UNIVERSITI PUTRA MALAYSIA

MICROARRAY-BASED GENOMIC ANALYSIS IDENTIFIES GERMLINE AND SOMATIC COPY NUMBER VARIANTS AND LOSS OF HETEROZYGOSITY IN ACUTE MYELOID LEUKAEMIA

A ANGELI A/P AMBAYYA @ AMPIAH

FPSK(m) 23
MICROARRAY-BASED GENOMIC ANALYSIS IDENTIFIES GERMLINE AND SOMATIC COPY NUMBER VARIANTS AND LOSS OF HETEROZYGOSITY IN ACUTE MYELOID LEUKAEMIA

By

A ANGELE A/P AMBAYYA @ AMPIAH

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fullfilment of the Requirements for the Degree of Master of Science

November 2015
COPYRIGHT

All materials contained within the thesis, including without limitation text, logos, icons, photographs and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia
MICROARRAY-BASED GENOMIC ANALYSIS IDENTIFIES GERMLINE
AND SOMATIC COPY NUMBER VARIANTS AND LOSS OF
HETEROZYGOSITY IN ACUTE MYELOID LEUKAEMIA

By

A ANGELI A/P AMBAYYA @ AMPIAH

November 2015

Chair : Sabariah Md Noor, PhD
Faculty : Medicine and Health Sciences

Acute myeloid leukaemia (AML) is characterized by the overproduction of immature
myeloid cells that accumulate in blood and bone marrow. While the specific cause of
AML is usually unknown, several factors including chromosomal aberrations and
 genetic mutations have been implicated in the pathogenesis of this aggressive disease.
Integration of genetic findings and clinicopathological information is crucial in
establishing the diagnosis, prognosis and determining the therapeutic approach in the
management of AML patients. The AML classification has evolved from morphology
to cytogentic/molecular genetics-based findings in recent years. Cytogenetic
information is important in the detection of chromosomal abnormalities and has
provided the framework for the diagnosis and risk-stratification in AML over the past
decade. However, conventional cytogenetics is a technically demanding method. The
success rate of chromosomal analysis is largely dependent on the availability of
optimal and viable cells for culturing and the expertise with experience in identifying
chromosomal aberrations at a limited resolution. Insights into molecular karyotyping
using comparative genomic hybridization (CGH) and single nucleotide polymorphism
(SNP) arrays enable the identification of copy number variations (CNVs) at a higher
resolution and facilitate the detection of copy neutral loss of heterozygosity (CN-LOH)
otherwise undetectable by conventional cytogenetics. The applicability of a customised
CGH+SNP 180K DNA microarray with additional additional custom probes for 49
genes; every exon of eleven of these genes (TP53, DNMT3A, TET2, ASXL1, MLL,
IKZF1, PAX5, EZH2, FLT3, NOTCH1 and ATM) was covered in the diagnostic
evaluation of AML was assessed in this study. Paired tumour and germline (remission
sample obtained from the same patient after induction) DNA were used to delineate
germline variants in 41 AML samples. The prognosis based on karyotyping and
molecular genetics was correlated with demographic (age, gender, ethnicity) and
laboratory findings (WBC, aberrant antigen expression of CD2, CD4, CD7, CD19 and
CD56). After comparing the tumour versus germline DNA, a total of 55 imbalances (n
5-10 MB = 21, n 10-20 MB = 8 and n >20 MB = 26) were identified. Gains were most
common in chromosome 4 (26.7%) whereas losses were most frequent in chromosome
7 (28.6%) and X (25.0%). CN-LOH was mostly seen in chromosome 4 (75.0%).
Excellent agreements between the karyotype and CGH+SNP analyses were observed in
20 cases, with CGH+SNP analyses providing more precise breakpoint definition.
Karyotype was not in agreement with CGH+SNP in 13 cases. In another three cases,
array CGH+SNP detected aberrations which were missed by conventional karyotyping. Translocations were not detected by CGH+SNP in six cases. Correlation between prognosis on karyotyping and molecular genetics based on the clinical and laboratory findings showed statistically significant association between CD19 expression and a favourable prognosis. Statistically significant differences were observed between genders (P < 0.05 by Fisher’s exact test); females had a more favourable prognosis compared to males. Chromosomal abnormalities with breakpoint coordinates were identified more accurately as compared to conventional cytogenetics with the use of the combined array CGH+SNP platform in this study. In summary, a combined platform of CGH+SNP provides invaluable insights into the elucidation of large spectrum of genomic aberrations in AML which may have prognostic implications.
Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk Ijazah Master Sains

**PENEMUAN VARIAN COPY NUMBER DAN KEHILANGAN HETEROZYGOSITY GERMA DAN SOMATIK BAGI LEUKEMIA MYELOID AKUT MELALUI ANALISIS GENOMIK BERASASKAN MIKROARRAY**

Oleh

A ANGELI A/P AMBAYYA @ AMPIAH

November 2015

Pengerusi : Sabariah Md Noor, PhD
Fakulti : Perubatan dan Sains Kesihatan

Leukemia myeloid akut (AML) dicirikan oleh pengumpulan sel-sel myeloid yang tidak matang di dalam darah dan sum-sum tulang. Walaupun sebab khusus untuk AML kebanyakannya tidak diketahui, beberapa faktor seperti aberasi kromosom dan mutasi genetik diakitaikan dengan patogenesis penyakit agresif ini. Integrasi maklumat genetik dan klinikopatologikal adalah penting untuk membuat diagnosis, prognosis dan penentuan hala tuju rawatan terapeutik pesakit AML. Klasifikasi AML telah berevolusi daripada morfologi kepada berdasarkan sitogenetik dan genetik molekular sejak beberapa tahun kebelakangan ini. Maklumat sitogenetik adalah penting dalam pengesanan keabnormalan kromosom dan telah menjadi kerangka bagi diagnosis dan stratifikasi risiko AML sejak sedekad yang lalu. Walau bagaimanapun, teknik sitogenetik konvensional adalah rumit dan kadar kejayaan analisis kromosom bergantung kepada sel hidup yang optimum untuk pengkulturan serta memerlukan kemahiran yang tinggi dalam pengesanan aberasi kromosom pada resolusi yang terhad. Kariotip molecular menggunakan kaedah *comparative genomic hybridization* (CGH) dan *single nucleotide polymorphism* (SNP) membolehkan identifikasi *copy number variations* (CNVs) pada resolusi yang lebih tinggi dan membantu pengesanan of *copy neutral loss of heterozygosity* (CN-LOH) yang tidak dapat dikesan melalui kariotip konvensional. Penggunaan gabungan CGH+SNP 180K DNA mikroarray yang diubahsuai dgn tambahan 49 gen yang merangkumi sebelas gen setiap exon (TP53, DNMT3A, TET2, ASXL1, MLL, IKZF1, PAX5, EZH2, FLT3, NOTCH1 dan ATM) untuk evaluasi diagnostik AML telah dikaji. DNA tumor and germline ( sampel remission yang diperoleh daripada pesakit yang sama setelah induksi) dianalisis secara berpasangan untuk membezakan varian germline di dalam 41 sampel pesakit AML.Prognosis berdasarkan kariotip dan genetik molecular dikorelasi dgn klaimat klinikal (umur, jantina and kumpulan etnik) serta penemuan makmal ( bilangan sel darah putih, aberasi antigen CD2,CD4, CD7, CD19 dan CD56). Setelah membandingkan DNA tumor dengan germline, sebanyak 55 ketidakseimbangan dikesan (n 5-10 MB = 21, n 10-20 MB = 8 and n >20 MB = 26). Penambahan paling banyak berlaku pada kromosom 4 (26.7%) manakala kehilangan paling banyak didapati pada kromosom 7 (28.6%) dan X (25.0%). CN-LOH paling kerap dilihat pada kromosom 4 (75.0%). Penemuan kariotip dan CGH+SNP adalah bertepatan di dalam 20 kes, di mana CGH+SNP menunjukkan breakpoint yang lebih tepat. Penemuan kariotip dan CGH+SNP tidak bertepatan di dalam 13 kes. Array CGH+SNP telah
menunjukkan aberasi pada tiga kes yang tidak dapat dikesan menggunakan kariotip konvensional. Seperti yang sedia maklum, translokasi tidak dapat dikesan menggunakan CGH+SNP seperti yang dilihat berlaku ke atas enam kes. Korelasi telah berdasarkan kariotip dan genetik molekular dengan penemuan klinikal dan makmal menunjukkan hubungan yang signifikan secara statistik di antara expresi CD19 dan kumpulan telahan memuaskan. Perbezaan signifikan secara statistik dilihat di antara jantina ((P < 0.05 by ujian Fisher’s exact); perempuan lebih banyak di kumpulan telahan memuaskan manakala lelaki lebih banyak di kumpulan telahan pertengahan. Keabnormalan kromosom dengan koordinat yang lebih tepat telah berjaya dikesan menggunakan gabungan array CGH+SNP di dalam kajian ini. Kesimpulannya, gabungan array CGH+SNP memberikan pencerahan untuk merungkai spektrum aberasi genomik AML yang berkemungkinan memberikan implikasi terhadap prognosis.
ACKNOWLEDGEMENTS

First and foremost I would like to thank my Lord and saviour Jesus Christ for His guiding hands throughout my life. Without Him, nothing is possible in my life.

Immeasurable appreciation and heartfelt gratitude to the following people who had been there in one way or another throughout my research journey.

**Dr Sabariah Md Noor**, for advice, support, guidance, and suggestion to improve this research from day one. Thank you for the trust you had in me, it meant a lot to me. There were times when I was rather quiet performing experiments and you trusted me so much without any doubt that I was doing my research work even in my silence. It was definitely a blessing to have someone who trusted me implicitly. **Dr Zainina Seman**, for always being there, supporting me and for all your ideas and feedback. Thank you for your time and encouragement.

**Dato Dr Chang Kian Meng** and **Dr Subramanian Yegappan**, words are never good enough to thank you both. Thank you so much for being wonderful bosses, for the encouragements and understanding. Thank you for the help during research grant application, also for being there in MOH’s proposal and grant defence. Dr Mani, thank you for your patience in answering all my endless questions. And again, words will never be good enough to express my gratitude to both of you and the Department of Haematology, Hospital Ampang for all the support.

**Dr Lim Soo Min**, thank you for being our co-investigator right when I needed more case for my studies. That was a tough moment because most of my subjects passed away before I could obtain their remission after induction samples and you willingly helped me out.

**My parents**, for your love, understanding and upholding me in your prayers. Thank you mummy for staying awake with me during my thesis write up, I am blessed beyond measure to have such a loving mum. **My friends and family**. Thank you for the encouragements and affection.

**Fellow colleagues in Haematology Department**, sincere appreciation for always giving me priority in using laboratory facilities and equipment though I was on my study leave.

Last but not least, I would like to thank the Malaysian government for giving me the opportunity to pursue my studies with scholarship and National Institute of Health for the research grant. I am overwhelmed with vivid memories of kindness of the people who were there during my research journey and may God bless you all.
I certify that a Thesis Examination Committee has met on 16 November 2015 to conduct the final examination of A Angeli a/p Ambayya @ Ampiah on her thesis entitled "Microarray-Based Genomic Analysis Identifies Germline and Somatic Copy Number Variants and Loss of Heterozygosity in Acute Myeloid Leukaemia" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science.

Members of the Thesis Examination Committee were as follows:

**Eusni Rahayu binti Mohd Tohit, PhD**
Senior Lecturer
Faculty of Medicine and Health Science
Universiti Putra Malaysia
(Chairman)

**Abhimanyu a/l Veerakumarasivam, PhD**
Senior Lecturer
Faculty of Medicine and Health Science
Universiti Putra Malaysia
(Internal Examiner)

**N. Veera Sekaran a/l Nadarajan, PhD**
Associate Professor
University of Malaya
Malaysia
(External Examiner)

---

**ZULKARNAIN ZAINAL, PhD**
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 15 December 2015
This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

**Sabariah Md Noor, M.Path.**
Senior Lecturer  
Faculty of Medicine and Health Sciences  
Universiti Putra Malaysia  
(Chairman)

**Zainina Seman, M. Path.**
Senior Lecturer  
Faculty of Medicine and Health Sciences  
Universiti Putra Malaysia  
(Member)

**Subramanian Yegappan, MBBS**  
Consultant Haematopathologist  
Department of Haematology  
Hospital Ampang  
(Member)

**Chang Kian Meng, MRCP, FRCP, FRCPA**  
Head and Consultant Haematologist  
Department of Haematology  
Hospital Ampang  
(Member)

**BUJANG BIN KIM HUAT, PhD**  
Professor and Dean  
School of Graduate Studies  
Universiti Putra Malaysia  

Date:
Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _______________________   Date: __________________

Name and Matric No.:  A Angeli a/p Ambayya @ Ampiah (GS33730)
Declaration by Members of Supervisory Committee

This is to confirm that:
- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ____________________
Name of Chairman of Supervisory Committee: Sabariah Md Noor, M.Path.

Signature: ____________________
Name of Member of Supervisory Committee: Zainina Seman, M.Path.

Signature: ____________________
Name of Member of Supervisory Committee: Subramanian Yegappan, MBBS

Signature: ____________________
Name of Member of Supervisory Committee: Chang Kian Meng, MRCP, FRCP, FRCPA
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRAK</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>v</td>
</tr>
<tr>
<td>APPROVAL</td>
<td>vi</td>
</tr>
<tr>
<td>DECLARATION</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiv</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xvi</td>
</tr>
</tbody>
</table>

## CHAPTER

### 1 INTRODUCTION

### 2 LITERATURE REVIEW

2.1 Leukaemia
2.2 Types of Leukaemia
2.3 Incidence of Acute Leukaemia
2.4 Aetiology of Leukaemia
2.5 Leukaemogenesis
2.6 Clinical Presentation
2.7 AML Diagnosis
2.8 Genome wide analysis using array based techniques
  2.8.1 Comparative Genomic Hybridization (CGH)
  2.8.1.1 Array CGH Protocol
  2.8.2 Single Nucleotide Polymorphism (SNP)
  2.8.3 Single Nucleotide Polymorphism (SNP) Array
    2.8.3.1 Copy Neutral Loss of Heterozygosity (CN-LOH)
  2.8.4 CGH+ SNP in a single array
  2.8.5 Limitations of SNP array and CGH array
2.9 Matched Normal Genomic DNA
2.10 Studies on AML using array CGH and SNP Array
  2.11 Classification of Leukaemia
    2.11.1 Earlier Classification of AML
    2.11.2 WHO Classification of Leukaemia
2.12 Prognostic factors
2.13 Treatment in leukaemia

### 3 METHODOLOGY

3.1 Study Design
3.2 Type of Study
  3.2.1 Sample size calculation
3.3 Ethical consideration
3.4 Data collection
3.5 DNA Extraction
3.5.1 gDNA Quantitation and Quality Analysis
   3.5.1.1 NanoDrop ND-1000 UV-VIS Spectrophotometer
   3.5.1.2 Fluorometry
   3.5.1.3 Gel Electrophoresis
3.6 CGH+SNP Genomic DNA Analysis Protocol
3.7 CGH+SNP Analyses
   3.7.1 Paired tumour and germline DNA analyses
   3.7.2 Exclusion of germline lesions for International System for Human Cytogenetic Nomenclature (ISCN), 2013
   3.7.3 Exon resolution gene analysis
3.8 Conventional Karyotyping, FISH Testing, Molecular and Immunophenotyping
   3.8.1 Karyotyping and Molecular Genetics
   3.8.2 Flow cytometry Immunophenotyping
3.9 Association of prognosis with clinical and other laboratory findings
3.10 Statistical Analyses

4 RESULTS
4.1 Sample Collection
4.2 DNA integrity, quality and purity
4.3 CGH+SNP Sample Processing
   4.3.1 Sample labelling and QC Metrics
4.4 Data analysis using Cytogenomics software
4.5 Consideration in data interpretation
   4.5.1 Percentage of leukaemic cells and clonal fraction of Tumour DNA in CGH+SNP analysis
   4.5.2 Gains in the telomeric regions of chromosome 14q32.33
   4.5.3 Region of losses and gains in chromosome 9p13.1-p11.2
   4.5.4 Region of undulating “waves” in chromosome 16q21
   4.5.5 Delineation of CN-LOHs from losses in chromosomes
4.6 Paired tumour and germline DNA analysis
   4.6.1 Somatic Genomic Aberrations
   4.6.2 Germline Only Genomic Aberrations
4.7 Findings in the additional custom probes
4.8 Exclusion of Germline Lesions in the CGH+SNP Analysis
4.9 Comparison of CGH+SNP Analysis and Karyotyping
4.10 CGH+SNP Findings and Other Laboratory Findings of the Subjects
4.11 Correlation with clinical and other laboratory findings

5 DISCUSSION
5.1 Distribution of subjects in this study
5.2 DNA Quality and Sample Processing
5.3 Data Analysis
5.4 Array CGH+SNP findings in this study
5.5 Comparison of CGH+SNP Findings and Karyotyping
5.6 Findings from this study, comparison with other studies
5.7 Frequent and recurrent genomic aberration detected in this study
5.8 Prognosis and association with other clinic-biological parameters

6 CONCLUSION AND RECOMMENDATIONS FOR FUTURE STUDIES

REFERENCES 84
APPENDICES 94
BIODATA OF STUDENT 150
LIST OF PUBLICATIONS 151
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Cytogenetics classification of AML</td>
</tr>
<tr>
<td>2.2</td>
<td>WHO classification of acute myeloid leukaemia (2008)</td>
</tr>
<tr>
<td>2.3</td>
<td>Prognostic Groups in AML</td>
</tr>
<tr>
<td>3.1</td>
<td>Custom Agilent SurePrint G3 CGH+SNP 180K Design</td>
</tr>
<tr>
<td>3.2</td>
<td>Additional probes for custom CGH+SNP array (genes covered at every exon)</td>
</tr>
<tr>
<td>3.3</td>
<td>Additional probes for custom CGH+SNP array</td>
</tr>
<tr>
<td>4.1</td>
<td>Characteristic of the samples collected in this study</td>
</tr>
<tr>
<td>4.2</td>
<td>Characteristics of included subjects</td>
</tr>
<tr>
<td>4.3</td>
<td>Laboratory findings of included subjects</td>
</tr>
<tr>
<td>4.4</td>
<td>Summary of Quality Control for Tumour and Germline DNA</td>
</tr>
<tr>
<td>4.5</td>
<td>Summary of percentage of leukaemic cells in each subject and clonal fraction generated by Cytogenomics software</td>
</tr>
<tr>
<td>4.6</td>
<td>Total chromosomal aberrations and types seen in tumour</td>
</tr>
<tr>
<td>4.7</td>
<td>Recurrent genomic aberrations detected in tumour</td>
</tr>
<tr>
<td>4.8</td>
<td>Chromosomes and types of aberrations detected in germline only</td>
</tr>
<tr>
<td>4.9</td>
<td>Additional custom probes findings</td>
</tr>
<tr>
<td>4.10</td>
<td>CGH+SNP findings and subjects with normal karyotype</td>
</tr>
<tr>
<td>4.11</td>
<td>CGH+SNP findings and subjects with abnormal karyotype</td>
</tr>
<tr>
<td>4.12</td>
<td>CGH+SNP findings and subjects with no analysable chromosome spread</td>
</tr>
<tr>
<td>4.13</td>
<td>CGH+SNP findings and subjects with cases with no karyotype available</td>
</tr>
<tr>
<td>4.14</td>
<td>Prognosis groups of the subjects</td>
</tr>
<tr>
<td>5.1</td>
<td>Comparison of findings from various studies with this study</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Mutation categories reported by Cancer Genome Atlas Research Network organized into groups of associated genes</td>
<td>7</td>
</tr>
<tr>
<td>2.2</td>
<td>Oligo aCGH detected additional complexity missed by BAC aCGH</td>
<td>10</td>
</tr>
<tr>
<td>2.3</td>
<td>CGH gDNA Labelling and Hybridization</td>
<td>11</td>
</tr>
<tr>
<td>2.4</td>
<td>Schematic view of Chromosome View and CGH Data Output</td>
<td>12</td>
</tr>
<tr>
<td>2.5</td>
<td>Clonal Heterogeneity Indicators</td>
<td>13</td>
</tr>
<tr>
<td>2.6</td>
<td>SNP Array Workflow and Interpretation</td>
<td>15</td>
</tr>
<tr>
<td>2.7</td>
<td>Schematic view of CN-LOH in the q arm of chromosome 11</td>
<td>16</td>
</tr>
<tr>
<td>2.8</td>
<td>Partial karyogram of chromosome 17 and X from case 7</td>
<td>20</td>
</tr>
<tr>
<td>2.9</td>
<td>Prognostic groups in younger adults and the frequency</td>
<td>24</td>
</tr>
<tr>
<td>3.1</td>
<td>Study Design</td>
<td>27</td>
</tr>
<tr>
<td>3.2</td>
<td>Array CGH+SNP Workflow for Tumour DNA</td>
<td>32</td>
</tr>
<tr>
<td>3.3</td>
<td>Array CGH+SNP Workflow for Germline DNA</td>
<td>33</td>
</tr>
<tr>
<td>4.1</td>
<td>Distribution of subjects according to age groups of below 60 and above 60</td>
<td>38</td>
</tr>
<tr>
<td>4.2</td>
<td>Gel Image of electrophoretic separation of samples using TapeStation</td>
<td>40</td>
</tr>
<tr>
<td>4.3</td>
<td>Electrophoretogram of Sample S01 using TapeStation</td>
<td>40</td>
</tr>
<tr>
<td>4.4</td>
<td>Electrophoretogram of Sample S01, S02 and S03 showing regions of digested DNA</td>
<td>41</td>
</tr>
<tr>
<td>4.5</td>
<td>Spot finding of the four corners of the array four spots</td>
<td>42</td>
</tr>
<tr>
<td>4.6</td>
<td>Histograms of signals plot (red) displaying the signal level and the distribution of the signal</td>
<td>42</td>
</tr>
<tr>
<td>4.7</td>
<td>Histograms of signals plot (green) displaying the signal level and the distribution of the signal</td>
<td>43</td>
</tr>
<tr>
<td>4.8</td>
<td>Spatial distributions of the positive and negative log/ratios</td>
<td>43</td>
</tr>
<tr>
<td>4.9</td>
<td>Linear normalization</td>
<td>44</td>
</tr>
<tr>
<td>4.10</td>
<td>Chromosome 7 of subject S30</td>
<td>45</td>
</tr>
<tr>
<td>4.11</td>
<td>Chromosome 4 of Subject S08</td>
<td>46</td>
</tr>
<tr>
<td>4.12</td>
<td>Chromosome 12 of subject S09</td>
<td>47</td>
</tr>
<tr>
<td>4.13</td>
<td>Gain in 14q32.33 as seen in a male patient (S29) and a deletion in the Agilent male reference DNA at the same region.</td>
<td>50</td>
</tr>
<tr>
<td>4.14</td>
<td>Loss in chromosome 9p13.1-p11.2 seen in case S29</td>
<td>52</td>
</tr>
<tr>
<td>4.15</td>
<td>Loss in chromosome 9p13.1-p11.2 seen in case S06</td>
<td>52</td>
</tr>
<tr>
<td>4.16</td>
<td>Reproducible regions of equal copy numbers with small variations with an increase in the average CGH log ratio (AvgCGHLR) seen in tumour DNA of patient S06</td>
<td>54</td>
</tr>
<tr>
<td>4.17</td>
<td>Reproducible regions of equal copy numbers with small variations with an increase in the average CGH log ratio (AvgCGHLR) seen in germline DNA of patient S09</td>
<td>55</td>
</tr>
<tr>
<td>4.18</td>
<td>Areas in green shades depicts regions of CN-LOH seen in patient S26 where the 2n copy numbers retained but heterozygosity was lost</td>
<td>57</td>
</tr>
</tbody>
</table>
4.19 Areas in green shades depicts regions LOH seen in patient S30 where loss of copy numbers seen in CGH data and supported by SNP data

4.20 Types of aberrations and sizes present in tumour

4.21 Types of aberrations and sizes present in germline only

4.22 Heatmap depicting prognosis based on molecular and cytogenetic findings and other laboratory findings in this study

4.23 Prognosis groups according to gender

4.24 Prognosis groups in below 60 group
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>aCGH</td>
<td>array CGH</td>
</tr>
<tr>
<td>ACMG</td>
<td>American College of Medical Genetics</td>
</tr>
<tr>
<td>ADM-2</td>
<td>Aberration Detection Method-2</td>
</tr>
<tr>
<td>AML</td>
<td>Acute myeloid leukaemia</td>
</tr>
<tr>
<td>AML</td>
<td>Acute myeloid leukaemia</td>
</tr>
<tr>
<td>AML M0</td>
<td>Undifferentiated acute myeloblastic leukemia</td>
</tr>
<tr>
<td>AML M1</td>
<td>Acute myeloblastic leukemia with minimal maturation</td>
</tr>
<tr>
<td>AML M2</td>
<td>Acute myeloblastic leukemia with maturation</td>
</tr>
<tr>
<td>AML M3</td>
<td>Acute promyelocytic leukemia (APL)</td>
</tr>
<tr>
<td>AML M4</td>
<td>Acute myelomonocytic leukemia</td>
</tr>
<tr>
<td>AML M5</td>
<td>Acute monocytic leukemia</td>
</tr>
<tr>
<td>AML M6</td>
<td>Acute erythroid leukemia</td>
</tr>
<tr>
<td>AML M7</td>
<td>Acute megakaryoblastic leukemia</td>
</tr>
<tr>
<td>AML-NK</td>
<td>Normal karyotype AML</td>
</tr>
<tr>
<td>AvgCGHLR</td>
<td>Average CGH log ratio</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial artificial chromosomes</td>
</tr>
<tr>
<td>BM</td>
<td>Bone marrow</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>CD</td>
<td>Cluster of differentiation</td>
</tr>
<tr>
<td>CGH</td>
<td>Comparative genomic hybridization</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin ImmunoPrecipitation</td>
</tr>
<tr>
<td>Chr</td>
<td>Chromosome</td>
</tr>
<tr>
<td>CN</td>
<td>Copy number</td>
</tr>
<tr>
<td>CN-LOH</td>
<td>Copy neutral loss of heterozygosity</td>
</tr>
<tr>
<td>CNV</td>
<td>copy number variation</td>
</tr>
<tr>
<td>CR</td>
<td>Complete remission</td>
</tr>
<tr>
<td>Cy3</td>
<td>Cyanine 3-dUTP</td>
</tr>
<tr>
<td>Cy5</td>
<td>Cyanine 5-dUTP</td>
</tr>
<tr>
<td>dCTP</td>
<td>dCTP-coupled fluorophores with dUTP-c</td>
</tr>
<tr>
<td>DFS</td>
<td>disease-free survival</td>
</tr>
<tr>
<td>DFS</td>
<td>Disease-free survival</td>
</tr>
<tr>
<td>DGV</td>
<td>Database of Genomic Variants</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>DIC</td>
<td>disseminated intravascular coagulation</td>
</tr>
<tr>
<td>DLRSD</td>
<td>Derivative Log2 Ratio Standard Deviation</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dUTP</td>
<td>Deoxyuridine triphosphate</td>
</tr>
<tr>
<td>FFPE</td>
<td>Formalin fixed paraffin embedded</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence in situ Hybridization</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>gDNA</td>
<td>gDNA</td>
</tr>
<tr>
<td>GRCh37</td>
<td>Genome Reference Consortium 37</td>
</tr>
<tr>
<td>HSC</td>
<td>Haematopoietic stem cell</td>
</tr>
<tr>
<td>ISCN</td>
<td>International System for Human Cytogenetic Nomenclature</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>LSC</td>
<td>Leukaemic stem cells</td>
</tr>
<tr>
<td>Mb</td>
<td>Megabase</td>
</tr>
<tr>
<td>MDS</td>
<td>Myelodysplastic syndrome</td>
</tr>
<tr>
<td>miRNA</td>
<td>micro RNA</td>
</tr>
<tr>
<td>MLPA</td>
<td>Multiplex ligation-dependent probe amplification</td>
</tr>
<tr>
<td>MPO</td>
<td>Myeloperoxidase</td>
</tr>
<tr>
<td>MRD</td>
<td>Minimal residual disease</td>
</tr>
<tr>
<td>NSE</td>
<td>Non-specific esterase</td>
</tr>
<tr>
<td>OS</td>
<td>Overall survival</td>
</tr>
<tr>
<td>PAC</td>
<td>P1-derived artificial chromosomes</td>
</tr>
<tr>
<td>PAS</td>
<td>Periodic acid Schiff</td>
</tr>
<tr>
<td>PB</td>
<td>Peripheral Blood</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>QC</td>
<td>Quality control</td>
</tr>
<tr>
<td>qPCR</td>
<td>quantitative PCR</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcription polymerase chain reaction</td>
</tr>
<tr>
<td>SBB</td>
<td>Sudan Black B</td>
</tr>
<tr>
<td>SEER</td>
<td>Surveillance Epidemiology and End Results</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>SNP-A</td>
<td>SNP arrays</td>
</tr>
<tr>
<td>SPSS</td>
<td>Statistical Package of Social Science</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>tCGH</td>
<td>Translocation comparative genomic hybridization</td>
</tr>
<tr>
<td>UCSC</td>
<td>University of California, Santa Cruz</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>UPD</td>
<td>Uniparental disomy</td>
</tr>
<tr>
<td>US</td>
<td>United States</td>
</tr>
<tr>
<td>WBC</td>
<td>White blood cell</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>YAC</td>
<td>Yeast artificial chromosome</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Acute myeloid leukaemia (AML) is a heterogeneous malignant haematopoietic disorder that is characterized by an increase in immature myeloid cells. It is a result of an arrest of normal cell differentiation in the bone marrow. AML is also known as a disease marked by heterogeneity in diagnosis, classification, response to therapy and survival. Recent insights into the genomic landscapes of AML have led to tremendous advancement in understanding the molecular pathogenesis of this disease. Currently, the state-of-the-art in the diagnosis of AML relies on the integration of clinicopathological findings which include morphologic assessment, immunophenotyping analysis, and genetic studies.

Over the past decade, karyotyping of a minimum of 20 metaphase cells has been considered imperative in the classification and prognostication of AML (Simons et al., 2012). Karyotypically, AML can be stratified into three risk based categories: favourable, intermediate and unfavourable based on specific genetic abnormalities detected by conventional cytogenetics and/or fluorescence in situ hybridization (FISH) (Foran, 2010; Kadia et al., 2014). Cytogenetic information is crucial in identifying translocations, inversions, duplications, deletions and aneuploidies in order to establish diagnosis, prognosis and adjusting specific therapies (Akagi et al., 2009; Le Scouarnec & Gribble, 2011).

Despite providing a genome-wide chromosomal assessment, there are several drawbacks in conventional cytogenetics, in terms of sampling and technical issues. The major issue is the requirement of viable cells that necessitates proper and expedited sample processing. Thus sample quality is critical for this method. Bone marrow is preferable to peripheral blood and the first draw of marrow is recommended whenever possible for karyotyping. Peripheral blood may only yield informative result if the blast counts are higher than 10-20 %. The challenges also lie in dissecting the complex genetic changes due to poor chromosome morphology and indistinct banding. Careful selection of the best metaphases on a slide is likely to bias the analysis towards cell s with a normal karyotype and subsequently missing the abnormal clone(s) which could be potentially important contributor to AML pathogenesis (Maciejewski et al., 2009; Vermeesch et al, 2012; Simons et al., 2012).

The accuracy of karyotyping is largely dependent on the number of metaphase and chromosomal banding resolution. Optimal analysis is not possible for the cases with poor metaphases as the location of the chromosomal lesion cannot be clearly defined resulting in underestimation of the degree of chromosomal changes. Next, conventional cytogenetics require intricate and laborious procedures which involve cell cultures that are tailored to specific cell types. It is also technically demanding as leukaemic cells react differently to various stimuli and this requires more
than one culture. This is due to the varied sensitivity of tumour cells to culture conditions and synchronization procedures. In addition, neoplastic cells are often affected by low mitotic index caused by altered cell kinetics. Leukaemic blast are inclined to undergo apoptosis in culture thus hampering efforts to elucidate chromosomal aberrations. Expertise and meticulous observation are needed to distinguish various random losses from actual aberrations due to technical issues in sample processing (Maciejewski et al., 2009; Vermeesch et al., 2012; Simons et al., 2012; Eklund, 2010).

Progresses in molecular cytogenetics techniques in the last decade have enabled the interrogation of the AML genomic knowledge. The advent of molecular cytogenetics using comparative genomic hybridization (CGH) and single nucleotide polymorphism (SNP) microarrays have permitted comprehensive genome-wide assessments at resolutions higher than conventional cytogenetics. Molecular karyotyping enables the elucidation of genetic alterations that may have a significant role in the pathogenesis of AML and could lead to better stratification of diagnosis and prognosis. Chances are that tailored or optimized therapy based on genetic findings could pave the way for improvement in responses to treatment, disease-free survival (DFS) and overall-survival rates (OS).

In contrast to conventional cytogenetics, molecular karyotyping does not depend on mitotically dividing cells, as genomic DNA is used instead of metaphase chromosomes (Heinrichs et al., 2010; Vermeesch et al., 2012; Simons et al., 2012; Heinrichs & Look, 2007). For resolution, molecular karyotyping allows the detection of genomic lesions of ~ 400 kb in size, this surpassing conventional cytogenetics (3-5 Mb) (Wiznieswka et al., 2014). Microarray platforms facilitate customized probes that are designed down to a single exon resolution permitting detection of submicroscopic genetic lesions including microduplications and microdeletions that may be clinically relevant. Moreover, all DNA, tumour and non-tumour is represented and so there is no selection or bias. Besides copy number variations (CNVs), SNP – array facilitates the detection of copy neutral loss of heterozygosity (CN-LOH), also known as uniparental disomy (UPD), otherwise undetectable by conventional cytogenetics (Heinrichs et al., 2010; O’Keefe et al., 2010; Vermeesch et al., 2012; Maciejewski et al., 2009; Heinrichs & Look, 2007).

Therefore, molecular karyotyping using a combination of CGH+SNP DNA microarray can complement conventional cytogenetics not only in the diagnosis but also in the classification and prognostication of AML. In addition, the discovery of cryptic chromosomal aberrations and novel disease related to genomic regions is possible through the utilization of CGH+SNP DNA microarray in a clinical setting.
Objectives of this study

Main Objective
The major objective of this study is to identify somatically-acquired genetic aberrations and their clinical association or significance in AML by using a CGH+SNP DNA microarray platform.

Specific Objectives
This study embarked with the following specific objectives:

I. to detect chromosomal aberrations (CNVs and CN-LOHs) on a genome-wide scale using CGH+SNP DNA microarray in AML
II. to delineate somatic related variants from germline variants in AML
III. to identify recurrent genomic aberrations in AML
IV. to compare the findings from karyotyping and that from CGH+SNP DNA microarray analysis
V. to correlate the prognosis based on karyotyping and molecular genetics with other clinical (age, gender, ethnicity) and laboratory findings (WBC, aberrant antigen expression of CD2, CD4, CD7, CD19 and CD56)

The hypotheses of this study are:
1. CGH+SNP DNA microarray will enable the detection of submicroscopic chromosomal aberrations at a higher resolution as compared to conventional cytogenetics method
2. by comparing tumour versus normal DNA, it would be possible to delineate somatically-acquired genetic aberrations from germline variants in AML
3. CGH+SNP DNA microarray will enable the elucidation of new regions of recurrent genetic aberrations including CNVs and CN-LOHs
REFERENCES


Raghavan, M., Smith, L.L., Lillington, D.M., Chaplin, T., Kakkas, I., Molloy, G., Chelala, C., Cazier, J.B., Cavenagh, J.D., Fitzgibbon, J., Lister, T.A., Young,


92


