met.

- Processes are discussed with the relevant NHLS departments (i.e. Information Management Unit and Operations Office) and are agreed upon.
- Confidentiality is maintained at participant and institutional level and there is no disclosure of personal information or confidential information as described by the NHLS policy.
- NHLS Data cannot be used to track patients as no pre-approval/consent is obtained from Patients.
- All data requested should be in accordance with the research protocol submitted and approved by the relevant Ethics Committee.
- Request for the inclusion of the NHLS as a source of data in the original protocol to be approved by Ethics as NHLS does not have a Human Research Ethics Committee.

- A final report of the research study and any published paper resulting from this study are submitted and addressed to the NHLS Academic Affairs and Research office and the NHLS has been acknowledged appropriately.
- Carmen Swanepoel is noted as NHLS collaborator for this study.

Please note that this letter constitutes approval by the NHLS Academic Affairs and Research Office. Any data related queries may be directed to NHLS Corporate Data Warehouse, contact number: 011 386 6074 email: zarina.sabat@nhls.ac.za

Yours Sincerely,

p.p. (Acting National Manager: Phuti Ngwepe) Dr Babatyi Malope-Kgokong
National Manager: AAR

Chairperson: Prof Eric Buch CEO: Dr Karmani Chetty
Physical Address: 1 Modderfontein Road, Sandringham, Johannesburg, South Africa Postal Address: Private Bag X8, Sandringham, 2131, South Africa
Tel: +27 (0) 11 386 6000/ 0860 00 NHLS(6457) www.nhls.ac.za
Practice number: 5200296

Appendix IX Department of Health TBH Ethics approval

TYGERBERG HOSPITAL
REFERENCE:
Research Projects
ENQUIRIES: **Dr GG Marinus**
TELEPHONE:021938 5752

Western Cape Government
Health

Project ID: 11509

Ethics Reference: S19/10/272

TITLE: **Screening and characterization of BCR/ABL kinase domain mutations in Chronic Myeloid Leukaemia patients at Tygerberg Hospital, South Africa.**

Dear Miss Shareefa Isaacs

PERMISSION TO CONDUCT YOUR RESEARCH AT TYGERBERG HOSPITAL.

- In accordance with the Tygerberg Hospital Health Research Policy and Protocol of **April 2018,** permission is hereby granted for you to conduct the above-mentioned research here at Tygerberg Hospital for a year based on your HREC approval.

- Researchers, in accessing Provincial health facilities, are expressing consent to provide the Department with an electronic copy of the final feedback within six months of completion of research. This can be submitted to the Provincial Research Co-Ordinator (Health.Research@westerncape.gov.za).

DR GG MARINOS
MANAGER: MEDICAL SERVICES

Date:

19/5/2021

Private Bag X3, Tygerberg, 7505 www.capegateway.go.v.za