POLYMORPHISMS IN ESTROGEN METABOLIZING GENES AND THEIR ASSOCIATION WITH ESTROGEN RECEPTOR-POSITIVE BREAST CANCER AMONG PATIENTS ATTENDING AGA KHAN UNIVERSITY HOSPITAL NAIROBI, AND AFRICA INLAND CHURCH HOSPITAL KIJABE

MARY KANYIRI MURITHI

DOCTOR OF PHILOSOPHY IN BIOCHEMISTRY

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

2026

Polymorphisms in Estrogen Metabolizing Genes and Their Association with Estrogen Receptor-Positive Breast Cancer among patients attending Aga Khan University Hospital Nairobi, and Africa Inland Church Hospital Kijabe

Mary Kanyiri Murithi

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy in Biochemistry of the Jomo Kenyatta University of Agriculture and Technology

2026

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University.

Signature……Date……...

Mary Kanyiri Murithi

This thesis has been submitted for examination with our approval as University Supervisors:

Signature……Date……...

Dr. Steven Ger Nyanjom, PhD

JKUAT, Kenya

Signature……Date……...

Dr. Francis Makokha, PhD

MKU, Kenya

Signature……Date……...

Dr. Victor Mobegi, PhD

UoN, Kenya

Signature……Date……...

Dr. Sayed Shahin, PhD

Aga Khan University Hospital, Kenya

DEDICATION

I dedicate this thesis to my beloved parents, Stephen Murithi M'Ngaruthi and Esther Mwari Murithi, whose love, wisdom, and sacrifices have shaped the person I am today. Your unwavering belief in me has been my greatest source of strength.

To my incredible siblings, Sammy Gitonga, Eunice Kinya, and Phyllis Kanana, your constant encouragement and love have made this journey less daunting and more meaningful.

To my partner, Fredrick Okeyo, your patience, support, and unwavering faith in me have been a guiding light, reminding me that I am never alone in this pursuit.

And to my best friend Teresa Kerubo, your kindness and steadfast encouragement have been a source of great inspiration.

This work is a testament to the love and support of each of you, and I am forever grateful to have you in my life.

ACKNOWLEDGEMENT

First and foremost, I am deeply grateful to God for granting me wisdom, good health, and resilience throughout this project. His guidance and strength have been my anchor during this journey.

I extend my heartfelt appreciation to my family for their unwavering support, encouragement, and patience throughout my studies. Their belief in me has been a constant source of motivation.

I sincerely thank my supervisors, Dr. Steven Ger, Dr. Victor Mobegi, Dr. Francis Makokha, and Dr. Sayeed Shahin, for their invaluable guidance, mentorship, and constructive feedback during this project. Their expertise and dedication have greatly contributed to the success of my work.

I am also deeply grateful to Mr. Erick Ouko and Dr. Patrick Njage for their mentorship, patience, and invaluable guidance in Data Analysis Using R Programming. Their expertise, willingness to provide support, and insightful explanations significantly enhanced my data analysis skills, contributing to the overall success of this research.

I extend my sincere appreciation to Aga Khan University Hospital, Nairobi, for allowing me to collect patient data and samples for this project. Special thanks to the nurses, Ms. Ann Karanu and Ms. Viviane Oluoch, for their assistance in obtaining hospital data and samples. I also extend my gratitude to Ms. Veronica Ngundo for her assistance and guidance with the ethical clearance application, ensuring compliance with ethical standards for this study.

I am grateful to Jomo Kenyatta University of Agriculture and Technology (JKUAT) and the Pan African University of Science, Technology, and Innovation (PAUSTI) for providing laboratory resources that were crucial for conducting this research.

I appreciate the support of the African-ai-Japan Research Grant – Project Innovation Research Fund (2020/2021), which provided initial funding for troubleshooting and optimizing study protocols.

I extend my deepest gratitude to the Mawazo Institute for providing a training platform through the Mawazo Learning Exchange Fellowship (2021), where I received valuable training in grant writing, budgeting and financial management, research ethics, science communication, public engagement, and leadership. This training significantly enhanced my research skills and professional development. I am also grateful to Mawazo Institute for their research grant support, which contributed to the successful completion of this study.

I am also thankful to the L'Oréal-UNESCO For Women in Science Sub-Saharan Africa (2020) for their leadership training and research grant, which significantly contributed to my professional growth.

I am deeply thankful to my colleagues at the School of Pharmacy, Kabarak University, for their unwavering moral support, encouragement, and camaraderie throughout my study. Their collaboration and positive energy created an enriching academic environment.

Lastly, I acknowledge all friends, research collaborators, and institutions that contributed to this work in one way or another. Their support, insights, and encouragement have been invaluable.

This journey has been made possible through the collective efforts of many individuals and organizations, and I am truly grateful for each contribution.

TABLE OF CONTENTS

DECLARATION ..... **ii**

DEDICATION ..... **iii**

ACKNOWLEDGEMENT ..... **iv**

TABLE OF CONTENTS ..... **vi**

LIST OF TABLES ..... **xiii**

LIST OF FIGURES ..... **xv**

LIST OF APPENDICES ..... **xvii**

ACRONYMS AND ABBREVIATIONS ..... **xviii**

ABSTRACT ..... **xxii**

CHAPTER ONE **1**

INTRODUCTION **1**

1.1 Background 1

1.2 Statement of the Problem 4

1.3 Justification of the Study 5

1.4 Research Questions 5

1.5 Objectives 6

1.5.1 General Objective 6

1.5.2 Specific Objectives....6

1.6 Scope of the Study....7

1.7 Limitations of the Study....8

CHAPTER TWO....9

LITERATURE REVIEW....9

2.1 General Overview of Cancer....9

2.2 Overview of Breast Cancer....9

2.3 Epidemiology....10

2.4 Breast Cancer Classification....13

2.4.1 Molecular Classification....13

2.4.2 Histopathological Classification....15

2.5 Benign Breast Disease....17

2.6 Breast Cancer Risk Factors....17

2.6.1 Non-Modifiable Factors....18

2.6.2 Modifiable Factors....20

2.7 Molecular Mechanisms of Breast Cancer....21

2.7.1 Hereditary and Somatic Genetic Alterations....21

2.7.2 Systemic and Metabolic Drivers: The Role of Obesity....22

2.7.3 Estrogen Metabolism and Genetic Susceptibility 22

2.8 Breast Cancer Screening and Diagnosis 24

2.8.1 Breast Cancer Screening 24

2.8.2 Diagnostic Imaging 25

2.8.3 Molecular Testing and Biochemical Markers 25

2.9 Breast Cancer Treatment 26

2.9.1 Surgery 26

2.9.2 Chemotherapy 26

2.9.3 Radiotherapy 27

CHAPTER THREE 28

MATERIALS AND METHODS 28

3.1 Study Sites 28

3.2 Study Design 29

3.3 Ethical Consideration 30

3.4 Study Participants 30

3.5 Sampling Method 30

3.6 Eligibility Criteria 31

3.6.1 Inclusion Criteria 31

3.6.2 Exclusion Criteria....32

3.7 Sample Size Determination....32

3.8 Data Collection....32

3.9 DNA Extraction....33

3.10 Polymerase Chain Reaction....34

3.11 Genotyping Using Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP)....35

3.12 Genotyping of CYP1B1 Polymorphisms by Sanger Sequencing....37

3.13 Bioinformatics Analysis of Sequences....37

3.14 Statistical Analysis....38

3.15 Association Analysis....38

3.15.1 Association with Epidemiological and Clinical Factors....38

3.15.2 Association of Genetic Variants with Clinicopathological Characteristics....38

3.15.3 Genotype Association Analysis....39

3.16 In-Silico Analysis of CYP1B1 Haplotypes....39

3.16.1 Linkage Disequilibrium and Haplotype Construction....39

3.16.2 Retrieval of the Wild-Type Cyp1b1 Sequence from UniProtKB and Generation of Haplotype Variants....40

3.16.3 Tertiary Structure Analysis....40

CHAPTER FOUR **42**

RESULTS **42**

4.1 Genomic DNA Yield and Quality 42

4.2 PCR Amplicons of Target Regions 43

4.3 Restriction Fragment Sizes of PCR amplicons 47

4.3.1 Msp1 Restriction Digestion of rs4646903 47

4.3.2 BsrD1 Restriction Digestion of rs1048943 48

4.3.3 AcuI Restriction Digestion of rs1056836 49

4.3.4 SspI Restriction Digestion of rs776746 50

4.3.5 NlaIII Restriction Digestion of rs4680 51

4.4 Sanger Sequencing of CYP1B1 Gene 52

4.5 Descriptive Analysis of Socio-Demographic Characteristics, Risk Factors, and SNPs in Study Participants 54

4.6 Descriptive Analysis of CYP1B1 SNPs in ER+ Breast Cancer and Control Participants 60

4.7 Descriptive Analysis of CYP1B1 SNPs in ER+ Breast Cancer and Benign Breast Diseases Participants 61

4.8 Tumor Characteristics in Estrogen Receptor-Positive Breast Cancer Cases 62

4.9 Association between Socio-Demographics and Risk factors with ER+ Breast Cancer Compared to Controls 64

4.10 Association between Selected Variables and ER+ Breast Cancer Compared to Benign Breast Diseases....65

4.11 Association of rs4646903, rs1048943, rs1056836, rs776746 and rs4680 with ER+ Breast Cancer Compared to Controls....66

4.12 Associations of rs4646903, rs1048943, rs1056836 SNPs with ER+ Breast Cancer Compared to Benign Breast Diseases....68

4.13 Association of CYP1B1 SNPs with ER+ Breast Cancer Risk Compared to Controls....70

4.14 Association of CYP1B1 SNPs with ER+ Breast Cancer Risk Compared to Benign Breast Diseases....71

4.15 Association of rs4646903, rs1048943, rs1056836, rs776746 and rs4680 SNPs with ER+ Breast Cancer Tumor Characteristics....73

4.16 Association of CYP1B1 SNPs with ER+ Breast Cancer Tumor Characteristics.76

4.17 In-Silico Analysis of CYP1B1 Haplotypes....77

4.17.1 Linkage Disequilibrium Analysis....77

4.17.2 Haplotype Frequencies of CYP1B1 Polymorphisms....78

4.17.3 Wild-Type Sequence Retrieval and Haplotype Variant Generation....79

4.17.4 Tertiary Structure Analysis....82

CHAPTER FIVE....87

DISCUSSION....87

5.1 Overview 87

5.2 Prevalence of the Socio-Demographics, Risk Factors and Single Nucleotide Polymorphisms in the Study Participants 88

5.3 Prevalence of Tumor Characteristics 91

5.4 Associations of Socio-Demographics and Risk Factors with ER^+ Breast Cancer 93

5.5 Associations between Single Nucleotide Polymorphisms and ER^+ Breast Cancer 94

5.6 Associations between Single Nucleotide Polymorphisms and Tumor Characteristics 96

5.7 In-Silico Analysis of CYP1B1 Haplotypes 97

5.7.1 Linkage Disequilibrium 97

5.7.2 Haplotype Distribution and Breast Cancer Risk 97

5.7.3 Functional Domain Mapping of Polymorphic Residues 98

5.7.4 Tertiary Structure Analysis 99

CHAPTER SIX 101

CONCLUSIONS AND RECOMMENDATIONS 101

6.1 Conclusions 101

6.2 Recommendations 101

REFERENCES 103

APPENDICES 125

LIST OF TABLES

Table 2.1: Description of rs4646903, rs1048943, rs1056836, rs776746 and rs4689 SNPs 24

Table 3.1: Primers and PCR Products of the Target Regions 35

Table 3.2: Restriction Enzymes and Conditions Used for PCR-RFLP Genotyping.....36

Table 4.1: Socio-Demographic Characteristics, Risk Factors, and SNPs Distributions in ER+ Breast Cancer Cases and Controls 57

Table 4.2: Genotype Frequencies of CYP1B1 SNPs among ER+ Breast Cancer Cases and Controls 61

Table 4.3: Genotype Frequencies of CYP1B1 SNPs among ER+ Breast Cancer Cases and BBD Participants 62

Table 4.4: Tumor Characteristics of ER+ Breast Cancer Cases 63

Table 4.5: Association between Socio-Demographics and Risk Factors with ER+ Breast Cancer Compared to Controls 64

Table 4.6: Association between Socio-Demographics and Risk Factors with ER+ Breast Cancer Compared to BBD Participants.....65

Table 4.7: Genotype Frequencies and Associations of rs4646903 and rs1056836 Variants in ER+ Breast Cancer Versus Control Participants 67

Table 4.8: Genotype Frequencies and Associations of rs4646903, rs1048943 and rs1056836 variants in ER+ Breast Cancer Versus Benign Breast Disease Participants 69

Table 4.9: Genotype Frequencies and Associations of rs10012, rs1056827 and rs1056836 Variants in ER+ Breast Cancer Versus Control Participants 70

Table 4.10: Genotype Frequencies and Associations of rs10012, rs1056827 and rs1056836 Variants in ER+ Breast Cancer Versus Benign Breast Disease Participants 72

Table 4.11: Frequencies and Association of rs4646903 Variants With Tumor Characteristics in ER+ Breast Cancer Patients 74

Table 4.12: Summary of Pairwise Linkage Disequilibrium among rs10012, rs1056827 and rs1056836 Polymorphisms 77

Table 4.13: Haplotype Frequencies of CYP1B1 SNPs in Cases, Controls, and Benign Samples 78

Table 4.14: Functional Domain Mapping of Target Residues 79

LIST OF FIGURES

Figure 2.1: Global Map Representing Absolute Numbers of Incidence of Cancer Types in Females 11

Figure 2.2: Global Map Representing Absolute Numbers of Mortalities of Cancer Types in Females 12

Figure 3.1: Map Showing Sampling Sites 29

Figure 4.1: Gel Photo of Extracted Genomic DNA 42

Figure 4.2: Gel Photo of rs4646903 Amplicons 43

Figure 4.3: Gel Photo of rs1048943 Amplicons. 44

Figure 4.4: Gel Photo of rs1056836 Amplicons. 44

Figure 4.5: Gel Photo of rs776746 Amplicons. 45

Figure 4.6: Gel Photo of rs4680 Amplicons 45

Figure 4.7: Gel photo of rs1056836 and rs1800440 Amplicons. 46

Figure 4.8: Gel Photo of rs10012 and rs1056827 Amplicons. 47

Figure 4.9: Gel Photo of PCR-RFLP Analysis for rs4646903 48

Figure 4.10: Gel Photo of RFLP-PCR for rs1048943 49

Figure 4.11: Gel Photo of RFLP-PCR for rs1056836 50

Figure 4.12: Gel Photo of PCR-RFLP for rs776746 51

Figure 4.13: Gel Image of RFLP-PCR for rs4680 52

Figure 4.14: Sanger Sequences of rs10012 and rs1056827 SNPs in CYP1B1.....53

Figure 4.15: Sanger Sequences of rs1800440 and rs1056836 SNPs in CYP1B1.....53

Figure 4.16: Boxplot Comparison of Age, BMI, and Age at Menarche among ER+ Breast Cancer Cases, Benign Breast Disease and Healthy Participants 59

Figure 4.17: CYP1B1 Haplotype Amino Acid Sequences 82

Figure 4.18: Tertiary Structures and Ramachandran Plot of CYP1B1 Highlighting the R48, A119, L432 and N453 Residues.....83

Figure 4.19: Tertiary Structures and Ramachandran Plot of CYP1B1 Highlighting the G48, S119, V432 and N453 Residues.....84

Figure 4.20: Tertiary Structures And Ramachandran Plot of CYP1B1 Highlighting the R48, A119, V432 and N453 Residues 85

Figure 4.21: Tertiary Structures and Ramachandran Plot of CYP1B1 Highlighting the R48, S119, V432 and N453 Residues.....86

LIST OF APPENDICES

Appendix I: Ethical Approval.....125

Appendix II: NACOSTI Permit.....126

Appendix III: Consent Page 127

Appendix IV: Data Abstraction Form 128

Appendix V: Participants Group Classification and Identifiers 130

Appendix VI: Participants Demographic, Clinical and Lifestyle Characteristics 132

Appendix VII: Genotypes for rs4646903, rs1048943, rs1056836, rs7767746 and rs4680 138

Appendix VIII: Sanger Sequences for rs1800440 and rs1056836 in exon 3 of CYP1B1 142

Appendix IX: Sanger Sequences for rs10012 and rs1056827 exon 2 of CYP1B1.....209

Appendix X: Genotypes for rs1800440, rs1056836, rs1056827 and rs10012.....284

Appendix XI: Clinicopathological Characteristics of Case Group Participants.....287

Appendix XII: Publication Abstracts 289

ACRONYMS AND ABBREVIATIONS

A	Adenine
AAM	Age at Menarche
AIs	Aromatase Inhibitors
AKUHN	Aga Khan University Hospital Nairobi
Ala	Alanine
Arg	Arginine
ASIR	Age-Standardized Incidence Rate
Asn	Asparagine
ATM	Ataxia-Telangiectasia Mutated
BARD1	BRCA1 Associated RING Domain 1
BC	Breast Cancer
BCS	Breast-Conserving Surgery
BBD	Benign Breast Disease
BMI	Body Mass Index
BRCA1	Breast Cancer Gene 1
BRCA2	Breast Cancer Gene 2
BRIP1	BRCA1 Interacting Protein C-Terminal Helicase 1
C	Cytosine
CBE	Clinical Breast Examination
CDH1	Cadherin-1
CHEK2	Checkpoint Kinase 2
CI	Confidence Interval
COMT	Catechol-O-Methyltransferase
CT	Computed Tomography

CYP	Cytochrome P450
CYP1A1	Cytochrome P450 Family 1 Subfamily A Member 1
CYP1B1	Cytochrome P450 Family 1 Subfamily B Member 1
CYP3A5	Cytochrome P450 Family 3 Subfamily A Member 5
DCIS	Ductal Carcinoma In Situ
df	Degrees of Freedom
DNA	Deoxyribonucleic Acid
ER+	Estrogen Receptor-Positive
ER+	Estrogen Receptor-Positive
FFPE	Formalin-Fixed Paraffin-Embedded
FISH	Fluorescence In Situ Hybridization
G	Guanine
GSTs	Glutathione S-transferases
HAAs	Heterocyclic Aromatic Amines
HDI	Human Development Index
HER2+	Human Epidermal Growth Factor Receptor 2 Positive
IARC	International Agency for Research on Cancer
IC	Invasive Carcinoma
IDC	Invasive Ductal Carcinoma
IHC	Immunohistochemistry
ILC	Invasive Lobular Carcinoma
Ile	Isoleucine
IMCS	Internal Mammary Chain Sentinel Node
IGF-1	Insulin-Like Growth Factor-1
IQR	Interquartile Range
IPC	Intracystic Papillary Carcinoma

KAIC	Africa Inland Church Hospital Kijabe
LD	Linkage Disequilibrium
Leu	Leucine
Lnlike	Log-Likelihood
LR	Likelihood Ratio
LVI	Lymphovascular Invasion
MAPK	Mitogen-Activated Protein Kinase ;pathway
MEGA	Molecular Evolutionary Genetics Analysis
Met	Methionine
MC	Mucinous Carcinoma
MRI	Magnetic Resonance Imaging
mRNA	Messenger Ribonucleic Acid
NATs	N-acetyltransferases
NST	Invasive Breast Cancer of No Special Type
OR	Odds Ratio
PAHs	Polycyclic Aromatic Hydrocarbons
PALB2	Partner and Localizer of BRCA2
PDCIS	Papillary Ductal Carcinoma In Situ
PCR-RFLP	Polymerase Chain Reaction-Restriction Fragment Length Polymorphism
PET	Positron-Emission Tomography
PTEN	Phosphatase and Tensin Homolog
RAS	Renin-Angiotensin System
rÂ²	Correlation Coefficient
Ser	Serine
SERDs	Selective Estrogen Receptor Degraders
SERMs	Selective Estrogen Receptor Modulators

SHBG	Sex Hormone-Binding Globulin
SNP	Single Nucleotide Polymorphism
SPECT	Single-Photon Emission Computed Tomography
SSA	Sub-Saharan Africa
STK11	Serine/Threonine Kinase 11
T	Thymine
TNBC	Triple-Negative Breast Cancer
TP53	Tumor Protein p53
Val	Valine
VEGF	Vascular Endothelial Growth Factor
WHO	World Health Organization

ABSTRACT

Breast cancer continues to be the most prevalent malignancy in women worldwide, with estrogen receptor-positive (ER+) tumors accounting for approximately 70% of cases. GLOBOCAN 2022 data reveal a significant global burden, with 2.3 million new cases annually, including 198,553 in Africa. Kenya reports 7,243 new cases and 3,107 deaths yearly, reflecting urgent needs for improved early detection and prevention strategies. This hospital-based case-control study aimed to determine the associations between socio-demographics, medical history, reproductive history, lifestyle factors and single nucleotide polymorphisms (SNPs) in estrogen-metabolizing genes with ER+ breast cancer risk among Kenyan women. The study compared 64 ER+ breast cancer cases with 79 benign breast disease (BBD) and 19 healthy controls from Aga Khan University Hospital Nairobi and Africa Inland Church (AIC) Kijabe Hospital. Socio-demographic and clinical data were abstracted from the questionnaires and medical records review. Estrogen plays a pivotal role in the pathogenesis of ER+ breast cancer. Individual genetic variation in estrogen-metabolizing enzymes can significantly alter the production, activity, and clearance of estrogen and its metabolites, thereby modifying cancer risk and progression. Single nucleotide polymorphisms (SNPs) in genes such as CYP1A1, CYP1B1, CYP3A5, and COMT can lead to differential enzymatic activity, influencing the critical balance between carcinogenic catechol estrogens and their detoxified forms. Consequently, these SNPs are considered key biomarkers for understanding interindividual susceptibility and prognosis in ER+ breast cancer. Five functionally relevant single nucleotide polymorphisms (SNPs) in key estrogen metabolism genes including; rs4646903 and rs1048943 of CYP1A1; rs1056836 of CYP1B1; rs776746 of CYP3A5 and rs4680 of COMT were analyzed using polymerase chain reaction restriction fragment length polymorphism (PCR-RFLP). Four additional SNPs (rs10012, rs1056827, rs1056836 and rs1800440 of CYP1B1) were analyzed via Sanger sequencing methods. Key findings demonstrated that women aged 50 years and older and postmenopausal women faced significantly elevated breast cancer risk. Genetic analysis revealed complex patterns: the alternative C allele of rs4646903 in CYP1A1 showed a protective effect against ER+ cancer development (OR=0.44, 95% CI [0.19-0.99], p = 0.048), but was also paradoxically associated with increased risk of aggressive Luminal B subtypes (OR=3.83, 95% CI [1.35-10.84]). Two CYP1B1 variants – alternative C allele in rs1056836 (OR = 0.34, 95% CI [0.19–0.62], p = 0.0003) and alternative A allele in rs1056827 (OR=0.43, 95% CI [0.19-0.98], p = 0.045) – were associated with reduced likelihood of malignant transformation from benign breast disease. Linkage disequilibrium analysis revealed strong association between rs10012 and rs1056827 ($D' = 0.9466$, $r^2 = 0.5767$, $p = 1.08 \times 10^{-13}$), confirming that these loci form a haplotype block. Haplotype frequency analysis identified eight distinct CYP1B1 haplotypes, with the G–A–C–T haplotype (R–S–V–N) predominating in benign samples (42.3%) suggesting a protective role, while the G–C–C–T haplotype (G–A–V–N) showed elevated frequency in cases (12.5%) compared to controls (8.3%), indicating potential risk association. Tertiary structure modeling using AlphaFold Server revealed that the four polymorphic residues—R48G, A119S, L432V, and N453S—occupy functionally distinct domains: R48 in the N-terminal membrane

interaction region, A119 in the structural core, L432 near the substrate-binding pocket, and N453 in the heme-binding region. Comparative structural analysis of the wild-type (R–A–L–N), reference (G–S–V–N), risk-associated (G–A–V–N), and protective (R–S–V–N) haplotypes demonstrated that the protective haplotype uniquely retains the wild-type R48 residue while incorporating S119 and V432, suggesting that proper membrane anchoring may be critical for maintaining protective function. Conversely, the risk-associated haplotype combines loss of R48 with the high-activity V432 variant, potentially synergistically enhancing carcinogenic estrogen metabolism. Ramachandran plot analysis confirmed structural reliability, with >90% of residues in favored regions across all models. These findings provide important insights into the complex interplay between socio-demographic factors, genetic predisposition, and structural consequences of CYP1B1 haplotypes in ER+ breast cancer development among Kenyan women. The study's findings underscore the importance of developing population-specific risk assessment tools that incorporate both genetic and structural information to combat Kenya's disproportionate breast cancer mortality burden. Future research should focus on molecular dynamics simulations to elucidate the dynamic mechanisms underlying haplotype-specific functional differences, validation in larger cohorts, and exploration of potential clinical applications in preventive strategies and personalized treatment approaches.

CHAPTER ONE

INTRODUCTION

1.1 Background

Breast cancer continues to be the most commonly diagnosed malignancy among women worldwide. According to GLOBOCAN 2022, an estimated 2,308, 897 new cases were reported, accounting for 23.8% of all female cancer cases, with 665,684 deaths, representing 15.4% of all female cancer-related deaths. In Africa, the trend is similar, with breast cancer ranking highest in both incidence - 198,553 new cases (29.2%) and 91,252 deaths (21.9%). Breast cancer is the most common female malignancy in Sub-Saharan Africa (SSA) (133,520 annual cases; 26.5% of cancers), yet ranks second in mortality (68,036 deaths; 21.1%) after cervical cancer. Despite lower incidence than high-income countries, SSA experiences disproportionately poor survival due to advanced-stage diagnosis and limited treatment access (Parenté et al., 2025).

In Kenya breast cancer accounts for 7,243 new cases making, it the most common cancer. It is ranked second in mortalities after cervical cancer and accounts for 3,398 deaths. The incidence and mortalities represent 25.5% and 18.9% of all the female cancers, respectively (International Agency for Research on Cancer, 2022). Majority of the breast cancer cases in Kenya were reported to have late-stage disease with a 5-year relative survival of 40.1 to 64% (Ekpe et al., 2019). The low rates of survival are due to patient and health-system barriers such as low awareness, high cost of screening, diagnosis as well as treatment. The disease therefore impacts negatively on the economic growth of the country, and also leads to loss of productivity due to the cancer-related premature deaths (Hutchinson et al., 2024).

Breast cancer incidence distribution based on the four-tiers Human Development Index (HDI) countries including; very high, high, medium, and low HDI level (excluding China and India in high and medium HDI, respectively) varies (Bray et al., 2024)The age-

standardized incidence rate (ASIR) for breast cancer is 54.1 per 100,000 in transitioned (high or very high HDI) countries and 38.1 per 100,000 in transitioning (low or medium HDI) countries (Sung et al., 2021). The higher incidence in transitioned countries is associated with increased exposure to breast cancer-associated risk factors, including obstetric and gynecological factors (age at menarche, menopause, and parity), race/ethnicity, lifestyle choices, diet, physical inactivity, obesity, medical history, and genetic predispositions (Al-Shami et al., 2023; Coughlin & Smith, 2015).

Despite the higher incidence in transitioned countries, mortality rates tend to be lower due to widespread screening, early detection, and advanced treatment options. In contrast, transitioning countries in South America, Africa, and Asia continue to experience both rising incidence and high mortality rates, largely due to late-stage diagnosis, inadequate healthcare infrastructure, and limited access to treatment. In SSA, where healthcare resources remain constrained, breast cancer mortality remains disproportionately high (Bray et al., 2024).

Estrogen exposure plays a critical role in breast cancer pathogenesis, particularly in estrogen receptor-positive (ER+) breast cancer (Yager et al., 2006). Estrogens undergo metabolism through two primary pathways: the oxidative pathway, mediated by cytochrome P450 (CYP) enzymes, and the conjugation pathway, regulated by enzymes such as catechol-O-methyltransferase (COMT) (Nebert & Dalton, 2006). Genetic polymorphisms in CYP1A1, CYP1B1, CYP3A5, and COMT influence estrogen metabolism efficiency and the balance between protective (2-hydroxyestradiol) and carcinogenic (4-hydroxyestradiol) estrogen metabolites (Coughlin & Piper, 1999).

Single nucleotide polymorphisms (SNPs) in estrogen metabolizing genes may affect individuals' susceptibility to estrogen receptor-positive (ER^+) breast cancer through altered enzyme activity, promoting oxidative stress, DNA damage, and activating xenobiotics such as polycyclic aromatic hydrocarbons (PAHs) and endogenous compounds such as estrogen (Abdelmonem et al., 2024).

The CYP1A1 gene metabolizes estradiol to 2-hydroxy estradiol, a weak carcinogen. Two nonsynonymous mutations described in the CYP1A1 include; the rs4646903 thymine (T) to cytosine (C) point mutation at nucleotide 6235 in 3' untranslated region (3' UTR)) and rs1048943 adenine (A) to guanine (G) base substitution at nucleotide 2455 resulting in substitution of isoleucine (Ile) to valine (Val) at codon 462 in exon 7 of CYP1A1 (Zhan et al., 2011).

The CYP1B1 gene preferentially converts estradiol to 4-hydroxy estradiol which is highly carcinogenic. The most common CYP1B1 gene polymorphisms may influence breast cancer risk (Wen et al., 2017a). They include; rs10012 which results in substitution of arginine (Arg) to Glycine (Gly) at codon 48, rs1056827 which substitutes alanine (Ala) to serine (Ser) at codon 119, rs1056836 which substitutes leucine (Leu) to Valine (Val) at codon 432 and rs1800440 which substitutes asparagine (Asn) to serine (Ser) at codon 453 (Jiao et al., 2010)

Polymorphisms in the CYP3A gene involved in carcinogen metabolism may result in variations in breast cancer susceptibility between individuals. The CYP3A5, rs776746 is one of the predominant polymorphisms in CYP3A gene. Substitution of guanine (G) to adenine (A) in intron 3 of CYP3A5 may lead to a splice defect of the mRNA and result in the formation of nonfunctional protein (Badavi et al., 2015). Catechol-O-methyl transferase (COMT) is a phase II gene involved in conjugation and inactivation of catechol estrogens. Insufficient or lack of COMT due to transition of guanine (G) to adenine (A) of rs4680 and consequent substitution of valine (Val) to methionine (Met) amino acid at codon 158 may lead to accumulation of quinones leading to DNA adducts formation and tumor initiation (Almeida et al., 2021; Qin et al., 2012).

Understanding the role of these genetic variations in ER+ breast cancer patients in Kenya is crucial, given the regional differences in genetic predisposition and environmental exposures. This study investigated polymorphisms in low-penetrance genes involved in estrogen metabolism, specifically those in phase I (cytochrome P450 family) and phase II (catechol-O-methyltransferase) pathways. Phase I gene polymorphisms analyzed included

rs4646903 and rs1048943 in CYP1A1; rs10012, rs1056827, rs1056836, and rs1800440 in CYP1B1; and rs776746 in CYP3A5. In phase II, the rs4680 polymorphism in COMT was examined. These variants were studied in women diagnosed with estrogen receptor-positive (ER^+) breast cancer, women with benign breast disease, and healthy controls attending Aga Khan University Hospital, Nairobi, and Africa Inland Church (AIC) Kijabe Hospital between 2019 and 2021. The study also assessed associations between ER^+ breast cancer risk and demographic or probable risk factors.

1.2 Statement of the Problem

Breast cancer imposes a significant health burden in Kenya, particularly among women of reproductive and middle age. The prevalence is estimated at 34 per 100,000 women and it contributes to 23% of female cancer cases and 11% of total cancer-related mortality expected to rise (Hutchinson et al., 2024).

Early diagnosis and screening remain major challenges in low-resource settings, leading to late-stage detection when treatment is often less effective and more costly. This contributes to high mortality, psychological distress, disability, reduced productivity, and financial strain for affected individuals and their families. The increasing burden of breast cancer also strains the healthcare system, diverting government resources from prevention to expensive curative care. Households not covered by health insurance face a high economic burden due to high out-of-pocket spending. At the community level, the disease disrupts family dynamics, as many women serve as primary caregivers and income earners. Its impact on mental health is significant—higher cancer severity in Kenyan women has been linked to increased psychological distress (Ndetei et al., 2018).

Kenya has a high prevalence of risk factors for cancer including genetic predisposition, lifestyle, environmental and reproductive factors (Ministry of Health, 2022). Although understanding these factors in the local context could improve prevention, early detection, and guide treatment strategies, they remain largely unexplored.

1.3 Justification of the study

There is low uptake of screening services in Kenya with only 1% of women screened for breast cancer with mammography (Ministry of Health, 2020). There is also low awareness of risk factors, hence the need to incorporate risk factors, screening and awareness in national health surveys. This will improve on early diagnosis and will reduce on the costs of treatment by the government, reduce the emotional distress in families and improve on the quality of life for the survivors.

In Kenya, late-stage breast cancer diagnosis contributes to poor prognosis and high mortality. Understanding the risk factors associated with breast cancer is crucial for improved monitoring, evaluation. This will also guide on implementation of policies and guidelines for cancer prevention.

Estrogen metabolism plays a crucial role in breast cancer initiation and progression. This may be attributed to deregulation of genes which codify enzymes involved in estrogen metabolism and exposure to high levels of estrogen and its metabolites. However, the association between polymorphisms in estrogen-metabolizing genes and breast cancer risk remains understudied in African populations. Most genetic studies have focused on Western and Asian populations, leaving a knowledge gap in African-specific genetic risk factors. Understanding the association of polymorphisms in the local population with breast cancer may contribute to development of molecular panels that can be used for screening and early diagnosis of the disease

1.4 Research questions

i. What is the association between socio-demographics with ER+ breast cancer patients attending Aga Khan University Hospital, Nairobi and Africa Inland Church Hospital Kijabe
ii. What is the association between reproductive, medical history, lifestyle, and single nucleotide polymorphisms in estrogen-metabolizing genes with ER+ breast cancer

patients attending Aga Khan University Hospital, Nairobi and Africa Inland Church Hospital Kijabe

iii. What is the association of single nucleotide polymorphisms in estrogen metabolizing genes with the clinicopathological characteristics in estrogen receptor positive breast cancer patients attending Aga Khan University Hospital, Nairobi and Africa Inland Church Hospital Kijabe

iv. What are the structural and functional consequences of the CYP1B1 haplotypes defined by the R48G (rs10012), A119S (rs1056827), L432V (rs1056836), and N453S (rs1800440) polymorphisms as predicted by in silico approaches?

1.5 Objectives

1.5.1 General Objective

To determine association of polymorphisms in estrogen-metabolizing genes, in conjunction with socio-demographics, medical history, lifestyle and reproductive factors, with estrogen receptor-positive breast cancer in patients attending Aga Khan University Hospital, Nairobi and Africa Inland Church Hospital Kijabe

1.5.2 Specific Objectives

i. To determine the association of socio-demographics with estrogen receptor-positive breast cancer patients attending Aga Khan University Hospital, Nairobi and Africa Inland Church Hospital Kijabe

ii. To assess the relationship between medical, lifestyle, environmental, reproductive factors and polymorphisms in estrogen metabolizing genes with estrogen receptor-positive breast cancer patients attending Aga Khan University Hospital, Nairobi and Africa Inland Church Hospital Kijabe

iii. To determine the association of single nucleotide polymorphisms in estrogen-metabolizing genes on clinicopathological characteristics in estrogen receptor positive breast cancer

iv. To investigate the potential structural and functional effects of four common CYP1B1 mutations—R48G (rs10012), A119S (rs1056827), L432V (rs1056836), and N453S (rs1800440) using in-silico approaches

1.6 Scope of the Study

This study investigated the influence of socio-demographic characteristics, risk factors, and single nucleotide polymorphisms (SNPs) in estrogen-metabolizing genes on the risk and progression of estrogen receptor-positive (ER^+) breast cancer.

Specifically, it examined SNPs involved in the phase I and phase II pathways of estrogen metabolism. Phase I polymorphisms analyzed included rs4646903 and rs1048943 in CYP1A1; rs10012, rs1056827, rs1056836, and rs1800440 in CYP1B1; and rs776746 in CYP3A5. In phase II, rs4680 in the COMT gene was evaluated. The study assessed the potential associations of these variants with breast cancer susceptibility and tumor characteristics.

In addition, it explored the distribution of socio-demographic characteristics and risk factors among three groups: women diagnosed with ER^+ breast cancer, individuals with benign breast diseases, and healthy volunteers. This analysis provided insights into the prevalence and impact of modifiable and non-modifiable risk factors within the study population.

Furthermore, the study evaluated the association between the genetic polymorphisms and tumor characteristics, such as tumor stage, histological type and grade, lymph node involvement, and luminal subtype, to determine whether specific SNPs influenced disease progression and prognosis.

The study participants included women diagnosed with ER^+ breast cancer, individuals with benign breast diseases, and healthy volunteers from Aga Khan University Hospital, Nairobi, and Africa Inland Church Hospital, Kijabe, in the period between 2019 and 2021. The study period (January 2019–December 2021) was selected because this study was

nested within a larger study, hence electronic medical records and biorepository samples were consistently available and complete during this period at both sites.

1.7 Limitations of the Study

This study provides important insights into risk factors for ER+ breast cancer, though several limitations warrant consideration. The retrospective design introduces potential recall bias in self-reported exposures, mitigated through standardized data collection by trained interviewers using REDCap. Hospital-based sampling at two facilities yielded a modest sample size that precluded stratified analyses, with cases drawn from diverse regions but controls predominantly from Nairobi, introducing potential geographic bias. Funding constraints limited polymorphism coverage and prevented functional validation of findings.

Regarding the in silico structural analyses, AlphaFold 3 models represent static snapshots and do not capture dynamic conformational changes during substrate binding or catalysis; molecular dynamics simulations are required to address this limitation.

Despite these constraints, this study offers novel insights into estrogen-metabolizing gene polymorphisms, haplotype architecture, and structural consequences of CYP1B1 variants in ER+ breast cancer among Kenyan women. Future multicenter studies with population-based sampling, expanded genomic profiling, molecular dynamics simulations, and in vitro functional validation are needed to enhance generalizability and clinical utility.

CHAPTER TWO

LITERATURE REVIEW

2.1 General Overview of Cancer

Cancer is a non-communicable disease (NCD) which can be categorized based on the body part affected. It occurs when abnormal cells divide uncontrollably, infiltrate and destroy body tissues (Ministry of Health, 2020). In Kenya, Cancer is the third leading cause of NCDs after infectious diseases and cardiovascular diseases. The five common adult cancers in Kenya include cervix, breast, oesophageal, prostate and colorectal while in children the most common cancers include lymphomas, leukemias. Retinoblastomas, renal and bone (Kagiri H et al., 2020)

2.2 Overview of Breast Cancer

Breast cancer is a complex heterogenous disease characterized by the uncontrolled proliferation of mammary epithelial cells, leading to tumor formation and potential metastatic dissemination. While clinically defined by histology and receptor status, its fundamental etiology lies in the accumulation of genetic and epigenetic alterations that disrupt normal cellular homeostasis, including DNA repair, cell cycle control, and apoptotic signaling.

Globally, breast cancer represents a major public health challenge, exhibiting the highest incidence and a leading cause of cancer mortality among women (Bray et al., 2024). Although significant advancements in screening and treatment have improved outcomes in high-resource settings, the rising global burden underscores the imperative to move beyond generalized approaches and elucidate the specific molecular mechanisms driving individual risk and tumor behavior (Nolan et al., 2023).

The search for these mechanisms has historically focused on rare, high-penetrance germline mutations in genes such as BRCA1 and BRCA2, which account for a small

fraction (5-10%) of cases, predominantly within high-risk families. The vast majority of breast cancer cases are considered sporadic, with their etiology attributed to the complex interplay of common, low-penetrance genetic variants with environmental, reproductive, and lifestyle factors.

This polygenic risk architecture suggests that susceptibility is often modulated not by single, highly penetrant mutations, but by the combined effect of polymorphisms in genes governing critical biochemical pathways. A pathway of particular relevance is xenobiotic and hormone metabolism, mediated by enzymes such as the cytochrome P450 (CYP) superfamily. Polymorphisms in these genes can significantly alter enzyme activity, influencing the biotransformation of endogenous estrogens and environmental carcinogens, thereby modulating individual exposure to DNA-damaging intermediates.

Consequently, while hereditary syndromes provide crucial mechanistic insights, a comprehensive understanding of breast cancer etiology requires a shift in focus toward functional polymorphisms in metabolic genes and their interaction with environmental exposures. Elucidating these associations is critical for advancing beyond descriptive epidemiology to a mechanistic, biochemical understanding of risk, with the potential to inform more personalized strategies for risk prediction, early detection, and prevention.

2.3 Epidemiology

Breast cancer is the second most common malignancy globally after lung cancer, accounting for approximately 2,296,840 new cases annually, equivalent to 11.5% of all cancer cases (Bray et al., 2024). Among women, breast cancer ranks first in incidence at 23.8%. According to Global Cancer Statistics 2022, breast cancer was the most commonly diagnosed cancer in 158 countries (Figure 2.1). The incidence rates are notably high in both transitioned (54.1 per 100,000) and transitioning (30.8 per 100,000) countries (Sung et al., 2021).

Globally, breast cancer is responsible for approximately 666,000 deaths, constituting 6.8% of all cancer cases and 15.4% of female cancers. It was reported as the primary cause of cancer deaths in 111 of 187 countries (Figure 2.2), with the highest mortality rates observed in Melanesia, Western Africa, and Micronesia/Polynesia (Bray et al., 2024).

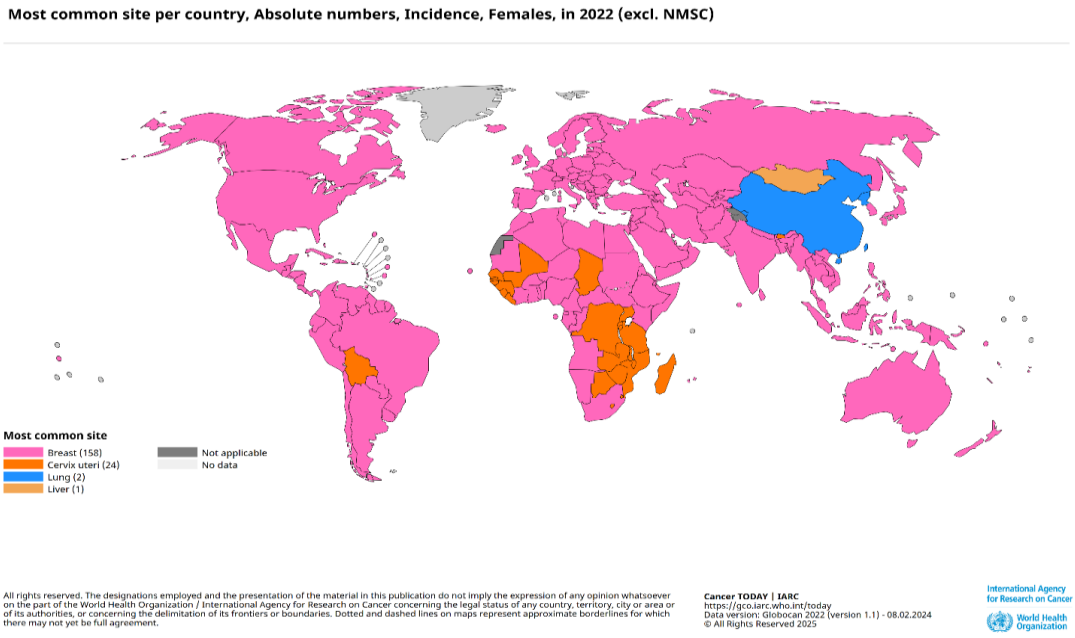


Figure 2.1: Global Map Representing Absolute Numbers of Incidence of Cancer Types in Females

Source: (Bray et al., 2024)

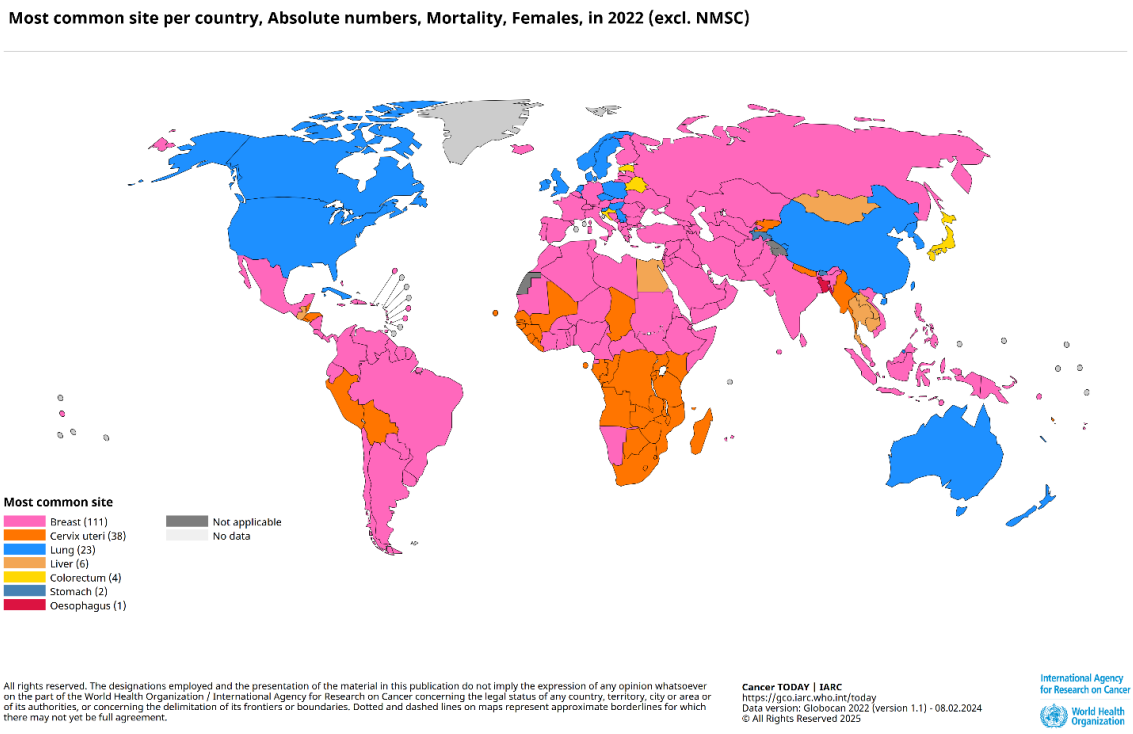


Figure 2.2: Global Map Representing Absolute Numbers of Mortalities of Cancer Types in Females

Source: (Bray et al., 2024).

In Africa, breast cancer leads in both incidence and mortality, with approximately 198,553 new cases and 91,252 deaths annually. The majority of these cases occur in Sub-Saharan Africa (SSA), where breast cancer ranks second in cancer-related mortality, with approximately 68,036 deaths (Anyigba et al., 2021). Despite a relatively low incidence compared to global figures, mortality rates remain disproportionately high, largely due to late-stage diagnosis, inadequate healthcare infrastructure, and limited access to treatment (Black & Richmond, 2019; Joko-Fru et al., 2020). The burden of breast cancer in SSA is expected to increase, highlighting the urgent need for improved screening, early diagnosis, and treatment strategies (Anyigba et al., 2021).

Breast cancer is among the five commonest cancers in Kenya. It is the most common malignancy among women with an incidence of 7243 and the second in causing deaths

with a mortality of 3398 mortalities with age standardized rates of 40.8/100000 and 19.6/100,000 respectively (International Agency for Research on Cancer, 2022). In Kenya, women are diagnosed with breast cancer at a younger age of 30-50 years compared to 50-55 years in the western countries. 7Seven out of ten breast cancers are diagnosed at late stages (stage III and IV) (Ministry of Health, 2021).

2.4 Breast Cancer Classification

2.4.1 Molecular Classification

Breast cancer is a complex disease with different molecular subtypes. Using the immunohistochemistry (IHC) technique breast cancer can be characterized by estrogen receptor (ER). Progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). Invasive breast cancer can be categorized into molecular subtypes based on mRNA gene expression profiles (Łukasiewicz et al., 2021). In 2000, Perou and colleagues initially classified breast cancer into four molecular subtypes: Luminal, HER2-enriched, Basal-like (triple-negative breast cancer), and Normal Breast-like (Prat & Perou, 2011). Further studies refined this classification by identifying two subcategories of the Luminal subtype, namely Luminal A and Luminal B (Prat et al., 2013) . The Normal Breast-like subtype was later omitted. Hence, the four widely recognized breast cancer subtypes based on mRNA gene expression are Luminal A, Luminal B, HER2-enriched, and Basal-like (triple-negative breast cancer) (Koboldt et al., 2012).

2.4.1.1 Luminal Breast Cancers

The luminal breast cancers which constitute the estrogen receptor-positive (ER+) tumors, account for nearly 70% of all breast cancer cases (Brufsky & Dickler, 2018). In ER+ breast cancer, cancer cells have receptors for the hormone estrogen, which binds to these receptors and promotes tumor growth (Murphy & Dickler, 2016). Proliferation-related and luminal-regulated pathways differentiate Luminal-like tumors into Luminal A and Luminal B subtypes; each associated with distinct clinical outcomes. The advantage of

ER+ breast cancer lies in its responsiveness to hormonal therapies, including selective estrogen receptor modulators (SERMs) and aromatase inhibitors (AIs), which block estrogen's effects on cancer cells, slowing or halting tumor progression (Murphy & Dickler, 2016).

Despite its better prognosis compared to other subtypes, ER+ breast cancer poses challenges due to potential recurrence and the need for long-term treatment, necessitating continuous research to improve treatment efficacy (Swain et al., 2023).

2.4.1.1.1 Luminal A Breast Cancer

Luminal A tumors are defined by the presence of estrogen receptor (ER) and/or progesterone receptor (PR) and the absence of HER2. These tumors exhibit low expression of proliferation-related genes (Ki-67 low of <14%), making them slow-growing with a favorable prognosis and late recurrence. Most luminal A breast cancers can be treated with endocrine therapies such as tamoxifen (Inic et al., 2014; Prat et al., 2013; Weigelt et al., 2010)

2.4.1.1.2 Luminal B Breast Cancer

Luminal B tumors are ER-positive but may be PR-negative and/or HER2-positive and Ki-67 high. The high Ki-67 has shown positive correlation with young age women, large tumors and positive lymph nodes. Hence, most Luminal B tumors are of higher grade, have high lymph node involvement, have a worse prognosis and early recurrence (Inic et al., 2014)

2.4.1.2 HER2-Enriched Breast Cancer

The HER2-enriched subtype accounts for 10–15% of breast cancer cases and is characterized by high HER2 expression in the absence of estrogen receptor (ER) and progesterone receptor (PR). This subtype is associated with aggressive tumor growth, rapid proliferation, and a poorer prognosis compared to luminal subtypes. Before the

introduction of HER2-targeted therapies, HER2-enriched breast cancer had one of the worst survival outcomes. However, advancements in targeted treatments, such as trastuzumab and pertuzumab, have significantly improved prognosis and treatment response (Łukasiewicz et al., 2021).

2.4.1.3 Basal-Like/Triple-Negative Breast Cancer

Triple-negative breast cancer (TNBC) is a highly heterogeneous group of breast cancers characterized by the absence of ER, PR, and HER2 expression. TNBC accounts for approximately 20% of all breast cancers and is more common in younger women, particularly those under 40, as well as in African-American women. Notably, around 80% of breast cancers in individuals with BRCA1 germline mutations are classified as TNBC, with 11–16% of all TNBC cases carrying BRCA1 or BRCA2 mutations. Due to its aggressive nature and lack of targeted therapies, TNBC is associated with high recurrence rates and poor prognosis, making it one of the most challenging breast cancer subtypes to treat (Plasilova et al., 2016).

2.4.2 Histopathological Classification

Breast cancer is categorized primarily by its histological appearance based on whether it arises from the ducts (ductal carcinoma) or the lobules (lobular carcinoma). The morphology is also considered based on whether the tumor is in situ (limited to the epithelial component of the breast) or invasive (invaded the stroma) (Makki, 2015).

2.4.2.1 Ductal Carcinoma In-Situ (DCIS)

DCIS is a non-invasive breast cancer characterized by proliferation of atypical epithelial cells within breast milk ducts and do not infiltrate the adjacent breast tissue. Histologically, DCIS is classified by architectural pattern (comedo, solid, cribriform, papillary, micropapillary) and nuclear grade (low, intermediate, high). Majority of the DCIS show a mixture of histological subtypes making it difficult to categorize. High grade

DCIS as well as untreated are precursors for development of invasive carcinomas (Allred, 2010).

2.4.2.2 Lobular Carcinoma in Situ (LCIS)

Histopathological examination reveals a proliferation of monomorphic, dyshesive atypical cells filling and distending the terminal duct-lobular units, with an intact basement membrane. It cannot be identified through visual examination and is found accidentally in samples or biopsies being screened for other reasons. Microscopic examination reveals intact lobular architecture (Christgen et al., 2021)

2.4.2.3 Invasive Ductal Carcinoma (IDC)

IDC are breast cancers with malignant ductal proliferation along with stromal invasion in the presence or absence of DCIS. It is the most predominant type of breast cancer constituting of 80% of all breast cases in women. They are classified based on cell type (apocrine), amount, type and location of secretion (mucinous), architectural features (papillary, tubular and micropapillary) and immunohistochemical profile (neuroendocrine). IDCs are heterogeneous and have a wide range of morphological feature variation. About 25% of invasive breast cancers are identified as 'special types,' characterized by unique growth patterns and cytological features. The specific subtypes include invasive lobular carcinoma (ILC), adenoid cystic carcinoma, apocrine carcinoma, invasive ductal carcinoma (IDC) with osteoclastic giant cells, medullary carcinoma, metaplastic carcinoma, micropapillary carcinoma, mucinous carcinoma, cribriform carcinoma, neuroendocrine carcinoma, and tubular carcinoma (Nascimento & Otoni, 2020; Weigelt et al., 2010). The majority (75%) are designated as IDC not otherwise specified (NOS) as they fail to exhibit sufficient morphological characteristics. The term no special type (NST) can also be used (Makki, 2015)

2.4.2.4 Invasive Lobular Carcinoma (ILC)

ILC is the second major invasive mammary cancer after IDC. It comprises of small, spherical, loose tumor cells growing in single files. It is characterized by unique growth patterns whereby the stromal invasion disrupts the normal architecture, loss of E-cadherin and cytoplasmic accumulation of p120 catenins (Pereslucha et al., 2023)

2.5 Benign Breast Disease

Benign breast disease (BBD) refers to a group of non-cancerous conditions that affect the breast. These conditions include fibrocystic changes, fibroadenomas, and ductal hyperplasia, all of which involve the growth of abnormal breast tissue but are not cancerous (Guray & Sahin, 2006). Benign breast disease is quite common, with many women experiencing benign breast lumps at some point in their lives (Galea, 2019).

While BBD itself does not lead to cancer, certain types, such as atypical hyperplasia, may slightly increase the risk of developing breast cancer in the future (Dyrstad et al., 2015). Despite its non-cancerous nature, BBD can have significant psychological, physical, and social impacts on women. The presence of breast lumps or abnormal tissue can lead to anxiety and uncertainty, as women often worry about the possibility of cancer (Liao et al., 2008). Additionally, the need for ongoing monitoring and, in some cases, biopsies to rule out malignancy can contribute to stress. The emotional toll is often compounded by social stigma or misconceptions surrounding breast abnormalities, leading to further isolation or concern (Srivastava et al., 2020).

2.6 Breast Cancer Risk Factors

Breast cancer is a complex disease with a multifactorial etiology, influenced by both non-modifiable and modifiable risk factors. Non-modifiable risk factors include gender, age, genetic predispositions, race or ethnicity, family history of breast cancer, and proliferative breast disease. Modifiable risk factors encompass menstrual and reproductive factors, radiation exposure, hormonal replacement therapy, alcohol consumption, smoking,

physical activity, and dietary habits, including a high-fat diet. Additionally, environmental exposures such as organochlorine chemicals and electromagnetic fields have been implicated in breast cancer risk (Łukasiewicz et al., 2021).

2.6.1 Non-Modifiable Factors

2.6.1.1 Gender

Among the non-modifiable factors, female gender is a major determinant, primarily due to heightened hormonal stimulation. Unlike men, who have negligible estrogen levels, women possess breast cells that are highly sensitive to estrogen and progesterone, making them more susceptible to hormonal imbalances that can trigger tumor development (Łukasiewicz et al., 2021). Estrogen and androgen levels have been positively correlated with breast cancer risk, and alterations in endogenous hormone levels contribute to increased susceptibility in both premenopausal and postmenopausal women (Folkerd & Dowsett, 2013). Breast cancer remains rare in men, with fewer than 1% of all cases occurring in males, translating to a lifetime risk of approximately 1 in 1,000 compared to 1 in 8 for women (Giordano, 2018).

2.6.1.2 Age

Increasing age is another well-established risk factor, with nearly 80% of breast cancer cases diagnosed in individuals over 50 and over 40% in those above 65. The risk escalates with age, from 1.5% at 40 years, to 3% at 50 years, and over 4% at 70 years. Early-onset breast cancers are often attributed to inherited factors, whereas late-onset cases result from prolonged exposure to carcinogenic influences that accumulate over time (Benz, 2008). A relationship between age and molecular subtypes of breast cancer has also been observed, with the more aggressive triple-negative breast cancer (TNBC) predominantly diagnosed in younger individuals, while the less aggressive luminal A subtypes are more prevalent in older populations (McGuire et al., 2015).

2.6.1.3 Family History

A family history of breast cancer increases risk, with 13–19% of cases linked to first-degree relatives (Łukasiewicz et al., 2021). The risk is higher when breast or ovarian cancer occurs at a younger age. While genetic predisposition plays a key role, not all individuals with a family history carry pathogenic mutations (Collaborative Group on Hormonal Factors in Breast Cancer, 2001).

2.6.1.4 Hormonal and Reproductive

Hormonal and reproductive factors are closely linked to breast cancer development. Several studies have demonstrated that prolonged exposure to estrogen and progesterone increases the risk. The timing of reproductive milestones, such as age at first menstruation, pregnancy, and menopause, plays a crucial role. Early menarche, late menopause, and a delayed first pregnancy are associated with increased risk, whereas early childbirth and prolonged breastfeeding have protective effects. Pregnancy, particularly before the age of 25, is linked to a reduced risk of breast cancer, while nulliparity is associated with a higher likelihood of developing the disease (Arnold et al., 2022).

2.6.1.5 Comorbidities

Hyperinsulinemia and type II diabetes have also been implicated in breast cancer risk due to insulin's role in cell proliferation and survival pathways (Gallagher et al., 2012). Insulin acts as an anti-apoptotic factor and promotes tumor cell invasiveness by stimulating adipokines such as leptin and vascular endothelial growth factor (VEGF). Additionally, insulin interacts with estrogen to induce adipose stromal aromatase and upregulate sex steroid hormone receptors, further increasing cancer risk (Rose & Vona-Davis, 2012). Obesity-related insulin resistance enhances mitogenic signaling via the mitogen-activated protein kinase (MAPK) pathway, thereby facilitating cancer progression. Hypertension, particularly due to the dysregulation of the renin-angiotensin system (RAS), has also been

linked to breast cancer through increased adiposity, inflammation, and angiogenesis (Draznin, 2011; Rasha et al., 2019).

2.6.2 Modifiable Factors

2.6.2.1 Alcohol Consumption

Lifestyle factors such as alcohol consumption, tobacco use, physical inactivity, and obesity significantly contribute to breast cancer risk. Alcohol is one of the primary modifiable carcinogens implicated in breast cancer development (Lachenmeier et al., 2012). The International Agency for Research on Cancer (IARC) classifies both ethanol and acetaldehyde as Group 1 human carcinogens (Tramacere et al., 2010). Alcohol influences carcinogenesis through oxidative stress, altered cellular proliferation, disruption of hormonal pathways, and DNA methylation (Freudenheim, 2020). Acetaldehyde, a major metabolite of alcohol, induces DNA adduct formation and genetic instability (Lachenmeier et al., 2010).

Alcohol consumption, even at moderate levels (15–30 grams per day), leads to elevated serum estrogen levels, which can contribute to breast cancer development (Shield et al., 2016). Additionally, alcohol impacts one-carbon metabolism and folate pathways, both of which are critical for DNA methylation and genomic stability (Shield et al., 2016). Beyond initiation, alcohol consumption may also affect breast cancer progression by promoting angiogenesis, metastasis, and the expansion of cancer stem cells (Wang et al., 2017).

2.6.2.2 Tobacco

Tobacco smoke contains more than 70 known carcinogens, including polycyclic aromatic hydrocarbons (PAHs), nitrosamines, and aromatic amines, which contribute to DNA damage and tumorigenesis (International Agency for Research on Cancer, 2004). While some studies suggest that tobacco smoking has anti-estrogenic effects that may counteract the carcinogenic potential of tobacco-related compounds, others indicate an increased risk

of breast cancer among smokers. The dual effects of tobacco on breast cancer risk appear to be influenced by menopausal status, with premenopausal women experiencing a stronger carcinogenic impact due to higher estrogen levels. Postmenopausal women may have a more balanced risk, with the anti-estrogenic effects of smoking potentially offsetting its carcinogenic properties (Cassidenti et al., 1990). Other factors influencing the impact of tobacco on breast cancer include the age at smoking initiation, duration, and intensity of use (Peñalver-Argüeso et al., 2023).

2.6.2.3 Physical Activity

Physical activity has been consistently associated with a reduced risk of breast cancer. Although the precise mechanisms remain unclear, regular exercise is believed to lower breast cancer risk by reducing exposure to endogenous sex hormones, modifying immune system responses, and lowering insulin-like growth factor-1 (IGF-1) levels (Chen et al., 2019). Obesity, on the other hand, significantly increases the likelihood of developing breast cancer, particularly in postmenopausal women.

2.7 Molecular Mechanisms of Breast Cancer

The pathogenesis of breast cancer is driven by a complex interplay of hereditary genetic defects, acquired somatic mutations, and systemic metabolic dysregulation. Understanding these molecular mechanisms is essential for elucidating disease etiology, prognostic stratification, and the development of targeted therapies.

2.7.1 Hereditary and Somatic Genetic Alterations

Hereditary breast cancer is frequently attributed to high-penetrance, loss-of-function mutations in tumor suppressor genes essential for maintaining genomic integrity. The most prominent of these are mutations in BRCA1 and BRCA2, which are critical for homologous recombination-mediated DNA double-strand break repair (Sokolova et al., 2023). While BRCA1-mutated tumors are strongly associated with the triple-negative phenotype, BRCA2 mutations are also linked to sporadic, hormone receptor-positive

cases. Although these mutations are rare in the general population (<0.1%), they account for approximately 15% of hereditary breast cancer cases (Jara et al., 2017).

Other high-penetrance genes implicated in hereditary syndromes include TP53 (Li-Fraumeni syndrome), PTEN (Cowden syndrome), CDH1 (hereditary diffuse gastric cancer syndrome), and STK11 (Peutz-Jeghers syndrome). Furthermore, moderate-penetrance genes such as PALB2, BRIP1, CHEK2, ATM, and BARD1 confer a more modest but significant increase in lifetime risk. In contrast, common low-penetrance somatic mutations and polymorphisms, particularly in genes governing estrogen metabolism (e.g., CYPs, NATs, GSTs), contribute to a polygenic risk architecture that influences overall susceptibility in the general population (Çelik et al., 2015) (Łukasiewicz et al., 2021).

2.7.2 Systemic and Metabolic Drivers: The Role of Obesity

Beyond direct genetic alterations, systemic metabolic states, particularly obesity, establish a tumor-promoting microenvironment through several integrated pathways. Obesity-induced metabolic dysregulation leads to a state of chronic low-grade inflammation, characterized by elevated levels of pro-inflammatory cytokines such as TNF-α and IL-6. Furthermore, increased aromatase activity within adipose tissue enhances the local conversion of androgens to estrogens, providing a potent mitogenic signal for hormone receptor-positive cells. Concurrent elevations in insulin and insulin-like growth factor-1 (IGF-1) levels, associated with insulin resistance, further stimulate oncogenic PI3K/AKT and MAPK signaling pathways (Rasha et al., 2019).

2.7.3 Estrogen Metabolism and Genetic Susceptibility

Estrogen and its metabolites are central to the initiation and progression of hormone-sensitive breast cancer (Yim et al., 2001). Estrogen metabolism is a two-phase enzymatic process. In Phase I, cytochrome P-450 (CYP) enzymes catalyze the hydroxylation of estrogen to form catechol estrogens, primarily 2-hydroxyestradiol (2-OHE2) and 4-

hydroxyestradiol (4-OHE2). In Phase II, conjugation and detoxification occur via enzymes such as catechol-O-methyltransferase (COMT), which methylates these reactive catechols into less genotoxic methoxyestradiols (Wen et al., 2017a). A failure in this detoxification pathway results in the accumulation of reactive quinones, elevated reactive oxygen species (ROS), and the formation of mutagenic DNA adducts, thereby initiating carcinogenesis (Cavalieri et al., 2012).

Genetic polymorphisms in the genes encoding these metabolizing enzymes are critical determinants of interindividual variation in estrogen metabolism and significantly modulate breast cancer susceptibility by altering the balance between estrogen activation and detoxification (Al-Shami et al., 2023). Key polymorphic genes include CYP1A1, CYP1B1, CYP3A5, and COMT, whose functional variants are summarized in Table 2.1.

The CYP1A1 gene encodes a Phase I enzyme involved in estrogen hydroxylation, predominantly yielding the weaker estrogenic metabolite 2-OHE2. Polymorphisms such as rs4646903 (T6235C) in the 3' UTR and the exonic rs1048943 (A4889G; Ile462Val) have been associated with increased enzyme activity and upregulated gene expression, potentially enhancing the bioactivation of procarcinogens and influencing breast cancer risk (N. Shimada et al., 2009).

Conversely, CYP1B1 catalyzes the formation of the highly reactive and carcinogenic 4-OHE2. Several functional polymorphisms, including rs10012 (G142C; Arg48Gly), rs1056827 (C355A; Ala119Ser), and the extensively studied rs1056836 (G4326C; Leu432Val), increase the enzyme's catalytic efficiency and substrate binding affinity. This leads to greater production of genotoxic 4-OHE2 and enhanced activation of environmental carcinogens such as polycyclic aromatic hydrocarbons (PAHs), thereby elevating breast cancer risk (Jiao et al., 2010; Wen et al., 2017b).

The CYP3A5 gene contributes to estrogen and tamoxifen metabolism. The common polymorphism rs776746 (G6986A) is a splicing defect that defines the CYP3A5*3 allele, resulting in a truncated, non-functional protein. Individuals homozygous for this allele

lack CYP3A5 activity, which may confer a protective effect. In contrast, carriers of the functional CYP3A5*1 allele exhibit higher enzyme expression and are potentially at increased risk (Lolodi et al., 2017).

The COMT gene encodes the critical Phase II detoxification enzyme. The seminal rs4680 (A472G; Val158Met) polymorphism results in a valine-to-methionine substitution that reduces enzyme thermostability and catalytic activity by 3- to 4-fold. Individuals homozygous for the low-activity Met/Met genotype exhibit impaired inactivation of catechol estrogens, leading to prolonged exposure to reactive metabolites, increased oxidative stress, and a well-documented elevated risk of breast cancer, with effects often more pronounced in postmenopausal women (Qin et al., 2012).

Table 2.1: Description of rs4646903, rs1048943, rs1056836, rs776746 and rs4689 SNPs

rs ID	Gene	Chromosomal Location	SNP description	Base change/position	Amino acid substitution
rs4646903	CYP1A1	15q22 - q24.1	3' UTR	T6235C	None
rs1048943	CYP1A1	15q22 - q24.1	Exon 7	A4889G	Ile462Val
rs776746	CYP3A5	7q22.1	Intron	G6986A	None
rs4680	COMT	22q11.21	Exon 4	A472G	Val158Met
rs10012	CYP1B1	2p22.2	Exon 2	G142C	Arg48Gly
rs1056827	CYP1B1	2p22.2	Exon 2	C355A	Ala119Ser
rs1056836	CYP1B1	2p22.2	Exon 3	G4326C	Leu432Val
rs1800440	CYP1B1	2p22.2	Exon 3	A4390G	Asn453Ser

Description of rs4646903, rs1048943, rs776746, rs4689, rs10012, rs1056827, rs1056836 and rs1800440 SNPs including their genes and chromosomal locations, functional regions, base change and amino acid substitutions.

2.8 Breast Cancer Screening and Diagnosis

2.8.1 Breast Cancer Screening

The World Health Organization (WHO) recommends that women at average risk for breast cancer between the ages of 50 and 69 undergo biennial mammography screening

in organized, population-based programs. However, in resource-limited settings, where late-stage diagnosis is prevalent and mammography is not cost-effective or widely accessible, early detection through clinical breast examination (CBE) and public awareness campaigns is prioritized (World Health Organization, 2014).

2.8.2 Diagnostic Imaging

Early-stage diagnosis significantly improves breast cancer prognosis. Imaging techniques such as mammography, magnetic resonance imaging (MRI), positron-emission tomography (PET), computed tomography (CT), and single-photon emission computed tomography (SPECT) are commonly used for diagnosis, screening, and monitoring therapeutic responses (Jafari et al., 2018).

Mammography remains the most widely used screening tool due to its high sensitivity, specificity, and cost-effectiveness. However, it carries risks of false positives, radiation exposure, and patient anxiety (Wellings et al., 2016). Ultrasound is often used as an adjunct to mammography, particularly in young, pregnant, or breastfeeding women, due to its high sensitivity and lack of ionizing radiation (Alcantara et al., 2014). MRI is commonly employed for high-risk patients, preoperative staging, and monitoring treatment response, although its lower specificity can result in higher false-positive rates (Sardanelli et al., 2010). Mammography combined with MRI is currently the most sensitive imaging strategy for high-risk populations (Wellings et al., 2016).

2.8.3 Molecular Testing and Biochemical Markers

While imaging techniques are useful for diagnosis and monitoring, molecular testing provides additional insights into tumor biology, prognosis, and treatment selection. Molecular markers, including DNA, mRNA, and microRNAs, have revolutionized breast cancer diagnosis and therapy by enhancing understanding of tumor heterogeneity (Duffy et al., 2017). Early-stage ER+ breast cancers are assessed using molecular tests that evaluate tumor size, nodal involvement, histological grade, proliferation markers (e.g.,

Ki-67), and hormone receptor status (ER, PR, and HER2). Immunohistochemistry (IHC) and fluorescence *in situ* hybridization (FISH) are the standard methods for determining these molecular features, which guide treatment planning and prognosis (Goldhirsch et al., 2013).

2.9 Breast Cancer Treatment

Cancer treatment is a component of cancer control aimed at curing, prolonging life and improving quality of life. Although treatment is most effective during early detection, it is also dependent on cancer type and quality of care. Cancer treatment methods include; surgery, chemotherapy (endocrine or targeted). Radiotherapy, cryotherapy, loop electrosurgical excision procedure (LEEP), brachytherapy, therapeutic nuclear medicine, end of life care and survivorship. Multi-disciplinary management is recommended over independent management for improved patient outcomes (Ministry of Health, 2017)

2.9.1 Surgery

Breast cancer surgery includes breast-conserving surgery (BCS) and mastectomy. BCS, also known as partial or segmental mastectomy, involves removing the cancerous tissue while preserving intact breast tissue, often combined with oncoplastic reconstruction techniques. Mastectomy, which entails complete removal of the breast, is typically recommended for large tumors, multicentric disease, or patients at high genetic risk. Immediate or delayed breast reconstruction is an option for women undergoing mastectomy to improve cosmetic and psychological outcomes (Keelan et al., 2021).

2.9.2 Chemotherapy

Chemotherapy is a systemic treatment administered as neoadjuvant therapy (before surgery) or adjuvant therapy (after surgery). Neoadjuvant chemotherapy is particularly beneficial for locally advanced breast cancer, inflammatory breast cancer, and triple-negative or HER2-positive subtypes, as it reduces tumor size and improves surgical outcomes. Common chemotherapeutic agents include carboplatin, cyclophosphamide, 5-

fluorouracil, taxanes (paclitaxel, docetaxel), and anthracyclines (doxorubicin, epirubicin). Although effective, chemotherapy is associated with side effects such as nausea, fatigue, hair loss, immunosuppression, and bone marrow suppression (Rouzier et al., 2005).

2.9.2.1 Endocrine Therapy

Endocrine therapy is primarily used in luminal breast cancer to block estrogen stimulation of cancer cells. It is administered as neoadjuvant, adjuvant, or metastatic treatment. Common endocrine therapy drugs include: selective estrogen receptor modulators (SERMs) like tamoxifen and toremifene, selective estrogen receptor degraders (SERDs) like fulvestrant, and aromatase inhibitors (AIs) such as letrozole, anastrozole, and exemestane. These therapies significantly reduce recurrence risk in ER+ breast cancer (Jones & Buzdar, 2004) survival (Jones & Buzdar, 2004).

2.9.2.2 Targeted Therapy

Targeted therapies, particularly for HER2-positive breast cancer, have dramatically improved patient outcomes. HER2 inhibitors such as trastuzumab, pertuzumab, lapatinib, and neratinib specifically block HER2 signaling, reducing tumor progression and recurrence risk (Ishii et al., 2019).

2.9.3 Radiotherapy

Radiation therapy is commonly used after breast-conserving surgery and/or following chemotherapy to eliminate residual cancer cells and lower recurrence risk. Despite side effects such as skin irritation, fibrosis, and fatigue, radiation therapy has significantly improved survival rates in breast cancer patients (Łukasiewicz et al., 2021).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study Sites

The study participants were recruited from Aga Khan University Hospital, Nairobi (1°15'55.0"S, 36°48'48.0"E) and Africa Inland Church Kijabe Hospital (0° 56' 46" S, 36° 35' 42" E). AKUHN is a private, non-profit institution located in Parklands, Nairobi, offering both tertiary and secondary healthcare services. It attracts patients from across Kenya and diverse socio-economic backgrounds. The hospital diagnoses and treats a significant number of breast cancer cases and has a well-equipped oncology unit.

AIC Kijabe Hospital, situated in Kijabe, Kiambu County, primarily serves patients from all regions of the country. Like AKUHN, it has a diverse patient population and provides a wide range of cancer related services such as screening, chemotherapy, surgery and palliative care for different cancer types. Both hospitals have a specialized team managing breast cancer cases. They maintain extensive patient records and adhere to strict ethical guidelines, ensuring high-quality data for research.

All genomic work was conducted at the Biotechnology and Molecular Laboratory at the Pan African University of Science, Technology, and Innovation (PAUSTI), Thika.

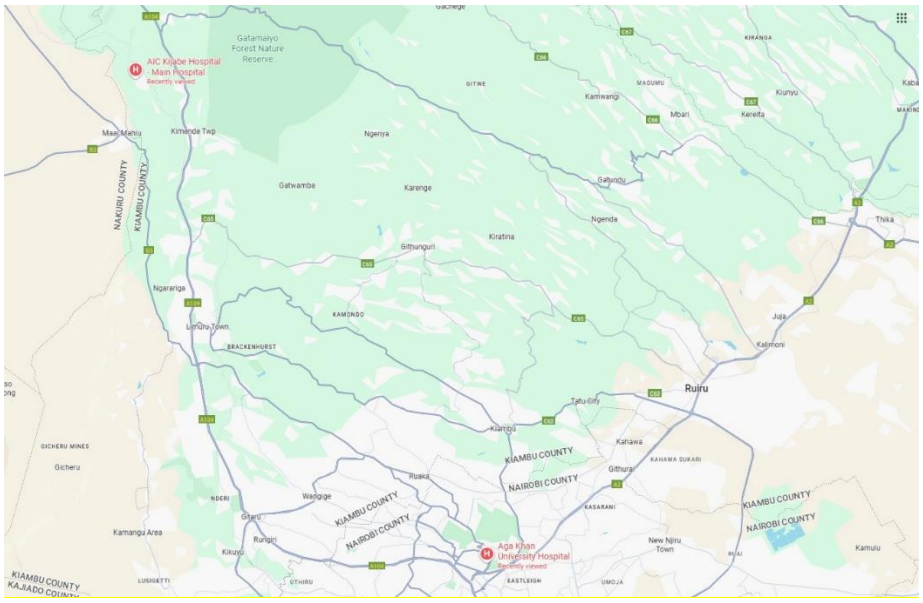


Figure 3.1: Map Showing Sampling Sites

Kenyan map showing sampling sites: Aga Khan University Hospital, Nairobi

(AKUH, N) and Africa Inland Church Kijabe Hospital (KAIC) (National Geographic Society, 2019).

3.2 Study Design

This study utilized a hospital-based, unmatched case-control design. This approach is particularly effective for retrospectively investigating associations between a defined outcome and prior exposures (Andrade, 2022). Three participant groups were established to facilitate a stratified analysis of risk: cases (women with histologically confirmed ER^+ breast cancer), women with histologically confirmed benign breast disease, BBD, and healthy controls (women with no personal history of breast cancer or any malignancy). A primary comparison of cases versus healthy controls aimed to identify factors associated with malignancy, while a secondary comparison of cases versus BBD controls enabled exploration of factors potentially distinguishing malignant progression from benign proliferative states. The inclusion of a BBD group thus allowed investigation across a

pathological continuum, refining the assessment of genetic and environmental influences on breast disease progression.

3.3 Ethical Consideration

The study protocol was reviewed and approved by the Aga Khan University Institutional Research Ethical Committee (AKU-REC) Ref: 2020/IERC-26 (v4) **(Appendix 1)**. In addition, a research permit was obtained from the National Commission for Science, Technology and Innovation (NACOSTI) (License No: NACOSTI/P/21/13890) **(Appendix 2)**. Data and biospecimens for this secondary analysis were derived from the parent study, "A Genomic Approach for Understanding Breast Cancer Progression in Kenya". All participants in the original study provided written informed consent for their samples and clinical data to be stored and used in future genetic and molecular studies **(Appendix 3).** For the current project, the institutional biobank granted access to these de-identified resources. No patient identifiers were accessible to the research team at any stage, ensuring confidentiality and anonymity throughout the current analysis.

3.4 Study Participants

The study participants comprised of the cases, individuals with benign breast diseases (BBD), and controls. Cases were women diagnosed with estrogen receptor-positive (ER+) breast cancer. For this study, 'ER^+ breast cancer' was operationally defined as a tumor showing positive immunohistochemical staining for the estrogen receptor (ER), irrespective of its progesterone receptor (PR) status. BBD participants were diagnosed with benign breast conditions, while controls were healthy volunteers working at the hospital.

3.5 Sampling Method

This study employed a purposive (non-probability) sampling strategy to select archived samples and associated clinical data that met predefined eligibility criteria. Purposive sampling was appropriate because the research specifically required samples from

individuals with defined pathological features: ER^+ breast cancer cases, benign breast disease (BBD) controls, and healthy controls.

Given the specific requirements for case confirmation (histologically verified ER^+ status) and the relative scarcity of such well-characterized samples, purposive sampling enabled the efficient selection of a focused and appropriate study cohort from the available biobank inventory. While this approach ensured the sample's relevance to the research objectives, a key limitation is the potential for selection bias and reduced generalizability, as the selected samples reflect the characteristics of the original biobanked population rather than the broader, general population.

3.6 Eligibility Criteria

The study enrolled participants from three distinct cohorts: Estrogen Receptor-Positive (ER+) Breast Cancer Cases, Benign Breast Disease (BBD) Controls, and Healthy Controls.

3.6.1 Inclusion Criteria

The study included three participant groups. For cases, eligible participants were women aged 18 years or older with a histologically confirmed diagnosis of estrogen receptor-positive breast cancer (ER/PR positive) at either AKUHN or AIC Kijabe Hospital, who provided blood samples and gave informed consent. Healthy control volunteers were required to be women aged 18 years or older with no prior diagnosis of breast cancer or any other malignancy, who provided a blood sample, gave informed consent, and were employed at AKUHN with confirmed health status at recruitment. The benign breast disease group comprised women aged 18 years or older with a histologically confirmed benign breast disease diagnosis at either hospital who provided blood samples and signed informed consent.

3.6.2 Exclusion Criteria

Patients were excluded from the cases if they had other breast cancer subtypes (such as triple-negative or HER2-positive tumors) or failed to provide a blood sample. Participants in the BBD and control groups were excluded if they did not provide a blood samples and/or did not sign the informed consent.

3.7 Sample Size Determination

The a priori sample size calculation, based on parameters from published literature (Kelsey et al., 1996), indicated a requirement of 134 participants per group to detect an odds ratio of 1.8 with 80% power. However, final recruitment yielded 64 ER^+ breast cancer cases, 79 benign breast disease (BBD) controls, and 19 healthy controls **(Appendix 5).** This shortfall was primarily due to post-COVID-19 restrictions on hospital-based recruitment and a lower-than-anticipated eligible patient flow during the study period. Consequently, the achieved statistical power for detecting the initially hypothesized effect size (OR=1.8) was reduced. To ensure rigorous analysis within this constraint, analytical methods robust to smaller sample sizes were prioritized, including exact statistical tests and careful control for confounding.

3.8 Data Collection

Existing data were abstracted from two primary sources: (1) the completed questionnaire forms from the original study database, and (2) the corresponding clinical and histopathology reports filed in the hospital records. A standardized abstraction form was used to extract the relevant variables (**Appendix 4**). No direct interaction with participants or new data collection occurred for this project. Data collected comprised of; socio-demographics, anthropometrics, reproductive and hormonal history, lifestyle factors, medical and family history, and environmental exposures. The abstraction form was used to collect verified clinical information from medical and histopathology reports. This process captured detailed clinicopathological characteristics (histology, grade, stage, and

receptor status) for breast cancer cases, and the specific benign diagnosis for BBD participants **(Appendix 6).** All collected data were de-identified using a unique study code and systematically entered into a secure electronic database for subsequent analysis.

Archived blood samples were originally collected from all participants following informed consent under the prior study protocol. For cases and BBD participants, samples were drawn during clinical visits; for healthy controls, samples were collected at a hospital occupational health unit. A standardized protocol was used: 5.0 mL of venous blood was collected into EDTA tubes, centrifuged at 744 × g for 15 minutes, with plasma and buffy coat fractions aliquoted and stored at -20°C in the institutional biobank. For this secondary analysis, de-identified buffy coat aliquots were retrieved from the biobank following ethical approval for reuse. No new samples were collected, and all samples remained anonymized throughout subsequent DNA extraction and genotyping.

3.9 DNA Extraction

Genomic DNA was isolated from buffy coat samples using the ISOLATE II Genomic DNA Kit (Bioline, Meridian Life Science Inc., USA) following the manufacturer's instructions. Briefly, 200 µL of buffy coat sample was mixed with 200 µL of lysis buffer G3 and 25 µL of Proteinase K, vortexed vigorously, and incubated at 70 °C for 10 minutes to facilitate cell lysis and protein digestion. Following incubation, 210 µL of 100% ice-cold ethanol was added, and the mixture was vortexed thoroughly to precipitate DNA. The lysate was then transferred into a 2 mL centrifuge tube containing a DNA binding column, allowed to stand for 2 minutes, and centrifuged to facilitate DNA binding.

Subsequently, the column was washed sequentially with buffers GW1 and GW2, each followed by centrifugation at 10,000 × g using a Thermo Scientific MicroCL 21 microcentrifuge with a fixed-angle rotor for 1 minute to remove contaminants. In the final step, 50 µL of Elution Buffer was added to the column, incubated at room temperature for 10 minutes, and centrifuged at 10,000 × g using a Thermo Scientific MicroCL 21 microcentrifuge with a fixed-angle rotor for 1 minute to elute the purified DNA. The

extracted DNA was dissolved in 50 µL of sterile distilled water. DNA integrity was assessed by 1.5 % agarose-1X Tris acetate-EDTA (TAE) gel stained with GelRed® nucleic acid gel stain (Biotium, Inc., Fremont, CA, USA). 8µl of the samples were resolved alongside 1 kb DNA marker (HyperLadder™ 100 bp, Bioline, Meridian Biosciences, BIO-33025). The gel was run at 70 volts 80mA and 50 Watts for one hour and the bands visualized using a UV transilluminator. The DNA samples were subsequently stored at -20 °C for downstream applications.

3.10 Polymerase Chain Reaction

Polymerase chain reaction (PCR) was used to amplify gene regions containing the polymorphic sites using specific primers designed with Primer3Plus (Untergasser et al., 2007). The primer sequences are presented in Table 3.1.

Amplifications were performed using conventional PCR in 30 µL reactions, with each reaction containing 50 ng/µL genomic DNA, 0.3 µM of each primer, 1.5 U MyTaq DNA polymerase (BIO-21105, Bioline, Meridian Life Science Inc., USA), 6 µL 5× MyTaq reaction buffer, and nuclease-free water. Thermocycling conditions consisted of an initial denaturation at 95 °C for 3 minutes, followed by 30 cycles of denaturation at 95 °C for 30 seconds, annealing at 55 °C for 45 seconds, and extension at 72 °C for 45 seconds, with a final extension step at 72 °C for 3 minutes.

PCR reactions were conducted using a ProFlex™ 3 × 32-well PCR System (Applied Biosystems, Thermo Fisher Scientific, Waltham, UK).

Polymerase chain reaction (PCR) was used to amplify regions harboring the SNPs rs4646903 (340 bp), rs1048943 (204 bp), rs1056836 (245 bp), rs776746 (251 bp), and rs4680 (304 bp) (Table 3.1) as well as rs10012 (202 bp), rs1056827 (202 bp), rs1056836 (567 bp) and rs1800440 (567 bp) (Table 3.2). The resulting amplicons were resolved on a 2% agarose-Tris borate-EDTA (1X TBE) gel stained with GelRed® nucleic acid gel stain (Biotium, Inc., Fremont, CA, USA) alongside 100 bp DNA marker (HyperLadder™ 100

bp, Bioline, Meridian Biosciences, BIO-33056). While TAE buffer was used for initial DNA quality checks, TBE buffer was specifically selected for gels analyzing restriction digestion products and PCR amplicons due to its ability to produce sharper, better-resolved bands—a requirement for reliable genotyping interpretation. The gel was run at 100 volts 80mA and 50 Watts for 45 mins. The bands were visualized using a UV transilluminator to confirm the presence and size of the PCR products.

Table 3.1: Primers and PCR Products of the Target Regions

rs ID	Gene	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Product Size (bp)
rs4646903	CYP1A1	CAGTGAAGAGGTGTAGCCGCT	TAGGAGTCTTGTCTCATGCCT	340
rs1048943	CYP1A1	CTGTCTCCCTCTGGTTACAGGAAGC	TTCCACCCGTTGCAGCAGGATAGCC	204
rs1056836	CYP1B1	TGTCCTGGCCTTCCTTTATG	TCATCACTCTGCTGGTCAGG	245
rs776746	CYP3A5	CCATACAGGCAACATGACTT	GGTTAGAAATGACAGTAGAGCA	251
rs4680	COMT	ACAGGCAAGATCGTGGAC	CAGTGAACGTGGTGTGAAC	304
rs10012	CYP1B1	GAAACACACGGCACTCAT	ACGCTCCTGCTACTCCTGT	567
rs1056827	CYP1B1	GAAACACACGGCACTCAT	ACGCTCCTGCTACTCCTGT	567
rs1056836	CYP1B1	ATCATCACTCTGCTGGTCA	TGCCTGTCACTATTCCTCA	202
rs1800440	CYP1B1	ATCATCACTCTGCTGGTCA	TGCCTGTCACTATTCCTCA	202

The table lists the primer sequences used for PCR amplification of rs4646903, rs1048943, rs776746 and rs4680, rs10012, rs1056827, rs1056836 and rs1800440 along with the corresponding product sizes (bp).

3.11 Genotyping Using Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP)

Amplified DNA fragments containing the polymorphic regions of interest namely: rs4646903, rs1048943, rs1056836, rs776746 and rs4680 were subjected to restriction enzyme digestion using the PCR-RFLP technique (Tarach, 2021). The following

restriction enzymes, obtained from New England Biolabs (UK) were used: *Msp*I (R0106) for rs4646903, *Bsr*D1/*Bse*MI (R0574) for rs1048943, *Acu*I (R0641S) for rs1056836, *Ssp*I (R3132) for rs776746 and *Nla*III (R0125S) for rs4680. Each digestion was carried out under enzyme-specific incubation conditions. The resulting restriction fragments were resolved on 3% agarose-Tris Borate EDTA (1X TBE) gels stained with GelRed. The samples were resolved resolved alongside 50 bp DNA marker (HyperLadder™ 50 bp, Bioline, Meridian Biosciences, BIO-33054). The resulting bands were visualized under a UV transilluminator. The expected fragment sizes corresponding to the wild type, heterozygous, and variant genotypes are shown in Table 3.2. A subset of samples were verified by Sanger sequencing to eliminate any ambiguity associated with gel-based fragment size discrimination.

Table 3.2: Restriction Enzymes and Conditions Used for PCR-RFLP Genotyping

SNP (rs ID)	Restriction Enzyme (NEB Catalog #)	Incubation Conditions	Fragment Sizes (bp)
rs4646903	*Msp*I (R0106)	37 °C (1hr), 80 °C (20 min)	wt: 340 het: 340,200,140 var: 200,140
rs1048943	*Bsr*D1/*Bse*MI (R0574)	65 °C (1hr), 80 °C (20 min)	wt: 150,54 het: 204,150,54 var: 204
rs1056836	*Acu*I (R0641S)	37 °C (1hr), 65 °C (20 min)	wt: 245 het: 245,176,69 var: 176,69
rs776746	*Ssp*I (R3132)	37 °C (1hr), 65 °C (20 min)	wt: 137,114 het: 251,137,114 var: 251
rs4680	*Nla*III (R0125S)	37 °C (1hr), 65 °C (20 min)	wt: 114,93,54,43 het: 304,114,93,54,43 var: 304

Summary of restriction enzymes, incubation conditions and expected fragment sizes for genotyping five SNPs using PCR-RFLP.**"wt"** = Wild type, **"het"** = Heterozygous**, "var"** = Variant

3.12 Genotyping of CYP1B1 Polymorphisms by Sanger Sequencing

Amplified PCR products for the CYP1B1 gene regions harboring the polymorphisms rs10012, rs1056827, rs1056836, and rs1800440 were purified using the QIAquick PCR Purification Kit (QIAGEN, Hilden, Germany), according to the manufacturer's protocol, to eliminate unincorporated primers, nucleotides, and other contaminants. During purification, DNA binds to a silica membrane under high-salt conditions, while impurities pass through the column. The membrane-bound DNA was then washed thoroughly, and purified DNA was eluted using Tris buffer. The resulting clean amplicons were subjected to bidirectional Sanger sequencing by Macrogen Europe BV.

3.13 Bioinformatics Analysis of Sequences

Polymorphism analysis was performed using *Geneious Prime* 2025.1.0 software (Biomatters Ltd, Auckland, New Zealand). Raw Sanger sequencing data, obtained in .ab1 format, were imported into the software for quality assessment and processing. Chromatograms were visually inspected, and low-quality base calls at the 3′ and 5′ ends were trimmed to ensure sequence accuracy. High-quality consensus sequences were generated for each sample.

The reference sequence for the CYP1B1 gene (accession number: NG_008386.2) was retrieved from the NCBI database and imported into *Geneious Prime* for alignment. Multiple sequence alignment (MSA) was conducted in MEGA using MUSCLE program (Edgar, 2004). Single nucleotide polymorphisms (SNPs) were detected using the "Find Variations/SNPs" tool by comparing each consensus sequence to the reference. All identified SNPs were validated by cross-checking with chromatogram peaks, and filtering was applied based on base quality and allele frequency to retain high-confidence variants. The final SNP dataset was compiled in Microsoft Excel for downstream statistical analysis.

3.14 Statistical Analysis

Statistical analyses were performed using R software (Version R4.3.3) (The R Core Team, 2023) and the tidyverse package (2.0.0) (Wickham et al., 2019) for data preparation and visualization. Descriptive analysis of the independent variables among the study participants was carried out. Tests of independence were employed using Student's t or Wilcoxon rank-sum tests for categorical data and Fisher's exact, or Pearson's Chi-squared tests for continuous data.

3.15 Association Analysis

3.15.1 Association with Epidemiological and Clinical Factors

The analytical strategy employed both epidemiological and genetic association approaches to assess factors related to estrogen receptor-positive (ER+) breast cancer. Bivariate logistic regression was first conducted to explore associations between various independent variables and ER+ breast cancer case status. The variables examined included socio-demographic factors such as age, and body mass index; reproductive factors including age at menarche, menopausal status, parity, and history of contraceptive use; lifestyle factors such as alcohol consumption and tobacco use; medical history factors such as personal history of other cancers, family history of breast cancer, and comorbidities including hypertension and diabetes; and environmental exposures such as self-reported history of significant radiation. Variables with a bivariate p-value less than 0.25 were considered potential confounders and were included in subsequent multivariate logistic regression models to estimate adjusted odds ratios with 95% confidence intervals, thereby assessing the independent association of each risk factor while adjusting for potential confounding.

3.15.2 Association of Genetic Variants with Clinicopathological Characteristics

A secondary analytical objective was to evaluate the association between the genotyped polymorphisms and key clinicopathological characteristics within the ER+ breast cancer

case group. This analysis focused on tumor characteristics abstracted from histopathology reports, which were dichotomized for analysis. These included molecular subtype, categorized as Luminal A versus Luminal B based on immunohistochemistry for ER, PR, HER2, and Ki-67; cancer stage, dichotomized as Early stage (Stage I & II) versus Late stage (Stage III & IV); tumor grade, classified as Low grade (Grade 1 & 2) versus High grade (Grade 3); histopathological type, defined as Invasive Ductal Carcinoma versus other subtypes; and the presence or absence of lymphovascular invasion.

3.15.3 Genotype Association Analysis

For both epidemiological and tumor characteristic analyses involving genetic data, genotype association tests were performed. All genetic association analyses were conducted in R software (v4.3.3) using the SNPassoc package (v2.0-0). The association between each polymorphism and a binary outcome was evaluated under five genetic models—allelic, codominant, dominant, recessive, and log-additive—using unconditional logistic regression to calculate odds ratios and 95% confidence intervals. A two-sided p-value less than 0.05 was considered statistically significant.

3.16 In-Silico Analysis of CYP1B1 Haplotypes

3.16.1 Linkage Disequilibrium and Haplotype Construction

Haplotypes All statistical analyses were performed using R (v4.3.3). Pairwise linkage disequilibrium (LD) among the four CYP1B1 SNPs (rs10012, rs1056827, rs1056836, and rs1800440) was assessed using the LD function from the genetics package. The strength of association between each SNP pair was quantified using D' (D-prime) and r^2 (correlation coefficient). Statistical significance was determined using a chi-square (χ^2) test, with p-values < 0.05 considered statistically significant. These LD metrics were used to confirm the haplotype block structure and validate the co-inheritance patterns of the alleles prior to haplotype reconstruction.

Subsequently, haplotype probabilities were estimated using the haplo.em function from the haplo.stats package (Sinnwell & Schaid, 2024), which infers the most likely pair of haplotypes for each individual based on the four CYP1B1 SNPs. For each participant, the combination of alleles at these four loci was determined in the order presented (rs10012, rs1056827, rs1056836, rs1800440), and haplotypes were designated based on the specific allele present at each position. Only haplotypes with an estimated frequency greater than 1% in at least one study group (cases, controls, or benign samples) were retained for further analysis. The frequency distribution of each haplotype was calculated separately for each group and expressed as counts (n) and percentages (%).

3.16.2 Retrieval of the Wild-Type Cyp1b1 Sequence from UniProtKB and Generation of Haplotype Variants

The wild-type amino acid sequence of human cytochrome P450 1B1 (CYP1B1) was retrieved from UniProt Knowledgebase (UniProt ID: Q16678, https://www.uniprot.org/uniprotkb/Q16678/entry). This 543-amino acid sequence, representing the canonical isoform of CYP1B1, served as the template for all subsequent structural and functional analyses, including homology modeling, molecular dynamics simulations, and in silico mutagenesis. The retrieved sequence was modified using point mutations corresponding to the haplotypes. The specific amino acid substitutions were as follows: R48G, A119S, L432V, and N453S.

3.16.3 Tertiary Structure Analysis

The three-dimensional structures of wild-type CYP1B1 and its haplotypes were generated using AlphaFold Server (https://alphafoldserver.com/welcome), powered by the AlphaFold 3 model. For each haplotype, the corresponding amino acid sequence was submitted, and the top-ranked model based on the composite ranking_score was selected for visualization in PyMOL. Four representations were prepared for comparative analysis including: ribbon diagrams to illustrate the overall protein fold; surface models to visualize solvent accessibility and active site topology; colour-coded ribbon diagrams

highlighting the side chains of the four polymorphic residues (R48G, A119S, L432V, N453S); and Ramachandran plots generated using PROCHECK to assess the stereochemical quality of each model. Prediction confidence was evaluated using the pLDDT scores from AlphaFold, with regions scoring above 90 considered highly reliable.

CHAPTER FOUR

RESULTS

4.1 Genomic DNA Yield and Quality

Genomic DNA extracted from buffy coat samples of 64 ER+ breast cancer cases, 79 BBDs and 19 healthy controls showed bands of greater than 10 Kb. The yield and concentration was analysed using a NanoDrop 2000 spectrophotometer (Biospec-mini, Shimadzu Corporation, Tokyo). DNA concentration was determined spectrophotometrically, with all samples yielding sufficient quantities for downstream analysis (concentrations > 350 ng/µL). Purity was assessed by the absorbance ratio at 260 nm and 280 nm (A260/A280), which fell within the optimal range of 1.8 to 2.0 for all samples, indicating minimal protein or phenolic contamination. DNA integrity was evaluated by electrophoresis on a 1.5 % agarose gel (1X TAE buffer). The majority of samples displayed a single, high-molecular-weight band exceeding 10 kilobases, confirming high integrity. A subset of samples exhibited slight smearing below the primary band, suggesting partial shearing or degradation; however, the DNA quality in these samples was deemed acceptable for PCR amplification and subsequent genotyping analyses. (Figure 4.1)

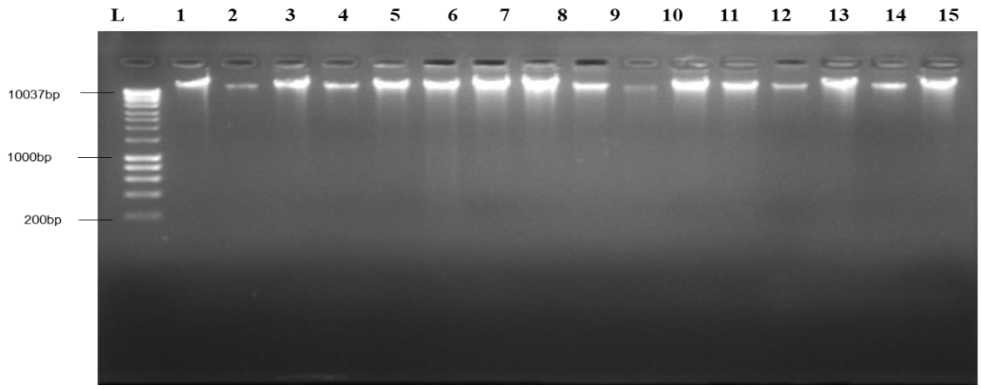


Figure 4.1: Gel Photo of Extracted Genomic DNA

Agarose gel electrophoresis of genomic DNA extracted from buffy coat samples in ER+ breast cancer cases. L: 1Kb DNA marker, 1: AKU057, 2: AKU060, 3: AKU073, 4:

AKU076, 5: AKU087, 6: AKU088, 7: AKU094, 8: AKU097, 9: AKU098, 10: AKU099, 11: AKU100, 12: AKU103, 13: KAIC005, 14: KAIC006, 15: KAIC007.

4.2 PCR Amplicons of Target Regions

Amplicons of the target gene regions were resolved on a 2% TBE agarose gel. The resulting amplicons for rs4646903 (340 bp), rs1048943 (204 bp), rs1056836 (245 bp), rs776746 (251 bp), rs4680 (304 bp), rs10012 and rs1056827 (202 bp), rs1056836 and rs1800440 (567 bp) are as shown in the agarose gel photos (Figures 4.2 – 4.8).

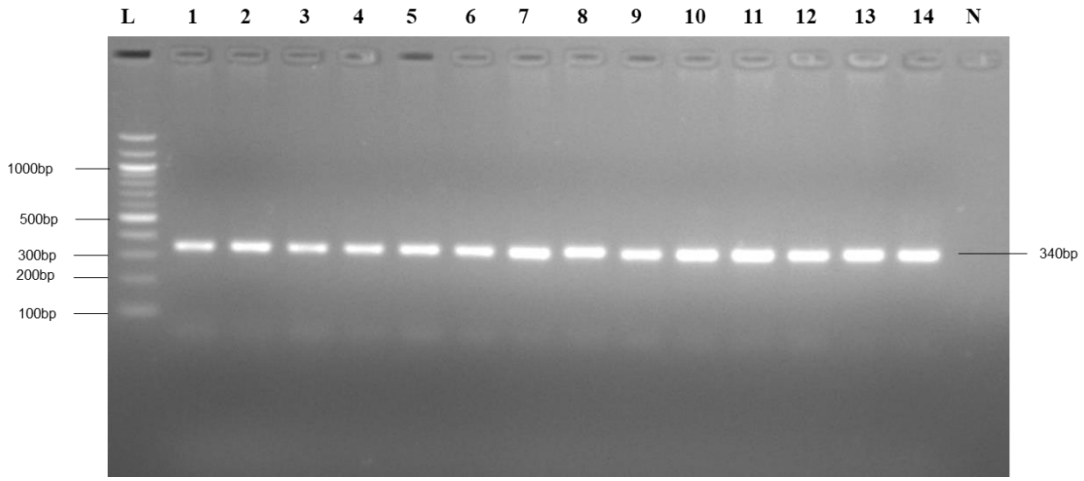


Figure 4.2: Gel Photo of rs4646903 Amplicons

Agarose gel electrophoresis of rs4646903 in ER+ breast cancer cases. L: 100 bp DNA marker, N: Negative control. 1: AKU057, 2: AKU060, 3: AKU073, 4: AKU076, 5: AKU087, 6: AKU088, 7: AKU094, 8: AKU097, 9: AKU098, 10: AKU099, 11: AKU100, 12: AKU103, 13: KAIC005, 14: KAIC006.

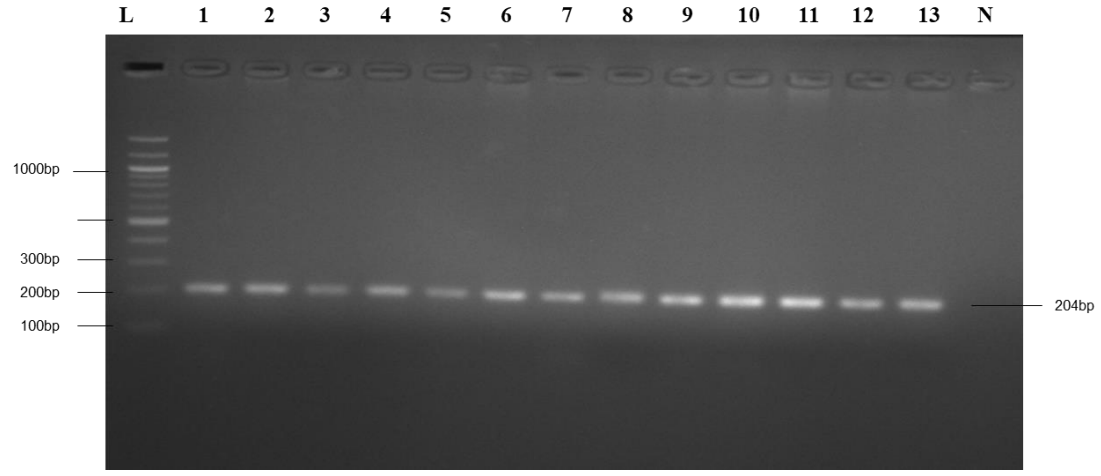


Figure 4.3: Gel Photo of rs1048943 Amplicons

Agarose gel electrophoresis of rs1048943 in ER+ breast cancer cases. L: 100 bp DNA marker, N: negative control. 1: AKU057, 2: AKU060, 3: AKU073, 4: AKU076, 5: AKU087, 6: AKU088, 7: AKU094, 8: AKU097, 9: AKU098, 10: AKU099, 11: AKU100, 12: AKU103, 13: KAIC005.

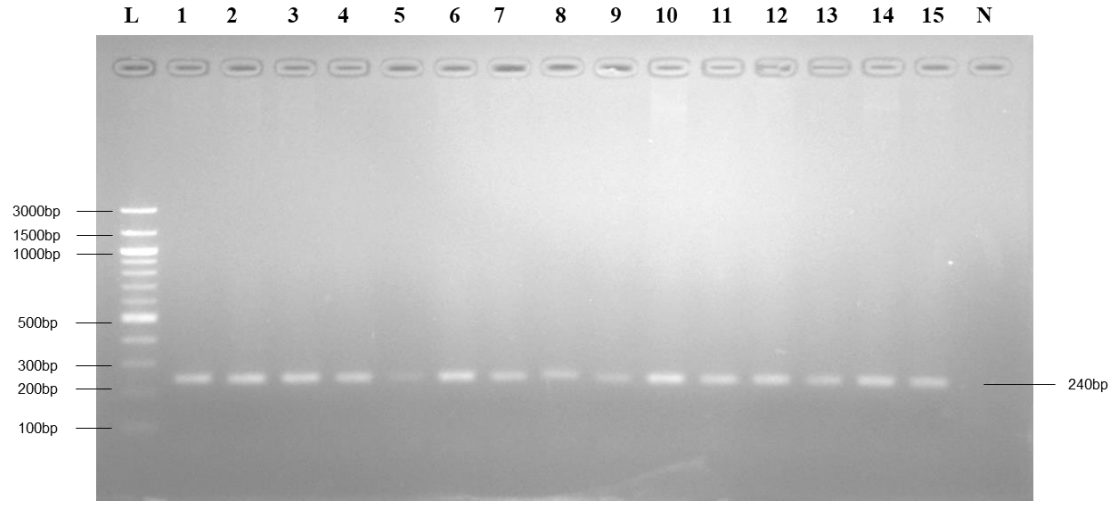


Figure 4.4: Gel Photo of rs1056836 Amplicons

Agarose gel electrophoresis of rs1056836 in ER+ breast cancer cases. L: 100 bp DNA marker, N: negative control. 1: AKU057, 2: AKU060, 3: AKU073, 4: AKU076, 5:

AKU087, 6: AKU088, 7: AKU094, 8: AKU097, 9: AKU098, 10: AKU099, 11: AKU100, 12: AKU103, 13: KAIC005, 14: KAIC006, 15: KAIC007.

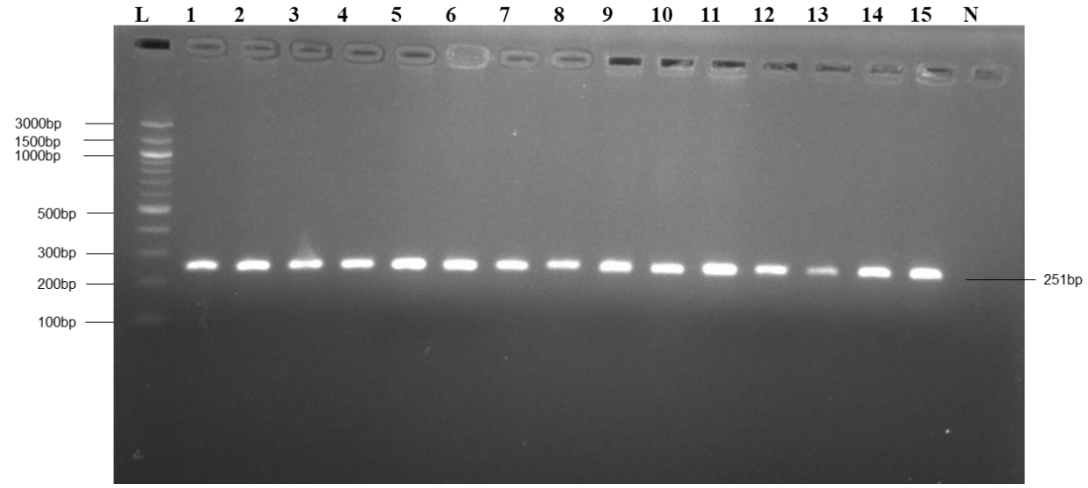


Figure 4.1: Gel Photo of rs776746 Amplicons

Agarose gel electrophoresis of rs776746 amplicons in ER+ breast cancer cases. L: 100 bp DNA marker, N: negative control. 1: AKU057, 2: AKU060, 3: AKU073, 4: AKU076, 5: AKU087, 6: AKU088, 7: AKU094, 8: AKU097, 9: AKU098, 10: AKU099, 11: AKU100, 12: AKU103, 13: KAIC005, 14: KAIC006, 15: KAIC007.

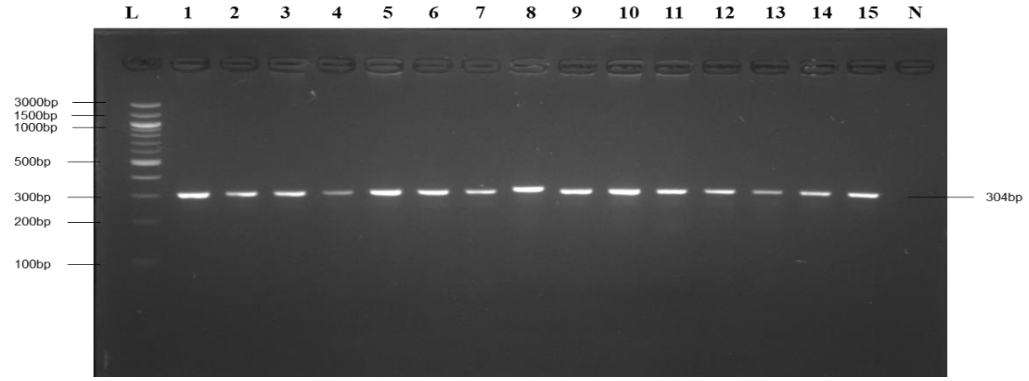


Figure 4.6: Gel Photo of rs4680 Amplicons

Agarose gel electrophoresis of rs4680 amplicons in ER+ breast cancer cases. L: 100 bp DNA marker, N: negative control. 1: AKU057, 2: AKU060, 3: AKU073, 4: AKU076, 5: AKU087, 6: AKU088, 7: AKU094, 8: AKU097, 9: AKU098, 10: AKU099, 11: AKU100, 12: AKU103, 13: KAIC005, 14: KAIC006, 15: KAIC007.

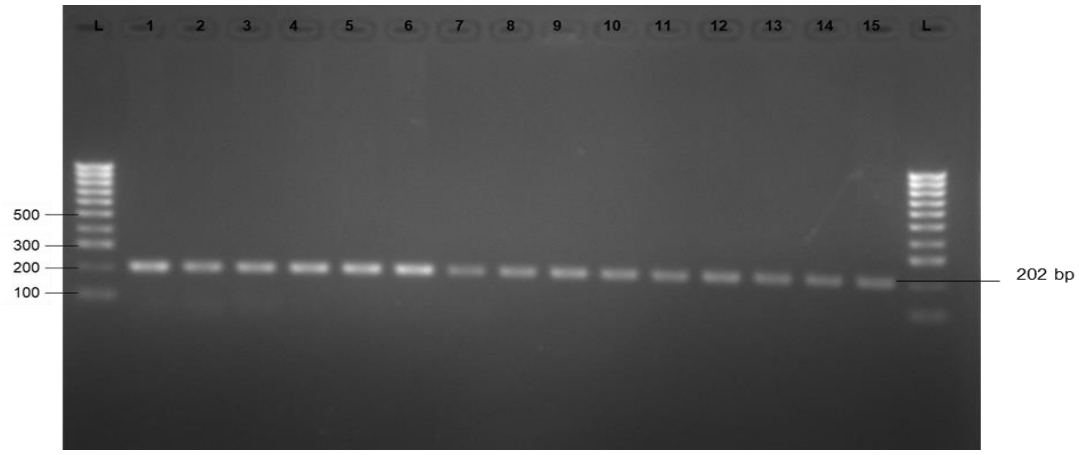


Figure 4.7: Gel photo of rs1056836 and rs1800440 Amplicons

Agarose gel electrophoresis of rs1056836 and rs1800440 target region in ER+ breast cancer cases. L: 100 bp DNA marker, 1: AKU057, 2: AKU060, 3: AKU073, 4: AKU076, 5: AKU087, 6: AKU088, 7: AKU094, 8: AKU097, 9: AKU098, 10: AKU099, 11: AKU100, 12: AKU103, 13: KAIC005, 14: KAIC006, 15: KAIC007.

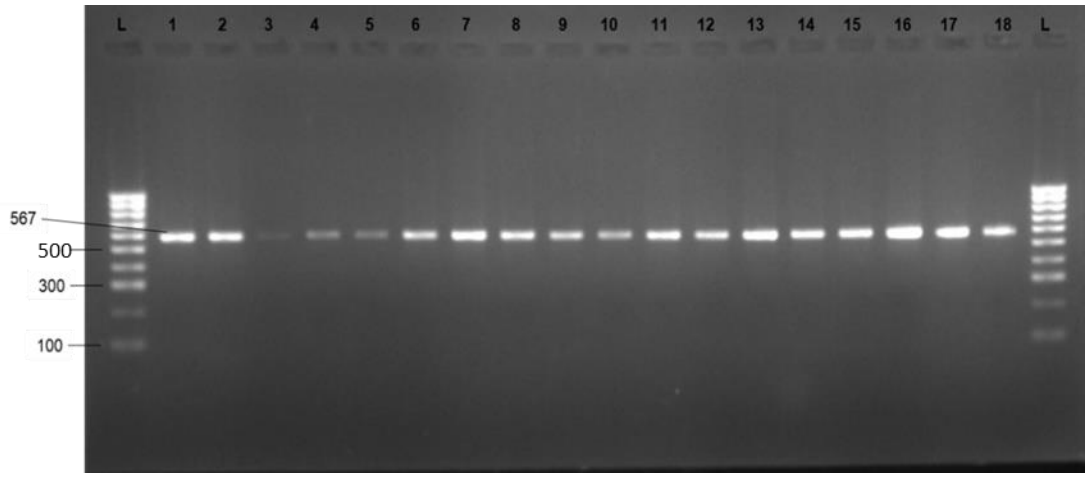


Figure 4.8: Gel Photo of rs10012 and rs1056827 Amplicons

Agarose gel electrophoresis of rs10012 and rs1056827 target region in ER+ breast cancer cases. L: 100 bp DNA marker, 1: AKU057, 2: AKU060, 3: AKU073, 4: AKU076, 5: AKU087, 6: AKU088, 7: AKU094, 8: AKU097, 9: AKU098, 10: AKU099, 11: AKU100, 12: AKU103, 13: KAIC005, 14: KAIC006, 15: KAIC007, 16: KAIC008, 17: KAIC011, 18: KAIC014.

4.3 Restriction Fragment Sizes of PCR amplicons

PCR amplicons digested with specific restriction enzymes were resolved on a 3% 1X TBE agarose gel. Agarose gel photos are shown in Figures 4.9 to 4.13. The photos project the observed banding patterns for each SNP based on restriction digestion. The complete genotype dataset is outlined in **Appendix 7**

4.3.1 Msp1 Restriction Digestion of rs4646903

Figure 4.9 displays a agarose gel photo of MspI restriction digestion for the rs4646903 (T > C) genetic variant. The T/T wild type genotype is represented by a single 340 bp fragment. The T/C heterozygous genotype is characterized by three fragments: 340 bp, 200 bp, and 140 bp. The C/C variant genotype yields two fragments: 200 bp and 140 bp. A 50 bp DNA ladder was included for size estimation.

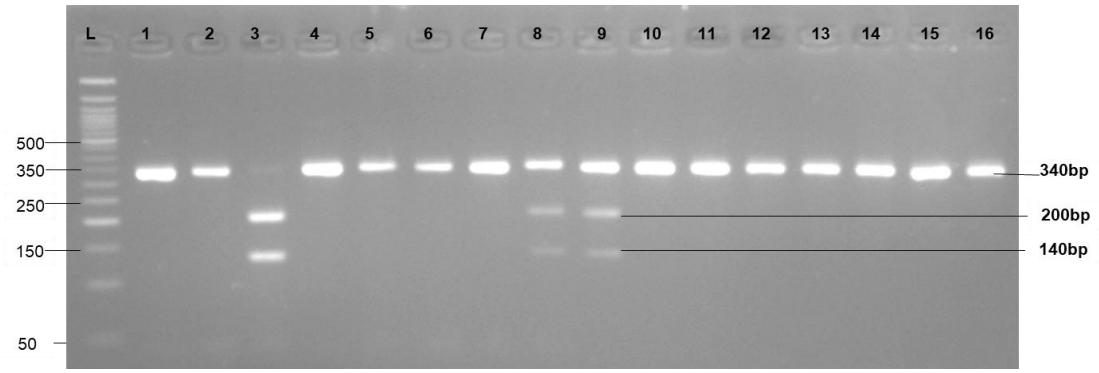


Figure 4.9: Gel Photo of PCR-RFLP Analysis for rs4646903

Agarose gel electrophoresis of MspI restriction digestion of rs4646903 in BBDs. L: 50 bp DNA marker, 1: Positive control (340 bp); 2: AKU011, 4: AKU031, 5: AKU034, 6: AKU039, 7: AKU043, 10: AKU044, 11: AKU046, 12: AKU047, 13: AKU048, 14: AKU049, 15: AKU054, and 16: AKU056 show wild type profiles (340 bp); Lanes 8: AKU059 and 9: AKU069 show heterozygous profiles (340 bp, 200 bp, 140 bp); 3: AKU196 shows a variant profile (200 bp and 140 bp).

4.3.2 BsrD1 Restriction Digestion of rs1048943

Figure 4.10 shows a gel photo of BsrDI digestion for the rs1048943 (A>G) variant. The A/A wild type genotype yields two fragments: 150 bp and 54 bp. The A/G heterozygous genotype produces three fragments: 204 bp, 150 bp, and 54 bp. The G/G variant genotype, along with the control, is represented by a single 204 bp fragment.

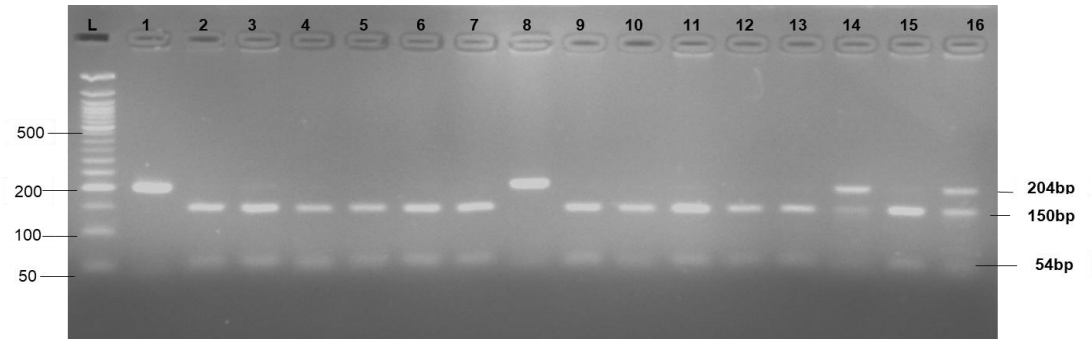


Figure 4.10: Gel Photo of RFLP-PCR for rs1048943

Agarose gel electrophoresis of BsrDI restriction digestion of rs1048943 in ER+ breast cancer cases. L: 50 bp DNA marker, 1: Undigested control (204 bp), 2: AKU055, 3: AKU073, 4: AKU104, 5: AKU105, 6: AKU110, 7: AKU112, 9: AKU116, 10: AKU122, 11: AKU124, 12: AKU127, 13: AKU128 and 15: AKU152 shows a wild type profile (150 bp and 54 bp); 14: AKU149 and 16: AKU173 show a heterozygous profile (204 bp, 150 bp, 54 bp); 8: AKU113 show a variant profile (204 bp).

4.3.3 AcuI Restriction Digestion of rs1056836

Figure 4.11 displays a gel photo of AcuI digestion for the rs1056836 (C > G) variant. The C/C wild type genotype yields a single 245 bp fragment. The C/G heterozygous genotype shows three fragments: 245 bp, 176 bp, and 69 bp. The G/G variant genotype is characterized by two fragments: 176 bp and 69 bp.

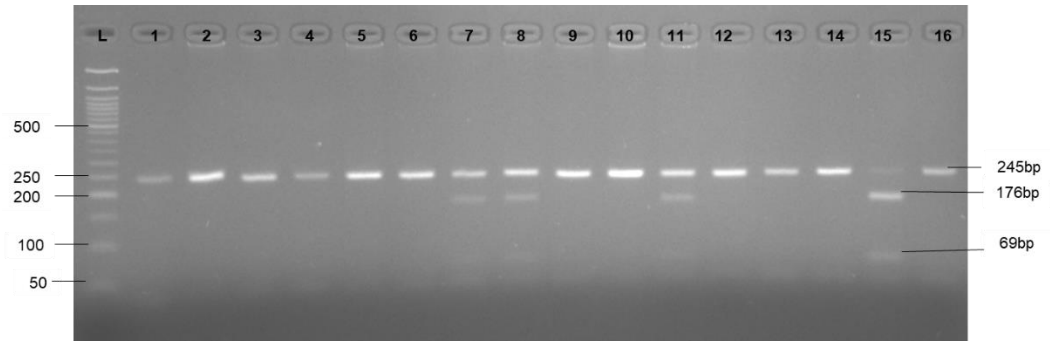


Figure 4.11: Gel Photo of RFLP-PCR for rs1056836

Agarose gel electrophoresis of AcuI restriction digestion of rs1056836 in BBDs. L: 50 bp DNA marker, 1: Undigested control (245 bp), 2: AKU083, 3: AKU187, 4: AKU107, 5: AKU119, 6: AKU126, 9: AKU139, 10: AKU146, 12: AKU153, 13: AKU166, 14: AKU170 and 14: AKU181 showed wild type profiles (245 bp); 7: AKU132, 8: AKU135 and 11: AKU148 showed heterozygous profiles (245 bp, 176 bp, 69 bp) and 15: AKU180 showed variant profile (176 bp and 69 bp).

4.3.4 SspI Restriction Digestion of rs776746

Figure 4.12 presents the gel photo of SspI digestion for rs776746 (A > G). The A/A wild type genotype produces two fragments: 137 bp and 114 bp. The A/G heterozygous genotype would be expected to show three fragments: 251 bp, 137 bp, and 114 bp. The

G/G variant genotype and the control yield a single 251 bp fragment.

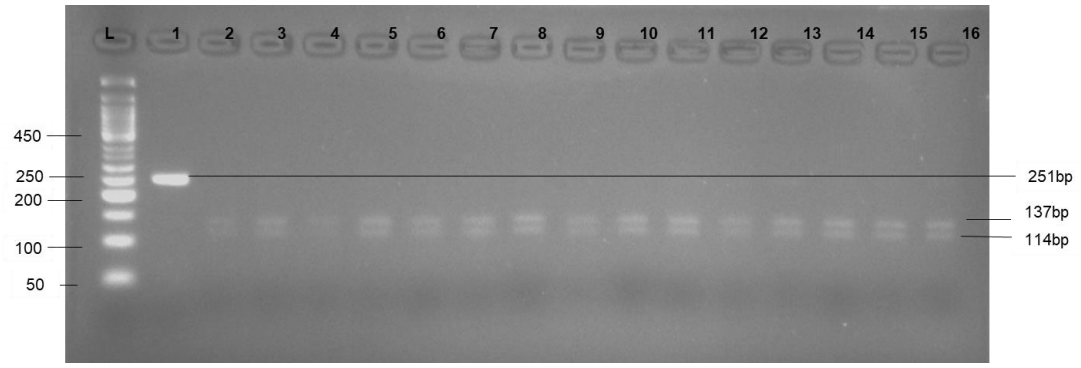


Figure 4.12: Gel Photo of PCR-RFLP for rs776746

Agarose gel electrophoresis of SspI restriction digestion of rs776746 in ER+ breast cancer cases. L: 50 bp DNA marker,1: Undigested control (251 bp), 2: AKU041, 3: AKU042, 4: AKU073, 5: AKU086, 6: AKU104, 7: AKU105, 8: AKU106, 9: AKU110, 10: AKU112, 11: AKU113, 12: AKU116, 13: AKU118, 14: AKU122, 15: AKU124 and 16: AKU127 showed wild type profiles (137 bp and 114 bp).

4.3.5 NlaIII Restriction Digestion of rs4680

Figure 4.13 shows the gel photo of NlaIII digestion for the rs4680 (A > G) variant. The A/A wild type genotype yields four fragments: 114 bp, 93 bp, 54bp and 43bp. The A/G heterozygous genotype would be expected to show all five fragments: 304,114,93,54,43. The G/G variant genotype produces one fragment: 304 bp.

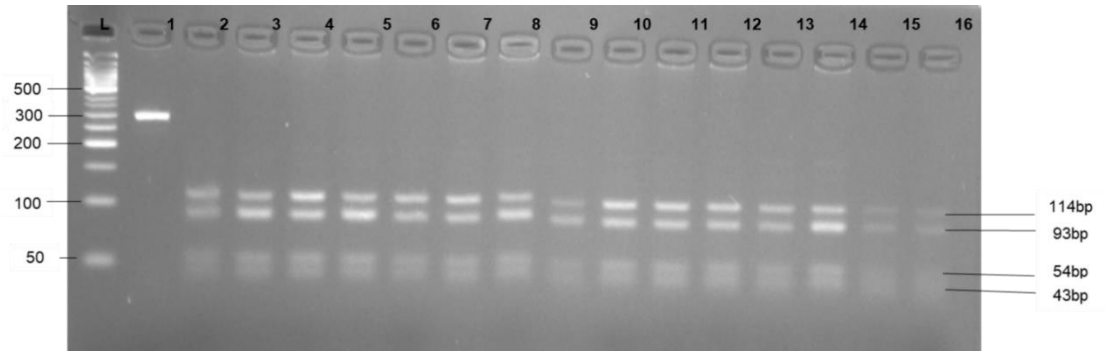


Figure 4.13: Gel Image of RFLP-PCR for rs4680

Agarose gel electrophoresis of NlaIII restriction digestion of rs4680 in BBDs. L: 50 bp DNA marker, 1: Undigested control (304 bp), 2: AKU075, 3: AKU077, 4: AKU083, 5: AKU107, 6: AKU119, 7: AKU126, 8: AKU132, 9: AKU135, 10: AKU139, 11: AKU146, 12: AKU148, 13: AKU153, 14: AKU166, 15: AKU170 and 16: AKU180 showed the wild type profiles (114 bp, 93 bp, 54 bp, and 43 bp).

4.4 Sanger Sequencing of CYP1B1 Gene

A total of 123 samples consisting of 64 cases, 19 controls and 40 benigns were sequenced.

Single nucleotide polymorphisms (SNPs) in CYP1B1 variants (rs10012, rs1056827, rs1056836, and rs1800440) were analyzed using *Geneious Prime* 2025.1.0 (Biomatters Ltd.). Sequence variants were mapped to the GRCh38.p14 reference genome (NCBI accession: NC_000002.12). The sequences were aligned and viewed in MEGA12. Figures 4.14–4.15. The list of Sanger sequences for are detailed in **Appendices VIII and IX.**

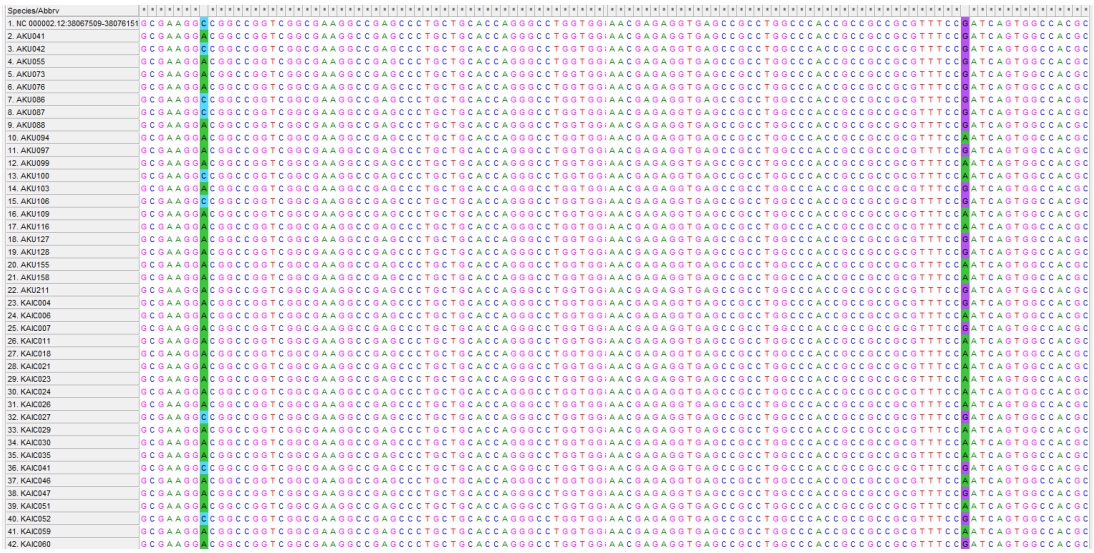


Figure 4.14: Sanger Sequences of rs10012 and rs1056827 SNPs in CYP1B1

The figure displays the CYP1B1 exon 2 sequences. Some samples exhibited wild type allele (C) and others variant allele (A) for the rs10012. For the rs1056827, some samples exhibited the wild type allele (G) and others variant allele (A).

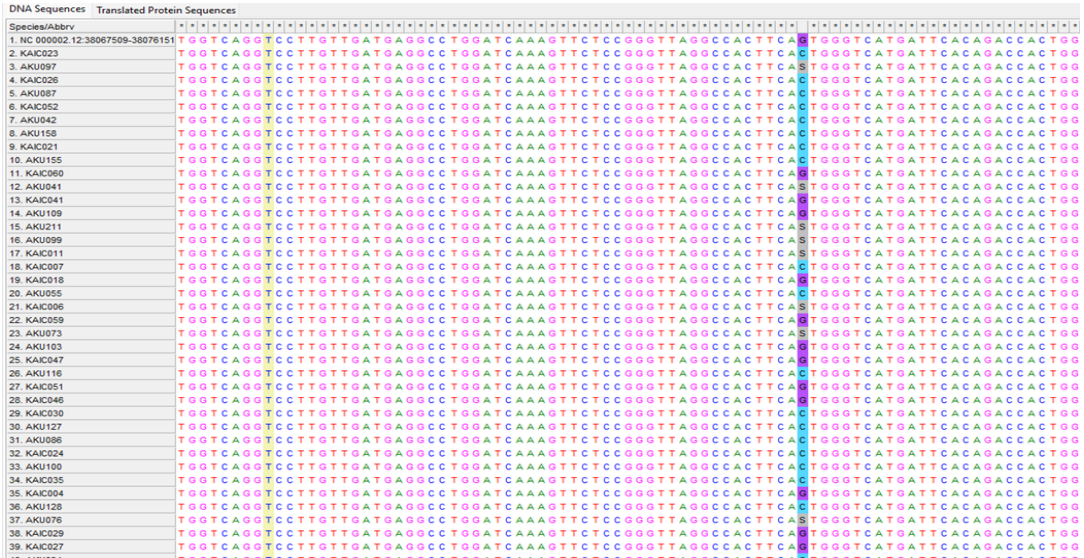


Figure 4.15: Sanger Sequences of rs1800440 and rs1056836 SNPs in CYP1B1

The figure displays the CYP1B1 exon 3 sequences. All the samples exhibited wild type allele (T) for the rs1800440. For the rs1056836 some samples exhibited wild type allele (G), majority exhibited variant allele (C) and a few heterozygous (C/G=S).

4.5 Descriptive Analysis of Socio-Demographic Characteristics, Risk Factors, and SNPs in Study Participants

The data used was obtained from Genomic Breast Cancer Progression Study. Age distribution varied significantly between the study participants. The ages of cases ranged had a median age of 52 years (IQR: 28–78), the controls had a median age of 41 years (IQR: 25–52), while that in the BBD was 42 years (IQR: 23-74). Notably, 58% of the cases were aged 50 years or older, compared to just 5.3% of controls ($p < 0.001$), and 28% of the BBDs ($p < 0.001$). indicating that ER+ breast cancer was more common among older women.

Although not statistically significant, regional distribution showed a notable trend: 84% of the controls were from Nairobi compared to 41% of the cases and 54% of the BBDs. Cases exhibited a wider geographic distribution across the seven regions while the controls and BBDs were majorly from three regions.

BMI distribution did not differ significantly between the groups. While a higher proportion of cases (52%) and BBDs (43%) were overweight or obese, the differences were not statistically significant.

Menopausal status showed a statistically significant difference between cases and controls ($p < 0.001$) as well as between cases and BBDs ($p < 0.001$). Most controls (95%) were premenopausal. This pattern may reflect the older age profile of the cases.

A family history of breast cancer was reported by 20% of cases but none for the controls ($p = 0.033$). Similarly, 33% of cases reported a history of other cancers, compared to none among controls ($p = 0.002$).

Alcohol consumption was significantly more common among controls (74%) and the BBDs (65%) compared to only 38% in the cases. The differences were statistically significant with $p = 0.005$ and $p = 0.001$ respectively. This was a potential reflection of differing lifestyle behaviors among the three groups.

For rs4646903 (CYP1A1*2A), the wild-type (TT) was the most common genotype across all groups, being most prevalent in the BBD group (77%), followed by Cases (67%), and least in Controls (47%). The heterozygous (TC) frequency was similar in Cases (28%) and Controls (37%) but lower in BBDs (22%). The variant homozygous (CC) genotype was rare overall, most frequent in Controls (16%), then Cases (4.7%), and very rare in the BBD group (1.3%).

The rs1048943 (CYP1A1*2C) SNP showed a predominant wild-type (AA) genotype in the vast majority of participants: 100% of Controls, 91% of Cases, and 87% of the BBD group. Variant genotypes were uncommon; the AG genotype was present in Cases (6.3%) and BBDs (10%) but absent in Controls, while the GG genotype was found only in Cases (3.1%) and BBDs (2.5%).

At the rs1056836 (CYP1B1*3) locus, the wild-type (GG) was most frequent in the BBD (71%) and Control (58%) groups but was less common in Cases (47%). The heterozygous (GC) frequency was similar in Cases (41%) and Controls (42%) but lower in BBDs (29%). Critically, the variant homozygous (CC) genotype was found exclusively in the Case group, present in 13% of participants and absent in both Controls and BBDs, which likely drives the significant statistical difference observed.

The analysis for rs776746 (CYP3A5*3) revealed no genetic variation in this cohort. All participants across Cases, Controls, and the BBD group (100%) were homozygous for the A allele (AA). This complete lack of polymorphism makes this SNP non-informative for case-control analysis within the studied population.

For rs4680 (COMT Val158Met), the wild-type (AA) was the predominant genotype in all groups, present in 100% of Controls and BBDs and 94% of Cases. The variant alleles were found only in the Case group: the heterozygous (AG) genotype was present in 4.7% of Cases, and the homozygous (GG) genotype was present in 1.6% of Cases. This exclusive presence in cases contributed to a statistically significant difference compared to the BBD group. (Table 4.1).

Table 4.1: Socio-Demographic Characteristics, Risk Factors, and SNPs Distributions in ER+ Breast Cancer Cases and Controls

Variable	Characteristic	Cases (n = 64[1]) n (%)	Controls (n = 19[1]) n (%)	Cases vs. Controls p-value[2]	BBD (n = 79[1]) n (%)	Cases vs. BBDs p-value[2]
Age (Years)	≥50	37 (58)	1 (5.3)	**<0.001**	18 (23)	**<0.001**
	40-49	17 (27)	10 (53)		34 (43)	
	30-39	9 (14)	4 (21)		24 (30)	
	20-29	1 (1)	4 (21)		3 (3.8)	
	median, IQR	52, 28-78	41, 25-52		42, 23-74	
Region	Nairobi	26 (41)	16 (84)	0.059	43 (54)	0.4
	Central	21 (33)	2 (11)		21 (27)	
	Rift Valley	9 (14)	1 (5.3)		6 (7.6)	
	Eastern	5 (7.8)	0 (0)		8 (10)	
	Coast	1 (1.6)	0 (0)		1 (1.3)	
	Nyanza	1 (1.6)	0 (0)		0 (0)	
	Western	1 (1.6)	0 (0)		0 (0)	
BMI (Kg/M^2)	≥30	33 (52)	6 (32)	0.2	34 (43)	0.5
	25-29	19 (30)	6 (32)		29 (37)	
	18-24	11 (17)	7 (37)		16 (20)	
	<17	1 (1)	0 (0)		0 (0)	
	median, IQR	29, 10-42	29, 19-34		29, 19-43	
Age at menarche (Years)	<12	10 (16)	6 (32)	0.2	53 (67)	0.1
	13-15	32 (50)	10 (53)		14 (18)	
	≥16	22 (34)	3 (16)		12 (15)	
	median, IQR	13, 8-22	13, 10-17		14, 11-17	
Menopausal status	No	30 (47)	18 (95)	**<0.001**	61 (77)	**<0.001**
	Yes	34 (53)	1 (5.3)		18 (23)	
Parity	No	7 (11)	6 (32)	0.065	61 (77)	0.064
	Yes	57 (89)	13 (68)		18 (23)	
Contraceptives	No	21 (33)	2 (11)	0.057	63 (80)	0.088
	Yes	43 (67)	17 (89)		16 (20)	
History of breast cancer	No	51 (80)	19 (100)	**0.033**	62 (78)	0.9
	Yes	13 (20)	0 (0)		17 (22)	
History of other cancers	No	43 (67)	19 (100)	**0.002**	60 (76)	0.2

Variable	Characteristic	Cases (n = 64[1]) n (%)	Controls (n = 19[1]) n (%)	Cases vs. Controls p-value[2]	BBD (n = 79[1]) n (%)	Cases vs. BBDs p-value[2]
	Yes	21 (33)	0 (0)		19 (24)	
Diabetes	No	57 (89)	19 (100)	0.3	72 (91)	0.7
	Yes	7 (11)	0 (0)		7 (8.9)	
Hypertension	No	44 (69)	14 (74)	0.7	61 (77)	0.3
	Yes	20 (31)	5 (26)		18 (23)	
Radiation therapy	No	63 (98)	19 (100)	>0.9	76 (96)	0.6
	Yes	1 (1.6)	0 (0)		3 (4)	
Tobacco	No	62 (97)	19 (100)	>0.9	77 (97)	>0.9
	Yes	2 (3.1)	0 (0)		2 (2.5)	
Alcohol consumption	No	40 (63)	5 (26)	**0.005**	51 (65)	**0.001**
	Yes	24 (38)	14 (74)		28 (35)	
rs4646903	TT	43 (67)	9 (47)	0.13	61 (77)	0.3
	TC	18 (28)	7 (37)		17 (22)	
	CC	3 (4.7)	3 (16)		1 (1.3)	
rs1048943	AA	58 (91)	19 (100)	0.7	69 (87)	0.7
	AG	4 (6.3)	0 (0)		8 (10)	
	GG	2 (3.1)	0 (0)		2 (2.5)	
rs1056836	GG	30 (47)	11 (58)	0.3	56 (71)	**<0.001**
	GC	26 (41)	8 (42)		23 (29)	
	CC	8 (13)	0 (0)		0 (0)	
rs776746	AA	64 (100)	19 (100)	-	79 (100)	-
rs4680	AA	60 (94)	19 (100)	>0.9	79 (100)	**0.038**
	AG	3 (4.7)	0 (0)		0 (0)	
	GG	1 (1.6)	0 (0)		0 (0)	

[1]n (%): Number of participants, percentage
[2]Fisher's exact test; Pearson's Chi-squared test
Bold: Significant p-value of <0.05

Descriptive comparison of socio-demographics, reproductive, lifestyle, and genetic characteristics in ER+ breast cancer and controls. Statistical significant differences were found in age, menopausal status, history of breast cancer, history of other cancers and alcohol consumption. No significant differences were observed in BMI, region, age at menarche, parity, contraceptive use, diabetes, hypertension, radiation therapy, tobacco use, rs4646903, rs1048943, rs1056836 and rs4680. (Data source: Genomic Breast Cancer Progression Study)

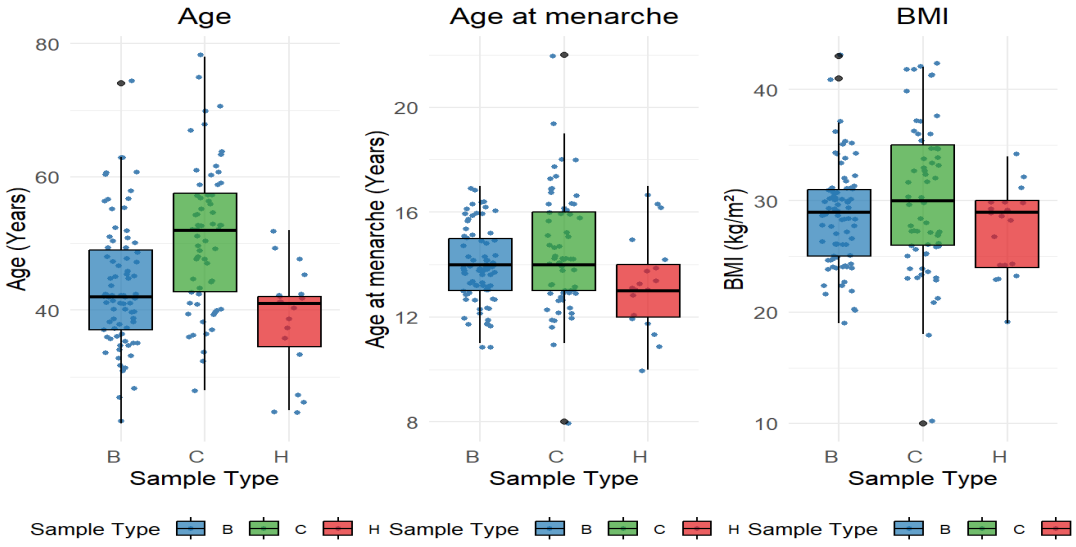


Figure 4.16: Boxplot Comparison of Age, BMI, and Age at Menarche among ER+ Breast Cancer Cases, Benign Breast Disease and Healthy Participants

The figure shows the distributions of age, age at menarche and body mass index (BMI) across the three study groups. The ER+ breast cancer cases (C) were significantly older than both benign breast disease (BBD) and healthy participants (H). No significant differences were observed in the distribution of age at menarche and BMI in the three groups.

4.6 Descriptive Analysis of CYP1B1 SNPs in ER+ Breast Cancer and Control Participants

Table 4.3 presents the distribution of CYP1B1 SNPs (rs10012, rs1056827, rs1056836, and rs1800440) among ER+ breast cancer cases and control participants. The data used was obtained from Genomic Breast Cancer Progression Study. No statistically significant associations were observed between these SNPs and breast cancer risk ($p > 0.05$ for all comparisons).

For rs10012, the heterozygous GC genotype was more prevalent in cases (55%) than controls (37%), while the variant CC genotype was more frequent in controls (53%) compared to the cases (28%). However, the differences were not statistically significant ($p = 0.300$).

For rs1056827, the variant AA genotype was predominant in both cases (72%) and controls (89%) – without significant difference ($p = 0.300$).

Similarly, for rs1056836, genotype distributions were comparable between groups, with the GG genotype slightly more frequent in controls (68%) than in cases (61%), but this difference was not significant ($p = 0.900$).

The rs1800440 SNP was monomorphic across all participants, with everyone exhibiting the TT genotype. (Table 4.2). The genotype profiles of rs10012, rs1056827, rs1056836 and rs1800440 are detailed in **Appendix 10.**

Table 4.2: Genotype Frequencies of CYP1B1 SNPs among ER+ Breast Cancer Cases and Controls

	Characteristic	Cases (n = 64[1]) n (%)	Controls (n = 19[1]) n (%)	p-value[2]
rs10012	GG	11 (17)	2 (11)	0.2
	GC	35 (55)	7 (37)	
	CC	18 (28)	10 (53)	
rs1056827	CC	12 (19)	2 (11)	0.4
	CA	6 (9.4)	0 (0)	
	AA	46 (72)	17 (89)	
rs1056836	GG	39 (61)	13 (68)	0.7
	GC	11 (17)	4 (21)	
	CC	14 (22)	2 (11)	
rs1800440	TT	64 (100)	19(100)	-

[1] n (%) frequencies and percentage
[2] Fisher's exact test; Pearson's Chi-squared test
bold high frequencies

Descriptive analysis of rs10012, rs1056827, rs1056836, and rs1800440 genotypes among ER+ breast cancer versus controls. No statistically significant differences were observed for all SNPs ($p > 0.05$). (Data source: Genomic Breast Cancer Progression Study)

4.7 Descriptive Analysis of CYP1B1 SNPs in ER+ Breast Cancer and Benign Breast Diseases Participants

Table 4.4 compares CYP1B1 genotypic distributions between ER+ breast cancer and participants with BBD. No significant differences in the distribution of the genotypes were found across the genotypes ($p > 0.05$ for all SNPs).

For rs10012, the variant CC genotype was slightly more frequent in BBD cases (37%) compared to ER+ breast cancer (28%). Conversely, the wild type GG genotype was more common in ER+ breast cancer cases (17%) than in BBD participants (8%). However, the differences were not statistically significant ($p = 0.300$).

For rs1056827, the variant AA genotype appeared more frequently in BBD participants (87%) than in breast cancer cases (72%), while the wild type CC genotype was slightly

higher in ER+ breast cancer cases (19%) compared to BBD (11%). Again, this difference was not significant (p = 0.300).

For rs1056836, genotype frequencies were similar across groups. The wild type GG genotype was observed in 63% of BBD and 61% in ER+ breast cancer cases, with no statistically significant difference (p = 0.900) (Table 4.3). The genotype profiles of rs10012, rs1056827, rs1056836 and rs1800440 are detailed in **Appendix 10.**

Table 4.3: Genotype Frequencies of CYP1B1 SNPs among ER+ Breast Cancer Cases and BBD Participants

	Characteristic	Cases (n = 64[1]) n (%)	BBD (n = 38[1]) n (%)	p-value[2]
rs10012	GG	11 (17)	3 (8)	0.4
	GC	35 (55)	21 (55)	
	CC	18 (28)	14 (37)	
rs1056827	CC	12 (19)	4 (11)	0.2
	CA	6 (9)	1 (3)	
	AA	46 (72)	33 (87)	
rs1056836	GG	39 (61)	24 (63)	>0.9
	GC	11 (17)	6 (16)	
	CC	14 (22)	8 (21)	
rs1800440	AA	64 (100)	38 (100)	

[1] n (%) Frequency and percentage
[2] Fisher's exact test; Pearson's Chi-squared test

Descriptive comparison of rs10012, rs1056827, rs1056836, and rs1800440 between ER+ breast cancer and BBDs. No statistically significant differences were observed (p > 0.05).

4.8 Tumor Characteristics in Estrogen Receptor-Positive Breast Cancer Cases

Data was obtained from "A Genomic Approach for Understanding Breast Cancer Progression in Kenya". Among the estrogen receptor positive (ER+) breast cancer participants, the predominant histological type was invasive ductal carcinoma (IDC), accounting for 76.7% of cases. Other histological types included ductal carcinoma in situ (DCIS) (6.3%), invasive carcinoma (6.3%), and invasive lobular carcinoma (ILC) (4.7%).

Most tumors were moderately differentiated (Grade 2: 62.5%), followed by poorly differentiated (Grade 3: 32.8%) and well-differentiated (Grade 1: 4.7%) subtypes

Lymph node involvement was observed in 56.2% of cases (N1-N3), with 43.8% node-negative (N0). Stage II disease was most prevalent (45.3%), followed by Stage I (28.1%) and Stage III (26.6%). Luminal A tumors comprised 82.4% of cases, while Luminal B accounted for 15.6%. (Table 4.4). The clinicalpathological characteristics of the case group participants is outlined in **Appendix 11.**

Table 4.4: Tumor Characteristics of ER+ Breast Cancer Cases

Tumor characteristic	**Levels**	**n = 64**	**%**
Histological type	IDC (Invasive ductal carcinoma)	49	76.7
	ILC (Invasive lobular carcinoma)	3	4.7
	DCIS (Ductal carcinoma in situ)	4	6.3
	IPC (Intracystic papillary carcinoma)	1	1.5
	PDCIS (Papillary ductal carcinoma in situ)	1	1.5
	MC (Mucinous carcinoma)	1	1.5
	IC (Invasive carcinoma)	4	6.3
	IMCS (Internal mammary chain sentinel node)	1	1.5
Grade	Grade 1 (Well differentiated)	3	4.7
	Grade 2 (Moderately differentiated)	40	62.5
	Grade 3 (Poorly differentiated)	21	32.8
Lymph node involvement	N0	28	43.8
	N1	18	28.1
	N2	13	20.3
	N3	5	7.8
Clinical stage	Stage I	18	28.1
	Stage II	29	45.3
	Stage III	17	26.6
Molecular subtypes	Luminal A	54	82.4
	Luminal B	10	15.6

Tumor characteristics ER+ breast cancer cases. This table summarizes the histological types, tumor grades, lymph node involvement, clinical stages, and molecular subtypes of ER+ breast cancer cases. The majority of tumors were invasive ductal carcinoma (76.7%)

and moderately differentiated (Grade 2, 62.5%). Most cases were classified as Stage II (45.3%) and belonged to the Luminal A molecular subtype (82.4%). (Data Source: "A Genomic Approach for Understanding Breast Cancer Progression in Kenya").

4.9 Association between Socio-Demographics and Risk factors with ER+ Breast Cancer Compared to Controls

In the adjusted logistic regression model, menopausal status emerged as a significant predictor of ER+ breast cancer. Postmenopausal women had significantly higher odds of developing ER+ disease compared to premenopausal women (OR = 7.16, 95% CI [1.19, 43.13], p = 0.031).

Other variables—including age category (p = 0.897), BMI (p = 0.476), parity (p = 0.430), and contraceptive use (p = 0.245)—were not significantly associated with ER+ breast cancer. However, trends suggested a possible increased risk with higher BMI and a reduced risk with parity. These findings underscore the importance of hormonal factors—particularly menopausal status—over age alone in determining susceptibility to ER+ breast cancer (Table 4.5).

Table 4.5: Association between Socio-Demographics and Risk Factors with ER+ Breast Cancer Compared to Controls

Variable		OR	95% CI	p-value
Age (years)	≤40	Ref		
	≥41	1.094	0.276-4.330	0.897
BMI (Kg/m^2)	≤24	Ref		
	≥25	1.771	0.367-8.547	0.476
Age at menarche	≥ 12	Ref		
	≤ 13	0.787	0.176-3.526	0.755
Menopausal status	No	Ref		
	Yes	7.157	1.187-43.131	**0.031**
Parity	No	Ref		
	Yes	0.453	0.063-3.233	0.430
Contraceptives use	No	Ref		
	Yes	0.306	0.042-2.256	0.245

OR: odds ratio, CI: confidence interval, bold: -statistical significance of $p < 0.05$, BMI: body mass index

Association of risk factors with ER+ Breast cancer risk compared to controls. Adjusted odds ratios (ORs), 95% confidence intervals (CIs), and p-values from multivariate logistic regression assessing associations between selected variables and ER+ breast cancer. Menopausal status was the only statistically significant predictor.

4.10 Association between Selected Variables and ER+ Breast Cancer Compared to Benign Breast Diseases

In the comparison between ER+ breast cancer cases and participants with benign breast diseases (BBD), menopausal status was again a significant risk factor. Postmenopausal women had over twice the odds of developing ER+ breast cancer (OR = 2.37, 95% CI [1.00, 5.70], p = 0.049).

Other variables—including age at menarche, parity, contraceptive use, history of other cancers, and hypertension—were not significantly associated with ER+ breast cancer. These results suggest that menopausal status is a consistent risk factor (Table 4.6).

Table 4.6: Association between Socio-Demographics and Risk Factors with ER+ Breast Cancer Compared to BBD Participants

Variable		OR	95% CI	p-value
Age (Years)	≤40	Ref		
	≥41	1.144	0.448-2.965	0.777
Age at menarche (years)	≥ 12	Ref		
	≤ 13	1.234	0.436-3.463	0.687
Menopausal status	No	Ref		
	Yes	2.367	1.000-5.698	**0.049**
Parity	No	Ref		
	Yes	0.637	0.201-1.892	0.421
Use of contraceptives	No	Ref		
	Yes	0.776	0.303-1.973	0.593
History of other cancers	No	Ref		
	Yes	1.864	0.845-4.194	0.123
Hypertension	No	Ref		
	Yes	1.166	0.501-2.676	0.717

Key: OR-odds ratio, CI- confidence interval, bold*statistical significance

Association of demographics and factors with ER+ Breast cancer compared to BBD. Adjusted odds ratios (ORs), 95% confidence intervals (CIs), and p-values from multivariate logistic regression comparing ER+ breast cancer cases and BBD participants. Menopausal status increased risk.

4.11 Association of rs4646903, rs1048943, rs1056836, rs776746 and rs4680 with ER+ Breast Cancer Compared to Controls

The rs4646903 polymorphism showed a significant association with breast cancer risk at the allele level. Specifically, the alternative C allele was found to be less frequent among ER+ breast cancer cases compared to controls, suggesting a potential protective effect (OR = 0.44, 95% CI [0.19, 0.99], p = 0.048). To further assess this association, several genotype models were tested. These included the codominant model, which compares each genotype separately (e.g., TT vs. TC vs. CC); the dominant model, which groups individuals carrying at least one copy of the variant allele (TC and CC) and compares them to wild type individuals (TT); the recessive model, which compares variant carriers (CC) to all other genotypes (TT and TC); the overdominant model, which evaluates whether heterozygous individuals (TC) differ in risk compared to both homozygotes (TT and CC); and the log-additive model, which assumes a linear change in risk with each additional copy of the variant allele. However, none of these genotype models reached statistical significance ($p > .05$), indicating that while the allele-level association is notable, the genotypic distribution did not strongly support a model-specific effect.

In contrast, for rs1056836, no statistically significant associations were observed in either allele-based or genotype-based analyses ($p > .05$), suggesting that this variant may not influence breast cancer risk in the study population. Due to limited genotype variation, model convergence issues, and lack of statistical significance, rs1048943 and rs4680 were excluded from the final genotype association analysis (Table 4.7)

Table 4.7: Genotype Frequencies and Associations of rs4646903 and rs1056836 Variants in ER+ Breast Cancer Versus Control Participants

SNP ID	Model	Variants	Cases n (%)	Controls n (%)	OR	95% CI	p-value
rs4646903	Alleles	T	104 (81.3)	25 (65.8)	Ref		
		C	24 (18.7)	13 (34.2)	0.44	0.19-0.99	**0.048**
	Codominant	T/T	43 (67.2)	9 (47.4)	Ref		
		T/C	18 (28.1)	7 (36.8)	0.54	0.17-1.67	0.283
		C/C	3 (4.7)	3 (15.8)	0.21	0.04-1.21	0.081
	Dominant	T/T	43 (67.2)	9 (47.4)	Ref		
		T/C-C/C	21 (32.8)	10 (52.6)	0.44	0.16-1.24	0.121
	Recessive	T/T-T/C	61 (95.3)	16 (84.2)	Ref		
		C/C	3 (4.7)	3 (15.8)	0.26	0.05-1.42	0.121
	Overdominant	T/T-C/C	46 (71.9)	12 (63.2)	Ref		
		T/C	18 (28.1)	7 (36.8)	0.67	0.23-1.97	0.469
	log-Additive	0,1,2	64 (77.1)	19 (22.9)	0.48	0.22-1.05	0.067
rs1056836	Alleles	G	86 (67.2)	30 (79)	Ref		
		C	42 (32.8)	8 (21)	0.54	0.23-1.29	0.169
	Codominant	G/G	30 (46.9)	11 (57.9)	Ref		
		GC	26 (40.6)	8 (42.1)	0.83	0.29-2.40	0.74
		C/C	8(12.5)	0 (0.0)	0.15	0.01-2.92	0.21
	Dominant	G/G	30 (46.9)	11(57.9)	Ref		
		G/C- C/C	34 (53.1)	8 (42.1)	0.64	0.22-1.80	0.40
	Recessive	G/G -G/C	56 (87.5)	19 (100)	Ref		
		C/C	8 (12.5)	0 (0)	0.17	0.01-3.09	0.231
	Overdominant	G/G-C/C	38 (59.4)	11 (57.9)	Ref		
		G/C	26 (40.6)	8 (42.1)	1.06	0.37-3.00	0.908
	log-Additive	0,1,2	64 (77.1)	19 (22.9)	1.82	0.76-4.31	0.298

OR: odds ratio, CI: confidence interval, bold statistical significance ($p < 0.05$)

Genotype and allele frequencies of rs4646903 and rs1056836 polymorphisms and their associations with ER+ breast cancer versus controls. The alternative C allele of rs4646903 was significantly associated with a lower risk of breast cancer (OR = 0.44, p = 0.048), suggesting a protective effect. No significant associations were observed for rs1056836 across all models.

4.12 Associations of rs4646903, rs1048943, rs1056836 SNPs with ER+ Breast Cancer Compared to Benign Breast Diseases

This analysis assessed the association of three SNPs (rs4646903, rs1048943, and rs1056836) with ER+ breast cancer in comparison to benign breast diseases (BBD). rs776746 and rs4680 were not analysed because they were monomorphic wild type in both the controls and BBDs. Among these, rs1056836 emerged as the only SNP with statistically significant associations across multiple genetic models.

In the allelic model, the alternative C allele of rs1056836 was significantly associated with reduced risk of ER+ breast cancer (OR = 0.34, 95% CI [0.19–0.62], p = 0.0003), suggesting a potential protective effect. This finding was consistent in the codominant model, where individuals with the G/C and C/C genotypes exhibited lower risk of ER+ breast cancer at OR = 0.47, 95 CI [0.23–0.96], p = 0.040 and OR = 0.03, 95 CI: 0.00–0.57, p = 0.019 respectively. Similarly, in the dominant model, carriers of either the G/C or C/C genotypes showed a significantly reduced risk compared to G/G homozygotes (OR = 0.36, 95 CI [0.18–0.72], p = 0.004). The recessive model also revealed a protective association for the C/C genotype (OR = 0.04, 95 CI [0.00–0.74], p = 0.030), and the log-additive model (per G allele) showed a dose-dependent increase in risk of ER+ breast cancer (OR = 2.94, 95% CI [1.60–5.39], p = 0.0003).

In contrast, rs4646903 and rs1048943 did not show any statistically significant associations with ER+ breast cancer or BBD. Across all tested genetic models, *p*-values exceeded the standard threshold for significance ($p > 0.05$) (Table 4.8)

Table 4.8: Genotype Frequencies and Associations of rs4646903, rs1048943 and rs1056836 variants in ER+ Breast Cancer Versus Benign Breast Disease Participants

SNP ID	Model	Variants	Cases n (%)	BBDs n (%)	OR	95% CI	p-value
rs4646903	Alleles	T	104 (81)	139 (88)	Ref		
		C	24 (19)	19 (12)	1.68	0.87 - 3.24	0.116
	Codominant	T/T	43 (67.2)	61 (77.2)	Ref		
		T/C	18 (28.1)	17(21.5)	1.50	0.70 - 3.24	0.299
		C/C	3 (4.7)	1 (1.3)	4.26	0.43-42.30	0.216
	Dominant	T/T	43 (67.2)	61(77.2)	Ref		
		T/C-C/C	21 (32.8)	18 (22.8)	1.66	0.79-3.47	0.182
	Recessive	T/T-T/C	61 (95.3)	78 (98.7)	Ref		
		C/C	3 (4.7)	1 (1.3)	3.84	0.39-37.80	0.249
	Overdominant	T/T-C/C	46 (71.9)	62 (78.5)	Ref		
		T/C	18 (28.1)	17 (21.5)	1.43	0.66-3.07	0.362
	log-Additive	0,1,2	64 (44.8)	79 (55.2)	1.66	0.87-3.18	0.121
rs1048943	Alleles	A	120 (94)	146 (92)	Ref		
		G	8 (6)	12 (8)	0.81	0.32-2.04	0.657
	Codominant	A/A	58 (90.6)	69 (87.3)	Ref		
		A/G	4 (6.2)	8 (10.1)	0.59	0.17-2.08	0.415
		G/G	2 (3.1)	2 (2.5)	1.19	0.16-8.71	0.864
	Dominant	A/A	58 (90.6)	69 (87.3)	Ref		
		A/G-G/G	6 (9.4)	10 (12.7)	0.71	0.24-2.08	0.537
	Recessive	A/A-A/G	62 (96.9)	77 (97.5)	Ref		
		G/G	2 (3.1)	2 (2.5)	1.24	0.17-9.07	0.830
	Overdominant	A/A-G/G	60 (93.8)	71 (89.9)	Ref		
		A/G	4 (6.2)	8 (10.1)	0.59	0.17-2.06	0.410
	log-Additive	0,1,2	64 (44.8)	79 (55.2	0.86	0.39-1.90	0.701
rs1056836	Alleles	G	86 (67.2)	135 (85.4)	Ref		
		C	42 (32.8)	23 (14.5)	0.34	0.19-0.62	**0.0003**
	Codominant	G/G	30 (46.9)	56 (70.9)	Ref		
		G/C	26 (40.6)	23 (29.1)	0.47	0.23-0.96	**0.040**
		C/C	8 (12.5)	0 (0)	0.03	0.00-0.57	**0.019**
	Dominant	G/G	30 (46.9)	56 (70.9)	Ref		
		G/C- C/C	34 (53.1)	23 (29.1)	0.36	0.18-0.72	**0.004**
	Recessive	G/G -G/C	56 (87.5)	79 (100)	Ref		
		C/C	8 (12.5)	0 (0)	0.04	0.00-0.74	**0.030**
	Overdominant	G/G-C/C	38 (59.4)	56 (70.9)	Ref		
		G/C	26 (40.6)	23 (29.1)	0.60	0.29-1.20	0.150
	log-Additive	0,1,2	64 (44.8)	79 (55.2)	2.94	1.60-5.39	**0.0003**

Key: OR: odds ratio, CI: confidence interval, bold: statistical significance (p=<0.05)

Association of rs4646903, rs1048943 and rs1056836 variants with ER+ breast cancer versus BBD. This table presents the distribution of genotypes and alleles for three SNPs and their association with ER+ breast cancer versus benign breast diseases, evaluated under different genetic models (allelic, codominant, dominant, recessive, overdominant,

and log-additive). Odds ratios (OR), 95% confidence intervals (CI), and p-values indicate the strength and significance of each association.

4.13 Association of CYP1B1 SNPs with ER+ Breast Cancer Risk Compared to Controls

The association between rs10012, rs1056827, and rs1056836 SNPs acquired via Sanger sequencing and estrogen receptor-positive (ER+) breast cancer risk was evaluated using multiple genetic models. The analysis revealed no statistically significant associations for any of the SNPs studied. All tested models—including allelic, codominant, dominant, recessive, overdominant, and log-additive—produced p-values greater than 0.05. Thus, none of these polymorphisms were significantly associated with ER+ breast cancer risk in this cohort (Table 4.9).

Table 4.9: Genotype Frequencies and Associations of rs10012, rs1056827 and rs1056836 Variants in ER+ Breast Cancer Versus Control Participants

SNP ID	Model	Variants	Controls	Cases	OR	95% CI	p-value
rs10012	Alleles	G	11(28.9)	57(44.5)	Ref		
		C	27(71.7)	71(55.5)	0.5	0.23 - 1.11	0.09
	Codominant	G/G	2 (10.5)	11(17.2)	Ref		
		G/C	7 (36.8)	35(54.7)	0.91	0.16 - 5.03	0.913
		C/C	10 (52.6)	18 (28.1)	0.33	0.06 - 1.78	1.196
	Dominant	G/G	2 (10.5)	11 (17.2)	Ref		
		G/C-C/C	17 (89.5)	53 (82.8)	0.57	0.11 - 2.81	0.488
	Recessive	G/G-G/C	9 (47.4)	46 (71.9)	Ref		
		C/C	10 (52.6)	18 (28.1)	0.35	0.12 - 1.01	0.052
	Overdominant	G/G-C/C	12 (63.2)	29 (45.3)	Ref		
		G/C	7 (36.8)	35(54.7)	2.07	0.72 - 5.94	0.176
	log-Additive	0,1,2	19 (22.9)	64 (77.1)	0.49	0.21 - 1.1	0.073
rs1056827	Alleles	C	34 (89.5)	98 (76.6)	Ref		
		A	4 (10.5)	30 (23.4)	2.6	0.85 - 7.92	0.09
	Codominant	C/C	17 (89.5)	46 (71.9)	Ref		
		C/A	0 (0.0)	6 (9.4)	4.89	0.26 - 91.49	0.288
		A/A	2 (10.5)	12 (18.8)	2.22	0.45 - 10.95	0.328
	Dominant	C/C	17 (89.5)	46 (71.9)	Ref		
		C/A-A/A	2(10.5)	18 (28.1)	3.33	0.7 - 15.88	0.131
	Recessive	C/C-C/A	17 (89.5)	52 (81.2)	Ref		
		A/A	2 (10.5)	12 (18.8)	1.96	0.4 - 9.66	0.41
	Overdominant	C/C-A/A	19 (100)	58 (90.6)	Ref		0.328
			0 (0.0)	6 (9.4)		0	
	log-Additive	0,1,2	19 (22.9)	64 (77.1)	1.73	0.74 - 4.04	0.373

SNP ID	Model	Variants	Controls	Cases	OR	95% CI	p-value
rs1056836	Alleles	G	30 (78.9)	89 (69.5)	Ref		
		C	8 (21.1)	39 (30.5)	0.60	0.25-1.44	0.260
	Codominant	G/G	13 (68.4)	39 (60.9)	Ref		
		G/C	4 (21.1)	11(17.2)	1.09	0.29-4.02	0.896
		C/C	2 (10.5)	14 (21.9)	0.42	0.09-2.14	0.302
	Dominant	G/G	13 (68.4)	39 (60.9)	Ref		
		G/C-C/C	6 (31.6)	25 (39.1)	0.72	0.24-2.14	0.550
	Recessive	G/G-G/C	17 (89.5)	50 (78.1)	Ref		
		C/C	2 (10.5)	14 (21.9)	0.42	0.09-2.04	0.280
	Overdominant	G/G-C/C	15 (78.9)	53 (82.8)	Ref		
		C/G	4 (21.1)	11 (17.2)	1.28	0.35-4.62	0.701
	log-Additive	0,1,2	19 (22.9)	64 (77.1)	0.68	2.79	

Key: OR: odds ratio, CI: confidence interval

This table presents the distribution of alleles and genotypes for rs10012, rs1056827 and rs1056836 and their association with ER+ breast cancer cases versus controls. The odds ratios (OR), confidence intervals (CI), and p-values for various models are reported for various genetic models. No statistically significant associations were observed for any of the SNPs assessed.

4.14 Association of CYP1B1 SNPs with ER+ Breast Cancer Risk Compared to Benign Breast Diseases

This study investigated the association between three polymorphisms - rs10012, rs1056827, and rs1056836 - with the risk of ER+ breast cancer risk compared to BBD.

No significant association was observed for rs10012.Although the C allele appeared more frequently in BBD (64.5%) than in breast cancer cases (55.5%), the difference was not statistically significant. Across various genetic models (codominant, dominant, recessive, overdominant, and log-additive), the associations remained non-significant ($p > 0.05$), indicating no clear link between rs10012 and ER+ breast cancer risk.

Similarly, rs1056836 showed no statistically significant associations across all genetic models ($p > 0.05$), suggesting that this SNP does not contribute meaningfully to differentiating breast cancer from BBD.

However, rs1056827 demonstrated a potential protective effect. The A allele was significantly associated with a reduced risk of ER+ breast cancer (OR = 0.43, 95% CI [0.19 – 0.98], p = 0.045). Although the recessive model (A/A vs. C/C-C/A) also showed a trend toward a protective effect, this was not statistically significant. These findings suggest that rs1056827, particularly the A allele, may be linked to a lower risk of developing ER+ breast cancer. (Table 4.10).

Table 4.10: Genotype Frequencies and Associations of rs10012, rs1056827 and rs1056836 Variants in ER+ Breast Cancer Versus Benign Breast Disease Participants

SNP ID	Model	Variants	BBD	Cases	OR	95% CI	p-value
rs10012	Alleles	G	27 (35.5)	57 (44.5)	Ref		
		C	49 (64.5)	71 (55.5)	0.68	0.382 - 1.232	0.207
	Codominant	G/G	3 (7.9)	11 (17.2)	Ref		
		G/C	21 (55.3)	35 (54.7)	0.45	0.11 - 1.82	0.265
		C/C	14 (36.8)	18 (28.1)	0.35	0.08- 1.50	0.158
	Dominant	G/G	3 (7.9)	11 (17.2)	Ref		
		G/C-C/C	35 (92.1)	53 (82.8)	0.41	0.11 - 1.59	0.197
	Recessive	G/G-G/C	24 (63.2)	46 (71.9)	Ref		
		C/C	14 (36.8)	18 (28.1)	0.67	0.29 - 1.58	0.360
	Overdominant	G/G-C/C	17 (44.7)	29 (45.3)	Ref		0.954
		G/C	21 (55.3)	35 (54.7)	0.98	0.44 - 2.19	0.955
	log-Additive	0,1,2	38 (37.3)	64 (62.7)	0.65	0.34 - 1.22	0.172
rs1056827	Alleles	C	9 (11.8)	30 (23.4)	Ref		
		A	67 (88.2)	98 (76.6)	0.43	0.19 - 0.98	**0.045**
	Codominant	C/C	4 (10.5)	12 (18.8)	Ref		
		C/A	1 (2.6)	6 (9.4)	2.00	0.18 - 22.06	0.571
		A/A	33 (86.8)	46 (71.9	0.46	0.14 - 1.57	0.216
	Dominant	C/C	4 (10.5)	12 (18.8)	Ref		
		C/A-A/A	34 (89.5)	52 (81.2)	0.51	0.15 - 1.71	0.275
	Recessive	C/C-C/A	5 (13.2)	18 (28.1)	Ref		
		A/A	33 (86.8)	46 (71.9)	0.39	0.13 - 1.15	0.087
	Overdominant	C/C-A/A	37 (97.4)	58(90.6)	Ref		
		C/A	1(2.6)	6 (9.4)	3.83	0.44- 33.08	0.222
	log-Additive	0,1,2	38 (37.3)	64 (62.7)	0.63	0.34 - 1.16	0.115
rs1056836	Alleles	G	54 (75)	89 (69.5)	Ref		
		C	18 (25)	39 (30.5)	0.76	0.39-1.46	0.411
	Codominant	G/G	24 (63.2)	39 (60.9)	Ref		
		G/C	6 (15.8)	11 (17.2)	0.88	0.29-2.70	0.832
		C/C	6 (21.1)	14 (21.9)	0.69	0.23-2.05	0.512
	Dominant	G/G	24 (63.2)	39 (60.9)	Ref		
		G/C-C/C	14 (36.8)	25 (39.1)	0.91	0.39-2.08	0.823
	Recessive	G/G-G/C	30 (78.9)	50 (78.1)	Ref		
		C/C	8 (21.1)	14 (21.9)	0.95	0.35-2.53	0.922
	Overdominant	G/G-C/C	32 (84.2)	53 (82.8)	Ref		
		C/G	6 (15.8)	11 (17.2)	0.90	0.30-2.67	0.854
	log-Additive	0,1,2	38 (37.3)	64 (62.7)	1.05	0.64-1.71	0.855

Key: OR-odds ratio, CI- confidence interval, bold*statistical significance

This table presents OR, 95% CI, and p-values for rs10012, rs1056827, and rs1056836. While rs10012 and rs1056836 showed no significant associations ($p > 0.05$), the rs1056827 A allele was linked to a lower breast cancer risk (OR = 0.43, $p = 0.045$), suggesting a possible protective effect.

4.15 Association of rs4646903, rs1048943, rs1056836, rs776746 and rs4680 SNPs with ER+ Breast Cancer Tumor Characteristics

The rs4646903, rs1048943, rs1056836, and rs4680 polymorphisms were genotyped using PCR-RFLP. Their associations with tumor characteristics in ER+ breast cancer were evaluated using five genetic models: allelic, dominant, recessive, additive, and codominant. This section specifically focusses on the relationship between rs4646903, rs1048943, and rs1056836 and key tumor characteristics, including molecular subtype, grade, histopathological type, lymphovascular invasion (LVI), and cancer stage. However, only the results for rs4646903 are presented, as it was the only polymorphism that showed statistically significant associations with tumor features.

The alternative allele C of rs4646903 was significantly associated with an increased risk of the Luminal B subtype (OR=3.83, 95 CI [1.35–10.84], $p = 0.011$). In the codominant model, the heterozygous T/C genotype was significantly associated with an increased risk for the Luminal B subtype (OR=6.67, 95 CI [1.45–30.75], $p = 0.015$). Similarly, the dominant model (T/C-C/C genotypes) indicated a significantly increased risk of the Luminal B subtype (OR = 6.67, 95 CI [1.51–29.38], $p = 0.012$). The overdominant model (T/T-C/C vs. T/C genotypes) showed significant association with Luminal B (OR = 5.25, 95 CI [1.27–21.69], $p = 0.022$). The log-additive model also revealed a statistically significant association between rs4646903 and the Luminal B subtype (OR=3.7, 95 CI [1.23–11.07], $p = 0.017$). These findings suggest that the T/C genotype of rs4646903 may predispose individuals to a higher likelihood of developing Luminal B subtype ER+ breast cancer.

Significant associations were also observed between rs4646903 and breast cancer stage. In both codominant and dominant models, the T/C genotype was significantly associated with an increased risk of late-stage disease (codominant model: OR = 3.50, 95% CI [1.05–11.69], p = 0.041; dominant model: OR = 3.28, 95% CI [1.03–10.43], p = 0.044). The overdominant model showed a similar trend (OR = 3.29, 95% CI [1.01–10.71], p = 0.048), indicating that the T/C variant may be linked to a higher risk of progression to advanced-stage ER+ breast cancer.

In contrast, rs4646903 did not show significant associations with tumor grade, histopathological type (IDC vs. other subtypes), or LVI. Across all genetic models, p-values for these features were >0.05, indicating that rs4646903 may not influence these tumor characteristics in ER+ breast cancer. Table 4.11.

Furthermore, no statistically significant associations were found between rs1048943 and rs1056836 polymorphisms and any of the tumor characteristics assessed. Across all models, p-values were >0.05 for tumor grade, histopathological type, and LVI (data not shown), suggesting these polymorphisms do not significantly impact these aspects in the study population.

Table 4.11: Frequencies and Association of rs4646903 Variants with Tumor Characteristics in ER+ Breast Cancer Patients

Association of rs4646903 with molecular subtype in ER+ breast cancer						
Model	**Variants**	**lum A n (%)**	**lum B n (%)**	**OR**	**95% CI**	**p-value**
Allele	T	92 (85)	12 (60)	Ref		
	C	16 (15)	8 (40)	3.83	1.35- 10.84	**0.011**
Codominant	T/T	40 (74.1)	3 (30)	Ref		
	T/C	12 (22.2)	6 (60)	6.67	1.45-30.75	**0.015**
	C/C	2 (3.7)	1 (10)	6.67	0.46-96.44	0.164
Dominant	T/T	40 (74.1)	3 (30)	Ref		
	T/C-C/C	14 (25.9)	7 (70)	6.67	1.51-29.38	**0.012**
Recessive	T/T-T/C	52 (96.3)	9 (90)	Ref		
	C/C	2 (3.7)	1 (10)	2.89	0.24-35.28	0.406
Overdominant	T/T-C/C	42 (77.8)	4 (40)	Ref		
	T/C	12 (22.2)	6 (60)	5.25	1.27-21.69	**0.022**
log-Additive	0,1,2	54 (84.4)	10 (15.6)	3.7	1.23-11.07	**0.017**

Association of rs4646903 with stage of ER+ breast cancer						
Model	**Variants**	**early stage n (%)**	**late stage n (%)**	**OR**	**95% CI**	**p-value**
Allele	T	80 (77)	24 (70.6)	Ref		
	C	24 (23)	10 (29.4)	1.38	0.58 - 3.30	0.457
Codominant	T/T	35 (74.5)	8 (47.1)	Ref		
	T/C	10 (21.3)	8 (47.1)	3.50	1.05-11.69	**0.041**
	C/C	2 (4.3)	1 (5.9)	2.19	0.18-27.2	0.542
Dominant	T/T	35 (74.5)	8 (47.1)	Ref		
	T/C-C/C	12 (25.5)	9 (52.9)	3.28	1.03-10.43	**0.044**
Recessive	T/T-T/C	45 (95.7)	16 (94.1)	Ref		
	C/C	2 (4.3)	1 (5.9)	1.41	0.12-16.58	0.786
Overdominant	T/T-C/C	37 (78.7)	9 (52.9)	Ref		
	T/C	10(21.3)	8 (47.1)	3.29	1.01-10.71	**0.048**
log-Additive	0,1,2	47 (73.4)	17 (26.6)	2.28	0.9-5.78	0.082
Association of rs4646903 with grade in ER+ breast cancer						
Model	**Variants**	**high grade n (%)**	**low grade n (%)**	**OR**	**95% CI**	**p-value**
Allele	T	99 (81)	5 (83)	Ref		
	C	23 (19)	1 (17)	1.16	0.12 - 10.42	0.893
Codominant	T/T	41 (67.2)	2 (66.7)	Ref		
	T/C	17 (27.9)	1 (33.3)	0.82	0.07 - 9.76	0.881
	C/C	3 (4.9)	0 (0)	0.42	0.01 - 10.62	0.599
Dominant	T/T	41 (67.2)	2 (66.7)	Ref		
	T/C-C/C	20 (32.8)	1 (33.3)	0.97	0.09 - 11.99	0.984
Recessive	T/T-T/C	58 (95.1)	3 (100)	Ref		
	C/C	3 (4.9)	0 (0)	0.418	0.01 - 9.80	0.588
Overdominant	T/T-C/C	44 (72.1)	2(66.7)	Ref		
	T/C	17 (27.9)	1 (33.3)	0.77	0.06 - 9.08	0.837
log-Additive	0,1,2	61 (95.3)	3 (4.7)	0.87	0.1 - 7.25	1
Association of rs4646903 with histopathology type in ER+ breast cancer						
Model	**Variants**	**IDC n (%)**	**Others n (%)**	**OR**	**95% CI**	**p-value**
Allele	T	80 (81.6)	24 (80)	Ref		
	C	18 (18.4)	6 (20)	0.90	0.32 - 2.52	0.841
Codominant	T/T	34 (69.4)	9 (60)	Ref		
	T/C	12 (24.5)	6 (40)	1.89	0.55-6.43	0.308
	C/C	3 (6.1)	0 (0)	0.51	0.02-10.94	0.673
Dominant	T/T	34 (69.4)	9 (60)	Ref		
	T/C-C/C	15 (30.6)	6 (40)	1.51	0.46-5.01	0.499
Recessive	T/T-T/C	46 (93.9)	15 (100)	Ref		
	C/C	3 (6.1)	0 (0)	0.42	0.02-8.76	0.582
Overdominant	T/T-C/C	37 (75.5)	9 (60)	Ref		
	T/C	12 (24.5)	6 (40)	2.06	0.61-6.97	0.248
log-Additive	0,1,2	49 (76.6)	15 (23.4)	1.1	0.41-2.97	0.418

Association of rs4646903 with lymphovascular invasion (LVI) in ER+ breast cancer						
Model	**Variants**	**LVI absent n (%)**	**LVI present n (%)**	**OR**	**95% CI**	**p-value**
Allele	T	46 (82)	48 (77)	Ref		
	C	10 (18)	14 (23)	1.34	0.54- 3.32	0.525
Codominant	T/T	20 (71.4)	23 (63.9)	Ref		
	T/C	6 (21.4)	12 (33.3)	1.74	0.55-5.49	0.345
	C/C	2 (7.1)	1 (2.8)	0.43	0.04-5.16	0.509
Dominant	T/T	20 (71.4)	23 (63.9)	Ref		
	T/C-C/C	8 (28.6)	13 (36.1)	1.41	0.49-4.1	0.524
Recessive	T/T-T/C	26 (92.9)	35 (97.2)	Ref		
	C/C	2 (7.1)	1 (2.8)	0.37	0.03-4.32	0.413
Overdominant	T/T-C/C	22 (78.6)	24 (66.7)	Ref		
	T/C	6 (21.4)	12 (33.3)	1.83	0.59-5.72	0.297
log-Additive	0,1,2	28 (43.8)	36 (56.2)	1.1	0.46-2.62	0.825

Key: OR-odds ratio, CI- confidence interval, bold*statistical significance

This table presents the associations between rs4646903 genotypes and key tumor characteristics—molecular subtype (Luminal A vs. Luminal B), stage (early vs. late), and grade (high vs. low)—under different genetic models (allelic, codominant, dominant, recessive, overdominant, and log-additive). Odds ratios (ORs), 95% confidence intervals (CI), and p-values are provided for each model.

4.16 Association of CYP1B1 SNPs with ER+ Breast Cancer Tumor Characteristics

The rs10012, rs1056827, and rs1056836 polymorphisms were genotyped using Sanger sequencing, and their association with tumor characteristics in ER+ breast cancer were evaluated under five genetic models: allelic, dominant, recessive, additive, and codominant. Although CYP1B1 plays a recognized role in estrogen metabolism, no statistically significant associations were observed between these SNPs and tumor size, histological grade, lymph node involvement, or other clinicopathological features. These findings suggest that rs10012, rs1056827, and rs1056836 may not independently influence tumor characteristics in ER+ breast cancer among the study participants. However, further research with larger and more diverse populations is warranted to validate these findings and clarify the role of CYP1B1 polymorphisms in breast cancer progression (data not shown).

4.17 In-Silico Analysis of CYP1B1 Haplotypes

4.17.1 Linkage Disequilibrium Analysis

The pairwise linkage disequilibrium (LD) analysis of three genetic variants (rs10012, Pairwise linkage disequilibrium (LD) analysis of the three CYP1B1 variants (rs10012, rs1056827, and rs1056836) revealed distinct patterns of association among the loci (Table 4.12). Strong and statistically significant LD was observed between rs10012 and rs1056827, with a D' value of 0.9466 and a moderate correlation ($r^2 = 0.5767$). The chi-square test for this pair yielded a highly significant p-value (1.08×10^{-13}), confirming that these two loci are strongly linked and likely inherited together with minimal recombination.

In contrast, the LD analysis between rs10012 and rs1056836 revealed very weak association (D' = 0.0498, $r^2 = -0.026$, p = 0.7369), indicating that these loci segregate independently. Similarly, rs1056827 and rs1056836 showed weak LD (D' = 0.3416, $r^2 = -0.109$, p = 0.1604), with no statistical significance.

These findings confirm that rs10012 and rs1056827 form part of the same haplotype block, while rs1056836 appears to be independent of these two loci, validating the haplotype structure used for subsequent frequency and structural analyses.

Table 4.12: Summary of Pairwise Linkage Disequilibrium among rs10012, rs1056827 and rs1056836 Polymorphisms

SNP Pair	D′	r^2	Chi-square (X^2)	P-value	LD Strength
rs10012 – rs1056827	**0.9466**	**0.5767**	**55.22**	**1.08×10^{-13}**	**Strong LD (Highly significant)**
rs10012 – rs1056836	0.0498	-0.0260	0.11	0.7369	Weak LD (Not significant)
rs1056827 – rs1056836	0.3416	-0.1090	1.97	0.1604	Weak LD (Not significant)

4.17.2 Haplotype Frequencies of CYP1B1 Polymorphisms

Following confirmation of the linkage patterns, eight distinct CYP1B1 haplotypes were identified across the study groups, with notable differences in distribution observed among breast cancer cases (n=40 chromosomes), healthy controls (n=12 chromosomes), and benign breast disease participants (n=26 chromosomes) (Table 4.13).

The C–A–C–T haplotype (reference) was most frequent in cases (25.0%), while G–A–C–T (protective) predominated in benign samples (42.3%). The wild-type haplotype (G–C–G–T) exhibited a frequency of 5.0% in cases and 3.8% in benign samples, but was absent in controls. Notably, three haplotypes—C–A–G–T, G–C–G–T (wild-type), and G–A–S–T—were absent in controls but collectively accounted for 20.0% of case chromosomes. The risk-associated haplotype (G–C–C–T) showed its highest frequency in cases (12.5%) compared to controls (8.3%) and benign samples (3.8%).

Table 4.13: Haplotype Frequencies of CYP1B1 SNPs in Cases, Controls, and Benign Samples

Haplotype (rs10012–rs1056827–rs1056836–rs1800440)	Amino acids (48-119-432-453)	CASES n (%)	CONTROLS n (%)	BENIGNS n (%)
C – A – C – T**	G-S-V-N	10 (25.0%)	3 (25.0%)	9 (34.6%)
G – A – G – T	R-S-L-N	7 (17.5%)	2 (16.7%)	3 (11.5%)
C – A – S – T	G-S-?-N	6 (15.0%)	3 (25.0%)	2 (7.7%)
G – C – C – T***	R-A-V-N	5 (12.5%)	1 (8.3%)	1 (3.8%)
G – A – C – T****	R-S-V-N	4 (10.0%)	3 (25.0%)	11 (42.3%)
C – A – G – T	G-S-L-N	4 (10.0%)	0 (0%)	2 (7.7%)
G – C – G – T *	R-A-L-N	2 (5.0%)	0 (0%)	1 (3.8%)
G – A – S – T	R-S-?-N	2 (5.0%)	0 (0%)	0 (0%)

Key: *=wild-type haplotype, **=most common haplotype, ***=risk-associated haplotype, ****=protective haplotype, R; Arginine, A; Alanine, V; Valine, N; Asparagine, ?; Leucine or Valine. (ambiguous call); [1]The ambiguous S–?–N indicates that the rs1056836 genotype could not be definitively assigned to either L432 or V432.

4.17.3 Wild-Type Sequence Retrieval and Haplotype Variant Generation

The human CYP1B1 wild-type amino acid sequence (NP_000095.2) was successfully retrieved from the NCBI database and verified against the UniProtKB canonical sequence (Q16678). The 543-amino acid sequence corresponds to the wild-type haplotype (G–C–G–T) identified in the frequency analysis.

Domain mapping confirmed the functional relevance of the four polymorphic residues targeted in this study, as summarized in Table 4.14. Residue R48 is located in the N-terminal hinge region involved in membrane interaction and protein orientation, while A119 resides in the N-terminal structural region critical for protein folding and integrity. Residue L432 is positioned near the substrate-binding pocket, where it influences substrate recognition and binding, and N453 is located in the C-terminal heme-binding region, essential for heme coordination and electron transfer.

Table 4.14: Functional Domain Mapping of Target Residues

Residue Position	Amino Acid	Domain/Region	Functional Role
48	Arginine (R)	Hinge/flexibility region	Membrane interaction and protein orientation
119	Alanine (A)	N-terminal / structural region	Protein folding and structural integrity
432	Leucine (L)	Near substrate-binding pocket	Substrate recognition and binding
453	Asparagine (N)	C-terminal / heme-binding region	Heme coordination and electron transfer

Based on the haplotype frequency data, four representative haplotypes were selected for structural analysis, each defined by a specific combination of amino acids at positions 48, 119, 432, and 453. The wild-type haplotype (G–C–G–T) contains R48, A119, L432, and N453 (R–A–L–N). The reference haplotype (C–A–C–T), which was most common in cases (25.0%), contains G48, S119, V432, and N453 (G–S–V–N). The risk-associated haplotype (G–C–C–T), elevated in cases (12.5%) compared to controls (8.3%) and benign samples (3.8%), contains G48, A119, V432, and N453 (G–A–V–N). The protective

haplotype (G–A–C–T), which predominated in benign samples (42.3%), contains R48, S119, V432, and N453 (R–S–V–N).

The complete amino acid sequences for these four haplotypes, generated through in silico mutagenesis by introducing the corresponding substitutions into the wild-type template, are presented in Figure 4.19. The four polymorphic residues are highlighted in bold and underlined at their respective positions.

A. Wild-Type (R–A–L–N) Sequence

MGTSLSPNDPWPLNPLSIQQTTLLLLLSVLATVHVGQRLL**R**QQRRRQLRSAPPG
PFAWPLIGNAAAVGQAAHLSFARLARRYGDVFQIRLGSCPIVVLNGERAIHQAL
VQQGSAFADRPAFASFRVVSGGRSMAFGHYSEHWKVQRR**A**AHSMMRNFFTRQ
PRSRQVLEGHVLSEARELVALLVRGSADGAFLDPRPLTVVAVANVMSAVCFGC
RYSHDDPEFRELLSHNEEFGRTVGAGSLVDVMPWLQYFPNPVRTVFREFEQLNR
NFSNFILDKFLRHCESLRPGAAPRDMMDAFILSAEKKAA**L**GDSHGGGARLDLEN
VPATITDIFGASQDTLSTALQWLLLLFTRYPDVQTRVQAELDQVVGRDRLPCMG
DQPNLPYVLAFLYEAMRFSSFVPVTIPHATTANTSVLGYHIPKDTVVFVNQWSV
NHDPLKWPNPENFDPARFLDKDGLINKDLTSRVMIFSVGKRRCIGEELSKMQLF
LFISILAHQCDFRANPNEPAKMNFSYGLTIKPKSFKVNVTLRESMELLDSAVQNL
QAKETC**N**Q

B. Reference (G–S–V–N) Haplotype Sequence

MGTSLSPNDPWPLNPLSIQQTTLLLLLSVLATVHVGQRLL**G**QQRRRQLRSAPPG
PFAWPLIGNAAAVGQAAHLSFARLARRYGDVFQIRLGSCPIVVLNGERAIHQAL
VQQGSAFADRPAFASFRVVSGGRSMAFGHYSEHWKVQRR**S**AHSMMRNFFTRQ
PRSRQVLEGHVLSEARELVALLVRGSADGAFLDPRPLTVVAVANVMSAVCFGC
RYSHDDPEFRELLSHNEEFGRTVGAGSLVDVMPWLQYFPNPVRTVFREFEQLNR
NFSNFILDKFLRHCESLRPGAAPRDMMDAFILSAEKKAA**V**GDSHGGGARLDLEN
VPATITDIFGASQDTLSTALQWLLLLFTRYPDVQTRVQAELDQVVGRDRLPCMG

DQPNLPYVLAFLYEAMRFSSFVPVTIPHATTANTSVLGYHIPKDTVVFVNQWSV
NHDPLKWPNPENFDPARFLDKDGLINKDLTSRVMIFSVGKRRCIGEELSKMQVF
LFISILAHQCDFRANPNEPAKMNFSYGLTIKPKSFKVNVTLRESMELLDSAVQNL
QAKETC**N**Q

C. Risk-Associated (G–A–V–N) Haplotype Sequence

MGTSLSPNDPWPLNPLSIQQTTLLLLLSVLATVHVGQRLL**G**QQRRRQLRSAPPG
PFAWPLIGNAAAVGQAAHLSFARLARRYGDVFQIRLGSCPIVVLNGERAIHQAL
VQQGSAFADRPCFASFRVVSGGRSMAFGHYSEHWKVQRR**A**AHSMMRNFFTRQ
PRSRQVLEGHVLSEARELVALLVRGSADGAFLDPRPLTVVAVANVMSAVCFGC
RYSHDDPEFRELLSHNEEFGRTVGAGSLVDVMPWLQYFPNPVRTVFREFEQLNR
NFSNFILDKFLRHCESLRPGAAPRDMMDAFILSAEKKAA**V**GDSHGGGARLDLEN
VPATITDIFGASQDTLSTALQWLLLLFTRYPDVQTRVQAELDQVVGRDRLPCMG
DQPNLPYVLAFLYEAMRFSSFVPVTIPHATTANTSVLGYHIPKDTVVFVNQWSV
NHDPCKWPNPENFDPARFLDKDGLINKDLTSRVMIFSVGKRRCIGEELSKMQVF
LFISILAHQCDFRANPNEPAKMNFSYGLTIKPKSFKVNVTLRESMELLDSAVQNL
QAKETC**N**Q

D. Protective (R–S–V–N) Haplotype Sequence

MGTSLSPNDPWPLNPLSIQQTTLLLLLSVLATVHVGQRLL**R**QQRRRQLRSAPPG
PFAWPLIGNAAAVGQAAHLSFARLARRYGDVFQIRLGSCPIVVLNGERAIHQAL
VQQGSAFADRPAFASFRVVSGGRSMAFGHYSEHWKVQRR**S**AHSMMRNFFTRQ
PRSRQVLEGHVLSEARELVALLVRGSADGAFLDPRPLTVVAVANVMSAVCFGC
RYSHDDPEFRELLSHNEEFGRTVGAGSLVDVMPWLQYFPNPVRTVFREFEQLNR
NFSNFILDKFLRHCESLRPGAAPRDMMDAFILSAEKKAA**V**GDSHGGGARLDLEN
VPATITDIFGASQDTLSTALQWLLLLFTRYPDVQTRVQAELDQVVGRDRLPCMG
DQPNLPYVLAFLYEAMRFSSFVPVTIPHATTANTSVLGYHIPKDTVVFVNQWSV
NHDPLKWPNPENFDPARFLDKDGLINKDLTSRVMIFSVGKRRCIGEELSKMQVF

LFISILAHQCDFRANPNEPAKMNFSYGLTIKPKSFKVNVTLRESMELLDSAVQNL
QAKETC**N**Q

Figure 4.17: CYP1B1 Haplotype Amino Acid Sequences

The four haplotypes selected for structural analysis are shown with the polymorphic residues (positions 48, 119, 432, 453) highlighted in bold. Substitutions relative to the wild-type sequence are highlighted in bold and underlined. A; Wild-type (R–A–L–N), B. Reference (G–S–V–N), C. Risk-associated (G–A–V–N) and D. Protective (R–S–V–N)

4.17.4 Tertiary Structure Analysis

The ribbon structures, surface models and colour coded ribbons diagrams of the four haplotype models were visualized in PyMol. Ramachandran plot analysis was performed to evaluate the stereochemical quality of the wild-type and variant CYP1B1 models. All the models demonstrated >90% of residues in favored regions, with no significant outliers in disallowed regions, confirming the structural reliability of the generated models for subsequent MD simulations.

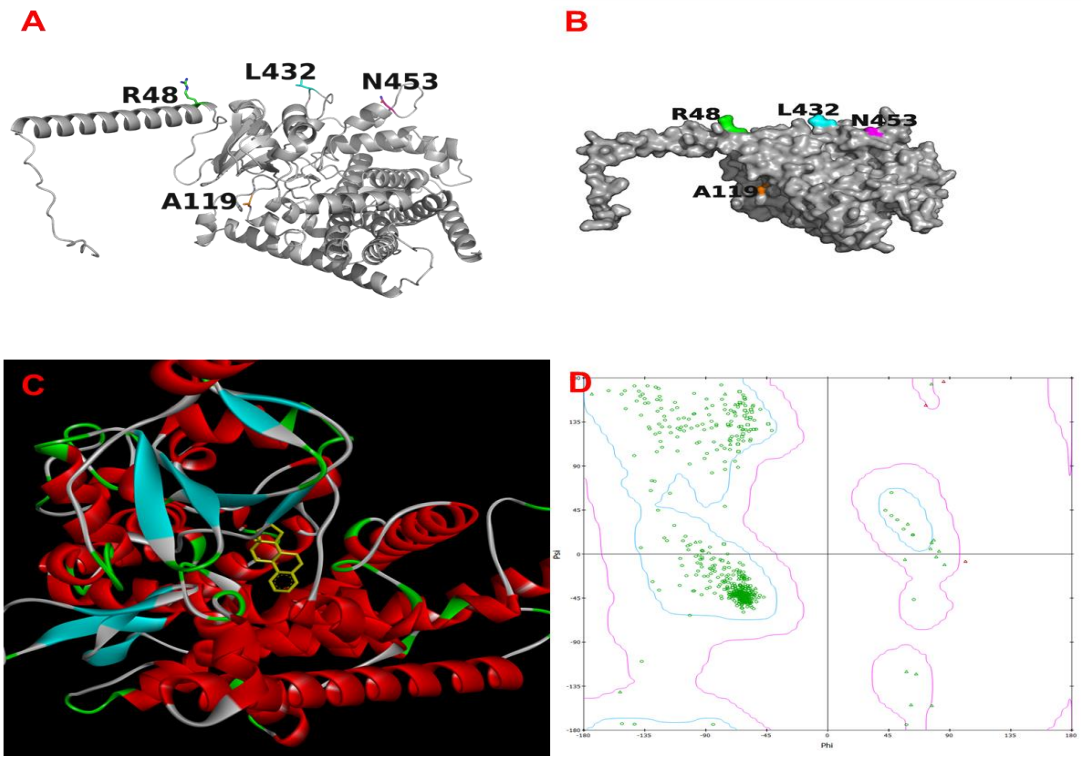


Figure 4.18: Tertiary Structures and Ramachandran Plot of CYP1B1 Highlighting the R48, A119, L432 and N453 Residues.

A; Ribbon structure, B; Surface model, C; Colour coded ribbon diagram and D; Ramachandran plot

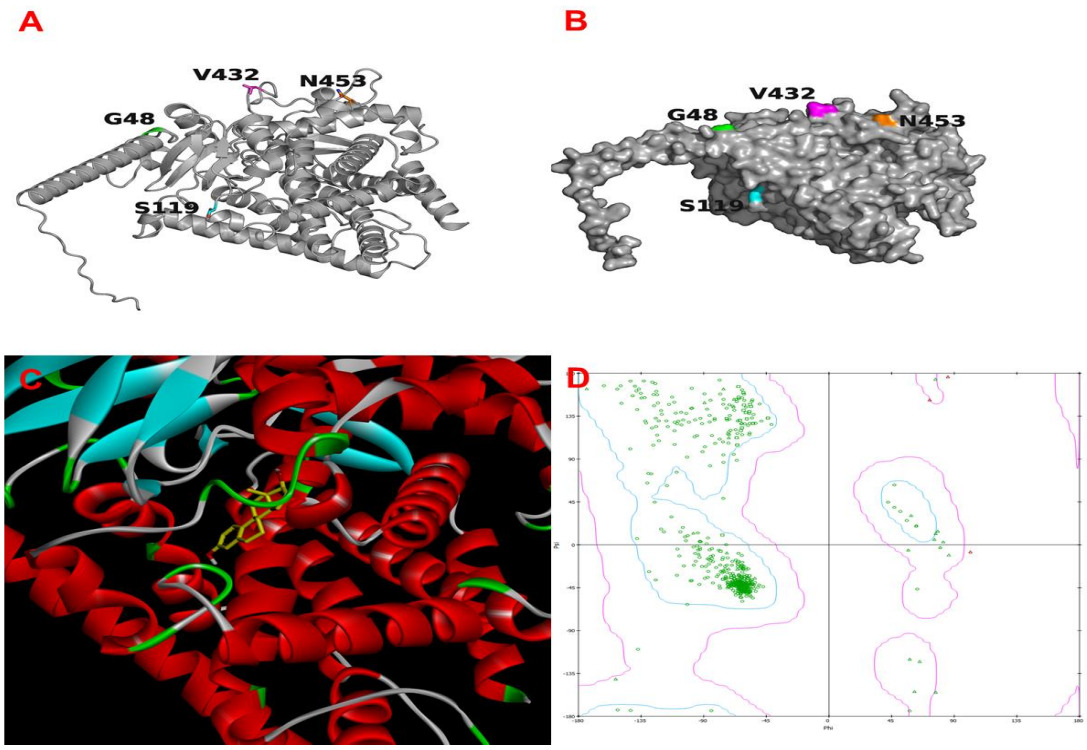


Figure 4.19: Tertiary Structures and Ramachandran Plot of CYP1B1 Highlighting the G48, S119, V432 and N453 Residues

A; Ribbon structure, B; Surface model, C; Colour coded ribbon diagram and D; Ramachandran plot

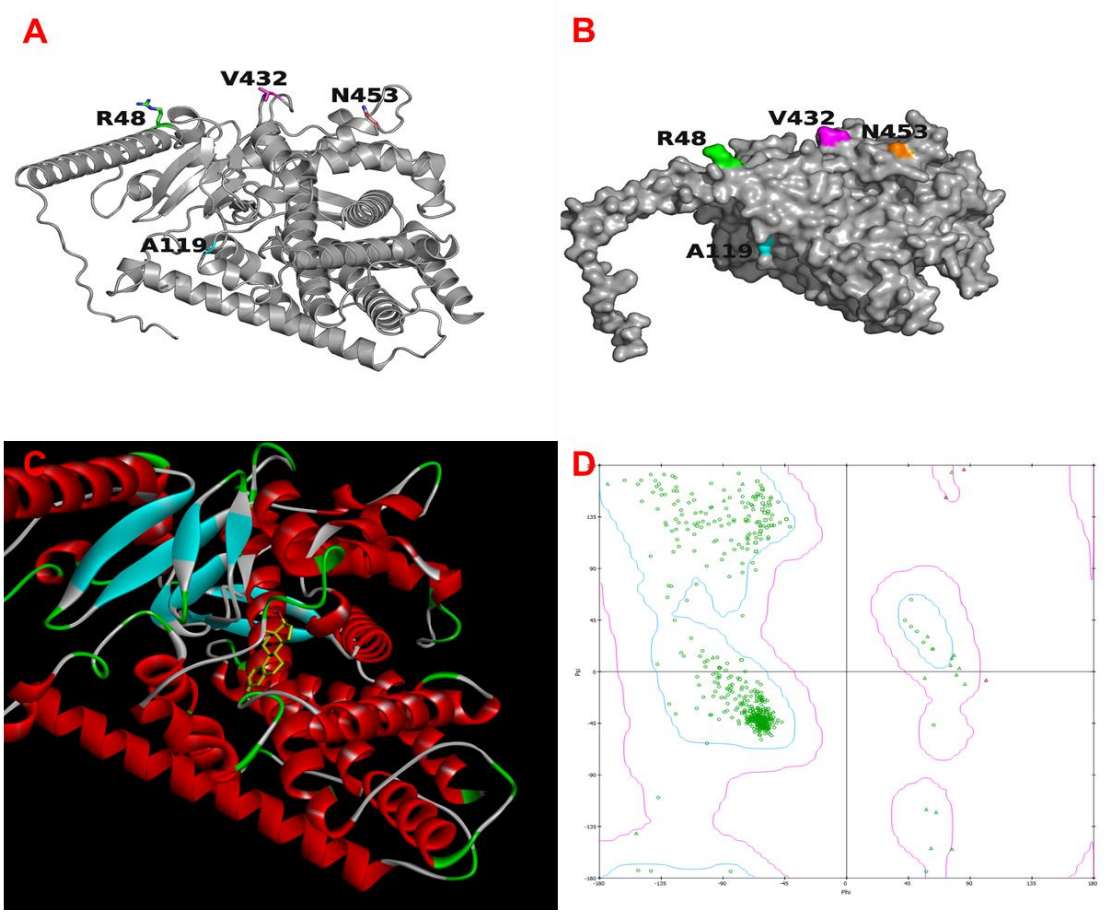


Figure 4.20: Tertiary Structures And Ramachandran Plot of CYP1B1 Highlighting the R48, A119, V432 and N453 Residues

A; Ribbon structure, B; Surface model, C; Colour coded ribbon diagram and D; Ramachandran plot

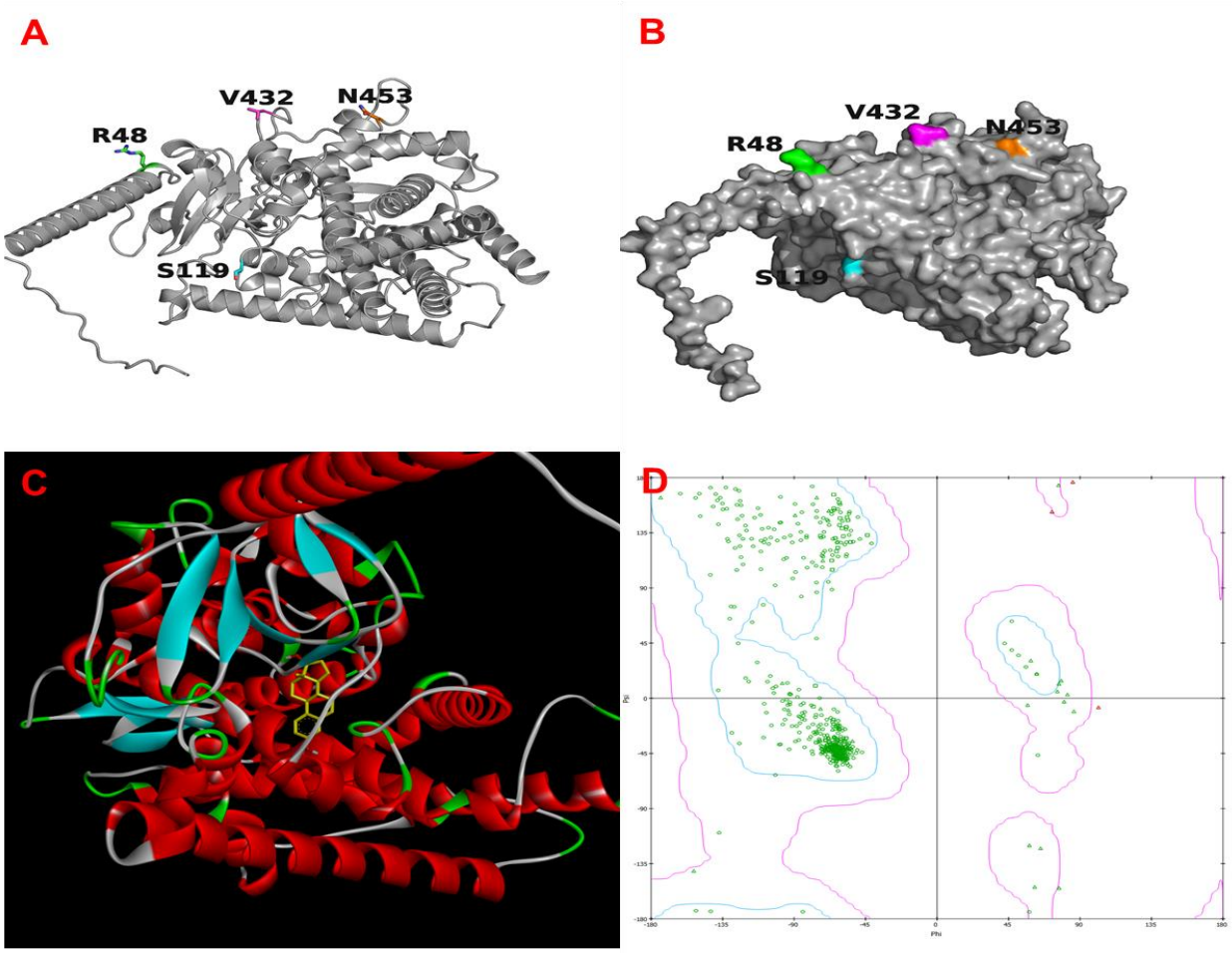


Figure 4.21: Tertiary Structures and Ramachandran Plot of CYP1B1 Highlighting the R48, S119, V432 and N453 Residues

A; Ribbon structure, B; Surface model, C; Colour coded ribbon diagram and D; Ramachandran plot

CHAPTER FIVE

DISCUSSION

5.1 Overview

The increasing global burden of breast cancer, driven by rising incidence and mortality rates in developing countries, underscores its significance as a public health challenge (Cai et al., 2025). Estrogen receptor-positive (ER+) breast cancer accounts for 70–80% of all breast cancer cases and relies on estrogen for tumor growth (Özdemir et al., 2018). This type of breast cancer may be influenced by genetic variations involved in estrogen metabolism. Polymorphisms in genes such as CYP1A1, CYP1B1, CYP3A5, and COMT can alter estrogen metabolism efficiency, therefore potentially modulating the risk of developing ER+ breast cancer (Tsuchiya et al., 2005). However, associations between these polymorphisms and breast cancer risk remain inconsistent, possibly due to differences in population-specific genetic susceptibility, environmental factors, lifestyle choices and reproductive histories (Sun et al., 2015). In addition to genetic influences, socio-demographic disparities may further contribute to risk variability. Thus, investigating these polymorphisms alongside reproductive, lifestyle, and socio-demographic factors is essential to understanding the multifaceted etiology of estrogen ER+ breast cancer in the local population.

This study examined the associations between socio-demographic characteristics, medical history, lifestyle behaviors, and genetic polymorphisms specifically rs4646903 and rs1048943 in CYP1A1, rs10012, rs1056827, rs1056836, and rs1800440 in CYP1B1, rs776746 in CYP3A5, and rs4680 in COMT. These genes were selected based on their previously reported roles in breast cancer susceptibility, with breast cancer risk. The analysis involved a comprehensive approach, including descriptive statistics, association analysis, and haplotype analysis. The comparisons were conducted between ER+ breast cancer cases and two reference groups - healthy controls and individuals with benign

breast disease (BBD). These comparisons are critical for identifying factors that may influence breast cancer development and progression.

Notably, comparing ER+ breast cancer cases with the BBD group provides distinct insights. BBD shares several clinical and pathological features with breast cancer but carries a distinct risk profile. It is well established that women diagnosed with BBD are at an elevated risk of developing breast cancer (Figueroa et al., 2021). Supporting this link, Schilling and Silva (2020) reported a significant association between family history of breast cancer and BBD, reinforcing the importance of genetic and familial factors in breast cancer susceptibility (Schilling & Silva, 2020).

5.2 Prevalence of the Socio-Demographics, Risk Factors and Single Nucleotide Polymorphisms in the Study Participants

Age is a well-documented risk factor for ER+ breast cancer. These findings in this study are consistent with existing evidence stating that advancing age, particularly postmenopausal status, increases the risk of ER+ breast cancer (Diab et al., 2000).

Globally, the median age at breast cancer diagnosis varies across regions. Studies in Ghana, Saudi Arabia, and Egypt reported higher median ages - 55 years, 53.7 years, and 50–59 years, respectively (Bosompem et al., 2024; El-Moselhy, 2017; Khabaz, 2014) - aligning with the findings of this study. In contrast, breast cancer in Sub-Saharan Africa, North Africa, and parts of Asia tends to affect younger populations. For instance, in Kenya, the most affected age group was 45–49 years (Makanga et al., 2013), with other studies reporting median ages of 48 and 47.5 years (Sayed et al., 2014, 2018). In Western Kenya, a median age of 48.5 years was reported (Sawe et al., 2016). Similarly, studies from Uganda (Galukande et al., 2014), Libya (Abousahmeen et al., 2023) and Eastern Morocco (Errahhali et al., 2017) reported median ages ranging between 47.3 and 48.7 years. A recent study examining a heterogeneous African population (Nigeria, Cameroon, and Uganda), reported a median age of 48 years, compared to 31 years in the BBD group (Omoleye et al., 2023).

These age disparities likely reflect regional differences in population demographics, with younger population structures in Sub-Saharan Africa, as well as possible variations in genetic susceptibility, environmental exposures, and access to diagnostic services (Errahhali et al., 2017). In contrast, Western countries tend to report older ages at diagnosis of breast cancer, which may be attributed to higher life expectancy and later onset of menopause (Leong et al., 2010).

The majority of ER+ breast cancer cases in this study were postmenopausal. This aligns with previous findings that identify menopause as a significant risk factor for the disease (Collaborative Group on Hormonal Factors in Breast Cancer, 2001). The increased risk in postmenopausal women is largely attributed to hormonal changes, particularly the shift from ovarian estrogen production to peripheral estrogen synthesis in adipose tissue (Cleary & Grossmann, 2009) . Postmenopausal breast cancer risk is closely linked to increased adiposity, which not only elevates circulating estrogen levels but also reduces sex hormone-binding globulin (SHBG), leading to higher bioavailable estrogen. This hormonal imbalance fosters carcinogenic processes, including enhanced cell proliferation, pro-angiogenesis, and suppression of apoptosis, all of which contribute to tumor development and progression (Rinaldi et al., 2006).

The association between menopausal status and BBD differs from its link with ER+ breast cancer. While postmenopausal women are at a higher risk of developing ER+ breast cancer, BBD is more common in younger, premenopausal women (Hasan et al., 2020). In this study, the majority of BBD cases occurred in premenopausal women, reinforcing the idea that age and hormonal fluctuations influence both benign and malignant breast conditions, but in distinct ways.

This contrast suggests that menopause is more strongly associated with an increased risk of breast cancer, whereas BBD tends to occur earlier in life when hormonal fluctuations are more frequent (Johansson et al., 2021). These findings highlight the importance of age-specific screening and risk assessment to distinguish benign from malignant conditions effectively.

Family history remains an important risk factor for breast cancer. While previous studies show that 10–30% of patients have affected relatives (Brewer et al., 2017). This study highlights that family history of breast or other cancers influences breast cancer risk. This suggests broader genetic or shared lifestyle factors could play a role in disease development.

Similar trends have been observed in previous studies. In Morocco, Tazzite et al. (2013) found that 18.4% of breast cancer patients had a family history of the disease (Tazzite et al., 2013). In the United States, Haber et al. (2012) reported that 11.3% of women with breast cancer had a family history of breast cancer, while 46% had a family history of other cancers (Haber et al., 2012). In China, Kilfoy et al. (2008) found an increased risk of breast cancer in individuals with a family history, although no association was observed between a history of other cancers and breast cancer risk (Kilfoy et al., 2008). However, a subgroup analysis by Zhou et al. (2014) in Chinese women linked a family history of lung cancer to an increased risk of ER+ breast cancer (Zhou et al., 2014).

Findings from Kenya also align with this study. Ekpe et al. (2019) reported that 23% of breast cancer patients had a family history of cancer, closely matching 20% prevalence of familial breast cancer cases in this study (Ekpe et al., 2019). While family history suggests a combination of genetic and environmental influences, the precise mechanisms linking breast cancer with other familial cancers remain unclear (Kilfoy et al., 2008).

Moderate alcohol consumption has been associated with a 30–50% increased risk of breast cancer, a relationship consistently observed in both case-control and cohort studies (McDonald et al., 2013). However, in this study, alcohol intake was assessed only over the past 12 months, which may not fully capture the cumulative effects of long-term alcohol consumption. Individuals in the ER+ breast cancer group may have reduced their alcohol intake after diagnosis due to health concerns or treatment side effects, potentially leading to lower reported consumption compared to the control and BBD groups. The 12-month assessment period may not accurately reflect the drinking patterns that contribute to breast cancer development over the long term. Alcohol's carcinogenic effects

accumulate over time, suggesting that past consumption, beyond the last year, could play a more significant role in breast cancer risk. Additionally, the type of alcoholic beverage consumed may also influence disease development (Liu et al., 2015). Given these factors, further research is needed to explore the timing, duration, and patterns of alcohol consumption in relation to breast cancer risk.

This study characterized the baseline distribution of key polymorphisms in estrogen metabolism genes, including CYP1A1 (rs4646903 and rs1048943), CYP1B1 (rs10012, rs1056827, rs1056836, and rs1800440), CYP3A5 (rs776746), and COMT (rs4680), across breast cancer cases, benign breast disease controls, and healthy controls. These specific variants were selected based on their established roles in estrogen biosynthesis and metabolism pathways. Notably, variation in rs1056836 distribution was observed between the cases and the controls. This suggests possible effects that may require consideration in subsequent analyses. The findings contribute to building a more complete understanding of the genetic architecture underlying estrogen metabolism in breast cancer pathogenesis

5.3 Prevalence of Tumor Characteristics

Breast cancer is a heterogeneous disease with significant variability in tumor characteristics across different ethnic and geographic populations. These tumor features influence disease progression, treatment response, and overall prognosis, making them essential for clinical decision-making. Key tumor characteristics include histological grade, histological type, lymph node status, molecular subtype, and disease stage, all of which play a critical role in determining patient outcomes (Abousahmeen et al., 2023).

With regards to histological subtype, this study established invasive ductal carcinoma (IDC) to be the most prevalent. It accounted for 84.2% of ER+ breast cancer cases. This finding is consistent with the report by Sayeed et al. (2014) (Sayed et al., 2014). Similarly, studies conducted in Western Kenya and Eastern Morocco have also identified IDC as the predominant histological subtype (Errahhali et al., 2017; Sawe et al., 2016).

The predominance of invasive ductal carcinoma (IDC) among ER+ breast cancer cases in this cohort aligns with established global patterns, as demonstrated in studies from North Africa (Errahhali et al., 2017), East Africa (Sawe et al., 2016; Sayed et al., 2014). This consistent histological distribution across diverse populations reinforces the importance of prioritizing IDC-focused research in both basic and translational settings. Particularly for ER+ disease, understanding the interplay between ductal morphology and hormone receptor signaling may yield clinically actionable insights.

The high frequency of lymph node involvement observed in this study reinforces its established role as a critical prognostic indicator in breast cancer. This pattern aligns with findings from Kenya (Sayed et al., 2014), Libya (Abousahmeen et al., 2023), and Morocco (Errahhali et al., 2017), though with notable variations that likely reflect differences in tumor biology, stage at diagnosis, and healthcare access. The particularly high rates reported in these African studies - compared to global averages - may indicate regional challenges in early detection systems. Research from Eastern India further confirms that tumor characteristics including grade, stage and lymphovascular invasion remain universal predictors of nodal metastasis, regardless of geographical setting (Chakraborty et al., 2016). These collective findings highlight the need for both improved diagnostic infrastructure and tailored treatment approaches across diverse healthcare environments.

Research from Kenya, Uganda, and Libya consistently shows that breast cancer is often detected at advanced stages in these regions (Abousahmeen et al., 2023; Galukande et al., 2014; Sayed et al., 2014). This pattern likely results from two key factors: the types of breast cancer being studied and differences in healthcare systems. Studies that include aggressive subtypes like triple-negative breast cancer naturally report more late-stage diagnoses, as these tumors progress rapidly. In contrast, ER-positive breast cancers—which tend to grow more slowly—are more likely to be caught earlier when screening is available (Nakashima et al., 2019). Healthcare access plays an equally important role; regions with established screening programs and greater public awareness typically detect cancers at earlier, more treatable stages (Hoveling et al., 2025). Together, these biological

and systemic factors explain why diagnosis timing varies significantly across different populations.

The predominance of luminal A breast cancer observed in this study aligns with findings from other studies, including research from Kenya (Sayed et al., 2014) and Uganda (Galukande et al., 2014). However, significant regional variations exist, as demonstrated by a Libyan study reporting luminal B as the most frequent subtype (Abousahmeen et al., 2023). A recent systematic review of African breast cancer phenotypes noted that luminal subtypes collectively represent the most common classification across the continent, with Kenya showing particularly high frequencies of luminal A tumors (Onyia et al., 2023).

These discrepancies in subtype distribution may reflect both true biological differences and methodological challenges in molecular classification. Historically, inconsistent biomarker panels and varying laboratory protocols have complicated reliable distinction between luminal A and B subtypes in many settings. This technical variability underscores the need for standardized molecular profiling approaches in breast cancer research, particularly for hormone receptor-positive disease where accurate subtyping carries significant therapeutic implications.

5.4 Associations of Socio-Demographics and Risk Factors with ER^+ Breast Cancer

While demographic characteristics provide a foundation for understanding disease patterns, identifying other factors associated with ER+ breast cancer is crucial for improving risk assessment and prevention strategies.

Existing literature consistently demonstrates that postmenopausal status significantly elevates risk for hormone receptor-positive breast cancer, primarily through prolonged estrogen exposure and altered hormonal metabolism (Johansson et al., 2021; Rinaldi et al., 2006). This biological mechanism is particularly relevant in ER+ tumors, where estrogen signaling plays a central role in tumorigenesis. This biological mechanism is central to ER+ tumorigenesis. In our study, this association was strongly supported.

Postmenopausal status was significantly more prevalent in ER+ breast cancer cases compared to the controls. The odds of an individual being a case was seven times that of control. This finding aligns with the established literature, reinforcing the role of menopausal status as a key risk determinant in our study population. Furthermore, it underscores the relevance of endocrine factors in the pathogenesis of ER+ breast cancer within the local context, consistent with global epidemiological patterns.

The relationship between benign breast disease (BBD) and subsequent cancer risk warrants particular attention. Proliferative breast lesions, especially those with atypia, are established risk factors (Alamri et al., 2020), yet the underlying mechanisms remain incompletely understood. The high prevalence of BBD compared to breast cancer (Hatim et al., 2017) underscores the importance of better characterizing these lesions for risk stratification.

These findings collectively emphasize the multifactorial nature of ER+ breast cancer etiology, where biological, behavioral, and healthcare access factors interact. Moreover, even though the participants were purposively selected, some participants suffered from other complications such as diabetes and hypertension that might have influenced the results.

5.5 Associations between Single Nucleotide Polymorphisms and ER^+ Breast Cancer

Research on estrogen-related gene variants reveals striking geographic and ethnic variations in breast cancer risk associations. The CYP1A1 rs4646903 variant demonstrates particularly divergent effects, showing risk associations in Korean and Mexican populations (Moreno-Galván et al., 2018; Shin et al., 2007), while exhibiting no significant effects in Iraqi populations (Ibrahem et al., 2021). Similarly, large-scale analyses of Caucasian and Chinese groups found no overall risk associations (Economopoulos & Sergentanis, 2010), highlighting how genetic background modifies variant effects. These population-specific patterns likely stem from differences in allele frequencies, local environmental exposures, and variations in linkage disequilibrium

patterns across ethnic groups. In the present study we observed a significant protective association of the alternative allele C of the rs4646903. There was a reduced odds of individuals association with ER+ breast cancer risk by 56%. Mechanistically, this may be explained by the enzyme's dual function; CYP1A1 contributes to breast cancer risk through a complex balance between the activation of procarcinogens and the detoxification of harmful compounds (Androutsopoulos et al., 2009). In our cohort, the variant may favor detoxification pathways or interact with local environmental factors in a way that attenuates risk, aligning it with the subset of populations where non-risk or protective effects have been suggested.

The influence of rs1056836 on breast cancer risk demonstrates striking variability across populations and disease stages. While this variant has been associated with increased risk in Iranian (Adab et al., 2017), Nigerian (Okobia et al., 2009), and Han Chinese populations (Jiao et al., 2010), protective effects during cancer initiation were observed in Chinese women (Qiu et al., 2018), with neutral effects reported in Southern Indian (Francis et al., 2019), Iraqi (Ibrahem et al., 2021), and Egyptian cohorts (El-Moselhy, 2017). In the present study, the CYP1B1 rs1056836 polymorphism exhibited distinct associations depending on the control group. While it showed no significant association when cases were compared to healthy controls, it demonstrated a protective effect against ER+ breast cancer when cases were compared to the Benign Breast Disease (BBD) group. This suggests that rs1056836 may influence later stages of carcinogenesis, potentially affecting the progression from benign proliferative states to invasive malignancy rather than initial cancer development. Such stage-specific influences could explain the variability across studies, as the relative importance of genetic factors may differ between initial transformation and malignant progression. These patterns highlight the complex interplay between genetic background, disease stage, and susceptibility mechanisms. Further research is needed to clarify whether these population differences reflect true biological variation or methodological factors in study design.

The current understanding of rs1056827's role in breast cancer risk remains inconclusive, with existing literature suggesting potential modulation of estrogen metabolism that may

influence carcinogenesis. The variant's effects appear to be highly context-dependent, with studies reporting varying associations across different populations. This variability extends to other CYP1A1 and CYP1B1 polymorphisms, where ethnic differences in allele frequencies, lifestyle factors, and environmental exposures may collectively contribute to divergent risk profiles. This study observed a protective association of the ER+ breast cancer compared to the BBDs.

The potential for gene-environment interactions further complicates risk prediction, particularly regarding exposures to endocrine-disrupting compounds or hormonal therapies that may differentially affect individuals based on their genetic background. Such interactions could explain why certain variants demonstrate protective effects in some populations but neutral or even risk-enhancing effects in others. These observations underscore the importance of considering both genetic and environmental contexts when evaluating breast cancer susceptibility markers.

5.6 Associations between Single Nucleotide Polymorphisms and Tumor Characteristics

This study identified significant associations between the rs4646903 variant and more aggressive ER+ breast cancer characteristics, specifically Luminal B molecular subtype and late-stage disease. These findings contrast with reports from Iraqi populations where no significant associations were observed for these variants (Ibrahem et al., 2021). The connection between rs4646903 and advanced disease in this study aligns with documented relationships to high-grade tumors in Southern Indian women (Francis et al., 2019), while differing from patterns seen in Brazilian cohorts where certain variants occurred more frequently in low-grade (I-II) versus high-grade (III) tumors (Martins de Oliveira et al., 2015).

The population-specific nature of these genetic effects may reflect variations in allele frequencies or environmental exposures. This study found no significant associations for rs1048943 or rs1056836 with tumor characteristics, suggesting variant-specific rather

than generalized effects across estrogen metabolism genes. The consistent observation of rs4646903 associations with aggressive disease features across multiple studies warrants further investigation into its potential role in tumor progression pathways. Additional research is needed to elucidate the biological mechanisms underlying these population-dependent patterns and their clinical relevance for risk stratification.

5.7 In-Silico Analysis of CYP1B1 Haplotypes

5.7.1 Linkage Disequilibrium

The analysis of CYP1B1 haplotypes provides valuable insights into ER+ breast cancer susceptibility patterns. The strong linkage between rs10012 and rs1056827 suggests these variants are typically inherited together and may function as a combined genetic unit, while rs1056836 appears to segregate independently. These findings align with previous cancer research demonstrating that haplotype effects often differ from individual variant associations, (Trubicka et al., 2010; Yu et al., 2015). The moderate r^2 value of 0.5767 suggests that while these two SNPs are strongly linked, some degree of historical recombination has occurred, which is typical for genes with moderate recombination rates (Ardlie et al., 2002). The independent segregation of rs1056836 (L432V) aligns with its location approximately 1.3 kb downstream, outside the core haplotype block, indicating that this variant may have arisen on a different evolutionary background (Hanna et al., 2000).

5.7.2 Haplotype Distribution and Breast Cancer Risk

The identification of eight distinct haplotypes across the study groups reflects the genetic diversity of CYP1B1 in the population studied. The wild-type haplotype (G–C–G–T) characterized by the combination of R48 (from rs10012), A119 (from rs1056827), L432 (from rs1056836) and N453 (from rs1800440) was notably absent in controls but present at 5.0% in cases, suggesting that individuals carrying this haplotype may have been underrepresented in the control group due to selection or sampling bias. Alternatively, this

distribution may indicate that the wild-type haplotype is not necessarily protective, a finding that contrasts with some previous studies where the wild-type allele has been associated with reduced risk (Sissung et al., 2006).

The risk-associated haplotype (G–C–C–T) is characterized by the combination of R48 (from rs10012), A119 (from rs1056827), V432 (from rs1056836) and N453 (from rs1800440) demonstrated a progressive increase in frequency from benign samples (3.8%) to controls (8.3%) to cases (12.5%), consistent with a dose-dependent effect on breast cancer risk. Notably, the V432 variant has been previously associated with increased catalytic activity toward estradiol, leading to elevated 4-hydroxyestradiol production, which is a known genotoxic metabolite (Hanna et al., 2000). The combination of this high-activity variant with R48 and A119 may further modulate enzyme function through altered membrane association and protein stability (Banday et al., 2014).

5.7.3 Functional Domain Mapping of Polymorphic Residues

The domain mapping analysis placed the four polymorphic residues within functionally critical regions of CYP1B1, providing a structural rationale for their potential impact on enzyme activity. R48 is located in the hinge region between the N-terminal membrane anchor and the globular catalytic domain. This position is critical for proper membrane insertion and orientation, and the R48G substitution may alter membrane binding affinity, potentially affecting subcellular localization (Sissung et al., 2006). The A119S substitution resides in a structural core region; while alanine to serine is a conservative substitution, the introduction of a hydroxyl group at this position could subtly alter hydrogen bonding patterns within the protein core, affecting overall stability (Banday et al., 2014). The R48G and A119S substitutions located near the hinge region of CYP1B1 are not known to directly impact CYP1B1 catalytic activity when assessed in isolation; rather, their functional significance is likely mediated through synergistic interactions with other polymorphic residues within the haplotype structure (Sissung et al., 2006)

L432 is positioned near the substrate-binding pocket, a region that directly influences substrate specificity and catalytic efficiency. The L432V substitution has been extensively studied and is known to alter the regiospecificity of estradiol hydroxylation, shifting the balance toward the genotoxic 4-hydroxyestradiol pathway (Hanna et al., 2000). This substitution may also affect the binding of other endogenous substrates and procarcinogens(T. Shimada et al., 1999). The L432V and N453 substitutions encodes the heme-binding domain, the region critical to the catalytic function of CYP1B1 (Bailey et al., 1998). N453 is located in the C-terminal region near the heme-binding domain. While the N453S substitution is relatively conservative, its proximity to the heme prosthetic group suggests it could influence electron transfer efficiency from cytochrome P450 reductase, thereby modulating overall catalytic turnover (Gajjar et al., 2012).

5.7.4 Tertiary Structure Analysis

The generation of high-confidence three-dimensional models for the four haplotypes using AlphaFold Server represents a significant advancement in our ability to visualize the structural consequences of these polymorphisms. The AlphaFold 3 model demonstrated excellent prediction quality, as evidenced by the high pLDDT scores (>90) for the ordered regions of all models, consistent with the model's strong performance on single-chain protein predictions (Abramson et al., 2024). The Ramachandran plot analysis confirmed the stereochemical quality of all models, with >90% of residues in favored regions, validating their suitability for downstream molecular dynamics simulations.

Visual inspection of the ribbon structures and surface models revealed that the four polymorphic residues are distributed across distinct structural domains, with R48 located in an extended loop region, A119 and L432 buried within the hydrophobic core and substrate-binding pocket respectively, and N453 positioned near the C-terminal helix in proximity to the heme-binding site. The surface models suggest that while R48 and A119 are relatively solvent-exposed, L432 and N453 are more buried, consistent with their predicted roles in substrate binding and heme coordination (Sissung et al., 2006)

The colour-coded ribbon diagrams allowed for direct comparison of the local environments surrounding each polymorphic residue. The wild-type haplotype (R–A–L–N) exhibited the expected distribution of charged and hydrophobic residues, with R48 potentially forming electrostatic interactions with membrane phospholipids (McLellan et al., 2000). In the reference haplotype (G–S–V–N), the R48G substitution removes a positive charge, potentially reducing membrane affinity, while the A119S substitution introduces a potential hydrogen bond donor in the structural core. The risk-associated haplotype (G–A–V–N) combines the loss of membrane interaction (R48G) with the highly active V432 variant, potentially creating a "gain-of-function" phenotype through enhanced substrate access and reduced membrane tethering (Banday et al., 2014). The protective haplotype (R–S–V–N) retains the wild-type R48 residue, which may maintain proper membrane localization, while incorporating both S119 and V432, suggesting that the R48 residue may be critical for maintaining the protective phenotype despite other substitutions (Sissung et al., 2006).

The tertiary structure analysis revealed that the protective haplotype (R–S–V–N) maintains the wild-type R48 residue while incorporating S119 and V432. This suggests that the presence of R48 may be critical for proper membrane anchoring, potentially outweighing any deleterious effects of the other substitutions. Conversely, the risk-associated haplotype (G–A–V–N) combines the loss of R48 with the high-activity V432 variant, which may synergistically enhance carcinogenic estrogen metabolism while reducing proper subcellular localization (Hanna et al., 2000). These structural observations provide a mechanistic framework for understanding the differential breast cancer risk associated with these haplotypes.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This study established that:

1. ER+ breast cancer cases predominantly affect older, postmenopausal women and presents primarily as invasive ductal carcinomas (76.7%) with Luminal A features (82.4%), frequently at Stage II (45.3%) with moderate differentiation (Grade II: 62.5%) and lymph node involvement (56.2%).
2. The study revealed low odds of CYP1A1 rs4646903-C allele against overall ER+ cancer development (OR=0.44, p=0.048) and CYP1B1 rs1056836-C (OR=0.34, p=0.0003) and rs1056827-A (OR=0.43, p=0.045) against malignant transformation from benign lesions. On the other hand, G-A-C haplotype (rs10012-G/rs1056827-A/rs1056836-C) conferred 4.75-fold increased ER+ cancer risk (p=0.046)
3. The rs4646903-C allele was associated with Luminal B subtypes (OR=3.83) and late tumor stage
4. Tertiary structure modeling revealed that the four polymorphic residues occupy functionally critical domains: R48 in the membrane interaction region, A119 in the structural core, L432 near the substrate-binding pocket, and N453 in the heme-binding region.

6.2 Recommendations

Based on the findings, the following recommendations are proposed:

1. Develop targeted screening and educational policies for postmenopausal women and those with a family history of breast cancer.

2. Investigate the mechanistic basis of rs4646903's dual role—protective against ER+ breast cancer yet associated with aggressive Luminal B subtypes—to determine its clinical utility.
3. Perform molecular dynamics simulations to elucidate the dynamic conformational changes, active site flexibility, and substrate binding affinities of the four CYP1B1 haplotypes (R–A–L–N, G–S–V–N, G–A–V–N, R–S–V–N).
4. Develop population-specific risk assessment tools integrating demographic risk factors with validated genetic markers and haplotype structural information.

REFERENCES

Abdelmonem, B. H., Abdelaal, N. M., Anwer, E. K. E., Rashwan, A. A., Hussein, M. A., Ahmed, Y. F., Khashana, R., Hanna, M. M., & Abdelnaser, A. (2024). Decoding the Role of CYP450 Enzymes in Metabolism and Disease: A Comprehensive Review. *Biomedicines*, *12*(7). https://doi.org/10.3390/biomedicines12071467

Abousahmeen, A., Saud, M. A. Ben, Elrais, S., Kashbour, M., Abuhlaiga, M., Al-Aqmar, D. M., & Al-Shareef, J. M. (2023). Molecular subtypes and clinicopathological features of breast cancer in Libya: a first glance. *Therapeutic Radiology and Oncology*, *7*(15). https://doi.org/10.21037/tro-23-1

Adab, F. K., Fard, Z. T., & Akbari, M. E. (2017). Association between cytochrome 1B1*3 polymorphism and the breast cancer in a group of Iranian women. *International Journal of Cancer Management*, *10*(1). https://doi.org/10.17795/ijcp-6428

Alamri, A. M., Alsareii, S. A., Al-Wadei, H., Al-Qahtani, A. M., Sultan, S. A. A., Alshamrani, S. A., Almakrami, A. H., Daiel, A. A., Alyami, A. Y., Hommadi, A. M., & Ali, Y. M. T. (2020). Epidemiological Pattern of Breast Diseases among Females in the South-Western Region, Saudi Arabia. *International Journal of Clinical Medicine*, *11*(05), 257–269. https://doi.org/10.4236/ijcm.2020.115027

Alcantara, D., Leal, M. P., García-Bocanegra, I., & García-Martín, M. L. (2014). Molecular imaging of breast cancer: Present and future directions. *Frontiers in Chemistry*, *2*(112), 1–11. https://doi.org/10.3389/fchem.2014.00112

Allred, D. C. (2010). Ductal carcinoma in situ: Terminology, classification, and natural history. *Journal of the National Cancer Institute - Monographs*, (41), 134–138. https://doi.org/10.1093/jncimonographs/lgq035

Almeida, M., Soares, M., Fonseca-Moutinho, J., Ramalhinho, A. C., & Breitenfeld, L. (2021). Influence of estrogenic metabolic pathway genes polymorphisms on

postmenopausal breast cancer risk. *Pharmaceuticals*, *14*(2), 1–9. https://doi.org/10.3390/ph14020094

Al-Shami, K., Awadi, S., Khamees, A., Alsheikh, A. M., Al-Sharif, S., Ala' Bereshy, R., Al-Eitan, S. F., Banikhaled, S. H., Al-Qudimat, A. R., Al-Zoubi, R. M., & Al Zoubi, M. S. (2023). Estrogens and the risk of breast cancer: A narrative review of literature. *Heliyon*, *9*(9). https://doi.org/10.1016/j.heliyon.2023.e20224

Androutsopoulos, V. P., Tsatsakis, A. M., & Spandidos, D. A. (2009). Cytochrome P450 CYP1A1: Wider roles in cancer progression and prevention. In *BMC Cancer* (Vol. 9). https://doi.org/10.1186/1471-2407-9-187

Anyigba, C. A., Awandare, G. A., & Paemka, L. (2021). Breast cancer in sub-Saharan Africa: The current state and uncertain future. *Experimental Biology and Medicine*, *246*(12), 1377–1387. https://doi.org/10.1177/15353702211006047

Ardlie, K. G., Kruglyak, L., & Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. In *Nature Reviews Genetics* (Vol. 3, Number 4, pp. 299–309). https://doi.org/10.1038/nrg777

Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J. R., Cardoso, F., Siesling, S., & Soerjomataram, I. (2022). Current and future burden of breast cancer: Global statistics for 2020 and 2040. *Breast*, *66*, 15–23. https://doi.org/10.1016/j.breast.2022.08.010

Badavi, E., Safavi, B., Jalali, A., Shahriary, G. M., Mohammadi-Asl, J., & Babaei, J. (2015). Association of CYP3A4 and CYP3A5 polymorphisms with Iranian breast cancer patients. *Egyptian Journal of Medical Human Genetics*, *16*(3), 219–225. https://doi.org/10.1016/j.ejmhg.2015.03.004

Bailey, R. L., Rood, N., Dupont, W. D., & Pari2, F. F. (1998). Association of Cytochrome P450 IBI (CYPIBI} Polymorphism with Steroid Receptor Status in Breast Cancer1.

Cancer Research, (58), 5038–5041. http://aacrjournals.org/cancerres/article-pdf/58/22/5038/2468165/cr0580225038.pdf

Benz, C. C. (2008). Impact of aging on the biology of breast cancer. *Critical Reviews in Oncology/Hematology*, *66*(1), 65–74. https://doi.org/10.1016/j.critrevonc.2007.09.001

Black, E., & Richmond, R. (2019). Improving early detection of breast cancer in sub-Saharan Africa: Why mammography may not be the way forward. *Globalization and Health*, *15*(1). https://doi.org/10.1186/s12992-018-0446-6

Bosompem, K., Yorke, J., Buckman, T. A., Brenu, S. G., Nyantakyi, M., Aitpillah, F. S. K., Kyei, I., Adinku, M. O., Yorke, D. A., Obirikorang, C., & Acheampong, E. (2024). Comparative analysis of breast cancer characteristics in young premenopausal and postmenopausal women in Ghana. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-52129-w

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *74*(3), 229–263. https://doi.org/10.3322/caac.21834

Brewer, H. R., Jones, M. E., Schoemaker, M. J., Ashworth, A., & Swerdlow, A. J. (2017). Family history and risk of breast cancer: an analysis accounting for family structure. *Breast Cancer Research and Treatment*, *165*(1), 193–200. https://doi.org/10.1007/s10549-017-4325-2

Brufsky, A. M., & Dickler, M. N. (2018). Estrogen Receptor-Positive Breast Cancer: Exploiting Signaling Pathways Implicated in Endocrine Resistance. *The Oncologist*, *23*(5), 528–539. https://doi.org/10.1634/theoncologist.2017-0423

Cai, Y., Dai, F., Ye, Y., & Qian, J. (2025). The global burden of breast cancer among women of reproductive age: a comprehensive analysis. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-93883-9

Cassidenti, D. L., Vijod, A. G., Vijod, M. A., Stanczyk, F. Z., & Lobo, R. A. (1990). Short-term effects of smoking on the pharmacokinetic profiles of micronized estradiol in postmenopausal women. *American Journal of Obstetrics and Gynecology*, *163*(6 PART 1), 1953–1960. https://doi.org/10.1016/0002-9378(90)90780-B

Cavalieri, E., Saeed, M., Zahid, M., Cassada, D., Snow, D., Miljkovic, M., & Rogan, E. (2012). Mechanism of DNA depurination by carcinogens in relation to cancer initiation. *IUBMB Life*, *64*(2), 169–179. https://doi.org/10.1002/iub.586

Çelik, A., Acar, M., Moroski Erkul, C., Gunduz, E., & Gunduz, M. (2015). Relationship of Breast Cancer with Ovarian Cancer. In *A Concise Review of Molecular Pathology of Breast Cancer*. InTech. https://doi.org/10.5772/59682

Chakraborty, A., Bose, C. K., Basak, J., Sen, A. N., Mishra, R., & Mukhopadhyay, A. (2016). Determinants of lymph node status in women with breast cancer: A hospital based study from eastern India. *Indian Journal of Medical Research, Supplement*, *143*, 45–51. https://doi.org/10.4103/0971-5916.191761

Chen, X., Wang, Q., Zhang, Y., Xie, Q., & Tan, X. (2019). Physical Activity and Risk of Breast Cancer: A Meta-Analysis of 38 Cohort Studies in 45 Study Reports. *Value in Health*, *22*(1), 104–128. https://doi.org/10.1016/j.jval.2018.06.020

Christgen, M., Cserni, G., Floris, G., Marchio, C., Djerroudi, L., Kreipe, H., Derksen, P. W. B., & Vincent-Salomon, A. (2021). Lobular breast cancer: Histomorphology and different concepts of a special spectrum of tumors. In *Cancers* (Vol. 13, Number 15). MDPI AG. https://doi.org/10.3390/cancers13153695

Cleary, M. P., & Grossmann, M. E. (2009). Minireview: Obesity and breast cancer: The estrogen connection. *Endocrinology*, *150*(6), 2537–2542. https://doi.org/10.1210/en.2009-0070

Collaborative Group on Hormonal Factors in Breast Cancer. (2001). Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. *The Lancet*, *358*(9291), 1389–1399. https://doi.org/10.1016/S0140-6736(01)06524-2

Coughlin, S. S., & Piper, M. (1999). Genetic Polymorphisms and Risk of Breast Cancer. *Cancer Epidemiology, Biomarkers & Prevention*, *8*, 1023–1032. http://aacrjournals.org/cebp/article-pdf/8/11/1023/3256051/ce119901023p.pdf

Coughlin, S. S., & Smith, S. A. (2015). The Impact of the Natural, Social, Built, and Policy Environments on Breast Cancer. *J Environ Health Sci*, *1*(3), 1–7. http://community.cochrane.org/cochrane-reviews

Diab, S. G., Elledge, R. M., & Clark, G. M. (2000). Tumor Characteristics and Clinical Outcome of Elderly Women With Breast Cancer. *Journal of the National Cancer Institute*, *5*(92), 550–556. https://doi.org/10.1093/jnci/92.7.550

Draznin, B. (2011). Mechanism of the mitogenic influence of hyperinsulinemia. *Diabetology and Metabolic Syndrome*, *3*(1). https://doi.org/10.1186/1758-5996-3-10

Duffy, M. J., Harbeck, N., Nap, M., Molina, R., Nicolini, A., Senkus, E., & Cardoso, F. (2017). Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). *European Journal of Cancer*, *75*, 284–298. https://doi.org/10.1016/j.ejca.2017.01.017

Dyrstad, S. W., Yan, Y., Fowler, A. M., & Colditz, G. A. (2015). Breast cancer risk associated with benign breast disease: systematic review and meta-analysis. *Breast*

Cancer Research and Treatment, *149*(3), 569–575. https://doi.org/10.1007/s10549-014-3254-6

Economopoulos, K. P., & Sergentanis, T. N. (2010). Three polymorphisms in cytochrome P450 1B1 (CYP1B1) gene and breast cancer risk: a meta-analysis. *Breast Cancer Research and Treatment*, *122*(2). https://doi.org/10.1007/s10549-009-0728-zï

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Ekpe, E., Jamal Shaikh, A., Shah, J., Jacobson, J. S., Sayed, S., & ChB, M. (2019). Metastatic Breast Cancer in Kenya: Presentation, Pathologic Characteristics, and Patterns-Findings from a Tertiary Cancer Center. *J Global Oncol*, *5*, 1–11. https://doi.org/10.

El-Moselhy, E. A. (2017). Female Breast Cancer: Socio-Demographic, Lifestyle, and Clinical Risk Factors - A Hospital-Based Study in Two Egyptian Governorates. *Biomedical Journal of Scientific & Technical Research*, *1*(4). https://doi.org/10.26717/bjstr.2017.01.000328

Errahhali, M. E., Errahhali, M. E., Ouarzane, M., El Harroudi, T., Afqir, S., & Bellaoui, M. (2017). First report on molecular breast cancer subtypes and their clinico-pathological characteristics in Eastern Morocco: Series of 2260 cases. *BMC Women's Health*, *17*(1). https://doi.org/10.1186/s12905-016-0361-z

Figueroa, J. D., Gierach, G. L., Duggan, M. A., Fan, S., Pfeiffer, R. M., Wang, Y., Falk, R. T., Loudig, O., Abubakar, M., Ginsberg, M., Kimes, T. M., Richert-Boe, K., Glass, A. G., & Rohan, T. E. (2021). Risk factors for breast cancer development by tumor characteristics among women with benign breast disease. *Breast Cancer Research*, *23*(1). https://doi.org/10.1186/s13058-021-01410-1

Folkerd, E., & Dowsett, M. (2013). Sex hormones and breast cancer risk and prognosis. *Breast*, *22*(S2). https://doi.org/10.1016/j.breast.2013.07.007

Francis, A. M., Ramya, R., Ganesan, N., Kumarasamy, P., Paul, S. F. D., Munirajan, A. K., & Divya, M. (2019). Breast cancer susceptibility genes in estrogen metabolizing pathway in a southern Indian population. *Meta Gene*, *19*, 225–234. https://doi.org/10.1016/j.mgene.2018.12.009

Freudenheim, J. L. (2020). Alcohol's effects on breast cancer in women. *Alcohol Research: Current Reviews*, *40*(2). https://doi.org/10.35946/arcr.v40.2.11

Gajjar, K., Owens, G., Sperrin, M., Martin-Hirsch, P. L., & Martin, F. L. (2012). Cytochrome P1B1 (CYP1B1) polymorphisms and ovarian cancer risk: A meta-analysis. *Toxicology*, *302*(2–3), 157–162. https://doi.org/10.1016/j.tox.2012.09.009

Galea, M. (2019). Benign breast disorders. *Surgery (United Kingdom)*, *37*(3), 151–156. https://doi.org/10.1016/j.mpsur.2019.01.003

Gallagher, E. J., Fierz, Y., Vijayakumar, A., Haddad, N., Yakar, S., & Leroith, D. (2012). Inhibiting PI3K reduces mammary tumor growth and induces hyperglycemia in a mouse model of insulin resistance and hyperinsulinemia. *Oncogene*, *31*(27), 3213–3222. https://doi.org/10.1038/onc.2011.495

Galukande, M., Wabinga, H., Mirembe, F., Karamagi, C., & Asea, A. (2014). Molecular breast cancer subtypes prevalence in an indigenous Sub Saharan African population. *The Pan African Medical Journal*, *17*, 249. https://doi.org/10.11604/pamj.2014.17.249.330

Giordano, S. H. (2018). Breast Cancer in Men. *New England Journal of Medicine*, *378*(24), 2311–2320. https://doi.org/10.1056/NEJMra1707939

Goldhirsch, A., Winer, E. P., Coates, A. S., Gelber, R. D., Piccart-Gebhart, M., Thürlimann, B., Senn, H. J., Albain, K. S., André, F., Bergh, J., Bonnefoi, H., Bretel-Morales, D., Burstein, H., Cardoso, F., Castiglione-Gertsch, M., Coates, A. S., Colleoni, M., Costa, A., Curigliano, G., … Wood, W. C. (2013). Personalizing the treatment of women with early breast cancer: Highlights of the st gallen international expert consensus on the primary therapy of early breast Cancer 2013. *Annals of Oncology*, *24*(9), 2206–2223. https://doi.org/10.1093/annonc/mdt303

Guray, M., & Sahin, A. A. (2006). Benign Breast Diseases: Classification, Diagnosis, and Management. *The Oncologist*, *11*(5), 435–449. https://doi.org/10.1634/theoncologist.11-5-435

Haber, G., Ahmed, N. U., & Pekovic, V. (2012). Family history of cancer and its association with breast cancer risk perception and repeat mammography. *American Journal of Public Health*, *102*(12), 2322–2329. https://doi.org/10.2105/AJPH.2012.300786

Hanna, I. H., Dawling, S., Roodi, N., Guengerich, F. P., & Parl, F. F. (2000). Cytochrome P450 1B1 (CYP1B1) Pharmacogenetics: Association of Polymorphisms with Functional Differences in Estrogen Hydroxylation Activity 1. In *CANCER RESEARCH* (Vol. 60).

Hasan, S., Azeez, & Rafiq, A. (2020). Immunological Status in Breast Cancer Women's in Erbil Province-Iraq. *P J M H S*, *14*(2), 924–929.

Hatim, K. S., Laxmikant, N. S., & Mulla, T. (2017). Patterns and prevalence of benign breast disease in Western India. *International Journal of Research in Medical Sciences*, *5*(2), 684. https://doi.org/10.18203/2320-6012.ijrms20170174

Hoveling, L. A., Heuken, L. P., Harfst, T., Schuurman, M. S., van Asselt, K. M., Siesling, S., & Bode, C. (2025). Understanding delays in breast cancer diagnosis in Africa:

Key insights and contributing factors. *International Journal of Cancer*, 1–11. https://doi.org/10.1002/ijc.70008

Hutchinson, B., Watts, R., Nyangasi, M., Anderson, B. O., Chepchumba, J., Wangia, E., Jalang'o, R., Mwenda, V., Yerramilli, P., Lee Kuguru, T., Kabubei, K. M., Gordillo-Tobar, A., Meheus, F., Meyer, C., Ilbawi, A., & Nugent, R. (2024). An economic evaluation of breast cancer interventions in Kenya. *EClinicalMedicine*, *77*(102894). https://doi.org/10.1016/j.eclinm.2024.102894

Ibrahem, S. Q., Ahmed, H. Q., & Amin, khalida M. (2021). Genetic Variations in Cytochrome P450 1A1 and 1B1 Genes in a Cohort of Patients from Iraq Diagnosed with Breast Cancer. *Breast Cancer (Auckl)*, *16*(15:11782234211050727). https://doi.org/10.1177/11782234211050727

Inic, Z., Zegarac, M., Inic, M., Markovic, I., Kozomara, Z., Djurisic, I., Inic, I., Pupic, G., & Jancic, S. (2014). Difference between Luminal A and Luminal B subtypes according to Ki-67, tumor size, and progesterone receptor negativity providing prognostic information. *Clinical Medicine Insights: Oncology*, *8*, 107–111. https://doi.org/10.4137/CMO.S18006

International Agency for Research on Cancer. (2004). *IARC monographs on the evaluation of carcinogenic risks to humans: tobacco smoke and involuntary smoking*. IARC Press. https://monographs.iarc.fr/wp-content/uploads/2018/06/mono83.pdf

International Agency for Research on Cancer. (2022). *Global Cancer Observatory: Cancer Today – Kenya*. https://gco.iarc.who.int/media/globocan/factsheets/populations/404-kenya-fact-sheet.pdf

Ishii, K., Morii, N., & Yamashiro, H. (2019). Pertuzumab in the treatment of HER2-positive breast cancer: an evidence-based review of its safety, efficacy, and place in therapy . *Core Evidence*, *Volume 14*, 51–70. https://doi.org/10.2147/ce.s217848

Jafari, S. H., Saadatpour, Z., Salmaninejad, A., Momeni, F., Mokhtari, M., Nahand, J. S., Rahmati, M., Mirzaei, H., & Kianmehr, M. (2018). Breast cancer diagnosis: Imaging techniques and biochemical markers. *Journal of Cellular Physiology*, *233*(7), 5200–5213. https://doi.org/10.1002/jcp.26379

Jara, L., Morales, S., de Mayo, T., Gonzalez-Hormazabal, P., Carrasco, V., & Godoy, R. (2017). Mutations in BRCA1, BRCA2 and other breast and ovarian cancer susceptibility genes in Central and South American populations. In *Biological Research* (Vol. 50, Number 1). BioMed Central Ltd. https://doi.org/10.1186/s40659-017-0139-2

Jiao, H., Liu, C., guo, W., Peng, L., chen, Y., & Martin, F. L. (2010). Association of CYP1B1 Polymorphisms with Breast Cancer: A Case-Control Study in the Han Population in Ningxia Hui Autonomous Region, P.R. China. *Biomarker Insights*, *5*, 21–27. http://www.la-press.com.http://www.la-press.com.

Johansson, A., Christakou, A. E., Iftimi, A., Eriksson, M., Tapia, J., Skoog, L., Benz, C. C., Rodriguez-Wallberg, K. A., Hall, P., Czene, K., & Lindström, L. S. (2021). Characterization of Benign Breast Diseases and Association with Age, Hormonal Factors, and Family History of Breast Cancer among Women in Sweden. *JAMA Network Open*, *4*(6). https://doi.org/10.1001/jamanetworkopen.2021.14716

Joko-Fru, W. Y., Jedy-Agba, E., Korir, A., Ogunbiyi, O., Dzamalala, C. P., Chokunonga, E., Wabinga, H., Manraj, S., Finesse, A., Somdyala, N., Liu, B., McGale, P., Jemal, A., Bray, F., & Parkin, D. M. (2020). The evolving epidemic of breast cancer in sub-Saharan Africa: Results from the African Cancer Registry Network. *International Journal of Cancer*, *147*(8), 2131–2141. https://doi.org/10.1002/ijc.33014

Jones, K. L., & Buzdar, A. U. (2004). A review of adjuvant hormonal therapy in breast cancer. *Endocrine-Related Cancer*, *11*(3), 391–406. https://doi.org/10.1677/erc.1.00594

Kagiri, H., Kiarie, H., & Karagu, A. (2020). Kenya Policy Brief 2020 - Assessment of demand versus supply of cancer services on Kenya. In *BMC Cancer* (Vol. 14, Number 1). BioMed Central Ltd. https://doi.org/10.1186/1471-2407-14-224

Keelan, S., Flanagan, M., & Hill, A. D. K. (2021). Evolving Trends in Surgical Management of Breast Cancer: An Analysis of 30 Years of Practice Changing Papers. *Frontiers in Oncology*, *11*(622621). https://doi.org/10.3389/fonc.2021.622621

Kelsey, J. L., Whittemore, A., Evans, A. S., & Thompson, W. D. (1996). *Methods in observational epidemiology*. https://global.oup.com/academic/product/methods-in-observational-epidemiology-9780195083774#.W0TlTeiT44Y.mendeley

Khabaz, M. N. (2014). Immunohistochemistry subtypes (ER/PR/HER) of breast cancer: Where do we stand in the west of Saudi Arabia? *Asian Pacific Journal of Cancer Prevention*, *15*(19), 8395–8400. https://doi.org/10.7314/APJCP.2014.15.19.8395

Kilfoy, B. A., Zhang, Y., Shu, X. O., Gao, Y. T., Ji, B. T., Yang, G., Li, H. L., Rothman, N., Chow, W. H., & Zheng, W. (2008). Family history of malignancies and risk of breast cancer: Prospective data from the Shanghai women's health study. *Cancer Causes and Control*, *19*(10), 1139–1145. https://doi.org/10.1007/s10552-008-9181-2

Koboldt, D. C., Fulton, R. S., McLellan, M. D., Schmidt, H., Kalicki-Veizer, J., McMichael, J. F., Fulton, L. L., Dooling, D. J., Ding, L., Mardis, E. R., Wilson, R. K., Ally, A., Balasundaram, M., Butterfield, Y. S. N., Carlsen, R., Carter, C., Chu, A., Chuah, E., Chun, H. J. E., … Palchik, J. D. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, *490*(7418), 61–70. https://doi.org/10.1038/nature11412

Lachenmeier, D. W., Lima, M. C., Nóbrega, I. C., Pereira, J. A., Kerr-Corrêa, F., Kanteres, F., & Rehm, J. (2010). Cancer risk assessment of ethyl carbamate in alcoholic

beverages from Brazil with special consideration to the spirits cachaça and tiquira. *BMC Cancer*, *10*(266). http://www.biomedcentral.com/1471-2407/10/266

Lachenmeier, D. W., Przybylski, M. C., & Rehm, J. (2012). Comparative risk assessment of carcinogens in alcoholic beverages using the margin of exposure approach. *International Journal of Cancer*, *131*(6). https://doi.org/10.1002/ijc.27553

Leong, S. P. L., Shen, Z. Z., Liu, T. J., Agarwal, G., Tajima, T., Paik, N. S., Sandelin, K., Derossis, A., Cody, H., & Foulkes, W. D. (2010). Is Breast cancer the same disease in Asian and Western countries? *World Journal of Surgery*, *34*(10), 2308–2324. https://doi.org/10.1007/s00268-010-0683-1

Liao, M.-N., Chen, M.-F., Chen, S.-C., & Chen, P.-L. (2008). Uncertainty and Anxiety During the Diagnostic Period for Women With Suspected Breast Cancer. *Cancer Nursing TM*, *31*(4). https://doi.org/10.1097/01.NCC.0000305744.64452.fe.

Liu, Y., Nguyen, N., & Colditz, G. A. (2015). Links between alcohol consumption and breast cancer: A look at the evidence. *Women's Health*, *11*(1), 65–77. https://doi.org/10.2217/whe.14.62

Lolodi, O., Wang, Y.-M., Wright, W. C., & Chen, T. (2017). Differential Regulation of CYP3A4 and CYP3A5 and its Implication in Drug Discovery. *Current Drug Metabolism*, *18*(12), 1095–1105. https://doi.org/10.2174/1389200218666170531112038

Łukasiewicz, S., Czeczelewski, M., Forma, A., Baj, J., Sitarz, R., & Stanisławek, A. (2021). Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—An updated review. *Cancers*, *13*(17). https://doi.org/10.3390/cancers13174287

Makanga, W., Wasike, R., & Saidi, H. (2013). A Profile of Female Breast Cancer Patients in a Kenyan Urban Private Hospital. *The Annals of African Surgery*, *10*(1). https://doi.org/https://doi.org/10.4314/AAS.V10I1

Makki, J. (2015). Diversity of breast carcinoma: Histological subtypes and clinical relevance. *Clinical Medicine Insights: Pathology*, *8*(1), 23–31. https://doi.org/10.4137/CPath.s31563

Martins de Oliveira, C. B., Cardoso-Filho, C., Bossi, L. S., Lourenço, G. J., Costa-Gurgel, M. S., & Passos Lima, C. S. (2015). Association of CYP1A1 A4889G and T6235C polymorphisms with the risk of sporadic breast cancer in Brazilian women. *Clinics*, *70*(10), 680–685. https://doi.org/10.6061/clinics/2015(10)04

McDonald, J. A., Goyal, A., & Terry, M. B. (2013). Alcohol intake and breast cancer risk: Weighing the overall evidence. *Current Breast Cancer Reports*, *5*(3), 208–221. https://doi.org/10.1007/s12609-013-0114-z

McGuire, A., Brown, J. A. L., Malone, C., McLaughlin, R., & Kerin, M. J. (2015). Effects of age on the detection and management of breast cancer. *Cancers*, *7*(2), 908–929. https://doi.org/10.3390/cancers7020815

Ministry of Health. (2017). *National Cancer Control Strategy 2017-2022*.

Ministry of Health. (2020). *Kenya Cancer Policy 2019-2030*. https://doi.org/https://arua-ncd.org/wp-content/uploads/2022/10/Kenya-Cancer-Policy-2020.pdf

Ministry of Health. (2021). *Breast Cancer Screening and Early Diagnosis Action Plan 2021-2025*.

Ministry of Health. (2022). *National Cancer Taskforce Report*.

Moreno-Galván, M., Herrera-González, N. E., Robles-Pérez, V., Velasco-Rodríguez, J. C., Tapia-Conyer, R., & Sarti, E. (2018). Impact of Cyp1A1 and Comt Genotypes on

Breast Cancer Risk in Mexican Women: A Pilot Study. *The International Journal of Biological Markers*, *25*(3), 157–163. https://doi.org/10.1177/172460081002500306.

Murphy, C. G., & Dickler, M. N. (2016). Endocrine resistance in hormone-responsive breast cancer: Mechanisms and therapeutic strategies. *Endocrine-Related Cancer*, *23*(8), R337–R352. https://doi.org/10.1530/ERC-16-0121

Nakashima, K., Uematsu, T., Takahashi, K., Nishimura, S., Tadokoro, Y., Hayashi, T., & Sugino, T. (2019). Does breast cancer growth rate really depend on tumor subtype? Measurement of tumor doubling time using serial ultrasonography between diagnosis and surgery. *Breast Cancer*, *26*(2), 206–214. https://doi.org/10.1007/s12282-018-0914-0

Nascimento, R. G. do, & Otoni, K. M. (2020). Histological and molecular classification of breast cancer: what do we know? *Mastology*, *30*(e20200024). https://doi.org/10.29289/25945394202020200024

Ndetei, D., Musyimi, C., Tele, A., Musau, A., Gitonga, I., & Mutiso, V. (2018). Psychological Well-Being and Social Functioning Across the Cancer Stages: Implications for Palliative Care. *Journal of Social Work in End-of-Life and Palliative Care*, *14*(2–3), 194–208. https://doi.org/10.1080/15524256.2018.1472718

Nebert, D. W., & Dalton, T. P. (2006). The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. *Nature Reviews Cancer*, *6*(12), 947–960. https://doi.org/10.1038/nrc2015

Okobia, M. N., Bunker, C. H., Garte, S. J., Zmuda, J. M., Ezeome, E. R., Anyanwu, S. N. C., Uche, E. E. O., Osime, U., Ojukwu, J., Kuller, L. H., Ferrell, R. E., & Taioli, E. (2009). Cytochrome P450 1B1 Val432Leu polymorphism and breast cancer risk in Nigerian women: A case control study. *Infectious Agents and Cancer*, *4*(SUPPL. 1). https://doi.org/10.1186/1750-9378-4-S1-S12

Omoleye, O. J., Freeman, J. Q., Oluwasanu, M., Adeniji-Sofoluwe, A., Woodard, A. E., Aribisala, B. S., Adejumo, P. O., Ntekim, A., Makumbi, T., Ndom, P., Ajayi, I. O., Olopade, O. I., & Huo, D. (2023). Benign breast disease and breast cancer risk in African women: A case-control study. *Research Square*. https://doi.org/10.21203/rs.3.rs-3301977/v1

Onyia, A. F., Nana, T. A., Adewale, E. A., Adebesin, A. O., Adegboye, B. E., Paimo, O. K., De Campos, O. C., Bisi-Adeniyi, T. I., Rotimi, O. A., Oyelade, J. O., & Rotimi, S. O. (2023). Breast Cancer Phenotypes in Africa: A Scoping Review and Meta-Analysis. *JCO Global Oncology*, (9). https://doi.org/10.1200/go.23.00135

Özdemir, B. C., Sflomos, G., & Brisken, C. (2018). The challenges of modeling hormone receptor-positive breast cancer in mice. *Endocrine-Related Cancer*, *25*(5), R319–R330. https://doi.org/10.1530/ERC-18-0063

Parenté, A., Gnangnon, F. H. R., Kinsou, D., Chkeir, M., Adou, C., Robin Sacca, H., Ivanga, M., Ngoungou, E., Darré, T. P., Houinato, D., Preux, P.-M., & Blanquet, V. (2025). Prognostic factors of breast cancer and survival in Sub-Saharan Africa: A systematic review and meta-analysis. *Bulletin Du Cancer*. https://doi.org/10.1016/j.bulcan.2025.05.009

Peñalver-Argüeso, B., García-Esquinas, E., Castelló, A., Fernández de Larrea-Baz, N., Castaño-Vinyals, G., Amiano, P., Fernández-Villa, T., Guevara, M., Fernández-Tardón, G., Alguacil, J., Obón-Santacana, M., Gómez-Acebo, I., Pinto-Carbó, M., Marcos-Gragera, R., Aragonés, N., Aizpurua, A., Martín-Sánchez, V., Ardanaz, E., Dierssen-Sotos, T., … Pérez-Gómez, B. (2023). Smoking history and breast cancer risk by pathological subtype: MCC-Spain study. *Tobacco Induced Diseases*, *21*(157), 1–16. https://doi.org/10.18332/tid/174132

Pereslucha, A. M., Wenger, D. M., Morris, M. F., & Aydi, Z. B. (2023). Invasive Lobular Carcinoma: A Review of Imaging Modalities with Special Focus on Pathology

Concordance. In *Healthcare (Switzerland)* (Vol. 11, Number 5). MDPI. https://doi.org/10.3390/healthcare11050746

Plasilova, M. L., Hayse, B., Killelea, B. K., Horowitz, N. R., Chagpar, A. B., & Lannin, D. R. (2016). Features of triple-negative breast cancer Analysis of 38,813 cases from the national cancer database. *Medicine (Baltimore)*, *95*(35). https://doi.org/10.1097/MD.0000000000004614

Prat, A., Cheang, M. C. U., Martín, M., Parker, J. S., Carrasco, E., Caballero, R., Tyldesley, S., Gelmon, K., Bernard, P. S., Nielsen, T. O., & Perou, C. M. (2013). Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal a breast cancer. *Journal of Clinical Oncology*, *31*(2), 203–209. https://doi.org/10.1200/JCO.2012.43.4134

Prat, A., & Perou, C. M. (2011). Deconstructing the molecular portraits of breast cancer. *Molecular Oncology*, *5*(1), 5–23. https://doi.org/10.1016/j.molonc.2010.11.003

Qin, X., Peng, Q., Qin, A., Chen, Z., Lin, L., Deng, Y., Xie, L., Xu, J., Li, H., Li, T., Li, S., & Zhao, J. (2012). Association of COMT Val158Met polymorphism and breast cancer risk: An updated meta-analysis. *Diagnostic Pathology*, *7*(1). https://doi.org/10.1186/1746-1596-7-136

Qiu, J., Du, Z., Liu, J., Zhou, Y., Liang, F., & Lü, Q. (2018). Association between polymorphisms in estrogen metabolism genes and breast cancer development in Chinese women A prospective case-control study. *Medicine (United States)*, *97*(47). https://doi.org/10.1097/MD.0000000000013337

Rasha, F., Ramalingam, L., Gollahon, L., Rahman, R. L., Rahman, S. M., Menikdiwela, K., & Moustaid-Moussa, N. (2019). Mechanisms linking the renin-angiotensin system, obesity, and breast cancer. *Endocrine-Related Cancer*, *26*(12), R653–R672. https://doi.org/10.1530/ERC-19-0314

Rinaldi, S., Key, T. J., Peeters, P. H. M., Lahmann, P. H., Lukanova, A., Dossus, L., Biessy, C., Vineis, P., Sacerdote, C., Berrino, F., Panico, S., Tumino, R., Palli, D., Nagel, G., Linseisen, J., Boeing, H., Roddam, A., Bingham, S., Khaw, K. T., … Kaaks, R. (2006). Anthropometric measures, endogenous sex steroids and breast cancer risk in postmenopausal women: A study within the EPIC cohort. *International Journal of Cancer*, *118*(11), 2832–2839. https://doi.org/10.1002/ijc.21730

Rose, D. P., & Vona-Davis, L. (2012). The cellular and molecular mechanisms by which insulin influences breast cancer risk and progression. In *Endocrine-Related Cancer* (Vol. 19, Number 6). https://doi.org/10.1530/ERC-12-0203

Rouzier, R., Perou, C. M., Symmans, W. F., Ibrahim, N., Cristofanilli, M., Anderson, K., Hess, K. R., Stec, J., Ayers, M., Wagner, P., Morandi, P., Fan, C., Rabiul, I., Ross, J. S., Hortobagyi, G. N., & Pusztai, L. (2005). Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clinical Cancer Research, 11*(16), 5678–5685. https://doi.org/10.1158/1078-0432.CCR-04-2421

Sardanelli, F., Boetes, C., Borisch, B., Decker, T., Federico, M., Gilbert, F. J., Helbich, T., Heywang-Köbrunner, S. H., Kaiser, W. A., Kerin, M. J., Mansel, R. E., Marotti, L., Martincich, L., Mauriac, L., Meijers-Heijboer, H., Orecchia, R., Panizza, P., Ponti, A., Purushotham, A. D., … Wilson, R. (2010). Magnetic resonance imaging of the breast: Recommendations from the EUSOMA working group. *European Journal of Cancer*, *46*(8), 1296–1316. https://doi.org/10.1016/j.ejca.2010.02.015

Sawe, R. T., Kerper, M., Badve, S., Li, J., Sandoval-Cooper, M., Xie, J., Shi, Z., Patel, K., Chumba, D., Ofulla, A., Prosperi, J., Taylor, K., Stack, M. S., Mining, S., & Littlepage, L. E. (2016). Aggressive breast cancer in western Kenya has early onset, high proliferation, and immune cell infiltration. *BMC Cancer*, *16*(1). https://doi.org/10.1186/s12885-016-2204-6

Sayed, S., Moloo, Z., Wasike, R., Bird, P., Oigara, R., Govender, D., Kibera, J., Carrara, H., & Saleh, M. (2014). Is breast cancer from Sub Saharan Africa truly receptor poor? Prevalence of ER/PR/HER2 in breast cancer from Kenya. *Breast*, *23*(5), 591–596. https://doi.org/10.1016/j.breast.2014.06.006

Sayed, S., Moloo, Z., Wasike, R., Bird, P., Oigara, R., Njoroge, F. W., Shaikh, A. J., Prasad, S. V., Vinayak, S., Gierach, G. L., Dawsey, S. M., Palakal, M., Fan, S., Mullooly, M., Chauhan, R., Okiro, P., Gakinya, S., Nzioka, A., Kyobutungi, C., … Saleh, M. (2018). Ethnicity and breast cancer characteristics in Kenya. *Breast Cancer Research and Treatment*, *167*(2), 425–437. https://doi.org/10.1007/s10549-017-4511-2

Schilling, M. P. R., & Silva, I. F. da. (2020). Family history of breast cancer and risk of benign breast diseases: an integrative literature review. *Mastology*, *30*. https://doi.org/10.29289/25945394202020200039

Shield, K. D., Soerjomataram, I., & Rehm, J. (2016). Alcohol Use and Breast Cancer: A Critical Review. In *Alcoholism: Clinical and Experimental Research* (Vol. 40, Number 6, pp. 1166–1181). Blackwell Publishing Ltd. https://doi.org/10.1111/acer.13071

Shimada, N., Iwasaki, M., Kasuga, Y., Yokoyama, S., Onuma, H., Nishimura, H., Kusama, R., Hamada, G. S., Nishimoto, I. N., Iyeyasu, H., Motola, J., Laginha, F. M., Kurahashi, N., & Tsugane, S. (2009). Genetic polymorphisms in estrogen metabolism and breast cancer risk in case-control studies in Japanese, Japanese Brazilians and non-Japanese Brazilians. *Journal of Human Genetics*, *54*(4), 209–215. https://doi.org/10.1038/jhg.2009.13

Shimada, T., Watanabe, J., Kawajiri, K., Sutter, T. R., Guengerich, F. P., Gillam, E. M. J., & Inoue, K. (1999). Catalytic properties of polymorphic human cytochrome P450 1B1 variants. *Carcinogenesis*, *20*(8), 1607–1613.

Shin, A., Kang, D., Choi, J.-Y., Lee, K.-M., Park, S. K., Noh, D.-Y., Ahn, S.-H., & Yoo, K.-Y. (2007). Cytochrome P450 1A1 (CYP1A1) polymorphisms and breast cancer risk in Korean women. *Experimental and Molecular Medicine*, *39*(3), 361–366. https://doi.org/10.1038/emm.2007.40

Sissung, T. M., Price, D. K., Sparreboom, A., & Figg, W. D. (2006). Pharmacogenetics and regulation of human cytochrome P450 1B1: Implications in hormone-mediated tumor metabolism and a novel target for therapeutic intervention. In *Molecular Cancer Research* (Vol. 4, Number 3, pp. 135–150). https://doi.org/10.1158/1541-7786.MCR-05-0101

Sokolova, A., Johnstone, K. J., McCart Reed, A. E., Simpson, P. T., & Lakhani, S. R. (2023). Hereditary breast cancer: syndromes, tumour pathology and molecular testing. In *Histopathology* (Vol. 82, Number 1, pp. 70–82). John Wiley and Sons Inc. https://doi.org/10.1111/his.14808

Srivastava, V., Meena, R. K., Ansari, M. A., Kumar, D., & Kumar, A. (2020). A study of anxiety and depression in benign breast disease. *Journal of Mid-Life Health*, *11*(4), 200–209. https://doi.org/10.4103/jmh.JMH_85_20

Sun, M.-Y., Du, H.-Y., Zhu, A.-N., Liang, H.-Y., de Garibay, G. R., Li, F.-X., Li, M., & Yang, X.-X. (2015). Genetic polymorphisms in estrogen-related genes and the risk of breast cancer among Han Chinese women. *International Journal of Molecular Sciences*, *16*(2), 4121–4135. https://doi.org/10.3390/ijms16024121

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Swain, S. M., Shastry, M., & Hamilton, E. (2023). Targeting HER2-positive breast cancer: advances and future directions. *Nature Reviews Drug Discovery*, *22*(2), 101–126. https://doi.org/10.1038/s41573-022-00579-0

Tarach, P. (2021). Application of polymerase chain reaction-restriction fragment length polymorphism (RFLP-PCR) in the analysis of single nucleotide polymorphisms (SNPs). *Acta Universitatis Lodziensis. Folia Biologica et Oecologica*, *17*, 48–53. https://doi.org/10.18778/1730-2366.16.14

Tazzite, A., Jouhadi, H., Saiss, K., Benider, A., & Nadifi, S. (2013). Relationship between family history of breast cancer and clinicopathological features in Moroccan patients. *Ethiopian Journal of Health Sciences*, *23*(2), 150–157.

The R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. https://doi.org/https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf

Tramacere, I., Scotti, L., Jenab, M., Bagnardi, V., Bellocco, R., Rota, M., Corrao, G., Bravi, F., Boffetta, P., & La Vecchia, C. (2010). Alcohol drinking and pancreatic cancer risk: A meta-analysis of the dose-risk relation. *International Journal of Cancer*, *126*(6), 1474–1486. https://doi.org/10.1002/ijc.24936

Trubicka, J., Grabowska-Kłujszo, E., Suchy, J., Masojć, B., Serrano-Fernandez, P., Kurzawski, G., Cybulski, C., Górski, B., Huzarski, T., Byrski, T., Gronwald, J., Złowocka, E., Kładny, J., Banaszkiewicz, Z., Wiśniowski, R., Kowalska, E., Lubinski, J., & Scott, R. J. (2010). Variant alleles of the CYP1B1 gene are associated with colorectal cancer susceptibility. *BMC Cancer*, *10*. https://doi.org/10.1186/1471-2407-10-420

Tsuchiya, Y., Nakajima, M., & Yokoi, T. (2005). Cytochrome P450-mediated metabolism of estrogens and its regulation in human. *Cancer Letters*, *227*(2), 115–124. https://doi.org/10.1016/j.canlet.2004.10.007

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, *35*(SUPPL.2). https://doi.org/10.1093/nar/gkm306

Wang, Y., Xu, M., Ke, Z. ji, & Luo, J. (2017). Cellular and molecular mechanisms underlying alcohol-induced aggressiveness of breast cancer. *Pharmacological Research*, *115*, 299–308. https://doi.org/10.1016/j.phrs.2016.12.005

Weigelt, B., Baehner, F. L., & Reis-Filho, J. S. (2010). The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: A retrospective of the last decade. *Journal of Pathology*, *220*(2), 263–280. https://doi.org/10.1002/path.2648

Wellings, E., Vassiliades, L., & Abdalla, R. (2016). Breast Cancer Screening for High-Risk Patients of Different Ages and Risk - Which Modality Is Most Effective? *Cureus*. https://doi.org/10.7759/cureus.945

Wen, C., Wu, L., Fu, L., Wang, B., & Zhou, H. (2017a). Unifying mechanism in the initiation of breast cancer by metabolism of estrogen (Review). *Molecular Medicine Reports*, *16*(2), 1001–1006. https://doi.org/10.3892/mmr.2017.6738

Wen, C., Wu, L., Fu, L., Wang, B., & Zhou, H. (2017b). Unifying mechanism in the initiation of breast cancer by metabolism of estrogen (Review). *Molecular Medicine Reports*, *16*(2), 1001–1006. https://doi.org/10.3892/mmr.2017.6738

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

World Health Organization. (2014). *WHO position paper on mammography screening.* 4. https://www.who.int/publications/i/item/who-position-paper-on- mammography-screening

Yager, J. D., Davidson, N. E., & Hopkins Bloomberg, J. (2006). Estrogen carcinogenesis in breast cancer. *N Engl J Med*, *354*(3), 270–282. https://doi.org/10.1056/NEJMra050776

Yim, D.-S., Park, S. K., Yoo, K.-Y., Yoon, K.-S., Chung, H., Kang, H. J., Ahn, S.-H., Noh, D.-Y., Choe, K.-J., Jang, I.-J., Shin, S.-G., Strickland, P. T., Hirvonen, A., & Kang, D. (2001). Relationship between the Val 158 Met polymorphism of catechol O-methyl transferase and breast cancer. *Pharmacogenetics*, *11*(4), 279–286. https://doi.org/10.1097/00008571-200106000-00001

Yu, P. J., Chen, W. G., Feng, Q. L., Chen, W., Jiang, M. J., & Li, Z. Q. (2015). Association between CYP1B1 gene polymorphisms and risk factors and susceptibility to laryngeal cancer. *Medical Science Monitor*, *21*, 239–245. https://doi.org/10.12659/MSM.893084

Zhan, P., Wang, Q., Qian, Q., Wei, S. Z., & Yu, L. K. (2011). CYP1A1 MspI and exon7 gene polymorphisms and lung cancer risk: An updated meta-analysis and review. *Journal of Experimental and Clinical Cancer Research*, *30*(1). https://doi.org/10.1186/1756-9966-30-99

Zhou, W., Ding, Q., Pan, H., Wu, N., Liang, M., Huang, Y., Chen, L., Zha, X., Liu, X., & Wang, S. (2014). Risk of breast cancer and family history of other cancers in first-degree relatives in Chinese women: A case control study. *BMC Cancer*, *14*(1), 1–7. https://doi.org/10.1186/1471-2407-14-662

APPENDICES

Appendix I: Ethical Approval

THE AGA KHAN UNIVERSITY

Ref: 2020/IERC-26 (v2)
March 3, 2020

Dr. Shahin Sayed, Faculty & AKU Site Investigator
Department of Pathology, Aga Khan University

Ms. Mary Muriithi – Doctor of Philosophy Degree
Jomo Kenyatta University of Science and Technology
Nairobi, Kenya

Dear Dr. Shahin,

RE: Genetic Polymorphisms in Estrogen Metabolising Enzymes in Patients Diagnosed with Estrogen Receptor Positive Breast Cancer at the Aga Khan University Hospital, Nairobi

The Aga Khan University, Nairobi (AKUN), Institutional Ethics Review Committee (IERC) is in receipt of your protocol and application form resubmitted to the Research Office on March 2, 2020. With reference to the AKU-IERC letter Ref: 2020/IERC-26 (v1) dated February 19, 2020, the committee notes that the researcher has addressed concerns earlier raised. This is a retrospective study that seeks to profile genetic polymorphisms (SNPs) in estrogen metabolizing enzymes and determine possible association with the prognosis and outcome in breast cancer patients diagnosed at Aga Khan University Hospital, Nairobi. Data will be collected from retrospective review of patient's charts, clinical notes and through molecular experiments for the period of June 2012 and June 2015. There will be no patient contact.

As per the AKU IERC regulations, the committee records that this study qualifies for exemption from a full committee review process. The committee has thus approved your exemption request *{as per attached official stamped protocol and attachments - version 2020/IERC-26 (v2)}* This approval is valid from **March 3, 2020** to **March 2, 2021.** As applicable, prior to export of biological specimens, ensure an MTA is in place as well as seek shipment authority/permit from the relevant government ministry.

To maintain anonymity and confidentiality of data no patient identifiers will be collected. The study should be conducted in full accordance with all the applicable institutional policies of the AKUHN, especially with regards to **patient records management**. You should notify the IERC immediately of any changes that may affect your project. Further, you must provide an interim report **60 days before expiration** of the validity of this approval and request extension if additional time is required for study completion. You must advise the IERC when this study is completed or discontinued and a final report submitted to the Research Office. The hospital management should be notified of manuscripts emanating from this work.

If you have any questions, please contact Research Office - research.supportea@aku.edu or call 020-366 2148/1136.

With best wishes,

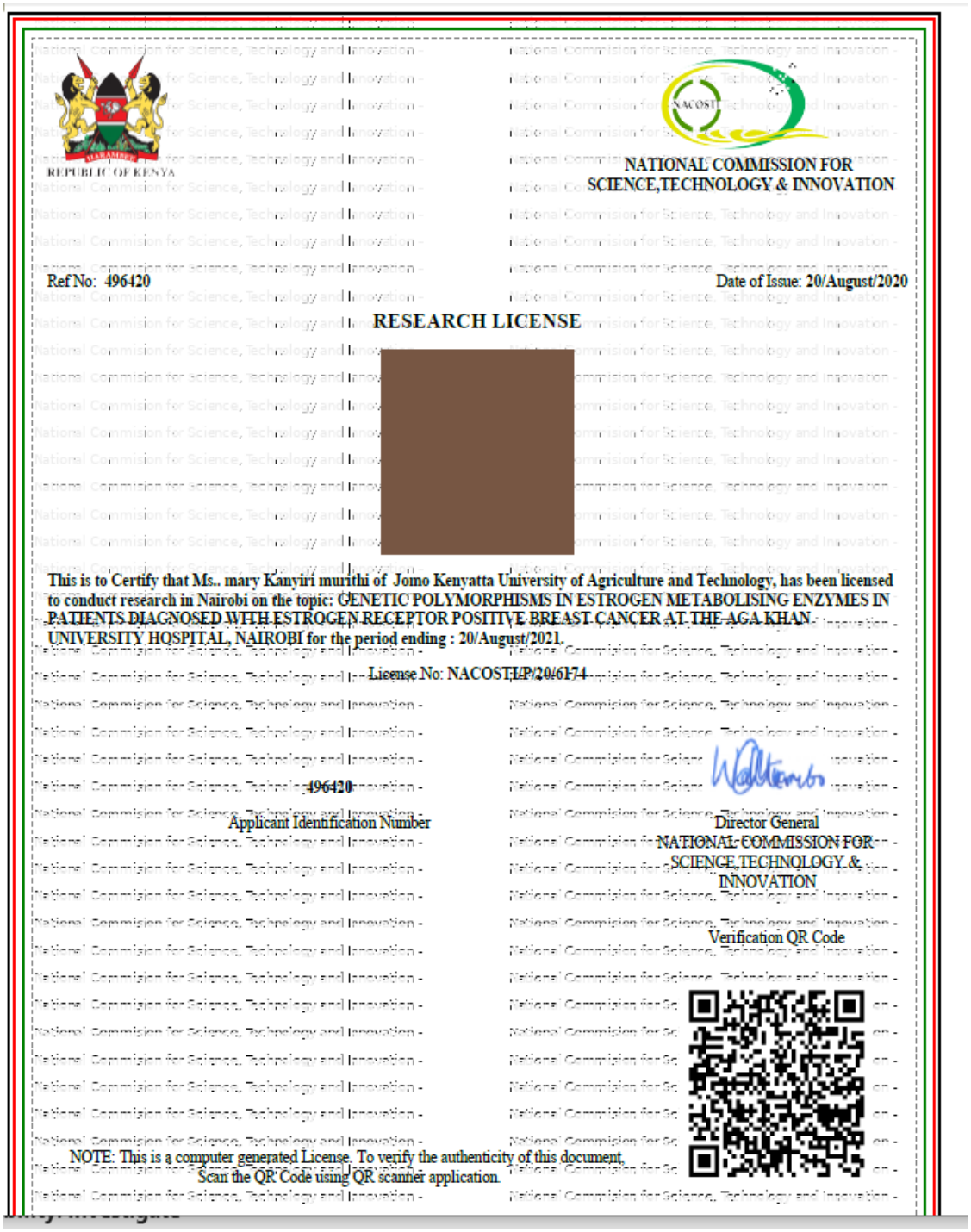
Wangari.

Dr. Wangari Waweru-Siika,
Chair - Institutional Ethics Review Committee (IERC)
Aga Khan University, (Kenya)

3rd Floor, Park Place Building, 2nd Parklands Avenue, Off Limuru Road
P. O. Box 30270, GPO 00100, Nairobi, Kenya
Tel: +254 20 366 1200; Website: www.aku.edu

AK 958

Appendix II: NACOSTI Permit



REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: 496420

Date of Issue: 20/August/2020

RESEARCH LICENSE

This is to Certify that Ms.. mary Kanyiri murithi of Jomo Kenyatta University of Agriculture and Technology, has been licensed to conduct research in Nairobi on the topic: GENETIC POLYMORPHISMS IN ESTROGEN METABOLISING ENZYMES IN PATIENTS DIAGNOSED WITH ESTROGEN RECEPTOR POSITIVE BREAST CANCER AT THE AGA KHAN UNIVERSITY HOSPITAL, NAIROBI for the period ending : 20/August/2021.

License No: NACOSTI/P/20/6174

496420

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

Appendix III: Consent Page

Confidential

Introduction and Consent

Record Number	_____
Facility/Institution	◯ AKU ◯ KAIC ◯ ACC
Study ID	_____ (This is the study ID such as AKU001)
Date of Interview	_____ (DDMMYYY)

Hello my name is ..... I am here on behalf of ..... As you know, you have been asked to participate in a study, and your doctor/nurse has explained the study to you. This study involves your breast tissue which will be used for testing and research. You were given the opportunity to sign the INFORMED CONSENT, and you willingly signed the INFORMED CONSENT and agreed to participate in the study. Today, I would like to invite you to participate in filling out this important questionnaire which is related to the study. I will be asking you questions on your socio-demographics, medical history, reproductive history and other health behaviours. I will also ask you which ethnicity you belong to. This will not in any way affect the management and treatment of your condition. In addition to these questions, we will also collect some blood, saliva and tissue samples for further testing in our laboratory. No HIV testing will be performed as part of this study. Your participation is voluntary, and if you choose not to participate, you will not be treated with prejudice. Please note that if at any point you feel uncomfortable, you can stop the interview at any time if you do not want to continue. Your answers to the questions asked will be completely confidential and will be used for research purposes only. Your responses and those of others who agree to participate in the study will be used by researchers for this study, and other partner organisations to understand the problem of breast cancer in Kenyan women and how this may differ among the three major ethnic groupings.

Do you have any questions about the survey?	◯ Yes ◯ No
Do I have agreement to proceed?	◯ Yes ◯ No
Completed Consent Document	
Does the respondent agree to be interviewed?	◯ Yes ◯ No
Interviewer's Signature	

Appendix IV: Data Abstraction Form

The standardized form below was used to systematically record data regarding the participants. The form has seven section capturing information on: sample and identification, demographic and anthropometric, reproductive and hormonal history, medical and family history, lifestyle and environmental exposures, tumor clinicopathology data (cases only) and genetic data

	Variable Name	**Description/ Definition**	**Data Type / Categories**
I. SAMPLE AND IDENTIFICATION	Sample id	Unique study identification code	Text (Alphanumeric)
	Sample type	Participant group classification	1 = Case (ER+ Breast Cancer), 2 = BBD, 3 = Healthy Control
II. DEMOGRAPHIC & ANTHROPOMETRIC DATA	Age	Age at recruitment (years)	Continuous (Integer)
	Body mass index (BMI)	Body Mass Index (kg/m^2)	Continuous (Decimal)
III. REPRODUCTIVE & HORMONAL HISTORY	Age at menarche	Age at first menarche (years)	Continuous (Integer)
	Menopausal status	Menopausal status at diagnosis/recruitment	1 = Premenopausal, 2 = Postmenopausal
	Parity	Number of live births	Continuous (Integer)
	Hormonal contraceptives	Ever use of hormonal contraceptives	0 = No, 1 = Yes
IV. MEDICAL & FAMILY HISTORY	Family history of breast cancer	First-degree relative with breast cancer	0 = No, 1 = Yes
	Family history of other cancers	First-degree relative with other cancer	0 = No, 1 = Yes
	Diabetes	Diagnosed with diabetes mellitus	0 = No, 1 = Yes
	Hypertension	Diagnosed with hypertension	0 = No, 1 = Yes
V. LIFESTYLE & ENVIRONMENTAL EXPOSURES	Tobacco	Ever smoked tobacco	0 = No, 1 = Yes
	Alcohol	Ever consumed alcohol	0 = No, 1 = Yes
	Radiation therapy	History of therapeutic radiation exposure	0 = No, 1 = Yes
VI. TUMOR CLINICOPATHOLOGY DATA (Cases only)	Histological type	Primary histopathological diagnosis	1 = Invasive Ductal Carcinoma, 2 = Other

	Grade	Tumor histological grade (Nottingham)	1 = Grade 1, 2 = Grade 2, 3 = Grade 3
	Stage	Clinical/pathological TNM stage (simplified)	1 = Stage I/II (Early), 2 = Stage III/IV (Late)
	Lymph node invasion	Presence of lymph node metastasis	0 = No, 1 = Yes
	Molecular subtype	Immunohistochemistry-based subtype	1 = Luminal A, 2 = Luminal B
VII. GENETIC DATA	rs4646903	CYP1A1 rs4646903 genotype	0 = Wild type, 1 = Heterozygous, 2 = Variant
	rs1048943	CYP1A1 rs1048943 genotype	0 = Wild type, 1 = Heterozygous, 2 = Variant
	rs776746	CYP3A5 rs776746 genotype	0 = Wild type, 1 = Heterozygous, 2 = Variant
	rs4680	COMT rs4680 genotype	0 = Wild type, 1 = Heterozygous, 2 = Variant
	rs1056827	*CYP1B1* rs1056827 genotype	0 = Wild type, 1 = Heterozygous, 2 = Variant
	rs10012	*CYP1B1* rs10012 genotype	0 = Wild type, 1 = Heterozygous, 2 = Variant
	rs1800440	*CYP1B1* rs1800440 genotype	0 = Wild type, 1 = Heterozygous, 2 = Variant
	rs1056836	*CYP1B1* rs1056836 genotype	0 = Wild type, 1 = Heterozygous, 2 = Variant

Appendix V: Participants Group Classification and Identifiers

The table below lists the 162 participants group classification (sample type) along with their unique sample identifiers. Coding: **C** = Case (ER+ Breast Cancer), **B** = Benign Breast Disease (BBD), **H** = Healthy Control. Samples are grouped by classification.

	BBD Group (B)	**Case Group (C)**	**Healthy Control Group (H)**
1.	AKU011	AKU041	AKU 182
2.	AKU025	AKU042	AKU 183
3.	AKU026	AKU057	AKU 191
4.	AKU031	AKU060	AKU 193
5.	AKU034	AKU073	AKU 194
6.	AKU039	AKU076	AKU 195
7.	AKU043	AKU086	AKU 196
8.	AKU044	AKU087	AKU 197
9.	AKU046	AKU088	AKU 198
10.	AKU047	AKU094	AKU 199
11.	AKU048	AKU097	AKU 200
12.	AKU049	AKU099	AKU 201
13.	AKU054	AKU100	AKU 202
14.	AKU056	AKU103	AKU 204
15.	AKU059	AKU104	AKU 205
16.	AKU061	AKU105	AKU 206
17.	AKU062	AKU106	AKU 207
18.	AKU063	AKU109	AKU 208
19.	AKU064	AKU110	AKU 209
20.	AKU067	AKU112	
21.	AKU069	AKU113	
22.	AKU070	AKU116	
23.	AKU071	AKU118	
24.	AKU075	AKU122	
25.	AKU077	AKU124	
26.	AKU078	AKU127	
27.	AKU081	AKU128	
28.	AKU083	AKU133	
29.	AKU107	AKU142	
30.	AKU115	AKU149	
31.	AKU119	AKU152	
32.	AKU126	AKU158	
33.	AKU132	AKU164	
34.	AKU135	AKU173	
35.	AKU136	AKU211	
36.	AKU139	AKU155	
37.	AKU144	KAIC004	
38.	AKU145	KAIC005	
39.	AKU146	KAIC006	

40.	AKU148	KAIC007	
41.	AKU151	KAIC008	
42.	AKU153	KAIC011	
43.	AKU154	KAIC016	
44.	AKU159	KAIC018	
45.	AKU162	KAIC019	
46.	AKU163	KAIC021	
47.	AKU165	KAIC023	
48.	AKU166	KAIC024	
49.	AKU167	KAIC025	
50.	AKU170	KAIC026	
51.	AKU172	KAIC027	
52.	AKU175	KAIC029	
53.	AKU176	KAIC030	
54.	AKU177	KAIC032	
55.	AKU180	KAIC033	
56.	AKU181	KAIC035	
57.	AKU182	KAIC041	
58.	AKU183	KAIC046	
59.	AKU185	KAIC047	
60.	AKU187	KAIC051	
61.	AKU188	KAIC052	
62.	AKU189	KAIC057	
63.	AKU190	KAIC059	
64.	AKU191	KAIC060	
65.	AKU192		
66.	AKU194		
67.	AKU196		
68.	AKU197		
69.	AKU198		
70.	AKU200		
71.	AKU202		
72.	AKU204		
73.	AKU205		
74.	AKU206		
75.	AKU207		
76.	AKU208		
77.	AKU210		
78.	AKU212		
79.	AKU213		

Appendix VI: Participants Demographic, Clinical and Lifestyle Characteristics

This table presents the coded values for key demographic, reproductive, and clinical variables collected from study participants. Variable abbreviations are defined as follows: Demographic/Clinical: Participant Identifier (PID), Body Mass Index (BMI), Diabetes Mellitus (DM), Hypertension (HTN); Reproductive: Menopausal Status (Meno_Stat), Hormonal Contraceptive Use (Horm_Contracept), Parity (Parity), Age at Menarche (Menarche_Age); Family History: Family History of Breast Cancer (FH_BC), Family History of Other Cancers (FH_OtherCa); Lifestyle/Exposure: Smoking (Smoking), Alcohol Consumption (Alcohol), Radiation Therapy (RTx).

PID	Age	Menarche_Age	BMI	Meno_Stat	Parity	Horm_Contracept	FH_BC	FH_OtherCa	DM	HTN	Smoking	Alcohol	RTx
AKU041	40	13	27	No	Yes	Yes	No	Yes	No	No	Yes	Yes	No
AKU042	57	12	40	Yes	Yes	No	No	Yes	Yes	Yes	No	No	No
AKU057	52	12	25	No	Yes	No	No	No	No	No	No	Yes	No
AKU060	28	14	18	No	No	No	No	No	No	No	Yes	Yes	No
AKU073	43	14	33	No	Yes	Yes	No	Yes	Yes	Yes	No	No	No
AKU076	42	12	32	No	Yes	Yes	No	Yes	No	No	No	No	No
AKU086	52	12	21	Yes	Yes	No	No	Yes	No	No	No	Yes	No
AKU087	63	14	34	Yes	Yes	No	No	No	No	No	No	Yes	No
AKU088	34	14	21	No	No	No	Yes	Yes	No	No	No	No	No
AKU094	56	16	24	Yes	Yes	Yes	No	Yes	No	No	No	No	No
AKU097	32	14	23	No	No	Yes	No	No	No	No	No	Yes	No
AKU099	60	13	27	Yes	Yes	No	No	No	No	No	No	No	No
AKU100	75	16	37	Yes	Yes	No	No	No	No	No	No	Yes	No
AKU103	64	13	23	Yes	Yes	No	Yes	No	No	Yes	No	No	No
AKU104	57	14	33	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No
AKU105	41	17	30	No	Yes	Yes	Yes	Yes	No	No	No	Yes	No
AKU106	48	8	27	No	Yes	Yes	No	Yes	No	No	No	Yes	No
AKU109	52	15	41	No	Yes	Yes	No	No	No	No	No	Yes	No

AKU110	44	11	26	No	No	No	No	Yes	No	Yes	No	Yes	No
AKU112	57	13	36	Yes	Yes	Yes	No	No	No	Yes	No	No	No
AKU113	39	14	25	No	Yes	Yes	No	Yes	No	No	No	Yes	No
AKU116	40	16	28	No	Yes	Yes	Yes	No	No	Yes	No	Yes	No
AKU118	70	13	25	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes	No
AKU122	53	16	33	Yes	Yes	Yes	Yes	No	No	No	No	No	No
AKU124	62	14	30	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No
AKU127	49	14	32	Yes	No	Yes	No	No	No	Yes	No	Yes	Yes
AKU128	53	17	27	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No
AKU133	78	17	26	Yes	Yes	Yes	No	No	No	Yes	No	No	No
AKU142	51	16	27	Yes	Yes	Yes	Yes	No	No	No	No	Yes	No
AKU149	54	15	36	Yes	Yes	Yes	No	Yes	No	No	No	Yes	No
AKU152	44	15	26	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU158	54	16	27	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No	No
AKU164	38	17	36	Yes	Yes	Yes	No	No	No	No	No	Yes	No
AKU173	48	13	30	Yes	Yes	Yes	No	Yes	No	No	No	No	No
AKU211	55	16	35	No	Yes	Yes	Yes	No	No	No	No	No	No
AKU155	53	13	41	Yes	Yes	Yes	Yes	No	No	No	No	No	No
KAIC004	59	15	34	Yes	Yes	Yes	No	Yes	No	No	No	No	No
KAIC005	55	14	42	Yes	Yes	No	No	No	No	No	No	No	No
KAIC006	36	13	26	No	Yes	Yes	No	Yes	No	No	No	No	No
KAIC007	50	14	23	Yes	Yes	No	No	No	No	No	No	No	No
KAIC008	71	13	33	Yes	Yes	No	No	No	Yes	Yes	No	No	No
KAIC011	68	14	37	Yes	Yes	No	No	No	No	Yes	No	No	No
KAIC016	67	17	10	Yes	Yes	Yes	No	No	No	Yes	No	No	No
KAIC018	43	16	42	No	Yes	Yes	No	No	No	No	No	No	No
KAIC019	41	13	32	No	Yes	Yes	No	No	No	No	No	No	No
KAIC021	56	18	23	No	Yes	Yes	No	No	No	No	No	No	No
KAIC023	45	18	23	No	Yes	Yes	No	No	No	No	No	No	No

KAIC024	57	18	35	Yes	Yes	Yes	No	No	No	No	No	No	No
KAIC025	47	17	32	No	Yes	No	No	No	No	No	No	No	No
KAIC026	39	16	35	No	No	No	No	No	No	No	No	No	No
KAIC027	37	12	42	No	Yes	No	No	No	No	No	No	No	No
KAIC029	59	19	26	No	Yes	Yes	No	No	No	Yes	No	No	No
KAIC030	59	22	28	Yes	Yes	Yes	Yes	No	No	No	No	No	No
KAIC032	53	12	27	Yes	Yes	Yes	No	Yes	No	No	No	Yes	No
KAIC033	60	12	24	Yes	Yes	No	No	No	No	No	No	No	No
KAIC035	61	13	33	Yes	Yes	Yes	No	No	Yes	Yes	No	No	No
KAIC041	49	16	38	Yes	Yes	No	No	Yes	No	Yes	No	No	No
KAIC046	53	13	42	No	Yes	Yes	Yes	No	No	Yes	No	No	No
KAIC047	36	14	28	No	Yes	Yes	No	No	No	No	No	No	No
KAIC051	50	12	35	No	Yes	Yes	No	No	No	No	No	No	No
KAIC052	36	13	24	No	No	Yes	No	No	No	No	No	Yes	No
KAIC057	48	16	26	No	Yes	No	No	Yes	No	No	No	No	No
KAIC059	61	15	30	Yes	Yes	No	No	No	No	No	No	No	No
KAIC060	40	15	35	No	Yes	Yes	No	No	No	No	No	No	No
AKU011	51	15	22	Yes	Yes	Yes	No	No	No	No	No	No	No
AKU025	42	14	32	No	Yes	Yes	No	Yes	No	Yes	No	Yes	No
AKU026	47	14	28	No	Yes	Yes	No	Yes	No	No	No	Yes	No
AKU031	28	13	27	No	No	No	Yes	Yes	No	No	Yes	Yes	No
AKU034	35	13	35	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU039	42	13	30	No	Yes	Yes	No	No	No	No	No	No	No
AKU043	34	14	19	No	Yes	No	Yes	Yes	No	No	No	Yes	No
AKU044	40	14	26	No	Yes	Yes	No	No	No	No	No	No	No
AKU046	55	16	24	Yes	Yes	No	No	No	No	No	No	Yes	No
AKU047	57	14	22	Yes	Yes	No	No	No	No	No	No	Yes	No
AKU048	60	12	32	Yes	Yes	Yes	No	No	No	Yes	No	Yes	Yes
AKU049	61	15	28	Yes	Yes	No	No	No	No	No	No	Yes	No

AKU054	40	12	41	No	Yes	Yes	Yes	Yes	No	No	No	Yes	No
AKU056	42	14	24	No	No	No	Yes	No	No	No	No	No	No
AKU059	74	16	26	Yes	No	No	Yes	Yes	No	Yes	No	No	No
AKU061	57	16	24	Yes	Yes	Yes	No	No	No	No	No	No	No
AKU062	61	16	34	Yes	Yes	Yes	No	Yes	No	No	No	No	No
AKU063	51	16	28	No	Yes	Yes	No	No	Yes	Yes	No	Yes	No
AKU064	63	14	24	Yes	Yes	Yes	No	No	No	No	No	No	Yes
AKU067	42	16	33	No	Yes	Yes	No	No	No	No	No	No	No
AKU069	42	14	30	No	Yes	Yes	No	Yes	No	No	No	No	No
AKU070	45	14	30	Yes	Yes	Yes	Yes	No	No	No	No	Yes	No
AKU071	38	13	22	No	No	Yes	Yes	No	No	No	No	No	No
AKU075	48	15	35	No	Yes	Yes	No	No	No	No	No	No	No
AKU077	41	14	20	No	No	Yes	Yes	No	No	No	No	Yes	No
AKU078	52	16	35	Yes	Yes	Yes	No	No	No	No	No	Yes	No
AKU081	47	16	25	No	Yes	Yes	No	Yes	No	Yes	No	Yes	No
AKU083	50	12	28	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU107	31	13	34	No	Yes	Yes	Yes	No	Yes	Yes	No	Yes	No
AKU115	45	15	25	Yes	Yes	Yes	No	No	No	No	No	No	No
AKU119	35	14	25	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU126	46	14	31	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU132	38	15	31	No	Yes	Yes	No	Yes	No	No	No	Yes	No
AKU135	35	16	31	No	Yes	Yes	Yes	No	No	No	No	No	No
AKU136	36	14	37	No	Yes	Yes	No	No	Yes	Yes	No	No	No
AKU139	36	14	34	No	No	No	Yes	Yes	No	No	No	Yes	No
AKU144	35	13	43	No	No	Yes	No	No	Yes	Yes	No	Yes	No
AKU145	27	14	30	No	No	Yes	No	No	No	No	No	Yes	No
AKU146	33	12	29	No	No	Yes	Yes	Yes	No	No	Yes	Yes	No
AKU148	23	13	24	No	No	No	No	No	No	No	No	Yes	No
AKU151	44	17	31	No	Yes	Yes	No	No	No	No	No	No	Yes

AKU153	49	13	36	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	No
AKU154	41	13	29	No	Yes	Yes	No	No	No	No	No	No	No
AKU159	37	13	30	No	No	No	No	No	No	No	No	Yes	No
AKU162	63	11	28	No	Yes	No	No	No	No	No	No	No	No
AKU163	55	14	31	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes	No
AKU165	45	13	22	No	Yes	Yes	No	No	No	No	No	No	No
AKU166	32	12	25	No	No	Yes	No	Yes	No	No	No	Yes	No
AKU167	47	15	31	No	Yes	Yes	Yes	No	No	No	No	Yes	No
AKU170	45	14	26	No	Yes	Yes	Yes	No	No	No	No	Yes	No
AKU172	35	14	25	No	No	Yes	No	No	No	No	No	Yes	No
AKU175	31	14	20	No	No	No	No	No	No	No	No	No	No
AKU176	40	14	30	No	Yes	Yes	No	Yes	No	No	No	Yes	No
AKU177	36	15	26	No	Yes	Yes	No	Yes	No	Yes	No	Yes	No
AKU180	37	14	28	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU181	49	14	29	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU182	37	13	29	No	Yes	Yes	No	No	No	No	No	No	No
AKU183	39	16	31	No	Yes	Yes	No	No	No	No	No	No	No
AKU185	33	11	30	No	No	Yes	No	No	No	Yes	No	No	No
AKU187	42	12	31	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU188	38	14	24	No	Yes	Yes	No	Yes	No	No	No	No	No
AKU189	39	13	23	No	Yes	Yes	No	Yes	No	No	No	Yes	No
AKU190	41	12	26	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU191	42	13	34	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU192	56	15	30	Yes	Yes	No	No	No	Yes	Yes	No	Yes	No
AKU194	42	13	27	No	Yes	Yes	No	No	No	No	No	No	No
AKU196	49	16	30	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU197	42	14	24	No	Yes	Yes	No	No	No	Yes	No	Yes	No
AKU198	48	13	24	No	Yes	Yes	No	No	No	Yes	No	Yes	No
AKU200	41	14	30	No	Yes	Yes	No	No	No	Yes	No	No	No

AKU202	52	17	30	Yes	Yes	Yes	No	No	No	Yes	No	Yes	No
AKU204	45	12	29	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU205	41	15	29	No	No	No	No	No	No	Yes	No	Yes	No
AKU206	42	12	28	No	No	Yes	No	No	No	No	No	Yes	No
AKU207	40	13	32	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU208	36	13	29	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU210	50	14	26	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
AKU212	58	16	25	Yes	Yes	Yes	No	No	No	No	No	Yes	No
AKU213	34	12	27	No	No	No	No	No	No	No	No	No	No
AKU 182	37	13	29	No	Yes	Yes	No	No	No	No	No	No	No
AKU 183	39	16	31	No	Yes	Yes	No	No	No	No	No	No	No
AKU 191	42	13	34	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU 193	26	11	24	No	No	Yes	No	No	No	No	No	Yes	No
AKU 194	42	13	27	No	Yes	Yes	No	No	No	No	No	No	No
AKU 195	27	14	23	No	No	No	No	No	No	No	No	No	No
AKU 196	49	16	30	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU 197	42	14	24	No	Yes	Yes	No	No	No	Yes	No	Yes	No
AKU 198	48	13	24	No	Yes	Yes	No	No	No	Yes	No	Yes	No
AKU 199	33	12	23	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU 200	41	14	30	No	Yes	Yes	No	No	No	Yes	No	No	No
AKU 201	25	11	23	No	No	Yes	No	No	No	No	No	Yes	No
AKU 202	52	17	30	Yes	Yes	Yes	No	No	No	Yes	No	Yes	No
AKU 204	45	12	29	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU 205	41	15	29	No	No	No	No	No	No	Yes	No	Yes	No
AKU 206	42	12	28	No	No	Yes	No	No	No	No	No	Yes	No
AKU 207	40	13	32	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU 208	36	13	29	No	Yes	Yes	No	No	No	No	No	Yes	No
AKU 209	25	10	19	No	No	Yes	No	No	No	No	No	Yes	No

Appendix VII: Genotypes for rs4646903, rs1048943, rs1056836, rs7767746 and rs4680

PID	rs4646903	rs1048943	rs1056836	rs776746	rs4680	PID	rs4646903	rs1048943	rs1056836	rs776746	rs4680
AKU041	TT	AA	GG	GG	AA	**AKU133**	CC	AA	GC	GG	AA
AKU042	TT	AA	GG	GG	AA	**AKU142**	TT	AA	CC	GG	AA
AKU057	TT	AA	GC	GG	AA	**AKU149**	TT	AG	GG	GG	AA
AKU060	TC	AA	CC	GG	AA	**AKU152**	TT	AA	GG	GG	AA
AKU073	TC	AA	GC	GG	AA	**AKU158**	TT	AA	GC	GG	AA
AKU076	TT	AA	GC	GG	AA	**AKU164**	TT	GG	GG	GG	AA
AKU086	TC	AA	GG	GG	AA	**AKU173**	TT	AG	GG	GG	AA
AKU087	TC	AA	GG	GG	AA	**AKU211**	TC	AA	GC	GG	AA
AKU088	TT	AA	CC	GG	AA	**AKU155**	TT	AA	GC	GG	AA
AKU094	TT	AA	GC	GG	AA	**KAIC004**	CC	AG	GC	GG	AA
AKU097	TT	AA	GG	GG	AA	**KAIC005**	TT	AA	GG	GG	AA
AKU099	TC	AA	GC	GG	AA	**KAIC006**	TT	AA	GG	GG	AA
AKU100	TT	AA	GC	GG	AA	**KAIC007**	TC	AA	GG	GG	AA
AKU103	TT	AA	CC	GG	AA	**KAIC008**	TC	AA	GC	GG	AA
AKU104	TT	AA	GG	GG	AA	**KAIC011**	TT	AA	GG	GG	AA
AKU105	TC	AA	GC	GG	AA	**KAIC016**	TT	AA	GG	GG	AA
AKU106	TC	AA	GG	GG	AA	**KAIC018**	TT	AG	GG	GG	AA
AKU109	TC	AA	CC	GG	AA	**KAIC019**	TT	AA	GC	GG	AA
AKU110	TT	AA	GC	GG	AA	**KAIC019**	TT	AA	GC	GG	AA
AKU112	TT	AA	GG	GG	AA	**KAIC021**	TT	AA	GG	GG	AG
AKU113	TT	GG	GG	GG	AA	**KAIC023**	TC	AA	GG	GG	AG
AKU116	TT	AA	GG	GG	AA	**KAIC024**	TT	AA	GC	GG	AA
AKU118	CC	AA	GC	GG	AA	**KAIC025**	TT	AA	GG	GG	AA
AKU122	TC	AA	GC	GG	AA	**KAIC026**	TC	AA	GG	GG	AA
AKU124	TT	AA	GG	GG	AA	**KAIC027**	TC	AA	GC	GG	AA
AKU127	TT	AA	GG	GG	AA	**KAIC029**	TT	AA	GC	GG	AA

AKU128	TT	AA	GC	GG	AA	**KAIC030**	TT	AA	GG	GG	AA
PID	**rs4646903**	**rs1048943**	**rs1056836**	**rs776746**	**rs4680**	**PID**	**rs4646903**	**rs1048943**	**rs1056836**	**rs776746**	**rs4680**
KAIC032	TC	AA	GC	GG	AA	**AKU064**	TT	AA	GG	GG	AA
KAIC033	TT	AA	GC	GG	AA	**AKU067**	TT	AA	GG	GG	AA
KAIC035	TC	AA	GG	GG	AA	**AKU069**	TC	AA	GG	GG	AA
KAIC041	TT	AA	GC	GG	AA	**AKU070**	TT	AA	GG	GG	AA
KAIC046	TT	AA	GC	GG	GG	**AKU071**	TC	GG	GC	GG	AA
KAIC047	TT	AA	CC	GG	AA	**AKU075**	TT	AA	GG	GG	AA
KAIC051	TT	AA	CC	GG	AA	**AKU077**	TT	AA	GC	GG	AA
KAIC052	TT	AA	GC	GG	AG	**AKU078**	TC	AA	GC	GG	AA
KAIC057	TC	AA	GG	GG	AA	**AKU081**	TC	AA	GG	GG	AA
KAIC059	TT	AA	CC	GG	AA	**AKU083**	TT	AA	GG	GG	AA
KAIC060	TT	AA	GG	GG	AA	**AKU107**	TT	AA	GG	GG	AA
AKU011	TT	AG	GC	GG	AA	**AKU115**	TT	AA	GG	GG	AA
AKU025	TC	AA	GC	GG	AA	**AKU119**	TT	AA	GG	GG	AA
AKU026	TC	AA	GG	GG	AA	**AKU126**	TT	AA	GG	GG	AA
AKU031	TT	AG	GG	GG	AA	**AKU132**	TC	AA	GC	GG	AA
AKU034	TT	AA	GC	GG	AA	**AKU135**	TT	AA	GC	GG	AA
AKU039	TT	AA	GG	GG	AA	**AKU136**	TT	AG	GC	GG	AA
AKU043	TT	AA	GG	GG	AA	**AKU139**	TT	AG	GG	GG	AA
AKU044	TT	AA	GG	GG	AA	**AKU144**	TC	AA	GC	GG	AA
AKU046	TT	AA	GG	GG	AA	**AKU145**	TT	AA	GG	GG	AA
AKU047	TT	AA	GG	GG	AA	**AKU146**	TT	AA	GG	GG	AA
AKU048	TT	AA	GG	GG	AA	**AKU148**	TT	AA	GC	GG	AA
AKU049	TT	AA	GG	GG	AA	**AKU151**	TT	AA	GG	GG	AA
AKU054	TT	AA	GG	GG	AA	**AKU153**	TT	AA	GG	GG	AA
AKU056	TT	AA	GG	GG	AA	**AKU154**	TT	AA	GC	GG	AA
AKU059	TC	AA	GG	GG	AA	**AKU159**	TT	AA	GC	GG	AA
AKU061	TT	AA	GG	GG	AA	**AKU162**	TT	AA	GC	GG	AA

AKU062	TT	AA	GG	GG	AA	**AKU163**	TT	AA	GG	GG	AA
AKU063	TT	AA	GG	GG	AA	**AKU165**	TT	AA	GC	GG	AA

PID	rs4646903	rs1048943	rs1056836	rs776746	rs4680	PID	rs4646903	rs1048943	rs1056836	rs776746	rs4680
AKU166	TT	AG	GG	GG	AA	**AKU208**	TC	AA	GC	GG	AA
AKU167	TT	AA	GG	GG	AA	**AKU210**	TT	AG	GG	GG	AA
AKU170	TT	AA	GG	GG	AA	**AKU212**	TT	AA	GC	GG	AA
AKU172	TT	AA	GC	GG	AA	**AKU213**	TT	AA	GG	GG	AA
AKU175	TC	AA	GG	GG	AA	**AKU 182**	TT	AA	CC	GG	AA
AKU176	TT	AA	GG	GG	AA	**AKU 183**	TC	AA	CC	GG	AA
AKU177	TC	AA	GG	GG	AA	**AKU 191**	TT	AA	GC	GG	AA
AKU180	TT	AG	GC	GG	AA	**AKU 193**	TT	AA	CC	GG	AA
AKU181	TT	AA	GG	GG	AA	**AKU 194**	TT	AA	CC	GG	AA
AKU182	TT	AA	GG	GG	AA	**AKU 195**	CC	AA	GC	GG	AA
AKU183	TC	AG	GG	GG	AA	**AKU 196**	CC	AA	GC	GG	AA
AKU185	TT	AA	GC	GG	AA	**AKU 197**	CC	AA	CC	GG	AA
AKU187	TC	AA	GG	GG	AA	**AKU 198**	TC	AA	CC	GG	AA
AKU188	TT	AA	GG	GG	AA	**AKU 199**	TT	AA	GC	GG	AA
AKU189	TT	AA	GG	GG	AA	**AKU 200**	TC	AA	GC	GG	AA
AKU190	TT	AA	GG	GG	AA	**AKU 201**	TT	AA	GC	GG	AA
AKU191	TT	AA	GC	GG	AA	**AKU 202**	TT	AA	CC	GG	AA
AKU192	TT	AA	GG	GG	AA	**AKU 204**	TT	AA	CC	GG	AA
AKU194	TT	AA	GG	GG	AA	**AKU 205**	TC	AA	GC	GG	AA
AKU196	CC	AA	GC	GG	AA	**AKU 206**	TT	AA	GC	GG	AA
AKU197	TC	GG	GG	GG	AA	**AKU 207**	TC	AA	CC	GG	AA
AKU198	TT	AA	GC	GG	AA	**AKU 208**	TC	AA	CC	GG	AA
AKU200	TC	AA	GC	GG	AA	**AKU 209**	TC	AA	CC	GG	AA
AKU202	TT	AA	GG	GG	AA						
AKU204	TT	AA	GG	GG	AA						
AKU205	TT	AA	GG	GG	AA						
AKU206	TT	AA	GC	GG	AA						
AKU207	TC	AA	GG	GG	AA						

Appendix VIII: Sanger Sequences for rs1800440 and rs1056836 in exon 3 of CYP1B1

The sequences below contain the forward (F) and reverse (R) raw Sanger sequencing reads of the rs1800440 and rs1056836 SNPs in the CYP1B1 gene. These sequences were obtained from case, control, and benign group participants, with all variants located in exon 3 of CYP1B1.

rs1800440 and rs1056836 sequences of the case group

AKU103_F

CTATCCAAGTTCTCCGGGTTAGGCCACTTCAGTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGT
GTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGGTGACAGGCAAAATG
TTTATCTTGTGATCTGTGTCTGAGGAAACTGCTCGCAGTGTGTCACCATGAGATGCTGTTTGTGTTTTT

AKU103_R

GTCGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAACTTTGATCCA
GCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGACAGTGATGATAGAGGATCTTTCGGC
TCTTTACTTTACCATCCGATCTGATAGACGATAACCCTGATAAACCCTTCACCCTCTTGGTGGCTTCCTGCTCCT
ATCCTGCACGGACCAAGTTATTGTAGATATCC

AKU104_F

TTATTCCACAGACCACTGGTTAGACAAAAACCACAGTTGTCCTTCGGGAATGTGGTAGCCCAAGACAAGAGGT
GTTGGGCAGTGGTGGCATGAGGATAAGTGACAGGCTCA

AKU104_R

GTCGCGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAACTTTGATCC
AGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATAGACTGCG

AKU105_F

CCCATTCATTGGGTTTCATGGATTTCACAGACCACTGGTTAGACAAAAACCACAAGTGTCCTTGGGAATGTGGT
AGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGGTGACAGGCAAA

AKU105_R

AATGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATCGA

AKU106_F

TTTTCCGGTTAGGCCACTTCACTGGGTCATTGATTCACAGACCACTGGTTGACAAAAACCACAGTTGTCCTTGG
GAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGCAAAGGA

AKU106_R

CATCGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGGTGTGATGATAAAACAA
GCTCACCTCATGGTTTACTGCTGGCTATAAAGGATAAATCTGAGACAAATTCACCCTCTTGATGGCTTCCTGCTC
CTATCTCTGACTGACCCACTGATTATAGATATAATTCTTCCCTCCC

AKU109_F

GAGTCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCAGTGGGTCATGATTCACAGACCACTGGTTGACAA
AAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACA
GGCA

AKU109_R

TTACACTTCCCAAGGACACTGTGGTGTTTTATTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCT
AACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGCAGGA
GTGATGATCGCAACAATCTCCTGCTCCAACTGCTGCCCACAGATGCCCCTCGGGCTCCCTGCTACACTGGTGCC
ACTTGTCATATTT

AKU110_F

AAATCCAGACCATAGTTAGACAAAAACCACAGTGTCCTTGGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGC
AGTGGTGGCATGAGGATAAGTTGACAGGCAAAACA

AKU110_R

CACGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATTGAA

AKU112_F

TTATTCCACAGACCACTAGTTAGACAAAAACCACAGTGTCCTTTGGGAATGTGGTAGCCCAAGACAGAGGTGTT

GGCAGTGGTGGCATGAGGATAATTGACAGGCACAACAAAAATACCAAGACCTATTTCTGAGGATTTTGCTGGT
TTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

AKU112_R

ACACGGGGGTTTTTATTCAACCAGTGTGCTGGGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTTGA
TCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAGAAGTGATGAAAAAG

AKU113_F

TTTTCCGGGTAGGCCACTTCACTAGGGTCATTGATTCACAGACCACTGGTTGACAAAAACCACAGTTGTCCTTG
GGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGCAAAAAAAG

AKU113_R

AAGGACACTGTGGTTTTTGGGCAACCAGTGGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAA
CTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAAAAGTGATGA

AKU116_F

ATCCAAAGTTTTTCCGGGGTTAAGGCCACTTTCACTGGGTCATTGATTCACAGACCACTGGTTGACAAAAACCA
CAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGCAA
A

AKU116_R

CTAGTAGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTTGATC
CAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATTGCTTCC

AKU118_F

CCTTCCCTGGTACATAGATTCACAGACCACTCGGTTCGACAAAAACCACAGTGTCCTTCGGGAATGTGGTAGCC
CAAGACAGAGGTCGTTGGCAGTGGTGGCATGAGAGAATTGACAGGCGAAAGCCCGGAAAAAAAAAAAAAAAA
AAAAAAA

AKU118_R

TTTTGGACAACCAGTGGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCG
ATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAAAAGTGATGAAAAAGGGGGACTTGGACATT
GTTTTTTACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAGAAAGT
ATTAGAAACAATCCCGGGATTTCGTTTGGGAAATAAGAAAGTCGTGAGTGTAGTTTC

AKU122_F

AAAATGTTCCTTGCCGGGTTAAGGCCACTTCACTTGGGTCATTGATTCACAGACCACTGGTTGACAAAAACCAC
AGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGGTGACAGGCAAA
ATA

AKU122_R

GTAGAGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAACTTTGATCC
AGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGTTAAACGTGTGCCACT
TCCTGGTTTTTTGTCTGAGCAGATAGACGATAAGCCTGATGCCCCTTCACGCAATTGTGGGCTTCCTGCTCCTAC
CTCTGAA

AKU124_F

ATGAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTGT
CCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGCA

AKU124_R

ATTCCCCAAGGACACTAGTGTTTTTAGGCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCG
GAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAGAAGTGATGA
TATAAAG

AKU127_F

AACAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTGT
CCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAAATTCTT

AKU127_R

CACGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATAGACT

AKU128_F

CGATCCAGATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGT
TGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATA
AGTGACAGGCAAAGAAAAATTCCTATGAGAAGTGAAATGAGGAACTTGCCAGAAGTCTGACAAGCTTGAAGTT
TTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAGAGAAGTACGCC

ATACCCTAAAGGAATAAGATGCCGTGGGGAGAGAGGATGTTTGGGAGATGATGGCCCAGTTTAAGACTGTGGA
AAGGTTGAGAGG

AKU128_R

TTACACTTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCCTAA
CCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGCAGGAGT
GATAGT

AKU133_F

TCGATGTCAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGG
TTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAT
AGTGACAGGCAA

AKU133_R

AGTTGCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATA
A

AKU142_F

AATTCTCCGGTTAGGCCACTTCAGTGGGTGCATTGATTCACAGACCACTGGTTGACAAAAACCACAGTTGTCCT
TAGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGCAAA

AKU142_R

TTGGTTTTTAGACAACCATGTCTAGGGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAACTTTGATCCAG
CTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAAAAGTGATGAAAAAGG

AKU149_F

CCCCTTTCATTGGGTCCATAGATTACACAGACCACTGGGTTAGACAAAAACCACAGTTGTCCTTGGGAATGTGG
TAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGCAAAACAAAAAGTGGGGTTTTG
GTGTTGATTGTGGGAAAAAAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

AKU149_R

CTAGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTTGATCC
AGCTCGATTCTTGGACAAGGACGGCTCATCAACAAGGACCTGACCAGAGGAGTGATGATACGAATGGAAAGGG
TGTTTGTTTTTTTTTAGGAAAAAAATAAAAAGGAGGGTTGATAAATGGGATGAAGAAAAGTAATAAGAGGGGGG
GGGGGGGGGGGGGGGGGGGG

AKU152_F

AATCATTCTCCGGTTAGGCCACTTACACTGGGTCATGATTCACAGACCACTCGGTTCGACAAAAACCACAGTCG
TCCTTGGGAATGCGGTAGCTCCAAGTACAGAGGTCTCTGGCAGTGCTCGGCATCGAGGTAAATTGACAGGCAA
AACAGG

AKU152_R

AAGGACATTGTGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTT
GATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAGAAGTGATGAAGA

AKU155_F

TTTCAGATCGAGCTGGATCAATTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGAC
AAAAACCACAGTGTCCCTTCGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAG
TGACAGGCAAAAATAATATATGAGTGCATTAATAGGATAATAATTATCCCCATGACTGTCGGCGGTCTTGATTT
TTTT

AKU155_R

ATTCCCCAGGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGG
AGAACTTTCGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAAGAAGTGATG
ATAAAAGAAAAAATTTTTTTAACCCCCCACCCCCCTCCACACCCTACAACCTCTCCCTATTAATCCTCTATTTTT
TTTTT

AKU158_F

CTGTGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAA
CCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGG
CAA

AKU158_R

ACTTCCCAGGACACTGTGTTTTTAGGCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGA
GAACTTTCGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAGAAGTGATGAA
A

AKU164_F

CCCCATTCATGGGTACATGGATTCACAGACCACTGGGTTAGACAAAAACCACAGTGTCCCTTCGGGAATGTGGT
AGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGCAAAAGCTAAGTCCTGGGATCTTT
GTTTGAGGGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTTTTT

AKU164_R

CATAGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATCGA

AKU173_F

TCCTGTCCAGAATCCGAGCTGGGACAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCAC
TCGGTTCGACAAAAACCACAGTGTCCCTTCGGGAATGTCGGTAGCCCAAGTACAGTAGGTCGTTCGGCAGTGGT
GGCATGAGGATAAGTGACAGGCAGAATTAATGACCCCCCCCCCCCCACCTGATGGGCCCCCCCCCCCGCCGTG
AGGAGGATGTTTTTTTTTTTTTTTT

AKU173_R

GATATACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCG

GAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAAGAGTGATGA
TAGAAATATAAAAATAAATTAAACCCCCCACCCCGCCCCCACAGCGGAACCCAGGAGGAGGATACGGCCGGA
GGTTTTTTTTTTTTTTTTTTT

AKU211_F

ATAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAAC
CACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGC
AAACTATTGCCCTACGACTCTATGATATGTGTAATACTTGCTCAGCATGGACTGTTTTCTCCTGGTAGGATGGCT
AT

AKU211_R

TACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGAGA
ACTTTGATCCAGCTCGATTCTTAGGACAAGGACGGCCTCATCAACAAGGACCTGACCCCAAAAAGTGATGATA
AAATCCCTTTCCTGTCCGTCTCTTCATCCCCCCCATATTGATTCCCTCTTACCCTACCTCTCCTATTCAAAAAAAA
AAAAAAGGGCTCTTTCTGCCACACTCCTAGGCATTTTAAGTTACCAGAAAAAGGCTACTCGTTCAGCGTGCTTA
TTCCGAACTTGCCATTAAGTTTATTTCTGCGCTAAGACCTGA

KAIC004_F

AGATAGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACC
ACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGCA
AAATATTATCAAATAAAGTCTATACATGAGGAACTTGCAGAAAACTTACCTCCATGAAGAAGGCTGAATGGTTT
TTTTTTTTGCAAAGAAAAGAAACAGACCTGCAATAGATGAAGGGAGA

KAIC004_R

TCCCCAAGGACACTTGTGGGTTTTTCGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCG
GAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTCGACTCAGAAAAAGTGA
TGAACCAAATGTGATAGCGCTGTATCGAGGTCTTTTGCCGGCCAGATAGACGATAAAGCTCATAGACGTTTCGC
TCTCTTGATGGCTTCTTGCTCGTGGTGGGGGGGATCCACTTATT

KAIC005_F

CGCGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAG
TGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAA

KAIC005_R

ACACTGGGGGGTTTTTATTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTT
GATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAGGAGTGATGATAAGACGGG
AGCTTTTGTGTTTTTTTCTCCCATCGATCAATAGTCATAACATCATAAGATACCCTCTGAGCATTGCTTTTTTTTT
TGTTTTTACCAAAAGCGTGTATGGCCTGAGCTGCCTACTTAT

KAIC006_F

AAAGTTGGAACAAAAGTTTCTCCCGGGGTTAGGCCACTTACACTGGGGTCATGATTCACAGACCACTGGTTGAC
AAAAACCACAGTGTCCCTTCGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAG
TGACAGGCAAATTGTTTATCTTATACCTCTGTGATATGAGAGAAATTGCTCAGCAGGCCTGACACGGCTGAGGA
TGATGTT

KAIC006_R

GATTGCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGTTA
GATAGCATCTTTTGTCTCTTAACTTTACCATCCGATCTCATATCGCTAAAGCGCCGAACCCCTTAACTCTCCTTG
TTTCACATTGCACTCCTCCTGCAGGGACTCAGTGATTGTAGATATAG

KAIC007_F

TCACTGGACAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCAC
AGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGCAAA
ATATTTTTCTTATCCCTCACAAAACCGCTGCTCCTCCCCTGAGGCCGGATGGGGATGAGGAGTTTTTTTTTTTTT
TTTTT

KAIC007_R

CACAGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAACT
TTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGATAGTGATGATTGGGAAA
TTCTTCTCTTCGTTTATTCCGATCGATAGAGGATAAACCTCCTAAAGCTTCCCCCTCCCGATGTCGTTTGCTCCCA
ACAGCAGTGAAAGTGATTGTAGATATCCGTCGTCGCTCATGCACTAGATTGCACGTGTCTCCTATGACAAACTC
AGAAGTCCCAGAGGTCCAGCTCAAGATGGGGGGGCCG

KAIC008_F

CGATCAAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAG
TGTCCTTCGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTTGACAGGCAA

KAIC008_R

TTCCCAAGGACACTGTGGTTTTTGTTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGA
GAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGAACCACAGAAGTGATGAT
AAAAAAGTAGTTGTCTATGAGACCGCCTCGCCTCCCGCCCGCTTCTCATCAGGATTACGTGAATTCTTGTGCTT

KAIC011_F

AAAGCTGGAACAAATTCTCCGGGTTACGGCCACTTCCACTCGGGTCATGATTCACAGACCACTCGGTTCGACAA
AAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAATTGAC
AGGCAAAATAACTTTTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
CAAAAAAA AAAAAAAAA

KAIC011_R

CACTACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATC
AA

KAIC012_F

AATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAA
ACCACAGTCGTCCTTGGGAATCGTCGGTAGCTCCAAGTACAGTGGTGTTCGGCAGTGGTCGGCATGAGGTATAA
TTGACAGGCTAAGA

KAIC012_R

CACCACATTTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCT
AACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGCAGAA
GTGATGATAGAAAGATTGATTTCTTCCTGCTCTTTTTCCCGCACAGAGACGATAACCCTCATGACCATTCCTCTC
TTTTTTCATTTGCTTT

KAIC014_F

AATCGAGTGGACAAATTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAACC
ACAGTGTCCTTCGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCCATGAGGATAAGTGACAGG
CAAAATTTTTTTTTTTTCCCCCCCCCCCCCCCCCCCCCCCCTCCCCGAACCCTCTTCAAACATGTTTTTTTTTTTTTT
TT

KAIC014_R

GATTACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGCAGGAGTGATGA
TAAAGAATTTTATAGTCTGTGCCTATCTGGGGTACTTGAGGCAAGCGCAAGAAGATCCGTCTTTATTTTTTTTT
TTTTT

KAIC016_F

ATACCAATTCTCCGGGTTAGGCCACTTACACTGGGTCATGATTCACAGACCACTCGGTTCGACAAAAACCACAG
TGTCCTTCGGGAATGTGGTAGCCCAAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGTAGTAAGTGACAGGCA

AGAAAGTTTATCCTTATGCTCTCCCCCTCCCAGCTCCTCCCTCGCACCCCCTGCTCGCGCCCCGAACAACGATAT
AATGTTTTTATTTTTTTTTTTT

KAIC016_R

TTGCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGA
GAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAAAAGTGATTAAA
ACGCCCCTCTTACTCTTTTCTTTTTTTCCCGCAGATAGTGATACCCTCCTTCCCATCGCCTCTTGTTCTTTTCTCCC
CCTCAGGATCATTTTCTCCTAGTGTTGCATGCATTGAATTTACTAAAATAAACTTATTCCCCAGTGTAGTTCCGG
CGATTCCAAGATTTTTTATCACTAGTGTTCTATTGTAGTTTTTTTTGTAGTCCTTCGTTC

KAIC018_F

AAATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCAGTGGGTCATGATTCACAGACCACTGGTTGACAA
AAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGAC
AGGCAACCTTTATCCTAGGACTCTATAATATGAGGAATAGTGACAGGCAAA

KAIC018_R

CACACATTCCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAAC
CCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAAAGTGAT
GATAAA

KAIC021_F

ATCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCA

CAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAA
A

KAIC021_R

AAGACTTGTGGTTTTTGACAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTTG
ATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAAAAGTGATGAAG

KAIC023_F

TCAGAATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGA
CAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGT
GACAGGCAAAATCAATCTTGCCAAAAAAAAAAAAAAAAAAAAACACCCAAAACAACGAAAAGAGGACGGG
TTTTTTTTTTTTTTT

KAIC023_R

GAGCACATTCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCCTAA
CCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGCAGAGTG
ATGATAGA

KAIC011_R

CACTACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATC
AA

KAIC012_F

AATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAA
ACCACAGTCGTCCTTGGGAATCGTCGGTAGCTCCAAGTACAGTGGTGTTCGGCAGTGGTCGGCATGAGGTATAA
TTGACAGGCTAAGA

KAIC012_R

CACCACATTTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCT
AACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGCAGAA
GTGATGATAGAAAGATTGATTTCTTCCTGCTCTTTTTCCCGCACAGAGACGATAACCCTCATGACCATTCCTCTC
TTTTTTCATTTGCTTT

KAIC014_F

AATCGAGTGGACAAATTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAACC
ACAGTGTCCTTCGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCCATGAGGATAAGTGACAGG
CAAAATTTTTTTTTTTTCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCGAACCCTCTTCAAACATGTTTTTTTTTTTTTT
TT

KAIC014_R

GATTACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGCAGGAGTGATGA
TAAAGAATTTTATAGTCTGTGCCTATCTGGGGTACTTGAGGCAAGCGCAAGAAGATCCGTCTTTATTTTTTTTT
TTTTT

KAIC016_F

ATACCAATTCTCCGGGTTAGGCCACTTACACTGGGTCATGATTCACAGACCACTCGGTTCGACAAAAACCACAG
TGTCCTTCGGGAATGTGGTAGCCCAAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGTAGTAAGTGACAGGCA
AGAAAGTTTATCCTTATGCTCTCCCCCTCCCAGCTCCTCCCTCGCACCCCCTGCTCGCGCCCCGAACAACGATAT
AATGTTTTTATTTTTTTTTTTT

KAIC016_R

TTGCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGA
GAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAAAAGTGATTAAA
ACGCCCCTCTTACTCTTTTCTTTTTTTCCCGCAGATAGTGATACCCTCCTTCCCATCGCCTCTTGTTCTTTTCTCCC
CCTCAGGATCATTTTCTCCTAGTGTTGCATGCATTGAATTTACTAAAATAAACTTATTCCCCAGTGTAGTTCCGG
CGATTCCAAGATTTTTTATCACTAGTGTTCTATTGTAGTTTTTTTTGTAGTCCTTCGTTC

KAIC018_F

AAATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCAGTGGGTCATGATTCACAGACCACTGGTTGACAA
AAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGAC
AGGCAACCTTTATCCTAGGACTCTATAATATGAGGAATAGTGACAGGCAAA

KAIC018_R

CACACATTCCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAAC
CCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAAAGTGAT
GATAAA

KAIC021_F

ATCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCA
CAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAA
A

KAIC021_R

AAGACTTGTGGTTTTTGACAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTTG
ATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAAAAGTGATGAAG

KAIC023_F

TCAGAATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGA
CAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGT
GACAGGCAAAATCAATCTTGCCAAAAAAAAAAAAAAAAAAAAAACACCCAAAACAACGAAAAGAGGACGGG
TTTTTTTTTTTTTTTT

KAIC023_R

GAGCACATTCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCCTAA
CCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGCAGAGTG
ATGATAGA

KAIC024_F

CGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAACC
ACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCA
A

KAIC024_R

AAGGCACACTTACCCAAGGACACTGTGGGTTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGG
CCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGCA
GAGTGATGATCCA

KAIC025_F

ACCCTGTCCAGAATCCGAGCTGGGACAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCAC
TCGGTTCGACAAAAACCACAGTGTCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGA
GGATAAGTGACAGGCAGAACAATCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTTCCCCCC
CCCCGGCCATTTTTTTTTTTTTTTTTTTTTT

KAIC025_R

AAGTTCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGCAAGAGTGATGT
GTGACGGGAAGTGGCTACAACCCCACACACACACACAAAAAACCCCCAACCAAACGCCGATACAACCCGTCCT
CGACGTAGTTCTTTTTCTT

KAIC026_F

CAATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAA
AAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACA
GGCAAAAG

KAIC026_R

GATTACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAAGTGATGAT
AGAAAGCGGGTACTTCTCCCACCTCGACAGGCCAACACTCTACTCGAGCCGCACTTCCCGATTACCTATAACAA
TTTTATTTGTATCGTTTGTGGGTATACACAATCATTATAGACCTAAGACGTGGTCATTGTCTGTTTCGTACAGTA
GT

KAIC027_F

ATCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCAC
AGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAAA
GAG

KAIC027_R

TTCCCAGGACACTGTGGTTTTTAGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGA
GAACTTTGATCCAGGTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAGAAGTGATGATA
GATAATGTAAAAAAAAAAAAACCGCCCCATCCCACCCCACACCGACCCCACCCTACCCCTCCCTATGCCAGCCG
ACTTTGTATTCTGTGCTTT

KAIC029_F

CGATGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAACC
ACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAA
A

KAIC029_R

ACCCTTCCCAGGACACTGTGGTTTTTGGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCC
GGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGTACCTGACCAGAAAAGGTGA
TGATCTAAAATACACCCCCCCCCCCACCCACTCCCCCACCCCACACCCATCAACCGCTATCCCCCCCTTTACCCC
TTCTCATCTCTCTATTTATATATTAAATTATTAAGAGGGGGGTCTCACTCGATTCTTCTATAGGACATTGCTAGT
CTATGTCTCATATCAGTGACCCCTGTGGCGGACCAGTGTTTGACTTCTGTTTCTCCAATGACCTGCCCATGATGT
A

KAIC030_F

TCCTGTCAGATCGAGCTGGACAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTT
GACAAAAACCACAGTGTCCTTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAAA
AGTGACAGGCAAAAT

KAIC030_R

AATATGCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCG
GAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGTT
AGA

KAIC032_F

AGATGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCA
CAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAA

KAIC032_R

CCCTGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATAAACC

KAIC033_F

CCCTTCATTGGTGCATAGATTTCACAGACCACTGGGTTAGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGC
CCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGGTGACAGGCAAACACAAAACTTTTTTTTTTTATTTT
TTTTTTTTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

KAIC033_R

AAGGACATTGGGTTTTTAGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGAGAACTT
TGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAAAAGTGATGAAAAT

KAIC035_F

TGATGAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGT
GTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAAAGTG
C

KAIC035_R

CCAGGACCGTGTTTTTAGTCAACCAGTGGGTCTGTGAATCATGACCCACTGAAGTAGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTAGGACAAGGATGGCCTCATCAACAAGGACCTGACCCAAAAGTGATGAAAA

KAIC041_F

CGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCAC
AGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAAAA
CTTTATCCTAGGACTCTTTGATATGAGGAATAGTGACAGGCAATTGGTGTTTTCGTGCTGCTGTGTGACTGTAGA
ATGAGC

KAIC041_R

TCCCAAGGACACTGTGGTTTTTAGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGA
GAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAGAAGTGATGATA
GAAAATCATATCTTCTTCTTCCTTTACCACCCGGCCACTTAGACGAAAACTGCAAAATTTCCCCTCTCGAGCTTT
TTT

KAIC046_F

TGTTACTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTGTCTCT
TGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGCAAA

KAIC046_R

CTAGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGAGAACTT
TGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATAGATA

KAIC047_F

CATGCAAAGTTCTCCGGGTTAGGCCACTTCACGTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAG
TGTCCTTTGGGAATGTGGTAGCCCAAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCA

KAIC047_R

AACGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGACAGTGATGATAAAAC

KAIC051_F

CAGGGATCAAAGTTCTCCGGGTTAGGCCACTTCACGTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCA
CAGTGTCCTTGGGAATGTGGTAGCCCAAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCA
A

KAIC051_R

CGCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGAGA
ACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGACAGTGATGATAAAA
AGCACCCGGTCCATCGACGAGCTTGAGACTTTATTCGGGAAGGCCCATTACTATGTCCGACCCCACAACTCCCT
GCAACAA

KAIC052_F

AGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCA
CAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAA

KAIC052_R

CTACACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAA
CCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAAGTG
ATGATAGAAGTATGTACCTACTAAACATCTATAGCCAGCAGATAACGATACCGGGAGAAATATTCCTCTTTTTT
TTTTTTTTTTTTCA

KAIC057_F

TTATTCCAGACCACTGGTTAGACAAAAACCACAGTGTCCTTCGGGAATGTGGTAGCCCAAAGACAGAGGTGTT
GGCAGTGGTGGCATGAGAGAATTGACAGGCAC

KAIC057_R

TTGGTTTTTAGACAACCAAGTGGTCTGGGAATTCATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTTGATCC
AGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAAAAAGTGATGAAGAA

KAIC059_F

GGGGCCACTTCAGTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTA
GCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAAAACCTGTAATTCTCCG

KAIC059_R

TTTGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAAAATGATGATTAAT

KAIC060_F

TCTTGTCAGATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCAGTGGGTCATGATTCACAGACCACTGG
TTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAA
AAGTGACAGGCAAAACAAGACCCCCCCAAGTGGTTAACACCCCCGCCCTCGGGGCTGGCGATTGCGGCAGCTT

KAIC060_R

TCGGCTACACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGC
CTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGCAG
GAGTGATGATCACATGACCAAAAAAAATGAAAATCAAACGATTACAGTTCGTCGGATAGGCTCTCCTGGCGCC
CGTTTTT

KAIC024_F

CGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAACC
ACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCA
A

KAIC024_R

AAGGCACACTTACCCAAGGACACTGTGGGTTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGG
CCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGCA
GAGTGATGATCCA

KAIC025_F

ACCCTGTCCAGAATCCGAGCTGGGACAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCAC
TCGGTTCGACAAAAACCACAGTGTCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGA
GGATAAGTGACAGGCAGAACAATCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTTCCCCCC
CCCCGGCCATTTTTTTTTTTTTTTTTTTTT

KAIC025_R

AAGTTCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGCAAGAGTGATGT
GTGACGGGAAGTGGCTACAACCCCACACACACACAAAAAACCCCCAACCAAACGCCGATACAACCCGTCCT
CGACGTAGTTCTTTTTCTT

KAIC026_F

CAATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAA
AAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACA
GGCAAAAG

KAIC026_R

GATTACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAAGTGATGAT
AGAAAGCGGGTACTTCTCCCACCTCGACAGGCCAACACTCTACTCGAGCCGCACTTCCCGATTACCTATAACAA
TTTTATTTGTATCGTTTGTGGGTATACACAATCATTATAGACCTAAGACGTGGTCATTGTCTGTTTCGTACAGTA
GT

KAIC027_F

ATCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCAC
AGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAAA
GAG

KAIC027_R

TTCCCAGGACACTGTGGTTTTTAGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGA
GAACTTTGATCCAGGTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAGAAGTGATGATA
GATAATGTAAAAAAAAAAAACCGCCCCATCCCACCCCACACCGACCCCACCCTACCCCTCCCTATGCCAGCCG
ACTTTGTATTCTGTGCTTT

KAIC029_F

CGATGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAACC
ACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAA
A

KAIC029_R

ACCCTTCCCAGGACACTGTGGTTTTTGGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCC
GGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGTACCTGACCAGAAAAGGTGA
TGATCTAAAATACACCCCCCCCCCCACCCACTCCCCCACCCCACACCCATCAACCGCTATCCCCCCCTTTACCCC
TTCTCATCTCTCTATTTATATATTAAATTATTAAGAGGGGGGTCTCACTCGATTCTTCTATAGGACATTGCTAGT
CTATGTCTCATATCAGTGACCCCTGTGGCGGACCAGTGTTTGACTTCTGTTTCTCCAATGACCTGCCCATGATGT
A

KAIC030_F

TCCTGTCAGATCGAGCTGGACAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTT
GACAAAAACCACAGTGTCCTTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAAA
AGTGACAGGCAAAAT

KAIC030_R

AATATGCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCG
GAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGTT
AGA

KAIC032_F

AGATGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCA
CAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAA

KAIC032_R

CCCTGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATAAACC

KAIC033_F

CCCTTCATTGGTGCATAGATTTCACAGACCACTGGGTTAGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGC
CCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGGTGACAGGCAAACACAAAACTTTTTTTTTTTATTTT
TTTTTTTTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

KAIC033_R

AAGGACATTGGGTTTTTAGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGAGAACTT
TGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAAAAGTGATGAAAAT

KAIC035_F

TGATGAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGT
GTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAAAGTG
C

KAIC035_R

CCAGGACCGTGTTTTTAGTCAACCAGTGGGTCTGTGAATCATGACCCACTGAAGTAGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTAGGACAAGGATGGCCTCATCAACAAGGACCTGACCCAAAAGTGATGAAAA

KAIC041_F

CGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCAC
AGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAAAA
CTTTATCCTAGGACTCTTTGATATGAGGAATAGTGACAGGCAATTGGTGTTTTCGTGCTGCTGTGTGACTGTAGA
ATGAGC

KAIC041_R

TCCCAAGGACACTGTGGTTTTTAGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGA
GAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAGAAGTGATGATA
GAAAATCATATCTTCTTCTTCCTTTACCACCCGGCCACTTAGACGAAAACTGCAAAATTTCCCCTCTCGAGCTTT
TTT

KAIC046_F

TGTTACTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTGTCTCT
TGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGCAAA

KAIC046_R

CTAGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGAGAACTT
TGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATAGATA

KAIC047_F

CATGCAAAGTTCTCCGGGTTAGGCCACTTCACGTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAG
TGTCCTTTGGGAATGTGGTAGCCCAAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCA

KAIC047_R

AACGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGACAGTGATGATAAAAC

KAIC051_F

CAGGGATCAAAGTTCTCCGGGTTAGGCCACTTCACGTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCA
CAGTGTCCTTGGGAATGTGGTAGCCCAAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCA
A

KAIC051_R

CGCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGAGA
ACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGACAGTGATGATAAAA
AGCACCCGGTCCATCGACGAGCTTGAGACTTTATTCGGGAAGGCCCATTACTATGTCCGACCCCACAACTCCCT
GCAACAA

KAIC052_F

AGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCA
CAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAA

KAIC052_R

CTACACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAA
CCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAAGTG
ATGATAGAAGTATGTACCTACTAAACATCTATAGCCAGCAGATAACGATACCGGGAGAAATATTCCTCTTTTTT
TTTTTTTTTTTTCA

KAIC057_F

TTATTCCAGACCACTGGTTAGACAAAAACCACAGTGTCCTTCGGGAATGTGGTAGCCCAAAGACAGAGGTGTT
GGCAGTGGTGGCATGAGAGAATTGACAGGCAC

KAIC057_R

TTGGTTTTTAGACAACCAAGTGGTCTGGGAATTCATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTTGATCC
AGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAAAAAGTGATGAAGAA

KAIC059_F

GGGGCCACTTCAGTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTA
GCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAAAACCTGTAATTCTCCG

KAIC059_R

TTTGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAAAATGATGATTAAT

KAIC060_F

TCTTGTCAGATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCAGTGGGTCATGATTCACAGACCACTGG
TTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAA
AAGTGACAGGCAAAACAAGACCCCCCCAAGTGGTTAACACCCCCGCCCTCGGGGCTGGCGATTGCGGCAGCTT

KAIC060_R

TCGGCTACACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGC
CTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGCAG
GAGTGATGATCACATGACCAAAAAAAATGAAAATCAAACGATTACAGTTCGTCGGATAGGCTCTCCTGGCGCC
CGTTTTT

rs1800440 and rs1056836 sequences of the control group

AKU182_F

TATCGAGTTGGACAAATTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAA
CCACAGTGTCCTTCGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAATTGACAG
GCAAAAAAAGGCTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAT
TTTTTTTTTTTTTTTTTTTTT

AKU182_R

TTACACTTCCCAGGACACTGTGGTTTTTAGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAAC
CCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAGAGTGAT
GATAAAC

AKU183_F

CGACCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACC
ACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAA
AGGTTTATCCTAGGACTCTATAATATGAGGAATAGTGACAGGCAAGGACTGATTACGCCTTGAAGAATGACTA
GGCAACA

AKU183_R

ACACGGTGGTTTTTAGACAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTTGA
TCCAGGTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAAAAGTGATGTGTCGAAAGTAC
GCTTACATTTGGTATTTTGCCAGCAA

AKU184_F

TTACTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCA
CAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAAA
ATAATATCCTACGACTCTATAATATGAGGATATGACAGGCATGGTCTGATTCCCCATGAAGGATGCTGGCACTG
AGGTATT

AKU184_R

GACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGA
ACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGACAGTGATGTTTGAAA
AGATGTCGCTCGGTTTTTATCCCCATCTGATAGTCGATACCCCTCCTTAACCCCTCACCTCTCCCGGTGTCTCATT
GCTTCCTCCCGCACGGAGACACTTATTTTCAGACATAGGGGGTGGGTCGG

AKU191_F

GATGCAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTG
ACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAG
TGACAGGCAAAAATAATGCCTATGACTCTATAATATGAGGAATAGTGACAGGCAGTAGGGTCAACATTGGTGT
CCAGAACGTTATGGG

AKU191_R

AATTACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGACAGTGATGATA
GAATAGAATTCTTCTGTGGTTTAAAAAACGGCCAGATAGAGGATAAACCTGATAGACATTCGCCTCTTGGTGGC
TTCATGCTCCCTT

AKU193_F

CTCCATGTCAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTG
GTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGA
TAAGTGACAGGCAA

AKU193_R

AAGGGACACTTGTGGTTTTTGTGCAACCAGTTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGA
ACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCTCATCAACAAGGACCTGACCACGGAAGTGATGATAAAT
ATTTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCACCCCCCAAATGATTTTTTTTT
TTTTTTTT

AKU194_F

TGGATTGCACAGACCACTGGTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTT
GGCAGTGGTGGCATGAGGATAGTGACAGGCAATATCTTTATCCTAGGACTCTATAATATGAGGAATAGTGACA
GGCA

AKU194_R

TTTTGGACAACCAGTGGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCG
ATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAAAAGTGATGAGTGAATCGTTGTCACTTTCTG
GTTTTTTGACCAGCAGAGAGATGATAAACCTGATACAACATCACCCTCTTGATGCTTCCTGCTC

AKU196_F

TAAGAGCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACA
GTGTCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGCCCGCACCCTT
ACGAATCCTAGGACTCTATAATATGAGGAATAGTGACAGGCAGGGC

AKU196_R

CCCCAAGGACACCGTGGTTTTTAGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGA
GAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAGAAGTGATGAGT
GAATAATGTATATGGTTCTGGCTTTTTTGACGAGCACAGAGATGGATAACCCTGATTCACCATCACCCTCTTAC
GGCTTCTGCTCCCTCAGCAGAAAAAGTATT

AKU197_F

ATAAAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTG
TCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAAAATAC
CTG

AKU197_R

CACGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAACT
TTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATAGATGATT
GCTCACATCCTGGTATTTTTGATCACCTGAGAGATGATAACCCTGATAGACCTTCCTCTCTTGATGTCTTCTTGC
TCCCCT

AKU198_F

CTAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACC
ACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGCA

AKU198_R

GGCTACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGATAGTGATGATA
GA

AKU199_F

CCGTTCATGTCAGATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCA
CTGGTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGA
GGAATAAGTGACAGGCAAAA

AKU199_R

TGGTTTTTAGGCAACCAGTGTACTAGAGAATTCATAGACCCACTGAAGTGGCCTAACCCGGAGAACTTTGATCC
AGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAGAAGTGATGATCAAAAG

AKU200_F

AATGCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTG
TCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAAAATATG
TC

AKU200_R

CCATGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGTTAGACGG

AKU201_F

TCGTTACTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTGTCCT
TGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAAAAACCC

AKU201_R

TCGACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGA
GAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGATAGTGATAATTAA

AKU202_F

AACCAAATTCTTCCGGGTTAGGCCACTTACACTGGGTCATTGATTCACAGACCACTGGTTGACAAAAACCACAG
TGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTTGACAGGCAAAGT
CTTTCGCATAGGACTCAATAAAAAAAAAAAGAACGAAAAGCAAAAAAAAAAAAACAAACGCCAA

AKU202_R

CATGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGACAGTGATGATAAAATA
ATTTCTTCTCCTGGTTAGTATCCGATAGATAGTCGATAACCCTCATACCCCTCCACCCTCCCGGTGCCGACTGAT
ATAAA

AKU204_F

AGCATACGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTG

TCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAAAGAGTG
T

AKU204_R

ACATGGGGGTTTTTAGGCAACCAGTTGTGCTGGGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTTG
ATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAGAAGTGATGATA

AKU205_F

AGATATGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTGT
CCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAAATGCTGG

AKU205_R

CAGTAGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTTGATCC
AGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATAAACTATTGTCTGCT
CAATATTTTACAAACTGGGTAGTTATAGGAGAAAGCTTTAATAAAATTCACCCTCTTGATGGCTTCCTGCTCCG
AACTGTGACTGACGAGTGATTGTAGATATGGGCTTGGGTGCTTTT

AKU206_F

AGTTACTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTGTCTCT
TGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAAAGCGCACG

AKU206_R

GTTTTTAGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTC

GATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGACAGTGATGATAGAAACTACACTCCTTGGT
TTATTAGCCGGGCAGATAGACGATAAAGCTGAGAGAAAATTCACCCTCTTGATGGCTTCCTGCTCCTAACTCTG
ACCGGCAAAGTGATTATATATATC

AKU207_F

TAATTCTCCGGTTAGGCCACTTCATTGGGTCATTGATTCACAGACCACTGGTTGACAAAAACCACAGTGTCTCTT
GGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGGTGACAGGCAAA

AKU207_R

AAGGGACACTGTGTTTTTATTCAACCAGTGGTCTGGTAGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGA
ACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAAAAGTGATGATACA
AAAATGGGGGGGGGGGAGTAGAGGATTCTCACAATTAGCGTACACTGACGACCTCATACAGTGCTCTCGGGGT
GGGGGGGGGG

AKU208_F

AAGGTTGGACAAATTCTCCGGGTTAGGCCACTTACACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACC
ACAGTGTCCTTAGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTTGACAGGC
AAAATATATATCTTTATTAAAAAAAAAAGACGAAAACCCCACCCACCCCCCCCACCCGCGGACCGGGGGGGTT
TTTTTTTTTTTTTTTTT

AKU208_R

ACGCTAGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGAGAACTTTGA

TCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGTTAGAATGCTTTTG
CTTCATAGCTCAAACATCCCATCTGAGAGACGATAAACAGCCTTGACCCTCCGCTAACCCGGTGTCACATTGGA
CTCCTC

AKU209_F

CGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGT
GTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGGCAAAAAC
C

AKU209_R

TTAGTTTTTAGGCAACCAGTGTCTGTGAATCATGACCCAGTGAAGTGGCCCTAACCCGGAGAACTTTGATCCAG
CTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAAAAGTGATGAAGAAGTAATTTTTTTTT
TTTTTGGTCCAAAAAATGAGAAAAAGAAAAAAAGAAAATAAAAACTCTTTTTTTTTTTTTTTGCCCCAAACGGG
GT

rs1800440 and rs1056836 sequences of the benign group

AKU011_F

ATATGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCA
CAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAA

AKU011_R

AACGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGACAGTGATGATAAA

AKU025_F

TTGGGAACAAATTCTCCGGGTTAAGGCCACTTACACTGGGGTCATGATTCACAGACCACTGGTTGACAAAAACC
ACAGTGTCCTTCGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAGTAAGTGACAGG
CAAA

AKU025_R

GATTACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAGAAGTGATGAT
AGA

AKU026_F

GAATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAA
AAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACA
GGCAA

AKU026_R

AACTTGAAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGA
AGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACC

AGAGAGTGATGATAGACATGACCGTTGTCCTGATGATCGTCGAAAAGTTCTGCGGGGGAGGCCACTTCGGGGC
GCCCGACCCCAGGGCCGC

AKU031_F

AGCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTGTC
CTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATACTGACAGGCAAAGGTCTGC
C

AKU031_R

CCAGGACATTGTGGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAAAAAATTGATGAGACT
AA

AKU034_F

ACTTTGGAACAAAGTTTCTCCCGGGGTTAGGCCACTTACAGTGGGTCATGATTCACAGACCACTGGTTGACAAA
AACCACAGTTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGAC
AGGCAAAC

AKU034_R

ATTCCCCAAGGACACTGTGGTTTTTAGGCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCG
GAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAAGAAGTGATG
ATACGAATGTT

AKU039_F

AACGTGTCAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGG
TTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAT
AGTGACAGGCAAATAAGGCATAATAAAAATGATGAAAAATGAGAGACCCCCCAATGACTCCGGTCGCTTCTAA
AAGATGTTTTTTTTTTTTTTTTTTTTTTT

AKU039_R

TAGTACTTGTAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACT
GAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTG
ACCAGCAGAGTGATGATT

AKU043_F

AGCCTTGTCAGATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACT
GGTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGG
AATAGTGACAGGCAAAAGAAAATACACCCAGGTGGATGATAGTCGGGGTCCTGCTGACTGCGGATTTTCTATT
GATCGGGCGGGGGCGGGGGGATTTAAAAAAAAAAAAAA

AKU043_R

ATGTACTTGTAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGT
GAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTG
ACCAGCAGAGTGATGAT

AKU044_F

CGGACAGAATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGT
TGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAAT
AGTGACAGGCAA

AKU044_R

ACTTAGTAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGA
AGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGAC
CAGAGAGTGATGATTGAGGT

AKU045_F

AATTCCAGACCACTGGTTAGACAAAAACCACAGTTGTCCTTGGGGAATGTGGTAGCCCAAAGACAGAGGTGTT
GGCAGTGGTGGCATGAGGATAAGTGACAGGCAAAAGAAACACAAAAAATTGATAAAAGTCTTTTTTTTTTTTTT
TTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

AKU045_R

TTTTTTTGACAACCATTGTGCTGGAGAATTCATCGACCCAGTGAAGTGGCCTAACCCGGAGAACTTTGATCCAG
CTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAAAAGTGATGATAGAAAGGGAGTAAAG
TGATTATAATTTTTCAACCACACCCCCCACCCCCCCCCCCCCCCCCCCCCACCCACCAAAAGGGGGGGGGGGGG

AKU046_F

ATCGAGTTGGACAAATTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAAC

CACAGTGTCCTTCGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAGAATTGACAGG
CAAAACAATAAAGTTCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCATTTTTT
TTTTTTTTTTTTTTT

AKU046_R

TACTTGTAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGA
AGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACC
AGAGGAGTGATGATA

AKU047_F

GATCAGAATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTT
GACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATA
GTGACAGGCAAAAAAATAGGACACGGAGTGAGATATTGGACTATTGGCACGCAAGTGGGATTAGGGTTGTACA
TGGCTTTGGTTGTTG

AKU047_R

GCACCCGTACTGGACTTGTAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATC
ATGACCCACTGAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAAC
AAGGACCTGACCAGAGAGTGATGATAGACGCGTCACTTGGACTTAACTCTTGTTCCCATCTCAGAGTCTATACC
TGCCTTACCCCTCACCTATTCCTGCTTCACATTGCACTCGTCCTGCAGGGAGTCAGTTGTTTCAGACCTACTACG
TTGGTAAGTGGCTAGGCTTGTGC

AKU048_F

AAGCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACAGTGT
CCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAAATTGG

AKU048_R

AAGGACACTGTTGTTTTTATCCAACCAGTGGGTCTGTAGAATCATGACCCAGTGAAGTGGCCCTAACCCGGAGA
ACTTTACGATCCCAGCTACGATTCTTAGGACAAGGATGGCCTCATCAACAAGGACCTGACACAAAAAGTGATG
AAAAAAGGGGGGGGGGGGAAGGGGGGTACCCCCCCCCACACCCTACCCCCCCTCCCCTCCCCCCCCCCCTCTT
AAAAAAAAAAAAAAA

AKU049_F

CGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAAC
CACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCA
A

AKU049_R

CACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGA
ACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGATTA

AKU050_F

TAGCTTGTGCTTCCTGTGCACAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTC
ACAGACCACTGGTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGG

TGGCATGAGGAATAGTGACAGGACAATATAAACACCGTCTCTTTAGCTTCTTCCTCTCTATCACTCTTCTATTAT
CATCCTTTTTTTTTTTTTTTTTTTCAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAACAAACTA

AKU050_R

GGTGTCTGTGCTTGTAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGAC
CCACTGAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGG
ACCTGACCAGCAGAGTGATGATAGAGGAAGAAACATTTTAAATGAGAGGGTTGGGTTCTTTAAAGTGTTGTGG
TGGTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

AKU054_F

TTGTCCAAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGT
TGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAAT
AGTGACAGGCAAAAGAAATAAAATTACAAAAAAAAAAGAAACACCCCACCCCCCCACCCCCCCCCACACACA
AAACACACCTATATTTTTTTTTTTTTTTTAAAAAAAAAAAAACAAAAAA

AKU054_R

TTAAGCTACACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGG
CCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGAG
AGTGATGATAGAC

AKU056_F

ATGATGGCAGATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTG

GTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGA
ATAGTGACAGGCAA

AKU056_R

ATTCCCCAAGGACACAGTGTTTTTAGACAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTCGGCCTAACCCG
GAGAACTTTCGATCCAGCTACGATTCTTGGACAAGGATCGGCCTCATCAACAAGGACCTGACCAGAAAAAGTG
ATGTTACAAAAAAGGAGAAGGGATAAACCTCGGTCTTCCCATCACAGCGTCGCTACCCCTCCGAGCTCGTGCTC
TCTGCCGGCGCACAAAAAAAAAAAAAAAA

AKU059_F

AAACTCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACA
AAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGA
CAGGCAAAGCTATAAATAAACCCAACAAAACCCCCACCCCCCCCCCCCACCCCCCCACCCCCCACCCCCCCTT
TTTTTTTTTTTTT

AKU059_R

AAGTGTGGAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTG
AAGTGGCCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTG
ACCAGCAGGAGTGATGATCGATAAAAATTTAGGGGCAAGATAGCGAGTTGCGATCACCAAAGTCTGTAACGAG
TGCTTGGCCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAA
AA

AKU061_F

GACTGCGCGTCCTTGCACAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCAC
AGACCACTGGTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTG
GCATGAGGAATAGTGACAGGCAAAATAAAAGTAAAAGTGATTGAAAGTTGTGTTACAGGTGACTGCTGCGGTT
CGGTGCTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAA

AKU061_R

GGACGCTAAGGTTACTTGTAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATC
ATGACCCAGTGAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAAC
AAGGACCTGACCAGCAGAGTGATGTCCGAGTAACAAGTAGTGAAATGAGATAGGGAGGAGGTAGATTTTGGG
AGATTAGTGGGGTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAA
A

AKU062_F

GCTAGAGAGAATCCTTGTCAGATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTC
ACAGACCACTGGTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGG
TGGCATGAGGAATAGTGACAGGCAAAAAAAAAGAAAAATGTGCAAAAGTTTGATCATTTCCGCCAATAACTTC
CAATATTTTTTTTTTTTTTTTTTTTTTTTTTTCAAAAAAAAAAAAAAAAAAAAAAAA

AKU062_R

ATTCCCCAGGGACACTGTGGTTTTTAGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCG

GAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAGAAGTGATGA
AAAA

AKU063_F

GATACTGACCCTCCGTGTACAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTC
ACAGACCACTGGTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGG
TGGCATGAGGAAAAGTGACAGGCAAAAGCTAACGTCAAGGGGAAGATAGAGTGGGTTTTTACCTTGTGGGGTT
GTTGTGTTTTGTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAA

AKU063_R

AATAAAAAATAGAGTACCACATTACCCAAGGACACTGTGGTTTTTTGTCAACCAGTGGTCTGTGAATCATGACC
CAGTGAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGA
CCTGACCACAGAAGTGATGATACAAAGTCTGTGTATCTAATCATCTATATTCACACACTTAACTCTATATACATT
ACACTAAATCTTCAGCTCACCTTCATCCCTCC

AKU064_F

TGTCCAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTG
ACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGT
GACAGGCAA

AKU064_R

TTAAGTACACTTACCCAAGGACACTGTGGTTTTTAGTTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTG

GCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACA
GGAGTGATGATAGATGAAAATTTTAAAAAAAAAACCCCCTCCCCCCCTCCTTCTCCGCCAGCCCCGCCCGGGGG
CCGAGTTTTTTTTTTTTTTTTTT

AKU067_F

GGCTGAGCGTCCTTGCACAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCAC
AGACCACTGGTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTG
GCATGAGGAATAGTGACAGGCAAACGAACTCCTCTAAGAGTGTCGATGTTACGGATGTTGTAAGGTTGGGGGG
GGTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAA

AKU067_R

CCGTACGGTGCTTGTAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGA
CCCACTGAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGG
ACCTGACCAGCAGAGTGATGATTGGA

AKU069_F

ATGTCAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGTTCG
ACAAAAACCACAGTGTCCTTCGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAA
GTGACAGGCAAAATAGATATGCGCACCGTCAGCCCCCGGATGCACACGAGACGCC

AKU069_R

AATTACCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCG

GAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGAGAGTGATGAT
TGA

AKU070_F

ACTTGTCAGATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGG
TTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAT
AAGTGACAGGCAAATGAAAAACCGCAGCCTAGCCCTTCACGTAACGACGAGGCAGGATATTTTATTTTTTTTTT
TTTTTAAAAAAAAAAAAAACACAAAAAAACAGAAGAA

AKU070_R

TACTTGTAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGA
AGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACC
AGAGAGTGATGATTG

AKU077_F

TCTGTCAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTT
GACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATA
GTGACAGGCAA

AKU077_R

ATTACTTGTAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTG

AAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCTCATCAACAAGGACCTGAC
CAGAGAGTGATGATA

AKU081_F

TTCCAGATCGAGCTGGACAAGTTCTCCGGGTTAGGCCACTTACACTGGGTCATGATTCACAGACCACTCGGTTC
GACAAAAACCACAGTCGTCCTTCGGGAATGTCGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGT
ATAAGTGACAGGCAAG

AKU081_R

CCTGTCTTGGGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTG
AAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGA
CCAGAGGAGTGATGATAAACATTTCCCCCTGAAAGGGGATTCTTGGTGGAAAGAGATTGTGGGGGGGGGGGGG
GAGGGAGGTTTTTTTTTTTTTTTTTTTTTTTT

AKU132_F

ATCCTGTCAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGG
TTGACAAAAACCACAGTGTCCTTCGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGA
TAAGTGACAGGCAAGATCATTCCCCCCCCAGTGAAGGCCTGAATCTAATTGCCAGGGTCTTTCTACTGTTTTTTT
TTTTTTTTTTTTTT

AKU132_R

CTACTGAACTTGGGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCC

ACTGAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGAC
CTGACCAGAGAGTGATGATAGAGTAGCCGTCCTTGTCCAATATCGAGCTGCAGGAAAGTTCTACGGAAAAAGC
CCAATAGGGGGGAGAGCCCCTAAAAGGCGTTAAACACCCCCCCTTGCCTGGGAAAATTTTTTTATACCACAAA
AAATTTTTTTTTGCCGGTGTTTCTGGGGAACAAAATGGGACTGGGGAAAAAAGT

AKU135_F

GGGGCGTCCTTGTCAGATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGA
CCACTGGTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCA
TGAGGAATAGTGACAGGCAAATAAATA

AKU135_R

TTACTGACTTGGGGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCC
ACTGAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGAC
CTGACCAGAGAGTGATGATAG

AKU136_F

ACTACTGTCAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCAGTGGGTCATGATTCACAGACCACTG
GTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGA
ATAGTGACAGGCAAAAATAAGATCTCTTAAAGTAGTGGTTGTGGGTGTTGGTGGTTGGGGGGGGGGGCTTTTTT
TTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

AKU136_R

CTACTGTCTTGGGGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCA
CTGAAGTGGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCT
GACCAGAGAAGTGATGATACAAA

AKU144_F

TTCCAGATCGAGTGGGACAAAGTTCTCCGGGTTAGGCCACTTCAGTGGGTCATGATTCACAGACCACTGGTTGA
CAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGT
GACAGAGCAAAATCAATTTCCCCCCCC

AKU144_R

TACACACTTCCCAGGACACTGTGTTTTTAGTTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAA
CCCGGAGAACTTTCGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAGAAGT
GATGAAAAA

AKU148_F

TTTCCAGATCGACGCTGGGACAAAGTTCTCCGGGTTAGGCCACTTCAGTGGGTCATGATTCACAGACCACTGGT
TGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATA
AGTGACAGGCAAAACTAATTCCAAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCAC
CCCTTTTTTTTTTTTTTTTTTTTTT

AKU148_R

TTGCAGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGT
GGCCTAACCCGGAGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAG
AAAGTGATGATCA

AKU159_F

GAAATCGACCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAA
AAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGAC
AGGCAAAACAGGGGGGCCCCAAAAAAAAAAAAAAAAAA

AKU159_R

TCTGGGGTACACATTTCCCAAGGACACTAGTGGGTTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGA
AGTGGCCCTAACCCGGAGAACTTTGATCCAGGTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGA
CCAGCAGAGTAATGTTCCGATAAAAATTCTAAACAATCTTGGGGTGCACGGGGTGGGGGGTGGAAAACTGTGT
TGGGTTATATTGGCGGGCGGCCGTTTAAAAAAAAAAAAA

AKU163_F

ACTCCTGTCAGATCGAGCTGGATCAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTC
GGTTCGACAAAAACCACAGTGTCCTTCGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGA
GGATAAGTGACAGGCAAAACATCCTCCTATGCCTCATTATACCGGGACTTGCTGGCTGTCTGGAAGCTTTTTTTT
TTTTTTTTTTT

AKU163_R

CCATTCCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCAGCGTAGTGATGATT
GA

AKU180_F

AGCTGGATCAAAGTTCTCCCCGGGTTAGGCCACTTCAGTGGGTCATGATTCACAGACCACTGGTTGACAAAAAC
CACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAAGTGACAGGC
AAAATAAATAACTTATCCCCCCTCCTACCGAAACATCCCCAGAGGCCGGACGGCCTGAAGATCCTTTTTTTTTTT
TTTTTT

AKU180_R

TCCCAAGGACACCTGTGGTTTTTAGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGGA
GAACTTTGATCCAGCTCGATTCTTGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAAAAAGTGATGTA
TAAAATACCCTCTCGCATCCCCACACCAACATACGATCACAAAGTGTTATTCTGCCATTCCCCACACCTCATCTT
GCTTCTGTTTGCTCTCCTCTTGATGTGAATAGGTCGTTCCTGACATACGACGTCGCTCAGGCTCTATATCGCACA
TGTCTACTATTAGAGCTCAGAGTGCCTGACTTCTACTCAGAAGCTAGTCATGTCAACTCTGTGCAGAATGTGA

AKU182_F

CAAATCGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACA
AAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGA
CAGGCAAA

AKU182_R

AATTCCCCAAGGACACTGGTAGGTTTTTAGACAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTCGGCCCTA
ACCCGGAGAACTTTCGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCACAAAAG
TGATGATAGAAAGATTTTTTACGTCCCCGCGCCCCACCCCCTCGCACGCCTCTCCCCCCCCCCTCCCCCTCCCTAG
CCGCTCTATTTCTTT

AKU185_F

CGAAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAA
CCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGAATAGTGACAGG
CAAA

AKU185_R

ATTCCCCAAGGACACTGTAGGTTTTTATTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCC
GGAGAACTTTGATCCAGCTCGATTCTTAGGACAAGGACGGCCTCATCAACAAGGACCTGACCACAAAAGTGAT
GTATAGAAATTTATTTTGGCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCTCCCTCTCCTCTTTTT
TGTGTTTTTT

AKU191_F

GGGCCGTCACGTGTCACGATCGACGCTGGATCACAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCAC
AGACCACTGGTTGACAAAAACCACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTG
GCATGAGGATAAGTGACAGGCAAACACGAACCCCTAATGCCTGATAACCGGAGCCTGCTCGCTTGATGGGCTA
ACATCCTGCAGGAGCCT

AKU191_R

ACTGACTTGGGGCTACCACATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAC
TGAAGTGGCCTAACCCGGAGAACTTTGATCCAGGTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCT
GACCAGCAGTAGTGATGATAGACTTACCTTTGGGATAAAACTCCCCAGTGATCTTTAGTGTTACCTGCCTTCGG
ACCCTCACTGCTCGATACACCCCCCCCTTTCTTTTTCCCCTCTCTTTCTTTTTTGAGCTTCAGGAGCCATTTATCA
CTCTAGTCGTAATTTAGAGAACAGAACTTTTGATTACTAAACCGCCCCTGTAACGCATGCAAAGGCGGGGATAA
GGGTTTG

AKU196_F

CGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAAC
CACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCA
AAGATTTATCTTGTCCTCGTGTTTTGTGCTATCGCTCTGCTGGTGTGAACGAGCCTTGAAGGATGGCTAAAGCGT
GATGTATTCCGGGACTGCCAAGTAGATTGGCGCCCTGCATGGAC

AKU196_R

GATTCCCAAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGG
AGAACTTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGCAGGAGTGATGA
TAGGAAAAAACTTTATTCATCCCCCCCCAACAACCACCACAAAGTGTTACCCTGCCATACCCAGACCTCTCTTG
TTTTTTTTTTTTTTTTTTTTTAAGGGAAAAGAGTTCTGAATCCACGTGGCAGGTCAGATGCAGCGCGCCCGGAAG
TACAGGCCGACTCCACCCCAAATGTCAGGTGTATCTATAAGAAGGCCGAATGTAAATCTTTGTACCTGTCTGGG
GGGACGGCGGGCAG

AKU198_F

TGAGCTGGATCAAAGTTCTCCGGGTTAGGCCACTTCACTGGGTCATGATTCACAGACCACTGGTTGACAAAAAC
CACAGTGTCCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCA
A

AKU198_R

CCATGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCAGTGAAGTGGCCTAACCCGGAGAAC
TTTGATCCAGCTCGATTCTTGGACAAGGATGGCCTCATCAACAAGGACCTGACCAGACAGTGATGATTGAGGTG
T

AKU200_F

CAGGATCAAGTTCTCCGGGTTAGGCCACTTCACGTGGGTCATGATTCACAGACCACTGGTTGACAAAAACCACA
GTGTCTTGGGAATGTGGTAGCCCAAGACAGAGGTGTTGGCAGTGGTGGCATGAGGATAGTGACAGGCAAATG

AKU200_R

GATTACCCAGGACACTGTGGTTTTTGTCAACCAGTGGTCTGTGAATCATGACCCACTGAAGTGGCCTAACCCGG
AGAACT

Appendix IX: Sanger Sequences for rs10012 and rs1056827 exon 2 of CYP1B1

The sequences below contain the forward (F) and reverse (R) raw Sanger sequencing reads of the rs1056827 and rs10012 SNPs in the CYP1B1 gene. These sequences were obtained from case, control, and benign group participants, with all variants located in exon 2 of CYP1B1.

rs1056827 and rs10012 sequences of the case group

AKU041_F

GGTACCCGAACATATAGCGGCCCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCCGCGCCACCA
GCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCG
CATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGC
CGGACACCACACGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGT
GGATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCG
CCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCA
AACGGGCCCGGGGGCGCGGACCCGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGG
CCAGCACCGACAGGAGTAGCAAGGAGCGTA

AKU041_R

GGGGGCTGTATGTGGGCAGCGGCTGCTGAGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTG
CGTGGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCG
CTACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACC

AGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGC
CGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCA
ACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGT
GGCGCTGCTGGTGCGCGGCAGCGCGGACGGCCGCCTTCCTCGACCCGAGGCCGCTGACCGTCGTGGCCGTGGC
CAACGTCATGAGTGCGGGGGGGGTTTAA

AKU042_F

GGTCTAGACGTTCGCGGCCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTC
GCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATC
ATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGA
CACCACACGGAAGGAGGCGAAGGCCGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGAT
GGCGCGCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGC
GCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAAACG
GGCCCGGGGGCGCGGACCGGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAG
CACCGACAGGATAGACAAAGGGGAGCGTA

AKU042_R

TTACGGCCCGCCGGCTGCTGAGGCAACGGAGGCGGCAGCTCCGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCC
ACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTCGGCG
ACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTG
GTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCAT
GGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTC

ACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCT
GGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTCGTGGCCGTGGCCAACGTCATG
AGTGCGGGGGGGGGGTTTAAAGAA

AKU055_F

GCAACGACGTAGCGGCCCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCG
CGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCA
TGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGAC
ACCACACGGAAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGAT
GGCGCGCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGC
GCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAAACG
GGCCCGGGGGCGCGGACCGGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAG
CACCGACGGAGAGCCCCGGGGGAGCTTAA

AKU055_R

TAAGGGCAGCGGCTGCTGAGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACT
GATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTACGGCGAC
GTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGT
GCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGG
CTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACG
CGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGG

TGCGCGGCAGCGCGGACGGCCGCCTTCCTCGACCCGAGGCCGCTGACCGTCGTGGCCGTGGCCAACGTCATGA
GTGGGGGGGGGGGTTTAAAAA

AKU057_F

AGGATGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGA
GGACTTGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCC
AGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCC
GGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGG
GCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGC
CTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAAACGGGCCCGGGGGCGCGGACCGGAGCTGCCGC
CTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGAGCCAGGGGGGGAGTTTA
ATTAAACAAAAGGAGGGGGCGGGCGGCCTTGTAACCCTCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGGGGCGGGCGGCCGGCCCGGGCCGCCGGGGGGCCCGGGGGGCGCGGCGGCGCCGGGGGGGGGG
CGCCGGGCGCGGGGCGGGCCGGGCCGGCCGGGGCGGCGCGCCGGGGGGGGGGGGCGGGGGCGGGGGG

AKU057_R

GCCGCGGCTGCTGAGGCACGGAGGCGGCGCTCCGGTCCGCGCCCCCGGCCCGTTTGCGTGGCCCTGATCGGAA
ACGCGGCGGCGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGAT
CCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGC
TCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCA
CTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCG
CGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCA

GCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTCGTGGCCGTGGCCAACGTCATGAGTCGGGGGGG
GTTTCAAA

AKU060_F

GACGAACGGTCGCGGCCTCGGGTCGAGGAAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCG
CGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCA
TGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGAC
ACCACACGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATG
GCGCGCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCG
CCAGGCGAGCGAAAGAGAGGTGACCCGCCTGCCCCACCGCCCCCGGCGTTCAGATCAGTGGCCACACACACGG
GCGCGGGGGCGCAGACCGGAGCTGCCGCCTCCTTCGCCTCCTTCCCCCGCCCCACCACCTGCACGCTGGCCACC
ACCAACGGATAGTAGCATAATCATACGTGTTAAACATTTAATCTTCATAGTATGAGTGAAGTTGAGATGATGGT
TATAGAGTTGTAGTCTGTAGGTAGTTTGGATAAGTATATTGGATCTTGTGGTGGATAAAATAGGTTGACAGAGA
AAGAATAATGTACGTTGTCGATTGTATATGTGGTAAGAACTAATTGGCATTGGGTGGGTTAACATGAGTTGACG
GGTTTTAAAGACGGTATTACAGGAGATCTCTATGACATATGATGAGTTGTTTATCACAGATAAAAATGTGTTGG
ATGGGCAAGTAGAGAATGCAGGAGAATTCGAACATGGTAGCTATTACAGGCGTGTGGGTAGA

AKU060_R

TCTCCGTGTGTGGGCAGCGGCTGCTGAGGCAACGGAGGCGGCAGCTCCGGTCAGCGCCCCCGGGCCCGTTTGC
GTGGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGC
TACGGCGACGTTTTCCACCTCCGCCTGGGCAGCTGCCCCATATTGGTGCTGAATGGCGAGCGCGCCACCCACCA
CGCCCTGGTGCAGCAGGGATCGCCCTTCGCCCACCGGCCGCCCTCCGCCCCCTCCCGTGGGCCGGCAGGCGGCG

GCACCATGCTTTTCGCCCCCAACTCGAACACCAGCACGGGGAACTGCCGCAAACCCCACACCACAACCCCATC
ATCCTCCTCACGCAACGCCCACCCCCCGACCCCCCCCCGCCCCCCGTTGCAATAGGTGACCCACGAGCCCGCCCC
GTCGTACCTGCCCCCAACCCGCCCCCCCCTCCCCCTCCCCCCCCCCCCCCACCCACCCCCCCCCCCCCCCTGTG
AAACATCTGAAAATAAATTAAAAATAATTTAACAACAAACTAACTGTAAATACCGCATAAGTCTAGTTGATAG
TAATGGTAGTGGGTTGAACTACTATGTAATATTAAGGAATCTGTGAACGTGAGCGAGTACATTGGGCTGGTTCT
GGTTTATTTAGGTGTTGTCTTTGGTCTGTGAATTAGTAGCTTGTGACTTAGTAGTTAGTGTAGTTGAGCTATTAG
TATGATCGTTAAATTTTGAGAAAGATGACGGGAATTTCTTATAGGTCTTAAAATTGACAGTGAGTTTTGCTAGA
GTGTGATTATATGTATATGTTAAAGATGATTAGTTATGTAATGATGAAGTATAGTTATT

AKU073_F

GGGGGAAAGGGGGGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTG
GCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGC
TGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGA
AGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCAC
CACTATGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAG
GTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAAACGGGCCCGGGGGCGCGGACCCG
AGCTGCCGCCTCCGTTGCCTCAGCAGCCGTTGGCCCACATGCACAGTGGCCAGCACCGACAGGAGTAGCCAAG
GAGCGTAGAAATTGAAAA

AKU073_R

ACAGCGGCTCCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGTGGGCCAGG
CGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCA

TAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCC
GTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGG
TGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGA
GGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGAGCCCTTCCTCG
ACCCGAGGCCGCTGACCGTCGTGGCCGTGGCCAACGTCATGATGCCGGGGGGGGGTGTAAAAA

AKU076_F

TCGTGGGTCGAGGAAGGCTGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCGCTCAGCA
CGTGGCCCTCGAGGACTTGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGC
CGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGG
CGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAG
CACCACTATGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAG
AGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAAACGGGCCCGGGGGCGCGGACC
CGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGAATAAAACGG
GGGGGAGTTTAATTTCTCACAGAAGCGCTTAAAGCGACCGGAGGAGCAAGGATGAGCAGGAGAGAGGCGTGG
TG

AKU076_R

GGCCGCGGCTGCTGAGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTGATC
GGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCGGGCGCGGCGCTACGGCGACGTTT
TCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCA
GCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTT

TCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCG
CCAGCCGCGCAGCCGCCAAGTCCTCGAGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGC
GCGGCAGCGCGGACGGAGCCTTCCTCGACCCGAGGCCGCTGACCGTCGTGGCCGTGGCCAACGTCATGATGCG
GGGGGGGGTGTTAAAAAAAGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCGGGGGCGCCGGCGGCGGGCGGGCGGGGCG
GCGCGGGCGGGGGGGGGCCCGGCGCGGGGGGGCCGGGCCGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGCGCGCGGGGGGGGCGCGGGGGCCCGGGGCG

AKU086C_F

AGGATGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCG
AGGACTTGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTT
CCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGCCGG
CCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATG
GGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCC
GCCTGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAAACGGGCCCGGGGGCGCGGACCGGAGCTGCCG
CCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGAATAGGAGGGGGGGAGCTT

AKU086C_R

TGCAAGCGGCTGCTGAGGCACGGAGGCGGCAGCTCCGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTGAT
CGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTT
TTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCA
GCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTT

TCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCG
CCAGCCGCGCAGCCGCCAAGTCCTCGAGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGC
GCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTCGTGGCCGTGGCCAACGTCATGATGCG
GGGGGGGTGTTAAAATAAAGTTCTTTTTAAGGAGGAATCCGGGGAGTTGGGGGGGG

rs1056827 and rs10012 sequences of the control group

AKU182_F

AGGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCT
CGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCAC
CTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGCC
GGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTA
TGGGGCAGCTGCCCGGGCGGATCTGGAAAACGTCGCCTTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAG
CCGCCTGGCCCACCGCCGCCTCGTTTCCAATCAGTGGCCACGCATACGGGCCCGGGGGCGCGGACCGGAGCTG
CCGCCTCCTTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCACAGGGAAC
CTTAAATTTTTTTTTATCCGCGAGATAACAGTACACGCGGAGGGGGGAGGGGTAGGGGTAGGATGGGGGGGGA
AGGAGGAAGAGCGGAGGGAGAGGCGAGGGAGGAAGTGAGCTTTTTTTTTTTTTTTTTGTTTATTTTGAGATGGT
GAATGGAAGGAGAAGGGGCAGAACTGGGGTGTGTATTTGGGTTGTTGTTTGTTTTTTTTTTTTTCTTTTTTTTTT
TTTATGTGGGGGAGGAGGAGGGGGGGTGAGCGGCGGGGGTGGGGAGGGGTGAGAGAAGGGGGGGGCAGGAG
GAGGAAATGGTGTGGGGTTG

AKU182_R

GAAGCAGAGGCACGGAGGCGGCGCTCCGGTCCGCGCCCCCGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGG
CGGCGGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCC
TGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGC
CTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTC
GGAGCACTGGAGGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGC
CGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGG
ACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGGAGTGCGCTTGAGGGTTT
AAGAGGGCGGGGGGGGAAGAGGGAGGAAAGGAAGAGGGGCCGGGG

AKU183_F

ATAGGCGGGCCTTCTGATTCTAGAGAGAGCGCCATCCGCGCTTCCGCGCCCCCCCCCCCCCGTTTGGGCGCGC
CTGGGTTGCAATGCCCCTCAGGACGTGGCGGGGCCCGTCTGGTTTGTGGGATGTTGCGCCCATGCTTGGGCTGA
CCCCCACTCCCTGTTGCCGTGCCCCGATAAGGGGCAGAAAGCCGCGCCCCCCCCCCCAGACACCACTGTGAAA
GAAGAAATTGACCTTCGGCCGCCGTGCGCCCTTCCCTGCGTGGCGTGGGGGGGGGGGAGGGGGCGCCAGGGCT
TTGGCCCACTACTCGGACCGTTGGAATGTGGAAATGCGAAAACCAGACACCAGGACGGCTAAGTTCTTGGGCG
CCACCCCCGCACCCCCGAAGTCCTCTGCTCAACGTGCTGACGAAGGCCCGCCACCTGTTGCGGCTGCTAGTAAC
AACATCGTGCACCGACCCTTCCTCGACCCGAGGCCTCTAACCGTCGTGGCCGAGGCCAACGTCTGTGTTTTAAA
TTTTTTTTTTAAAA

AKU183_R

AGGCACGGATGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGA
AACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTC
GGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGC
GCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCC
TCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGG
AAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGC
AGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTG
GTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTG
GCCAACGTCATGAGTGCGGGGGGGGTTCAAAAAATTCATTCTTAATGAACGAAGAATCCC
TTGCAATTCCTGTTCTTAAAAATAATTGTCTCATGATTAAAAAATCAAAAGAAAACGGAG
GGGAAAGGAAAAAGGGGGCGGGTTGTGTAATCGTTCGGG

AKU184_F

ACGTACCGGGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACT
TGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTG
CTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCCGGTC
GGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAG
CTGCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCC

CACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGCTGCCGCCTCCGTT
GCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGAGTCCCGGGGGGGGGCTTT

AKU184_R

GCGCCCCCGGGCCCGTTTGCGTGGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTT
TCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAAT
GGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTT
CCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCA
GCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGA
GCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCCTTCCTCGACCCGAGGCCGC
TGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGGGGGGGGTTTCAAGAAGGGCTTCAAGAGGGCACGAGGA
TCCTTGACTTTCTGCGCGGGGGGGGGGG

AKU191_F

GCGGCGGCCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCG
CTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGG
CTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACG
GAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTC
GCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGA
GCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGG

CGCGGACCCGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGAT
AACAGGGGGGGGCGCTTTATTTTAGGTGGATAAAAAGGA

AKU191_R

TTACGAGCAGCGGCTGCTGAGGCAACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCA
CTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGC
GACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCT
GGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCA
TGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTC
ACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGC
TGGTGCGCGGCAGCGCGGACGGGCGCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCAT
GGTGCGGGGGGGGGTGTGTTTAAAAAAAAATTCCACGCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
CCGCGGGCGGCCG

AKU193_F

AGGGATGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCT
CGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCAC
CTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGAC
GGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTA
TGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAG

CCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGC
CGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCAAGGAACCGT
AAGTTTAAAATGATAAATA

AKU193_R

TAACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGTGGG
CCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCCTGGGCAGCT
GCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGA
CCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACT
GGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGT
CCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCC
TTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAATGCGCGGGGTAGGTTTAAAAAAA
CCATAATTATTGACCGAAGAATCCTTGCCCTTTCTGCATTCTAAAATGCTGTCTCCGCTGAGAAATCTGGGGGG
CGGGGGGGGGGGGGGGGGGGG

AKU194_F

CCTGTAGCGGGCTCGGGTCGAGGAAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGC
CTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGT
GGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCAC
ACGGAAGGAGGCGAAGGCCGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCG

CTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGG
CGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGG
GGGCGCGGACCGGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGAC
AGGGGCGGGGGGGGGGTTTTACCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCAACCAACACCAAACAAAA
AACACCACCACACACCAACATCAAACCCATCAAACTCAGAACGCAAAAATCACCTATACTAAATGACCCGTGA
ATCCTGATGCCTTGGGGGAGTACCTAGCTTGTCATGGATCAGCCACGAGGTCGGGAGGCTGAGACTAAACTTA
GAGAGAACCTGCAAGGTGTTGTGTTACGGAACACCCAGAGGTA

AKU194_R

GTAATGTGGGCAGCGGCTGCTGAGGCAACGGAGGCGGCAGCTCCGGTCCGCGCCCCCGGGCCCGTTTGCGTGG
CCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACG
GCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCC
CTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAG
CATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTC
TTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGC
TGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGT
CATGAGTGCGGTGGTGTTTCAG

AKU196_F

ACAGTTAGCGGGCTCGGGTCGAGGAAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCG
CCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTG
TGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCA
CACGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGC
GCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAG
GCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCG
GGGGCGCGGACCCGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGA
CGGAAAAAGGGGGGGGCCTTTAAT

AKU196_R

GGATTGTGGGGCAGCGGCTGCTGAGGCAACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTG
GCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTAC
GGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGC
CCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCA
GCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTT
CTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCG
CTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACG
TCATGGTGCCCGGGGGGGGTGTAAAAA

AKU197_F

CCGTACCGCGCTGCCGCGCACCAGCAGCGCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGG
CGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTC
CGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCCGGTCGGC
GAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGCTG
CCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCCC
ACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCCGCCTCCGTTG
CCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTCCCGGGGGGGAGGTTTAGTTTAA

AKU197_R

AGACTGCTGAGGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAAC
GCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGA
TCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGG
CTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCC
ACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCC
GCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGC
AGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCCGGGG
GGGTTCAAAGAATT

AKU198_F

GGCGGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCCCGCTCGCGCGCCTCGCTCAGCACGTGGCTCCTCGAGG
ACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCA
GTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCCG
GTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGAATGGCGCGCTCGCCATTCAGCACCACTATGGG
GCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGC
CTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAGACGGGCCCGGGGGCGCGGACCGGAGCTGCCGC
CTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGATTTTCCAAGGGGAGGTTTTA

AKU198_R

CAGGGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGG
CGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCGCCT
GGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCC
TTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCG
GAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCC
GCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGA
CGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGGGGGGGGGGTTC
AAG

AKU199_F

AGGGAAAGGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCGCTCGCGCGCCTCGCTCAGCACGTGGCCCT
CGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCAC
CTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGAC
GGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTA
TGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCTTAGCGCCGCGCCGGGCGAGCGAACGAGAGGTGAG
CCGCCTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCATACGGGCCCGGGGGCGCTGACCCGAGCTG
CCGCCTCCTTTGCCTCACCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCAAGGGACCTT
AAAGGGGGAAAA

AKU199_R

CGGACCGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGGTGGG
CCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCGCCTGGGCAGC
TGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCG
ACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCAC
TGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAG
TCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGC
CTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCAGAGTGCGATGGGGGGTTCAAGGGGGG
TTAA

AKU200_F

GCGTGCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTT
GGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGC
TCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCCGGTCG
GCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGC
TGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCC
CACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCCGCCTCCTTT
GCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCCAGGGAACCTTAATAGAA

AKU200_R

AAATGCGGCTGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGGTGGGC
CAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCCTGGGCAGCTG
CCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCCTTCGCCGA
CCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACT
GGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGT
CCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCC
TTCCTCGACCCGAGGCCGCTAACCGTGGTGGCCGTGGCCAACGTCTGAGTGCCGTGGGAGGTTCAAG

AKU201_F

CCCGACGTTAGCGGGCTCGGGTCGAGGAAGGCTCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGC
GCGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCAT
GCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGAC
ACCACACGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATG
GCGCGCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCG
CCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGG
CCCGGGGGCGCGGACCCGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCA
CCGACAGAGAAGGGGGGGGGGAGC

AKU201_R

TATGTGGGGCAGCGGCTGCTGAGGCAACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGC
CACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACG
GCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCC
CTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAG
CATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTC
TTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGC
TGCTGGTGCGCGGCAGCGCGGACGGAGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGT
CATATGCGGGGGGGGGTGTTTAAAAA

AKU202_F

AGGGAAAGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTC
GAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACC
TTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACG
GCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTAT
GGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGC
CGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGCTGCC
GCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGAGATTCCCCGGGGAACCTTT
ACCCCCGGTAAATTATACCATTAATTCACCTATTTGTGGGTTGGTGGGGGGGGGGCCGGGGGCGGGCGCGGGC
CGGGGGCCGCGCCCGGGG

AKU202_R

GTCGCCCCGGGCCCGTTTGCGTGGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTT
TCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAAT
GGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTT
CCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCA
GCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCTCGGAGGGCCACGTGCTGA
GCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCT
GACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGGGGGGGGGGTTTCAAGAAAACTGTTTTATTGACGAAGGA
TCCTTGGGCTTTGGTTCTTTGAAGGGG

AKU204_F

GGGAAGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGA
GGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTC
CAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGC
CGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGAATGGCGCGCTCGCCATTCAGCACCACTATG
GGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAAACGAGAGGTGAGC
CGCCTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCC
GCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGATTAGCCAAGGGAACGTT
AGAGAAAAAGACTCGTATTGTGCCCGTGGGGTGCTTGTTGCAGTGGGGAGATGGTGTCAGTAAAGTCTTGCGT
GTCGACGCTCGTTGGATGATGGGGGTGGTGGCTAAGAGTTGGAGGAGGGGAGAAGGGAAGAGCGGTGAAAGA
GAGTAAGAGTGATTGCAGGAATTGTGACGGGTGGTGGGGCGAGGGAATGCAGTAGAGTGAAGGAAAAATTAT
AATGTTTGGCTGCTGTGGGGTGAGAGGTAGTGGGGGGTTGGGGGTGGGGTGTGGGAGGGAGAGGGGACGGGG

AKU204_R

GACGCCCCGGGCCCGTTTGCGTGGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGT
TTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAA
TGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCT
TCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGC
AGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCTCGGAGGGCCACGTGCTG

AGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGC
TGACCGTGGTGGCCGTGGCCCACGTCATGAGTGCGCGGGGGGTGTAAA

AKU205_F

GGCTGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAG
GACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCC
AGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCC
GGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGG
GCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGC
CTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCCGCC
TCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCAGGAGCGTAAAA

AKU205_R

CCGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGGTGG
GCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCCTGGGCAGC
TGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCG
ACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCAC
TGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAG
TCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGAGC
CTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGGTGCGGGTGGAGTTTTCAAG

AKU206_F

GGGCGGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCCCGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGG
ACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCA
GTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCCG
GTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGAATGGCGCGCTCGCCATTCAGCACCACTATGGG
GCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGC
CTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCCGCC
TCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGAGAACCCAGGGGAACCTTAGTT
TAAACGAAGATTAAGCGAAGGAAAACATGATACGGGCGTTTCTTGTGAGGGTCTGTGGCGCGCTCTGAATCGA
CAGGGCGCGGACGGGATGATGTCGGGTAGCTAGTGGGGAGGTGTCGTGTGCGAGAGAGCTGAGGAGGTACAA
GGGGTGGTTGGGGTTTGCGGGGTGAGGAGAGCAAGTGTTGAGTGTGGGGG

AKU206_R

GCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGTGGGCCAGGC
GGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCAT
AGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCG
TCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGT
GCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAG
GGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCG

ACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGCGGGGGGTGTTAAAGAATTCATCCTT
ATGACGGA

AKU207_F

AGGGATGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCG
AGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTT
CCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGG
CCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATG
GGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCC
GCCTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGCTGCCG
CCTCCTTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCCAAGGAACCGTT

AKU207_R

CAGGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCG
TGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCCTGGGC
AGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCG
CCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAG
CACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCC
AAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGG

CGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGGGCCGCGGGGGGGTGTTAAAA
AAAGGGGTCTTATTGACGAAGAATCCTTGCACTTCTGGGAGTTTAAG

AKU208_F

TCCGGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGC
GGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCC
GAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCCGGTCGGCG
AAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGCTGC
CCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCA
CCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGCTGCCGCCTCCGTTGC
CTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGGGATGCCCCGGGG

AKU208_R

ACAAGCTGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGGTGGGCCAG
GCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCC
CCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCG
GCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGGA
AGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCT
CGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTC
CTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGCGGGGGGTGTTTAAAA

AKU209_F

TCCGGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGC
GGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCC
GAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCCGGTCGGCG
AAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGCTGC
CCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCA
CCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGCTGCCGCCTCCGTTGC
CTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGAGTTCCCCGGGGGGAGTTTT

AKU209_R

GCAGCGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGG
TGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCGCCTGGG
CAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTC
GCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGA
GCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGC
CAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACG
GAGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGGTGCGCGGGGGGTGTTAAAG
TGGTGAGTGGGGGGGGGGGGGGGTGGGGGGGTGGGGGGGGGGGGGGGGGTGTGGGGGTGGGGTGTGG
GGGTGTTGGGGGGGGGGGGGGGTGTTTGGGGTGTTTTTGTTTGGGGGTGTGGGGTGTGTGTTTTGGTTGTGG
GGTGTGGTGTGGGGGTTTGGGTTGGGGTTGGTGGGT

rs1056827 and rs10012 sequences for the benign group

AKU011_F

AGGGATTGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTC
GAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACC
TTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACG
GCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTAT
GGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGC
CGCCTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGGCGCGGACCGGAGCTGCC
GCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCAAAGAACGGTA
GG

AKU011_R

GAGGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGC
GGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCGCCTG
GGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCT
TCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCG
GAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCC
GCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGA

CGGCGCCTTCCTCGACCCGAGGCCGCTAACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGCGGGGAGGTTTA
A

AKU025_F

GGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCT
CGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCAC
CTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGCC
GGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTA
TGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAG
CCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGC
CGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGATAAGCCGGG

AKU025_R

GCTGGCTGAGGCAACGGAGGCGGCAGCTCCGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAAC
GCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGA
TCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGG
CTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCC
ACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCC
GCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGC
AGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGCGGG
GGGTGTTAAAA

AKU026_F

AGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCT
CGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCAC
CTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGAC
GGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTA
TGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAG
CCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGCTGC
CGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAAAAAAAGGGGGGGACG
TTTAGAATCCACAA

AKU026_R

GATCAGCGGGCTGGCTGGAGGCAACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCAC
TGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCG
ACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTG
GTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCAT
GGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTC
ACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGC
TGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCAT
GAGTGCGGGGGGGGTTTCA

AKU031_F

GCGCCATCCGCGCTGCCGCGCACCAGCAGCGCCCCGCTCGCGCGCCTCGCTAGCACGTGGCCCTCGAGGACTTG

GCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCT
CCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGCCGGCCGGTCGG
CGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGCT
GCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCC
CACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCCGCCTCCGTT
GCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGATAGACCCCAGGGAACCTTA

AKU031_R

AGAGACGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGGTGGG
CCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCCTGGGCAGCT
GCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGA
CCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACT
GGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGT
CCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCC
TTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGCGGGGGGGTTCAAGGGA

AKU034_F

GGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGG
ACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCA
GTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCCG
GTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGAATGGCGCGCTCGCCATTCAGCACCACTATGGG
GCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGC

CTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCCGCC
TCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGGGAGCCCACCGGGGAACCTTA
GACGTGGTTTTCAATAGGTGAAGGTCCGCAACGGTGGGTAGGCGTCTGGTTGGCGGGAAGAACATAGGGATGG
GGTGCGCTGTTTGGTGTACAGAGGAGCGCAGTTGTTGTTAGTATTGGGAGTGGGGAATTGGGGTTGGATGGGGT
GGGAGGGCTGTGGATTGCGATGGGGAGCGAGTTGTGGTAGAGTGGTGAGGGATGTGGGGATGTGGGGGGGGG
GGGCGGGAGGGGGGGCGGGGAGTGGGGGGGGGGCTAGGTGGAGGACATAATGTTGATGGGGATGTTGACTGG
AGGGGCGTGCTCAGAT

AKU034_R

GCATCGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGG
TGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCGCCTGGG
CAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTC
GCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGA
GCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGC
CAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACG
GCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGGAGTGCGGTGAGGTTTCAA

AKU039_F

CGAGGCGCGTAGCGGCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCG
CGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATG
CTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACAC
CACACGGAAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGG

CGCGCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGC
CAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGC
CCGGGGGCGCGGACCGGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCAC
CGACAGAGAGCACGGG

AKU039_R

GGGTTGGTCGTCGCAGCGGCTGCTGAGGCACGGAGGCGGCAGCTCCGGTCCGCGCCCCCGGGCCCGTTTGCGT
GGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCT
ACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAG
GCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCG
CAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAAC
TTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGG
CGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAA
CGTCATGGTGCGGGGGGGGGGTTTAAAA

AKU043_F

AATACCCAGACGTAGCGGCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTC
GCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATC
ATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGA
CACCACACGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGAA
TGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCG
CGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACG

GGCCCGGGGGCGCGGACCCGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAG
CACCGACAGGAGAGACACGGGAAG

AKU043_R

GTTTGTTCGTGTCGCAGCGGCTGCTGAGGCACGGAGGCGGCAGCTCCGGTCCGCGCCCCCGGGCCCGTTTGCGT
GGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCT
ACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAG
GCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCG
CAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAAC
TTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGG
CGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAA
CGTCATGAGTGCGGGGGGGGTTTCAA

AKU044_F

CATGAAAGGCGCCCAACTTCCGCCGCTCCAGAAAGGAACCAGGCGGTTTTGGGTGCAACCCCCCCCCCCTTTGG
GTGCCGCTGTTATCAGTGGCCCTTTGGAGTGGTGGGGTCGCCTCCTTTCCCTTGGAGATTTGCGCCTCCTGCTGT
GTTCTGCTTATCCTTCTGCCTTGCTGTGCTTCAAATGATGGTGAAAGAGGCGCGCGGCGTGCTCCAAGGACCAC
TGCAGAGGGAGTGAAAGCTCGGCCATCAGCGAAAGTTCACTTCTTTGTGTGTGGGGTTGTGGAGGGGGCGGGC
TTTTTTTTCTTCCTCATAGGATGCTAGCCCCGGGAAATCTAAAAAAAATGTATGTTGGGCGTTTTTGTGAGAAAA
GCCGGGGAGCCGACCACGCCCCCGCGCCCCCGTGTCTTTTATTATGATCGGACAAAAAGGCTGGTGGTTGAGA
AAGATATAGCTGCCCTTTTCTTACTTCGCCTATTGAACCAAGTTGTGGCCAGACCAAAACGTGGTGGGGGTTTG
TTTTAA

AKU044_R

CATTATAGCGGCCGCCCTCCTCCTCTCGAGGAAAGAGCCGATCAGCGCTGCCGCGCACCACCCCCGCCCCTTTT
GGTGGGCCTCGGTTAACACATGGCCCTCGAGGACTTGGGGGCGGCTCTGCTTGCCTTTGAAGGAGTTGCGCTCC
CGGTACTTTGTTGTATGCCATGGGTCGTTCTGCACTGGTGCTGAAGAGGGCGCCGGCCCCCCCCGGACCCTTAG
AAGGGGGGGAATGCGGCAGTGGGGAATTCAACCTTCTTGCGGGGGGGGGGTGAGGGGGGTTTTTTTTCAACAT
CATGGTGGCTGGTCCAGCGGGTGCTAAAACTCCACCATTCGCCCCGCCTTGCGAGCGAAAGAGAGGTAGCCCC
CCGGCCTAACGCGACCCGTTTTTTTATTAGTGGCCACCTAAAACAGCTCCGGTGGCGCCAAACACAAATGATGC
CTTCGTTGTCTCTCTCAGCCCCTGGCCCACATGGACAGTGGCCACCACCGAAGAGCGAGGTGGGGTTTTTAGAA
TATTTAAGGT

AKU045_F

GGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCCCGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACT
TGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCAGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCT
CCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCCGGTCGG
CGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGCT
GCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCC
CACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCCGCCTCCGTT
GCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGGACCCAAGGGAACCTTAATATAATG
GCAAATAGGGGATGAGCCTGCGTGGCACGGGGCGTTGATTGGCAGGCGTACTGTTACATCAGGGTGTAGGTCG
GAGTTGAGCTATTCTGG

AKU045_R

GAGGGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCG
GCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCGCC
TGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGC
CTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTC
GGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGC
CGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGG
ACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGGGGGGGGTTTC
AAGAT

AKU046_F

AGGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCC
TCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCA
CCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGA
CGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACT
ATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGA
GCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTG
CCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACAGGAGTAGCCAAAGGAAC
CGTA

AKU046_R

GCAAGCGGGCTGGCTGAGGCAACGGAGGCGGCAGCTCCGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTG

ATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGAC
GTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGT
GCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGG
CTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACG
CGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGG
TGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGGT
GCGGGGGGGGTGTTAAAAT

AKU047_F

AGGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTC
GAGGACTTGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCT
TCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACG
GCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTAT
GGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGC
CGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCC
GCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGGAGACCAA

AKU047_R

GCTGGAGGCACGGAGGCGGCAGCTCCGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGG
CGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGC
CTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGG
CCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACT

CGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAG
CCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCG
GACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGGCCCGGGGGGGGTG
TTAAAAA

AKU048_F

ACGGCTCGCCGCGCACCAGCGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGG
CTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGA
GTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAAGGAGGCGAAGGCCGGCCGGTCGGCGA
AGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGCTGCC
CAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCAC
CGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCCGCCTCCGTTGCC
TCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGATTTCCCCGGGAGAACCTT

AKU048_R

GCTGCTCCAGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGG
CTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAG
TGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGC
CTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGC
AGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGG
CCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGAC
CCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGGCGCGGGGGGGTGTTAAAAA

AKU049_F

GCGGGCCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCT
CAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCT
GCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGA
AAGGAGGCGAAGGCCGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCG
CCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAG
CGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGC
GCGGACCGGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGATTT
TCGGGGGGGGGTCTTTTACCAACCAACGCAGCAACCGAGCACGCAGGAGAGAGGAG

AKU049_R

TAAGGAGCAGCGGCTGCTGAGGCAACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCA
CTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGC
GACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCT
GGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCA
TGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTC
ACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGC
TGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCAT
GAGTGCGCGGGGGGTGTTAAAAA

AKU050_F

AGGGGATGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCC

CTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGC
ACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGG
ACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCCTGGTGGATGGCGCGCTCGCCATTCAGCACCA
CTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGT
GAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGC
TGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGATACGAGGGGGGGGG
CCTTTTACCCTCATTCACGTAGAACACGACGCCGACGTAAAGAGCGAGCCTTGTGCTAGAAGACGCCACACAC
ACCGATGGGGTTGGTCGGCGGGTCCCGTGGTCAGAGGGGGGGCTTGTCCTCTTTTAAATATTTTTATTTTCGTGT
CAACAGTGTCTGACGTCAGATGGCTATTTTGAATAGTCTTATTTTTTATATTTTATGATTTGACTGTTATTTTTT

AKU050_R

GCTTTGTTTGGGGCAGCGGCTGCTGAGGCAACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGT
GGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCT
ACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAG
GCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCG
CAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAAC
TTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGG
CGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAA
CGTCATGAGTGCCGTTGTGTTTCAGAA

AKU054_F

GACGGTGTAGCGGCTCGGGTCGAGGAAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGC

GCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCT
GTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACC
ACACGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCG
CGCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCA
GGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCC
GGGGGCGCGGACCGGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCG
ACAGGAGAGCCAAAGGAACCGT

AKU054_R

GCAGGAGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGG
CGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCG
CCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCG
GCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTA
CTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGC
AGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCG
CGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGGGGGGGGG
TTTCA

AKU056_F

CCGTGCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTT
GGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGC
TCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCCGGTCG

GCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGC
TGCCCAGGCGGATCTGAAAAACGTCGCCATAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCC
CACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCCGCCTCCTTT
GCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCCCAAGAACCTTAGAAA

AKU056_R

GACGCCCCCGGGCCCGTTTGCGTGGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCG
TTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGA
ATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCC
TTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGC
AGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTG
AGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGC
TGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGGGGGGGAGGTTTAA

AKU059_F

CCGTACCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTT
GGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGC
TCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCCGGTCG
GCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGC
TGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCC
CACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCCGCCTCCGTT
GCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGATTTACAAAGGGGAACGTT

AKU059_R

TAGAGCAGCGGCTGCTGAGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTG
ATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGAC
GTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGT
GCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGG
CTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACG
CGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGG
TGCGCGGCAGCGCGGACGGAGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAG
TGCGGGGGGGGTGTTTAA

AKU061_F

AGGGATGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCC
TCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCA
CCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGA
CGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACT
ATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGA
GCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTG
CCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCAAAGGAACC
GTA

AKU061_R

GGTTTGGTGTAGCAGCGGCTGCTGAGGCAACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGT

GGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCT
ACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAG
GCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCG
CAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAAC
TTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGG
CGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAA
CGTCATGGTGCGGGGGGGGGTTTAAAA

AKU062_F

ACGAGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCGCTCAGCACGT
GGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCG
CTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCG
AAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGAATGGCGCGCTCGCCATTCAGC
ACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAG
AGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCC
GAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGAAGCCAAG
GGGAAGGTTAAAACTGTGGGGGGGGGGGAGGGGGGGGGGGGGGGGGGG

AKU062_R

CACGCCCCCGGGCCCGTTTGCGTGGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCG
TTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAA
TGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCT

TCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGC
AGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTG
AGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGC
TGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCCGGGGGAGTGTTAAAG

AKU063_F

CGGAAAGGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCC
TCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCA
CCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGA
CGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGAATGGCGCGCTCGCCATTCAGCACCAC
TATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTG
AGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGCT
GCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGTGAAAGAGGGGGGGGG
CGTTT

AKU063_R

GCTTTGGGCGGATACATCTGGCTGCAGGCAACGGCAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGC
GTGGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCG
CTACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACC
AGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGC
CGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCA
ACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGT

GGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCC
AACGTCATGAGTGCGGGGGGTTTCAAAA

AKU064_F

AGGACTGGCGCCCGTCCCGCGCTGCCGCGCACCAGCGCGCTCACCAGCTCGCGCGCCCTCGCTCAGCACGTGGC
TCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTG
CACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAG
GACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGAATGGCGCGCTCGCCATTCAGCACC
ACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGG
TGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCATACGGGCCCGGGGGCGCGGACCGGA
GCTGCCGCCTCCTTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCCAGGA
ACCTTA

AKU064_R

TAGGGCAGCGGCTGCTGAGGCAACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGCCCGTTTGCGTGGCCACTGA
TCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTT
TTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCA
GCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTT
TCGGCCACTACTCGGAGCACTGGAGGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCG
CCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTG
CGCGGCAGCGCGTACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTG
CGGTGAGGTTTCAA

AKU067_F

GGGGTGACGAGCGCTAGCGGCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGC
TCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCA
TCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCG
GACACCACACGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGG
ATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCC
GCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAAC
GGGCCCGGGGGCGCGGACCCGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCA
GCACCGACAGGAGAGCCCGGGGAGCGT

AKU067_R

GCTGGGTTAGTGTGGCAGCGGCTGCTGAGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGC
GTGGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCG
CTACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACC
AGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGC
CGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCA
ACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGT
GGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCC
AACGTCATGAGTGCGGGGGGGGGTTTAAAA

AKU069_F

AGGACTGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCG

AGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTT
CCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGG
CCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATG
GGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCC
GCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGCTGCCG
CCTCCTTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGATAACCCCAAGGGACCTTAA
TTGTAAAACAACCACACCAATGACATACAGGACCTCTGGGGGGGGACCGAGGGGGCCGGTGGGAGGGGAGGG
GGGGGGGGGGGGAGGCGGCCTTTTTTTTTTTTTTTTTTTTTTTCTTTTGTTCTTGGGGGGAGTAACGTGGTTGT
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGGGGGGGGGGAGGGGGGGGGGAGGGCGGGG
GCGGGGGGGGAGGGGGGGTGGGGG

AKU069_R

ATGAGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGC
GGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCCTGG
GCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTT
CGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGG
AGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCG
CCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGAC
GGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGGAGTGCGGGTGAGGTTTCA
AGAATGGGGGGGGGG

AKU070_F

GAGCGGTATAGCGGCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGC
GCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCT
GTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACC
ACACGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCG
CGCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCA
GGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCC
GGGGGCGCGGACCGGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCG
ACGAATTCCGGGGGGGGACTTTT

AKU070_R

TGCAGGCGGCTGCTGAGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTGAT
CGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGT
TTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGC
AGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCT
TTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGC
GCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGT
GCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGT
GCGGTGGTGTTTCAG

AKU077_F

AGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCG

AGGACTTGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTT
CCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGG
CCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATG
GGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCC
GCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGCTGCCG
CCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGAACCAAGGGGAGGGTTA
AAAACACTAACAACCAACACACACACAAAAACAAAACCACACCCAACACCACACACAAAACAACCAACACCC
AACACCCAGAGGGCGGGGGGAGGGCCGGAGGACGGGGGGAAGCGACGGGAGAGCAAGGAACGACCGGACCA
CAAGCAACCACACGACGAAGGCGGGACAGGACGGCGGCGCCACACGCGAAAGGAGAGCAGGGGGCGGGGGG
GAGCAGAGCGGGGCGGCGGGCCGGG

AKU077_R

TAGCAGCGGCTGCTGAGGCAACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTGA
TCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACG
TTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTG
CAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGC
TTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGC
GCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGT
GCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGT
GCGGGGGGGGTTTCAA

AKU081_F

AGGATTGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCTCCT
CGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCAC
CTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGAC
GGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTA
TGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAG
CCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGC
CGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCCGGGGACCGT

AKU081_R

GTAGCGGCTGCTGAGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTGATCG
GAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTT
CCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAG
CAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTC
GGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCC
AGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCG
CGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCG
TTGAGTGTTTCAGAG

AKU132_F

GAGCAGTAAGCGGCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCG
CCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTG

TGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCA
CACGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGC
GCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAG
GCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCG
GGGGCGCGGACCCGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGA
CAGGAGAGCCAAGGGA

AKU132_R

GGGGGAGGTGGGAGCAGCGGCTGCTGAGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCG
TGGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCT
ACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAG
GCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCG
CAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAAC
TTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGG
CGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAA
CGTCATGAGTGCGGGGGGGGTTTAAA

AKU135_F

AGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCG
AGGACTTGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTT
CCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGG
CCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATG

GGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCC
GCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGGCGCTGACCGGAGCTGCCG
CCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGATTGAAGGGGGGGGGGCTTT
TAAAACCACCCCCACCAAACAAACCCACACCCAACACCAACAAAACAACACAAAACCCCACCCACCCACACAC
CCCCGCCCCCGGGGCAGGAACAACCCGACGACCCACAAAGACCAGGCCGCAGAGGCGCCCCAGGAGGCGCCA
CGCACGGAGAGCGGCGGAGAACGAGCGAACCGGGCCGCGGCGGGGGG

AKU135_R

GCTGCTGAGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCG
GCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCG
CCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCG
GCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTA
CTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGC
AGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCG
CGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGGTGGGGGT
TTTA

AKU136_F

AGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCG
AGGACTTGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTT
CCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGG
CCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATG

GGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCC
GCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCCG
CCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGATTTAACAGGGGAGGAGTTT

AKU136_R

TAGCGTGTTTAGACCGCGGCTGCGGAGGCACGGAGGCGGCGCTCCGTCCGCGCCCCCGGCGCCTTTGCGTGGCC
CTGATGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACTCTGTTCGCTCGCCTGGCGCGGCGCTCGGGCGACCTT
TTCCAGATCCGCCGGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGC
AGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCT
TTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGC
GCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGT
GCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGT
GCGTTGTGTTTCA

AKU144_F

TTAGGCGGGCCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCG
CTAGCACGTGGCCCTCGAGGACTTGGCGGCCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCAGCTGTGGGCT
GCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGA
AGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGC
CATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGC
GAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCG

CGGACCCGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACAGGAG
TAGCCAAGGAGCGT

AKU144_R

GCTGGCTGGAGGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTGATCGGAA
ACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCA
GATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAG
GGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGC
CACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGC
CGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGG
CAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGGTG
GTGTTTCAG

AKU148_F

AGGGAAAGGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCGCTCGCGCGCCTCGCTCAGCACGTGGCCCT
CGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCAC
CTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGAC
GGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTA
TGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAG
CCGCCTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGCTGC
CGCCTCCTTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGGAGTAGCAAGGAGCGTA
TAATAAGGTTTAGTGGGGGGGCTGTGGTGGTGAGGTGTGGAAAGGGTTAGGAGGAGGGGTAGGAGAGGGGGG

GGGGGGGGGGGGGAGGGGGGGGGGGGGAGAGGGTATGGTAGGGTTTGGTGTATGAGAATGTGGGGTGGGGG
GTGGGTGAGAGGGGGAGGAGGGTTTGGGGGTTGGGTGGGGGTGTGGGCGGGTGTTGATGCTTGGGTGCGGGG
GGGGGGGGAGGGGGGGGGGAGGGGGTGGAGGG

AKU148_R

GGAGCTGAGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCG
GCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCC
GCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTC
GGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTA
CTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGC
AGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCG
CGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGGTGGGGGT
TTAAAAAACCCCCAACACCCCCACAACAACACAACCCAACACAAAAAAAAACAACCAAAAAAAACCAACACA
AAAGGGGGGGGGGGGG

AKU159_F

TTTTCGGCCTCTTGGTCGAGGAAGGCGCCGTCCGCGCTGCCGCGCACCAGCCGCGCCACCTTGCTCGCGCGCCT
CGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTG
GGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACA
CGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGC
TCGCCATTCAGCACCACTATGGGGCAGCTGCCCAAGCAGATCTGAAAAACGTCGCCGTAGCGCCGCCGGCGAG
CGAACGAGAGGTGAGCCGCCTGGCCCACCGCCCCCGCTTTCCAATCAGTGGCCACGCAGACGGGCCCGGGGGC

GCGGACCCGAGCTGCCGCCTCCTTTGCCTCAGCACCCGCTGGCCCACATGCACGGTGGCCGGCACCGACGGGA
GTAGCACGAGCGTAAGTGGGAGGAGGGGGGGGAGGAGGGGAAGGGGGGGAAGGAAAGGGAAGGAGGAGAG
GGAGGAAGAAGGAGGAGAAGAAAAAAGGAGAGGGGAGGGAGGGGGGGGGGGGGGGGAGAGGGGAGGGGG
GGGAAGAGGGAGGGGAGGAGGGGGGGAGGGGGGGGGGGGGGGGGGGGGAGGGGAGGGAAAGAAGAAAAG
AAGGAAAAAGAAAAGAAGAGGGGAAGAGAGAAGAGAGGGAGGAGGAGGAGGGGGGGGGAAAGAGAGGGAAG
GG

AKU159_R

GTCGCGGCTGCTGAGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGCCCGTTTGCGTGGCCCTGATCGGAA
ACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAG
ATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGG
GCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCC
ACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCC
GCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGC
AGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTAACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGTTGA
TGTTTCAGAA

AKU163_F

CTAGCGGCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCGCTCGCGCGCCTCGCTCA
GCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGC
GCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAG
GAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCAT

TCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCGGGCGAGCGAA
CGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGCGCGG
ACCCGAGCTGCCGCCTCCTTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGTTGCCCC
GGGAGACCTT

AKU163_R

CCCCGCCGGCTGCTGAGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGCCCGTTTGCGTGGCCCTGATCGG
AAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCC
AGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCA
GGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCG
GCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCA
GCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGC
GGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGT
TAGTGTTTCAGATTTGGTGGGTGGATTTTGGCGGGGAGTGGTGGTGAGGGGAGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGGGGGGGGGGGGGGGGTGATGGAGGTTAAAGGGGGGGAGGGGGAGGCAAAGGGGGGGGGGGA
GAGG

AKU070_F

GAGCGGTATAGCGGCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGC
GCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCT
GTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACC
ACACGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCG

CGCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCA
GGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCC
GGGGGCGCGGACCGGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCG
ACGAATTCCGGGGGGGGACTTTT

AKU070_R

TGCAGGCGGCTGCTGAGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTGAT
CGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGT
TTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGC
AGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCT
TTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGC
GCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGT
GCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGT
GCGGTGGTGTTTCAG

AKU077_F

AGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCG
AGGACTTGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTT
CCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGG
CCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATG
GGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCC
GCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGCTGCCG

CCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGAACCAAGGGGAGGGTTA
AAAACACTAACAACCAACACACACACAAAAACAAAACCACACCCAACACCACACACAAAACAACCAACACCC
AACACCCAGAGGGCGGGGGGAGGGCCGGAGGACGGGGGGAAGCGACGGGAGAGCAAGGAACGACCGGACCA
CAAGCAACCACACGACGAAGGCGGGACAGGACGGCGGCGCCACACGCGAAAGGAGAGCAGGGGGCGGGGGG
GAGCAGAGCGGGGCGGCGGGCCGGG

AKU077_R

TAGCAGCGGCTGCTGAGGCAACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTGA
TCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACG
TTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTG
CAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGC
TTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGC
GCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGT
GCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGT
GCGGGGGGGGTTTCAA

AKU081_F

AGGATTGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCTCCT
CGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCAC
CTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGAC
GGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTA
TGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAG

CCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGC
CGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCCGGGGACCGT

AKU081_R

GTAGCGGCTGCTGAGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTGATCG
GAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTT
CCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAG
CAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTC
GGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCC
AGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCG
CGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCG
TTGAGTGTTTCAGAG

AKU132_F

GAGCAGTAAGCGGCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCG
CCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTG
TGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCA
CACGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGC
GCTCGCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAG
GCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCG
GGGGCGCGGACCCGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGA
CAGGAGAGCCAAGGGA

AKU132_R

GGGGGAGGTGGGAGCAGCGGCTGCTGAGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCG
TGGCCACTGATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCT
ACGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAG
GCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCG
CAGCATGGCTTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAAC
TTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGG
CGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAA
CGTCATGAGTGCGGGGGGGGTTTAAA

AKU135_F

AGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCG
AGGACTTGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTT
CCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGG
CCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATG
GGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCC
GCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCTGACCGGAGCTGCCG
CCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGATTGAAGGGGGGGGGGCTTT
TAAAACCACCCCCACCAAACAAACCCACACCCAACACCAACAAAACAACACAAAACCCCACCCACCCACACAC
CCCCGCCCCCGGGGCAGGAACAACCCGACGACCCACAAAGACCAGGCCGCAGAGGCGCCCCAGGAGGCGCCA
CGCACGGAGAGCGGCGGAGAACGAGCGAACCGGGCCGCGGCGGGGGG

AKU135_R

GCTGCTGAGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCG
GCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCG
CCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCG
GCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTA
CTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGC
AGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCG
CGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGGTGGGGGT
TTTA

AKU136_F

AGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCG
AGGACTTGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTT
CCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGG
CCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATG
GGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCC
GCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCGGAGCTGCCG
CCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGATTTAACAGGGGAGGAGTTT

AKU136_R

TAGCGTGTTTAGACCGCGGCTGCGGAGGCACGGAGGCGGCGCTCCGTCCGCGCCCCCGGCGCCTTTGCGTGGCC
CTGATGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACTCTGTTCGCTCGCCTGGCGCGGCGCTCGGGCGACCTT

TTCCAGATCCGCCGGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGC
AGCAGGGCTCGGCCTTCGCCGACCGGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCT
TTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGC
GCCAGCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGT
GCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGT
GCGTTGTGTTTCA

AKU144_F

TTAGGCGGGCCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCAGCTCGCGCGCCTCG
CTAGCACGTGGCCCTCGAGGACTTGGCGGCCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCAGCTGTGGGCT
GCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGA
AGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGC
CATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGC
GAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAACGGGCCCGGGGGCG
CGGACCCGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACAGGAG
TAGCCAAGGAGCGT

AKU144_R

GCTGGCTGGAGGGCACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTGATCGGAA
ACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCA
GATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAG
GGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGC

CACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGC
CGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGG
CAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGGTG
GTGTTTCAG

AKU148_F

AGGGAAAGGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCGCTCGCGCGCCTCGCTCAGCACGTGGCCCT
CGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCAC
CTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGAC
GGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTA
TGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAG
CCGCCTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGCGCGGACCCGAGCTGC
CGCCTCCTTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGGAGTAGCAAGGAGCGTA
TAATAAGGTTTAGTGGGGGGGCTGTGGTGGTGAGGTGTGGAAAGGGTTAGGAGGAGGGGTAGGAGAGGGGGG
GGGGGGGGGGGGGGAGGGGGGGGGGGGGAGAGGGTATGGTAGGGTTTGGTGTATGAGAATGTGGGGTGGGGG
GTGGGTGAGAGGGGGAGGAGGGTTTGGGGGTTGGGTGGGGGTGTGGGCGGGTGTTGATGCTTGGGTGCGGGG
GGGGGGGGAGGGGGGGGGGAGGGGGTGGAGGG

AKU148_R

GGAGCTGAGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCG
GCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCC
GCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTC

GGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTA
CTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGC
AGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCG
CGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGGTGGGGGT
TTAAAAAACCCCCAACACCCCCACAACAACACAACCCAACACAAAAAAAAACAACCAAAAAAAACCAACACA
AAAGGGGGGGGGGGGG

AKU159_F

TTTTCGGCCTCTTGGTCGAGGAAGGCGCCGTCCGCGCTGCCGCGCACCAGCCGCGCCACCTTGCTCGCGCGCCT
CGCTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTG
GGCTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACA
CGGAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGC
TCGCCATTCAGCACCACTATGGGGCAGCTGCCCAAGCAGATCTGAAAAACGTCGCCGTAGCGCCGCCGGCGAG
CGAACGAGAGGTGAGCCGCCTGGCCCACCGCCCCCGCTTTCCAATCAGTGGCCACGCAGACGGGCCCGGGGGC
GCGGACCCGAGCTGCCGCCTCCTTTGCCTCAGCACCCGCTGGCCCACATGCACGGTGGCCGGCACCGACGGGA
GTAGCACGAGCGTAAGTGGGAGGAGGGGGGGGAGGAGGGGAAGGGGGGGAAGGAAAGGGAAGGAGGAGAG
GGAGGAAGAAGGAGGAGAAGAAAAAAGGAGAGGGGAGGGAGGGGGGGGGGGGGGGGAGAGGGGAGGGGG
GGGAAGAGGGAGGGGAGGAGGGGGGGAGGGGGGGGGGGGGGGGGGGGAGGGGAGGGAAAGAAGAAAAG
AAGGAAAAAGAAAAGAAGAGGGGAAGAGAGAAGAGAGGGAGGAGGAGGAGGGGGGGGAAAGAGAGGGAAG
GG

AKU159_R

GTCGCGGCTGCTGAGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGCCCGTTTGCGTGGCCCTGATCGGAA
ACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAG
ATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGG
GCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCC
ACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCC
GCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGC
AGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTAACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGTTGA
TGTTTCAGAA

AKU163_F

CTAGCGGCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCGCTCGCGCGCCTCGCTCA
GCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGC
GCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAG
GAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCAT
TCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCGGGCGAGCGAA
CGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAACGGGCCCGGGGGCGCGG
ACCCGAGCTGCCGCCTCCTTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGTTGCCCC
GGGAGACCTT

AKU163_R

CCCCGCCGGCTGCTGAGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGCCCGTTTGCGTGGCCCTGATCGG

AAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCC
AGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCA
GGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCG
GCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCA
GCCGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGC
GGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTGGTGGCCGTGGCCAACGTCATGAGTGCGT
TAGTGTTTCAGATTTG

AKU180_F

AGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCCGCTCGCGCGCCTCGCTAGCACGTGGCCCTCGAG
GACTTGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCA
GTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGGCCG
GTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGGG
CAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCT
GGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAAACGGGCCCGGGGGCGCGGACCCGAGCTGCCGCCT
CCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGATATGCCAGGGGGAGCGTTATT
TCTTTTTATACATACCAGCCAAACTTGACTAAGACTTTTAATCGAGAATTCACTGCGGGGGGAGGTGGTGCAGG
GTGGGGTTGGGGGAGGGGGGCGGGTGGCATTTTGTTTTTTTGTTTTGTTTGTTGTTATGGCCGAGGTATGGGGG
CGGGTGCTGGCGTTAGGTGATTTTGATTTTGTTATTTTTGTCTT

AKU180_R

GCTGGCTGGAGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGCCCGTTTGCGTGGCCCTGATCGGAAACGC

GGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCG
CCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCG
GCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTAC
TCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCA
GCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGC
GGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTCGTGGCCGTGGCCAACGTCATGAGTGCGTGGGGGGTTT
CAAA

AKU182_F

TCCGCGGCTGCCGCGCACCAGCAGCGCCACCGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCGAGGACTTGGCG
GCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTTCCAGTGCTCCG
AGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGCCGGCCGGTCGGCGA
AGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATGGGGCAGCTGCC
CAGGCGGATCTGGAAAACGTCGCCTTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCCGCCTGGCCCACC
GCCGCCGCGTTTCCAATCAGTGGCCACGCAGACGGGCCCGGGGGCGCGGACCGGAGCTGCCGCCTCCTTTGCCT
CAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGTAGCCAAGGAACCTTA

AKU182_R

TGGACCGCGGCGCTCCGGTCCGCGCCCCCGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGGTGGGC
CAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGC
CCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCCTCGCCGACC
GGCCGGCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGG

AAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCC
TCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTT
CCTCGACCCGAGGCCGCTAACCGTCGTGGCCGTGGCCAACGTCATGAAGTGCGTTGAAGGTTCGAGA

AKU185_F

GGGAAAGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCG
AGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTT
CCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGG
CCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATG
GGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCC
GCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAAACGGGCCCGGGGGCGCGGACCCGAGCTGCC
GCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGGGTGCCCGGGGAACCTTA
TTTTTT

AKU185_R

AGAGCGGCTGCTGAGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGA
AACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCA
GATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAG
GGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGC
CACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGC
CGCGCAGCCGCCAAGTCCTCGAGGGCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGG
CAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTCGTGGCCGTGGCCAACGTCATGAGTGCGCGG

GGGGTGTTTAAAAAAGACGGCCCATTTTTAAAAAAAAGGCCTTGGACTTTCCGGAGAGTGTAGAGGGGGGGTG
TGATCCTGGCCGGAAAAATTGTTGAATTAAACCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG

AKU191_F

AGGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCG
AGGACTTGGCGGCGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTT
CCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGG
CCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATG
GGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCC
GCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAAACGGGCCCGGGGGCGCGGACCCGAGCTGCC
GCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAAAGAAAAGGGGAGGGG

AKU191_R

TCACCCAGCGGCTGCTGAGGCAACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTG
ATCGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTACGGCGACG
TTTTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTG
CAGCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGC
TTTCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGC
GCCAGCCGCGCAGCCGCCAAGTCCTCGAGGCCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGT
GCGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTCGTGGCCGTGGCCAACGTCATGAGT
GCGGTGTGTTTCAGA

AKU196_F

TTAGCGGCCTCGGGTCGAGGAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCG
CTCAGCACGTGGCCCTCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGG
CTGCGCGCCGCTGCACCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACG
GAAGGAGGCGAAGGACGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTC
GCCATTCAGCACCACTATGGGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGA
GCGAACGAGAGGTGAGCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAAACGGGCCCGGGG
GCGCGGACCCGAGCTGCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGG
AGAGCCAAAGGAACGTTA

AKU196_R

GCAGCGGGCTGCTGAGGCAACGGAGGCGGCAGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCACTGAT
CGGAAACGCGGCGGCGGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTT
TTCCAGATCCGCCTGGGCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCA
GCAGGGCTCGGCCTTCGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTT
TCGGCCACTACTCGGAGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCG
CCAGCCGCGCAGCCGCCAAGTCCTCGAGGCCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTG
CGCGGCAGCGCGGACGGCGCCTTCCTCGACCCGAGGCCGCTGACCGTCGTGGCCGTGGCCAACGTCATGAGTG
CGTGGGGGGTTCAAGAA

AKU198_F

AGGGACGGCGCCGTCCGCGCTGCCGCGCACCAGCGCGCCCAGCTCGCGCGCCTCGCTCAGCACGTGGCCCTCG

AGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCACCTT
CCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGACGG
CCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACTATG
GGGCAGCTGCCCAGGCGGATCTGGAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGAGCC
GCCTGGCCCACCGCCGCCGCGTTTCCAATCAGTGGCCACGCAAACGGGCCCGGGGGCGCGGACCGGAGCTGCC
GCCTCCTTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGATTAGCCAAAGGGAACGT
TAT

AKU198_R

GACTGCGGCTGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGCGGTGGGC
CAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTCGGCGACGTTTTCCAGATCCGCCTGGGCAGCTGC
CCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTTCGCCGACC
GGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGGAGCACTGG
AAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCGCCAAGTCC
TCGAGGCCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGACGGCGCCTT
CCTCGACCCGAGGCCGCTGACCGTCGTGGCCGTGGCCAACGTCTGAGGTGCGTGGGAGGTTTCAAG

AKU200_F

AGGAAAGGCGCCGTCCGCGCTGCCGCGCACCAGCAGCGCCACCAGCTCGCGCGCCTCGCTCAGCACGTGGCCC
TCGAGGACTTGGCGGCTGCGCGGCTGGCGCGTGAAGAAGTTGCGCATCATGCTGTGGGCTGCGCGCCGCTGCA
CCTTCCAGTGCTCCGAGTAGTGGCCGAAAGCCATGCTGCGGCCGCCGGACACCACACGGAAGGAGGCGAAGGA
CGGCCGGTCGGCGAAGGCCGAGCCCTGCTGCACCAGGGCCTGGTGGATGGCGCGCTCGCCATTCAGCACCACT

ATGGGGCAGCTGCCCAGGCGGATCTGAAAAACGTCGCCGTAGCGCCGCGCCAGGCGAGCGAACGAGAGGTGA
GCCGCCTGGCCCACCGCCGCCGCGTTTCCGATCAGTGGCCACGCAAACGGGCCCGGGGGCGCGGACCGGAGCT
GCCGCCTCCGTTGCCTCAGCAGCCGCTGGCCCACATGCACAGTGGCCAGCACCGACGGAGAGCCAGAGGAGCG
TAT

AKU200_R

GAGGGCACGGAGGCGGCGCTCGGGTCCGCGCCCCCGGGCCCGTTTGCGTGGCCCTGATCGGAAACGCGGCGGC
GGTGGGCCAGGCGGCTCACCTCTCGTTCGCTCGCCTGGCGCGGCGCTACGGCGACGTTTTCCAGATCCGCCTGG
GCAGCTGCCCCATAGTGGTGCTGAATGGCGAGCGCGCCATCCACCAGGCCCTGGTGCAGCAGGGCTCGGCCTT
CGCCGACCGGCCGTCCTTCGCCTCCTTCCGTGTGGTGTCCGGCGGCCGCAGCATGGCTTTCGGCCACTACTCGG
AGCACTGGAAGGTGCAGCGGCGCGCAGCCCACAGCATGATGCGCAACTTCTTCACGCGCCAGCCGCGCAGCCG
CCAAGTCCTCGAGGCCCACGTGCTGAGCGAGGCGCGCGAGCTGGTGGCGCTGCTGGTGCGCGGCAGCGCGGAC
GGCGCCTTCCTCGACCCGAGGCCGCTGACCGTCGTGGCCGTGGCCAACGTCATGAGTGCGTGGGGGGGTTCAA
GAAA

Appendix X: Genotypes for rs1800440, rs1056836, rs1056827 and rs10012

PID	rs1800440	rs1056836	rs1056827	rs10012	PID	rs1800440	rs1056836	rs1056827	rs10012
AKU041C	GG	GC	CA	GC	**AKU142C**	AA	CC	CA	GC
AKU042C	AA	CC	CA	GC	**AKU149C**	AA	CC	CC	GC
AKU057C	AA	CC	AA	GC	**AKU152C**	AA	CC	AA	GC
AKU060C	GG	GG	AA	GC	**AKU155C**	AA	CC	AA	GC
AKU073C	AG	GC	AA	CC	**AKU158C**	AA	CC	AA	GC
AKU076C	AG	GC	AA	CC	**AKU164C**	AA	CC	AA	CC
AKU086C	AA	CC	CA	GC	**AKU173C**	AA	CC	AA	GC
AKU087C	AA	CC	CC	GG	**AKU211C**	AG	GC	AA	GC
AKU088C	AA	CC	AA	GC	**KAIC004C**	AG	GG	AA	GC
AKU094C	AA	CC	AA	CC	**KAIC005C**	AA	CC	CC	GG
AKU097C	AA	GC	AA	GC	**KAIC006C**	AA	GC	AA	CC
AKU099C	AG	GC	AA	GC	**KAIC007C**	AG	CC	AA	CC
AKU100C	AA	CC	CA	GC	**KAIC008C**	AA	CC	CA	GC
AKU103C	AG	GG	AA	GC	**KAIC011C**	AG	GC	AA	CC
AKU104C	AA	CC	CC	GG	**KAIC014C**	GG	GC	CC	GG
AKU105C	AA	CC	AA	GC	**KAIC016C**	AA	CC	AA	CC
AKU106C	AA	CC	CC	GG	**KAIC018C**	GG	GG	AA	GC
AKU109C	AA	CC	AA	GC	**KAIC021C**	AA	CC	AA	GC
AKU110C	AA	CC	AA	CC	**KAIC023C**	AA	CC	AA	GC
AKU112C	AA	CC	CC	GG	**KAIC024C**	AA	CC	AA	GC
AKU113C	AA	GC	CC	GG	**KAIC025C**	AA	CC	AA	CC
AKU116C	AA	CC	AA	GC	**KAIC026C**	AA	CC	AA	GC
AKU118C	AA	CC	AA	CC	**KAIC027C**	AG	GG	CC	GG
AKU122C	AA	CC	AA	GC	**KAIC029C**	GG	GG	AA	CC
AKU124C	AA	CC	CC	GG	**KAIC030C**	AA	CC	AA	CC
AKU127C	AG	CC	AA	GC	**KAIC032C**	AG	GG	AA	CC

AKU128C	AA	CC	AA	GC	**KAIC033C**	AG	GC	AA	GC
AKU133C	AG	GG	AA	CC	**KAIC035C**	AA	CC	AA	CC
PID	**rs1800440**	**rs1056836**	**rs1056827**	**rs10012**	**PID**	**rs1800440**	**rs1056836**	**rs1056827**	**rs10012**
KAIC041C	AG	GG	CC	GG	**AKU077B**	AG	GC	AA	CC
KAIC046C	AG	GG	AA	CC	**AKU081B**	AG	GC	AA	GC
KAIC047C	GG	GG	AA	CC	**AKU132B**	GG	CC	AA	CC
KAIC051C	GG	GG	AA	GC	**AKU135B**	AG	CC	AA	GC
KAIC052C	AG	CC	CC	GG	**AKU136B**	AG	GG	AA	GC
KAIC057C	AA	CC	AA	GC	**AKU144B**	GG	GG	AA	CC
KAIC059C	AG	GG	AA	GC	**AKU148B**	AG	GG	AA	CC
KAIC060C	GG	GG	AA	GC	**AKU159B**	AA	GC	AA	CC
AKU011B	AG	GG	AA	GC	**AKU163B**	AA	CC	AA	CC
AKU025B	GG	GG	CC	GG	**AKU180B**	GG	GG	AA	CC
AKU026B	AG	CC	AA	CC	**AKU182B**	AA	CC	CC	GG
AKU031B	AA	CC	CA	GC	**AKU185B**	AG	GG	AA	CC
AKU034B	AG	GC	AA	GC	**AKU191B**	AA	CC	AA	CC
AKU039B	AG	CC	AA	GC	**AKU196B**	AG	CC	AA	CC
AKU043B	AA	CC	AA	GC	**AKU198B**	AA	CC	AA	GC
AKU044B	AG	GC	AA	CC	**AKU200B**	AG	GG	AA	GC
AKU046B	AA	CC	AA	GC	**AKU182N**	AA	CC	CC	GG
AKU047B	AG	CC	AA	GC	**AKU183N**	AA	CC	AA	CC
AKU048B	AA	CC	CC	GG	**AKU191N**	AG	GC	AA	GC
AKU049B	AA	CC	CC	GC	**AKU193N**	AA	CC	AA	CC
AKU054B	AG	CC	AA	GC	**AKU194N**	AA	CC	CC	GG
AKU056B	AA	CC	AA	GC	**AKU195N**	AA	CC	AA	CC
AKU059B	AA	CC	AA	GC	**AKU196N**	AG	CC	AA	CC
AKU061B	AA	CC	AA	GC	**AKU197N**	AA	CC	AA	GC
AKU062B	AG	GC	AA	CC	**AKU198N**	AA	CC	AA	GC
AKU063B	AA	CC	AA	CC	**AKU199N**	AA	CC	AA	CC

AKU064B	AA	CC	AA	GC	**AKU200N**	AG	GG	AA	GC
AKU067B	AG	CC	AA	GC	**AKU201N**	AG	GC	AA	CC
AKU069B	AA	CC	AA	GC	**AKU202N**	AG	GC	AA	CC
AKU070B	AA	CC	AA	GC	**AKU204N**	AG	GC	AA	GC
PID	**rs1800440**	**rs1056836**	**rs1056827**	**rs10012**					
AKU205N	AA	CC	AA	GC					
AKU206N	GG	GG	AA	GC					
AKU207N	AA	CC	AA	CC					
AKU208N	AA	CC	AA	CC					
AKU209N	AA	CC	AA	CC					

Appendix XI: Clinicopathological Characteristics of Case Group Participants

This table presents the detailed histopathological and molecular profile data for participants in the case group (ER+ breast cancer).

PID	Histo_Type	Grade	Stage	LN_Status	Mol_Sub type	PID	Histo_Type	Grade	Stage	LN_Status	Mol_Sub type
AKU041	IDC	3	1	N0	lum A	**AKU118**	IDC	2	2	N0	lum A
AKU042	IDC	2	3	N2	lum A	**AKU122**	IDC	2	2	N0	lum A
AKU057	IDC	2	2	N0	lum A	**AKU124**	IDC	2	2	N1	lum A
AKU060	IDC	2	3	N3	lum B	**AKU127**	ILC	1	2	N1	lum A
AKU073	IPC	3	1	N1	lum A	**AKU128**	IDC	2	1	N0	lum A
AKU076	DCIS	3	3	N2	lum A	**AKU133**	IDC	2	3	N2	lum B
AKU086	IDC	2	1	N0	lum A	**AKU142**	IDC	2	1	N0	lum A
AKU087	ILC	2	2	N0	lum A	**AKU149**	IDC	3	2	N1	lum A
AKU088	IDC	1	1	N0	lum A	**AKU152**	IC	2	1	N0	lum A
AKU094	DCIS	2	1	N0	lum A	**AKU158**	IDC	2	1	N0	lum A
AKU097	IDC	2	1	N0	lum A	**AKU164**	IDC	2	1	N0	lum A
AKU099	ILC	3	3	N2	lum A	**AKU173**	IC	2	1	N0	lum A
AKU100	IDC	3	2	N1	lum A	**AKU211**	MC	2	1	N0	lum A
AKU103	IC	3	2	N0	lum A	**AKU155**	IDC	2	1	N0	lum A
AKU104	IDC	3	2	N1	lum A	**KAIC004**	IDC	2	1	N0	lum A
AKU105	IDC	1	2	N0	lum A	**KAIC005**	IDC	2	2	N1	lum B
AKU106	IC	2	1	N0	lum B	**KAIC006**	IDC	3	2	N1	lum A
AKU109	IDC	3	3	N1	lum A	**KAIC007**	IDC	3	2	N1	lum A
AKU110	DCIS	2	2	N3	lum A	**KAIC008**	IDC	2	3	N2	lum A
AKU112	IDC	2	2	N1	lum A	**KAIC011**	PDCIS	2	2	N0	lum A
AKU113	IDC	2	1	N0	lum A	**KAIC016**	IDC	3	3	N2	lum A

AKU116	IDC	3	2	N1	lum A	**KAIC018**	IDC	2	3	N2	lum A

PID	**Histo_Type**	**Grade**	**Stage**	**LN_Status**	**Mol_Subtype**
KAIC019	IDC	3	3	N2	lum A
KAIC021	IDC	3	3	N3	lum A
KAIC023	IDC	3	3	N3	lum B
KAIC024	IDC	2	2	N1	lum A
KAIC025	IDC	2	3	N2	lum A
KAIC026	IDC	2	2	N1	lum B
KAIC027	IDC	2	3	N2	lum B
KAIC029	IDC	2	2	N2	lum A
KAIC030	IDC	3	2	N0	lum A
KAIC032	IDC	3	2	N1	lum A
KAIC033	IDC	2	2	N1	lum A
KAIC035	IDC	3	3	N3	lum A
KAIC041	IDC	2	3	N2	lum A
KAIC046	IDC	2	2	N0	lum B
KAIC047	DCIS	2	1	N1	lum A
KAIC051	IDC	2	2	N1	lum A
KAIC052	IDC	2	2	N0	lum B
KAIC057	IMCS	2	3	N2	lum B
KAIC059	IDC	3	2	N0	lum A
KAIC060	IDC	3	2	N0	lum A

Appendix XII: Publication Abstracts

Association of rs4646903 and rs1048943 CYP1A1 estrogen-metabolizing gene polymorphisms with estrogen receptor-positive breast cancer in Kenyan women

Mary Murithi[1,2,*], Steven Nyanjom[1], Victor Mobegi[3], Sayed Shahin[4] and Francis Makokha[5]

[1]*Biochemistry Department, Jomo Kenyatta University of Agriculture and Technology,* 62000-00200, Nairobi, Kenya
[2]*Pre-Clinical Department, Kabarak University,* Private Bag-20157, Nakuru, Kenya
[3]*Department of Biochemistry, University of Nairobi, 30197-00100*, Nairobi, Kenya
[4]*Department of Pathology, Aga Khan University Hospital,* 30270-00100, Nairobi, Kenya
[5]*Directorate of Research and Innovation, Mount Kenya University,* 342-01000, Thika, Kenya

***Corresponding author:** murithimary11@gmail.com

Received: January 15, 2023; **Revised:** January 31, 2023; **Accepted:** February 1, 2023; **Published online:** February 10, 2023

Abstract: Breast cancer is the most prevalent neoplasm and the second leading cause of death among females in Kenya. Estrogen and its metabolites are known risk factors for breast cancer. Polymorphisms in these genes and breast cancer susceptibility are unique among different populations. This study aimed to determine the probable associations between estrogen-metabolizing gene variations and other risk factors for breast cancer risk in Kenyan women. Buffy coat samples were obtained from patients diagnosed with estrogen receptor-positive breast cancer, benign breast disease, and healthy volunteers. Genotyping of target polymorphisms was conducted using polymerase chain reaction (PCR)-restriction fragment length polymorphism (RFLP) analysis. The rs4646903 variant genotype CC was associated with breast cancer in the case-control model (P=0.001); the heterozygous genotype TC (P=0.01) and the luminal B molecular subtype (P=0.02) showed increased odds of late-stage breast cancer. The rs1048943 variant genotype GG was associated with breast cancer in the case-benign model (P=0.04), whereas CG was associated with breast cancer in the case-control model (P=0.02). These findings imply that the rs4646903 and rs1048943 variant genotypes are involved in breast cancer risk in Kenyan women. Hence, they may be explored further as potential markers for the disease.

Keywords: breast cancer; gene polymorphisms; estrogen metabolizing gene; polymerase chain reaction-restriction fragment length polymorphism; genotype