UNIVERSITI PUTRA MALAYSIA

IDENTIFICATION AND ANALYSIS OF SINGLE NUCLEOTIDE POLYMORPHISMS IN MATRIX METALLOPEPTIDASE 2 AND 3 GENES IN MALAYSIAN BREAST CANCER PATIENTS

CHAN SOON CHOY

FPSK(p) 2013 5
IDENTIFICATION AND ANALYSIS OF SINGLE NUCLEOTIDE POLYMORPHISMS IN MATRIX METALLOPEPTIDASE 2 AND 3 GENES IN MALAYSIAN BREAST CANCER PATIENTS

CHAN SOON CHOY

DOCTOR OF PHILOSOPHY
UNIVERSITI PUTRA MALAYSIA

February 2013
IDENTIFICATION AND ANALYSIS OF SINGLE NUCLEOTIDE POLYMORPHISMS IN MATRIX METALLOPEPTIDASE 2 AND 3 GENES IN MALAYSIAN BREAST CANCER PATIENTS

By

CHAN SOON CHOY

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

February 2013
COPYRIGHT

All material 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
DEDICATION

This thesis is especially dedicated to my beloved father who passed away during my PhD candidature. Your hope and wish for me and my brother to be well-educated to earn a decent living will be fulfilled. Your teachings will be a reminder for me and pave the road to my success. I am what you have taught me in the past and will continue to uphold your teachings in the future. Loving you always and forever, from a son who missed many opportunities to serve my filial duty to you.
Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

IDENTIFICATION AND ANALYSIS OF SINGLE NUCLEOTIDE POLYMORPHISMS IN MATRIX METALLOPEPTIDASE 2 AND 3 GENES IN MALAYSIAN BREAST CANCER PATIENTS

By

CHAN SOON CHOY

February 2013

Chair: Professor Rozita Rosli, PhD

Faculty: Medicine and Health Sciences

Breast cancer is the most common cancer among women worldwide as well as in Malaysia. However, it is the process of metastasis in which cancerous cells spread to distant sites from its site of origin that has contributed up to 90% of cancer related mortality. In breast cancer, the matrix metallopeptidase 2 (MMP2) and matrix metallopeptidase 3 (MMP3) genes have been reported to be involved in metastasis and this is supported by both in vitro and in vivo studies. However, the existing literature has not addressed the influence of coding single nucleotide polymorphisms (SNPs) in these genes towards breast cancer metastasis. Hence, this study is designed to be exploratory in nature, utilising the candidate gene approach to investigate the influence of SNP and its haplotype in MMP2 and MMP3 genes on metastasis in Malaysian breast cancer patients.
The combination of high resolution melting (HRM) analysis and DNA sequencing was established as a SNP detection strategy, which successfully identified 26 known SNPs, 10 novel SNPs and 1 novel deletion in both MMP2 and MMP3 genes. All the novel SNPs and the novel deletion have been deposited into the SNP Database (dbSNP) and have been released in database version of Build 132. Comparison of SNP genotypes across three different sources of DNA consisting of blood, adjacent normal tissue, and carcinoma tissue shows 100% concordance. This finding suggests that no somatic mutation occurred in both the MMP genes. It could be implied that any statistically significant SNPs identified in subsequent analysis are inherited low-penetrant variants that can potentially serve as predictive markers.

Statistical analysis identified SNPs that may confer protective effect against metastasis of breast cancer patients. The identified SNPs are c.678G>C of MMP2 gene, and c.133A>G, c.288T>C, c.626-14A>G of MMP3 gene. In addition, a logistic regression model for predicting metastasis status of the patients was built and the overall accuracy of the model was 76.7%. Bioinformatics analysis predicted four SNPs in both MMP2 (c.678G>C, c.750C>T, c.1806C>T, and c.1842C>G) and MMP3 (c.133A>G, c.288T>C, c.306C>G, and c.*129T>C) to exert major effects in changing the secondary structure of its mRNA. Such mRNA structural changes could possibly lead to lower expression levels due to their instable structure.

Phylogenetics analysis showed that negative (purifying) selection acted upon both MMP2 and MMP3 genes in eliminating deleterious non-synonymous SNPs from the breast cancer patient population. This explained the identification of only three
non-synonymous SNPs (MMP2: c.344G>A and c.1499G>A; MMP3: c.133A>G) among the breast cancer patients. Additionally, it is suggested that deleterious synonymous SNPs that confer protective effect against metastasis may possibly be experiencing balancing (positive) selection. It is hoped that findings from this study have contributed towards new knowledge on the genetic basis of SNPs in MMP2 and MMP3 genes in breast cancer metastasis.
Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PENGENALPASTIAN DAN ANALISIS POLIMORFISME TUNGGAL NUKLEOTIDA DALAM GEN-GEN MATRIKS METALLOPEPTIDASE 2 DAN 3 DALAM PESAKIT-PESAKIT KANSER PAYUDARA MALAYSIA

Oleh
CHAN SOON CHOY

Februari 2013

Pengerusi: Professor Rozita bt. Rosli, PhD

Fakulti: Perubatan dan Sains Kesihatan

Kanser payudara merupakan kanser yang paling biasa di kalangan wanita di seluruh dunia dan juga di Malaysia. Proses metastasis di mana sel-sel kanser merebak ke lokasi yang jauh dari tapak asal telah menyumbang sehingga 90% kematian yang berkaitan dengan kanser. Bagi kanser payudara, gen-gen matriks metallopeptidase 2 (MMP2) dan matriks metallopeptidase 3 (MMP3) telah dilaporkan terlibat dalam metastasis dan ini disokong oleh kedua-dua kajian in vitro dan in vivo. Walau bagaimanapun, literatur yang sedia ada tidak mengemukakan pengaruh SNPs pengekodan dalam gen-gen ini terhadap metastasis kanser payudara. Oleh itu, kajian ini dirangka untuk menggunakan pendekatan gen calon untuk menyiasat pengaruh SNP dan haplotip bagi gen MMP2 dan gen MMP3 terhadap metastasis di kalangan pesakit-pesakit kanser payudara Malaysia.
Gabungan kaedah “high resolution melting” (HRM) dan penjujukan DNA telah ditubuhkan sebagai satu strategi pengesanan SNP yang berjaya mengenal pasti 26 “known SNPs”, 10 “novel SNPs” dan 1 “novel deletion” dalam kedua-dua gen MMP2 dan MMP3. Kesemua “novel SNPs” dan “novel deletion” telah didaftarkan ke dalam “SNP Database” (dbSNP) dan telah diterbitkan dalam versi pangkalan data “Build 132”. Perbandingan genotip SNP daripada tiga sumber DNA yang berlainan terdiri daripada darah, tisu normal bersebelahan, dan tisu karsinoma menunjukkan konkordans 100%. Penemuan ini menunjukkan bahawa tiada mutasi somatik berlaku dalam kedua-dua gen MMP. Ini menunjukkan bahawa sebarang SNPs yang dikenal pasti signifikan secara statistik dalam analisis yang berikutnya adalah varian dengan “low-penetrance” yang diwarisi di mana ia berpotensi berfungsi sebagai petanda ramalan.

Analisis “phylogenetics” menunjukkan bahawa pemilihan negatif (pembersih) bertindak ke atas kedua-dua gen MMP2 dan MMP3 dalam penghapusan “deleterious non-synonymous SNPs” daripada populasi pesakit-pesakit kanser payudara. Ini menjelaskan pengenalpastian hanya tiga “non-synonymous SNPs” (MMP2: c.344G>A and c.1499G>A; MMP3: c.133A>G) di kalangan pesakit-pesakit kanser payudara. Selain itu, ia mencadangkan bahawa “deleterious synonymous SNPs” yang memberikan kesan perlindungan terhadap metastasis mungkin mengalami pemilihan pengimbangan (positif). Adalah diharapkan bahawa penemuan-penemuan daripada kajian ini boleh menyumbang kepada ilmu baru tentang genetik asas bagi gen MMP2 dan gen MMP3 terhadap metastasis kanser payudara.
ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Dr. Rozita Rosli for giving me the opportunity to pursue my PhD study under her supervision. I would also like to extend my heartfelt appreciation to her for her invaluable guidance, constructive suggestions and continuous support throughout my research. She has given me freedom to explore and develop my own ideas in research. Going through the process of my PhD study, I have learned a lot and matured in both research and life.

My sincere gratitude and appreciation goes to my second supervisor, Associate Professor Dr. Sabariah Abdul Rahman for her advice, assistance and insightful comments. Thanks for all the detail explanations to me on the pathological aspect of breast cancer on many occasions. Thanks and gratitude is also extended to my third supervisor, Dr. Thilakavathy Karuppiah for her advice and support in my study.

I also wish to convey my thanks and gratefulness to Dr. Abhimanyu Veerakumarasivam and Dr. Michael Ling King Hwa for their brilliant and fruitful discussions. My ideas are more sound and firm through the exchange of views with them. I will not forget to thank Dr. Norshariza Nordin and Dr. Syahrilnizam Abdullah for their friendly support and encouragement to me in completing my study.

Special words of thanks to Professor Dr. Sharifah Noor Akmal Syed Hussain of HUKM for coordination of clinical samples collection. Besides, I would also like to
thank Clarence Ko, Chia Wai Kit, Lam Kah Yuen and Aisyah in assisting the collection of clinical samples. I am also grateful to the administration of Hospital Universiti Kebangsaan Malaysia (HUKM), Hospital Kuala Lumpur (HKL) and Hospital Putrajaya for granting access to clinical sample collection and allowing me to examine their patients’ medical records.

I would like to extend my thanks to Professor Dr. Maznah Ismail Head of Laboratory of Molecular Biomedicine (MOLEMED), Institute of Bioscience (IBS-UPM) for allowing me to use the Varian DHPLC and Beckman-Coulter Genetic Analyser. Thanks also go to Kak Siti Muskinah Mansor in arranging the schedule for working and utilising the facilities in MOLEMED laboratory. My warmest gratitude to all the staff in UPM-MAKNA Cancer Research Laboratory (CANRES), Institute of Bioscience (IBS-UPM) for the support and help especially in handling paperworks on purchasing of research reagents and consumables.

I will never forget to thank Professor Dr. Syed Tajudin Syed Hassan and Dr. Shamarina Shohaimi for their valuable suggestions on biostatistics analysis. I would also like to acknowledge the technical support and assistance by Miss Pung Chai Chin for DHPLC analysis, Mr. Chua Bok Hui and Mr. Chin Lib Kent for HRM analysis.

I am happy to have the luck to be in the Medical Genetics Laboratory during my PhD study with the companionship of my fellow lab mates. Their presence created a friendly and joyful working environment in the lab. I also enjoyed the casual chatting with my friends from Iran, Philippines, Indonesia, Nigeria and China.
for getting the presence of the local Chinese, Malay and Indian friends whom we have so much in common. I also wish to express my gratitude to Kak Salimah Mohd. Said, Kak Hazlen Salleh and Kak Puspa for the hard work in ensuring the laboratory is always well maintained.

I am greatly indebted to my father and mother for their unconditional love ever since I was born. Their love made me feel warm and safe. I also wish to express my heartiest thanks to my only brother for his undying support in pursuing my dream. Emotional support gained from him has helped me journey so far in my higher-level education. Not forgetting my cute nephew and lovely sister in-law. My heartfelt thanks go to Chin Fee Wai for her companionship in the lab till late at night. The memories of those days when we were doing our bench work together are still fresh in my mind. Her emotional support and practical discussion was helpful during the period when both of us were facing problems in our experiments. We have been through so many ups and downs together and I really treasure such friendship.

I would like to acknowledge the Ministry of Science, Technology and Information (MOSTI), Malaysia for funding this study through the National Biotechnology Directorate (NBD) research grant. Apart from that, I am grateful that Universiti Putra Malaysia (UPM) awarded me a postgraduate scholarship under the Graduate Research Fellowship (GRF) scheme.
Last but not least, I wish to express my special thanks to all the breast cancer patients who contributed their clinical samples for this study. Sadly, some of them might succumb to the cancer and never see their loved ones again. I believe that the main reason for them to contribute clinical samples is for the betterment of other breast cancer patients if not for themselves. I wish all of them to be well, happy and to survive the disease. May God always be at the side of those who are greatly in need of help.

Chan Soon Choy
I certify that a Thesis Examination Committee has met on 28th February 2013 to conduct the final examination of Chan Soon Choy on his thesis entitled "Identification and Analysis of Single Nucleotide Polymorphisms in Matrix Metallopeptidase 2 and 3 Genes in Malaysian Breast Cancer Patients" 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 Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

**Fauziah Othman, PhD**  
Professor  
Faculty of Medicine and Health Sciences  
Universiti Putra Malaysia  
(Chairman)

**Lye Munn Sann, DrPH**  
Professor Dato’  
Faculty of Medicine and Health Sciences  
Universiti Putra Malaysia  
(Internal Examiner)

**Cheah Yoke Kqueen, PhD**  
Faculty of Medicine and Health Sciences  
Universiti Putra Malaysia  
(Internal Examiner)

**Shoba Ranganathan, PhD**  
Professor  
Department of Chemistry and Biomolecular Sciences / Faculty of Science  
Macquarie University  
Sydney, Australia  
(External Examiner)

---

**NORITAH OMAR, PhD**  
Associate Professor and Deputy Dean  
School of Graduate Studies  
Universiti Putra Malaysia

Date: 23 May 2013
This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Rozita Rosli, PhD
Professor
Faculty of Medicine and Health Science
Universiti Putra Malaysia
(Chairman)

Sabariah Abdul Rahman, MBBS, M. Path
Professor
Faculty of Medicine
Universiti Teknologi MARA
(Member)

Thilakavathy a/p Karuppiah, PhD
Senior Lecturer
Faculty of Medicine and Health Science
Universiti Putra Malaysia
(Member)

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

Date:
DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

CHAN SOON CHOY

Date: 28 February 2013
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td></td>
<td>vi</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td></td>
<td>ix</td>
</tr>
<tr>
<td>APPROVAL</td>
<td></td>
<td>xiii</td>
</tr>
<tr>
<td>DECLARATION</td>
<td></td>
<td>xv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>xx</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>xxiv</td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td></td>
<td>xxvi</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td></td>
<td>xxvii</td>
</tr>
</tbody>
</table>

## CHAPTER 1

**INTRODUCTION**

1

## CHAPTER 2

**LITERATURE REVIEW**

2.1 Breast anatomy | 7

2.2 Breast cancer | 7

2.2.1 Incidence and risk factors | 9

2.2.2 Histological classification | 13

2.2.3 Hormone receptors | 15

2.2.4 Staging and grading | 17

2.2.5 Treatment | 21

2.3 Invasion and metastasis in breast cancer | 23

2.4 Matrix metallopeptidases and breast cancer | 29

2.4.1 Matrix metallopeptidase 2 gene | 32

2.4.2 Matrix metallopeptidase 3 gene | 35

2.4.3 Involvement of MMP2 and MMP3 genes in breast cancer | 36

2.5 Single nucleotide polymorphism | 42

2.5.1 Types and classes of single nucleotide polymorphism | 43

2.5.2 Evolution and mutation of single nucleotide polymorphism | 47

2.5.3 Single nucleotide polymorphism and haplotype | 51

2.5.4 Single nucleotide polymorphism and candidate gene approach | 54

2.6 Single nucleotide polymorphism detection technologies | 57

2.6.1 Direct DNA sequencing | 58

2.6.2 Denaturing high performance liquid chromatography | 60
2.6.3 High resolution melting analysis

2.7 Bioinformatics in single nucleotide polymorphism functional analysis

3 METHODOLOGY

3.1 General outline of the study

3.2 Data mining

3.2.1 Candidate gene identification

3.2.2 Information retrieval, editing and compilation

3.3 Specimens

3.3.1 Ethics approval

3.3.2 Sample size determination

3.3.3 Specimens collection and storage

3.3.4 Medical records collection

3.4 Genomic DNA extraction and assessment

3.4.1 Genomic DNA extraction from fresh frozen tissue

3.4.2 Genomic DNA extraction from whole blood

3.4.3 Assessment of genomic DNA by UV spectrometry

3.4.4 Assessment of genomic DNA by agarose gel electrophoresis

3.5 Experimental protocol optimisation

3.5.1 Denaturing high performance liquid chromatography

3.5.1.1 SNP masking for primer design

3.5.1.2 Primer design

3.5.1.3 Primer evaluation

3.5.1.4 Partially denaturing temperature prediction

3.5.1.5 Touchdown PCR optimisation

3.5.1.6 DHPLC temperature mapping

3.5.2 High resolution melting

3.5.2.1 Primer design

3.5.2.2 Optimisation of conventional PCR

3.5.2.3 Optimisation of Real-Time PCR

3.5.2.4 Assessment of amplification specificity

3.5.3 DNA sequencing

3.5.3.1 Sequencing reaction condition optimisation

3.5.3.2 Template volume standardisation

3.6 Single nucleotide polymorphism identification

3.6.1 Denaturing high performance liquid chromatography analysis

3.6.1.1 Touchdown PCR

3.6.1.2 Post-PCR reannealing
3.6.1.3 System initialisation
3.6.1.4 Column performance and oven temperature verification
3.6.1.5 Sample analysis
3.6.1.6 Normalised chromatogram cluster analysis

3.6.2 High resolution melting analysis
3.6.2.1 Real-Time PCR
3.6.2.2 Melting curve acquisition
3.6.2.3 Melting curve analysis

3.6.3 Confirmatory DNA sequencing analysis
3.6.3.1 Post-PCR purification
3.6.3.2 Dye terminator cycle sequencing
3.6.3.3 Post sequencing reaction purification
3.6.3.4 Automated capillary sequencing
3.6.3.5 Sequence processing using Pregap4
3.6.3.6 Sequence assembly and variation identification using GAP4

3.7 Single nucleotide polymorphism genotype concordance assessment

3.8 Nomenclature of novel SNP

3.9 Submission of novel SNP to dbSNP

3.10 Statistical analysis
3.10.1 SNPStats
3.10.2 PASW Statistics 18

3.11 Bioinformatics analysis of single nucleotide polymorphism
3.11.1 Integrated bioinformatics analysis software and database selection
3.11.2 Functional analysis implementation

3.12 Sequence analysis of single nucleotide polymorphism
3.12.1 Sequence preparation and input format conversion
3.12.2 Phylogenetic tree construction
3.12.3 Selection test and statistical test

4 SNP GENOTYPE CONCORDANCE STUDY
4.1 Introduction
4.2 Results and Discussion
4.2.1 Optimisation of experimental protocols
4.2.2 SNP genotype concordance
4.3 Conclusion

5 SNP IDENTIFICATION FOR MMP2 & MMP3 GENES
5.1 Introduction
5.2 Results and Discussion
5.2.1 Clinicopathological characteristics of patients 162
5.2.2 Genetic and statistical basis of SNPs identified in MMP2 & MMP3 genes 167
5.2.3 Genetic basis of SNP haplotypes identified in MMP2 & MMP3 genes 183
5.2.4 Consideration of confounder effect in analysis 191
5.2.5 Combinatorial effect of SNPs in MMP2 & MMP3 genes on metastasis and survival status 206

5.3 Conclusion 214

6 BIOINFORMATIC ANALYSIS OF MMP2 & MMP3 GENES 216
6.1 Introduction 216
6.2 Results and Discussion 217
   6.2.1 Functional analysis of SNP in gene regions 217
   6.2.2 Functional analysis of SNP in mRNA folding 230
6.3 Conclusion 238

7 SEQUENCE ANALYSIS OF MMP2 & MMP3 GENES USING PHYLOGENETIC APPROACH 240
7.1 Introduction 240
7.2 Results and Discussion 241
   7.2.1 SNP haplotypes and phylogenetic trees construction 241
   7.2.2 Investigation of selection on MMP2 and MMP3 genes 249
   7.2.3 Evolutionary view on genetic variation of MMP2 and MMP3 genes 254
7.3 Conclusion 258

8 GENERAL CONCLUSION 260
8.1 Research accomplishments 260
8.2 Research limitations 262
8.3 Future research focus 263

REFERENCES 266
APPENDICES 317
BIODATA OF STUDENT 371
LIST OF PUBLICATIONS 372