Childhood Acute Lymphoblastic Leukemia

Genetic and Epigenetic Analysis of Archived Samples

Master's Thesis by

Laeya Abdoli Najmi

Supervisor: Helge Klungland

Norwegian University of Science and Technology (NTNU)

The Faculty of Medicine

Department of Laboratory Medicine, Children's and Women's Health

Trondheim, August 2012

NTNU
Norwegian University of
Science and Technology

Acknowledgements

This thesis is the result of a Master degree at university of NTNU at the Medicine Faculty, Department of laboratory medicine, Children's and women's health in the spring 2012.

I would like to express my special thanks and appreciation to my supervisor, Professor Helge Klungland, for his invaluable advices, knowledge and experience which guided me a lot along the process of this thesis. I am also grateful to my co-supervisor Bendik Lund for his support, help, and cooperation to develop this work better. I also appreciate Veslemøy Malm Landsem for helping me in practical laboratory work and for her great feedbacks during writing of this thesis. I also would like to express my thanks to laboratory engineer Kristi Rain for her cooperativeness during laboratory works.

I am really obliged to my dear friends, thank you for all your support, help and positive energy.

To my great family, for all their love and support, for helping me when I couldn't, and for sending constant love and support across the world for the past two years!

August 2012

Laeya Abdoli Najmi

ABSTRACT

Acute lymphoblastic leukemia (ALL) is recognized as a fast-developing cancer originated from blood-progenitor cells. Blasts cells are immature cells which generate white blood cells (leukocytes), and it is the malignancy of the blast cells which lead to leukemias. The bone marrow is gradually filled up with these blasts and as a result, the production of healthy blood cells will be damaged. Malignant cells might also find their way to the blood circulation and have the ability to infiltrate vital organs as the brain and spinal cord. As the number of healthy bone marrow cells decrease, the development of severe organ failure will take place, and it will turn into a lethal disease.

Great advances in leukemia treatment have resulted in high cure rates of more than 80% in children. However, treatment related death for this disease is still 2-4%. For further treatment improvement, it is required to customize treatment for each individual patient. The interindividual differences in response to treatment and its toxicity are caused by many factors in which genetic variations including single-nucleotide polymorphisms (SNPs) seems to play an important role. The development of genome-based treatment is possible by making associations between an individual genetic make-up and the drug response. The uses of archived samples increase the feasibility of the retrospective study. In the present study, archived samples from patients who died because of treatment toxicity were used for multiple SNPs analysis and DNA methylation study.

DNA was extracted from smears and formalin fixed paraffin embedded bone marrow tissues. The quantity of isolated DNA was measured by UV spectroscopy and Fluorometric methods, and the quality of the isolated DNA was assayed by evaluation of the ability of samples that were amplified using DNA profile analysis. Generally, smears were able to amplify markers up to 234 bp and FFPE tissues up to 170 bp. In this study, multiple SNPs analysis failed in most of the samples with highly degraded DNA. Based on the findings, the average SNPs call rate was 91% for reference blood samples and 74% for smears with 4x sequencing depth.

In a parallel study, DNA methylation of IL-8 was analysed by methylation-specific PCR using archived samples. In this methylation analysis, all samples were amplified successfully to an amplicon size of 173bp. We detected IL-8 hypomethylation in 98% of bone marrow smears and in 96% of FFPE bone marrow tissues in patient with acute lymphoblastic leukemia.

In conclusion, amplifiable DNA was extracted from archived samples. The whole genome amplification was not efficacy for highly degraded DNA samples. The results obtained through this study confirm the possibility of doing multiple SNPs analysis and STR markers amplification by archived samples. However, they need to be optimized in terms of better quantity and quality control methods to get more successful results.

LIST OF ABBREVIATIONS AND SYMBOLS

5-MeC	5-Methylcytosine
Akt	Serine/threonine kinase
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
bp	base pairs
CBC	Complete blood cell count
CpG	Cytosine-phosphate-guanine
CXCL8	Cxc chemokine ligand 8
CXCR1	Cxc chemokine receptors
CXCR2	Cxc chemokine receptors
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
DTU	Danmarks Tekniske Universitet
EDTA	Ethylen ediamine tetraacetic acid
ER	Estrogen receptor
FFPE	Formalin Fixed Paraffin Embedded Tissue
g	Gravity
HCHO	Formaldehyde
HSC	Hematopoietic Stem Cells
ID	Identification
IL-8	Interleukin-8
MAPK	Mitogen-activated protein kinase
MDA	Multiple displacement amplification
MDR1	Multi-drug resistance gene 1
mL	Millilitres
MSP	Methylation-Specific PCR
ND-1000	NanoDrop TM 1000 spectrophotometer
NF-κB	Nuclear factor-Kb

ng	Nanograms
ng/µL	nanograms per microliter
NOPHO	Nordic Society for paediatric Haematology and Oncology
OD	Optical Density
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
pg/µL	picograms per microliter
PI3K	Phosphatidylinositol 3-Kinase
PKC	Protein kinase C
q	Chromosome long arm
qPCR	Quantitative polymerase chain reaction
RFU	Relative fluorescence units
RNA	Ribonucleic acid
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
T-ALL	T lymphocyte
TBE buffer	Tris-borate-EDTA buffer
TRD	Treatment related death
U	Units
UV	Ultraviolet light
VNTR	Varying number of tandem repeats
WGA	Whole genome amplification
µL	Microliters
µm	Micrometre
µM	Micromolar
Φ	Phi
°C	degrees Celcius

CONTENTS

Table of Figures

1 Introduction

In general, cancer is a group of different diseases characterized by unregulated cell growth. In cancer, division and growth of cells are out of control to form lumps or masses of tissue called tumors. The cancer may also move to distant parts of the body through the blood or lymph systems and destroy healthy tissues. Cancers are usually diseases of middle age and older. The incidence of the most types of cancer increase after age 50. Although childhood cancers are uncommon, they account for a substantial proportion of childhood deaths. About 1,545 children under age 15 die from cancer in United State [1].

The blood cells formation basically takes place in the bone marrow and comprises a balanced process of proliferation, differentiation and cell survival. In leukemia, uncontrolled proliferation of immature malignant cells, damages the reformation of healthy blood cells. More malignant development forces the leukemia cells to enter into blood circulation. Finally, this will result in infiltration of organs in various parts among which the most common ones include spleen, liver and kidney. It would turn into a lethal disease, if it was left without treatment.

All mature blood cells are generated from a relatively small number of Hematopoietic Stem Cells (HSCs) as a common ancestor. The pluripotent haematopoietic stem cells generate multiple committed stem cells, including lymphoid or myeloid progenitors. The lymphoid progenitors have the capacity to differentiate into B or T lymphocytes, and myeloid progenitors can give rise to red cells, platelets, monocytes and granulocytes.

Based on the origin of the cells, Leukemia is divided into lymphoid and myeloid leukemia. Lymphoid leukemia is separated into T- and B-lineage leukemia, while myeloid leukemia has several types based on the types of the involved cells. Finally both lymphoid and myeloid leukemia can be classified into chronic and acute conditions. One of the characteristics of acute leukemia is its rapid progress and accumulation of immature malignant cells. Acute leukemia mainly afflicts in children and young adults. While chronic leukemia progresses slowly and engages more mature blood cells. It also occurs in elder people and urgent treatment is not required, and consequently, it can be postponed to be sure that the maximum efficiency of the treatment is occurred.

Leukemia is the most prevalent cancer in childhood. It is the cause of around 30% cancers in children. Acute lymphoblastic leukemia (ALL) is the most common type; almost 80-85% of childhood leukemia and about 15-20% is acute myeloid leukemia (AML)[2-3]. In the Nordic countries (Norway, Denmark, Finland, Iceland and Sweden) about 175-200 children are diagnosed with ALL each year [4]. An annual incidence rate in Europe and US is approximately 3.5 per 100,000 children younger than 15 years old [5].

Progresses in the management of ALL has resulted in increasing the cure rate up to 80-85 % of the patients with ALL [6]. The most significant drawback of this great advance is that up to 3-5% of patients die due to toxic side effects of the anticancer treatments. Most of Treatment Related Death (TRD) occurs because of immunosuppresion and cytotoxic effects of anti-cancer drugs or by the leukemia which inhibits bone marrow recovery during induction therapy. Also, patients treated by the same protocol vary significantly in treatment-related toxicity. Usually all patients experience infections due to immunosuppression related to treatment, but only some suffer other severe complications such as thrombosis, hepatotoxicity, organ toxicity and other serious effects [7-8].

In order to improve efficiency of childhood leukemia treatment, clinical impact of genetic variations should be investigated. The responses of the patients to the drugs are different and could also be unpredictable because of host factor in the individual genome.

1.1 Genetic Polymorphism

The human genome is made of 3.2 billion base pairs. Approximately 99.9% of DNA sequence is similar among individuals across the population; the remainder (0.1%) represents genetic polymorphisms which arise from evolutionarily stable mutation in the genome. Frequent variation at a particular locus in the genome is described as a genetic polymorphism. In other words, a locus is polymorphic when there is more than one allelic form existing among individuals in the same population. An allele is usually described as polymorphic providing that it is observed with a relative frequency of more than 1% in the population. The considerable importance of Genetic polymorphism is its role as a tool to allocate and determine the human genome which is responsible for single gene disorders.

There are different types of genetic polymorphisms including tandem repeat polymorphism and base-substitution polymorphism.

Varying number of tandem repeats (VNTR) are highly polymorphic regions of DNA sequences which vary between individuals in terms of the repeated unite length and the number of repeated sequence times. A class of VNTR is short tandem repeats (STRs), also called microsatellites consisting of di-, tri- and tetra-nucleotide repeat units. STR is the most informative markers for gene mapping and other genetic analysis. The term "mini-satellite" is used when the length of the repeating unit is between 10 to 100 base pairs (bp) [9].

Single Nucleotide Polymorphisms (SNPs) are the most common form of DNA variation, arising from one single base pair substitution. For example, SNPs might alter DNA sequence namely AAGGC to ATGGC. SNPs account for 90% of all human polymorphisms and occur at the frequency of 1 in 1, 000 bp throughout of the human genome [10].

1.1.1 Coding Region SNPs

Coding regions comprise low percentage of human genome, thus the majority of SNPs have no significant functionality.

Synonymous: The substitution happens in the third variable position of the amino acid codon which does not cause amino acid alterations in the resulting protein. These synonymous SNPs are called silent because they do not alter amino acids.

Non-synonymous: The substitution leads to the change of encoded amino acid and alters the gene protein product which is called a missense mutation. If the substitution leads to a misplacement of a termination codon, it is called a nonsense mutation. Around half of the coding SNPs are non-synonymous.

1.1.2 Non-Coding Region SNPs

Vast majority of SNPs have no functional consequences when they occur in non-coding regions of the genome. Polymorphisms can also change transcription level and create splice variation when they occur in non-coding regions as in the promoter or splice sites,

respectively. SNPs occurring in regulatory regions of genes have the capability to affect the level of protein expression or the timing of protein production [11-13].

SNPs are not the main causes of disease. They can increase the disease susceptibility or resist to its development. SNPs could determine the level of severity or progress of a disease and they can change the body response to the drugs [14]. SNPs are progressively persistent and do not change among generations which provide a stable indicator in order to study genetic polymorphisms in population [12]. Sequence variations are typically recognized by doing DNA sequencing and the comparison of sequence reads among individuals and alignment to database entries. After any SNP discovery, frequency determination and association studies should be conducted to determine functional relevance of polymorphism at a statistically reliable level. For this purpose, high-throughput technologies are needed to handle massive amount of analyses. Recently, the development of second-generation technology has widely allowed the researchers to identify large number of SNPs in the genome. Those gathered information will make precise link between the genotype and the phenotype. These SNPs analyzing technology can be applied for identifying individual SNPs risk profiles and for individualizing and optimizing drug therapy [15-16].

1.2 SNPs and drug response

The role of genetic polymorphisms in genes coding for drug-metabolism has increased clearly since 20 years ago. Genetic polymorphisms of drug-metabolizing enzymes, their receptors and transporters cause inter-individual variation in drug responses. Therefore SNPs could affect absorption, transportation, metabolism and excretion of the drugs. Consequently, some drugs show better response in some patients compared to others but some are more toxic in certain individuals [17].

Large individual variations in drug disposition are responsible for treatment failures, severe and even lethal toxicities. There is a growing list of polymorphisms found in genes that affect drug targets metabolizing enzymes, drug transporters and disease-modifying genes. However this field faces many challenges to completely discover the contribution of genetic variation into inter-individual differences in drug effects and translate the new findings to clinical practice.

In most cases, candidate gene approaches are conducted in SNP screening. Candidate genes are chosen based on their functions, structures and locations. Then DNA sequencing is performed from these genes or their significant regions i.e. exon, promoter and enhancer. Although selected genes are important, it is technically difficult to understand the function of specific polymorphisms. Therefore, the study of pathways of genes is more important than the study of individual genes, because the effects of a polymorphism in the network of genes acting together to generate a single phenotype. The correlation of genomics and medicine has the potential to become a new diagnostic tool which can be utilized for optimization of drug therapy [11].

Reliable identification of the functions of SNPs is needed for better diagnosis, identification of new cancer genes and personalized treatment. Although translation of these findings into clinical application may not occur in short period, they will result in discovering of novel genes involved in pathophysiology of investigated traits [15, 18]. However extensive clinical research will consequently be needed before applying these new findings in treatment protocols.

The main challenge with regard to the study of clinical impact of genetic variation is a need for homogeneous patient populations treated by the same regimen and minimal puzzling variables. Childhood acute lymphoblastic leukemia (ALL) is one of the optimal models addressing these challenges. Based on a unique network between all pediatric oncology centers in the Nordic region, our study was planned to screen approximately 30,000 individual SNPs related to genes encoding proteins involved in pharmacology, immunology, DNA repair mechanisms, mitosis activity, genes affecting apoptosis, neurotoxicity and thrombosis. SNPs were chosen if they were within coding regions, splice sites and regulatory regions, with the aim of exploring the combined effects of the thousands of already known SNPs with the clinical outcome of childhood ALL within these biological domains. Multiple SNPs analyzing makes a definitive step towards individualized patient therapies.

1.3 Professional ethics

This master study project is a part of a large Nordic project was partly in collaboration with other ongoing projects at Bonkolab, Rigshospitalet in Copenhagen. All studies have been approved by the research ethics committees in Denmark and Norway. For all Norwegian participants, an additional written consent has been collected. The study has been performed in accordance with the Declaration of Helsinki.

1.4 Biological samples

Gathering and collecting of biological samples and their storage for future studies are significant aspects of biological research. It is imperative to have efficient storage procedures which could preserve sample integrity over time. Today, billions of biological samples are collected in hospitals, research and medical institutes. These samples are deployed for diagnosis. In addition, they might fit for research applications depending on sample nature, size, storage and ethical implications. In current experiments, blood samples, bone marrow smears and formalin fixed paraffin embedded bone marrow tissues were used.

1.4.1 Blood samples

Blood samples are frequently used in diagnostics and are convenient to take. They do often have high quality DNA even in samples stored for many years.

1.4.2 Archived samples

Although many institutions are equipped with frozen tissue banks to respond to the growing request for molecular analysis, few of them can support large scale of genetic analyses and often they do not have enough historical follow up information to get precise clinical data [19]. Collection of biological samples is a routine process to preserve samples in pathology laboratories as a virtual historical archive of each disease. Estimates show that there are more than 300 million tissue blocks in the United States with an increasing rate of 20 million samples every year. Paraffin blocks have been collected and maintained for a period of a century, representing a historical information base for diseases. Most of the samples contain valuable medical history of patients which makes them a precious source for identification and production of disease biomarkers [20].

Archived samples such as paraffin-embedded blocks frequently form the core of retrospective biomolecular analysis. Retrospective studies have multiple benefits. For example, in many situations in diagnostic pathology, there might not be necessity of analyzing at the time of gathering of samples; such necessity may arise after the samples' are being archived. On the other hand, retrospective study provides basis to study rare diseases in large number of groups compared with the prospective study which requires fresh samples.

Bone Marrow Smear:

A complete blood cell count (CBC) test is the first step in diagnosing of ALL. If the blood test results show blood abnormality such as increasing number of white blood cells, decreasing number of red blood cells and platelets or presence of blast cells in blood, a bone marrow aspiration is the next step. In normal conditions less than 5% of bone marrow cells are blast cells (the undifferentiated cells that normally develop into healthy blood cells). In acute leukemia, blasts cells constitute between 30 - 100% of the bone marrow [21].
Preparing Wright-Giemsa stained glass slides of bone marrow cells or peripheral blood is a routine procedure in Hematology clinics. Bone marrow clots rapidly and EDTA can be used to prevent clotting. For smear preparation one drop of aspirated bone marrow is expelled on a glass slide, a second glass microscope slide is put over the first one (longitudinally) and the top slide is smoothly pulled to the end of the bottom slide. The bone marrow smears picture was taken by Bendik Lund in our lab.

Gene amplification by PCR has significantly facilitated the analysis of DNA derived from archived samples. It does not require high quality DNA, the technique can support template derived from archived clinical specimens, since small quantities of degraded DNA can act as the template for the reaction. The capability to use routinely processed archived materials facilitates in large-scale retrospective studies to be conducted. Some studies have proved the possibility of deriving amplifiable DNA and RNA from archived air-dried unstained bone marrow slides as well as from archived Giemsa-stained peripheral blood smears and archival Giemsa-stained bone marrow slides [22-23].

Formalin Fixed Paraffin Embedded Tissue (FFPE):

In pathology archive, storage of tissue sample as formalin fixed embedded tissue is a common way to maintain specimen for a long time. Collection of paraffin blocks, as histology-based archives have been done for more than a century. Estimates show that there are more than one billion tissue blocks in the United States [19]. Fresh dissection tissues are usually fixed with formalin. This fixative not only maintains tissue efficiently but also preserves morphological structure relatively intact through induction of protein-nucleic acid cross-links. So, unfortunately, great majority of specimens are fixed in formaldehyde for a reason other than genetic analysis. The quality and status of the trapped DNA within these clinical specimens make them challenging to use, so their diagnostic and therapeutic significance should be investigated [24].The picture of FFPE tissue was taken by Bendik Lund in our lab.

1.5 Fixation effects on DNA quality

Different chemical fixation reagents are used to well-preserve morphology and chemically characteristics of biological specimens. Routinely used formalin in preparation of FFPE tissues adversely affects DNA quality. Multiple factors affect DNA such as fixation condition and time of storage of sample. The latter one has the most important effect on DNA quality because of interaction of residual formalin within FFPE tissue with DNA during storage period [25].

Formaldehyde (HCHO), a main component in formalin, reacts with nucleic acids and causes several modifications. Formaldehyde adds a hydroxymethyl group to a nitrogen atom of the nucleic acid, this step is reversible. At the next step, electrophilic attack of hydroxymethyl group on amino bases generates stable methylene bridges between two adjacent bases. The other important effect of formalin is hydrolysis of the phosphodiester bonds which causes breakage of the strands of DNA and results in DNA fragmentation [20, 26].

1.6 Quality and quantity assessment of isolated DNA

In processing archived samples in large-scale, DNA extraction step needs to be not only simple but also rapid and it must not affect the amplification of PCR. Due to poor quality and limited amounts of recovered DNA from archived material, accurate assessments of quantitative and qualitative points of view are significant.

Assaying quality of DNA is a critical step to achieve meaningful data from initial material to decide which kind of technique can be supported by these materials. It is necessary to have a reliable estimate of the quality of DNA prior to the time and resources invested for downstream processes. There are several methods for DNA quality assaying, for example gel electrophoresis and southern analysis. The use of gel electrophoresis does not predict the utilization of DNA for PCR-based methods, because of DNA cross-linked which is caused by fixation. Although these methods give information about DNA fragmentation, not all could predict the capability for successes in PCR. Several studies have shown usefulness of PCR-based assays for DNA quality-control from archived samples [24, 27].

1.6.1 UV spectroscopy

The most common method to determine DNA concentration and purity is measurement of absorbance at 260 nm. The maximum absorption of ultraviolet light (UV) occurs at 260 nm for nucleic acids, a property which is used to determine the concentration of nucleic acids in a sample by measuring Optical Density (OD). The potential contamination of a DNA extracted by organic compounds, e.g. polysaccharides, phenols or by proteins can be assessed by measuring OD at 230 nm and 280 nm respectively. A 260/230 nm absorbance ratio above 1.8 and a 260/280 nm absorbance ratio around 2.0 are considered to be acceptable [28-29].

1.6.2 Fluorescence spectroscopy

The extensive availability of fluorescent DNA binding dyes and fluorometers provide another popular option for measurement of DNA yield. Fluorescence base methods are more sensitive, especially for low concentration samples. It uses specific fluorescent dyes for DNA, RNA or Protein molecules separately. The dye molecules become intensely fluorescent upon binding to target molecules and the amount of the fluorescent signal is proportional to the concentration of the related components [30].

1.6.3 Gel electrophoresis

The purpose of the gel might be either to determine DNA concentration or to estimate the quality of DNA fragmentation. The DNA is visualized in the gel by adding intercalating fluorescent dye such as ethidium bromide. In the quality checking of the DNA, intact DNA should appear as compact, high molecular weight band while degraded DNA results in low-molecular weight smears [29].

1.6.4 Quality assay of isolated DNA by DNA profiling

Short Tandem Repeats (STRs) are highly polymorphic short segments in non-coding DNA regions with repeated sequence pattern of two or more nucleotides. The STRs repeated units range from 2 to 7 base pairs that are repeated for example $(CATG)_n$, one after another (in tandem). The differences in STR alleles are caused by size variation due to difference in the number of times the units are repeated. Creating a unique genetic profile is made possible by analyzing multiple STR loci and counting the number of STR sequence occurrences at a given locus [31].

Routinely DNA profiling is used for genotyping, human identity testing, forensic and paternity testing. But in the present study, DNA profile is used to assay the quality of isolated DNA from archived materials and to estimate the length of fragmented DNA. It could be a reliable method to check the quality of the recovered DNA in comparison with one single gene study. The DNA profile analyzing is also used to check that no contamination exists and that the sample belongs to the correct person. Due to limited amount of recovered DNA from archived material, capability of this method has the capability to assay the quality of DNA by using 1 ng of genomic DNA. For this purpose, multiplex PCR is performed, and the PCR product is screened via capillary electrophoresis.

The STR marker analysis evaluates ten different loci which are distributed in various loci in the human genome. The nine STRs are unlinked regions distributed through 9 autosomal chromosomes in the human genome as shown in Figure 1. Exceptions are the CSF1PO and D5S818 markers which are both on chromosome 5 in 5q33.3-34 and 5q21-31, respectively. One fragment from the Amelogenin gene is located on both X and Y chromosomes. The

amplified fragments of this gene are slightly longer on the Y chromosome compared to that on the X chromosome (113 bp and 107 bp respectively). A male genome shows two different lengths (107/113 bp) whereas a female genome displays two similar lengths, so this can be used for gender identification [32].

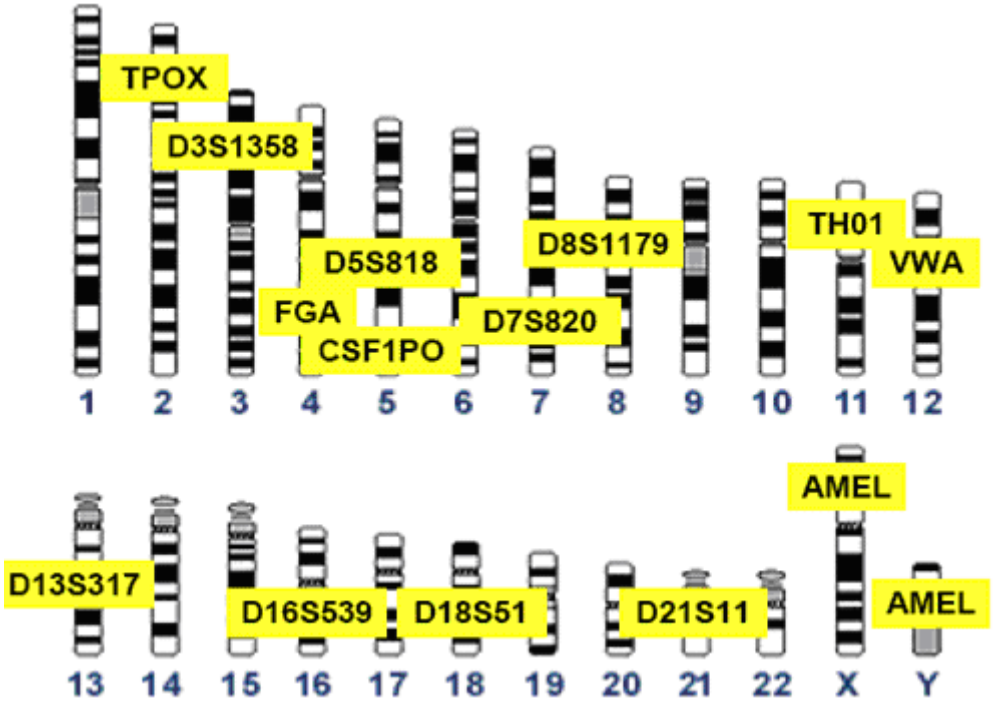


Figure1: The position of the STRs markers from the AmpFℓSTR® profiler kit in the genome. NB: the markers D8S1179, D16S539, D18S51 and D21S11 are not present in the kit. From Technology [31].

The method is fluorescence based PCR using multiple dye technology which enables co-amplification of loci with overlapping size within one multiplex PCR reaction. One primer of each locus –specific primer is labeled with 5-FAM, JOE or NED and ROX dye which are detected as blue, green, yellow and red (internal standard), respectively. The internal size standard normalizes difference in electrophoretic mobility between gel lanes or injections. The number of repeats is constant for every individual and is used to make a specific genetic profile. The Allelic ladder is an external standard used to genotype analyzed samples. Allelic ladder comprise of the most common alleles for each loci [32].

Amplified fragments are separated on fluorescence based electrophoresis on a capillary electrophoresis machine. Amplified fragments, which are fluorescently labelled, migrate through a 50 cm capillary filled with polymer (POP7). By applying high voltage DNA fragments with negative charge move toward cathode through the polymeric capillaries. DNA fragments with fluorescent labels separated by their size and move along the path of the laser beam just before getting to the cathode. The dyes on the fragments then are fluoresced by the effect of laser beam. This fluorescence effect is recorded by using an optical detection system and then converted into digital data by data acquisition software. The results appear as electropherograms which display florescent intensity indicated as relative fluorescence units (RFU) on Y-axis and base pair size on X- axis. Each peak represents a fluorescently DNA fragment with particular size and quantity based on the amount of fluorescent signal [32-33].

1.7 Whole Genome Amplification

Whole genome amplification (WGA) methods which are in vitro reactions are designed to non-specific amplification of whole materials involved within samples containing low amounts of DNA. These methods provide sufficient DNA template for molecular analysis. Ideally in WGA methods, every amplified DNA would be a true representative of the initial DNA and lead to identical results which are not distinguishable from the input DNA. Human DNA amplification is a challenging process through which more than 3 billion faithful amplifications of bases should be done without any loss or preferential amplification of each specific loci or alleles.

A great effort has been directed to improve whole genome amplifications techniques to provide sufficient amount of DNA to support robust high-throughput analysis. Highly degraded DNA isolated from FFPE tissues prevents successfully whole genome amplification through standard procedure. The REPLI-g FFPE principle combines multiple displacement amplification (MDA) with possessive DNA polymerase activity which result in much more reliable yield compared with PCR-based WGA methods. The MDA basis is the strand-displacing activity of the Φ 29 DNA polymerase by using random primers to amplify DNA in an isothermal temperature at 30 °C (Figure 2). DNA template is continually copied by branching mechanism, as Φ DNA polymerase synthesizes new strands while ‘strand displacement’ activity concurrently displaces previously extended strands. The Φ 29 DNA

polymerase performs a highly and continuous elongation of each individual DNA strand without disconnection from the template which leads to synthesis of long strand [34-35].

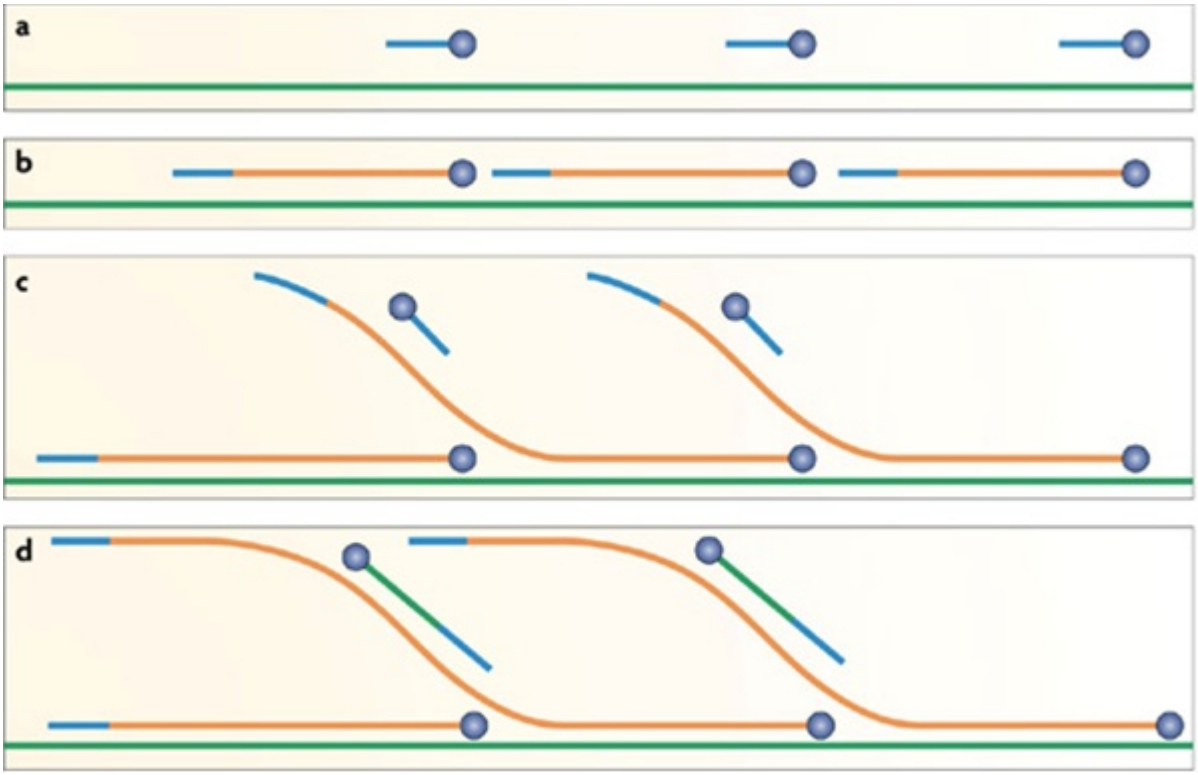


Figure 2: Schematic diagram of REPLI-g DNA amplification. Φ29 DNA polymerase amplification method "(1) The random hexamer primers (represented by a blue line) bind to the denatured DNA (represented by a green line); (2) The Φ29 DNA polymerase (represented by a blue circle) extends the primers until it reaches newly synthesized double-stranded DNA (represented by an orange line); (3) The enzyme proceeds to displace the strand and continues the polymerization, while primers bind to the newly synthesized DNA; (4) Polymerization starts on the new strands, forming a hyperbranched structure". From Spits [34].

The REPLI-g FFPE procedure is random ligation of DNA fragments followed by binding of random hexamer to denatured DNA and amplification by REPLI-g Polymerase (Figure 3).

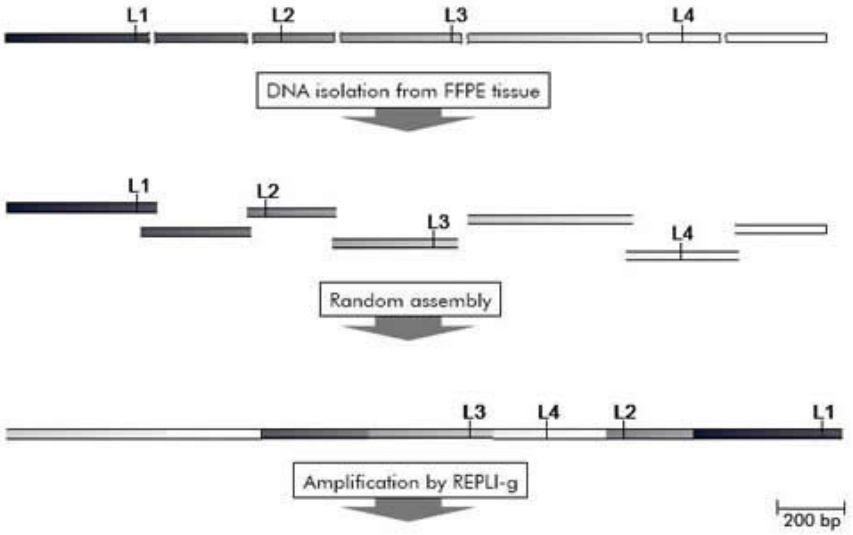


Figure 3: Random DNA ligation in REPLI-g FFPE procedure. Fragmented DNA isolated from FFPE tissues are randomly ligated and before amplification. From Qiagen [36].

1.8 Library preparation

The ability to read the sequence of bases comprising a polynucleotide has a significant impact on biological research. The invention of 'next generation' sequencing techniques has changed the development of DNA sequencing at a great extent. They could process thousands to millions of DNA templates simultaneously. As a result not only the cost of per generated sequence base will decrease but also the throughput will be on the gigabase scale. Ultimately, whole-genome sequencing provides more understanding about both full spectrum of genetic variation, and the pathogenesis of complex traits.

New techniques and protocols have been developed for next generation sequencing to provide diverse application including genetic polymorphism. The routine sequencing of large numbers of whole genomes has not been feasible yet, because it's still time consuming and implies high costs. Therefore, considerable effort has led to develop "target-enrichment" methods. This approach allows selecting genomic regions of interest from DNA samples and to enrich these regions prior to sequencing [37].

Selection of interesting regions of the genome for sequencing can reduce cost and efforts significantly compared with the whole genome sequencing. Several approaches to target enrichment have been developed. SureSelect target enrichment system (Agilent Technologies, Santa Clara, CA, USA) is based on a Hybrid capture approach. The Agilent SureSelect Target Enrichment system is based on hybridization capture method which permits us to sequence only genomic regions of interest. The Agilent SureSelect platform allows capturing all exons or custom design targets which could be subset of exon or other genome regions, and the rest of the genome is discarded.

Through the Hybrid capture, nucleic acid strands which are derived from input samples are hybridized to prepared DNA fragments as a complement to targeted regions of interest. Thus, the interested sequence could be physically captured and isolated. Short length fragments of library preparation are required for enrichment by hybrid capture (normally from 100 to 250 bp) which are synthesize prior to the hybridization step.

The SureSelect method is amongst the most efficient hybrid selection techniques to capture specific regions of the entire genome. The technology utilizes biotinylated RNA capture probes ("bait") which are complementary to target regions of the genome. Then all targeted sequences are captured in one hybridization reaction. After hybridization, streptavadin-coated magnetic beads were used to capture the oligos. Then nonspecific hybrids are washed away and targeted DNA is eluted. Targeted DNA ("catch") is amplified and then prepped libraries are ready for sequencing [37-38]. Experimental pipeline is shown in Figure 4.

The quality of the input DNA sample influences the performance of the targeted enrichment approach. Having enough DNA with good quality is required for any downstream processes. If low amounts of the genomic DNA are available, WGA is typically applied. While, WGA generate just a representation and not an intact copy of the genome, it could make bias in final results. This could be compensated by handling the samples in the control group in a similar way [38].

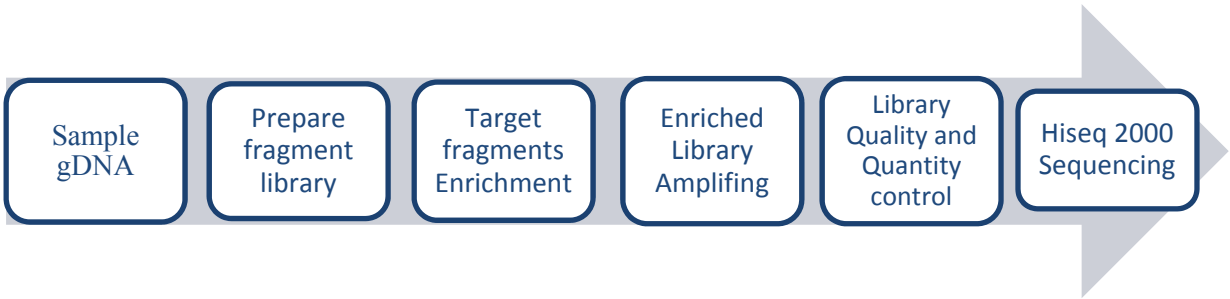


Figure 4: The experimental pipeline. The workflow includes the following steps: shearing genomic DNA into random fragments, enriching the target fragments of interest region by SureSelect Target Enrichment System protocol (Agilent Technologies) and this is followed by HiSeq 2000 sequencing technology. From Agilent Technologies [39].

1.9 DNA methylation analysis

The term "Epigenetic" describes a heritable change in gene expression without any changes in DNA sequence. Two main factors that promote epigenetic alterations are DNA methylation in cytosine bases in CpG dinucleotide and post-translational histone modification.

Disturbance of balance epigenetic arrangement may significantly impact the chromatin configuration and transcriptional activity. Patterns of DNA methylation and gene expression of various genes are extremely disruptive in human cancer. Almost half of the genes in the human genome contain CpG islands in the proximal regions of the promoters which are unmethylated in normal cells. These epigenetic characters serve as substitutions to mutations and deletions in inactivation of tumor suppressor genes. A huge number of genes involving fundamental cellular pathways may be influenced by unusual methylation of CpG islands in connection with transcriptional silencing in a variety of human malignancy[40]. Statistically speaking, conducted hypermethylation studies are much more compared to hypomethylation ones [41-42]. Hypermethylation has been found usually in CpG islands of genes. A large numbers of genes are subjected to hypermethylation in cancer such as DNA repair, cell cycle regulation, apoptosis, drug resistance, angiogenesis and metastasis.

More regions of the genome are subjected to second type of methylation, hypomethylation modification, rather than methylation. The biological significance of hypomethylation

modification is less understood in human malignancies. Global genomic hypomethylation has been observed in most of human cancer such as breast cancer, prostate cancer, cervical cancer, hapatocellular cancer and in hematologic malignancy as well [41].

DNA methylation pattern could be used not only as a biomarker in detection of cancer but also as a tool for prognosis evaluation and a therapeutic target. This specific feature of DNA methylation is due to the fact that it is heritable and reversible [41, 43].

ALL is a heterogeneous malignant disorder with various biological and clinical characteristics. Diagnose and therapy of ALL depends on various factors such as age of patients, chromosomal abnormality, immunophenotype and the risk of nervous system involvement. Aberrant methylation of several genes such as calcitonin genes, p21, Cip1/Waf1, cyclin-dependent kinase, multidrug resistance gene 1(MDR1), estrogen receptor gene (ER), p15 and P16 is found in Acute lymphoblastic leukemia [41, 44-46].

1.9.1 IL8 and human cancer biology

Interleukin-8 (IL-8), also known as CXCL8, is a member of the chemokine family produced by several normal cells (macrophages, neutrophils and endothelial) and malignant human cells. It has been observed that IL-8 contributes to human cancer progression through mitogeniec and angiogenic effects. Some studies show overexpression of IL-8 by tumor cells which are induced in response to chemotherapeutic drug or environmental factors such as hypoxia. Increasing production of IL-8 has significant effect on tumor microenvironment result in expression of IL-8 receptors CXCR1 and CXCR2 in cancer cells [47-49]. IL-8 activates several signaling pathways through two cell surface receptors, i.e. CXCR1 and CXCR2 (Figure 5). As a result of divers' effects of IL-8 in downstream targets, IL-8 promotes angiogenic, proliferation and survival in cancer cells as well as potentiates migration of tumor cells [47].

Most of the research regarding methylation is done on promoters with multiple CpG islands; however, analysis of promoters with sparse CpG site has been largely ignored. The IL-8 contains sparse CpG sites in the promoter; the selected CpG dinucleotides are located between -136 and +43 nucleotides in the IL-8 promoter. This region contains binding sites for

transcription factors NF-KB and activator protein-1 which are responsible for over-transcription and constitutive expression of IL-8 in malignancy condition [50-51].

IL-8 plays a vital role in human cancer progression; few studies have been carried out to investigate methylation status of this gene. Hypomethylated status of the IL-8 gene promoter have been shown in various human cancers including colorectal cancer, breast cancer, lung cancer, prostate cancer and cervical cancer [52]. IL-8 is a chemoattractant cytokine and plays a role in several hematopoietic malignancies as well. Several studies have reported high level of mRNA and gene expression of IL-8 in hematopoietic malignancy [48-49, 53]. Consequently, we decided to study IL-8 methylation status in childhood ALL by using archived materials.

.

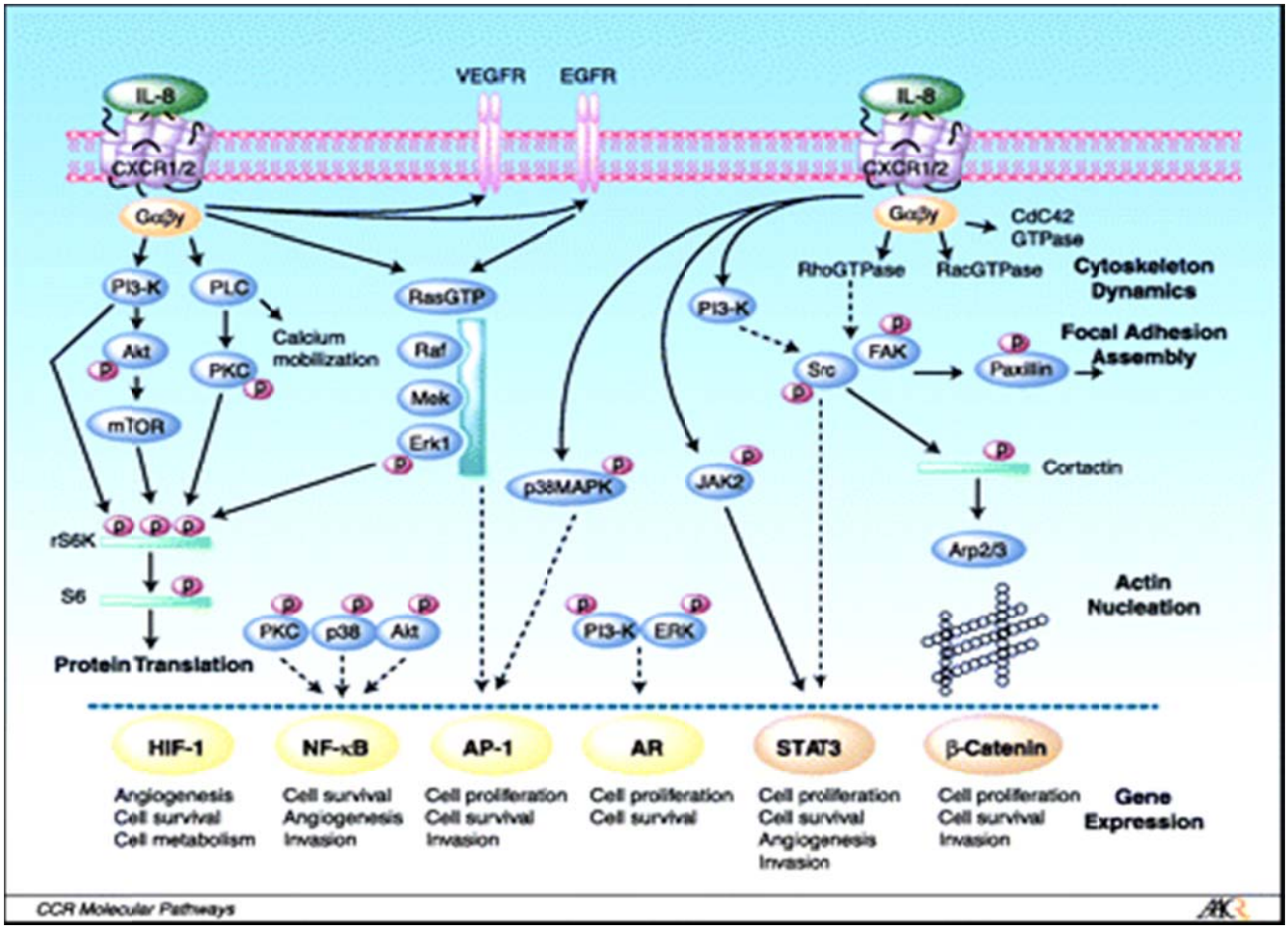


Figure 5: IL-8 Signaling Pathways. The figure illustrates the range of signaling pathways activated by CXCR1and CXCR2 receptors. IL-8 signaling pathway activates G proteins which lead to activation of PI3K, Akt, PKC and MAPK signaling pathways. These activated pathways have been shown to activate protein translation and regulate different transcription factors activities. As a consequence, IL-8 signaling pathway promotes proliferation, survival, angiogenesis, tumorigenicity and metastasis in cancer cell. From Waugh [47].

2 Aim of study

Genetic variations in human genome significantly influence the response to disease treatment. This genetic variation is a key determinant of interindividual differences in treatment resistance and toxic side effects. The present master study project is a part of a large Nordic project where the main goal is to analyze several thousand of known SNPs to determine genetic polymorphisms within immune response genes in childhood ALL, and to investigate whether they are associated with treatment related toxicity. For some patients who died following treatment, however, only archived samples are available. In the present study suitability of archived samples for multiple SNPs and methylation analysis have been evaluated.

We aimed to do this by performing DNA isolation from archived bone marrow slides and formalin fixed paraffin embedded bone marrow tissues. Quantity and quality control of isolated DNA were assessed. To overcome limited amount of isolated DNA, whole genome amplification was also applied. Major part of this study focused on quantification and qualification of isolated DNA from archived samples in high-throughput single nucleotide polymorphism analysis.

In parallel, the suitability of archived materials for epigenetic studies was investigated. In order to do so, methylation status of IL-8 was evaluated in patients with acute lymphoblastic leukemia. The overall aim of this study was to investigate the applicability of amplified DNA extracted from archived samples in multiple SNP and methylation analysis.

3 Materials and Methods

3.1 Study population

The present study was a part of the main project with title of "Genetic variation affecting treatment related to toxicity of childhood acute lymphoblastic leukemia" related to NOPHO (Nordic Society for pediatric Hematology and Oncology). The aim of the main study was to determine genetic polymorphisms within immune response genes in childhood ALL, and to investigate whether they are associated with treatment related to toxicity with special emphasis on treatment related death and infectious complications**.**

In the main project, approximately 2700 patients who were treated by the NOPHO-1992 and NOPHO-2000 ALL protocol were included as well as 90 cases of treatments related deaths. The study focused on clinical data from the NOPHO database, and additional data from a questionnaire collected from the different centers. Genetic analysis of approximately 30,000 SNPSs were carried out by using Illumina high-throughput sequencing. The selected candidates' genes were relevant to the immune system pharmacology, cell cycle, DNA repair, apoptosis, drug metabolism, neurotoxicity, and thrombosis. Stored DNA samples from 700 patients treated under ALL protocol from 1992 to 2007 in Denmark and Norway were analyzed. They have been treated according to NOPHO-ALL 1992 and NOPHO-ALL 2000 protocols. The SNP analysis is associated with clinical outcomes including toxic death and severe infectious complications in these patients. In this study, if any associations are identified the results will be used to carry out a prospective confirmatory study in the Nordic countries in order to be able to predict which patients are at greatest risk and may develop severe infectious and inflammatory complications. Based on these genetic studies, it may be possible to improve the individualization of chemotherapy in order to reduce treatment related mortality, thereby increasing overall survival. The targeted microarray may also provide a platform for other studies on genetic impact of therapy in other diseases where patients are immunocompromised.

In the main project, blood samples of the patients were used for multiple SNP analysis. However, the blood samples of some patients were not available, especially those who died during treatment (TRD). Therefore, we conducted an experimental study to evaluate archived samples as starting materials for multiple SNPs analysis. In the first setup, we included eleven stored archived samples (bone marrow smears and bone marrow biopsies) from St.Olavs hospital, Trondheim, Norway. To evaluate the quality of the SNP profiling, archived material was compared with fresh taken blood samples from patients who had finished treatment. Also, the same group of patients and samples were subjected to epigenetic study.

3.2 DNA isolation

DNA was extracted from the following samples: (1) bone marrow smears; (2) Formalin-fixed paraffin-embedded bone marrow tissues ;(3) Blood samples

3.2.1 DNA isolation from bone marrow smears

Giemsa-stained bone marrow smears of the patients which had been stored in the archives of St.Olavs hospital in Trondheim were used in this study. DNA was isolated from smears according to the following procedure. The cover slides were separated from the glass slides by immersion in xylene which was followed by putting the slides in ethanol bath for 5 minutes three times. Later, the slides were exposed in open area be dried completely. Volume of 20-30µL of PBS buffer was pipetted on the glass slide and the cells were carefully scraped from the slide surface with a sterile Razor blade. Then the mixture of buffer and scraped cell is pipetted into a 1.5mL Eppendorf tube (Hamburg, German) and the DNA was extracted using the QIAamp DNA Micro kit (QIAGEN, GmbH, Germany). The scraped material was re-suspended in buffer ATL to a final volume of 100µL, then 10µL of proteinase K and 100µL buffer AL was added. After vortexing, the mixture was incubated at 56°C for 10 minutes, 50µL of ethanol was added, and incubated for 3 minutes at room temperature after vortexing. Then the supernatant was added to QIAamp MinElute column and centrifugated for 1 minute at 6000 g. The flow-through liquid was discarded and 500µL wash buffer I, containing guanidine-hydrocholoride and ethanol, was added before centrifugating for 2 minutes at 6000g. The flow-through liquid was discarded and a second washing step using 500µL wash buffer II was performed. The next step was centrifugation for 2 minute at 8000g, again discarding the flow-through, and then centrifugating for 3 minutes at 20,000g. Finally,

DNA was eluted into a sterile 1.5mL Eppendorf tube by addition of 50µL AE buffer and centrifugation for 1 minute at 20,000g. At the end, 5mL of extracted DNA was transferred to a separate Eppendorf tube for DNA concentration measurements and both tubes containing DNA were frozen at -20 °C.

3.2.2 DNA isolation from formalin fixed paraffin embedded bone marrow tissues

Standard microtome machine with disposable blades was used for preparation of new cut section of FFPE block tissues with thickness of up to 10 µm. DNA was isolated using the QIAamp DNA FFPE Tissue Kit (QIAGEN, GmbH, Germany). QIAamp FFPE Tissue procedure consists of 6 steps:

- Removal of paraffin: paraffin is dissolved in xylene and removed
- Lyse: sample is lysed under denaturing conditions with a short proteinase K digestion
- Heat treatment: incubation at 90°C reverses formalin cross-linking
- Bind: DNA binds to the membrane and contaminants flow- through
- Wash: residual contaminants are washed away
- Elute: pure, concentrated DNA is eluted from the membrane

Briefly, five tissue sections of 10 µm were transferred into a 1.5 mL Eppendorf tube; 1 mL xylene was added to remove the paraffin from tissue sections. The tube was vortexed for 10 seconds and centrifuged for 2 minutes at maximum speed (20,000g). 1 mL ethanol was added after removing the supernatant to eliminate residual xylene, followed by centrifugation for 2 minutes at full speed (20,000g), then the supernatant was carefully removed and the tube was incubated at room temperature to completely evaporate all residual ethanol. The pellet was re-suspended by adding 180 µL buffer ATL and 20 µL proteinase K, vortexed before incubation at 56°C for 1 hour so that the sample would completely be lysed. After the lysing step, it was incubated at 90°C for another 1 hour; this heating step could reverse to some extent formaldehyde modification on nucleic acids [54-55]. This is progressed by adding 200 µL of AL buffer before vortexing, and then 200 µL ethanol was added. Samples were transferred to QIAamp MinElute column after vortexing, then centrifugation for 1 minute at 6000g. The flow-through liquid was discarded and 500 µL wash buffer I, containing guanidine-hydrocholoride and ethanol, was added before centrifugation for 1 minute at 6000g. The flow-through liquid was discarded and a second washing step using 500 µL wash buffer II was performed. It followed by centrifugation for 1 minute at 8000g, discarding the

flow-through, then centrifugation for 3 minutes at 20,000g to dry membrane completely. Finally, DNA was eluted into a sterile 1.5 mL Eppendorf tube by addition of 50 µL AE buffer and centrifugation for 1 minute at 20,000g. DNA concentration was measured and samples were frozen at -20 °C for later analyzing.

3.2.3 DNA isolation from Blood samples

DNA was extracted using the QIAamp DNA Mini kit (QIAGEN, GmbH, Germany) according to the manufacturer's instructions. For DNA extraction, 20 µl proteinase K was added into a 1.5 mL Eppendorf tube, followed by adding 200 µl of blood sample and 200 µl of AL buffer. The sample was vortexed before incubation at 56 °C for 10 minutes to completely lyse the cells. Then 200 µL of ethanol was added, finally DNA bonded to silica_based membrane and residual contaminants were washed away. Finally, DNA was eluted with 50 µl AE buffer or distilled water, and the DNA concentration was measured and the sample was stored at -20 °C for later analyzing.

3.3 Assessment of DNA concentration

Accurate quantification of isolated DNA is significant to make an approximation of the DNA yield and its suitability for future applications. DNA concentration can be assessed using various methods; two methods including ultraviolet light (UV) and fluorescence spectroscopy have been extensively used.

3.3.1 UV spectrophotometric measurements

The purity and concentration of DNA extracts were assessed by OD measurements using NanoDrop TM 1000 spectrophotometer; (Thermo Fisher Scientific, Waltham, MA, USA), referred to here as the ND-1000. Each sample was measured at least twice. Sterile water (Aqua B. Braun, Melsungen, Germany) was used as a blank. To avoid carry-over effect between the samples, the researcher wiped each sample compartment with lens paper before each measurement. UV scan in the range of 220 nm to 320 nm reveals potential DNA contamination. The detection limit of ND-1000 spectrophotometer is 2 ng/µL up to 3700 ng/µL without dilution [56].

3.3.2 Fluorometric mesurments

The Qubit dsDNA BR assay kit (Invitrogen, Carlsbad, CA, USA) referred to here as Qubit, was used to measure DNA concentration with Qubit™ Fluorometer. The assay is extremely selective for double stranded DNA and is accurate for 100 pg/µL to 1000 ng/µL of initial sample concentrations[57]. Concentration of DNA was measured according to the manufacture's recommendations. The thin-wall, clear, 0.5 mL PCR tube (500 tubes, Cat.no.Q32856) was used for Qubit measurement. Working solution was made by diluting dsDNA BR reagent 1/200 in dsDNA BR buffer. Each standard tube required 190 µL of working solution and 10 µL of each standard. For each assay of samples, 1 µL of sample was added to assay tube containing 199 µL of working solution. For each assay, final volume was 200 µL, followed by vortexing for 2-3 seconds. Then the tubes were incubated for two minutes at room temperature. Samples were read by Quibt 2.0 Fluorometer. The results are related to sample concentration after dilution; we calculated concentration of original samples.

3.4 Whole Genome Amplification procedure

The REPLI-g-FFPE kit (QIAGEN, GmbH, Germany) provides uniform amplification of the entire genome. The principle is based on randomly ligation of DNA fragments before amplification. WGA was performed using REPLI-g-FFPE kit (QIAGEN, GmbH, Germany), according to the manufacture's instruction. Briefly, 100 ng of DNA template was added to a tube and volume was adjusted to 10 µL with TE buffer, then sample was denatured at 95 °C for 5 minutes, and then cooled down an ice. A mixture containing 8 µL of FFPE Buffer, 1 µL of ligation Enzyme and 1 µL of FFPE Enzyme was added, mixed and then centrifuged briefly. The reaction was incubated at 24 °C for 30 minutes. In this step DNA fragments are ligated to form high molecular weight DNA. The reaction was stopped with incubation at 95°C for 5 minutes by using a Techne thermo-cycler (Tc-512,

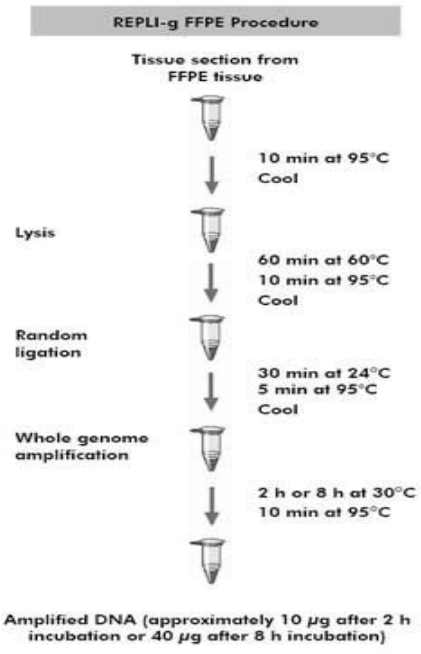


Figure 6: REPLI-g procedure from REPLI-g FFPE kit.

Burlington, NJ, USA). After that, a mix of 29 µL of REPLI-g Midi reaction Buffer and REPLI-g Midi DNA polymerase was added to the denatured DNA and then incubated at 30 °C for 8 hours by using the Appiled Biosystem Thermal Cycler 2720. The amplification step was ended by incubation at 95 °C for 10 minutes. The reaction was stopped before incubation at 95 °C to remove an aliquot to DNA quantification by Qubit.

3.5 Purification of REPLI-g amplified DNA

Purification of WGA products was carried out using the QIAamp Mini Kit. According to the Qiagen supplementary protocol, 50 µL of amplified DNA was added into a 1.5 mL Eppendorf tube, followed by adding 150 µL nuclease-free water. After vortexing, 200 µL of AL buffer was added and continued by briefly vortexing and centrifugation. The precipitation of DNA was performed by adding 200 µL of ethanol giving a pellet upon centrifugation, repeating vortexing and centrifugation step. Then the mixture was transferred to a QIAamp spin column and was centrifugated for 1 minute at 6000 g. The flow-through liquid was discarded and 500 µL wash buffer I, containing guanidine-hydrocholoride and ethanol, was added before centrifugation for 1 minute at 6000 g. The flow-through liquid was discarded and a second washing step using 500 µL washing buffer II was performed. The next step was centrifugation for 3 minutes at 20,000 g, again discarding the flow-through, and then centrifugation for 1 minute at 20,000 g. Finally, DNA was eluted into a sterile 1.5 mL Eppendorf tube by the addition of 100 µL AE buffer and centrifugation for 1 minute at 6000 g. At the end, DNA concentration was measured and samples were frozen at -20 °C.

3.6 Assessment of DNA quality

The ability to rapidly assay DNA quality is required before proceeding with downstream analysis. There are various methods to assay DNA quality. Gel electrophoresis is one of these methods through which DNA fragmentation is estimated shown. However, it cannot predict the ability of samples to support PCR. The previously published studies suggest using multiplex PCR analysis to estimate DNA quality precisely [24, 60]. Although most of the predicting assays using multiplex PCR require 100 ng of initial material, the amount of isolated DNA is a limiting factor in this analysis.

3.6.1 Gel electrophoresis

The DNA (250 ng) extracted from blood, smears and FFPE material was run on a 0.8 % agarose gel using 0.5xTBE buffer (Tris-Borat-EDTA) for 2 hours and the λ DNA Hind III Digest (New England Biolabs) was used as a molecular weight marker. The gel was stained by ethidium bromide and was visualized under UV illumination. Gel electrophoresis was performed after DNA isolation, after WGA and clean up of the WGA products. DNA from smears and FFPE tissue produced a slight smear (consistent) which indicated poor quality or degraded DNA.

3.6.2 DNA profile procedure

The quality of multiplex PCR amplified DNA was assayed by using AmpFℓSTR® Profiler kit (Applied Biosystem, Foster, CA, USA). For each PCR setup a mastermix consisting of reagents listed in Table 1 was prepared. 1ng of isolated DNA of each sample was used in reaction. Diluted DNA was used in 25 µl reaction mix, 1ng DNA in 10 µl dH2O. The following temperature cycle was programmed to Thermal Cycler GeneAmp ® PCR system 9700 (Applied Biosystem, USA): 95 °C for 11 minutes for initial strand separation, followed by 28 cycles of 94 °C for 1 minutes; primer annealing 59°C for 1 minute, extension step at 72 °C for 1 minute then final elongation at 60 °C for 45 minutes. After the completion of PCR reaction, amplified fragments were separated on ABI3730 capillary electrophoresis machine (Applied Biosystem, HITACHI, USA).

Application into the ABI3730 96 wells plate:

For running in capillary electrophoresis, 0.2 mL Non-skirted 96 well PCR tube (AB-0600, Thermo scientific, UK) was used. A mixture of 10.0 µl formamide and 0.5 µl of 500 lizTM internal line size standard was prepared .To each of the wells on the 96-well plate, 1.05 µl of prepared mixture was then added, then 1 µl of samples and allelic ladder were added to the designated wells on the plate. The plate was then covered with sealing tap and placed on a microplate shaker, with moderate shaking speed for 30 seconds. Finally the assay plate was placed on capillary machine. DNA fragments were separated based on the size using capillary electrophoresis and the smallest fragments move faster.

DNA fragments are excited through a laser while they move past a detector where they are identified and sized to a single base pair. The results were analyzed by GeneMapper v3.7 software and observed as electropherogram. Two sources of data obtained in generation of DNA profile, include retention time and signal strength. The retention time in comparison to allelic ladder define alleles to individual peaks. Every peak on the electropherogram stands for fluorescently labelled DNA fragment with an exact size as characterized by the number of base pairs, and a particular height based on the florescent signal strength. The strength of the signal generated shows the peak height which has positive linear correlation with DNA quality[31, 33].

Table 1: PCR amplification of DNA with the AmpFℓSTR® Profiler kit

Reagent	Amount
AmpFℓSTR PCR reaction mix	10.5 µL
AmpFℓSTR AmpliTaq Gold (DNA polymerase)	0.5 µL
AmpFℓSTR Profiler Primer Set	5.5 µL
Addition of diluted DNA sample	10.0 µL

3.7 Library preparation for sequencing using SureSelect Target Enrichment System

The availability of high-throughput of next generation sequencing platforms combined with high throughput of target capture methods provides the ability to screen thousands of SNPs simultaneously. The budgetary limitation for this kind of study is both cost of sample preparation and sequencing. In this method, pooling of eight samples before capture enrichment makes it a cost effective analysis platform to screen thousands of SNPs simultaneously, targeted by custom-designed baits.

DNA shearing and library preparation were done according to SureSelect Target Enrichment System protocol (Agilent Technologies, Santa Clara, CA, USA) with modification in pooling samples prior target enrichment after adding unique barcodes to each sample as shown in Figure 7. In brief, in the first step, 3 µg of input DNA was shared by Covaris S2 System (Covaris Inc., Woburn, MA, USA), followed by purification of sheared DNA using

Agencourt AMPure XP beads kit (Bekman Coulter, Inc). Purification step was repeated after each reaction. Then the quality was assessed with Agilent 2100 Bioanalyzer using high sensitivity DNA kit (Agilent Technology). In the next step, the end-repairing was performed by using T4 DNA polymerase, T4 phosphonucleotide kinase and klenow fragment enzyme, adding "A" bases to the 3'end of DNA fragments. In the next stage custom-made adapters were added. These adapters contained unique barcodes of four base pairs and directly ligated to each DNA fragment. Barcodes addition allowed pooling the samples which made them to be distinguishable after data gathering. The prepped DNA library was amplified by using Phusion High-Fidelity PCR Master Mix (Finnzymes, Espoo, Finland), the following temperature cycles were programmed: denaturation at 98 °C for 30 seconds, followed by 14 cycles of 94 °C for 10 seconds for denaturation, annealing at 65°C for 30 seconds and extension at 72 °C for 30 seconds. Final extension was performed at 72°C for 5 minutes. Quality and quantity of DNA was assessed with ND-1000 spectrophotometer and Agilent 2100 Bioanalyzer, respectively.

After pooling of eight DNA libraries – the above mentioned modification step in protocol – was carried out by mixing 62 ng of each sample in one tube. The pooled DNA library was hybridized with custom-designed SureSelect Oligo Capture library SureSelect (Agilent Technologies) for 24 hours according to manufacturer's instructions. After completing hybidiziation step, hybrid capture was purified by magnetic beads. It was followed by post-hybridization amplification step, Standard primers from SureSelect Target Enrichment System kit and Herculase II fusion DNA polymerase (Stratagene, Agilent Technologies) were used. The following temperature cycles were programmed: Initial denaturation at 98 °C for 30 seconds, followed by 18 cycles of 98 °C for 10 seconds, annealing at 57 °C for 30 seconds and extension at 72 °C for 30 seconds with a final extension step for 7 minute at 72 °C. The quality of DNA was checked with Agilent 2100 Bioanalyzer as a final step in library preparation before DNA sequencing. The electropherogram of Bioanalyzer showed a single peak in the range of 350 bp for amplified capture DNA [6, 58]. Finally, enriched and prepared libraries were ready to be sequenced. Sequencing was done in Århus on Denmark by Illumina HiSeq 2000 (Illumina Inc., San Diego, CA, USA). Sample preparation for DNA sequencing using SureSelected Target Enrichment System was performed in the center of biological sequence analysis, Department of systems biology, Technology University of Denmark (DTU).

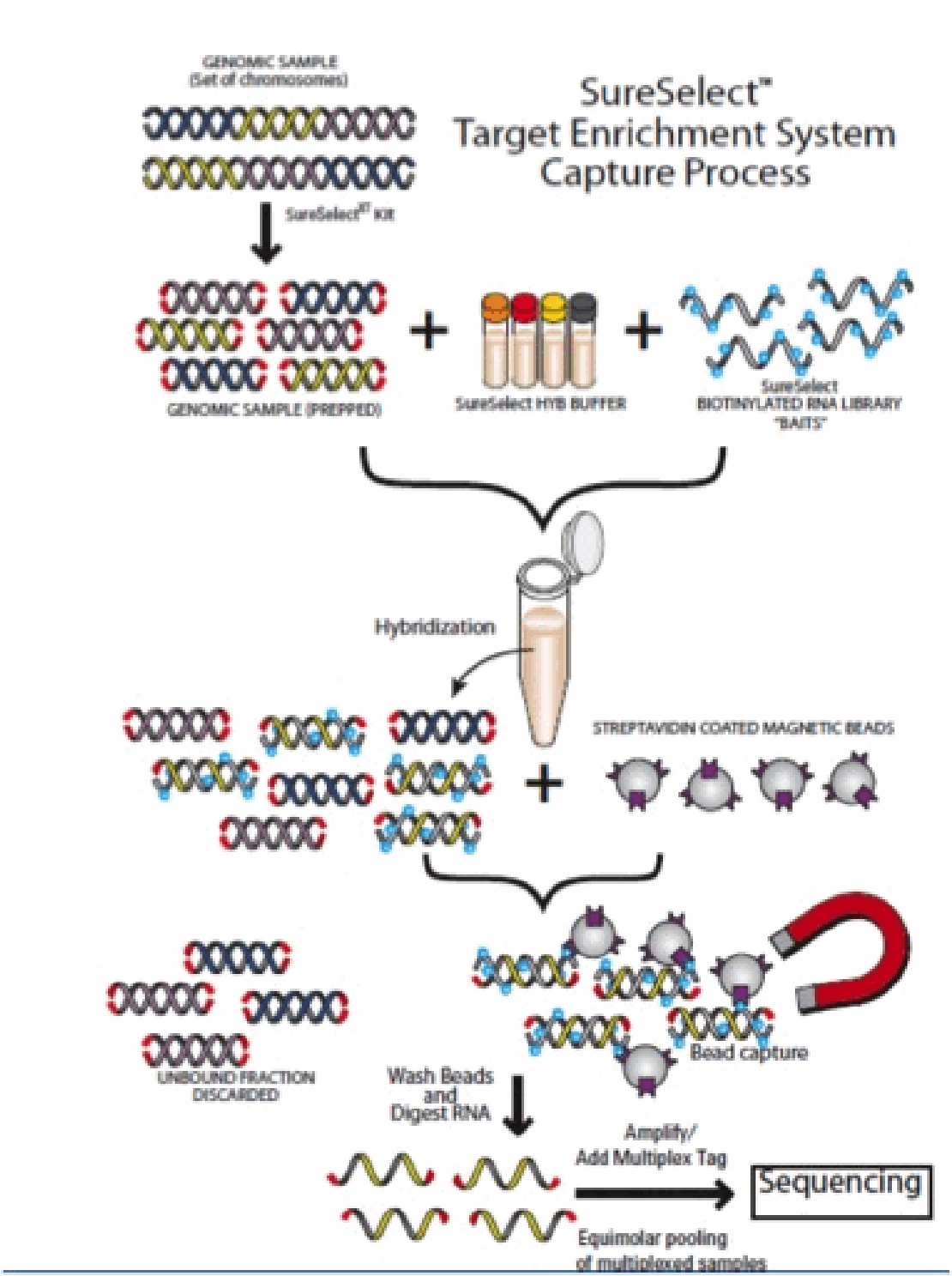


Figure 7: SureSelect Target Enrichment System Capture Process. Sample preparation steps in custom target enrichment involve DNA shearing, purification, repairing ends, ligating adapters and barcods, purification, prepared libraries amplification, quality assessment, Pooling DNA library, library hybridization and capture final quality assessment before sequencing. From Aglient Technologies [58] .

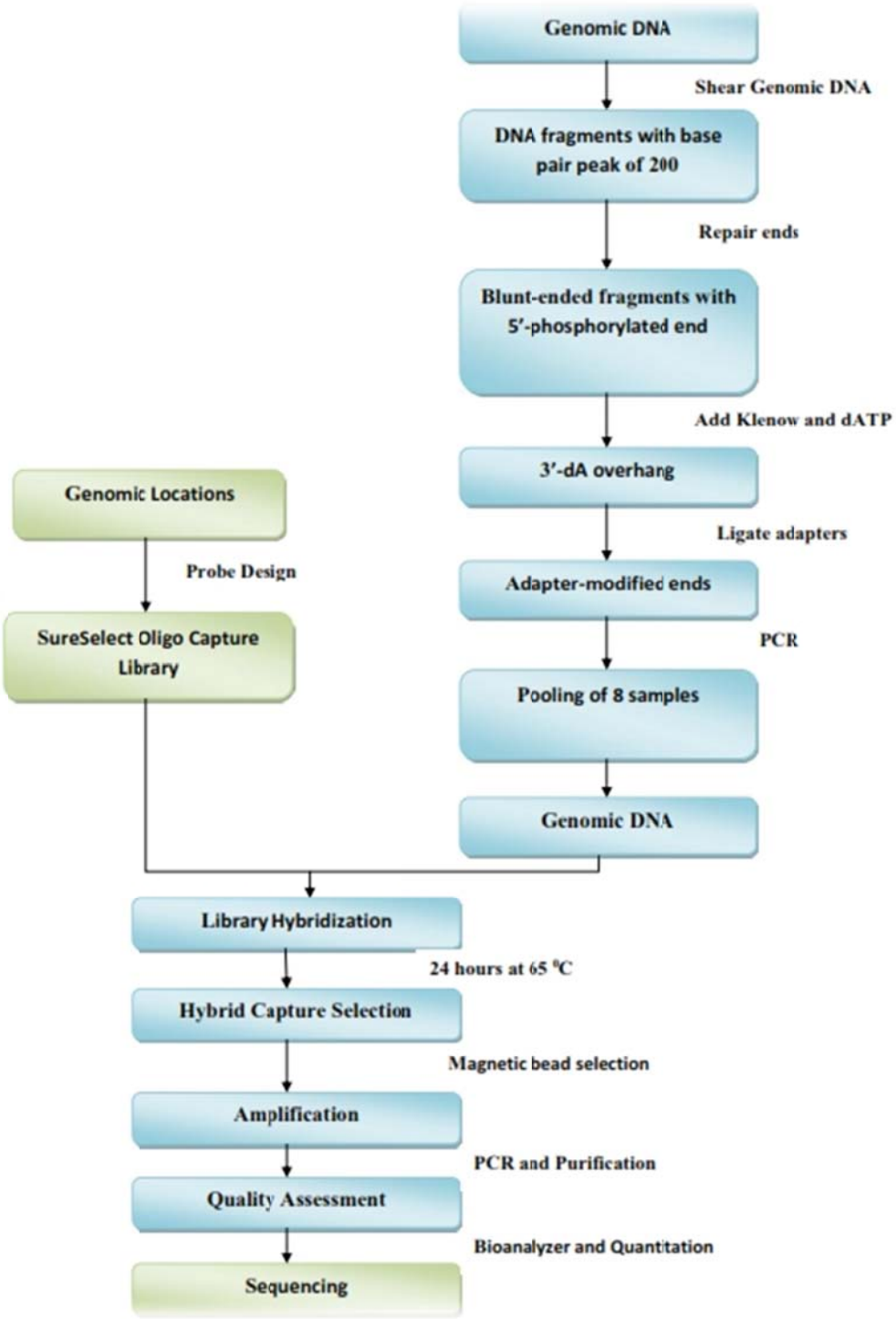


Figure 8: SureSelect Target Enrichment System workflow. From Aglient Technologies [58].

3.8 IL-8 methylation Assay

We have analyzed methylation status from blood, bone marrow smears and FFPE tissues in childhood ALL leukemia. Blood samples were taken freshly some years after treatments but smears and FFPE tissues were collected at the time of initial presentation of leukemia. Eleven blood samples of the patients without specific cancer were included as control group.

3.8.1 Bisulfite modification

Isolated DNA was subjected to sodium Bisulfite modification by using EZ DNA Methylation Gold kit (Zymo Research Corp, Irvine, CA, USA). The principle is based on different sensitivity of cytosine and 5-methycytosine against deamination through bisulphate under acidic conditions that lead to non- methylated cytosine residues which in turn are converted to uracil while 5-MeC remains unchanged as shown in Figure 9.

For bisulfite treatment, 300-500 ng of isolated DNA was added to a PCR tube, followed by the addition of 130 µL of CT conversion Reagent. The tube was vortexed and centrifuged, then incubated in a thermal cycler in the following steps: 98 °C for 10 minutes followed by 64 °C for 2.5 hours. Then samples were transferred to Zymo-Spin™ IC Column to be desulphonated and Clean up steps were performed according to manufacturer's instructions. Bisulfite treated DNA was eluted with 10 µL M-Elution buffer.

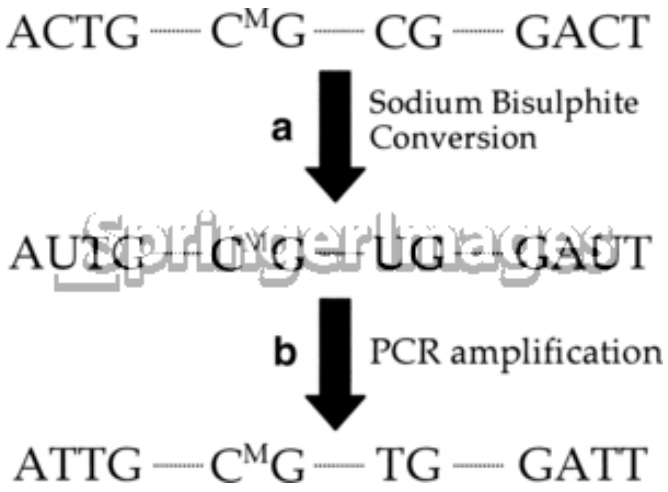


Figure 9: Schematic of the sodium bisulphite reaction. From SpringerImage [59].

3.8.2 Methylation-Specific PCR (MSP)

After Bisulfite modification, DNA was analyzed by Methylation-specific PCR. A fragment of 173bp was amplified with specific primer pairs which are presented in Table 2 (GenBank accession number M28130). The sequence of primer was specific for either methylated or unmethylated targets.

Table 2: MSP primers

Primer	**Forward (5'-3')**	**Revers (5'-3')**	Fluorescence dye
Methylated	aaaattttcgttatatttcg	tccgtaacttttttatatcat	FAM
Unmethylated	aaaatttttgttatattttg	tccaataacttttttatatcat	VIC

For each PCR setup, a mastermix consisting of the reagents listed in Table 3 was prepared. The primers were used at a concentration of 10 µM. The following parameters were used to program the Applied Biosystem 2720 Thermal Cycler: 95 °C for 5 minutes, followed by 35 cycles of 95 °C for 45 seconds, 47 °C for 45 seconds; 72 °C for 45 seconds and final elongation at 72 °C for 7min and 4 °C forever. After completion of PCR reaction, amplified pproducts were separated on ABI3730 capillary electrophoresis machine. (Applied Biosystem, HITACHI, USA). Each PCR product was mixed with 10.0 µL formamide and 0.5 µL of 500 lizTM standard before being separated in capillary electrophoresis. The PCR products, which were fluorescently labelled, separated through capillary. The laser beam caused the dyes on the fragments to be fluoresced. An optical detection device detects the fluorescence, and the software converts the fluorescence signal to digital data and was illustrated as electropherogram. The results were showed in two different color peaks represented as methylated and umethylated PCR products. The retention time defines fragment length and signal strength generates the peak height. The software generates quantities value for both height and area of the peak.

Table 3: PCR mastermix for each MSP reaction

Solution reagent	**Volume per reaction**
AmpliTaqGold 360 Buffer, 10x	2.5 µL
Magnesium Chloride, 25 mM	2 µL
dNTP mix, 10 mM	2 µL
Forward Methylated , 10 µM	1 µL
Revers Methylated, 10 µM	1 µL
Forwars Unmethylated, 10 µM	1µL
Revers Unmethylated, 10 µM	1 µL
AmpliTaq Gold 360 DNA polymerase, 5 units/µL	0.12 µL
Bisulfite -treated DNA	100 ng
Nuclease-free water	Variable

4 Results

4.1 DNA isolation

We extracted DNA from eleven bone marrow smears and nine FFPE tissues. The DNA concentration and purity of archived samples were measured by Qubit and ND-1000. The DNA concentration from all samples by ND-1000 and Qubit measurements can be seen in Table 4. The average DNA level from smears was 74.5 ± 61 ng/µL and from FFPE tissues was 103.6±73 ng/µL based on ND-1000. The smears DNA concentration averaged 39.8±45 ng/µL and from FFPE tissues averaged 5.1±3.7 ng/µL based on Qubit measurement. The range of 260/280 ratio was from 0.9 to 2.9 and it was from 0 to 2.0 concerning 260/230 ratio. ND-1000 showed higher DNA concentrations of samples compared with Qubit measurements.

Table 4 : Isolated DNA measurements by UV and Fluorescence spectroscopy

DNA Concentration of Smears				
Patient ID	**Qubit (ng/µL)**	**NanoDrop (ng/µL)**	**260/280**	**260/230**
1	25.4	54	1.8	0.9
2	10.7	29	1.9	0.8
3	21.8	69	1.7	1.4
4	53	109	1.8	1.5
5	24.5	56	1.9	1.1
6	2.3	14.3	2.8	0.7
7	150	193.4	1.8	1.8
8	100	180.5	1.8	1.7
9	20	46.2	1.8	1.0
10	10	25.7	1.6	0.7
11	20	42.4	1.9	0.7
DNA Concentration of FFPE tissues				
Patient ID	**Qubit (ng/µL)**	**NanoDrop (ng/µL)**	**260/280**	**260/230**
1	3.5	133	1.8	2.0
2	2.19	29.8	1.8	2.0
4	0.73	19.8	0.9	1.2
5	2.46	49.2	2.1	1.5
6	3.9	76.6	1.7	1.8
7	12	140.7	2.0	0.6
9	10	257.8	2.0	0.7
10	5	98.8	2.9	0
11	6	126.7	2.1	0.6

***** FFPE tissue samples were available from nine patients

It is not an accurate approach to compare DNA yields from archived samples. This inaccuracy is the result of deploying archived samples which prevent cell counting.DNA concentration of each smear sample was the amount of DNA obtained from one slide and for FFPE tissue it was obtained from 5 tissue sections with 10 μm thickness of each FFPE block.

The archived samples included in this study varied in the time of storage. They were divided in two groups to evaluate whether the samples ages would affect the quantity and quality of DNA. Group I, samples ID (1-6) was stored from 3 to 8 years, and group II, samples ID (7-11), was 1 to 2 years old of storage. Two smears in group II showed very high DNA concentration of 150 ng/ μL (7500 ng per slide) and 100 ng/ μL (500 ng per slide).

The average of DNA concentration of two groups is shown in Table 5. The difference in the mean of DNA concentration from smears was not statistically significant between group I and II (95% CI, p= 0.2; Figure10). However the comparison in the average of the DNA concentration of FFPE tissues between the two groups was statistically significant (95% CI, P=0.04; Figure11).

Table 5: DNA concentration between the two groups

Smears	**DNA concentration (ng/μL)**	**p-value**
3-8years	22.95 ± 17.25	0.2
1-2 years	60 ± 62.05	
FFPE	**DNA concentration (ng/μL)**	**p-value**
3-8years	2.5 ± 1.24	0.04
1-2 years	8.25 ± 3.30	

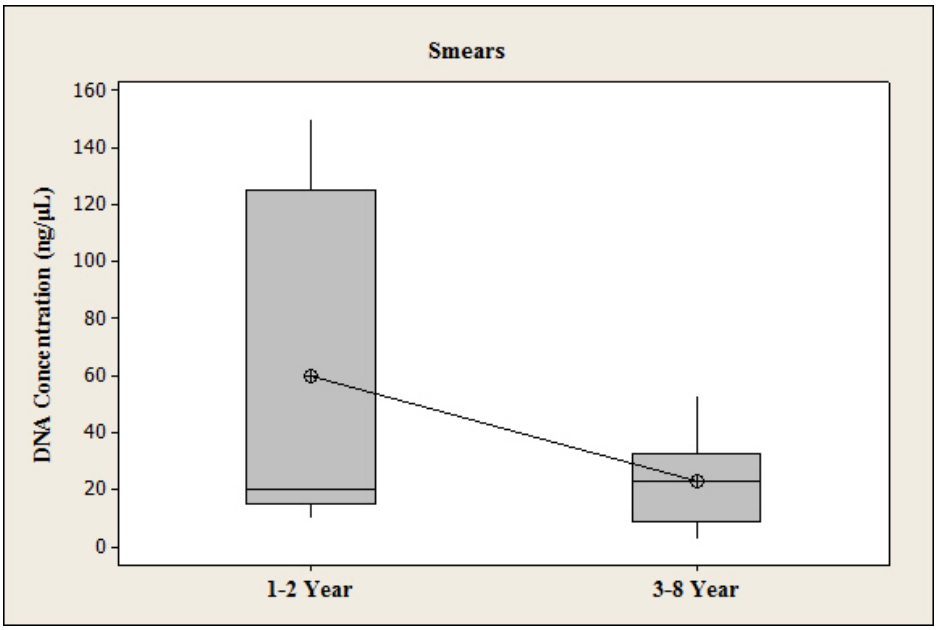


Figure 10: Difference in mean DNA concentration of smears, no statistical difference between the two groups. The horizontal bars show median value; the vertical lines represent the concentration range.

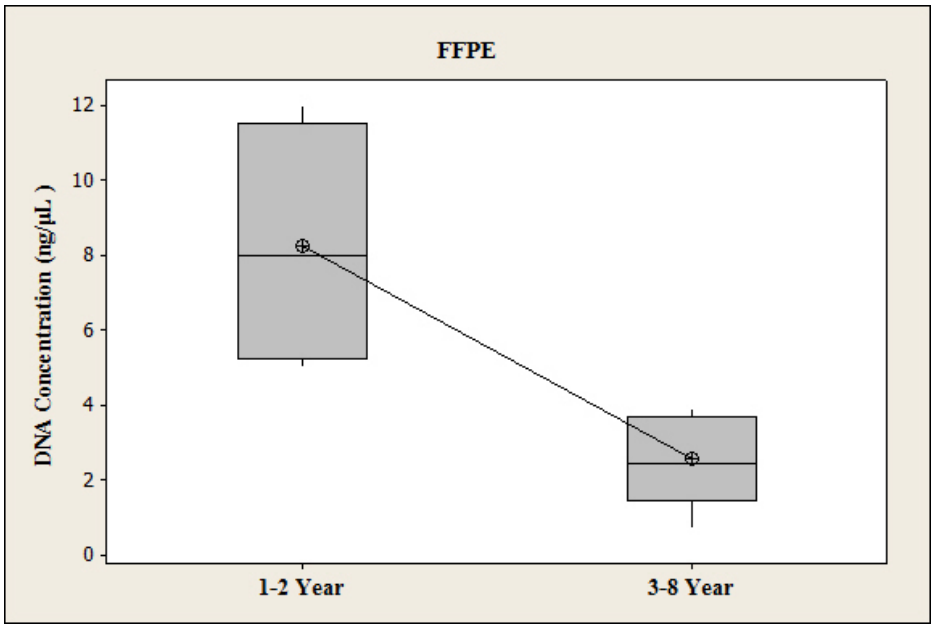


Figure 11: Difference in mean DNA concentration of FFPE tissues, there is statistical difference between the two groups. The horizontal bars show median value; the vertical lines represent the concentration range.

4.2 DNA concentration of WGA product

The DNA concentration after the WGA was measured by ND-1000 and Qubit respectively (Table 6). The average DNA yield from smears after WGA was 2512 ± 215ng/μL and from FFPE tissues was 2316 ± 513 ng/μL based on ND-1000. Based on Qubit measurements, the average DNA yield of smears was 329 ± 68 ng/μL and 388 ± 142 ng/μL from FFPE tissue. From here on, we have decided to use Qubit® dsDNA as the measurement.

Table 6: WGA yield measurements by UV and Fluorescence spectroscopy

DNA Concentration of WGA yield of Smears				
Patient ID	Qubit (ng/μL)	NanoDrop (ng/μL)	260/280	260/230
1	305	2358	1.8	2.0
2	412	2526	1.8	2.0
3	470	2743	1.8	2.0
4	320	2378	1.8	2.1
5	387	2414	1.7	2.1
6	427	2695	1.8	2.0
7	240	3008.7	1.7	2.0
8	360	2372.8	1.7	2.0
9	352	2417.5	1.7	2.0
10	320	2315.4	1.7	2.0
11	280	2405.7	1.7	2.1
DNA Concentration of WGA yield of FFPE tissues				
Patient ID	**Qubit (ng/μL)**	**NanoDrop (ng/μL)**	**260/280**	**260/230**
1	549	3133	1.8	1.4
2	333	1294	1.8	1.9
4	482	2410	1.8	1.7
5	517	2844	1.7	1.7
6	519	2406	1.6	1.5
7	218	2285	1.8	1.8
9	156	2118	1.8	1.8
10	418	2278	1.8	1.8
11	302	2077	1.8	1.8

** FFPE tissue samples were available from nine patients.

4.2.1 Differences in WGA product

Although there was no significant difference between groups I and II of smears (95% CI, P=0.4; Figure 12) in the average of WGA yield, there was a statistically significant difference between these two groups from FFPE samples (95% CI, P=0.02; Figure 13).

The WGA by using REPLI-g-FFPE kit provides amplification from DNA samples with low DNA concentration. The results from smears and FFPE tissues are illustrated in Figures 14 and 15. Based on the results of WGA yield, all smears were amplified from 1.6 to 180 fold. The highest amplification yield is concerned with the sample number 6 and the lowest one was related to sample number 7. Although the sample number 6 had the lowest DNA concentration (2.3 ng), it showed the maximum fold of amplification, i.e. 180 fold (417 ng). Sample number 7 followed the same trend in opposite direction. It means the sample number 7 had highest DNA concentration (150 ng), it amplified just 1.6 fold (240 ng). Samples with low DNA concentration seem to generate great amount of WGA product.

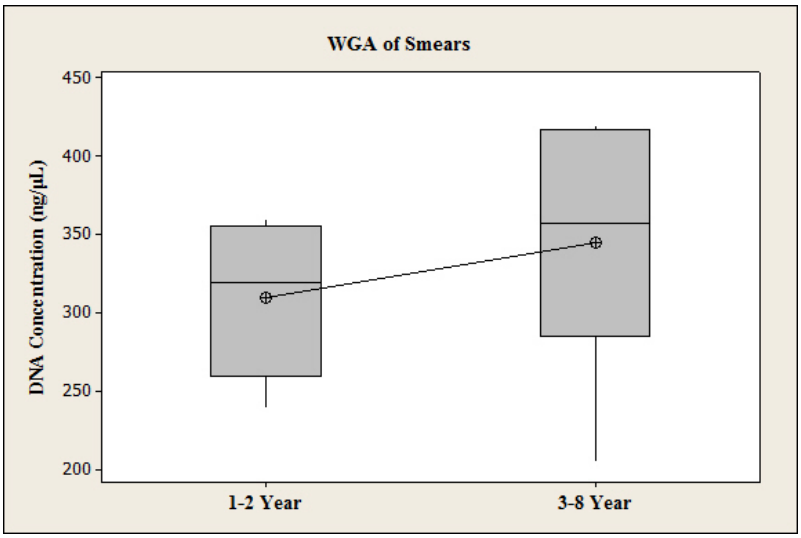


Figure 12: Difference in average of WGA product of smears, there is a statistical difference between two groups. The horizontal bars show median value; the vertical lines represent the concentration range.

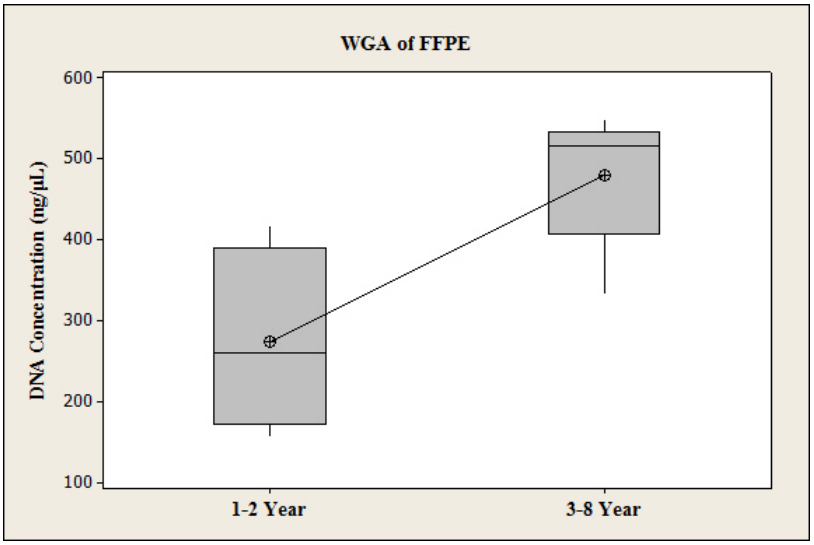


Figure 13: Difference in average of WGA product of FFPE tissues, there is a statistical difference between two groups. The horizontal bars show median value; the vertical lines represent the concentration range.

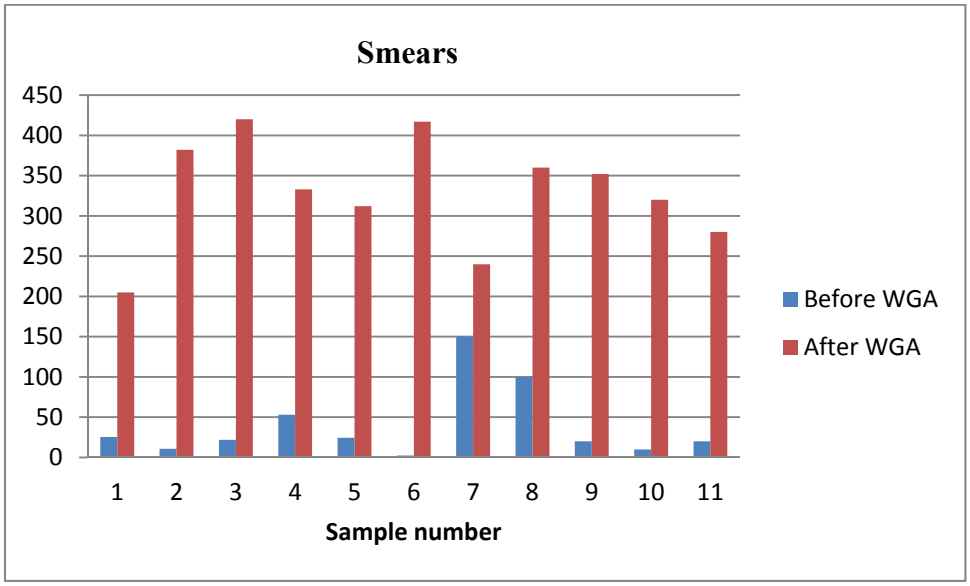


Figure 14: DNA concentration (ng/µL) of smears before and after WGA

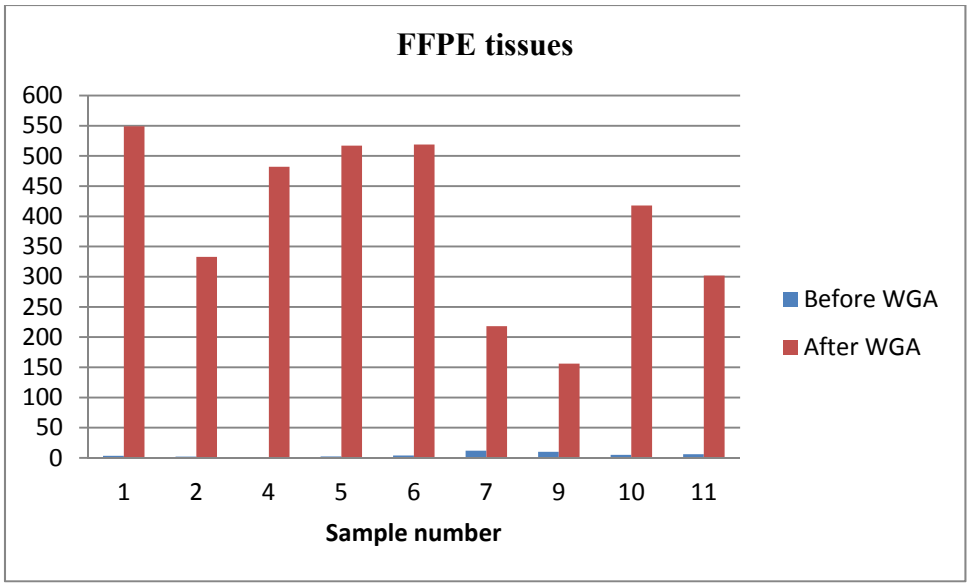


Figure 15: DNA concentration (ng/µL) of FFPE tissues before and after WGA

The DNA yield from WGA amplification indicates that FFPE tissues amplified from 15 to 660 fold. As for the smears, the highest degree of amplification was in sample number 4 with the lowest DNA concentration (0.73 ng). The other samples with low DNA concentration were numbers 5, 2 and 1 with amplified 210, 156 and 152 fold. The lowest amplification yield was seen in the sample with the highest DNA concentration. DNA concentration of sample 7 was 10 ng which amplified just 15 fold. Although correlation between DNA concentration and WGA yield is difficult to estimate because WGA is depend on the quality of the DNA.

4.3 Purification of WGA product

Purification was carried out for WGA products of samples in group I. Determination of DNA concentration showed substantial decrease in DNA concentration as shown in Figure16 and 17. Based on REPLI-g kit, purification of WGA products is not required for Illumina platform. In order to achieve accurate quantification, WGA products should be purified before measuring so that residual primers and protein components could be removed. The concentration of WGA products after purification of smears was reduced in the range of 2.8 to 9.8 fold and those from FFPE tissues were reduced in the range of 3.5 to 8.4 fold.

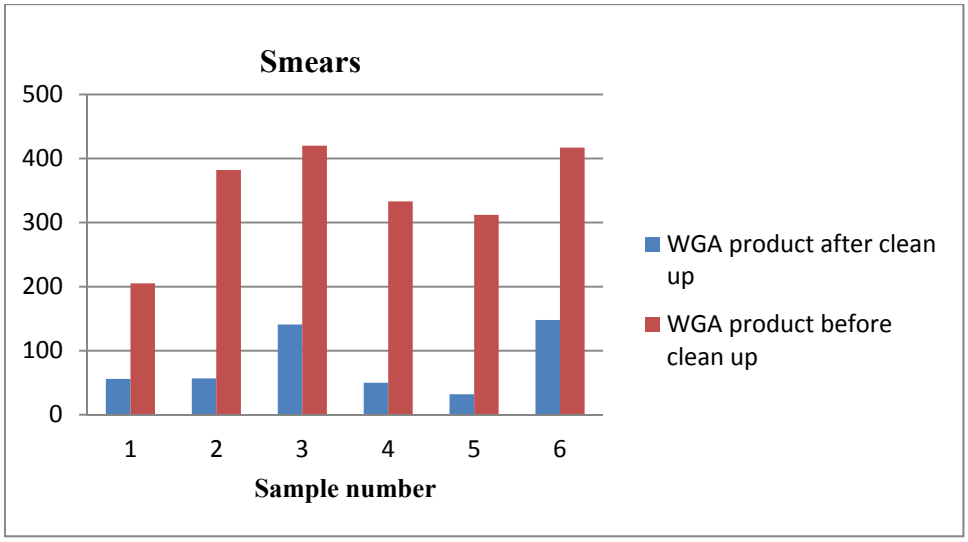


Figure 16: DNA concentration (ng/ µL) of WGA product of smears before and after purification.

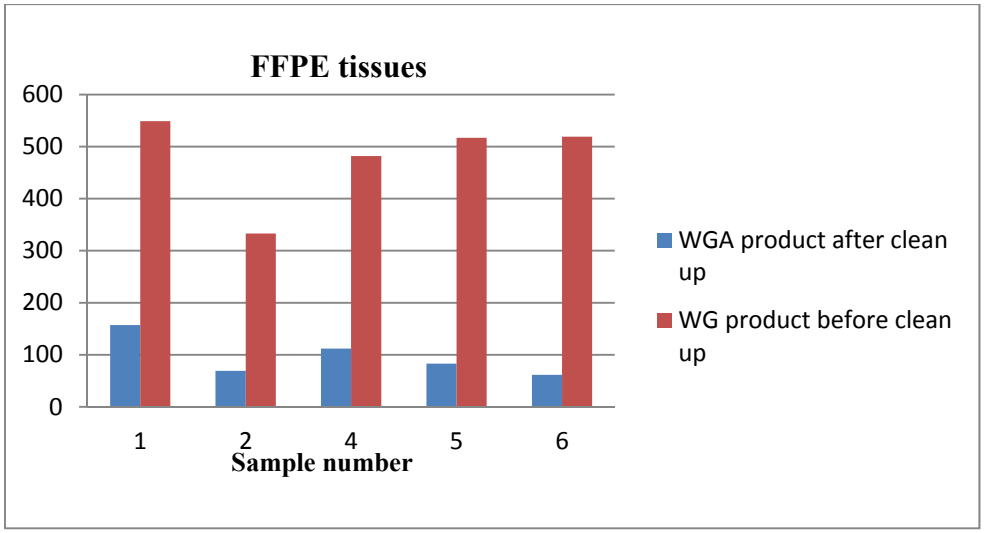


Figure 17: DNA concentration (ng/µL) of WGA product of FFPE before and after purification.

4.4 Gel electrophoresis analysis

An image of electrophoresis of FFPE tissues is shown in Figure 18. The same amount of DNA (250ng) after extraction, amplification using REPLI-g FFPE and clean up of WGA samples were separated on agarose gel. The Agarose gel analysis indicated that FFPE tissues made a faint smear consistent with poor or degraded DNA. It might be difficult to see in figure of gel. Samples after amplification showed DNA bands with size around 23kb which indicates that WGA amplification has occurred. We see the smear of DNA with sizes above and below 23kb. Substantial DNA amount decreased in the lanes after WGA clean up.

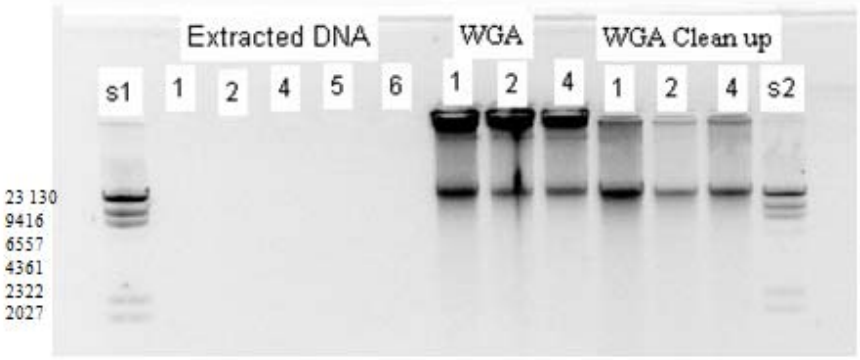


Figure 18: Agarose gel electrophoresis of FFPE tissues. S1 and S2: λ Hind III ladder. Lanes 1-6; after DNA extraction, sample 1, 2, 4 after WGA and sample 1, 2 and 4 after cleanup of WGA.

4.5 DNA profile analysis

The quality of PCR amplifiable DNA from different sources of samples was assessed by standard analyses using STR markers. The diagram in Figure 19 shows the percentage of the samples that amplified markers with different lengths.

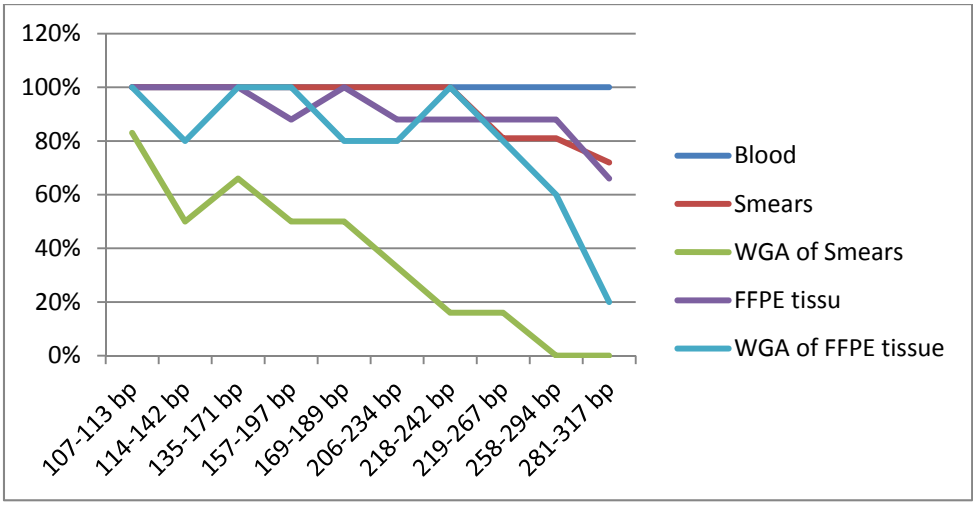


Figure 19: The samples ability of amplification of STR markers based on the size of the markers (Group I and II).

Applied Biosystem suggested a peak- height threshold of 150 RFU, but we calculated the peaks under this threshold since they indicate correct length size compared with allelic ladder [60]. See Appendix A for full results of STR markers amplification of the samples.

Based on the results concerning smears, the amplification failed in the marker with the largest size CSF1P0 (281-317 bp) in three cases and for FGA (219-267 bp) and D7S820 (206-234 bp) in two cases. The amplification was successful for the rest of the markers with the range of 100 to 242 bp. The number of the markers that are amplified through smears in group I is illustrated separately in Figure 20. The result shows that there is a progressive drop out in amplification of large marker size in this group. STR markers analyzed from FFPE tissue samples failed for the marker with the largest amplicone size, CSF1P0 (281-317 bp) in three cases. Eight out of nine cases were able to amplify the following markers: D7S820 (258-294 bp), FGA (219-267 bp), TPOX (218-242 bp), D13S317 (206-234 bp) and Vwa (157-197 bp). The rest of the markers were successfully amplified in all cases.

STR analyzing of WGA product was done just for samples in group I (3-8 years old of storage) which included six smears and five FFPE tissues. There were no amplification products from the markers with large amplicon size, CSF1P0 (281-317 bp) and D7S820 (258-294 bp) from the WGA products. Concerning WGA product of FFPE tissues, one sample was

able to amplify the largest marker CSF1P0 (281-317 bp) but with small amounts of PCR-product. See Appendix B for partial and full genetic profiles.

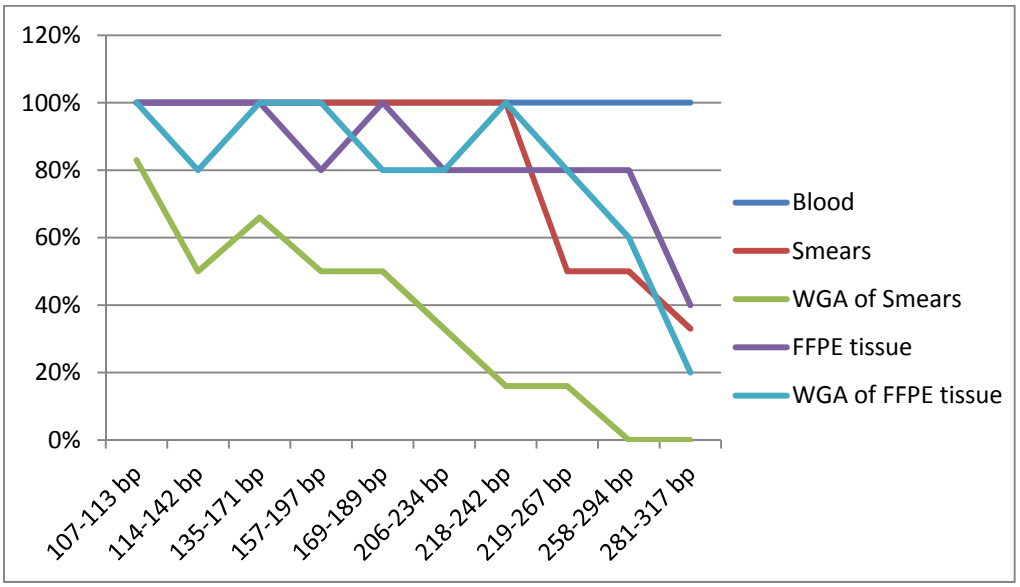


Figure 20: The samples ability of amplification of STR markers based on the size of the markers (Group I).

Most of the smears and FFPE tissues in group I showed partial genetic profile which indicate that little PCR- products were amplified from these samples as shown in Figure 21.

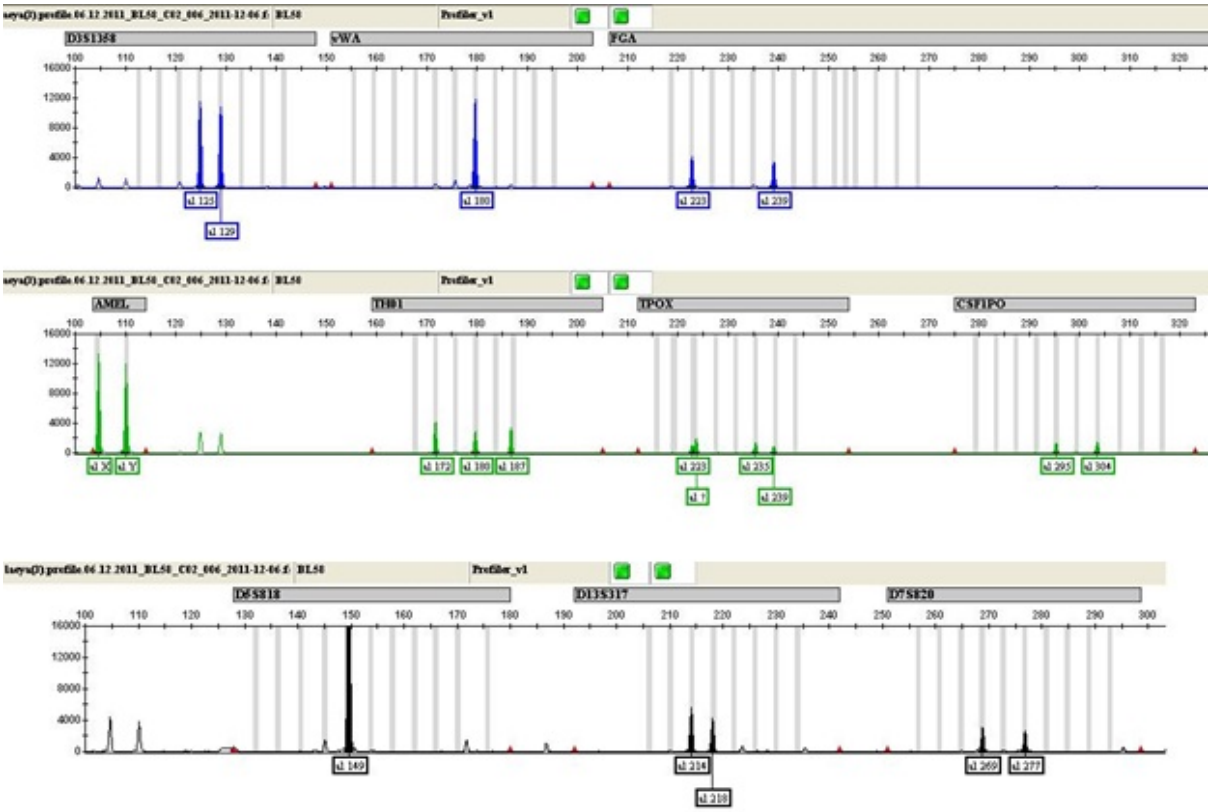


Figure 21: Partial genetic profile from smear of sample number 3 (group I).

All samples in group II were able to amplify all ten markers successfully (100%) in DNA profile analyzing, and all showed full genetic profile with high height of peaks as shown in Figure 22.

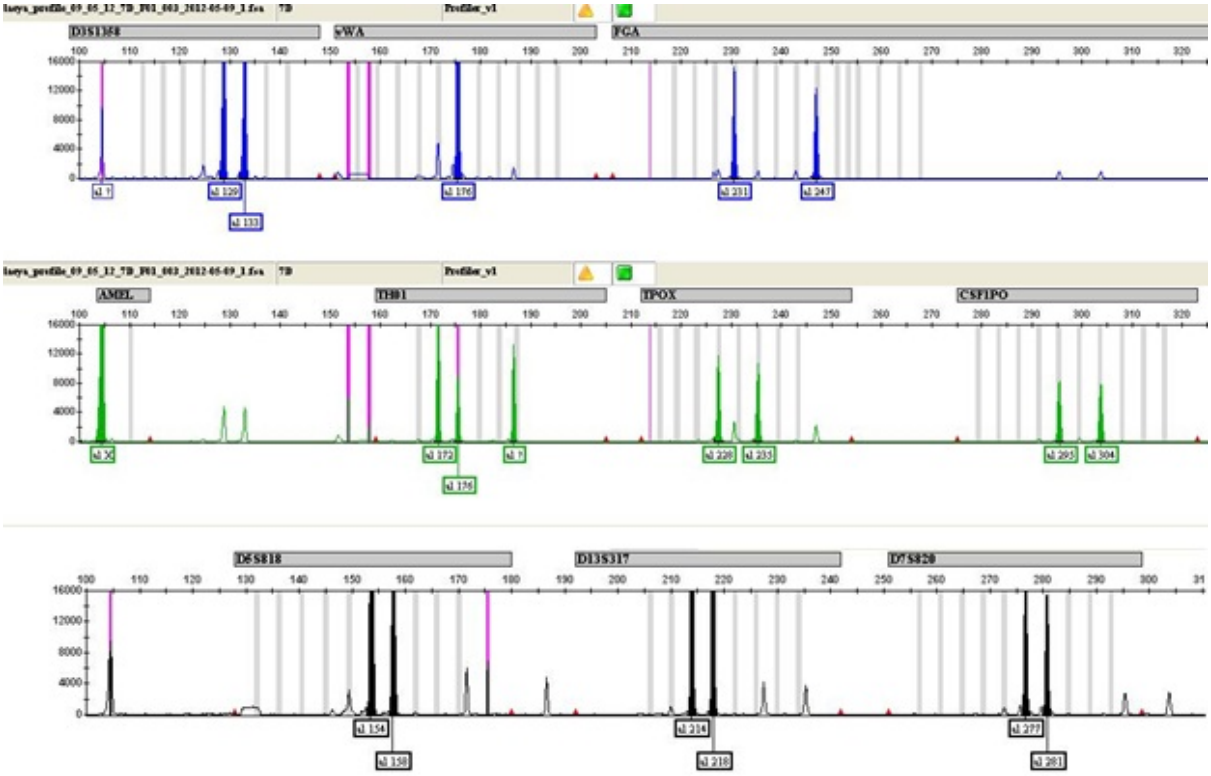


Figure 22: Full genetic profile from smear of sample number 7 (group II).

4.6 Multiple SNP Sequencing

In this experiment, we performed multiple SNPs analyses on DNA derived from all samples in group I and bone marrow smears in group II, the FFPE tissues from group II were not available at that time. To evaluate the quality of the SNPs profiling, archived samples were compared with fresh blood samples from patients which were taken some years after treatment. Both genomic and amplified forms of some samples were tested in this experiment.

SNP call rates with coverage depths of 1x, 4 x, 10x and 20x were analyzed in all samples. After strict quality control of the criteria (according to the consulting bioinformatician), acceptable call rates were achieved from 5 out of 35 smears. All five FFPE samples failed (data not shown). The average SNP call rate of fresh blood samples was 91% and of the matching archived smears was 74% with 4x sequencing depth (Table 7). As expected, with increasing sequencing depth, the number of compared SNPs decreased. Sample number 7 achieved higher call rate from genomic DNA than amplified DNA. The samples achieved

concordance above 85% at 10x sequencing depth. The average concordance between fresh blood samples and archived bone marrow smears was 89%.

Table 7: Performance of multiple SNPs sequencing

Sample ID	Sample type[a]	SNP call rate[b]				Concordance[c]
		1x	4x	10x	20x	
1	Smear. amplified	88%	70%	52%	76%	89%
1	Blood	98%	92%	81%	68%	
7	Smear. amplified	89%	68%	46%	26%	85%
7	Blood	98%	90%	78%	63%	
7	Smear. genomic	97%	90%	79%	65%	88%
7	Blood	98%	90%	78%	63%	
11	Smear. amplified	80%	70%	51%	31%	87%
11	Blood	98%	91%	80%	67%	
4	Smear. genomic	95%	81%	64%	43%	96%
4	Blood	98%	93%	83%	70%	

[a]Type of samples included in SNP analyzing, blood, bone marrow smear and smear after whole genome amplification after purification.

[b]Rate of successful genotype identification.

[c]Percentage of genotype calls concordant between the matching fresh and bone marrow smears.

4.7 Methylation Specific PCR Analysis

The results obtained from capillary electrophoresis visualized in two different color peaks represented as methylated and umethylated PCR products as shown in Figures 23 and 24. Analyzing based on area is more reliable because the two peaks might have same heights but different areas. The results were analyzed based on two parameters, the heights and the area of the peaks; however, our findings were almost the same. Unmethylated ratio in blood samples of patients was 88% and 83 % in the control group as shown in Table 8-9.

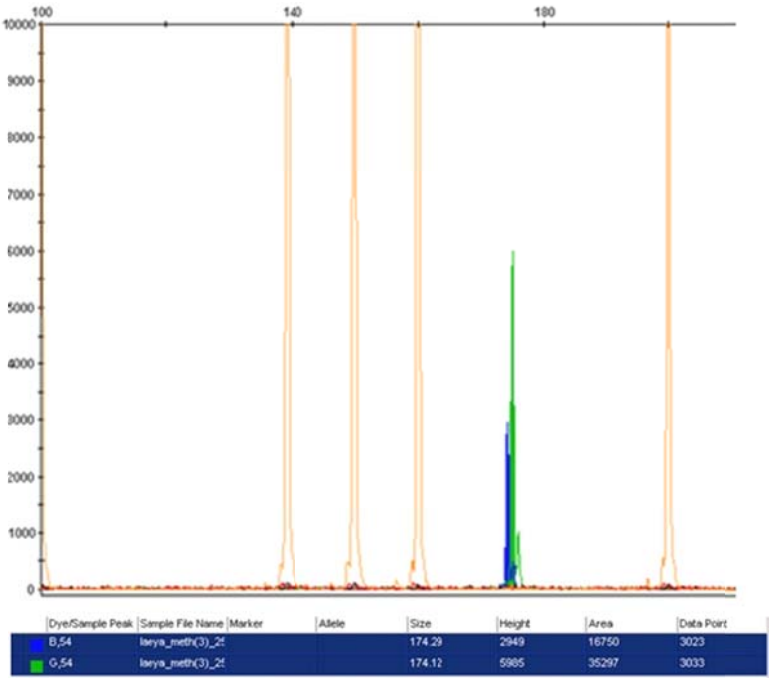


Figure 23: Methylated and unmethylated status of IL-8 in sample number 5 of control group. The blue peak and green peak illustrate methylated and unmethylated status, respectively.

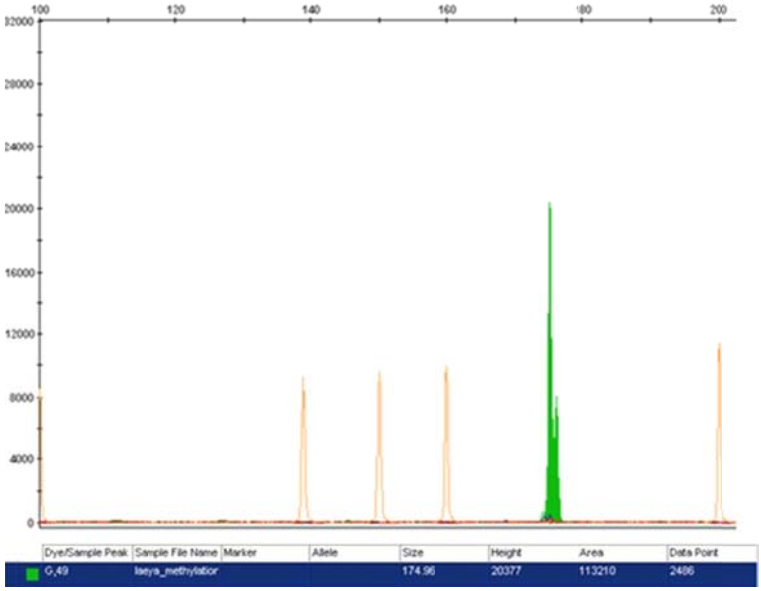


Figure 24: Only unmethylated status of IL-8 in patient number 8 of smear sample. The green peak illustrates the unmethlyted PCR product.

Table 8 : Methylated and unmethylated status of blood samples in the patient group

	Methylated		Unmethylated		Unmethylated%	
Sample	**Height**	**Area**	**Height**	**Area**	**Height**	**Area**
1	943	6186	20679	117080	96%	95%
2**	6965	39792	13925	81349	67%	67%
3*	629	4530	17685	100037	97%	96%
4*	0	0	13760	78165	100%	100%
5*	836	4890	6235	35479	88%	88%
6*	0	0	1907	11033	100%	100%
7*	1055	6626	10981	62589	91%	90%
8*	0	0	19535	114965	100%	100%
9*	0	0	1533	9283	100%	100%
10**	4717	26675	14333	83293	75%	76%
11*	6516	37496	6687	38437	51%	51%

*PCR product was diluted 1/10 in these samples.

**PCR product was diluted 1/100 for these samples.

- Patients number 7 and 8 were under treatment.

Table 9 : Methylated and unmethylated status of blood samples in the control group

Sample	Methylated		Unmethylated		Unmethylated%	
	Height	**Area**	**Height**	**Area**	**Height**	**Area**
1*	1965	11835	8536	50194	81%	81%
2*	581	3777	17825	100197	97%	96%
3*	0	0	6787	40029	100%	100%
4*	11601	68317	22769	133063	66%	66%
5*	2949	16750	5958	35297	67%	68%
6*	3208	18360	10068	57362	76%	76%
7*	1887	11241	8778	51090	82%	82%
8*	4653	27110	23594	138719	84%	84%
9*	1991	12187	13279	77668	87%	86%
10**	2409	14219	11183	64851	82%	82%
11*	1093	6658	15972	91029	94%	93%

*PCR product was diluted 1/10 in these samples.

**PCR product was diluted 1/100 for these samples.

Methylation status of IL-8 was analyzed from ten smears and seven FFPE tissues, as well; result can be observed in Table 10. Unmethylated status was detected in almost all of the smears samples. According to the results, IL-8 was almost completely unmethylated in all FFPE tissue samples except for sample number 4.

Table 10 : Methylated and unmethylated status of Smears and FFPE samples

Smears						
Sample	**Methylated**		**Unmethylated**		**Unmethylated%**	
	Height	**Area**	**Height**	**Area**	**Height**	**Area**
2	0	0	4749	27795	100%	100%
3	0	0	11746	69012	100%	100%
4	0	0	14625	90828	100%	100%
5	0	0	2595	14230	100%	100%
6	0	0	9948	56186	100%	100%
7*	393	2574	4442	26345	92%	91%
8	438	3442	20377	113210	98%	97%
9	0	0	2564	14014	100%	100%
10*	477	3210	14276	82738	97%	96%
11	0	0	2311	13110	100%	100%
FFPE Tissue						
Sample	**Methylated**		**Unmethylated**		**Unmethylated%**	
	Height	**Area**	**Height**	**Area**	**Height**	**Area**
1	0	0	634	4281	100%	100%
4*	1468	8874	6283	37075	81%	81%
5	112	884	6889	42446	98%	98%
6	0	0	7050	39175	100%	100%
7*	1159	8094	27853	158799	96%	95%
9	96	653	10695	70126	99%	99%
11	0	0	1454	8952	100%	100%

*PCR product was diluted 1/10 in these samples.

- Samples were not available from samples 1 of smears and samples 2, 3 and 8 from FFPE tissues.

5 Discussion

5.1 DNA isolation

The amount of the recovered DNA is a limiting factor in using archived samples. Although the range of analyses of nucleic acid from archived biological samples has increased, there is no standard operating protocol for nucleic acid isolation.

Ludyga et al. [61] compared two methods i.e. phenol-chloroform isoamyl alcohol and Qiagen kit (with some modification), for DNA isolation from FFPE tissues. They reported that higher yields were achieved from the phenol-chloroform protocol compared with the Qiagen kit. The reasons of the mentioned difference were explained as loss of DNA in the silica with the column-based method and during several processing steps. Although both methods showed DNA with high purity, they found longer PCR fragments from DNA isolated with phenol- chloroform method [61].

In the other study, Wei et al. [62] compared three methods for DNA isolation of FFPE tissues. They reported that both phenol-chloroform and simple boiling methods were more efficient for PCR amplification of the β-globin gene (256 bp) than the DNA Mini kit (Qiagen). However, Gilbert et al. [63] found that DNA micro kit is more effective than Tris-buffered proteinase k regarding DNA extraction from FFPE tissues.

5.1.1 DNA concentration based on ND-1000 and Qubit measurements

The accurate measurement of DNA concentration is significant in performing high throughput genotyping and sequencing successfully. The quantification of DNA yield from archived samples can be challenging due to low level of isolated DNA and potential contamination extracted together with the DNA. Both assays required similar amount of samples; the ND-1000 used 1.5 µL of samples and Qubit used 1 µL or 2 µL when the DNA concentration is low. The main difference between the two assays is the sensitivity. The lower limit of ND-1000 is 2 ng/µL, whereas the Qubit is sensitive to 100 pg/µL [56-57].
Our results showed that, ND-1000 spectrophotometer overestimated DNA concentration compared to Qubit measurement. UV absorbance is not an accurate measurement of DNA concentration because of varying amount of contamination with other molecules; it cannot

distinguish between DNA, RNA or free nucleic acids. All UV absorbance at 260 nm in the sample is calculated as DNA concentration. Also, the overestimation by ND-1000 might be explained by the presence of degraded DNA sample, because single stranded DNA absorbs 20-30 % more UV light at 260 nm than double stranded DNA [64-65]. The other issue that should be considered is co-purification of RNA with the DNA. We did not use RNAase for DNA isolation by QIAamp FFPE tissue kit. Presence of RNA will influence the ND-1000 measurements. The presence of RNA can also inhibit some enzymatic downstream applications, and the influence of RNA in each sample can still be different from sample to sample[55].

The 260/230 ratio was low for most of the samples. This indicates the presence of contamination such as chaotropic salts. The presence of chaotropic salts has potential for inhibition of WGA and PCR and often results in overestimation of the DNA concentration at 260 nm. Some samples had low 260/280 ratio which indicate contamination with protein or other components that are absorbed at 280 nm. Although purity ratio and spectral profile are important indicators of samples quality, the best indicator of DNA quality is functionality in downstream application.

Based on our results, there is a clear discrepancy, at least 2-fold, between Qubit and ND-1000 measurement of DNA concentration. There is usually not an agreement between the two methods, even with pure DNA samples. Holden et al. [65] has reported that PicoGreen measurement usually show lower DNA concentration than UV absorbance. They investigated the lack of correlation between these two measurements.

Qubit measurement is based on the fluorescence enhancement of the fluorescent dye upon binding to dsDNA. Therefore, protein and RNA cannot interfere with the obtained results. Regarding Qubit, we should consider how fragmented DNA affects the measurements. Holden et al. [65] reported that the PicoGreen measurement was dependent on the size of DNA fragment but only if the DNA were in pure water, as it would be less sensitive in buffer.

In the other study, Georgiou et al.[66] proposed that quantification of DNA is dependent on the degree of fragmented DNA, by using Invitrogen protocol with two fluorescent dyes Hoechst and PicoGreen.They found UV spectroscopy is independent of degree of

fragmentation.Georgiou et al.[66] in another study evaluated effects of DNA fragmentation on accurate measurements of DNA concentration. They reported that UV absorbance measures intact DNA and totally fragmented DNA at 260 nm equally. Only 30% of the concentration of the intact form is measured in fragmented DNA by Hoechst- and PicoGreen-based assays.

These studies are in contrast with some other relevant research in the literature which indicates that the size of dsDNA does not have any effects on DNA-dye complexes [67-68].
In communication with Life Technologies (Invitrogen), they claim that quantification of DNA by Qubit is not influenced by DNA size.

Haque et al. [69] evaluated the performance of three methods for DNA quantification of high DNA quality samples from a lymphoblastoid cell line. They found UV spectroscopy to be a precise DNA quantification method compared to two fluorometric methods, PicoGreen and novel real-time quantification genomic PCR assay.

DNA concentration differences between the two groups

The average of DNA yield obtained from smears and FFPE tissues of group II (1-2 years old) were higher than in group I (3-8 years old). Although this difference was not statistically significant in smears, it was considerable in FFPE tissues. This difference could be the result of damages to DNA which was exposed a longer time to formalin.

The interaction of fixative with chromatin proteins leads to the loss of DNA yield which was reported by Vince et al. [23]. Insufficient DNA yield from bone marrow smears could be explained by residual staining interference in cell destruction which prevents DNA to be liberated from the cell. Our results are consistent with a study that has been conducted by Ludyga et al. [61]. They evaluated the isolation of DNA from FFPE breast and colon cancer tissues between which are 10 to 40 years old. They observed that the age and origin of FFPE tissues influenced the DNA yield.

Consequently, to achieve optimal quality and quantity of extracted DNA, a method should be optimized before isolation of DNA from archived materials. An accurate and sensitive method for quantification also is required. The presence of impurities or small amounts of DNA in the sample may lead to inaccurate analyzing of DNA in downstream applications.

5.2 Whole genome amplification efficiency

Genotyping of DNA samples with limited quantity is possible by WGA and this capability results in increasing the number of samples for genetic analysis studies. Limited quantities of isolated DNA and its fragmented nature are two major problems in molecular analysis of archived samples. The REPLI-g FFPE kit provides amplification of this precious samples to overcome the shortages of initial material in downstream analyses.

Based on Qubit measurement smears and FFPE tissues showed 1.6 to 180 and 15 to 660 amplification fold, respectively. We found high yield of DNA concentration after WGA but most of the amplified samples failed in multiple SNPs and STR marker analyzing.

WGA product quantification based on ND-1000 could incorporate some inaccuracies, because it could measure the residual nucleotides and unused primers that influence the DNA concentration. For more accurate quantification of dsDNA, the Qubit® dsDNA BR Assay kit was used. Although it is highly selective for dsDNA, primer dimers can be measured in highly degraded DNA templates when the amplification reaction has not occurred properly.

The main reason of this failure could be explained by inability of these two physical methods for quantification of WGA yield. Because WGA generates various amount of nucleic acid side products which could be interfered by DNA quantification, both accuracy and further studies may severely be affected. These artefactual side products are detected by UV and fluorescent spectroscopy [70-71]. Thus, we can conclude that the number of amplification fold measured by these methods could not be accurate.

Hansen et al.[71] reported that quantitative PCR of human-specific *Alu* yd6 was an accurate method for the evaluation of suitability of WGA products for high-throughput sequencing in comparison to UV and fluorescent spectroscopy.

Therefore, it is difficult to investigate whether the REPLI-g FFPE was able to successfully amplify isolated DNA from archived samples. The efficiency of a WGA reaction depends on the quality and the number of genomic equivalents of template DNA. The yield of DNA after WGA is strongly dependent on the quality of the template DNA. This is critical especially in archived samples since the fixation with formalin leads to fragmentation of the nucleic acids and is depending on the incubation time and storage conditions, which can dramatically impact the quality of DNA.

In general, WGA based MDA could be able to amplify degraded DNA samples, but not highly fragmented samples. It is presumably because highly degraded fragments are too short to permit random hexamer primers binding and therefore limited primer binding sites are available. Furthermore, if primers bind to the middle of DNA fragments and then are extended by DNA polymerase, each strand displacement could make a short template, and limited binding of other primers so that extension cannot continue any further. This would lead to formation of primer-primer dimers and primer concatemers due to lack of amplifiable template[72]. According to the REPLI-g FFPE kit "is not suitable for using DNA fragments less than 500 bp in length or small number of genome equivalent less than 500" [36].In the other study, Gunn et al.[73] examined the efficacy of WGA on isolated DNA from low-yield samples. They reported low success rate in genotyping of four microsatellite loci from WGA of hair samples in comparison to fresh tissue.

An accurate quantification of WGA yield could be achieved by real-time PCR method that specifically amplifies human DNA sequence [36, 71]. Also sufficient quality of template DNA is required for successful WGA; it could be a helpful method to increase DNA quantity when the initial template has high DNA quality.

5.3 Evaluation of isolated DNA quality

Gel electrophoresis

Although currently there isn't any simple method to recognize the amount of cross-linking within a sample, valuable hints about the sample quality could be perceived from gel electrophoresis of isolated DNA. Gel electrophoresis provides additional confirmation about low quality of isolated DNA from archived samples. Because of limited amounts of DNA, there is no DNA visible after extraction on the gel picture. DNA remained in the well after WGA might be DNA-protein complexes which were removed after purification. The band around 23kb might be mitochondrial DNA or repetitive DNA.

DNA profiling

Here we used a novel application of routinely used PCR-based STR analyzing to assess DNA quality in multiple regions throughout the genome. All biopsies and smears in group II were able to amplify all ten markers successfully (100%) with almost full genetic profile, while most of the samples in group I had partial genetic profile (low height of peaks) for large marker size (larger than 200 bp). Small amount of PCR products lead to weak signals could be due to either degradation of the template DNA or the PCR inhibitors exists in the sample. Therefore, WGA was performed only for group I. It is evident from the results that only 36% of smears with WGA had the ability to amplify all markers, while 83% of the smears without WGA did the same. Samples without WGA have more ability in markers amplification compared to those samples with WGA. Just one sample of FFPE tissue showed better result from WGA than without WGA. The reason for this failure is explained in the following paragraph.

Based on our findings, we might conclude that applying REPLI-g FFPE ligated DNA fragments but not necessarily in the position which is needed to amplify intended markers.
The other issue that should be taken into consideration is the effect of several factors on the performance of downstream assays like PCR and WGA. An important factor could be the copy number of the DNA template. The performance of WGA is proportional to the input amount of DNA in the WGA reaction. The high amounts of DNA template and therefore the larger copy number of the genome, led to successful REPLI-g FFPE amplification [74]. Determination of DNA concentration from archived samples can be challenging (explained in

DNA concentration part). Fragmented DNA should contain multiple copy of each locus. Therefore, to ensure exact locus existence, the initial amount of DNA template should be increased [36]. The degree of cross-linking within a sample is the other significant factor that affects the performance of amplification reaction. The more cross-links inside DNA, the lower performance of amplification reaction [74].

Another advantage of using DNA profiling is the detection of any contamination that might happen during preparation of archived samples. We were able to match smears and FFPE tissues with blood samples as a reference in DNA profile analysis. Comparing DNA profile of three types of samples confirmed that no contamination exists. As a result, it could be concluded that the sample belongs to the correct person.

Bablo-Pokora et al. [75] found partial genetic profile from all FFPE tissue samples and were not able to match with reference samples based on marker analyzing. In the other study, Thomas Gillbert et al. [63] assayed the quality of DNA using multiplex PCR Minisequencing (MPMS) method. In this study 44 autosomal unlinked SNPs were amplified and resulted in PCR products between 19 and 115 bp. The percentages of SNPs that successfully amplified correlated to the quality of DNA.

Our study showed significant difference between quality and quantity of the two archived group I and II. While samples within group II had suitable DNA concentration as well as high quality, group I samples showed low result in both parameters. Our study is in accordance with the study of Ludyga et al. [61] which reported that DNA fragmentation was associated with storage time of samples; the older samples showed shorter fragments. While in the other study, Thomas et al. [22] has reported that neither the storage time nor staining of bone marrow slides affected PCR microsatellite typing even after long period storage.

5.4 Multiple SNPs analysis

As mentioned earlier, we included all smears and only FFPE tissues from group I for SNPs analyzing. Most of the FFPE samples failed and showed an unacceptable call rate. The reason could be low quality of isolated DNA extracted from archived samples especially from group I. Even WGA by REPLI-g FFPE kit was not efficient for those highly degraded samples.

Some smears especially from group II, led to high signal and call rate detection even from genomic samples without WGA amplification. Thompson et al. [76] used FFPE ovarian tumor tissues for whole genome SNPs in the Affymetrix 10 k mapping array. They reported the average call rate for fresh samples as 89% and for FFPE tissues as 83%.

There are different failure causes in our experiment which are included within the following lines. The main reason might be highly degraded materials which lead to low number of intact copies for each interested gene. To prevent failed or poor genotyping result, the quality control of whole genome amplified DNA is needed before genotyping. In addition, we used fewer bait than the recommended amount by the manufacturer. We were advised to do so by some of the molecular biologists at the lab (DTU).

5.5 Methylation analysis

All blood samples except sample 7 and 8 were taken a couple of years after patients treatment. So, blood samples could not reliably be compared to the control group in methylation status. The comparison will be reliable when the samples are taken during the diagnostic period. All of the smears and FFPE tissue samples in this study were taken at the diagnosis time of leukemia. The results indicated that the rate of hypomethylation status of IL-8 gene in most of the smears and FFPE tissues were 98% and 96%, respectively.

Chiaretti et al. [49] conducted a study to compare the gene expression profile between refractory patients and those who responded to induction chemotherapy of adult T-cell acute lymphoblastic leukemia. They identified high expression of IL-8 in refractory T-ALL cells.

In the other study, Garcia-Manero et al. [45] examined methylation pattern of five genes i.e. MDR, ER, P73, P15 and P16 in ALL by using FFPE tissues at the time of diagnosis and first relapse. They reported that methylation patterns were stable in most of the patients, in spite of the fact that a subgroup of the patients acquired novel methylation changes.

From the present study, it could be concluded that isolated DNA from smears and FFPE tissues lead to successful results of methylation analysis. Thus, as further development, it is

recommended that smears and FFPE tissues to be used for large scale methylation analysis by considering the size of amplicon.

5.6 Conclusion and future perspectives

Retrospective studies focusing on genes and genetic alternation imply the use of archived samples. Especially after the patients' death, these samples may represent the only possible way to get DNA from such participants under study. The present research study was set to investigate the suitability of isolated DNA, taken from archived samples of treated children suffering from ALL, for multiple SNP and methylation analysis.
To achieve this purpose, the quantity and quality of archived samples were examined. The two commonly used, UV and flourometric spectroscop were used to measure DNA quantity. Based on our results, these methods were not as accurate as needed for determination of DNA concentration and quality control of whole genome amplification.

In this study, DNA profiling was used to assess the quality and the ability of amplification of the isolated DNA from archived samples. The results indicated that 100% of smears were able to amplify markers with size up to 234 bp. Concerning FFPE tissues, they could generate markers with size up to 170 bp in 100% of cases. Our study has shown the feasibility of amplifiable DNA extracted from both bone marrow smears and paraffin embedded bone marrow tissues.

Recently, the spectrum of molecular analysis of archived samples has been increased in many retrospective studies. However, the quality of recovered nucleic acids is reduced and this is still problematic for future molecular analyses. Dealing with archived samples requires an optimized method for DNA extraction and quantification. This optimized protocol will open up vast archived samples for large scale genetic analysis and unlock a wealth of biological information.

High-throughput sequencing technologies facilitate better understanding of the inter-individual variation in genetic analysis and epidemiology studies. However, the suitability of the isolated DNA from tissue blocks and smears for multiple SNP analysis is not supported by firm evidences.

This study showed that bone marrow smears and FFPE tissues are suitable for methylation analysis as well. Therefore, large scale retrospective analysis could be possible by using the valuable archived samples available in pathology archives. Our study was limited in terms of number of patients, while generating comprehensive methylation pattern requires large number of samples. Methylation analysis from archived materials could face the same challenge of limited amounts of isolated DNA. Based on a study carried out by Sahoo et al. [77] some amount of DNA template were lost during conventional bisulphate modification. They have proposed *in situ* bisulphate treatment that could be an approach for overcoming most limiting factors in applying archived samples. In the present study, it was not possible to utilize *in situ* bisulphate because we needed untreated DNA for multiple SNPs analysis.

This study suggests that optimizing a method with high yield and quality is required for DNA extraction from archived samples. Also accurate and standard methods for quantification of fragmented DNA and WGA product are needed before applying downstream application. Based on our results, UV absorbance and DNA fluorescence have limited value in predicting WGA efficiency. However, further experiments are needed to be conducted to confirm it as a generalized expression.

It should be taken into consideration that our sample population was small and this exposed our results to some degree of uncertainty. Thus, to get more reliability over the results, further studies using a larger sample population is recommended.

6 References

1. NCI. Childhood Cancers. 2008; Available from: http://www.cancer.gov/cancertopics/factsheet/Sites-Types/childhood.
2. Pui CH, E.W., Treatment of acute lymphoblastic leukemia. N Engl J Med, 2006. 354(2): p. 166-78.
3. Susannah E. Koontz, P., BCOP, Common Pediatric Cancers. U.S. Pharmrmacist, 2004. 29(10).
4. Gustafsson G, K.A., Clausen N, Garwicz S, Kristinsson J, Lie SO, Moe PJ, Perkkiö M, Yssing M, Saarinen-Pihkala UM., *Intensified treatment of acute childhood lymphoblastic leukaemia has improved prognosis, especially in non-high-risk patients: the Nordic experience of 2648 patients diagnosed between 1981 and 1996. Nordic Society of Paediatric Haematology and Oncology (NOPHO).* Acta Paediatr, 1998. **87**(11): p. 1151-61.
5. Lisa Lyngsie Hjalgrim, K.R., Kjeld Schmiegelow, Stefan So¨derha¨ll,Svein Kolmannskog, Kim Vettenranta, Jon Kristinsson, Niels Clausen,, *Age- and Sex-Specific Incidence of Childhood Leukemia by Immunophenotype in the Nordic Countries.* Journal of the National Cancer Institute, , 2003. **95**(20): p. 1539-44.
6. Wesolowska A, D.M., Borst L, Gautier L, Bak M, Weinhold N, Nielsen BF, Helt LR, Audouze K, Nersting J, Tommerup N, Brunak S, Sicheritz-Ponten T, Leffers H, Schmiegelow K, Gupta R., *Cost-effective multiplexing before capture allows screening of 25□000 clinically relevant SNPs in childhood acute lymphoblastic leukemia.* Leukemia, 2011. **25**(6): p. 1001-6.
7. Prucker, C., et al., *Induction death and treatment-related mortality in first remission of children with acute lymphoblastic leukemia: a population-based analysis of the Austrian Berlin-Frankfurt-Munster study group.* Leukemia, 2009. **23**(7): p. 1264-9.
8. Lund B, Å.A., Heyman M, Kanerva J, Harila-Saari A, Hasle H, Söderhäll S, Jónsson ÓG, Lydersen S, Schmiegelow K, *Risk factors for treatment related mortality in childhood acute lymphoblastic leukaemia.* Pediatr Blood Cancer., 2011. **56**(4): p. 551-9.
9. David Hopkinson, D.W., *An Introduction to Genetic Polymorphism*, in *Human blood cells : consequences of genetic polymorphisms and variations* M.-J. King., Editor. 2000, River Edge, NJ : Imperial College Press, 2000.

10. Albert, P.R., *What is a functional genetic polymorphism? Defining classes of functionality.* J Psychiatry Neurosci, 2011. **36**(6): p. 363-5.
11. Evans, W.E. and M.V. Relling, *Moving towards individualized medicine with pharmacogenomics.* Nature, 2004. **429**(6990): p. 464-8.
12. Phillips, C., *Online resources for SNP analysis: a review and route map.* Mol Biotechnol, 2007. **35**(1): p. 65-97.
13. Smith, K. *Genetic Polymorphism and SNPs*. 2002 19 Feb; Available from: http://www.cs.mcgill.ca/~kaleigh/compbio/snp/snp_summary.html.
14. *SNP Fact Sheet*. 2003; Available from: http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml.
15. Innocenti, F., N.J. Cox, and M.E. Dolan, *The use of genomic information to optimize cancer chemotherapy.* Semin Oncol, 2011. **38**(2): p. 186-95.
16. Engle LJ, S.C., Landers JE., *Using high-throughput SNP technologies to study cancer* Oncogene, 2006. **25**(11): p. 1594-601.
17. Shastry, B.S., *SNPs in disease gene mapping, medicinal drug development and evolution.* J Hum Genet, 2007. **52**(11): p. 871-80.
18. Lee, P.H. and H. Shatkay, *F-SNP: computationally predicted functional SNPs for disease association studies.* Nucleic Acids Res, 2008. **36**(Database issue): p. D820-4.
19. Blow, N., *Tissue preparation: Tissue issues.* Nature, 2007. **448**(7156): p. 959-63.
20. Steve S. Michalik, S.R.a.D.S., Sigma-Aldrich Biotechnology, St. Louis, Mo., *Overcoming Poor Quality DNA.* Drug Discovery and Development, 2008. **11**(3).
21. Conter V, R.C., Sala A, Chiesa R, Citterio M, Biondi A, *Acute Lymphoblastic Leukemia* Orphanet Encyclopedia, 2004.
22. Pabst T, S.J., Tobler A, Fey MF., *Detection of microsatellite markers in leukaemia using DNA from archival bone marrow smears.* Br J Haematol, 1996. **95**: p. 135-137.
23. Vince, A., M. Poljak, and K. Seme, *DNA extraction from archival Giemsa-stained bone-marrow slides: comparison of six rapid methods.* Br J Haematol, 1998. **101**(2): p. 349-51.
24. Turner, B.J., et al., *Overcoming poor attendance to first scheduled colonoscopy: a randomized trial of peer coach or brochure support.* J Gen Intern Med, 2008. **23**(1): p. 58-63.
25. Falconi, M., et al., *Effect of fixative on chromatin structure and DNA detection.* Microsc Res Tech, 2007. **70**(7): p. 599-606.

26. Srinivasan, M., D. Sedmak, and S. Jewell, *Effect of fixatives and tissue processing on the content and integrity of nucleic acids.* Am J Pathol, 2002. **161**(6): p. 1961-71.
27. Wang, F., et al., *DNA degradation test predicts success in whole-genome amplification from diverse clinical samples.* J Mol Diagn, 2007. **9**(4): p. 441-51.
28. Technologies, N., *NanoDrop Technical Support Bulletin T009 260/280 and 260/230 Ratios*. 2007.
29. Maciver, I., *Methods for Determining DNA Yield and Concentration*. 2012, Promega Connections.
30. Nicklas, J.A. and E. Buel, *Quantification of DNA in forensic samples.* Anal Bioanal Chem, 2003. **376**(8): p. 1160-7.
31. Technology, N.M. *Genetic Profiling Using STR Analysis*. 2007; Available from: http://dspace.nmt.edu/dspace/handle/10136/46
32. Biosystems, A., *AmpFLSTR® Profiler User's Manual*. 2005, Applied Biosystems
33. Wickenheiser, R.A., *General Guidelines for Categorization and Interpretation of Mixed STR DNA Profiles.* Canadian Society of Forensic Science Journal, 2006. **39**(4): p. 179-216.
34. Spits, C., et al., *Whole-genome multiple displacement amplification from single cells.* Nat Protoc, 2006. **1**(4): p. 1965-70.
35. Dean, F.B., et al., *Comprehensive human genome amplification using multiple displacement amplification.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 5261-6.
36. QIAGEN, *REPLI-g FFPE Handbook*. 2007.
37. Mamanova, L., et al., *Target-enrichment strategies for next-generation sequencing.* Nat Methods, 2010. **7**(2): p. 111-8.
38. Mertes, F., et al., *Targeted enrichment of genomic DNA regions for next-generation sequencing.* Brief Funct Genomics, 2011. **10**(6): p. 374-86.
39. BGI. *Exome sequencing*. 2011; Available from: http://www.bgisequence.com/eu/services/sequencing-services/disease-research/exome-sequencing/.
40. Galm, O., J.G. Herman, and S.B. Baylin, *The fundamental role of epigenetics in hematopoietic malignancies.* Blood Rev, 2006. **20**(1): p. 1-13.
41. Das, P.M. and R. Singal, *DNA methylation and cancer.* J Clin Oncol, 2004. **22**(22): p. 4632-42.

42. Ehrlich, M., *DNA methylation in cancer: too much, but also too little.* Oncogene, 2002. **21**(35): p. 5400-13.
43. Jones, P.A. and S.B. Baylin, *The fundamental role of epigenetic events in cancer.* Nat Rev Genet, 2002. **3**(6): p. 415-28.
44. Gutierrez, M.I., et al., *Concurrent methylation of multiple genes in childhood ALL: Correlation with phenotype and molecular subgroup.* Leukemia, 2003. **17**(9): p. 1845-50.
45. Garcia-Manero, G., et al., *DNA methylation patterns at relapse in adult acute lymphocytic leukemia.* Clin Cancer Res, 2002. **8**(6): p. 1897-903.
46. Garcia-Manero G, J.S., Daniel J, Williamson J, Albitar M, Kantarjian HM, Issa JP., *Aberrant DNA methylation in pediatric patients with acute lymphocytic leukemia.* Cancer, 2003. **97**(3): p. 695-702.
47. Waugh, D.J. and C. Wilson, *The interleukin-8 pathway in cancer.* Clin Cancer Res, 2008. **14**(21): p. 6735-41.
48. Scupoli, M.T., et al., *Bone marrow stromal cells and the upregulation of interleukin-8 production in human T-cell acute lymphoblastic leukemia through the CXCL12/CXCR4 axis and the NF-kappaB and JNK/AP-1 pathways.* Haematologica, 2008. **93**(4): p. 524-32.
49. Sabina Chiaretti, X.L., Robert Gentleman,Antonella Vitale,Marco Vignetti,Franco Mandelli,Jerome Ritz, and Robin Foa. *Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival.* Blood, 2004. **103**(7): p. 2771-2778.
50. Dimberg, J., et al., *DNA promoter methylation status and protein expression of interleukin-8 in human colorectal adenocarcinomas.* Int J Colorectal Dis, 2012. **27**(6): p. 709-14.
51. Andia, D.C., et al., *DNA methylation status of the IL8 gene promoter in aggressive periodontitis.* J Periodontol, 2010. **81**(9): p. 1336-41.
52. Xie, K., *Interleukin-8 and human cancer biology.* Cytokine Growth Factor Rev, 2001. **12**(4): p. 375-91.
53. Francia di Celle P, M.S., Riera L, Stacchini A, Reato G, Foa R., *Interleukin-8 induces the accumulation of B-cell chronic lymphocytic leukemia cells by prolonging survival in an autocrine fashion.* Blood, 1996. **87**(10): p. 4382-9.

54. Hamatani, K., et al., *Improved RT-PCR amplification for molecular analyses with long-term preserved formalin-fixed, paraffin-embedded tissue specimens.* J Histochem Cytochem, 2006. **54**(7): p. 773-80.
55. QIAGEN, *FFPE Tissue Handbook* 2010.
56. NanoDrop Technologies, I., *ND-1000 Spectrophotometer v3.2 User Manual*. 2005.
57. Invitrogen, *Qubit® dsDNA BR Assay Kits*. 2011.
58. Technologies, A., *SureSelected Target Enrichment System Protocol* in *SureSelected Target Enrichment System* 2009, Agilent Technologies.
59. Springer, in *Springer Images*, Springer Images.
60. NFSTC. *STR Data Analysis*. DNA Analyst Training; Available from: http://www.nfstc.org/pdi/Subject06/pdi_s06_m02_02.htm.
61. Ludyga, N., et al., *Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses.* Virchows Arch, 2012. **460**(2): p. 131-40.
62. Wei Cao, M.D., Ph.D.a, Mia Hashibe, M.P.H.a, Jian-Yu Rao, M.D.b, Hal Morgenstern, Ph.D.a, Zuo-Feng Zhang, M.D., Ph.D, *Comparison of methods for DNA extraction from paraffin-embedded tissues and buccal cells.* Cancer Detection and Prevention, 2003. **27**(5): p. 397-404.
63. Gilbert, M.T., et al., *The isolation of nucleic acids from fixed, paraffin-embedded tissues-which methods are useful when?* PLoS One, 2007. **2**(6): p. e537.
64. Nielsen, K., et al., *Comparison of five DNA quantification methods.* Forensic Sci Int Genet, 2008. **2**(3): p. 226-30.
65. Holden, M.J., et al., *Factors affecting quantification of total DNA by UV spectroscopy and PicoGreen fluorescence.* J Agric Food Chem, 2009. **57**(16): p. 7221-6.
66. Georgiou, C.D. and I. Papapostolou, *Assay for the quantification of intact/fragmented genomic DNA.* Anal Biochem, 2006. **358**(2): p. 247-56.
67. S J Ahn, J.C., and J R Emanuel, *PicoGreen quantitation of DNA: effective evaluation of samples pre- or post-PCR.* Nucleic Acids Res, 1996. **24**(13): p. 2623–2625.
68. Kanter, P.M. and H.S. Schwartz, *A fluorescence enhancement assay for cellular DNA damage.* Mol Pharmacol, 1982. **22**(1): p. 145-51.
69. Haque KA, P.R., Beerman MB, Struewing JP, Chanock SJ, Bergen AW., *Performance of high-throughput DNA quantification methods.* BMC Biotechnol, 2003. **3**(20).

70. Bergen, A.W., et al., *Comparison of yield and genotyping performance of multiple displacement amplification and OmniPlex whole genome amplified DNA generated from multiple DNA sources.* Hum Mutat, 2005. **26**(3): p. 262-70.
71. Hansen, H.M., et al., *DNA quantification of whole genome amplified samples for genotyping on a multiplexed bead array platform.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(8): p. 1686-90.
72. Muharam, F.A., *Overcoming problems wiyh limiting DNA samples in forensics and clinical diagnostics using Multiple Displacement Amplification*, in *Science*. 2005, Queensland University of Technology: Brisbane Australia.
73. Melissa R. Gunn, K.H., et al,. *A test of the efficacy of whole-genome amplification on DNA obtained from low-yield samples.* Molecular Ecology Notes, 2007. **7**: p. 393-399.
74. QIAGEN. *Whole Genome Amplification (WGA)*. Available from: http://www.qiagen.com/products/bytechnology/wholegenomeamplification/tutorial/default.aspx.
75. Babol-Pokora K, B.J., *SNP-minisequencing as an excellent tool for analysing degraded DNA recovered from archival tissues.* Acta Biochim Pol, 2008. **55**(4): p. 815-819.
76. Thompson ER, H.S., Forrest SM, Campbell IG., *Whole genome SNP arrays using DNA derived from formalin-fixed, paraffin-embedded ovarian tumor tissue.* Hum Mutat, 2005. **26**(4): p. 384-389.
77. Sahoo R, B.A., Payal K, Wani S,. , *Improved Method For Detection Of Methylation Status Of Genes From Limited, Archived, FFPE And FNAC Samples.* Journal of Clinical and Diagnostic Research, 2009. **3**(3): p. 1493-1499.

Appendix A

		D3S1358	Height	Vwa	High	FGA	High	AMEL	High	TH01	High	TPOX	High	CSF1PO	High	D5S818	High	DI3S317	High	D7S820	High
Patient 1	Blood	129/137	6114/3882	176/180	4871/3935	235/247	2871/2392	x	18390	172/184	2517/1122	235	8923	295/299	1523/1498	149/154	12470/10358	210	25526	273/281	10909
	Smear	129/137	3687/3680	176/180	2124/2313	235/247	461/516	x	11255	172/184	1296/487	235	437	295/299	147/146	149/154	3961/3329	210	1435	273/281	402/305
	Smear.WGA	–	–	–	–	–	–	x	176	172	1029	–	–	–	–	149/154	193/158	–	–	–	–
	Biopsy	128/136	16540/12654	174/178	10838/9625	234/246	4044/3213	–	–	169/181	5603/4108	233	3516	294/298	421/242	147/151	32659/32659	208	19698	272/280	1137/1039
	Biosy.WGA	–	–	174/178	84/62	234	3886	x	65/156	169	66/60	233	30149	–	–	147	5193	–	–	–	–

		D3S1358	High	Vwa	High	FGA	High	AMEL	High	TH01	High	TPOX	High	CSF1PO	High	D5S818	High	DI3S317	High	D7S820	High
pateint2	Blood	121/125	3689/3323	184	5872	239	4289	x	11620	172/176	829/3253	224/239	4022/1032	295/304	2591/1961	149/154	6954/11325	218/222	11524/14701	265/273	5138/2171
	Smear	121/125	8879/7749	184	6627	239	2724	x	19640	172/176	1115/1422	224	1051	295/304	221/224	149/154	10157/8037	218/222	1903/1776	265/273	1484/998
	Smear.WGA	–	–	184	221	–	–	x	128	–	–	–	–	–	–	–	–	–	–	–	–
	Biopsy	120/124	24139/21351	183	25828	238	17649	x	658	169/173	9318/8031	222	8965	294/302	1038/1851	146/150	32444/32444	216/220	16404/15277	263/272	3319/2818
	Biosy.WGA	120/124	11878/10515	183	1205	238	678	x	6747	175	2281	222	464	295/304	68/146	149/154	2364/9147	216/220	428/456	264/272	189/164

	sample	D3S1358	High	Vwa	High	FGA	High	AMEL	High	TH01	High	TPOX	High	CSF1PO	High	D5S818	High	DI3S317	High	D7S820	High
pateint3	Blood	125/129	5850/5513	180	7810	223/239	2174/2005	xy	10913/3380	172	4486	223/235	3232/8136	295/304	2303/1341	149	32705	214/218	15243/14433	269/277	1701/2525
	Smear	125/129	6265/5405	180	4641	223/239	472/467	xy	7058/6204	172/186	1344/971	223/235	424/407	–	–	149	8272	214/218	664/723	269/277	393/284
	Smera.WGA	129	72	180	1315	223	244	–	–	180	328	223	59	–	–	–	–	–	–	–	–
	Biopsy	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

	sample	D3S1358	High	Vwa	High	FGA	High	AMEL	High	TH01	High	TPOX	High	CSF1PO	High	D5S818	High	DI3S317	High	D7S820	High
pateint4	Blood	129/137	4806/5538	172/184	4225/2787	231/243	2724/1606	xy	2749/4807	172/180	3897/2242	228	32697	295/304	986/860	149/154	18537/14452	226	21031/9988	277	6722
	Smear	129/137	3672/2942	172	904	–	–	xy	5589/4123	172/180	882/574	228	224	–	–	149/154	1710/1542	226	165	–	–
	Smera.WGA	129/137	50/52	172	90	–	–	xy	175/153	–	–	–	–	–	–	149/154	150/151	226	73	–	–
	Biopsy	129/137	1565/12717	170	20851	230/242	6263/498335	xy	32614/32658	171	10121	227	8263	293/304	995/899	147/151	32561/32561	225	24259	275	4632
	Biopsy.WGA	137	31847	172	5566	230	240	xy	15948/722	171	2145	227	6742	295/304	–	149/154	18388/15430	225	10467	275	3524

	sample	D3S1358	High	Vwa	High	FGA	High	AMEL	High	TH01	High	TPOX	High	CSF1PO	High	D5S818	High	DI3S317	High	D7S820	High
pateint5	Blood	121/133	1656/6087	180	3799	227/247	799/850	x	3089	172/186	518/870	223	5186	295/304	522/834	149	14576	214/222	6083/4270	277	4565
	Smear	121/133	7125/6134	180	7061	227/247	2793/2059	x	15576	172/186	2150/2232	223	1826	295/304	776/726	149	18998	214/222	3286/2762	277	3476
	Smear.WGA	–	–	–	–	–	–	x	191	186	234	–	–	–	–	149	824	241/222	50/147	–	–
	Biopsy	–	–	–	–	–	–	x	73	–	–	–	–	–	–	149	2041	–	–	–	–
	Biopsy.WGA	121/134	101/88	180	380	–	–	x	680	-	–	224	279	–	–	149	1296	223	146	–	–

	sample	D3S1358	High	Vwa	High	FGA	High	AMEL	High	TH01	High	TPOX	High	CSF1PO	High	M8	High	DI3S317	High	D7S820	High
pateint6	Blood	125/129	6754/3663	184/188	2590/2520	227/235	3255/1354	x	6828	176	8091	223/231	5253/1691	299/304	1619/871	145/149	19736/12164	210	18517	269/277	3469/1658
	Smear	125/129	1698/1481	184/188	389/222	–	–	x	6508	176	254	223	496	–	–	145/149	7674/5068	210	219	269/276	65/56
	Smera.WGA	129	1522	–	–	–	–	x	275	–	–	–	–	–	–	145/147	81/108	–	–	–	–
	Biopsy	125/128	95/58	180/184	10068/8382	225/233	3583/3033	x	312/1131	173	15894	222/231	4410/3578	298/302	621/519	143/147	32566/32566	208	27833	268/276	1433/1287
	Biopsy.WGA	124	272	180/184	119/223	233	601	x	89/207	173	139	230	5589	–	–	143/147	8082/9265	208	815	268/	249

	sample	D3S1358	High	Vwa	High	FGA	High	AMEL	High	TH01	High	TPOX	High	CSF1PO	High	D5S818	High	DI3S317	High	D7S820	High
pateint7	Blood	129/133	6081/6475	176	11680	230/247	5570/6061	x	9826	172/186	3774/3662	227/235	3846/2505	295/304	3806/4085	154/158	12454/12081	214/218	10833/10033	277/280	6114/5247
	Smear	129/133	26567/25101	176	32703	230/247	15223/12379	x	32527	172/186	18452/13316	227/235	11817/10683	295/304	8232/7785	154/158	32242/32492	214/218	31982/30082	277/280	18411/15451
	Biopsy	129/133	3873/4498	176	15591	230/247	6198/5872	x	23385	172/186	4551/4183	227/235	2391/2354	295/304	1608/1359	154/158	32563/28141	214/218	16832/16601	277/280	27422/1168

	sample	D3S1358	High	Vwa	High	FGA	High	AMEL	High	M5	High	M6	High	M7	High	M8	High	M9	High	M10	High
pateint8	Blood	132	32575	172/180	24272/19140	230/247	14389/12181	x	32594	172/186	14796/10014	223/235	11527/11444	230	16768	149/154	32123/32375	210/225	31832/26733	264/276	17416/14550
	Smear	132	32025	172/180	15334/13463	230/243	6223/5328	x	32615	172/186	11949/8533	223/235	6494/4769	230	4992	149/154	30240/27179	210/225	17219/14378	264/276	6948/5866
	Biopsy																				

	sample	D3S1358	High	Vwa	High	FGA	High	AMEL	High	TH01	High	TPOX	High	CSF1PO	High	D5S818	High	DI3S317	High	D7S820	High
pateint9	Blood	133	11516	180	11898	243/250	4939/6005	x	9748	175	6999	223	7318	299/304	3355/2651	144/149	12066/10511	205/222	9990/8793	272/280	7112/6416
	Smear	133	10459	180	8175	243/250	966/780	x	22824	175	3381	223	1229	299/304	285/177	144/149	13215/10656	205/222	3773/2861	272/280	1251/700
	Biopsy	133	2074	180	1851	243/250	803/388	x	5314	175	1114	223	667	–	–	144/149	5414/4149	205/222	1517/1136	272/280	291/227

	sample	D3S1358	High	Vwa	High	FGA	High	AMEL	High	TH01	High	TPOX	High	CSF1PO	High	D5S818	High	DI3S317	High	D7S820	High
pateint10	Blood	133	28309	187	26840	230/243	12372/13343	xy	14371/13564	186	8758	223/235	101947/9891	304/308	10807/10967	149/154	26129/30904	222	32708	267/272	20921/18906
	Smear	133	16334	187	9028	230/243	1165/1448	xy	16805/12290	186	2724	223/235	1199/1094	304/308	394/347	149/154	9185/9467	222	3529	267/272	943/1123
	Biopsy	133	199	187	59	230/243	566/689	xy	159/161	186	899	223/235	909/1005	304/308	289/302	149/154	123/191	222	2560	267/272	391/405

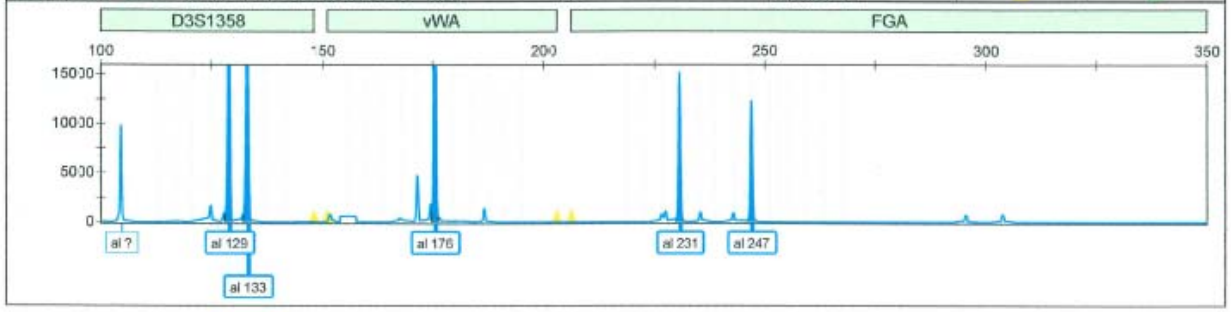
	sample	D3S1358	High	Vwa	High	FGA	High	AMEL	High	TH01	High	TPOX	High		High	D5S818	High	DI3S317	High	D7S820	High
pateint11	Blood	129	17460	180/184	9085/9359	226/245	8468/8239	xy	10257/9487	172	9889	227/235	6622/5832	295	12423	144/154	18307/18285	218/222	17336/15161	264/272	13041/12587
	Smear	129	20901	180/184	8855/5551	226/245	1938/1327	xy	18854/17431	172	5177	227/235	1087/1027	295	703	144/154	16050/17311	218/222	3889/4332	264/272	734/897
	Biopsy	129	4245	180/184	2293/1295	226/245	1672/1097	xy	3344/3430	172	2384	227/235	680/531	295	666	144/154	5674/5860	218/222	2433/2581	264/272	729/475

Appendix B

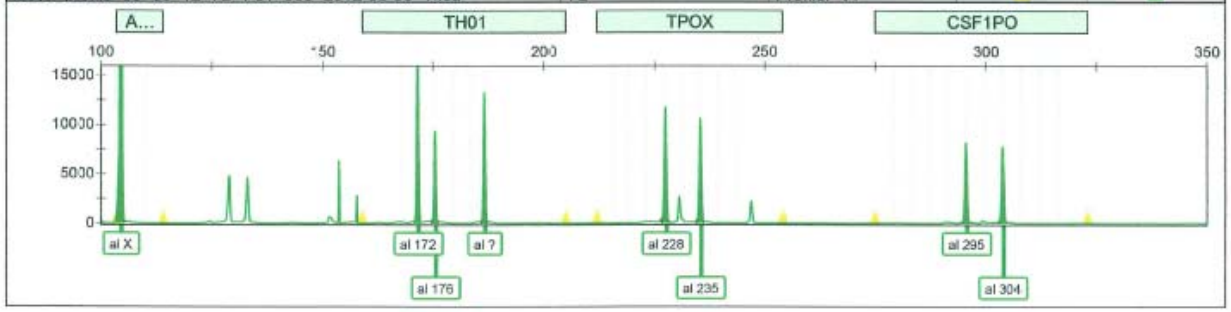
Applied Biosystems
GeneMapper v3.7

Laeya_profile_09_05_12

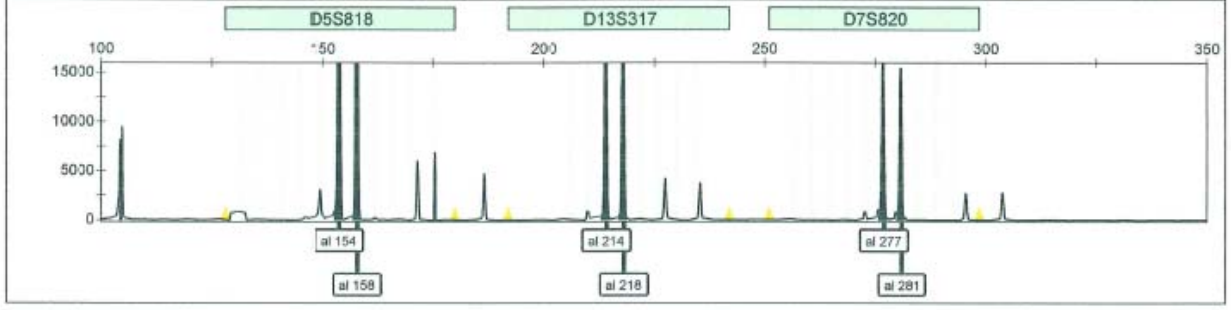
Sample File	Sample Name	Panel	OS	SQ
laeya_profile_09_05_12_7D_F01_003_2012-05-09_1.fsa	7D	Profiler_v1		



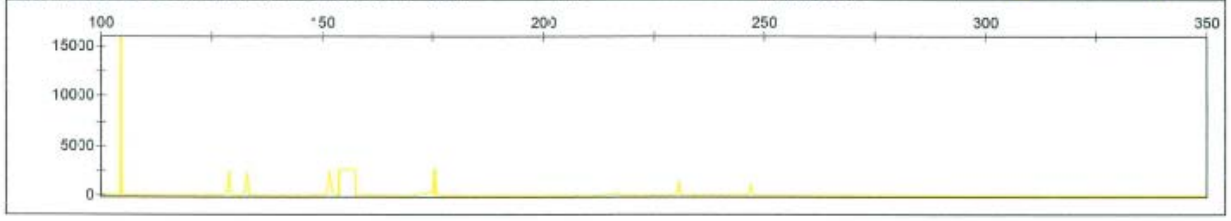
laeya_profile_09_05_12_7D_F01_003_2012-05-09_1.fsa | 7D | Profiler_v1



laeya_profile_09_05_12_7D_F01_003_2012-05-09_1.fsa | 7D | Profiler_v1



laeya_profile_09_05_12_7D_F01_003_2012-05-09_1.fsa | 7D | Profiler_v1

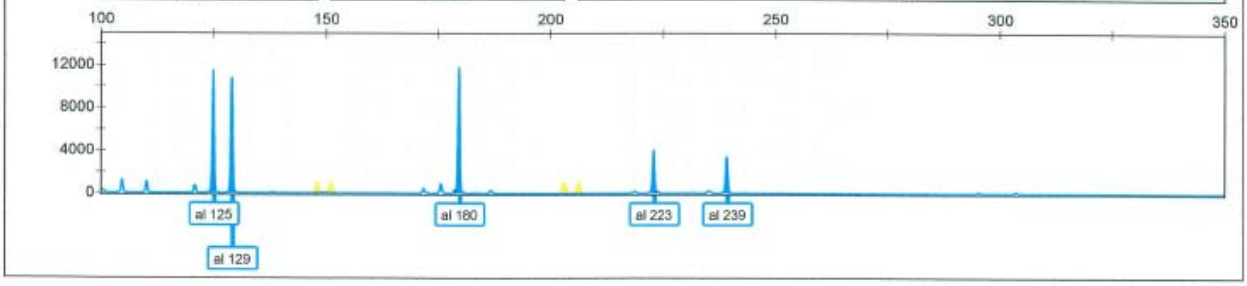


Applied Biosystems

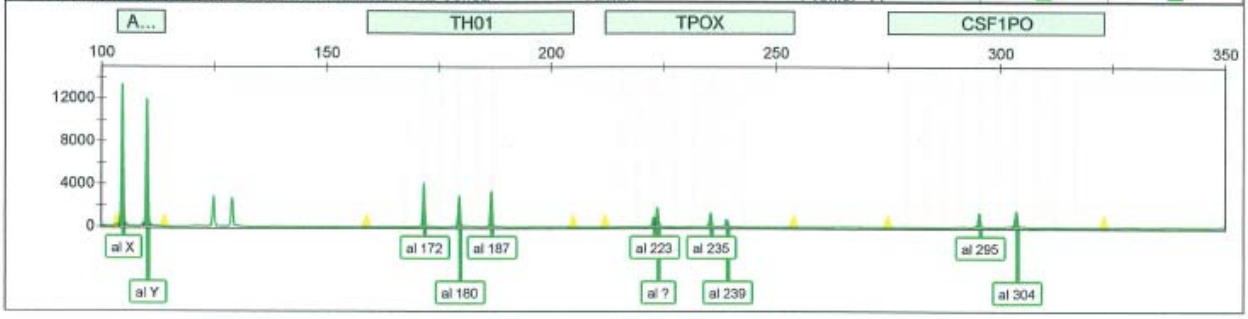
GeneMapper v3.7

laeya(3).06.12.2011

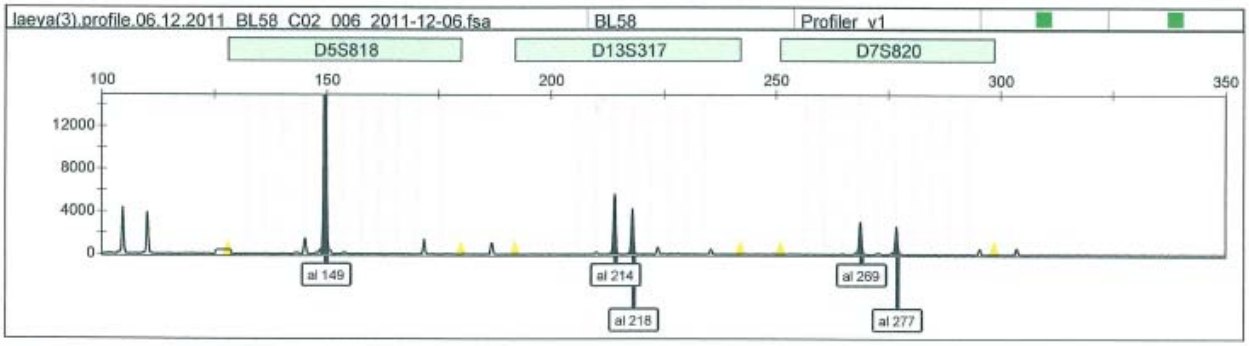
Sample File	Sample Name	Panel	OS	SQ
laeya(3).profile.06.12.2011_BL58_C02_006_2011-12-06.fsa	BL58	Profiler_v1		



Sample File	Sample Name	Panel	OS	SQ
laeya(3).profile.06.12.2011_BL58_C02_006_2011-12-06.fsa	BL58	Profiler_v1		



Sample File	Sample Name	Panel	OS	SQ
laeya(3).profile.06.12.2011_BL58_C02_006_2011-12-06.fsa	BL58	Profiler_v1		



Sample File	Sample Name	Panel	OS	SQ
laeya(3).profile.06.12.2011_BL58_C02_006_2011-12-06.fsa	BL58	Profiler_v1