# CYTOGENOMIC PROFILING OF CHRONIC LYMPHOCYTIC LEUKAEMIA PATIENTS USING DNA MICROARRAY

**DR WAN NORIZZATI BINTI WAN MOHAMAD ZAMRI**

**DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF PATHOLOGY (MEDICAL GENETICS)**


**UNIVERSITI SAINS MALAYSIA**

**2021**

# ACKNOWLEDGEMENT

First and foremost, praises and thanks to the Allah, the Almighty, for His showers of blessings to complete the dissertation project successfully.

I would like to express my deep and sincere gratitude to my main supervisor, Associate Prof. Dr. Sarina Sulong, Coordinator of Human Genome Center (HGC) for giving me the opportunity to do research and providing invaluable guidance throughout this research. Her dynamism, vision, sincerity and motivation have deeply inspired me. She also provided me ample freedom to express myself through this work and at the same time he also provided critical comments as well as valuable suggestions and recommendations to enable me to make necessary improvements in my research work. I could not imagine having a better advisor and mentor for my master programme.

I also would like to express my heartful thanks to Dr. Nazihah Mohd Yunus and Dr. Ahmad Aizat Abdul Aziz as my co-supervisors, for teaching me the methodology to carry out the research and to present the research works as clearly as possible. It was a great privilege and honor to work and study under their guidance.

I would like to take this opportunity to express my appreciation to Prof Zilfalil Alwi and Dr. Marahaini for their counsel in finishing this dissertation. My sincere thanks also go to all HGC lecturers and other HGC staff especially Encik Zaki, Puan Hafizah, Puan Siti Mariam, Puan Norhidayah, Puan Zulaikha Bakar, Encik Nik Mohd Zulfikri, Puan Nurul Alia, and Puan Fatin for expert technical assistance. Without their passionate participation and assistance, the lab work and data analysis could not have been successfully conducted. Special thanks to Puan Noratifah, Encik Md Ros and Mr. Chia, the science officers of HGC, for always ensuring optimization of workflow in the lab.

I also would like to thank my fellow MPath comrades for making so many ordinary moments extraordinary. All the laughters and beautiful memories we have shared are always remembered.

This study was financially supported by Universiti Sains Malaysia under Master of Medicine Research Grant Fund (GPSP/1001/PPSP/8070011). I would like to express my sincere thanks to Universiti Sains Malaysia for providing this grant.

Last but not least, I would like to express my deepest gratitude to my parents and siblings for their love, patience, support, prayers and understanding. Without their unwavering support, I could never have finished this project. I am completely indebted to them and dedicate this work to them.

Finally, it is difficult to name all, but I offer my sincere thanks to all those who have knowingly and unknowingly helped me in the successful completion of this project.

## ABSTRAK

## PEMPROFILAN SITOGENOMIK UNTUK PESAKIT LEUKEMIA LIMFOSITIK KRONIK MENGUNAKAN *DNA MICROARRAY*

**Pengenalan:** Leukemia limfositik kronik (CLL) menyumbang kira-kira 5 – 11 % daripada penyakit limfoproliferatif di negara Barat. Meskipun CLL mempunyai perkembangan penyakit yang indolen, ia dilaporkan lebih agresif dan mempunyai masa yang lebih singkat untuk menerima terapi pertama jika dibandingkan dengan negara Barat kerana mempunyai lanskap genomik yang rumit. Walaupun keabnormalan sitogenetik (CA) dalam CLL dapat dikesan mengunakan analisis *'Fluorescence in situ hybridization'* (FISH) and '*Multiplex Ligation- dependent Probe Amplification'*(MLPA) bersama dengan analisis sitogenetik konvensional (CCA), namun begitu, masih terdapat keabnormalan sitogenetik krip yang tidak dapat didiagnose disebabkan oleh kelemahan analysis tersebut. Kajian ini mengkaji tentang pengesanan keabnormalan sitogenetik menggunakan *DNA microarray* yang boleh mengatasi kelemahan *FISH* dan *MLPA* dengan mengesan keabnormalan krip yang boleh terlepas daripada pengesanan kedua-dua kaedah serta mengenalpasti calon gen yang terlibat dalam CLL untuk populasi kita.

**Kaedah:** Dalam kajian retrospektif, sampel arkib DNA daripada kajian lepas (Tajuk Geran Jangka Pendek USM: Analysis of Hypermethylation Status of Tumor Suppressor Genes *p16INK4a, p15INK4b, ADAM12* and *PCDHGB7* in Chronic Lymphocytic Leukemia Patients and Normal Individuals)(JEPeM code: USMKK/PPP/JEPeM/234.3(07) untuk 7 pesakit yang baru didiagnos sebagai leukemia limfositik kronik pada tahun 2012 telah digunapakai dalam analisis *DNA microarray*

menggunakan Affymetrix CytoScan® 750K mengikut piawaian pengeluar. Maklumat klinikal terperinci pesakit diekstrak daripada borang permohonan sitogenetik dan rekod perubatan pesakit.

**Keputusan:** Keabnormalan kromosom dikenalpasti dalam semua pesakit menggunakan *DNA microarray.* Penemuan menunjukkan terdapat pemotongan kromosom 13q dan 14q yang berlainan saiz antara 0.42MB hingga 39.97MB. Menariknya, terdapat juga KS tambahan seperti trisomi 3, pemotongan interstis kromosom yang krip pada kromosom 13q,14q,16p,Xp dan penduaan rantau kromosom 14q32.33 dan kromosom 22q11.2. Pemotongan kromosom pada rantau 14q32.33 adalah berkait rapat dengan status mutasi pada gen *IGH* manakala keabnormalan krip yang lain memerlukan kajian seterusnya untk mengkaji sumbangan mereka kepada pathogenesis leukemia limfositik kronik. Kesemua pemotongan dan penduan rantau kromosom yang krip tidak dapat dikesan menggunakan CCA kerana saiz yang sangat kecil (<5MB). Penemuan ini diharapkan dapat membantu doktor merawat untuk mengklasifikasikan pesakit sewajarnya serta dapat memberikan permerhatian yang lebih teliti.

**Kesimpulan:** *DNA microarray* merupakan satu teknik analisis yang sangat berharga dalam mengesan keabnormalan genomik untuk pesakit CLL terutamanya pengesanan keabnormalan krip. Penerapan *DNA microarray* dalam diagnosis CLL dapat membantu doktor untuk memberikan rawatan yang tepat bersesuaian dengan keperluan pesakit.

## ABSTRACT

## CYTOGENOMIC PROFILING OF CHRONIC LYMPHOCYTIC LEUKAEMIA PATIENTS USING DNA MICROARRAY

**Introduction:** Chronic lymphocytic leukaemia (CLL) accounts approximately 5 – 11% of lymphoproliferative disorders in Western countries. Despite the indolent disease progression, it had been reported that Asian CLL is more aggressive and had shorter time to first therapy compared to Western counterpart due to its complex genomic landscape. Even though chromosomal aberration of CLL can be detected by Fluorescence *in situ* Hybridization (FISH) and Multiplex Ligation-dependent Probe Amplification (MLPA) together with conventional cytogenetic analysis (CCA), however some of the abnormality can be missed due to method's limitation. This study investigated chromosomal aberration using DNA microarray to detect cryptic chromosomal abnormality that could escape detection of FISH and MLPA which able to overcome both methods limitation and subsequently identify candidate genes that could involve in our population CLL.

**Methodology:** In this retrospective study, archive genomic DNA of 7 newly diagnosed CLL in 2012 from previous study (USM Short Term grant title: Analysis of Hypermethylation Status of Tumor Suppressor Genes *p16INK4a, p15INK4b, ADAM12* and *PCDHGB7* in Chronic Lymphocytic Leukemia Patients and Normal Individuals)(JEPeM code: USMKK/PPP/JEPeM/234.3(07)) were used. They were subjected to DNA microarray analysis using Affymetrix CytoScan® 750K Array using manufacture procedure. Detailed clinical information was extracted from patient's cytogenetic request form and patient's medical records.

**Results:** Chromosomal aberration were identified in all patients using DNA microarray. Result showed deletion of chromosome 13q and 14q varied in size ranging 0.42MB to 39.97MB. Interestingly, there were also additional CAs in form of trisomy 3, cryptic interstitial deletion of chromosome 13q, 14q, 16p, Xp and duplication of 14q32.33 and 22q11.2. Deletion 14q32.33 region closely related to mutational status of *IGH* while other cryptic aberrations need further investigation to explore contribution to CLL pathogenesis. All the cryptic deletion and duplication of these chromosome unable to be detected by CCA due to small size (<5MB). The findings can help the clinicians to stratify patients accordingly and more close monitoring warranted.

**Conclusion:** DNA microarray are valuable tool in detecting relevant genomic aberrations in CLL patients especially cryptic abnormalities. Implementation of DNA microarray in CLL diagnosis will help clinicians to give precise treatment that tailored to patient's need.

# TABLE OF CONTENT

## LIST OF SYMBOLS/ABBREVIATION/NOMENCLATURE

**ALL** **:** Acute lymphocytic leukaemia

**AML** **:** Acute myeloid leukaemia

**BM** **:** Bone marrow

**BMAT** **:** Bone marrow aspiration and trephine biopsy

**CA/s** **:** Chromosomal aberration/s

**CCA** **:** Conventional cytogenetic analysis

**CLL** **:** Chronic lymphocytic leukaemia

**CLL-IPI** **:** Chronic lymphocytic leukaemia – International Prognostic Index

**Del/del** **:** Deletion

**DLBCL** **:** Diffuse large B cell lymphoma

**DNA** **:** deoxyribonucleic acid

**Dup/dup** **:** Duplication

**EDTA** **: E**thylene**d**iamine**t**etraacetic **a**cid

**FISH** **:** Fluorescence *in situ* hybridization

**IGH** **:** Immunoglobulin heavy locus

**iwCLL** **:** International Workshop on Chronic Lymphocytic Leukemia

**LN** : Lymph node

**lncRNA** **:** long non-coding RNA

**mRNA** **:** messenger RNA

**miRNA** **:** microRNA

**MLPA** **:** Multiplex Ligation-dependent Probe Amplification

**MZBCL** **:** Marginal zone B cell lymphoma

**NGS** **:** Next generation sequencing

**NHL** **:** Non-Hodgkin lymphoma

**p** **:** short arm of chromosome

**PB** **:** peripheral blood

**PCR** **:** polymerase chain reaction

**q** **:** long arm of chromosome

**RNA** **: R**ibo**n**ucleic **a**cid

**sIg** **:** surface immunoglobulin

**SNP** **:** single nucleotide polymorphism

**TTFT** **:** time to first treatment

# CHAPTER 1 : INTRODUCTION

## 1.1 INTRODUCTION

Chronic lymphocytic leukaemia (CLL) is a lymphoproliferative disorder defined by the proliferation of neoplastic mature B lymphocytes leading to their accumulation in the blood, bone marrow, lymph nodes and spleen (Hallek *et al.*,2019; Swerdlow *et al.*, 2016). CLL is the most frequent leukaemia in adults in Western countries, accounting for 5% to 11% of lymphoproliferative diseases. Every year, CLL affects about 4 to 6 people per 100,000 population. This disease primarily affects people in their sixties and seventies, with 70% of patients over 65 years old. According to America National Cancer Institute, approximately 3930 individuals died from CLL in 2019, accounting for 0.6% of cancer types. Males are more susceptible to CLL than females (M:F ratio of 1.5 - 2.1:1).

In contrast, this disease is infrequently observed in Asia, particularly in the Far East, and perpetuated by migrants and their descendants (Kawamata *et al.*, 2013). CLL accounted for just 1% to 3% of lymphoproliferative disorders in Asian nations. According to the Malaysian National Cancer Registry's 2007-2011 report, the total number of chronic lymphocytic leukaemia patients was 124 for those five years, implying that Malaysia has an average of 24.8 newly diagnosed CLL cases per year (Ministry of Health,2015).

Clinical manifestations of newly diagnosed CLL are highly diverse. CLL presents differently in each patient, ranging from a reasonably indolent disease with a

nearly average life expectancy to a quickly advancing disease with early death (Scarfo *et al.*, 2016). Most CLL cases are asymptomatic and are usually monitored until symptoms like cytopenia, lymphadenopathy, and splenomegaly occur. Some patients will develop an aggressive form of B lymphocyte malignancy, such as diffuse large B cell lymphoma (DLBCL) or, less frequently, Hodgkin lymphoma or another aggressive lymphoma (Hallek, 2019).

For decades, CLL was diagnosed using full blood picture with lymphocytes count more than $5x10^9$/L, marrow morphology, marrow immunophenotyping, marrow cytogenetics, as well as clinical evaluation to reveal nodal involvement. At the time of diagnosis, clinicians mostly use recognized risk stratification for CLL in the form of modified Rai classifications and/or the Binet classification. The modified Rai classifications focused on lymphocyte count, nodal involvement, and organ involvement, while the Binet classification focused on haemoglobin level, platelet count, and the number of nodal areas involved. However, the information gained from these classifications upon CLL diagnosis in patients will not forecast disease progression in each person.

Many factors that contribute to CLL heterogeneity have been discovered via research throughout the years (Karakosta *et al.*,2015). The development of CLL can be understood as a cooperative effort between risk factors and genetic abnormalities. Genetic factors play major role that influence the prognosis and progression of disease in CLL patient. Numerous studies have been conducted to uncover risk factors for CLL development however, no unique acquired factor for CLL development has been

established yet. There is compelling evidence that CLL can be caused by genetic predisposition (Stilgenbauer *et al.* 2002; Ripolles *et al.*, 2006). Family history of haematological malignancy (CLL and/or non-Hodgkin lymphoma (NHL)) is included as host factor. According to Slager *et al.* (2013), relatives of CLL patients have a two-to-eightfold increased chance of getting CLL and a two-fold increased risk of developing NHL compared to the general population. Goldin *et al.* corroborate this finding, stating that familial CLL was diagnosed at a younger age than random CLL. Additionally, there are also case reports involving familial CLL where two or more individual affected by CLL in same family.

While the incidence of CLL is higher in Western countries than in Asian countries, Asian patients have been shown to have a more aggressive disease progression and a shorter time to first treatment duration. This occurrence was speculated to occur due to the presence of distinct biomarkers and susceptibility factors in the Asian population. Numerous case reports demonstrated that Asian CLL patients have a few different chromosomal abnormalities than Western CLL patients (Kawamata *et al.*, 2013). The major copy number variations shared by Western and Asian CLL are del13q14, trisomy 12, deletion 17p, and deletion 11q (Lee *et al.*, 2019). Additionally, Kawamata *et al.* revealed that Asian CLL patients were more likely to have trisomy/duplication of 3q or trisomy 18/dup18q, but none of these chromosomal aberrations were observed in Western CLL patients. Another study conducted by Wu and his colleagues indicated that Asian CLL patients had a higher prevalence of tumor protein p53 (*TP53*) mutations than Western CLL patients (Wu *et al.*,2017).

When compared to the earlier modified Rai and Binet risk classification, genetic aberrations are taken into account in the most recent CLL risk stratification (CLL-IPI) which introduced by International Workshop on CLL in 2016. This inclusion demonstrated the need of early chromosomal aberration diagnosis in order to provide improvement to patient monitoring and treatment.

Cytogenomics can be defined as the study of the numerical and structural variation of the genome at the chromosomal and subchromosomal level as well as at molecular resolution using methods that cover the entire genome or specific DNA sequences (Hochstenbach, Liehr, and Hastings, 2020; Iourov,2019). It also evaluates chromosomes and their relation to disease (Shao,2017). Cytogenomic testing is not limited to conventional cytogenetic analysis and molecular cytogenomics methods i.e. *Fluorescence in situ Hybridization* (FISH), Polymerase Chain Reaction (PCR) or Multiplex Ligation-dependent Probe Amplification (MLPA), it also comprised of high-throughput cytogenomics technologies which includes applications of whole genome Copy Number Variation (CNV) analysis like DNA microarray.

Chromosomal abnormalities are found in up to 80% in CLL patients (Gadhia *et al.*,2019; Shao *et al.,* 2010). Among these, deletions of 11q, 13q, 17p, and trisomy 12 have been shown to have predictive significance and play an essential role in CLL pathogenesis and evolution, influencing patient outcome and therapy methods. Only a few studies have studied conventional cytogenetic analysis in CLL, in contrast to acute leukaemia or myelodysplastic syndromes due to the poor in vitro proliferative activity even in the presence of B-cell mitogens. This insufficient proliferative activity of

mature neoplastic B cells led to either the production of metaphases was completely inhibited or clonal aberrations were detected in 40–70% of cases due to low quality of the metaphases or because normal hematopoietic cells showed in vitro proliferation but CLL cells did not (Puiggros *et al.*,2014).

For the past ten years, Fluorescence *in situ* Hybridization (FISH) and Multiplex Ligation-dependent Probe Amplification (MLPA) have been used in addition to conventional cytogenetic analysis (CCA) to identify CLL chromosomal aberrations. Both have a high specificity and sensitivity for the diagnosis of genetic aberration in non-dividing cells. However, both approaches are limited to specific known genetic regions (Urbankova *et al.*,2014; Kolquist *et al.*, 2011; Patel *et al.*,2008). FISH testing is a relatively time and money-consuming procedure because each genetic aberration requires a different set of probes to be used. In addition, it is not able to identify any chromosomal abnormalities other than those at the known genomic locations that probes had identified (Alhourani *et al.*, 2014).

While MLPA can concurrently detect copy number variations, methylation pattern changes, and/or point mutations in several target locations (Homig-Holzel and Savola, 2012; Abdool *et al.*, 2010), it also has its own drawbacks. This method has is unable to identify copy neutral loss of heterozygosity, as well as having problem with mosaicism i.e. unable to get tumour heterogeneity and occasionally can cross-contaminate with normal cells (Stuppia *et al.*,2012).

In view of both methods limitations, genomic profiling utilising a microarray platform for CLL is becoming more popular as a diagnostic tool since it has a high-resolution array for examining cryptic aberrations in copy number variants. DNA microarray analysis bridges the gap between genome-wide low-resolution chromosome studies and loci-specific targeted FISH panels by offering genome-wide, high-resolution analysis that does not require cell culture or viable cells for testing (Peterson *et al*., 2018). In identifying aneuploidies, microdeletions, notably cryptic locus deletion and duplications, as well as amplification in CLL, microarray technique, particularly those whom utilising Copy number aberrations (CNA) +SNP chip technology, is superior. In addition, it has the capacity to detect additional confirmation of CNAs, copy-neutral loss of heterozygosity (CN-LOH), and various polyploidies. The use of microarray analysis into the cytogenetic diagnosis of haematological malignancies enhances patient care by providing doctors with extra information regarding possibly clinically relevant genomic changes (Peterson *et al.*, 2015).

This study aimed to investigate cytogenomic profiles among the newly diagnosed CLL patients at diagnosis and to stratify the patients based on their chromosomal aberrations and initial clinical presentation using the recent approach of DNA microarray. As there is a lack of data on chromosomal aberration in CLL patients in Malaysia due to the disease prevalence being lower than in Western counterparts and most patients being asymptomatic, this study will serve as a pilot study to collect data on chromosomal aberration in CLL patients using DNA microarray platform in Malaysia to help clinicians provide better treatment and counselling to patients and their relatives in explaining the disease and its risk on the relatives.

# CHAPTER 2: OBJECTIVES OF THE STUDY

## 2.1 GENERAL OBJECTIVE

To investigate cytogenomic profiling in CLL patients using DNA microarray approach.

## 2.2 SPECIFIC OBJECTIVES

1. To determine the genetic aberrations in CLL patients using CytoScan® 750K Array.
2. To identify candidate genes involved in cytogenomic regions with aberration in CLL patients.
3. To correlate the severity of initial presentation and staging of CLL with genomic aberration at diagnosis.

## CHAPTER 3: STUDY PROTOCOL

### 3.1 STUDY PROTOCOL

**JEPeM Code: USM/JEPeM/20060302**

**Research title: Cytogenomic Profiling of Chronic Lymphocytic Leukaemia Patients using DNA microarray.**

**Principal investigator:** Dr. Wan Norizzati Binti Wan Mohamad Zamri (MMC No: 64854)

**Co-researchers:** Assoc. Prof Dr. Sarina Sulong

Dr. Nazihah Mohd Yunus (MMC No: 41858)

Dr. Ahmad Aizat Abdul Aziz

**Introduction**

Chronic lymphocytic leukaemia (CLL) is a lymphoproliferative disorder which characterized by proliferation and accumulation of neoplastic mature B lymphocytes in the blood, bone marrow, lymph nodes and/or spleen (Swerdlow *et al.*, 2016). The clinical course of CLL is varied from patient to patient, ranging from a very indolent condition, with a nearly normal life expectancy to rapidly progressive leading to early death (Scarfo *et al.*, 2016). CLL is the most common leukaemia in Western countries especially in adults accounting for 5% - 11% of lymphoproliferative disease. Every year, CLL has an incidence rate of approximately ranging between 4 - 6 cases per 100,000. This disease mostly affected the older age group which 70% of cases reported to be older than 65 years old. According to America National Cancer Institute, approximately 3930 individuals die from CLL in 2019, accounting for 0.6% of all cancer type. Males are more predisposed to CLL compared to female (M:F ratio of 1.5 - 2.1:1). However, this disease is rarely reported in Asia especially at Far East and

maintained by migrants and their progeny (Kawamata *et al.*, 2013). In Asian countries, CLL constituted only 1% to 3% of lymphoproliferative disorder. Based on Malaysia National Cancer Registry Report 2007-2011 (Ministry of Health,2015), the total number of chronic lymphocytic leukaemia patients during those 5 years duration was 124 patients, which means on average, there is 24.8 of newly diagnosed CLL cases per year in Malaysia.

Although most of the CLL cases are asymptomatic and usually managed with watch and wait until development of symptoms i.e cytopenia, lymphadenopathy and splenomegaly, some of patients will transform into aggressive form of B lymphocytes malignancy such as diffuse large B cell lymphoma (DLBCL) or rarely transform into Hodgkin lymphoma or another type of aggressive lymphoma (Hallek, 2019).

In the era of precision medicine, detailed genomic data of patients contribute a lot in producing effective personalized treatment that tailor according to patients' need.

**Problem statement & Study rationale**

Genetic factors play major role that influence the prognosis and progression of disease in CLL patient. Host factor also can contribute to development of CLL. In latest classification of risk stratification of CLL (CLL-IPI) by International Workshop on CLL (iwCLL) ,genetic abnormalities are included into consideration compared to previous Rai and Binet risk classification. This inclusion proved the need of early chromosomal aberration detection to give better monitoring and treatment to patient.

Although in Asia, CLL are reported low incidence compared to Western countries, but the Asian CLL has been reported to have more aggressive clinical course and treatment outcome. It has been postulated that Asian CLL have different biological characteristic and some cases had different chromosomal abnormalities when compared to Western CLL (Gunawardana *et al.*, 2008).

In Hospital Universiti Sains Malaysia (Hospital USM), the newly diagnosed CLL patients will be subjected to marrow morphology assessment, marrow immunophenotying, marrow cytogenetic analysis, full blood picture and biochemical tests. As most of patients are asymptomatic, they are subjected to do full blood picture as a part of wait and watch management.

However, due to lack of database regarding common chromosome abnormalities in Malaysian CLL patient, the treatment given to CLL patients which progressed to aggressive form of B lymphocytes are used the Western CLL Protocol that may not suitable to our population because of different gene and chromosomal abnormalities.

This study is designed to identify the common cytogenomic abnormalities among Malaysian Chronic Lymphocytic Leukaemia using microarray genetic testing i.e CytoScan® 750K Array on Affymetrix platform. Even though 80% of CLL cases can be diagnosed using conventional cytogenetic analysis and fluorescence *in situ* hybridization, cryptic abnormalities of chromosomes can be missed (Shao *et al.*, 2010; Urbankova *et al.*, 2014). The microarray method can detect extra chromosomal aberration together with recurrent chromosomal abnormalities (Kolquist *et al.*, 2011).

Appropriate diagnosis and monitoring of patients are essential to ensure patient receive appropriate treatment and precise therapy in case disease progression. This study may serve as a pilot study to help in optimizing the treatment for CLL, more toward targeted therapies with the CLL cytogenomic profiling data produced.

**Research Question(s)**

1) Does CytoScan® 750K Array can detect the genomic abnormalities in CLL patients compared to conventional cytogenic analysis?

2) What are cytogenomic aberrations present in CLL?

3) Does genomic aberration contribute to severity of CLL during initial presention?

**Objective**

**General:**

To investigate cytogenomic profiling for CLL patients using microarray-based genomic profiling.

**Specific:**

1) To determine the genetic aberrations in CLL patients using CytoScan® 750K Array.
2) To identify candidate genes involved in cytogenomic aberration in CLL patients.
3) To correlate the severity of initial presentation and staging of CLL with genomic aberration at diagnosis.

**Literature review**

Chronic lymphocytic leukaemia (CLL) pathogenesis can be viewed as cooperation between patient's risk factor and genetic aberrations. There were several studies done in identifying risk factor for CLL development, however up to date, there still no specific acquired factor has been identified to disease development. However, there are strong evidence that genetic predisposition can lead to CLL (Ripolles *et al*., 2006; Stilgenbauer *et al.*, 2002). Host factor which included family history with haematological malignancy (CLL, and/or non – Hodgkin lymphoma (NHL)). The study done by Slager *et al.* revealed that relatives of CLL patients have a 2 –to – 8-fold increase in the risk of developing CLL and 2-fold increase NHL compared to general population (Slager *et al.*, 2013).This finding also supported by Goldin *et al.* which state that familial CLL was diagnosed at earlier age compared to sporadic CLL (Goldin *et al*., 2010). There are also case report involving familial CLL where two or more individual affected by CLL in same family. Latest study using genome wide association studies revealed more than 20 susceptibility loci which important in B lymphocytes and apoptotic pathway (Puiggros *et al.*, 2014).

Even though, the incidence of CLL in Western countries is higher than the Asian countries, the disease progression in Asian patient been reported more aggressive and shorter survival outcome. This event was postulated to happen due to different biomarker and susceptibility in Asian population. There were several case reports (Kawamata *et al.*, 2013), showed Asian CLL had reported a few different chromosomal aberrations than Western CLL. Western and Asian CLL shared the major copy number changes which are del13q14, trisomy 12, deletion 17p and deletion 11q (Lee *et al.*,

2019). Kawamata *et al.* also reported that Asian CLL patient more frequently to have either trisomy/duplication of 3q or trisomy 18/dup18q which none of these chromosomal aberrations were reported in Western CLL patients. Other study done by Wu and his teams revealed Asian CLL patient had high frequency of *TP53* mutation compared to Western CLL (Wu *et al.*, 2017).

For decades, diagnosis of CLL was done using full blood picture with the presence of lymphocytes more than $5x10^9$/L, examination of marrow morphology, marrow immunophenotyping, marrow cytogenetics and clinical examination to detect nodal involvement. However, for the past 10 years, the usage of Fluorescence *in situ* hybridization (FISH) and Multiplex Ligation-dependent Probe Amplification (MLPA) was used in diagnosing CLL. However, in spite the high sensitivity test for both methods, however, they are limited to specific known genomic loci (Patel *et al.*, 2008). Genomic profiling using microarray platform for CLL is increasing in trend for diagnostic purpose as it had high resolution array to examine the cryptic aberration in copy number variant.

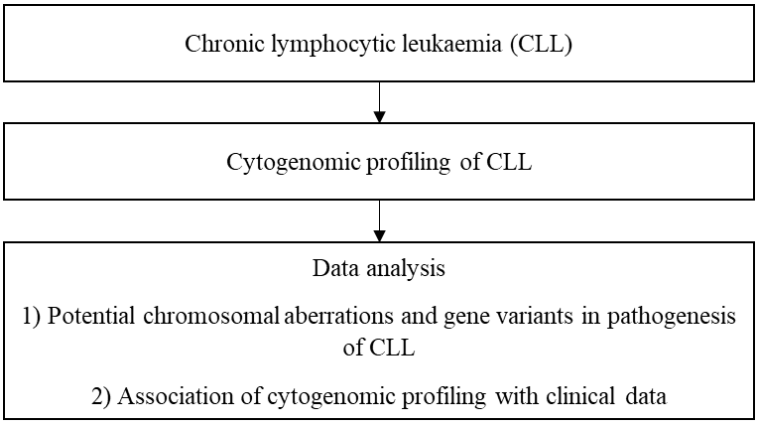
Clinicians mostly are using known risk stratification for CLL in form of Rai and modified Rai classification and/or Binet classification at early diagnosis. However, Leukemia and Lymphoma Society released more comprehensive prognostic index, also known as CLL International Prognostic Index (CLL-IPI) by which included known genetic abnormalities in their criteria. The CLL-IPI used as supplementary to the existing risk stratification of CLL.

Rai and modified Rai classification stressed more on the lymphocytes count, nodal involvement and organ involvement while Binet classification look more on haemoglobin level, platelet count and number of nodal areas involved. However, the information gain from these classification during diagnosis of CLL in patient will not able to predict the progression of disease in each individual. (Hallek, 2019).

The genetic aberration information collected from each CLL patient from conventional cytogenetic, FISH, DNA microarray and whole genome sequencing will help the clinician to tailor according to patient's need and reduce the complication of treatment on patient (Chun *et al*., 2018; Maciejewski and Mufti, 2008). The information may help to anticipate the progression of disease if the patient harbouring certain chromosomal aberrations such as del 17p and del 13q or monosomy 13/17 respectively.

As the data of chromosomal aberration of CLL patient in Malaysia is lacking due to disease prevalence is lesser compared to Western counterpart and most of patients are asymptomatic, this study will serve as pilot study to collect the data regarding chromosomal aberration in CLL patient in Malaysia in order to help clinician to give better treatment and counselling to patients and their relatives in explaining the disease and its risk on the relatives.

## Conceptual framework




## Research design

This study is a retrospective study to investigate the cytogenomic aberration in CLL using CytoScan® 750K Array.

## Study area

Hospital Universiti Sains Malaysia (Hospital USM), Kubang Kerian, Kelantan.

## Study population

Reference population – All CLL patients in Peninsular Malaysia.

Source population – All CLL patients who receiving treatment in Hospital USM.

Sampling frame : All CLL patients in Hospital USM who fulfilled the inclusion and exclusion criteria.

**Subject criteria**

Inclusion criteria

All patients who fulfil the diagnostic criteria for chronic lymphocytic leukaemia by International Workshop of Chronic Lymphocytic Leukemia (iwCLL) 2008 shall be included in this study.

The diagnostic criteria for chronic lymphocytic leukaemia defined by iwCLL are as follows:

1) Presence of lymphocytes at least 5 x $10^9$/L AND
2) Have peculiar immunophenotypic profile as detected in flow cytometry showing either;
    i) clonal light chain restriction (either kappa or lambda chain),
    ii) CD 5 expression ,
    iii) CD 23 expression ,
    iv) low level of CD 20, CD79b and surface immunoglobulin expression.

Exclusion criteria

1) Patients who underwent bone marrow transplant.
2) Patients who did not consent to participate.

**Vulnerability of subject**

There will be no issue of vulnerability as this study will be using only the archieved samples and no new information/data will be collected from the patients.

**Sample size estimation**

For specific objective 1, the sample size needed to determine the number of respondents needed for study recruitment using sample size calculator(version 2.0) (http://wnarifin.github.io). Using single proportion estimation as study design, the calculated sample size (n) is 274.

Input: p=80%, Precision = 5%, α = 0.05, drop-out=10%

p = prevalence of chromosomal aberration in CLL patient (Chun *et al*.,2018)

Precision = The probability of correctly rejecting the null hypothesis and Type I error probability α.

α = The Type I error probability for a two sided test. This is the probability that we will falsely reject the null hypothesis.

Drop-out = the possible number of subject that will withdraw their participance in the study.

However, as CLL is one of the leukaemia that underreported and undiagnosed at most times, the prevalence of CLL in Malaysia is 2.7% with total of 124 patients according to Malaysian National Cancer Registry Report 2007 -2011.

In view of this huge number of sample size and time limit constraint, we propose the method of CLL data analysis using CytoScan® 750K microarray platform as there is only one prior study done in Netherland using similar method. The study was done by Stevens-Kroef *et al.* was done on 28 CLL patients despite the prevalence of CLL disease in Europe is 10% according to 2017 European Subtype Report of Chronic Lymphocytic Leukaemia.

Here, this study will serve as pilot study for Malaysia population. This study will proceed with 6 sample as our CLL prevalence in one fourth of European CLL prevalence and also due to time and budget constraint.

For **specific objective 2 and specific objective 3,** no sample size required as this study is created to develop our own reference data for CLL using CytoScan® 750K microarray method.

**Sampling method and subject recruitment**

**Selection of potential subjects (available archived samples)**

Potential subjects will be screened from our previous study (USM Short Term grant title: Analysis of Hypermethylation Status of Tumor Suppressor Genes *p16INK4a, p15INK4b, ADAM12* and *PCDHGB7* in Chronic Lymphocytic Leukemia Patients and Normal Individuals)(JEPeM code: USMKK/PPP/JEPeM[234.3(07)). They can be also from a list of cytogenetic requests received at Human Genome Centre for cytogenetic analysis for CLL patients (if the cell pellets available). These potential subjects will be matched with this study inclusion and exclusion criteria.

**Data collection method**

1) Clinical information related to patient during initial presentation

   Relevant clinical information such as age, sex, race, occupation, clinical presentation at diagnosis, relevant laboratory results will be extracted from patient's request form for cytogenetic testing and also from patient's medical records. (Refer appendix: Data collection sheet) Data collection will be started after ethical approval. Once ethical approval is obtained and patients consented

to participate in the study, relevant data needed is obtained from patients' medical record.

The subjects will be labelled with study code to maintain their privacy and confidentiality.

**2) Cytogenomic analysis for CLL**

This study will be using archive genetic material (extracted DNA) from previous study: USM Short Term grant title: Analysis of Hypermethylation Status of Tumor Suppressor Genes *p16INK4a, p15INK4b, ADAM12* and *PCDHGB7* in Chronic Lymphocytic Leukemia Patients and Normal Individuals)(JEPeM code: USMKK/PPP/JEPeM[234.3(07)) and archive genetic material (cell pellets) from cytogenetic diagnostic test (Guidelines on retention of Pathology Records and Materials (part 1 (version 1/2015) will be used for this study. All archive genetic samples are stored in -20$^0$C in researcher's fridge in Human Genome Center Laboratory, Hospital University Sains Malaysia for privacy protection.

**Research tool**

**1) Cytogenomic profiling using DNA Microarray**

**I) DNA extraction**

DNA material was extracted from peripheral blood using commercialized QIAGEN QIAmp DNA Blood Mini Kit (QIAGEN, Hilden, Germany), following manufacturer's protocol during previous study (USM Short Term grant title: Analysis of Hypermethylation Status of Tumor Suppressor Genes *p16INK4a, p15INK4b, ADAM12* and *PCDHGB7* in Chronic Lymphocytic

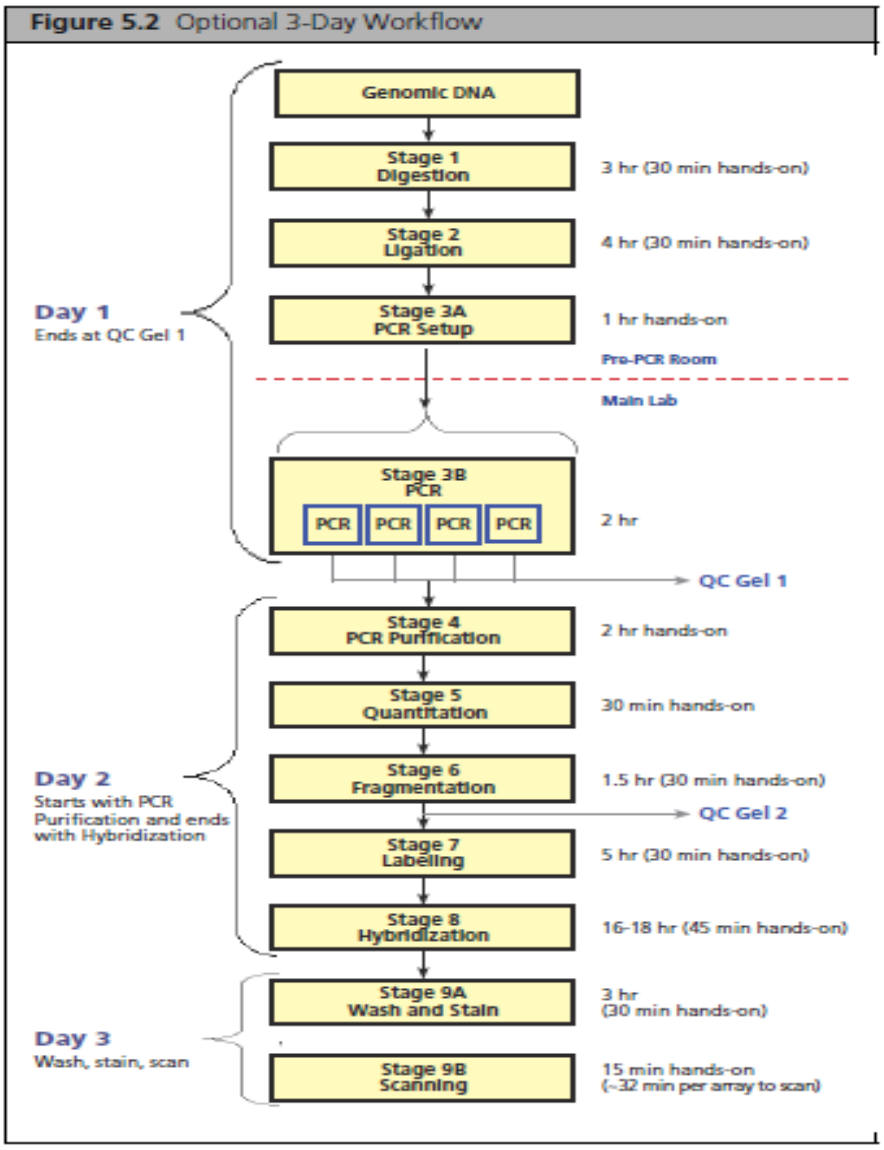
Leukemia Patients and Normal Individuals)(JEPeM code: USMKK/PPP/JEPeM[234.3(07)).The procedure QIAmp Blood Mini Kit is designed to extract the total genomic DNA without any prior separation of leukocytes from 200µl fresh whole blood which had been treated with EDTA. After extraction, the pure genomic DNA will be stored at -20$^{o}$C until used.

If archive cell pellets from bone marrow of CLL patient available, QIAGEN QIAmp DNA Blood Mini Kit (QIAGEN, Hilden, Germany), following manufacturer's protocol will be used to extract DNA material from the cell pellets.

**II) Qualitative and quantitative estimation of DNA**

Archive DNA material purity will be measured using NanoQuant , Qubit fluorometer and gel electrophoresis. The quantity of DNA material will be determined by Qubit fluorometer reading while the quality of DNA concentration will be determined by absorption ratio A260/280.

### III) Cytogenomic profiling data using CytoScan® 750K Array.




The genomic DNA will be analysed by CytoScan® 750K Affymetrix platform. The figure above has depicted a schematic laboratory workflow of CytoScan® 750K assay, The genomic DNA (250ng) will be digested by *Nsp1* and amplified using a ligation mediated PCR with adapters covalently linked to the restrictive fragment.

In the next step, PCR products will be purified using magnetic beads, fragmented using DNase1, labelled with biotin and hybridized overnight (17 hours) to a 49-format array.

After incubation, samples will be washed and stained with streptavidin using GeneChip Fluidic Station 450. Finally, arrays will be scanned by GeneChip Scanner 3000, using GeneChip Command Console Software (Thermo Fisher Scientific) to generate the CEL files that includes the intensity probe signal.

**IV) Data Analysis**

CEL files are analysed using the Chromosome Analysis Suite (ChAS) software (Thermo Fisher Scientific, Inc) and converted to CYCHP files containg information on copy number, loss of heterozygosity (LOH) and mosaicism cells.

From this study, we able to detect lost or gain of copy number of chromosomes, loss of heterozygozity and mosaic from microarray. However, this study unable to detect any reciprocal translocation, SNP genotype or gene expression due to 750K array limitation.

**2) Information on conventional cytogenetic analysis**

Data on cytogenetic analysis will be retrieved from cytogenetic database in Human Genome Centre USM. This data is collected and stored as per standard clinical indication for patients who underwent BMA for cytogenetic analysis.

Cytogenetic analysis was carried out as part of standard procedure in our laboratory for cytogenetic evaluation in CLL patients. For cytogenetic analysis, three (3)mls bone marrow sample were collected (as part of standard procedure for cytogenetic evaluation in CLL patients) in bone marrow transport media (containing 10% fetal bovine serum in RPMI media) and immediately subjected to overnight culture. The sample then harvested and put on slides for cytogenetic analysis. They were stained using GTG

banding. Metaphases were screened, analysed and the abnormalities detected were interpreted following International System for Human Cytogenomic Nomenclature (2016) based on standard operation procedure had been use in Cytogenetic Laboratory in Human Genome Centre in our centre.

**Operational definition**

**1)** Rai and modified Rai classification system calculates the CLL stage by 5 parameters namely level of total lymphocytes in blood/marrow, level of haemoglobin, level of platelets, presence of lymphadenopathy and presence of organomegaly i.e hepatomegaly or splenomegaly or both. This system uses these factors to divide CLL into 5 stages (Stage 0 – Stage 4) (Table 1).

**2)** Binet classification system (Table 2) calculates CLL stages by using 3 parameters namely level of haemoglobin, level of platelet and number of nodal areas involved. This system uses these factors to divide CLL into 3 stages.

**3)** CLL International Prognostic Index (CLL-IPI) combines genetic, biochemical and clinical parameters to categorize patients into four prognostic risk group. The independent prognostic group were identified and shown in Table 3.

## Table 1: Rai and modified Rai classification system

**Table 2. Rai and modified Rai classification system***

| Stage (Rai) | Description | Risk status (Modified Rai) | Median survival (years)† |
|---|---|---|---|
| 0 | Lymphocytosis, with lymphoid cells >30% in the blood and/or bone marrow | Low | 11.7 |
| I | Stage 0 with enlarged node(s) | Intermediate | 8.3 |
| II | Stage 0–1 with splenomegaly, hepatomegaly, or both | Intermediate | 5.8 |
| III | Stage 0–II with hemoglobin <110 g/L | High | 1.7 |
| IV | Stage 0–III with platelets <100 x $10^9$/L | High | 1.7 |

**Adapted from the 2008 IWCLL guidelines; 2008 NCI guidelines; BC Cancer Agency 2008 guidelines[3,5]*
*†These median survival estimates are based on earlier study data and do not take into account the revision of CLL diagnostic techniques and the improved efficacy of treatment. A recent retrospective study by Shanafelt, et al. examined median estimated survival times by Rai stage category in CLL patients from the Mayo Clinic patient database. Results showed that median survival times were not reached for low-risk, were approximately 10 years for intermediate-risk, and around 7 years for high-risk patients.[10]*

## Table 2: Binet classification system

**Table 2. Binet classification system*†**

| Stage | Description |
|---|---|
| A | Hemoglobin ≥10 g/dL and platelets ≥100,000/mm³ and <3 involved nodal areas |
| B | Hemoglobin ≥10 g/dL and platelets ≥100,000/mm³ and ≥3 involved nodal areas |
| C | Hemoglobin <10 g/dL and or platelets <100,000/mm³ and any number of involved nodal areas |

**Adapted from the 2008 NCI guidelines.[3]*
*†Areas of involvement considered for staging are as follows: (1) Head and neck, including the Waldeyer ring (this counts as one area, even if more than one group of nodes is enlarged). (2) Axillae (involvement of both axillae counts as one area). (3) Groins, including superficial femorals (involvement of both groins counts as one area). (4) Palpable spleen. (5) Palpable liver (clinically enlarged).*

## Table 3: CLL-IPI category

| CLL-IPI Category | Risk Score | Treatment Recommendations |
|---|---|---|
| Low Risk | 0-1 | Do not treat |
| Intermediate Risk | 2-3 | Do not treat unless the disease is highly symptomatic |
| High Risk | 4-6 | Treat unless the patient is asymptomatic |
| Very High Risk | 7-10 | If the decision is made to treat, use novel agents or treatment in a clinical trial rather than chemotherapy |