ROLE OF GENETIC VARIATIONS OF *PARP1* POLYMORPHISMS ON TRIPLE NEGATIVE BREAST CANCER (TNBC) SUSCEPTIBILITY RISK

DR. DURAR AQILAH BINTI ZAMRI

A Dissertation Submitted in Partial Fulfillment of the Requirements for

The Degree of Master of Pathology (Medical Genetics)

SCHOOL OF MEDICAL SCIENCES

UNIVERSITI SAINS MALAYSIA

2022

ACKNOWLEDGEMENTS

First and foremost, Alhamdulillah, I am grateful to Allah The Almighty, for granting me courage, patience and good health throughout my research journey.

Highest appreciation to my main supervisor, Dr. Ahmad Aizat Abdul Aziz, for his endless guidance, advices and support since the beginning till the accomplishment of this project. Thank you for everything you have taught me, and thank you for being patient and understanding at each step of this journey. Appreciation also goes to Prof. Dr. Ravindran Ankathil, my ex-co-supervisor, who had generously shared his ideas, knowledge and guidance up to his retirement. I am also indebted to my current co-supervisor, Dr. Wan Nur Amalina Zakaria for her non-stop supervision, continuous support and inspiration.

I would like to extend my sincere gratitude to Dr. Aziati Azwari Annuar, Dr. Nazihah Mohd Yunus, Dr. Marahaini Musa, Associate Prof. Dr. Sarina Sulong and Prof. Dr. Zilfalil Alwi, who have actively guided and shown me the exemplary figures of a dedicated and passionate researcher. They have been a true inspiration for me to continue explore the research world, especially in the medical genetics field.

My gratitude goes to Dr. Anis Kausar Ghazali, for her insightful guidance in terms of sample size calculation and projection of statistical analysis during this research proposal preparation. I would also like to take this opportunity to thank Mrs. Nur Shafawati Ab Rajab, Ms. Yeoh Hao Ing and Mrs Siti Mariam for their expert in technical assistance, which had helped a lot in improving my laboratory skills.

Thank you my fellow MPath comrades, Dr. Wan Nor Izzati Wan Mohamad Zamri, Dr. Hidayati Husainy Hasbullah and Dr. Mohd Ridzuan Hamid for the beautiful memories of sharing knowledge and valuable experiences together.

I wish to express my sincere thanks to all Human Genome Centre lecturers and all cytogenetic and molecular staff especially, Mrs Alia, Mrs Hidayah, Mrs Zulaikha, Mrs Fatin, Mrs NurHafizah, Mr Nik Zulfikri and Mr Zaki for assisting me during this project.

I am fortunate to have an exceptionally strong support system. Most importantly, I would like to thank my husband, Associate Prof. Dr. Mohd Rashdan Saad, for his endless encouragement, understandings and for taking care of our beautiful children. Heartiest gratitude to Aaliyah Rumaysa, Ayyash Rayyan and Adyan Raeed, for being the most supportive children a mother can have. A million thanks to my extraordinary parents, Associate Prof. Haji Zamri Ibrahim and Hajjah Zawiah Puteh, for the continuous love, prayers and strength that kept me steadfast in this journey. Special thanks to my parents-in-law, Associate Prof. Haji Saad Mohamed for the endless support, and Hajjah Zuraini Taib, who had supported and helped in many ways, up till she returned to Allah, in the midst of this dissertation submission. May Allah reward you abundantly.

Last but not least, I am thankful to the Ministry of Health for providing me the scholarship to fulfil my career goals, and Universiti Sains Malaysia for supporting this study under one of the grants.

Any omission in this brief acknowledgement does not mean lack of gratitude.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND SYMBOLS

ABBREVIATIONS

AS-PCR	Allele-Specific – Polymerase Chain Reaction
BC	Breast Cancer
BMI	Body Mass Index
CI	Confidence Interval
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
ER	Oestrogen Receptor
HER-2	Human Epidermal Growth Factor Receptor 2
$MgCl_2$	Magnesium Chloride
NSCLC	Non-Small Cell Lung Cancer
OR	Odds Ratio
PARP1	*Poly(ADP-Ribose)Polymerase1*
PCR	Polymerase Chain Reaction
PR	Progesterone Receptor
RFLP	Restriction Fragment Length Polymorphism
SD	Standard Deviation
SNP	Single Nucleotide Polymorphism
Taq	*Thermophilus aquaticus*
TBE	Tris-borate- Ethylenediamine Tetraacetic Acid
TNBC	Triple Negative Breast Cancer
UV	Ultraviolet

SYMBOLS

bp	basepairs
mL	millilitre
mM	millimolar
µl	microlitre
µM	micromolar
rpm	revolutions per minute
U	units
vs	versus
°C	degree Celsius
%	percent
≤	less or equal than
≥	more or equal than

ABSTRAK

PERANAN VARIASI GENETIK POLIMORFISME *PARP1* TERHADAP RISIKO KECENDERUNGAN KANSER PAYUDARA GANDA TIGA NEGATIF

Latar belakang: Kanser payudara merupakan kanser paling kerap ditemui di kalangan wanita dan penyebab kematian daripada kanser kedua tertinggi di serata dunia. Kanser payudara ganda tiga negatif (TNBC) mempunyai fenotip paling agresif dan prognosis paling teruk berbanding jenis kanser payudara yang lain. '*Poly(ADP-Ribose)Polymerase1*' (PARP1) ialah protein nukleus yang memainkan peranan dalam pembaikian kerosakan asid deoksiribonukleik (DNA) dengan bertindak sebagai pengesan bagi kerosakan bebenang DNA. Variasi genetik antara individu seperti Polimorfisme Nukleotida Tunggal (SNP) *PARP1* Val762Ala (rs1136410 T>C) dan Ala284Ala (rs1805414 T>C) berkemungkinan mengurangkan aktiviti *PARP1* dalam membaiki kerosakan DNA, dan seterusnya berkemungkinan meningkatkan risiko kecenderungan kanser. Memandangkan terdapat kekurangan maklumat mengenai kekerapan polimorfisme *PARP1* rs1136410 T>C dan rs1805414 T>C dalam kalangan pesakit TNBC, dan laporan sedia ada yang terhad berkenaan polimorfisme ini ke atas risiko kecenderungan TNBC, kajian ini bertujuan untuk menyiasat kesan polimorfisme *PARP1* rs1136410 T>C dan rs1805414 T>C dalam memodulasi risiko kecenderungan TNBC.

Kaedah: Sampel darah telah diambil daripada 70 pesakit TNBC yang telah dipastikan diagnosis melalui histopatologi dan 70 individu sihat dalam kumpulan kawalan, dan seterusnya DNA telah diekstrak. Pengenotipan polimorfisme *PARP1* telah dilaksanakan

melalui penggunaan teknik Tindak Balas Berantai Polimerase - Polimorfisme Pemotongan Panjang Cebisan (RFLP-PCR) bagi rs1136410 T>C, manakala Tindak Balas Berantai Polimerase – Alel Khusus (AS-PCR) digunakan untuk rs1805414 T>C. Kemudian, pola genotip telah dikategorikan kepada jenis liar homozigot, heterozigot dan varian homozigot. Frekuensi genotip dan alel kedua-dua polimorfisme *PARP1* telah dihitung dan dibandingkan antara pesakit TNBC dengan kumpulan individu sihat. Analisis regresi logistik yang menerbitkan nisbah ganjil (OR) dengan sela keyakinan 95% telah dilakukan untuk menilai risiko kecenderungan bagi pesakit TNBC.

Keputusan: Varian homozigot genotip (CC) polimorfisme rs1136410 *PARP1* dan genotip heterozigot (TC) polimorfisme rs1805414 *PARP1* memperlihatkan risiko lebih tinggi bagi kecenderungan TNBC dengan OR: 1.190, 95% CI: 0.300-4.721 dan OR: 1.103, 95% CI: 0.402-3.031. Tambahan lagi, genotip heterozigot *PARP1* rs1136410 (TC) didapati mempunyai perkaitan bermakna dengan risiko lebih rendah untuk keberulangan semula penyakit dalam kalangan pesakit-pesakit TNBC, dengan OR: 0.105, 95% CI: 0.029-0.374, p=0.001.

Kesimpulan: Keputusan kajian kami menunjukkan kekurangan perkaitan polimorfisme *PARP1* rs1136410 T>C dan rs1805414 T>C dalam memodulasikan risiko kecenderungan TNBC. Walau bagaimanapun, kajian ini menunjukkan *PARP1* rs1136410 genotip heterozigot (TC) dijumpai mempunyai perkaitan bermakna dengan risiko lebih rendah untuk keberulangan penyakit dalam kalangan pesakit TNBC. Kekurangan perkaitan berkemungkinan disebabkan oleh sampel saiz subjek kajian yang rendah telah mengganggu daya kaji dan mempengaruhi kerendahan frekuensi alel varian dalam kalangan pesakit TNBC. Kajian ini perlu diperluaskan kepada bilangan saiz sampel yang lebih besar supaya daya kaji dapat ditingkatkan, dan perlu meneliti gabungan genotip polimorfik *PARP1*.

Kata kunci: Kanser payudara ganda tiga negatif, polimorfisme gen PARP1, *PARP1* rs1136410, *PARP1* rs1805414, risiko kecenderungan

ABSTRACT

Background: Breast cancer is the most common cancer in females worldwide and the second most common cause of cancer deaths. Triple negative breast cancer (TNBC) has the most aggressive phenotype and worst prognosis compared to other breast cancer subtypes. *Poly(ADP-Ribose)Polymerase1* (PARP1) is a nuclear protein, that plays a role in deoxyribonucleic acid (DNA) damage repair by acting as sensor for DNA strand breaks. Inter-individual genetic variations such as single nucleotide polymorphisms (SNPs) of *PARP1* Val762Ala (rs1136410 T>C) and Ala284Ala (rs1805414 T>C) may reduce the *PARP1* activity in repairing DNA damage and thus may increase the cancer risk predisposition. Since there is paucity of information on the frequency of *PARP1* rs1136410 T>C and rs1805414 T>C polymorphisms in TNBC patients, and limited reports available on these polymorphisms on TNBC susceptibility risk, the present study is designed to investigate the impact of *PARP1* rs1136410 T>C and rs1805414 T>C polymorphisms in modulating TNBC susceptibility risk.

Methods: Blood sample of 70 histopathologically confirmed TNBC patients and 70 healthy control individuals were collected, and DNA was extracted. The genotyping of *PARP1* rs1136410 T>C and rs1805414 polymorphisms was performed by using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) and allele specific-polymerase chain reaction (AS-PCR) technique, respectively. Next, the genotype patterns were categorized into homozygous wild-type, heterozygous and homozygous variant. The genotype and allele frequencies of both *PARP1* polymorphisms were calculated and compared between TNBC patients and healthy individual group

using chi-square test. Next, logistic regression analysis deriving Odds Ratio (OR) with 95% confidence interval (CI) was performed to evaluate the TNBC patients' susceptibility risk.

Results: The homozygous variant (CC) genotype of *PARP1* rs1136410 polymorphism and heterozygous (TC) genotype of *PARP1* rs1805414 showed a higher TNBC susceptibility risk with OR: 1.190, 95% CI: 0.300-4.721 and OR: 1.103, 95% CI: 0.402-3.031. In addition, heterozygous (TC) genotype of *PARP1* rs1136410 was found to be significantly associated with lower risk of recurrence among TNBC patients with OR: 0.105, 95% CI: 0.029-0.374, p=0.001.

Conclusion: Our results showed the lack of association of *PARP1* rs1136410 T>C and rs1805414 T>C polymorphisms in modulating TNBC susceptibility risk. However, the present study observed that *PARP1* rs1136410 (TC) genotype was found to be significantly associated with lower risk of disease recurrence among TNBC patients. The lack of association could be due to low sample size of study subjects which interfered the power of the study and influence the low frequency of variant allele in TNBC patients. The study needs to be extended to larger number of sample size which can increase the power of the study and needs to explore the polymorphic genotype combinations of *PARP1*.

Keywords: Triple Negative Breast Cancer, PARP1 gene polymorphism, *PARP1* rs1136410, *PARP1* rs1805414, susceptibility risk

CHAPTER 1: INTRODUCTION

1.1 Introduction

Breast cancer is the most common cancer in females worldwide and the second most common cause of cancer deaths. Triple negative breast cancer (TNBC), a subtype of breast cancer, is characterized by negative staining of estrogen (ER), progesterone (PR) receptors and no amplification of human epidermal growth factor receptor 2 (HER2) hormones (Yin et al., 2020). There is variable prevalence of TNBC across different ethnics and population. The statistics revealed a shocking discovery that in African-American population, the incidence is surprisingly high, around 39%. In India, the number is also high which is around 31.9%, while the incidence in Malaysia stands at around 17.6% (Yeh et al., 2017).

TNBC is a molecular subtype of breast cancer with poor prognosis and aggressive disease course compared to other breast cancer subtypes. Diagnosis of TNBC is histopathologically confirmed by the absence of expression of estrogen receptor (ER), progesterone receptor (PR) and no amplification of human epidermal growth factor receptor 2 (HER-2). There is paucity of information on the nature of genetic factors which contribute to TNBC susceptibility risk in Malaysian population. Hence, there is the need to identify the genetic predisposition factors associated with TNBC susceptibility risk. Deoxyribonucleic acid (DNA) repair mechanisms play a major role in protecting cells against DNA damage and carcinogenesis. *Poly(ADP-Ribose)Polymerase1* (PARP1) enzyme is a nuclear protein, serving as sensor for DNA strand breaks. By repairing DNA damage and genetic stability, PARP1 plays a key role in the prevention of cancer (Bashir et al., 2018; Toss & Cortesi, 2013; Ye et al., 2012).

Single nucleotide polymorphisms (SNPs) which are substitution of single nucleotide along the DNA, are the most common forms of genetic variations associated with susceptibility to diseases. SNPs represent a potentially vast area for the detection of genetic alterations that can be related with different types of diseases across different populations. *PARP1* polymorphisms have been associated with increased or reduced risk of different types of cancer (Bashir et al., 2018; Cheng et al., 2019; Hua et al., 2014; Ma et al., 2016).

Poly(ADP-Ribose)Polymerase1 (*PARP1*), which is located at chromosome 1q41-q42, has at least 1287 reported SNPs, including 202 coding-region SNPs (cSNPs) (Bashir et al., 2018). Val762Ala (rs1136410) polymorphism is one of the most investigated cSNPs, and is reported to cause functional alteration of *PARP1*. The transition of nucleotide T>C changes the amino acid valine to alanine at position 762 in exon 17 of *PARP1*. The amino acid change resulting in the reduction of *PARP1* activity may further cause increased risk of several cancers (Bashir et al., 2018; Cheng et al., 2019; Ramezani et al., 2019).

Ma et al., (Ma et al., 2016) observed that the frequencies of the *PARP1* rs1136410 was significantly different between the case and control groups (p=0.024), and the C/C genotype showed reduced breast cancer risk compared to T/T genotype (OR=0.61, 95% CI: 0.43-0.89, p=0.007) in the Chinese population. However, Ramezani et al., (Ramezani et al., 2019) reported no association between rs1136410 and breast cancer susceptibility among Iranian women. Meta-analysis by Yin et al., (Yin et al., 2020) involving 65 studies, revealed rs1136410 C vs. T allele association with breast cancer odds ratio (95% CI) of 0.95 (0.83-1.09) with p=0.015, which confers significant reduced susceptibility.

In view of other cancer types, stratified analysis by Cheng et al., (Cheng et al., 2019) showed that rs1136410 TC/CC carriers were more likely to develop tumours originating from the mediastinum (TC/CC vs. TT: adjusted OR=1.65, 95% CI=1.06-2.56, p=0.028). Meanwhile, Bashir et al., (Bashir et al., 2018) found that the major allele homozygote (CC) of rs1136410 had a significant association with susceptibility risk of thyroid cancer (OR=1.30, 95% CI 0.99-1.77). Meta-analysis by Hua et al., (Hua et al., 2014) demonstrated evidence for an association between *PARP1* Val762Ala and cancer susceptibility risks within gastric cancer, brain tumour and Asian subgroups. A more recent study by Jin et al., (Jin et al., 2020) has revealed that haplotypes comprising of 5 SNPs on *PARP1* which included rs1136410 and rs1805414 were able to differentiate non-squamous cell lung cancer (NSCLC) tumour tissues from normal tissues.

Meanwhile, *PARP1* Ala284Ala (rs1805414) polymorphism is a synonymous variant resulting from silent mutation of single nucleotide change from T>C. This SNP has been reported to be positively associated with breast cancer susceptibility risk in the Saudi Arabia population (Alanazi et al., 2013), but negatively associated with colorectal cancer (Bashir et al., 2018). A study looking into novel SNPs in glioblastoma has found that the *PARP1* Ala284Ala polymorphism had positive association with the disease (Keller et al., 2011). Another study on brain tumour risk revealed that heterozygous mutant genotype of Ala284Ala SNP conferred reduced susceptibility risk for brain tumour in the Pakistani population (Khan et al., 2019). SNP-SNP interaction analysis done in a study conducted in America showed that interaction between *PARP1* SNPs rs1805414, rs3219073 and rs1805415 were significantly associated with non-squamous cell lung cancer (NSCLC) tumour with OR ranging from 3.61–6.75; 95%CI from 1.82 to 19.9; p-value<0.001 (Jin et al., 2020).

Reported data thus far has shown inconsistent susceptibility risks of *PARP1* genetic variations with different types of cancer across different population and ethnic groups. In regards to breast cancer risk association with *PARP1* variants specifically, only limited number of studies has been conducted, indicating more studies are required considering breast cancer as the most common cancer in females and the fifth most common cause of cancer deaths, worldwide ([IARC] International Agency for Research on Cancer, 2020).

Breast cancer being the most common cancer accounting for 34.1% of all cancers among females in Malaysia, and TNBC possessing the poor prognosis, studies on genetic risk markers would be undeniably crucial (Azizah et al., 2019; Yin et al., 2020). Study on the role of genetic variations of *PARP1* on TNBC susceptibility risk has not been conducted in the Malaysian population setting, hence the frequency of these genetic variants still needs to be discovered, and subsequently the association with TNBC susceptibility risk needs to be determined. Furthermore, since it has been reported that increased cancer risk was detected with *PARP1* polymorphisms in the subgroups of Asian, studies involving Malaysian population is deemed warranted (Yin et al., 2020). The identification of these genetic susceptibility risk candidate markers would be of paramount importance as to provide more evidence and insight to the pathogenesis mechanism, and consequently may contribute to the development of targeted therapies for TNBC.

PARP1 polymorphisms have been reported to influence cancer predisposition risk in several number of studies. However, the results were inconsistent for different ethnicity, population and diseases. Since no information is yet available on the distribution of *PARP1* polymorphisms in Malaysian TNBC population, and no reports available on the impact of genetic variations of *PARP1* on TNBC susceptibility risk, the

present study is designed to investigate the impact of *PARP1* polymorphisms on TNBC susceptibility risk.

CHAPTER 2: OJECTIVES OF THE STUDY

2.1 General objective

To investigate the contributions of two candidate SNPs of *PARP1* (rs1136410 T>C and rs1805414 T>C) on TNBC susceptibility risk.

2.2 Specific objectives

1. To determine the genotype and allele frequencies of *PARP1* rs1136410 T>C and rs1805414 T>C polymorphisms among TNBC patients and healthy normal controls.
2. To determine the association of *PARP1* rs1136410 T>C and rs1805414 T>C polymorphisms on TNBC susceptibility risk.
3. To evaluate whether data generated on the SNPs of *PARP1* (rs1136410 T>C and rs1805414 T>C) could serve as candidate biomarkers for TNBC susceptibility risk.

CHAPTER 3: STUDY PROTOCOL

3.1 Research proposal submitted for ethical approval

Research title: Role of genetic variations of *PARP1* polymorphisms on triple negative breast cancer susceptibility risk.

Principal investigator (MMC No. if applicable): Dr Durar Aqilah binti Zamri (MMC: 61611)

Co-researchers: (MMC No. if applicable):

Dr Ahmad Aizat bin Abdul Aziz

Prof Dr Ravindran Ankathil

Introduction

Breast cancer is a global health issue, with incidence of 2.2 million and mortality of 684,996 worldwide according to the 2020 GLOBOCAN statistics ([IARC] International Agency for Research on Cancer, 2020; Bray & Møller, 2006; Ferlay et al., 2019). Breast cancer is the most common cancer and accounted for 34.1% of all cancers among females in Malaysia, according to the Malaysian National Cancer Registry 2012-2016 (Azizah et al., 2019). Breast cancer risk is partly determined by genetic factors including rare pathogenic variants in susceptibility genes (BRCA1/2, PTEN, TP53 CDH1, STK11) and common low risk genetic variants. Other risk factors include alcohol use, smoking, reproductive factors, hormonal factors, family history, mammographic density, Body Mass Index (BMI) and body height.

Triple negative breast cancer (TNBC) is one of the molecular subtypes of invasive breast carcinomas which lacks the expression of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER-2). TNBC has the most aggressive phenotype and worst prognosis. There is paucity of information on the nature of genetic factors which contribute to TNBC susceptibility risk in Malaysian population. Hence, there is the need to identify the genetic predisposition factors associated with TNBC susceptibility risk.

Problem statement & Study rationale

DNA repair mechanisms play a major role in protecting cells against DNA damage and carcinogenesis. *Poly(ADP-Ribose)Polymerase1* (PARP1) enzyme is a nuclear protein, serving as sensor for DNA strand breaks. By repairing DNA damage and genetic stability, PARP1 plays a key role in prevention of cancer (Bashir et al., 2018; Hoeijmakers, 2001; Toss & Cortesi, 2013; Ye et al., 2012).

Single nucleotide polymorphisms (SNPs) which are substitution of single nucleotide along the DNA, are the most common forms of genetic variations associated with susceptibility to diseases. SNPs represent a potentially vast area for the detection of genetic alterations that can be related with different types of diseases across different populations. *PARP1* polymorphisms have been associated with increased or reduced risk of different types of cancer (Bashir et al., 2018; Cheng et al., 2019; Hua et al., 2014; Ma et al., 2016). However, these associations were inconsistent across different population and ethnicity (Bashir et al., 2018; Cheng et al., 2019; Hua et al., 2014; Ma et al., 2016). Effects of SNPs in Caucasian population cannot be translated directly to Asian and other populations, and vice versa. There are limited number of studies reported on the

distribution of *PARP1* polymorphisms in TNBC population and no reports available on the impact of genetic variations of *PARP1* on susceptibility risk in Malaysian TNBC patients. Hence, the present study is designed to investigate the impact of *PARP1* polymorphisms on TNBC susceptibility risk.

Research Question(s)

1. What is the distribution of *PARP1* SNPs (rs1136410 T>C and rs1805414 T>C) in TNBC patients and healthy normal controls?
2. Is there any association of the *PARP1* SNPs (rs1136410 T>C and rs1805414 T>C) with susceptibility risk in TNBC patients?

Objective

General:

To investigate the contributions of two candidate SNPs of *PARP1* (rs1136410 T>C and rs1805414 T>C) on TNBC susceptibility risk.

Specific:

1. To determine the genotype and allele frequencies of *PARP1* rs1136410 T>C and rs1805414 T>C polymorphisms among TNBC patients and healthy normal controls.
2. To determine the association of *PARP1* rs1136410 T>C and rs1805414 T>C polymorphisms on TNBC susceptibility risk.

3. To evaluate whether data generated on the SNPs of *PARP1* (rs1136410 T>C and rs1805414 T>C) could serve as candidate biomarkers for TNBC susceptibility risk.

Literature review

PARP1 genetic variations

PARP1 is a nuclear enzyme that functions mainly in DNA damage repair mechanism and genetic stability (Ye et al., 2012). PARP1 gene, which is located at chromosome 1q41-1q42, has at least 1287 reported SNPs across different population and ethnic groups, including 202 coding-region SNPs (cSNPs) (Hua et al., 2014; Wood et al., 2001). Study on the PARP1 gene polymorphisms has become an area of interest for intensive research as the resultant functional alterations that impair the DNA damage repair mechanism and genome stability, impose their possible role in cancer susceptibility risk.

Val762Ala (rs1136410) polymorphism is one of the most investigated cSNPs, and is reported to cause functional change of PARP1. The transition of nucleotide T>C changes the amino acid valine to alanine at position 762 in exon 17 of *PARP1*. This might cause significant reduction of poly (ADP-ribosylation) activities in an allele dosage-dependent manner and reduced PARP1 activity (Hua et al., 2014; Lockett et al., 2004). Meanwhile, *PARP1* SNPs Ala284Ala (rs1805414) is a synonymous variant resulting from silent mutation of single nucleotide change from T>C.

Ma et al., (2016) observed that the frequencies of the *PARP1* rs1136410 were significantly different between the case and control groups (p=0.024), and the C/C genotype showed reduced breast cancer risk compared to T/T genotype (p=0.007 OR=0.61, 95% CI: 0.43-0.89) in a Chinese population. However, Ramezani et al., (2019)

reported no association between rs1136410 and breast cancer susceptibility among Iranian women. Meta-analysis by Li et al., (2020) involving 65 studies, revealed rs1136410 C vs. T allele association with breast cancer odds ratio (95% CI) of 0.95 (0.83-1.09) with p=0.015, which confers significant reduced susceptibility.

In view of other cancer types, stratified analysis by Cheng et al., (2019) showed that rs1136410 TC/CC carriers were more likely to develop tumours originating from the mediastinum (TC/CC vs. TT: adjusted OR=1.65, 95% CI=1.06-2.56, p=0.028). Meanwhile, Bashir et al., (2018) found that the major allele homozygote (CC) of rs1136410 had a significant association with susceptibility risk of thyroid cancer (OR=1.30, 95% CI 0.99-1.77). Meta-analysis by Hua et al., (2014) demonstrated evidence for an association between *PARP1* Val762Ala and cancer susceptibility risks within gastric cancer, brain tumour and Asian subgroups. A recent study by Jin et al., (2020) has revealed that haplotypes comprising of 5 SNPs on PARP1 gene which included rs1136410 and rs1805414 were able to differentiate non-squamous cell lung cancer (NSCLC) tumour tissues from normal tissues.

PARP1 gene Ala284Ala (rs1805414) polymorphism has been reported to be positively associated with glioblastoma (Keller et al., 2011), but negatively associated with colorectal cancer (Bashir et al., 2018; Berndt et al., 2007; Ogino and Stampfer, 2010). SNP-SNP interaction analysis done in a study conducted in America showed that interaction between *PARP1* SNPs rs1805414, rs3219073 and rs1805415 were significantly associated with non-squamous cell lung cancer (NSCLC) tumour with OR ranging from 3.61–6.75; 95%CI from 1.82 to 19.9; p-value<0.001 (Jin et al., 2020). In regards to breast cancer susceptibility, it has been reported that rs1805414 SNP was associated with a significantly increased susceptibility to breast cancer, with significant

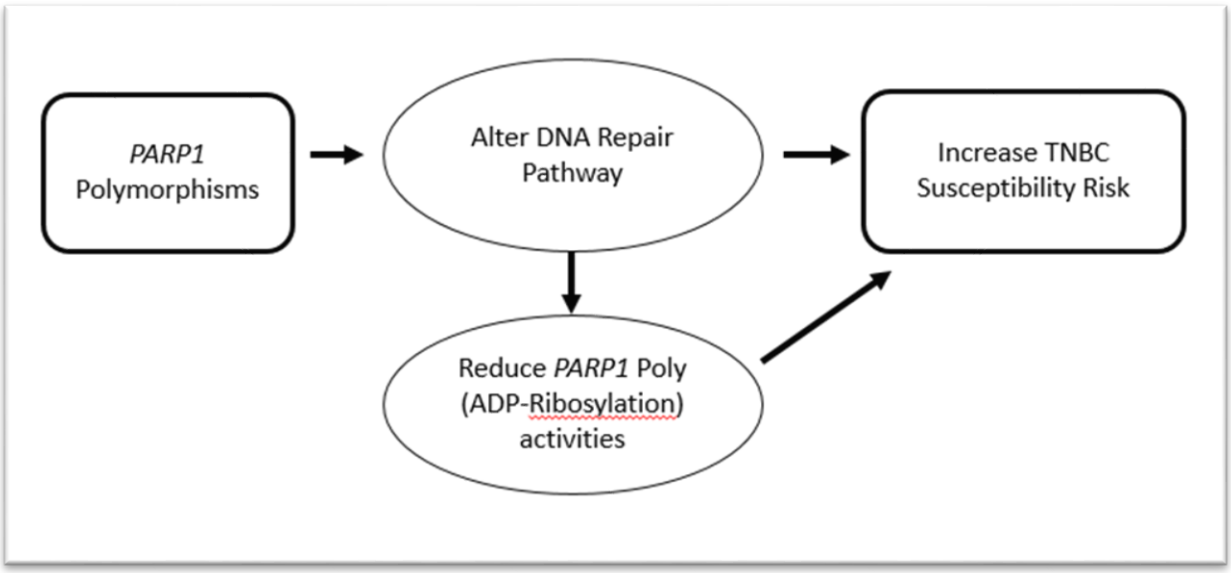
risk observed for the TC, CC and TC+CC genotypes in Saudi population (Alanazi et al., 2013).

Reported data thus far has shown inconsistent susceptibility risks of *PARP1* genetic variations with different types of cancer across different population and ethnic groups. In regards to breast cancer risk association with PARP1 gene variants specifically, only limited number of studies has been conducted, indicating more studies are required considering breast cancer as the most common cancer in females and the fifth most common cause of cancer deaths, worldwide ([IARC] International Agency for Research on Cancer, 2020).

Breast cancer being the most common cancer accounting for 34.1% of all cancers among females in Malaysia, and triple negative breast cancer (TNBC) possessing the poor prognosis, studies on genetic risk markers would be undeniably crucial (Azizah et al., 2019; Yin et al., 2020). TNBC with reported incidence of 17.6% in Malaysia was significantly associated with a higher grade of grade 3 carcinoma (Tan et al., 2009; Yeh et al., 2017). Study on the role of genetic variations of PARP1 gene polymorphisms on TNBC susceptibility risk has not been conducted in Malaysian population setting, hence the frequency of these genetic variants still needs to be discovered, and subsequently the association with TNBC susceptibility risk needs to be determined. Furthermore, since it has been reported that increased cancer risk was detected with PARP1 gene polymorphisms in the subgroups of Asian, studies involving Malaysian population is deemed warranted (Li et al., 2020). The identification of these genetic susceptibility risk candidate markers would be of paramount importance as to provide more evidence and insight to the pathogenesis mechanism, and consequently may contribute to the development of targeted therapies for TNBC.

PARP1 gene polymorphisms have been reported to influence cancer predisposition risk in several number of studies. However, the results were inconsistent for different ethnicity, population and diseases. Since there is no information exists on the distribution of PARP1 gene polymorphisms in Malaysian TNBC population, and no reports available on the impact of genetic variations of *PARP1* on TNBC susceptibility risk, the present study is designed to investigate the impact of PARP1 gene polymorphisms on TNBC susceptibility risk.

Conceptual framework



Research design

- Case-control study

Study area

- Hospital Universiti Sains Malaysia
- Human Genome Centre (PPSP)

Study population

- Reference population: Patients diagnosed with Triple Negative Breast Cancer (TNBC) in Kelantan
- Target population: Triple Negative Breast Cancer (TNBC) patients attending Hospital USM
- Healthy Normal Controls

Subject criteria

TNBC patients group

Inclusion criteria:

1. Histopathologically confirmed female breast cancer patients by negative estrogen receptor (ER), progesterone receptor (PR) and no amplification of HER-2 receptors.

Exclusion criteria:

1.Female breast cancer patients with histopathologically not TNBC subtype.

Controls group

Inclusion criteria:

1. Normal healthy female individual (age matched to the case group).

2. Biologically unrelated to case group.

3. Cancer free participants.

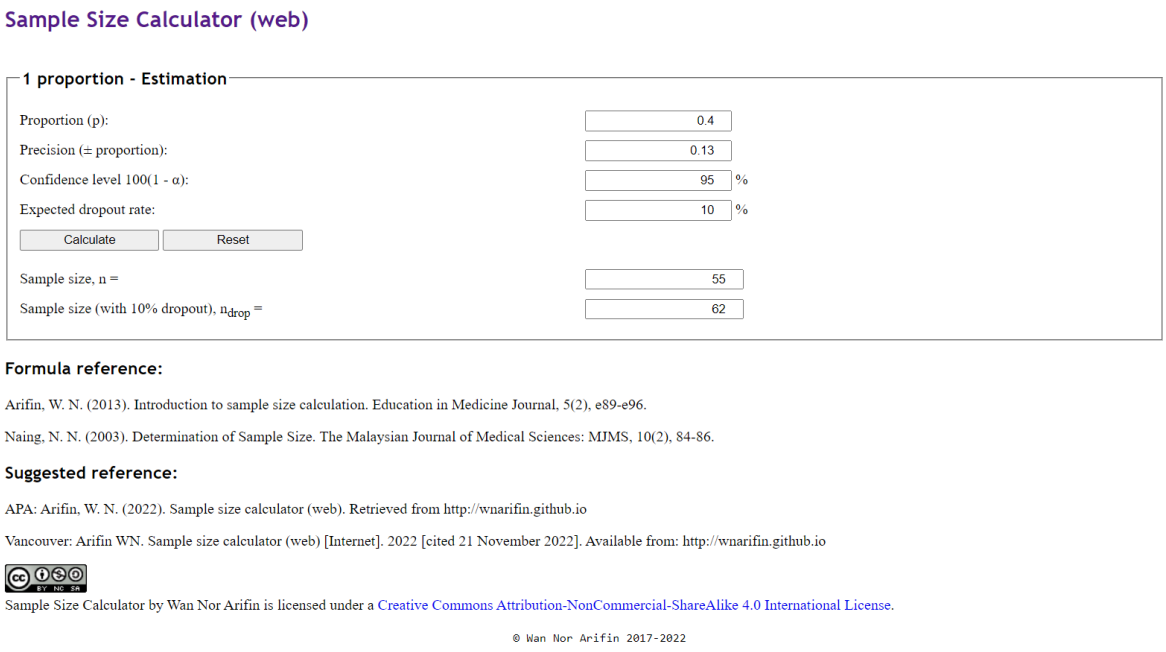
Exclusion criteria:

1. Normal healthy female with family history of breast cancer or any other cancers.

Sample size estimation

Objective 1: Sample size using single proportion calculator estimation.

The sample size for objective 1 was calculated by using single proportion estimation with proportion of 0.40 (Liao et al., (Liao et al., 2020)) and precision of 0.13 which gives the number of patients required for this study was 55. After adding anticipated 10% dropouts, the required sample size was 62 participants per group. Since estimation will be made based on each group, the required sample size is 62 TNBC patients and 62 healthy female normal controls (total number of participants required is 124).


Sample Size Calculator (web)

1 proportion - Estimation

Proportion (p):	0.4
Precision (± proportion):	0.13
Confidence level 100(1 - α):	95 %
Expected dropout rate:	10 %
Calculate Reset	
Sample size, n =	55
Sample size (with 10% dropout), n_{drop} =	62

Formula reference:

Arifin, W. N. (2013). Introduction to sample size calculation. Education in Medicine Journal, 5(2), e89-e96.

Naing, N. N. (2003). Determination of Sample Size. The Malaysian Journal of Medical Sciences: MJMS, 10(2), 84-86.

Suggested reference:

APA: Arifin, W. N. (2022). Sample size calculator (web). Retrieved from http://wnarifin.github.io

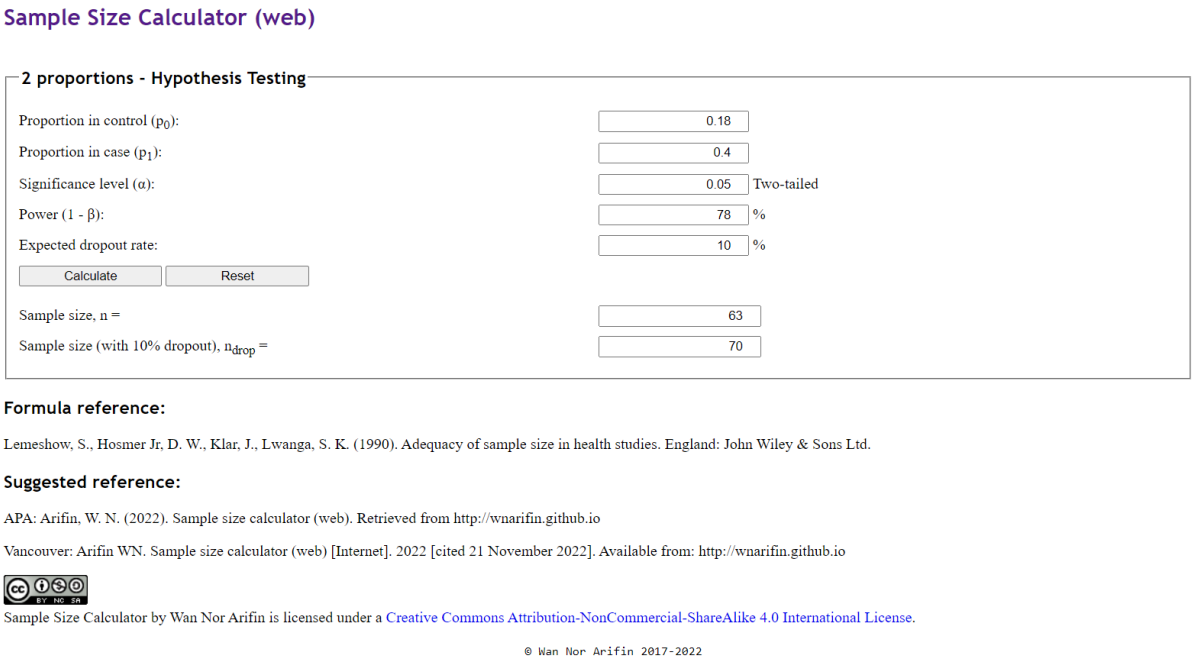
Vancouver: Arifin WN. Sample size calculator (web) [Internet]. 2022 [cited 21 November 2022]. Available from: http://wnarifin.github.io

Sample Size Calculator by Wan Nor Arifin is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

© Wan Nor Arifin 2017-2022

Objective 2: Sample size using two proportion calculator estimation.

The sample size was calculated using the two-proportion formula; with P_0=0.18, P_1=0.4, power=0.78 and significance level=0.05 (Liao et al., 2020; Ramezani et al., 2019). Thus, number of sample size is 63 each for control and case groups. Taking into account an estimated drop-out percentage of 10%, the total number of patients required for this study is 70 patients. Hence, 70 TNBC patients and 70 healthy female normal controls shall be used for genotyping and association analyses, amount up to a total of 140 subjects.



Sample Size Calculator (web)

2 proportions - Hypothesis Testing

Proportion in control (p_0):	0.18	
Proportion in case (p_1):	0.4	
Significance level (α):	0.05	Two-tailed
Power (1 - β):	78	%
Expected dropout rate:	10	%
Calculate / Reset		
Sample size, n =	63	
Sample size (with 10% dropout), n_{drop} =	70	

Formula reference:

Lemeshow, S., Hosmer Jr, D. W., Klar, J., Lwanga, S. K. (1990). Adequacy of sample size in health studies. England: John Wiley & Sons Ltd.

Suggested reference:

APA: Arifin, W. N. (2022). Sample size calculator (web). Retrieved from http://wnarifin.github.io

Vancouver: Arifin WN. Sample size calculator (web) [Internet]. 2022 [cited 21 November 2022]. Available from: http://wnarifin.github.io

Sample Size Calculator by Wan Nor Arifin is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

© Wan Nor Arifin 2017-2022

Summary of sample size calculation:

Combining the calculations for sample size of the two specific objectives, the largest sample size obtained was from estimation of sample size for objective 2 (140 subjects). Therefore, in this study, 70 TNBC patients and 70 healthy female normal controls will be recruited.

Sampling method and subject recruitment

Archived materials from our previous study, JEPeM code: (USMKK/PPP/JEPeM[260.39210]) shall be used, as this study has already been consented for storage and that their archived genetic materials shall be used in this study (Based on Guidelines on Retention of Pathology Records and Materials (Version 1/2005)).

Three (3) mL of peripheral blood samples have been collected from consented study subjects. Written informed consent has been obtained from the study subjects prior to blood collection, during enrolment into the study.

For case group, TNBC patients have been recruited, and peripheral blood samples have been collected from SOPD clinic, Hospital Universiti Sains Malaysia (HUSM) Kubang Kerian, Kelantan. Study subjects have been labelled with a study code to ensure the privacy and confidentiality. A set of questionnaires was designed to inquire on patients' personal details including socio-demographic data have been collected from study subjects. Whereas clinical and pathological variables including age at diagnosis, histological type, staging according to TNM classification, lymph node involvement, menopausal status, presence of metastases/recurrence and treatment response shall be retrieved from the patients' medical records and recorded with permission from respective hospital director.

For normal female healthy control recruitment, volunteers (USM staffs and students) or who visited HUSM as patients' bystanders, visitors, who come to treatment of simple medical problem which is unrelated to cancer have been recruited. Subjects have been labelled with a study code to ensure the privacy and confidentiality. A questionnaire

which includes socio-demographic status have been distributed to study subjects in order to obtain the epidemiological data.

Study subject recruitment:

TNBC patient recruitment

TNBC cases have been identified by routine immunohistochemistry through routine histopathology diagnosis. Seventy (70) histopathologically confirmed TNBC patients have been recruited into the study according to the inclusion and exclusion criteria mentioned previously.

Control group recruitment

Seventy (70) normal healthy females have been recruited into this study based on the inclusion and exclusion criteria as mentioned.

Research tool

Types	Description	Owner	Location
NanoQuant Infinate M200 (TECAN, Switzerland) spectrometer	For quantification of the extracted DNA	Human Genome Centre, School of Medical Sciences, Universiti Sains Malaysia (USM)	Human Genome Centre, School of Medical Sciences, Universiti Sains Malaysia (USM)
Mastercycle thermocylcer	For amplification of targeted gene	Human Genome Centre, School of Medical Sciences, Universiti Sains Malaysia (USM)	Human Genome Centre, School of Medical Sciences, Universiti Sains Malaysia (USM)

Data collection method

Data collection sheet: includes profile and socio-demographic data of the patients, that are recorded in the patients' files which contains date of diagnosis, breast cancer type, chemotherapy commencement date and the clinical response. Data collection will be started after ethical approval is received. Once ethical approval is obtained, relevant data required will be collected from patients' medical records with permission from the respective hospital director.

DNA isolation

DNA will be extracted from peripheral blood samples by using QIAGEN QIAamp DNA Blood Mini kit (QIAGEN, Hilden, Germany) according to the manufacturer's instruction. Briefly, 200 µl of blood will be mixed with 20 µl of proteinase K and Buffer AL (lysis buffer) to lyse the red blood cells and proteins that attach to DNA. Then, sample will be incubated at 56 ^{o}C for 10 minutes to ensure complete lysis of the red blood cells. Later, 70-100% of ethanol will be added to the sample to precipitate DNA, followed by addition of Buffer AW1 and AW2 (washing buffer) to wash the DNA. Finally, buffer AE will be added to elute the DNA and the DNA extract will be stored in -20 ^{o}C for long term storage.

Polymerase Chain Reaction – Allele-specific PCR

Genotype analysis of the *PARP1* rs1136410 (Val762Ala) and rs1805414 (Ala284Ala) polymorphisms will be carried out by using allele-specific PCR technique. Briefly, PCR reactions will be performed in 25 µl of final volume containing 200 ng of each primer, 50 ng genomic DNA, 1.5 mM $MgCl_2$, 200 µM dNTPs and 1.0 unit of Taq DNA polymerase. The primer sequences and sizes of *PARP1* amplicons to be studied are given in the Table 1. Amplification will be performed in Mastercycle thermocycler for the PCR reaction.

Table 1: List of primer sequences for *PARP1* polymorphisms

SNPs	**Primer Sequence**	**Amplicon Size (bp)**
PARP1 Val762Ala T>C (rs1136410)		236
*WF	TTGCTCCTCCAGGCCAAGTC	
**MF	TTGCTCCTCCAGGCCAAGTT	
***CR	CAGCTTTCCAGGAGATCCTA	
PARP1 Ala284Ala T>C (rs1805414)		153
WF	GCAGATCTTGGACCGAGTAGAT	
MF	GCAGATCTTGGACCGAGTAGAC	
CR	GGTGTCTGTGTCTTGACCAT	

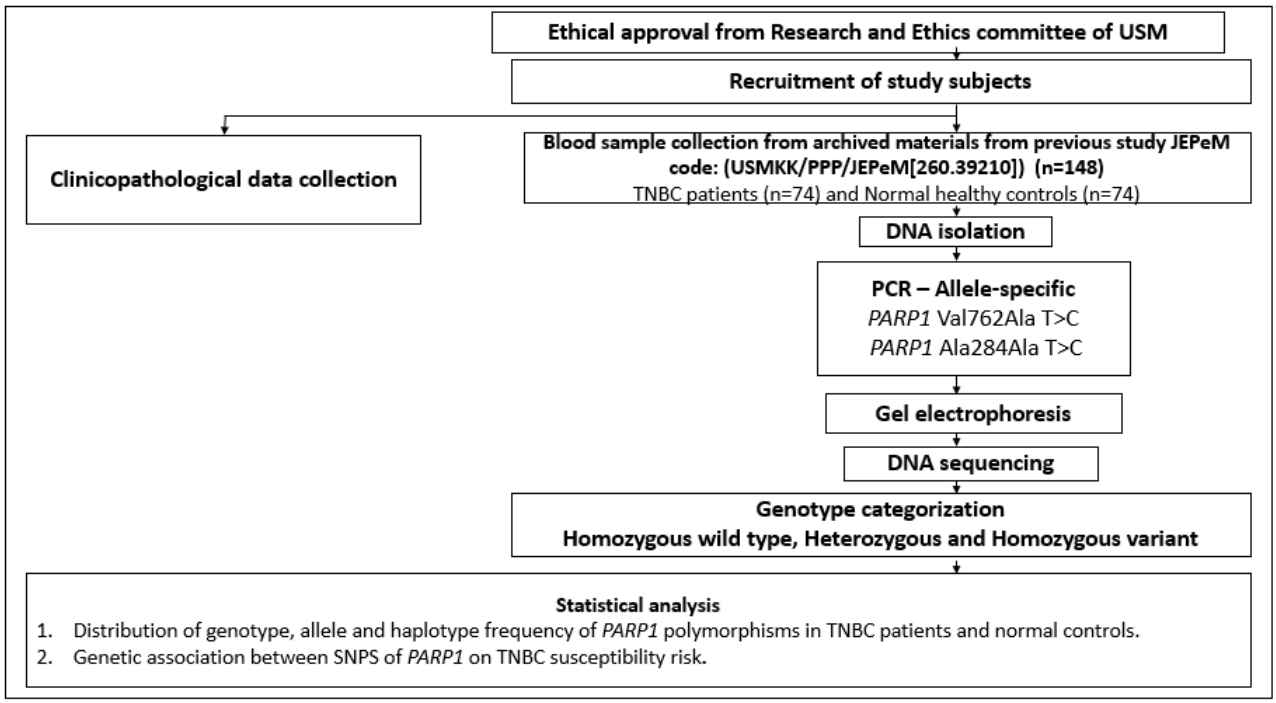
*Wild forward, **Mutant forward, ***Common reverse

Next, electrophoresis will be performed using 2% agarose gel to observe the presence of DNA band amplicons. After genotyping, the genotypes will be categorized into three groups; homozygous wild type (major), heterozygous and homozygous variant (minor).

Purification and sequencing of PCR products

A proportion of the PCR products will be chosen at random to be subjected to purification and sequencing for verification of the polymorphisms. PCR product will be purified using QIAquick PCR Purification Kit (QIAGEN, Germany). Briefly, 20 µl of PCR product will be added into 1.5 ml tube containing 100 µl of PB buffer. The mixture will be mixed and transferred into QIAquick column. Next, the column will be centrifuged at 13000 rpm for 30 seconds and the flow through out from the column will be discarded. Next, 750 µl of PE buffer will be added into the column and centrifuged at 13000 rpm for 30 seconds. Finally, 20 µl of buffer EB will be added to elute the DNA. Then, a total of 20 µl of purified PCR product will be sent to the Apical Scientific Sdn. Bhd. (Malaysia) for sequencing.

Study flowchart



Data analysis

Statistical analysis shall be performed by using SPSS version 27.0. For genotyping analysis in the first objective, genotype, allele and haplotype frequencies for each SNPs of *PARP1* shall be tested for deviation from Hardy-Weinberg equilibrium by using Pearson chi-square (X^2) test or Fisher exact test and will be compared among TNBC patients and normal healthy controls. As for the second objective, the association of genotype, allele and haplotype of *PARP1* (either singly and in combination) between TNBC patients and normal healthy controls shall be evaluated using logistic regression analysis by deriving odds ratio (ORs) and 95% confidence interval (CI) using unconditional logistic regression analysis. For all analyses, $P<0.05$ shall be considered as statistically significance.