FORENSIC SHORT TANDEM REPEAT MARKERS ALTERATION IN LEUKEMIA CELLS AND BREAST CANCER TISSUES

IZZAH SYAHIRA BINTI OMAR

UNIVERSITI SAINS MALAYSIA

2025

FORENSIC SHORT TANDEM REPEAT MARKERS ALTERATION IN LEUKEMIA CELLS AND BREAST CANCER TISSUES

by

IZZAH SYAHIRA BINTI OMAR

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

April 2025

ACKNOWLEDGEMENT

First and foremost, I am grateful to Allah SWT for granting me the opportunity and providing endless support throughout my Master's journey. My heartfelt thanks go to my mother, Noraskim Md Isa, and my siblings for their unwavering support and prayers for my success.

I am deeply thankful to my supervisor, Dr. Nur Waliyuddin Hanis Zainal Abidin, for his guidance and encouragement, and to my co-supervisors, Dr. Siti Norasikin Mohd Nafi (Pathology Department) and Prof. Dr. Zafarina Zainuddin (HID DNA), for their constructive feedback. I also extend my gratitude to Dr. Seoparjoo Azmel Mohd Isa and Dr. Shafini Mohamed Yusoff (Hematology Department, HUSM) for their expertise in sample collection.

I am sincerely thankful to my supervisor's fellow student, Nur Hafiza, for her invaluable support, especially during our attachment at the Department of Chemistry Johor Bahru. My appreciation also goes to those I met during the attachment for their guidance in lab analyses and insights into real-life work in the government sector.

This study would not have been possible without the cooperation and kindness of the participants involved, whose support has significantly contributed to my growth. Lastly, special thanks to my close friends—Norsaffarina, Nur Hanis Natasha, Sabrina George, Saiful Ammar, and Ahmad Sadikin—for their endless motivation, support, and ideas throughout this journey. Thank you.

"**Success is not the key to happiness. Happiness is the key to success. If you love what you are doing, you will be successful**." — Albert Schweitzer

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

mL	Millilitre
µL	Microlitre
rpm	Revolution per minute
°C	Degree celcius
g	Gram
x g	Times gram
mg/µL; ng/µL	Milligram per microlitre; nanogram pre microlitre
cm	Centimetre
mm	Millimetre
V	Voltage
xN	Paired leukemia normal tissue
bp	Base pairs
AC - TC	Breast cancerous tissues
AN - TN	Adjacent breast normal tissues
1C – 20C	Leukemia
1NC–20NC	Healthy controls

LIST OF ABBREVIATIONS

STR	Short tandem repeat
DNA	Deoxyribonucleic acid
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
CODIS	Combined DNA Index System
MMR	Mismatch repair
LOH	Loss of heterozygosity
MSI	Microsatellite instability
Aadd	Additional alleles
Anew	Occurrence of a new allele
BC	Breast cancer
IHC	Immunohistochemistry
ER	Estrogen receptor
PR	Progesterone receptor
HER2	Human epidermal growth factor receptor 2
ALL	Acute lymphoblastic leukemia
AML	Acute myeloblastic leukemia
CML	Chronic myeloid leukemia
CLL	Chronic lymphocytic leukemia
FFPE	Formalin-fixed paraffin embedded
STR_{GA}	Total number of LOH and MSI
CI	Confidence intervals
SD	Standard deviation
CT	Cycle threshold

LIST OF APPENDICES

PENGUBAHAN PENANDA FORENSIK ULANGAN TANDEM PENDEK DALAM SEL LEUKEMIA DAN TISU KANSER PAYUDARA

ABSTRAK

Kanser merupakan kebimbangan yang ketara di Malaysia, dimana ia dipengaruhi oleh kecacatan genetik dari keturunan dan persekitaran. Penyelidikan kanser terkini telah menumpukan kepada penanda ulangan tandem pendek (STR) untuk mengkaji pengubahan yang terdapat dalam kedua-dua kanser pepejal dan darah. STR, yang terdiri daripada urutan berulang dengan unit motif ulangan pendek, adalah berharga untuk pengecaman forensik dan ujian paternal oleh kerana kadar polimorfisme yang tinggi. Dalam patologi diagnostik, analisis STR boleh membantu dalam pengecaman tisu, menangani kesilapan pelabelan sampel dan pencemaran. Walau bagaimanapun, penggunaan penanda STR pada tisu kanser adalah mencabar disebabkan ketidakstabilan mikrosatelit (MSI) dan kehilangan heterozigositi (LOH), yang boleh menjejaskan penemuan genotip STR. Walaupun terdapat kegunaannya, terdapat jurang dalam memahami pengubahsuaian STR dalam pelbagai subjenis molekul kanser payudara dan leukemia. Oleh itu, kajian ini bertujuan untuk menyiasat prevalens pengubahsuaian STR pada tisu sampel kanser payudara dan leukemia menggunakan kit QIAGEN Investigator 24plex QS. Sebanyak 85 sampel telah dianalisis, merangkumi 20 pasangan tisu kanser payudara formalin-terfiksasi parafin-terbenam (FFPE) beserta tisu normal bersebelahan, 20 sampel darah dan lima sampel calitan bukal daripada pesakit leukemia, serta 20 sampel darah daripada individu sihat yang tidak berkaitan sebagai kawalan. Hasil kajian menunjukkan bahawa di antara empat subjenis molekul tisu kanser payudara subjenis Luminal A mempunyai kadar LOH tertinggi, manakala subjenis sel basal mempunyai kadar MSI tertinggi. Tiada

pengubahsuaian genetik yang ketara ditemui antara leukemia berpasangan dan sampel tisu normal mereka tetapi lokus D19S433 didapati berkait rapat dengan leukemia berbanding dengan kawalan normal yang tidak berkaitan. Secara tuntas, ketidakstabilan genetik dalam karsinoma manusia telah dikenalpasti dan boleh menimbulkan cabaran dalam pengenotipan forensik dan padanan profil DNA.

FORENSIC SHORT TANDEM REPEAT MARKERS ALTERATION IN LEUKEMIA CELLS AND BREAST CANCER TISSUES

ABSTRACT

Cancer is a significant concern in Malaysia, influenced by genetic defects from heredity and environment. Recent cancer research has focused on short tandem repeat (STR) markers for alteration studies in both solid and blood cancers. STRs, characterized by repetitive sequences with short repeat motif units, area valuable for forensic identification and paternity testing due to their high polymorphism. In diagnostic pathology, STR analysis could aid in tissue identification, addressing sample mislabelling and contamination. However, applying STR markers to cancerous tissues is challenging due to microsatellite instability (MSI) and loss of heterozygosity (LOH), which can affect STR genotyping. Despite their utility, there is a gap in understanding STR alterations across different molecular subtypes of breast cancer and leukemia. Therefore, this study aims to investigate the prevalence of STR alterations on breast cancer and leukemia sample tissues using QIAGEN Investigator 24plex QS kit. A total of 85 samples were genotyped, comprising 20 paired formalin-fixed paraffin-embedded (FFPE) breast cancer tissues and their corresponding normal adjacent tissues, 20 blood samples and five buccal swab samples from leukemia patients, and 20 blood samples from unrelated healthy individuals serving as controls. The findings revealed that among the four molecular subtypes of BC tissues, Luminal A subtype exhibited the highest LOH rate, while basal cell-like subtype displayed had highest MSI rate. No significant genetic alterations were found between the paired leukemia and their normal tissue samples but the D19S433 locus was significantly associated with leukemia compared to unrelated normal control. In summary, the

genetic instability observed in human carcinomas has been recognized and may present difficulties for forensic genotyping and DNA profile comparison.

CHAPTER 1

INTRODUCTION

This chapter serves a comprehensive summary of the thesis, beginning with an introduction to the study's background to the aims of research conducted.

1.1 Research Background

Cancer poses a significant concern in Malaysia, with various types affecting the population due to genetic defects influenced by both heredity and the environment. Solid cancer and blood cancer are two distinct types of cancer, with solid cancers, such as breast cancer, arising from epithelial, glandularm and connective tissues, typically forming solid masses that can spread to other parts of the body through metastatis (Koanga et al., 2015). Breast cancer, the most common solid cancer among women in Malaysia, originates in the epithelial tissues of the breast and can range from localized tumors to invasive forms that metastasize to other organs. Blood cancers, such as leukemia, originate from the blood and bone marrow and involve the uncontrolled proliferation of damaged blood cells, including white blood cells, red blood cells, and platelets, without forming solid masses (Bhattacharyya et al., 2011; Hassan & Seno, 2020). Instead, they disrupt normal blood cell production and circulate throughout the bloodstream. These distinct forms of cancer highlight the complexity and challenges associated with their diagnosis and treatment.

Recent attention in cancer research has focused on short tandem repeat (STR) markers for modification studies in both solid and blood cancers (Alharbi et al., 2022; Chen et al., 2021). The STRs or also known as microsatellites is characterized by repetitive sequences with short repeat motif units of two to seven base pairs (bp), exhibiting

variations in the number of repeats, the repeat patterns, and sequence variations (Wyner et al., 2020). Compared to other regions of the genome, STRs undergo mutations at a significantly higher rate, ranging from 10^{-6} to 10^{-2} mutations per locus, gamete and generation (Balzano et al., 2021). Various mechanisms, including replication slippage, sister chromatid exchange, unequal crossing-over, and gene conversion, contribute to the diversity of microsatellite (Song et al., 2011). The high polymorphism of STRs among individuals makes them valuable for forensic human identification to distinguish individuals (Dang et al., 2020).

Numerous methods such as amplification of polymerase chain reaction (PCR) and hybridization-based approaches, have been developed for STR genotyping. These methods enable accurate detection and analysis of STRs, leading to their widespread application in various field. Centralized STR databases have been successfully established worldwide, facilitated by the availability of multiplex amplification of STR markers. The Combined DNA Index System (CODIS) database, initiated in 1997 with 13 core STR loci, has evolved to include an expanded number of core loci and commercial STR kits, particularly in forensics and paternity testing over the years. By employing 16 STR loci, the rates of individual recognition can achieve 99.9%.

In diagnostic pathology labs, STR analysis has proven useful for tissue identification testing in addressing mislabelled samples, mixed-up tissues and cross-contamination (Much et al., 2014). This would be especially helpful in cancer diagnosis. In forensics, DNA analysis typically uses samples like blood, saliva, hair, or other bodily fluids, but if an individual has cancer, their cancerous tissue might also provide DNA for identification by analyzing specific genetic markers or profiles. While cancer tissue

itself is not typically used for forensics identification in the context of criminal investigations or personal identification, DNA from cancerous cells can potentially be utilized for such purposes.

However, challenges exist in applying STR markers to cancerous tissues, as microsatellite stability can be affected by factors such as defects in mismatch repair (MMR) function, frequent mutation incidences and chromosomal abnormalities (Vaderhobli et al., 2014; Vauhkonen et al., 2004). MMR corrects DNA mismatches generated during DNA replication, thereby preventing mutations from becoming permanent in dividing cells by reducing the number of replication-associated errors. The presence of loss of heterozygosity (LOH) and microsatellite instability (MSI) may alter the STR genotyping findings, especially if particular alleles have varying repeat counts. Some of these mutations have been associated with various cancers, impacting illness diagnosis and prognosis, such as lung, breast, gastric, colorectal, esophageal, and renal cell cancer (Vauhkonen et al., 2004; Qi et al., 2018; Zhang et al., 2018; Chen et al., 2021).

These MSIs could be in the form of additional alleles (Aadd) or the occurrence of a new allele (Anew) instead of those found in normal tissue (Tozzo et al., 2021; Vauhkonen et al., 2004). An additional allele refers to the presence of extra copies of an existing allele at a specific locus, potentially due to replication slippage or other mutational mechanisms (Zhang et al., 2018). In contrast, the occurrence of a new allele signifies the emergence of an allele not previously observed in the corresponding normal tissue, often resulting from mutations or microsatellite instability during tumour progression (Zhang et al., 2018). Regardless of the tumour type or STR marker,

allelic loss has remained the most frequently altered mutational type, followed by the insertion of an additional allele (Qi et al., 2018). It is well accepted that chromosomal status of a patient may undergo certain alterations as their disease advances.

Therefore, this study aims to investigate the prevalence of STR alteration in malignant tissues, specifically breast cancer (BC) and leukemia samples, using the forensic commercial QIAGEN Investigator 24plex QS kit. While other related investigations involving BC tissues have been carried out, none of them have addressed molecular subtypes. However, the least researched samples to date are those with leukemia (Alharbi et al., 2022; Omar et al., 2024). This study may shed light on the suitability of using cancerous tissue samples in STR profiling.

1.2 Problem Statement

Maintaining patient identity throughout laboratory processing is critical for any diagnostic pathology laboratory. However, mislabeling of cases, specimens, blocks, and slides, as well as cross-contamination and specimen mix-ups, continue to occur in surgical pathology (Nakhleh et al., 2011). A systematic strategy using STR profiling can assist idenfy cross-contamination and specimen mix-ups in surgical pathology, enhancing patient safety and decreasing cross-contammination hazards. Nonetheless, the genetic alterations in tumour DNA, affecting both coding sections and repetitive DNA sequences, pose challenges for utilising tumour samples in forensic studies (Vauhkonen et al., 2004; Zhang et al., 2018; Qi et al., 2018; Chen et al., 2021).

Although STRs are highly useful, there is still a notable lack of understanding about STR alterations in different molecular subtypes of breast cancer and leukemia. In

contrast to breast cancer, blood cancer like leukemia, is also highly prevalent in Malaysia but has received far less focus on the study of STR alterations. Moreover, blood is frequently encountered at crime scenes, emphasizing the need for forensic studies on blood cancers. Therefore, it is essential to enhance our understanding of the suitability of these cancers for forensic identification purposes.

1.3 Research Objectives

1.3.1 General Objective

To investigate the prevalence of STR alterations on breast cancer and leukemia sample tissues using QIAGEN Investigator 24plex QS kit.

1.3.2 Specific Objectives

1 To characterize STR alterations in the different molecular subtypes of breast cancers compared to the normal adjacent tissues.

2 To identify the STR alteration pattern on leukemia blood samples compared to to the paired normal tissues.

3 To assess the association of each STR locus between breast cancer sample and unrelated nomal control individuals.

4 To assess the association of each STR locus between leukemia sample and unrelated nomal control individuals.

1.4 Significant of Study

To the best of our knowledge, no studies have utilized forensic STR kits to analyze cancerous tissues specifically within the Malaysian population. Given Malaysia's diverse genetic background, investigating STR alterations in leukemia and breast

cancer could reveal unique genetic variations that may influence forensic DNA profiling. This diversity is crucial, as forensic STR panels are designed based on population genetics, and deviations in STR loci due to cancerous mutations could impact their reliability in DNA identification. Addressing such potential alterations ensures that forensic DNA profiling remains accurate, particularly in a legal context where blood and tissue samples are common forms of evidence at crime scenes.

Beyond forensic applications, this study has significant implications for medical diagnostics and cancer management. STR alterations such as loss of heterozygosity and microsatellite instability are frequently observed in malignancies and could serve as genetic biomarkers for cancer detection and classification. By identifying patterns of STR instability unique to leukemia and breast cancer, this research could provide insights into tumorigenesis while also enhancing forensic DNA analysis. Additionally, forensic STR profiling may offer an alternative approach for resolving disputes over cancerous tissue samples, such as mislabeling or contamination in pathology laboratories. Establishing a forensic-genetic link for cancerous samples could assist in both legal and medical scenarios where tissue identification is critical.

By bridging forensic science and medical genetics, this study enhances both field, ensuring the integrity of forensic DNA evidence while supporting advancements in cancer diagnostics and personalized medicine. The findings could inform forensic policies on handling cancerous samples and provide a foundation for integrating forensic STR analysis into broader biomedical research.

CHAPTER 2

LITERATURE REVIEW

This chapter provides a comprehensive summary and evaluation of existing research and scholarly articles relevant to STR and alteration in cancerous tissue. It serves as the foundation for understanding the current state of knowledge in a field and helps readers build upon existing research to contribute new insights or advancements.

2.1 Human Genome

The human genome comprises all the DNA sequences found in humans, organized into 23 pairs of chromosomes in the nucleus human cells (Oyelade et al., 2020). Deoxyribonucleic acid (DNA) is the molecular structure that encodes the essential genetic information necessary for the proper development, growth, and reproduction of all living beings, including humans. The Human Genome Project significantly contributed to mapping and sequencing the human genome, offering a foundational understanding of human DNA and genetic variation (Sebastiani & Palmer, 2017). The DNA molecule is structured as a double-stranded helix, with nucleotides comprising the fundamental building blocks. Four nucleotide bases (adenine, guanine, cytosine, and thymine) are arranged in a specific sequence to encode the genetic instructions for the production of proteins, the key functional molecules of biological systems (Neuman, 2010).

In human beings, autosomal chromosomes (1-22) inherit genetic material, with one member from each pair originating from the parent. In contrast, sex chromosomes differ between males (one X and one Y) and females (a pair of X chromosomes). Human DNA predominantly comprises coding (1.5%) and non-coding regions

(98.5%) with most of them have lacking genetic information and function (Zhang et al., 2019). Within DNA, coding regions contain genes, serving as the fundamental units of heredity responsible for messenger RNA encoding amino acid sequences or functional RNA molecules. Meanwhile, most of the polymorphic markers used for identity testing are found throughout the non-coding regions of the chromosome. The genes situated at specific locations on chromosomes are called loci, with each cell inheriting two copies, one from each parent, can be potentially identical or distinct, forming variant forms known as alleles.

When two alleles at a specific locus are identical, they are considered homozygous, while different alleles at the same locus are referred to as heterozygous. In DNA profiling, if only one allele is detected, it indicates that a person inherits the same allele from both parents, but if three or more alleles are detected, it shows the sample may contain more than two contributors. Despite a remarkable similarity in DNA sequences among individuals, the remaining variability contributes to what makes each person unique. This understanding has been pivotal in exploring population history, human adaptability, and genetic variations linked to various diseases.

2.1.1 Short Tandem Repeats (STRs)

Short tandem repeats (STRs) refer to the DNA region with repeat units of two to seven bp in length and have become popular DNA markers over the years. STR markers can be readily amplified using the polymerase chain reaction due to their short length, which helps avoid the risk of differential amplification (Butler & Hill, 2013). STR sequences exhibit considerable variation in length, number of repeats, and repeat patterns. In forensic genetics, most STR markers are tetranucleotide repeats (4 bp

repeat motifs (Goodwin et al., 2011) .

STRs are often categorized into a few groups, such as simple repeats, compound repeats and complex repeats. The structure of repeat STRs based on the three classifications is illustrated in Figure 2.1. Simple STR markers are those characterized by consistent repeat unit length and sequence composition. Compound STR markers are composed of two or more adjacent simple repeat structures. Complex STR markers may feature multiple repeat blocks with variable unit lengths, as well as intervening non-repetitive sequences (Butler & Hill, 2013). Interestingly, some STR locus could have incomplete repeat units due to the presence of microvariants, which further contribute to the individual identification.

Multiple STR databases have been developed, the most notable being the Combined DNA Index System (CODIS), which was established in the United States in 1997. The database was founded upon a panel of 13 core STR loci included CSF1PO, FGA, TH01, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, TPOX, D21S11 and with additional sex marker, AMEL. Analyses of these 13 CODIS markers have demonstrated that the average probability of a random match among unrelated individuals is less than one in a trillion. The core set of STR loci has grown over time, and commercially produced STR analysis kits are now extensively utilized, particularly in forensic applications and paternity testing.

2.1.2 Forensic STR Genotyping Kit

Forensic STR genotyping kits are extensively used in forensic laboratories for DNA profiling in criminal investigations. Several kits have been developed and validated

for their efficiency, accuracy, and reliability (Jiang et al., 2012; Ren et al., 2021; Xiong et al., 2022; Wang et al., 2023) These kits use highly polymorphic STR markers in the human genome to generate unique DNA profiles for identification purposes.

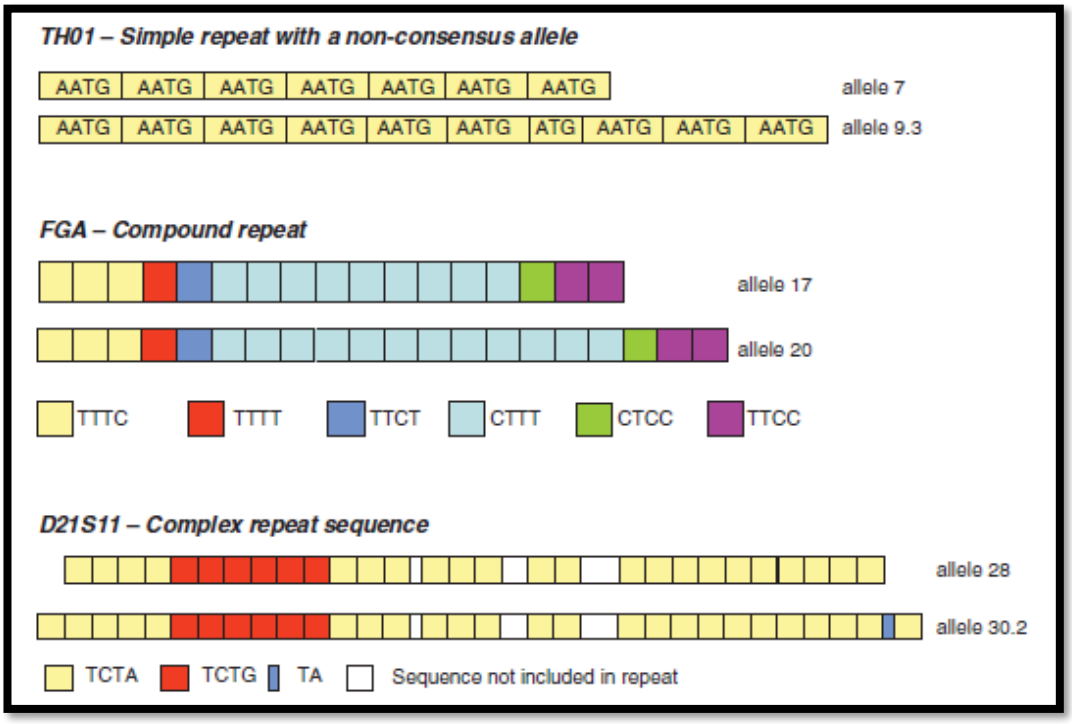

Figure 2.1 Structure of STR repeats that can be observed across three widely used STR loci: TH01 (simple repeat), FGA (compound repeat) and D21S11 (complex repeat), adapted from Goodwin et al. (2011).

The Y32 STR kit, for instance, is a multiplex amplification system designed specifically for profiling Y-chromosomal STR markers, providing a high level of discrimination capacity for identifying related male individuals (Xiong et al., 2022). The STRscan-17LC kit is a 6-dye STR panel that targets 16 STR markers, designed to be more tolerant of PCR inhibitors and capable of analysing moderately degraded samples (Jiang et al., 2012). The development of PCR-based STR typing methods was a crucial turning point in forensic DNA analysis, as it enabled more efficient analysis

of even the smallest amounts of DNA samples (Scudder et al., 2018). These kits have been shown to be reliable tools for forensic applications, providing accurate and informative results for DNA profiling in criminal investigations.

2.1.2(a) QIAGEN Investigator® 24plex QS Kit

The QIAGEN Investigator 24plex QS kit is a reliable and robust identification assay which is commonly used for forensic multiplex PCR in both casework analysis and database basing (Figure 2.2). This 6-dye multiplex assay is designed to allow multiplex amplification of the CODIS (Combined DNA Index System) core loci, the ESS (European Standard Set) markers, plus SE33, DYS391 and Amelogenin. The primer mix for the kit includes two internal PCR controls that provide valuable information about the efficiency of the PCR process and the presence of any inhibitors. Previous studies using this kit have demonstrated its ability to produce reliable and reproducible results, even with challenging samples. For instance, a previous study conducted by Zgonjanin et al. (2017) evaluated the Investigator 24plex QS kit's performance across a range of forensic samples, including bloodstains, skeletal remains, and degraded DNA. This study found that the kit consistently provided high-quality DNA profiles with minimal artifacts and dropout rates, even when working with limited or compromised DNA samples (Zgonjanin et al., 2017). Besides, it showed high concordance with other commonly used forensic kits, indicating its reliability and accuracy (Kangsanan et al., 2020).

Other studies by different research groups have also reported similar findings, confirming the kit's robustness and suitability for forensic applications (Hakim et al., 2020; Tao et al., 2022). The QS markers in the Investigator 24plex QS kit have been

found to be effective in confirming sample quality and improving STR profile completeness, especially in inhibited samples (Harrel et al., 2021). This improvement in discrimination power is attributed to the inclusion of additional and more informative genetic markers in this kit, which provides greater accuracy and resolution in differentiating between individuals.

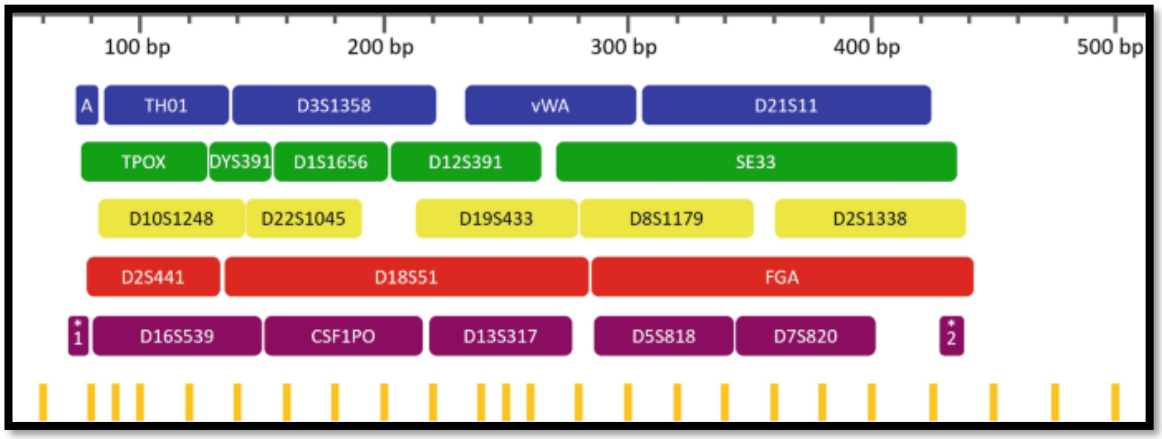

Figure 2.2 Twenty-two STR loci include two quality sensors for multiplex amplification by Investigator® 24plex QS Kit. Abbreviations: A, Amelogenin; 1, quality sensor (QS) 1; 2, quality sensor (QS) 2.

2.2 Cancer

According to projections from World Health Organization, the incidence of cancer in Malaysia is expected to significantly escalate by the year 2040, with 48,639 new diagnoses documented in 2020 alone. The 2012-2016 Malaysia National Cancer Registry Report revealed an 11% increase in new cancer cases and a 30% rise in cancer-related mortalities compared to the preceding 2007-2011 reporting period (National Cancer Registry, 2018). Primarily, cancer arises from a structural alterations

in the relation between cells and tissues, which leads to a dysfunctional cellular social structure (Cardiff et al., 2006).

2.2.1 Breast Cancer (BC)

Generally, breast is composed of a network of glandular tissues (also known a lobule) and fatty tissue that situated between the skin and the chest wall. Each breast is divided into 15 to 20 distinct lobes of milk-producing glands, with a single duct leading to each nipple. Breast cancer (BC) is a leading cause of cancer-related mortality worldwide, ranking fifth with an estimated 684,996 deaths annually, and is the most prevalent form of malignancy among women globally (Malaysia Globocan 2020, 2020). Similarly, BC is considered one of the primary contributors to cancer deaths among women in Malaysia (National Cancer Registry, 2018). BC develops when normally functioning cells in the lobules (milk-producing glands) or ducts become aberrant and divide uncontrolled (Figure 2.3). This cancer is highly complex, with many different biological subgroups, each displaying unique clinical, pathological, molecular, and genetic characteristics, responding differently to treatment, and having varying therapeutic results (Park et al., 2012).

2.2.1(a) Molecular Subtypes of Breast Cancer

Regarding metastatic potential and therapeutic response, it is well established that even breast tumours of the same histological type can behave differently. Immunohistochemistry (IHC) on tissue sections and in-situ hybridization are routinely employed in breast tumours to assess the estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) protein expression (Alizart et al., 2012). The presence of these markers used to classify BC into distinct

subtypes (Figure 2.4). These indicators are also helpful in predicting how a patient will respond to treatment. According to a large number of microarray studies, breast

subtypes (Figure 2.4). These indicators are also helpful in predicting how a patient will respond to treatment. According to a large number of microarray studies, breast tumours may be divided into five distinct subtypes: luminal A, luminal B, HER2-enriched, negative basal cell-like and normal breast types (Holm et al., 2010; Park et al., 2012; Sun et al., 2021). The association of molecular subtypes with IHC analysis can be summarised in Table 2.1. The Luminal A subtype is often associated with a better prognosis, whereas the basal cell-like subtype is more aggressive. Clinical stratification of BC has been greatly aided by the availability of simple biomarkers such as ER, PR, and HER2. It is possible to acquire a more comprehensive prediction by integrating the results from several biomarker tests.

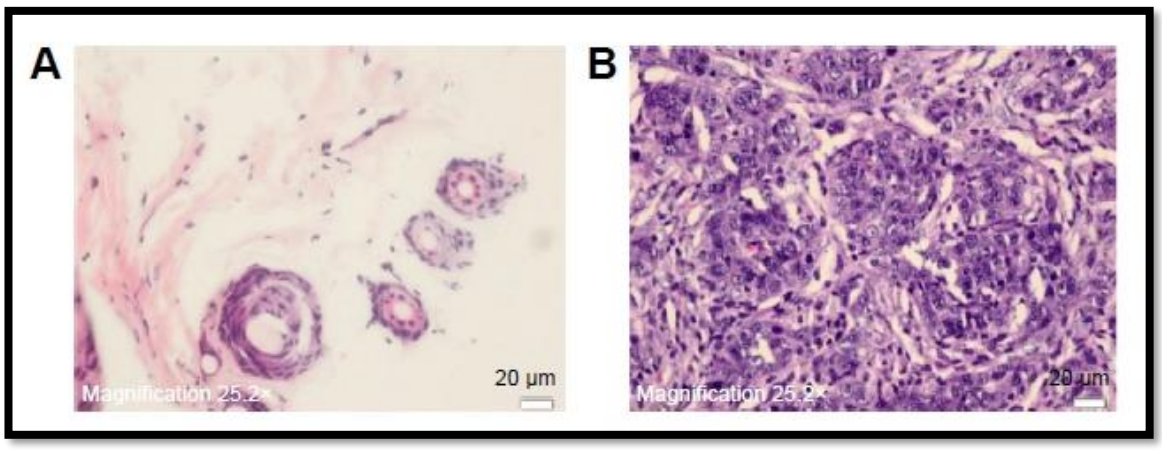

Figure 2.3 Histological examination of BC tissues: (A) normal breast tissue with conformity in cell size and shape, and lower levels off dividing cells, (B) tumour control with variation in cell size and shape, and disorganized growth and arrangement of cells adapted from Arya (2017).

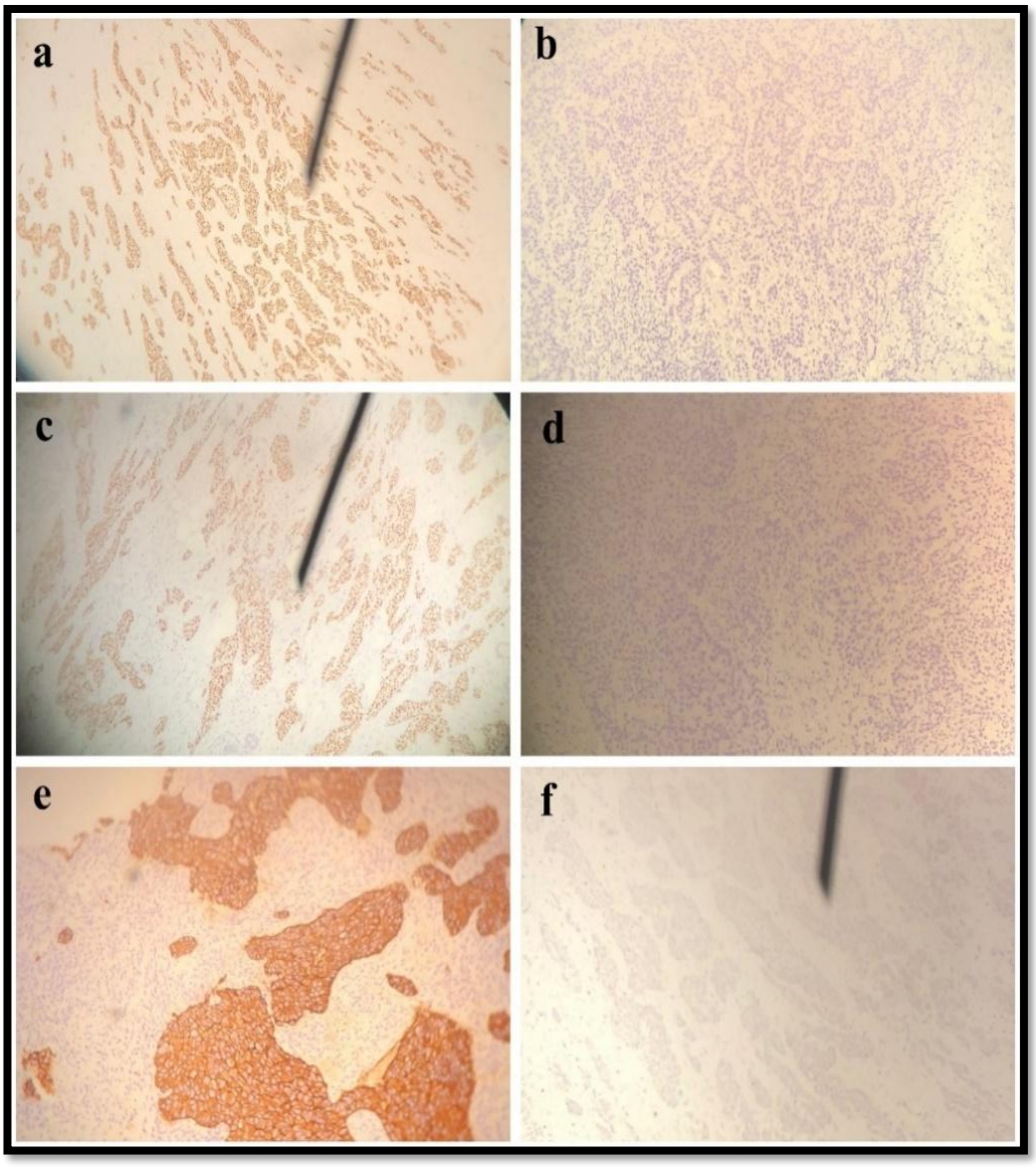

Figure 2.4 Slides of staining test under 100x magnification to determine the molecular subtypes of BC. Each slide was labelled as follows, a: ER+; b: ER-; c: PR+; d: PR-; e: HER2+. and f: HER2-.

2.2.2 Leukemia

Blood cancer, specifically leukemia, is a complex and challenging disease that affects the functioning of the body's blood cells. Leukemia is characterized by the uncontrolled production of abnormal white blood cells, which can lead to destroying the immune system of the human body. This can lead to a range of symptoms, including fatigue, frequent infections, and bleeding or bruising easily. Leukemia is classified into four main types based on the type of blood cells affected: acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and chronic myeloid leukemia (CML) (Figure 2.5).

Each type of leukemia has its own unique characteristics and treatment options. The incidence of leukemia in Malaysia varies depending on the type of leukemia. In children below the age of 15, acute leukemia is the most cancer, with acute lymphoblastic leukemia (ALL) accounting for 80% of cases and acute myeloblastic

Table 2.1 Classification of molecular subtypes of breast cancer based on Yersal & Barutca (2014).

Molecular subtypes	**Immunohistochemistry analysis**
Luminal A	ER +, PR +, HER2 -
Luminal B	Luminal B-like (HER2+): ER +, PR +/- and HER2 + Luminal B-like (HER2-): ER +, PR - and HER2 -.
HER2-enriched	ER -, PR -, HER2 +
Basal cell-like	ER -, PR -, HER2 -

Abbreviations: ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor 2; +, positive and -, negative.

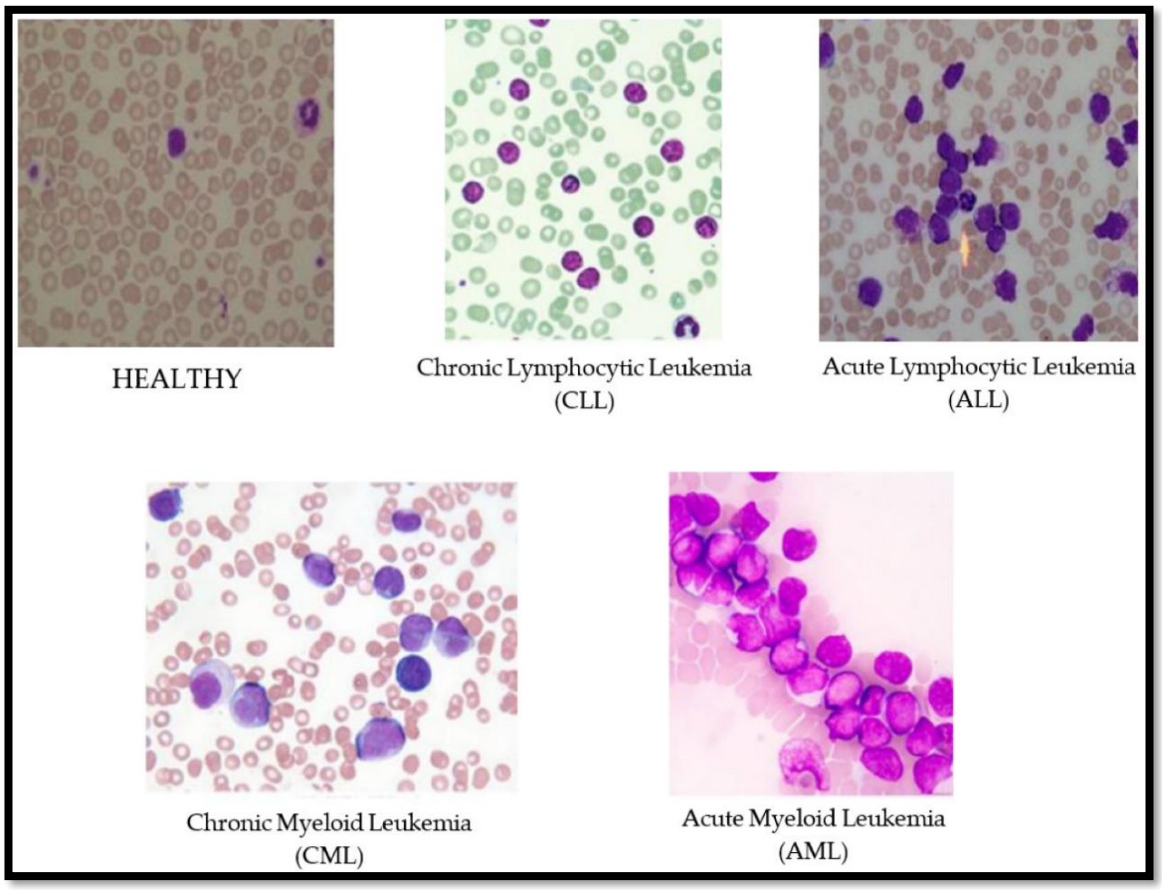

Figure 2.5 The histology of four different types of leukemia cancer compared to healthy blood based on Ahmed et al. (2019).

leukemia (AML) accounting for 20% (Abdul Rahman et al., 2008). Among adults, chronic myeloid leukemia (CML) is prevalent, with a total of 79 patients identified in a single centre in southern Sarawak (Kuan & Melaine Michael, 2018). The estimated prevalence of CML in southern Sarawak was 69.2 per 1,000,000 population in 2016, and the incidence was 8.0 per 1,000,000 population per year from 2011 to 2016. The incidence of AML in Malaysia increases with age, and different cytogenetic abnormalities show different frequencies with increasing age (Meng et al., 2013).

2.3 Importance of Genetic Markers (STRs) in Cancer Research

Genome-wide association studies, forensic medicine, population genetics and tumour genetics benefit from microsatellite markers, which are particularly useful for identifying disease causing-markers. In some situations, the only source of DNA that could be retrieved from a deceased person might be a cancerous tissue specimen from a pathology that can be used to represent individuals, whether in paternity testing or forensic identification. The biopsy material is may accidentally swapped, and lead to inappropriate treatment and psychological harm to the patient (Kalfoglou et al., 2006).

It is therefore critical to identify the specific tissue samples to prevent inaccurate DNA genotyping results. However, cancerous tissues usually harbour genetic alterations in defined coding regions and repetitive DNA sequences Long-term studies have consistently demonstrated the effectiveness of STR polymorphisms in identifying tissue samples, including those derived from malignant tissues. In forensic practice, only the complete loss of one allele and the occurrence of a new allele in the tumour tissue, instead of the one detected in normal tissue, would impacting forensic genetic accuracy (Tvedebrink et al., 2009). Moreover, it is worth noting that many STR

markers are shared among different commercial kits, allowing for the utilisation of common markers as internal controls.

2.3.1 STR Allelic Alteration in Cancerous Tissues

The failure of DNA replication in cancerous tissues frequently leads to the genomic instability with allelic losses known as loss of heterozygosity (LOH) and microsatellite instability (MSI) characterized by the accumulation of somatic alterations in the length of a microsatellite (Chen et al., 2020) (Figure 2.6).

2.3.1(a) Loss of Heterozygosity (LOH)

Allelic deletion or also known as LOH, could be further categorized into complete LOH (cLOH) and partial LOH (pLOH). It is possible to consider a cLOH only if that ratio is lower than 0.5 or higher than 2.0 (Chen et al., 2021). If the peak height of one allele is >50% decreased in the tumour, pLOH is considered (Dang et al., 2020). However, it is important to note that a reduction of one allele by more than 40–50% does not significantly impact forensic STR typing. This is particularly relevant because larger alleles may appear as very small peaks in the analysis, especially when dealing with low amounts or severely degraded DNA samples (Timken et al., 2014). As a result, this could cause heightened stochastic effects, such as diminished balance in peak heights between heterozygous alleles and a rise in the occurrence of undetected alleles (Timken et al., 2014).

2.3.2 Microsatellite Instability (MSI)

Microsatellite instability is a unique molecular alteration and hyper-mutable involving

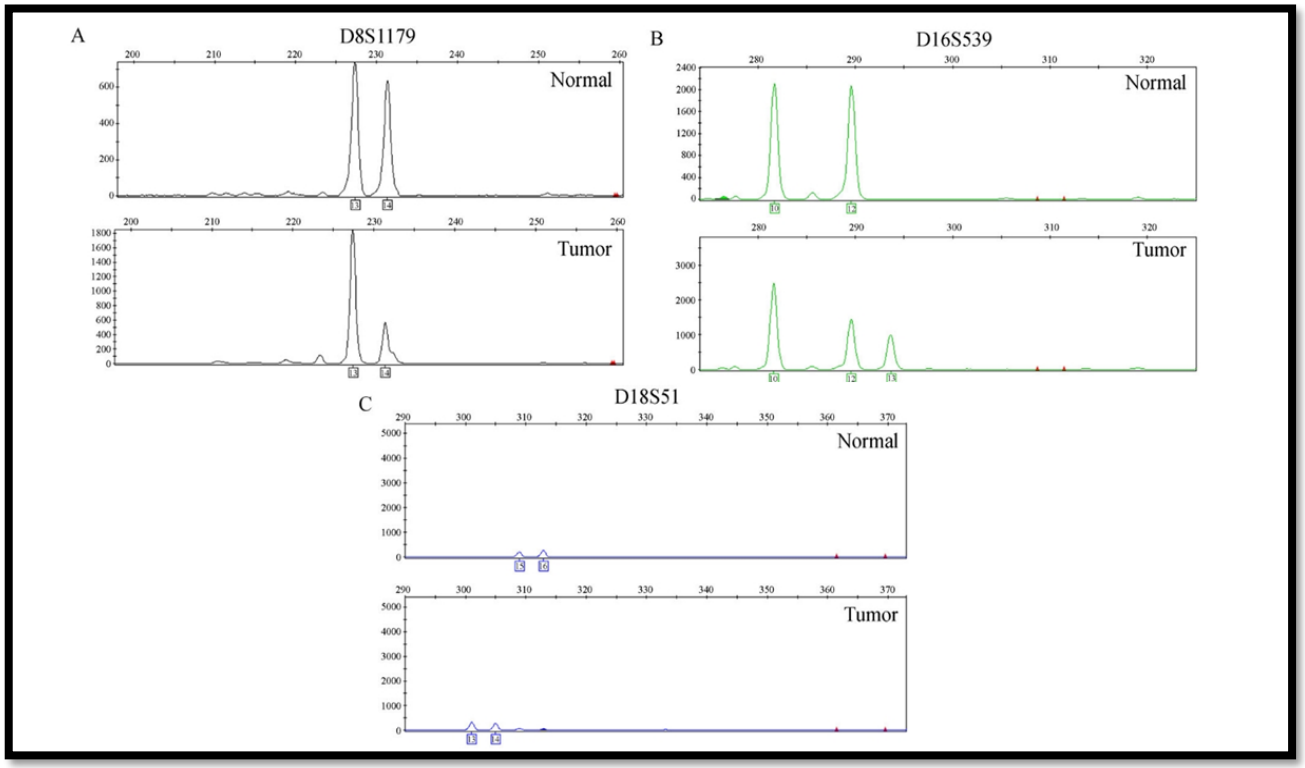

Figure 2.6 Electropherogram of STR alteration types. (A) Electropherogram at locus D8S1179 showing pLOH: upper panel, reference DNA showing a profile of alleles of 13 and 14; lower panel with half peak of allele 14. (B) Electropherogram at locus D16S539 of MSI in terms of Aadd: upper panel, reference DNA showing a profile allele of 10 and 12; lower panel with additional allele,13. (C) Electropherogram at locus D18S51 of MSI in terms of Anew: upper panel, reference DNA showing a profile of allele of 15 and 16; lower panel with new alleles of 13 and 14 adapted from Chen et al. (2020).

alternate-sized repetitive DNA sequences that are not present in the corresponding germline DNA, known to have resulted from a defective DNA mismatch repair system (Nojadeh et al., 2018). MSI in a carcinoma sample was identified by detection of allelic insertion at the STR loci compared to the control tissues (Ananian et al., 2011). The allelic insertion could be in the form of additional alleles and the occurrence of a new allele instead of those found in normal tissue.

According to (Dang et al., 2020) , a new allele was defined as the disapperance of an original allele, with it being replaced by a new allele; such as “9,10” alleles in the normal tissue mutating into “9,11” in the cancer tissue. An additional alelle is defined as an additional allele appearing with the original alleles in the cancer tissues; such as “9,11” alleles in the normal tissue mutating into “9,10,11” in the cancer tissue .

2.3.3 LOH and MSI Patterns in Cancerous Tissues

Regardless of the tumour type or STR loci, the LOH has remained the most frequently altered mutational type, followed by the insertion of an additional allele. In earlier studies, case-control studies were used to investigate the association between the STR markers and cancer. Peloso et al. (2003) examined 24 lung carcinomas, with most showing allele drop-out in at least one STR locus. Despite the LOH being detected at two loci, D5S818 and D13S317, this study had some limitations, including the small sample size and a lack of correlations between STR variation frequency and patient clinical background.

To improve this study, an examination of 75 lung cancer tissues was conducted in which 24 showed genetic alteration on at least one STR locus and a positive correlation

between STR variation frequencies with patients' age and clinical background. Using the theory of programmed onset for chronic diseases such as cancer, a cohort study was done to investigate the genetic risk of death associate with cancer. Accordingly, three loci could be confirmed as lung (D18S51-20) or liver cancer-related (D21S11-30.2 and D6S1043-18), and the genetic risk of death associated with cancer was able to be determined in this cohort study.

A recent study by Chen et al. (2021) reported that for overall forensic STR markers, the highest alteration was detected on D13S317 (19.86%), followed by FGA (16.62%), D18S51 (16.77%), D6S1043 (15.49%), D8S1179 (14.99%), and TPOX (4.34 ± 3.55%). They found that cancerous esophageal tissues obtained the highest STR alteration frequency of 25.73% ± 11.43%, while colorectal came in the second-highest STR alteration frequency with 16.85%. This finding supported their previous findings in a study of 250 gastrointestinal (GI) cancer where the most frequently altered loci observed were D18S51 (32.79%), CSF1PO (22.09%), and Penta E (22.45%) (Chen et al., 2020).

2.3.3(a) STR Alterations in Breast Cancerous Tissues

BC is considered to have the lowest mutation rates, ranging from 0% to 48.60% (Omar et al., 2024). The common altered markers for LOH are D13S790, D18S51, FGA, TH01, and D21S11, with mutation rates ranging from 5.97% to 36.00%. Furthermore, BC tissues show low rates of pLOH alterations, ranging from 1.23% to 1.43% for markers D16S539, D21S338, D22S1045, D2S441, D3S1358, D1S1656, and SE33. Additionally, low mutation rates of MSI (1.00% to 4.00%) were observed in FGA, D18S51, and D1S103 markers.

Hou et al. (2017) conducted a study on BC and found that STR variation was observed in most STR markers, except for the Amel and Penta D markers. Nonetheless, the autosomal STR variation was generally low in all stages of BC differentiation (Hou et al., 2017). Recent research has shown that BC contains fewer mutations compared to colorectal cancers and gastrointestinal neoplasms, indicating the relative genetic stability of BC (Kalfoglou et al., 2006) These findings have been supported by subsequent investigations, which consistently found that the mutation rates of STRs in BCs were generally lower than those of gyneacology cancer (ranging from 0% to 48.60%) (Ananian et al., 2011; Chen et al., 2021; dos Santos et al., 2017; Hou et al., 2017; Powierska-Czarny et al., 2003; Tozzo et al., 2021).

Although BC samples are generally claimed to have microsatellite stability, a few loci were observed to have a variable high percentage of mutation. This discrepancy in findings may be attributed to sample heterogeneity across studies. It is possible that some studies claiming microsatellite stability focused on specific BC subtypes that are generally considered to have a low frequency of MSI (Ananian et al., 2011; Hou et al., 2017; Powierska-Czarny et al., 2003). On the other hand, studies reporting loci with high mutation rates might have focused on different subtypes or specific genomic regions that are inherently prone to instability (Chen et al., 2021; dos Santos et al., 2017; Tozzo et al., 2021).

2.3.3(b) STR Alterations in Leukemia

STR analysis is a valuable tool in leukemia research, particularly for monitoring chimerism after allogeneic hematopoietic stem cell transplantation (Kim et al., 2014). It has been used to identify the origin of leukemia cells, such as in cases of donor-

derived leukemia following cord blood transplantations (Plesa et al., 2022), and to detect donor cell leukemia by revealing full donor chimerism in leukemic blast samples (Kurosawa et al., 2016). Additionally, STR analysis serves as a quality control measure to prevent sample mix-ups and monitor minimal residual disease in leukemia patients (Hoogeveen et al., 2018).

LOH and MSI are common alterations observed in leukemia samples (Das-Gupta et al., 2001). Defects in DNA repair pathways can lead to the accumulation of microsatellite alterations in genomic DNA, contributing to genomic instabilities in various cancers, including leukemia (Filoglu et al., 2014). On the other hand, MSI, has been associated with specific subsets of acute myeloblastic leukemia (AML), indicates mismatch repair defects and can lead to the occurrence of new alleles in STR loci, highlighting the genetic instability in leukemia pathogenesis (Das-Gupta et al., 2001). MSI and LOH alterations have been linked to poor outcomes in leukemia patients, underscoring their prognostic value which is used to predict the outcomes of a disease (Velasco et al., 2004).

Furthermore, the impact of leukemia on the detection of STR markers has been studied, revealing differences in allele number at specific loci between individuals with leukemia and healthy controls (Alharbi et al., 2022; Filoglu et al., 2014). These alterations in STR markers can occur due to defects in DNA repair pathways and the accumulation of changes in microsatellite loci, emphasize the need to consider the effects of leukemia on genetic analyses (Filoglu et al., 2014). However, the lack of studies on STR alterations in leukemia highlights the need for further research to