

# **ASSOCIATION OF GENETIC VARIANTS WITH BREAST CANCER INTRINSIC SUBTYPES AND SPLICE VARIANTS OF THE FIBROBLAST GROWTH FACTOR RECEPTOR 2 IN A SOUTH AFRICAN POPULATION**

**Thérèse Dix-Peek**

A thesis submitted to the Faculty of Health Sciences, University of the  
Witwatersrand, in fulfilment of the requirements for the degree of

**Doctor of Philosophy**

Johannesburg, 2023

## **DECLARATION**

I, Thérèse Dix-Peek, declare that this thesis is my own work. It is being submitted for the degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other university.

*T Dix-Peek*

---

\_\_15<sup>th</sup>\_\_ day of November 2023 in Johannesburg

## **DEDICATION**

This thesis is dedicated to my mother, Hilary Michele Dix-Peek for her encouragement, love and unwavering support.

## **AUTHOR'S CONTRIBUTION TO THE WORK**

I was fully involved in all the stages of the study. The original hypothesis was mine. I worked with the SABCHO study to recruit participants, and recruited control participants independently, obtained informed consent, collected samples and performed the laboratory work, was the primary author for papers and wrote the thesis, under the supervision of Raquel Duarte and Tanya Augustine.

Coauthors of the published papers have agreed to inclusion of the papers in the thesis, signed in Appendix A.

## **WORK SUBMITTED FOR PUBLICATION**

1. **Thérèse Dix-Peek**, Boitumelo P. Phakathi, Eunice J. van den Berg, Caroline Dickens, Tanya N. Augustine, Herbert Cubasch, Alfred I. Neugut, Judith S. Jacobson, Maureen Joffe, Paul Ruff, Raquel A.B. Duarte

**Title:** Discordance between PAM50 intrinsic subtyping and Immunohistochemistry in South African Women with Breast Cancer

**Status:** Published: Breast Cancer Res Treat. 2023 May;199(1):1-12. doi: 10.1007/s10549-023-06886-3. Epub 2023 Mar 3.

2. **Thérèse Dix-Peek**, Caroline Dickens, Tanya N. Augustine, Boitumelo P. Phakathi, Eunice J. van den Berg, Maureen Joffe, Oluwatosin A. Ayeni, Herbert Cubasch, Sarah Nietz, Christopher G. Mathew, Mahtaab Hayat, Alfred I. Neugut, Judith S. Jacobson, Paul Ruff, Raquel A.B. Duarte

**Title:** FGFR2 Genetic Variants in Southern African Women with Breast Cancer

**Status:** Published: Mol Med Rep. 2023 Dec;28(6):226. doi: 10.3892/mmr.2023.13113. Epub 2023 Oct 13.

3. **Title:** FGFR2 mRNA Expression and FGFR2 IIIb and IIIc Splice Isoforms Associated with Breast Cancer

**Status:** Preparing draft for publication

## **ABSTRACT**

African populations are more genetically diverse than other groups, but there is a paucity of information about breast carcinomas. South Africa uses immunohistochemistry (IHC) rather than multiparameter genomic assays, such as PAM50, to classify tumors. Genome-wide association studies have shown variants in *fibroblast growth factor receptor 2* (*FGFR2*) are associated with breast cancer, but this has not been examined in black, African women. This thesis investigated intrinsic subtypes; the effects of four *FGFR2* single nucleotide polymorphisms (SNPs), and *FGFR2* mRNA expression and splice variants associated with breast cancer subtypes.

In a cohort of 378 breast cancer patients, we investigated the concordance between tumor samples classified by IHC and PAM50. The SNPs rs2981582, rs35054928, rs2981578 and rs11200014 were examined in 1001 patients with, and 1005 participants without, cancer. The *FGFR2* mRNA expression and IIIb and IIIc isoforms were examined using cBioPortal, TCGA Splice Seq and TSVdb databases. Statistical analyses were performed using STATA v14.2. This study was approved by the Human Research Ethics Committee of the University of the Witwatersrand (clearance certificate no M161116).

IHC classified patients as estrogen receptor-positive (77.45%), progesterone receptor-positive (70.56%), and epidermal growth factor receptor 2 (HER2)-positive (32.28%). These results, together with ki67, were used as surrogates for intrinsic subtyping, and showed 7% IHC-A-clinical, 73% IHC-B-clinical, 5% IHC-HER2-clinical and 15% triple negative (TNBC). PAM50 gave 19% luminal-A, 32% luminal-B, 24% HER2-enriched and 25% basal-like. Concordance was highest between basal-like and TNBC and lowest between luminal-A and IHC-A.

There was no association with the SNPs, rs2981582, rs35054928, rs2981578 and rs11200014, and breast cancer in the black, South African population. However, rs2981578 was associated with invasive lobular carcinoma (ILC) and HER2-positive breast cancer was associated with rs11200014. ILC and ER-positive cancers were associated with higher *FGFR2*, while TNBC or HER2-

positive breast cancers were associated with lower *FGFR2*. The IIIb isoform was prevalent in ER-positive breast cancer and IIIc prevalent in HER2-positive breast cancer.

Genetic information from the black South African population can improve understanding of breast cancer in our population. We suggest that the cutoff for Ki67 be changed to 20-25% to better reflect the luminal subtype classifications. Lobular carcinoma is associated with rs2981578, and HER2-positive carcinoma is associated with rs11200014. Increased levels of *FGFR2* mRNA and IIIb isoforms are associated with ER-positive breast cancer, while lower levels of *FGFR2* and higher IIIc isoforms are associated with HER2 and TNBC.

## **ACKNOWLEDGEMENTS**

My sincere gratitude is expressed to the following people for their assistance, encouragement and support:

- My supervisor Professor Raquel Duarte, for your support, encouragement and many discussions.
- My supervisor Professor Tanya Augustine, for your help, encouragement and coffee talks.
- Dr Caroline Dickens, for your help with laboratory work and statistical analysis and enthusiasm for the work.
- Ms Confidence Makgoro, for your help with laboratory work and a superb help with recruiting participants to the study.
- Dr Boitumelo Phakathi - I could not have asked for a more involved lab mate and help with clinical questions.
- The participants at CHBAH and CMJAH who kindly and patiently sacrificed their time to enrol in this study. They are thanked for their generosity.
- The Strengthening Oncology research unit under Dr Maureen Joffe, with Dr Oluwatosin Ayeni, Victor Shandukani, Sihle Inge, Nontlantla (Noni) Mkwanazi, Mokgadi Mawela, Thandi Mtyapi, Sihle Sibiya, Yvonne Chaka, Olebogeng (Olé) Mokgadi, for help with patient and control recruitment.
- The SA MRC Common Epithelial Cancer Research Centre, under Professor Paul Ruff
- Dr Alfred Meremo and Dr Unati Ngebelele who helped with participant recruitment.
- Professor Herbert Cubasch at the Batho Pele Breast Unit, CHBAH, for allowing me to talk with the patients and staff in the unit.
- The Staff at the Surgical Breast Unit, CMJAH, under Dr Sarah Nietz and later Dr Boitumelo Phakathi.
- Dr Eunice van den Berg for her continuous help with the IHC.
- Dominique Dix-Peek whose belief in people's intrinsic value has helped guide the way that I would like to treat the participants in this study as well as my colleagues. If you are not part of the solution, you are part of the precipitate!
- My friends and family who have been kind and supportive of my studies over the years.

- I would like to thank the following for their financial assistance:
  - NIH of the National Cancer Institute (Grant no: 01-CA192627 and P30-CA136696) to J. Jacobson, M. Joffe, A. Neugut, P. Ruff
  - University of the Witwatersrand/ South African Medical Research Council Common Epithelial Cancer Research Center Grant (CERC) to P. Ruff;
  - South African National Research Foundation (NRF 105646) grant award to R. Duarte.
  - National Research Foundation (NRF 87935) to T Augustine
  - T Dix-Peek received partial funding from the Department of Internal Medicine, Wits University, to attend PAM50 training in Halle, Germany.

## **TABLE OF CONTENTS**

|                                         |      |
|-----------------------------------------|------|
| DECLARATION .....                       | ii   |
| DEDICATION .....                        | iii  |
| AUTHOR'S CONTRIBUTION TO THE WORK ..... | iv   |
| WORK SUBMITTED FOR PUBLICATION .....    | v    |
| ABSTRACT.....                           | vi   |
| ACKNOWLEDGEMENTS.....                   | viii |
| TABLE OF CONTENTS.....                  | x    |
| LIST OF FIGURES.....                    | xiv  |
| LIST OF TABLES.....                     | xv   |
| NOMENCLATURE.....                       | xvii |
| OVERVIEW OF THE THESIS.....             | xx   |

|                                                    |          |
|----------------------------------------------------|----------|
| <b>CHAPTER 1: INTRODUCTION.....</b>                | <b>2</b> |
| 1.1. BREAST CANCER .....                           | 2        |
| 1.2. HETEROGENEITY OF THE AFRICAN POPULATION ..... | 3        |
| 1.3. BREAST CANCER SUBTYPING .....                 | 5        |
| 1.3.1. Histopathological Subtyping .....           | 5        |
| 1.3.2. Intrinsic Subtyping .....                   | 7        |
| 1.3.3. Clinical Implications of Subtyping .....    | 8        |
| 1.4. GENES ASSOCIATED WITH BREAST CANCER .....     | 10       |
| 1.5. FIBROBLAST GROWTH FACTOR RECEPTOR 2.....      | 11       |
| 1.6. <i>FGFR2</i> POLYMORPHISMS .....              | 13       |
| 1.7. <i>FGFR2</i> IIIb AND IIIc ISOFORMS .....     | 16       |
| 1.8. RATIONALE AND OBJECTIVES.....                 | 18       |
| 1.8.1. Rationale: .....                            | 18       |
| 1.8.2. Objectives:.....                            | 18       |

|                                                          |           |
|----------------------------------------------------------|-----------|
| <b>CHAPTER 2: MATERIALS AND METHODS .....</b>            | <b>19</b> |
| 2.1. PARTICIPANT SELECTION AND DATA COLLECTION .....     | 19        |
| 2.2. IMMUNOHISTOCHEMISTRY (IHC) AND PAM50 SUBTYPING..... | 22        |

|                                                                                                                                              |           |
|----------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| 2.2.1. Immunohistochemical classification of tumors .....                                                                                    | 22        |
| 2.2.2. Intrinsic subtyping classification .....                                                                                              | 24        |
| 2.2.3. RNA Extraction .....                                                                                                                  | 25        |
| 2.2.4. PAM50 assay .....                                                                                                                     | 26        |
| 2.2.5. Analysis of the PAM50 Assay .....                                                                                                     | 28        |
| 2.3. FIBROBLAST GROWTH FACTOR RECEPTOR 2 SNP GENOTYPING (rs2981582, rs2981578, rs35054928, rs11200014) .....                                 | 29        |
| 2.3.1. Study Population .....                                                                                                                | 29        |
| 2.3.2. Genomic DNA extraction .....                                                                                                          | 30        |
| 2.3.3. Genotype rs2981582, rs2981578, rs35054928, rs11200014.....                                                                            | 30        |
| 2.3.4. Analysis of FGFR2 Variants.....                                                                                                       | 31        |
| 2.4. <i>FGFR2</i> mRNA EXPRESSION AND FGFR2 IIIB and IIIC SPLICE ISOFORMS ASSOCIATED WITH BREAST CANCER .....                                | 32        |
| 2.4.1. <i>FGFR2</i> gene expression levels in breast cancer .....                                                                            | 32        |
| 2.4.2. FGFR2 transcript variants.....                                                                                                        | 33        |
| 2.4.3. <i>FGFR2</i> splice variants .....                                                                                                    | 33        |
| <b>FORWARD TO CHAPTER 3.....</b>                                                                                                             | <b>36</b> |
| Authors' Contributions.....                                                                                                                  | 37        |
| <br><b>CHAPTER 3: DISCORDANCE BETWEEN PAM50 INTRINSIC SUBTYPING AND IMMUNOHISTOCHEMISTRY IN SOUTH AFRICAN WOMEN WITH BREAST CANCER .....</b> | <b>38</b> |
| 3.1. ABSTRACT.....                                                                                                                           | 38        |
| 3.2. INTRODUCTION.....                                                                                                                       | 39        |
| 3.3. METHODS.....                                                                                                                            | 41        |
| 3.3.1. Study participants .....                                                                                                              | 41        |
| 3.3.2. IHC classification of tumors .....                                                                                                    | 41        |
| 3.3.3. PAM50 intrinsic subtyping .....                                                                                                       | 42        |
| 3.3.4. Statistical analysis .....                                                                                                            | 44        |
| 3.4. RESULTS.....                                                                                                                            | 44        |
| 3.4.1. Characteristics of the study cohort.....                                                                                              | 44        |
| 3.4.2. Comparison of immunohistochemistry and intrinsic subtypes .....                                                                       | 48        |
| 3.4.3. Characteristics by intrinsic subtype .....                                                                                            | 51        |

|                                                                                                                               |           |
|-------------------------------------------------------------------------------------------------------------------------------|-----------|
| 3.4.4. Comparisons of Ki67 cutoff levels.....                                                                                 | 53        |
| 3.5. DISCUSSION AND CONCLUSIONS.....                                                                                          | 54        |
| 3.6. ADDITIONAL INFORMATION .....                                                                                             | 58        |
| Additional Files.....                                                                                                         | 58        |
| <br><b>FORWARD TO CHAPTER 4.....</b>                                                                                          | <b>63</b> |
| Authors' contributions .....                                                                                                  | 64        |
| <br><b>CHAPTER 4: FIBROBLAST GROWTH FACTOR RECEPTOR 2 GENETIC VARIANTS IN SOUTHERN AFRICAN WOMEN WITH BREAST CANCER .....</b> | <b>65</b> |
| 4.1. ABSTRACT.....                                                                                                            | 65        |
| 4.2. INTRODUCTION.....                                                                                                        | 65        |
| 4.3. MATERIALS AND METHODS .....                                                                                              | 68        |
| 4.3.1. Study population.....                                                                                                  | 68        |
| 4.3.2. Genomic DNA extraction, SNP selection and analysis .....                                                               | 69        |
| 4.3.3. Classification of tumors.....                                                                                          | 70        |
| 4.3.4. Statistics .....                                                                                                       | 71        |
| 4.5. RESULTS.....                                                                                                             | 72        |
| 4.6. DISCUSSION.....                                                                                                          | 80        |
| <br><b>FORWARD TO CHAPTER 5.....</b>                                                                                          | <b>82</b> |
| <br><b>CHAPTER 5: FGFR2 EXPRESSION AND FGFR2 SPLICE ISOFORMS ASSOCIATED WITH BREAST CANCER ..</b>                             | <b>83</b> |
| 5.1. ABSTRACT.....                                                                                                            | 83        |
| 5.2. INTRODUCTION.....                                                                                                        | 83        |
| 5.3. MATERIALS AND METHODS.....                                                                                               | 86        |
| 5.3.1. FGFR2 gene expression levels in breast cancer .....                                                                    | 86        |
| 5.3.2. FGFR2 transcript variants.....                                                                                         | 87        |
| 5.3.3. FGFR2 splice variants .....                                                                                            | 87        |
| 5.4. RESULTS.....                                                                                                             | 88        |
| 5.4.1. FGFR2 mRNA expression and subtypes .....                                                                               | 88        |

|                                                                              |            |
|------------------------------------------------------------------------------|------------|
| 5.4.2. Percent spliced in (PSI) of <i>FGFR2</i> IIIb and IIIc Isoforms.....  | 91         |
| 5.4.3. Expression of <i>FGFR2</i> exons (exon 8 and exon 9) with IHC.....    | 92         |
| 5.5. DISCUSSION AND CONCLUSIONS.....                                         | 94         |
| 5.6. SUPPLEMENTARY .....                                                     | 98         |
| <br><b>CHAPTER 6: DISCUSSION AND CONCLUSIONS.....</b>                        | <b>102</b> |
| 6.1. DISCUSSION.....                                                         | 102        |
| 6.2. FUTURE STUDIES .....                                                    | 109        |
| 6.3. CONCLUSIONS.....                                                        | 110        |
| <br><b>CHAPTER 7: REFERENCES.....</b>                                        | <b>113</b> |
| <br><b>APPENDICES .....</b>                                                  | <b>132</b> |
| Appendix A: Coauthors' agreement for papers to be used in thesis .....       | 132        |
| Appendix B: Permission from Prof S Tishkoff to use Fig 1.1 .....             | 140        |
| Appendix C: Ethics Clearance Certificate: M161116.....                       | 143        |
| Appendix D: Informed Consent.....                                            | 145        |
| Appendix E: Information Sheet for Participants .....                         | 147        |
| Appendix F: Inclusion/Exclusion Criteria.....                                | 149        |
| Appendix G: Prosigna® Package Insert .....                                   | 153        |
| Appendix H: Purification of Genomic DNA and Total RNA from FFPE Tissue ..... | 163        |
| Appendix I: PAM50 Gene list .....                                            | 170        |
| Appendix J: Modified Salting Out Procedure for DNA Extraction.....           | 172        |

## **LIST OF FIGURES**

|                                                                                                                                              |           |
|----------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>CHAPTER 1: INTRODUCTION .....</b>                                                                                                         | <b>2</b>  |
| Figure 1.1. Graphical representation of African language sub-families and patterns of migration. ....                                        | 4         |
| Figure 1.2. (A) Structure and (B) Activation of FGFR.....                                                                                    | 13        |
| Figure 1.3. Scheme showing <i>FGFR2</i> IIIb and IIIc splice variants.....                                                                   | 17        |
| <br><b>CHAPTER 2: MATERIALS AND METHODS .....</b>                                                                                            | <b>19</b> |
| Figure 2.1. General overview of the assays conducted for the thesis and the chapters where they have been analyzed. ....                     | 21        |
| Figure 2.2. Heatmap of PAM50 genes by subtype.....                                                                                           | 28        |
| Figure 2.3. Representative FGFR2 gene map.....                                                                                               | 34        |
| <br><b>CHAPTER 3: DISCORDANCE BETWEEN PAM50 INTRINSIC SUBTYPING AND IMMUNOHISTOCHEMISTRY IN SOUTH AFRICAN WOMEN WITH BREAST CANCER .....</b> | <b>38</b> |
| Figure 1. Cohort subtypes by (a) Intrinsic subtyping by the PAM50 assay and (b) Clinical IHC subtypes.....                                   | 48        |
| Figure 2. Comparison between immunohistochemical subtypes and molecular subtypes. ....                                                       | 51        |
| Figure 3. IHC Analysis subtypes with different Ki67 cutoff values.....                                                                       | 54        |
| Supplementary Figure S1. Representative heatmap of intrinsic subtypes.....                                                                   | 60        |
| <br><b>CHAPTER 5: FGFR2 EXPRESSION AND FGFR2 SPLICE ISOFORMS ASSOCIATED WITH BREAST CANCER ..</b>                                            | <b>83</b> |
| Figure 1. Scheme showing <i>FGFR2</i> IIIb and IIIc splice variants.....                                                                     | 85        |
| Figure 2. Representative FGFR2 gene map.....                                                                                                 | 87        |
| Figure 3. <i>FGFR2</i> mRNA expression in (A) Intrinsic subtypes and (B) IHC subtypes .....                                                  | 89        |
| Figure 4. <i>FGFR2</i> expression in HER2 Expression .....                                                                                   | 90        |
| Figure 5. <i>FGFR2</i> expression in ER+ and ER- breast cancers .....                                                                        | 90        |
| Figure 6. <i>FGFR2</i> expression in invasive ductal and invasive lobular carcinomas.....                                                    | 91        |
| Figure 7. PSI of <i>FGFR2</i> exons 7, 8, 9, 10 .....                                                                                        | 92        |
| Figure 8. Expression of FGFR2 Exon 8 (A) and Exon 9 (B) in ER+ and ER- tumors.....                                                           | 93        |
| Figure 9. Expression of FGFR2 Exon 8 (A) and Exon 9 (B) in HER2+ and HER2- tumors.....                                                       | 93        |
| Figure 10. Expression of FGFR2 Exon 8 (A) and Exon 9 (B) in TNBC and ER/PR+ .....                                                            | 94        |

## **LIST OF TABLES**

|                                                                                                                                                                              |           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>CHAPTER 1: INTRODUCTION .....</b>                                                                                                                                         | <b>2</b>  |
| Table 1.1. Molecular and IHC subtypes of breast cancer .....                                                                                                                 | 6         |
| Table 1.2. Allele frequency distribution in South African, African and Global populations .....                                                                              | 14        |
| <b>CHAPTER 2: MATERIALS AND METHODS .....</b>                                                                                                                                | <b>19</b> |
| Figure 2.1. General overview of the assays conducted for the thesis and the chapters where they have been analyzed. ....                                                     | 21        |
| Table 2.1. Allred score: Estrogen and progesterone receptor analysis .....                                                                                                   | 23        |
| Table 2.2. HER2 results by IHC .....                                                                                                                                         | 23        |
| Table 2.3. Immunohistochemical classifications of breast cancer.....                                                                                                         | 24        |
| Table 2.4. Number of required unmarked slides as per the tumor surface area (Prosigna) .....                                                                                 | 25        |
| Table 2.5. <i>FGFR2</i> Variants associated with breast cancer .....                                                                                                         | 31        |
| <b>CHAPTER 3: DISCORDANCE BETWEEN PAM50 INTRINSIC SUBTYPING AND IMMUNOHISTOCHEMISTRY IN SOUTH AFRICAN WOMEN WITH BREAST CANCER .....</b>                                     | <b>38</b> |
| Table 1. Immunohistochemical surrogates for expected PAM50 Intrinsic subtypes, showing IHC used for clinical decisions and Ki67 cutoff values for concordance analysis ..... | 42        |
| Table 2. Clinicopathological characteristics of the study cohort .....                                                                                                       | 46        |
| Table 3. Tumor features by intrinsic subtype .....                                                                                                                           | 52        |
| Table 4. Comparisons for PAM50 and IHC luminal subtypes based on Ki67 cutoff levels .....                                                                                    | 54        |
| <b>CHAPTER 4: FIBROBLAST GROWTH FACTOR RECEPTOR 2 GENETIC VARIANTS IN SOUTHERN AFRICAN WOMEN WITH BREAST CANCER .....</b>                                                    | <b>65</b> |
| Table 1: Primer sequences for the <i>FGFR2</i> SNPs.....                                                                                                                     | 70        |
| Table 2: Allele frequency distribution in South African, African and Global populations .....                                                                                | 70        |
| Table 3: Characteristics of the study participants.....                                                                                                                      | 73        |
| Table 4: Genotypic frequencies of <i>FGFR2</i> variants in case and control populations .....                                                                                | 74        |
| Table 5: Associations between genotypic frequencies of <i>FGFR2</i> variants and immunohistochemical subtyping of patients with Breast Cancer .....                          | 76        |
| Table 6: Associations between genotypic frequencies of <i>FGFR2</i> variants and immunohistochemical subtyping of patients with breast cancer .....                          | 78        |
| Table 7: Genotypic frequencies of <i>FGFR2</i> variants with histological diagnosis.....                                                                                     | 79        |

## **CHAPTER 5: FGFR2 EXPRESSION AND FGFR2 SPLICE ISOFORMS ASSOCIATED WITH BREAST CANCER ..83**

|                                                                                                                                                                                               |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Supplementary Table S5.1. <i>FGFR2</i> transcripts.....                                                                                                                                       | 98  |
| Supplementary Table S5.2. Median (IQR) values of <i>FGFR2</i> expression levels from cBioPortal with intrinsic subtypes, IHC subtype; HER2 score, ER, tumor subtype.....                      | 100 |
| Supplementary Table S5.3. Expression of <i>FGFR2</i> exon 8 and exon 9 expression levels relative to overall <i>FGFR2</i> expression levels with respect to ER, HER2 and TNBC expression..... | 101 |

## **NOMENCLATURE**

|       |                                                 |
|-------|-------------------------------------------------|
| AJCC  | American Joint Committee on Cancer              |
| as_id | alternatively spliced identification            |
| ASIR  | Age standardized incident rate                  |
| ASMR  | Age standardized mortality rate                 |
| ATM   | Ataxia-telangiectasia mutated                   |
| BP    | Before present                                  |
| BRCA1 | Breast cancer 1                                 |
| BRCA2 | Breast cancer 2                                 |
| BRIP1 | BRCA1 interacting helicase 1                    |
| CCDS  | Consensus coding sequences                      |
| CDH1  | Epithelial cadherin 1                           |
| CDS   | Coding sequences                                |
| CEPH  | Center d'Etude du Polymorphisme Humain          |
| CHBAH | Chris Hani Baragwanath Academic Hospital        |
| CHEK2 | Checkpoint kinase 2                             |
| CMJAH | Charlotte Maxeke Johannesburg Academic Hospital |
| DCIS  | Ductal carcinoma <i>in situ</i>                 |
| DNA   | Deoxyribonucleic acid                           |
| DNase | DNA nuclease                                    |
| EDTA  | Ethylenediaminetetraacetic acid                 |
| EMT   | Epithelial-mesenchymal transition               |
| ER    | Estrogen receptor                               |
| ESRP  | Epithelial splicing regulatory protein          |
| FFPE  | Formalin-fixed, paraffin-embedded               |
| FGF   | Fibroblast growth factor                        |
| FGFR2 | Fibroblast growth factor receptor 2             |
| GWAS  | Genome-wide association studies                 |
| H&E   | Haematoxylin & Eosin                            |

|       |                                          |
|-------|------------------------------------------|
| HDAC  | Histone deacetylase                      |
| HER2  | Human epidermal growth factor receptor 2 |
| HGDP  | Human Genome Diversity Panel             |
| HIC   | High income country                      |
| HIV   | Human immunodeficiency virus             |
| HR    | Hormone receptor                         |
| HSGAG | Heparan sulphate glycosaminoglycan       |
| HWE   | Hardy-Weinberg equilibrium               |
| IDC   | Invasive ductal carcinoma                |
| Ig    | Immunoglobulin                           |
| IHC   | Immunohistochemistry                     |
| ILC   | Invasive lobular carcinoma               |
| ISH   | <i>In situ</i> hybridization             |
| KASP  | Kompetitive allele specific PCR          |
| LCIS  | Lobular carcinoma <i>in situ</i>         |
| LD    | Linkage disequilibrium                   |
| LIC   | Low-income country                       |
| LR    | Lifetime risk                            |
| mAB   | monoclonal antibody                      |
| MAF   | Minor allele frequency                   |
| MAPK  | Mitogen activated protein kinase         |
| NCR   | National cancer registry                 |
| NHLS  | National health laboratory services      |
| ng    | Nanogram                                 |
| nm    | Nanometer                                |
| PALB2 | Partner and localizer of BRCA2           |
| PAM50 | Prediction Analysis of Microarray 50     |
| PI3K  | Phosphatidylinositol-3-kinase            |
| PLCy  | Phospholipase Cy                         |

|        |                                                  |
|--------|--------------------------------------------------|
| PR     | Progesterone receptor                            |
| PSI    | Percent spliced in                               |
| PTEN   | Phosphatase and tensin homolog                   |
| QC     | Quality control                                  |
| RAO    | Recent African origin                            |
| RBFOX1 | RNA binding Fox-1 homolog-1                      |
| RCC    | Reporter code counts                             |
| RNA    | Ribonucleic acid                                 |
| ROR    | Risk of recurrence                               |
| RUNX2  | Runt-related transcription factor 2              |
| SABCHO | South African Breast Cancer and HIV Outcome      |
| SNP    | Single nucleotide polymorphism                   |
| STAT   | Signal transducer and activator of transcription |
| STK11  | Serine/threonine kinase 11                       |
| TE     | Tris EDTA                                        |
| TCGA   | The cancer genome atlas                          |
| TNBC   | Triple negative breast cancer                    |
| TNM    | Tumor, nodal, metastasis                         |
| TP53   | Tumor protein P53                                |
| TSVdb  | TCGA Splicing Variants Database                  |
| UICC   | Union for International Cancer Control           |
| YY1    | Yin Yang 1                                       |
| μl     | Microlitre                                       |

## **OVERVIEW OF THE THESIS**

This PhD thesis is presented in a publishable format and is divided into six chapters:

- Chapter 1 consists of a detailed literature review.
- Chapter 2 discusses the methodology used to answer the research questions.
- Chapter 3 was written in the format for “Breast Cancer Research and Treatment”, and was published in May 2023.
- Chapter 4 was written in the format for “Molecular Medicine Reviews”. The paper has been accepted for publication for December 2023.
- Chapter 5 has been drafted in the format for “Frontiers in Genetics”. This paper has not yet been submitted.
- Chapter 6 comprises of the discussion, conclusions and recommendations for future research.

## **CHAPTER 1: INTRODUCTION**

### **1.1. BREAST CANCER**

Breast cancer is the leading cancer of African women and the most commonly diagnosed cancer among South African women; it accounts for 27.1% of all cancers diagnosed in South African women (2020 GLOBOCAN (1)). The burden of breast cancer on the African continent has been steadily increasing over the past few decades. The increase is attributed to longer lifespans, increased access to diagnosis, as well as changes in lifestyle associated with westernization. In Africa, mortality rates are higher than in Europe and the United States, largely due to late stage at diagnosis and treatment options (2,3). For example, the age standardized incident (ASIR) of breast cancer in West European women is 90.7/100 000; 89.4/100 000 in North American women and only 50.4/100 000 in African women, but the age standardized mortality rate (ASMR) is 15.6/100 000 in West European women; 12.5/100 000 in North American women and only 15.7/100 000 in African women (1,4). South Africa's history of African diaspora and movement of people from Europe and Asia during European colonization, have contributed to South Africans having different ancestral heritages (5).

The South African population is classified as 80.66% black; 8.11% white; 8.11% mixed-race; 2.47% Indian/Asian and <1% unknown (6,7). The National Cancer Registry (NCR) (8) found the percentage of breast cancer compared with all cancers to be similar in the black and white population groups (21.90% and 21.54%, respectively), with relatively higher preponderance in the mixed-race (28.43%) and Indian/Asian (38.63%) population groups (8). There are a number of risk factors strongly associated with breast cancer, including: early age at menarche; nulliparity; late age at first birth; late age at any birth; low parity; late menopause, and large weight gain after menopause (9). South African women, who are increasingly following a westernized lifestyle, are at increased lifestyle risks for breast cancer. These risks include moving to an urban area; increasingly sedentary lifestyles; a change in diet with increased consumption

of meat, fat and sugar and decreased consumption of fruits and vegetables, and South African women are less physically active than their male counterparts (10,11).

## **1.2. HETEROGENEITY OF THE AFRICAN POPULATION**

Africa, as the second largest continent, comprises 54 countries and is the second most populous continent. As of 2022, Africa had a population of 1.4 billion people (<https://www.statista.com/statistics/1224168/total-population-of-africa/>); approximately 17% of the world's population.

Africa is considered to be the cradle of humankind (12) and of the existing theories of human evolution, the Recent African Origin (RAO) model has the most support (13,14). According to this "Out of Africa" model non-African populations descended from a *Homo sapiens* ancestor in Africa 100-200 thousand year before present (BP)). This ancestor spread throughout the world to replace more primitive *Homo* species, with the first split between non-African and African populations occurring 44-200 thousand years BP (15).

Africa's heterogeneity is reflected in diverse linguistic differences (16) and the 2144 living languages of Africa have been categorized into 4 major linguistic families: Afro-Asiatic (in northeastern and eastern Africa, and the Sahara); Nilo-Saharan (predominant in Saharan, Sudanic and East Africa); Niger-Kordofanian (spoken widely across Africa) and the Khoisan (spoken by the Hada and Sandawe in east Africa and the San of southern Africa). The diversity of African language and well as key migrations in and out of Africa are represented in figure 1.1 (14,16).

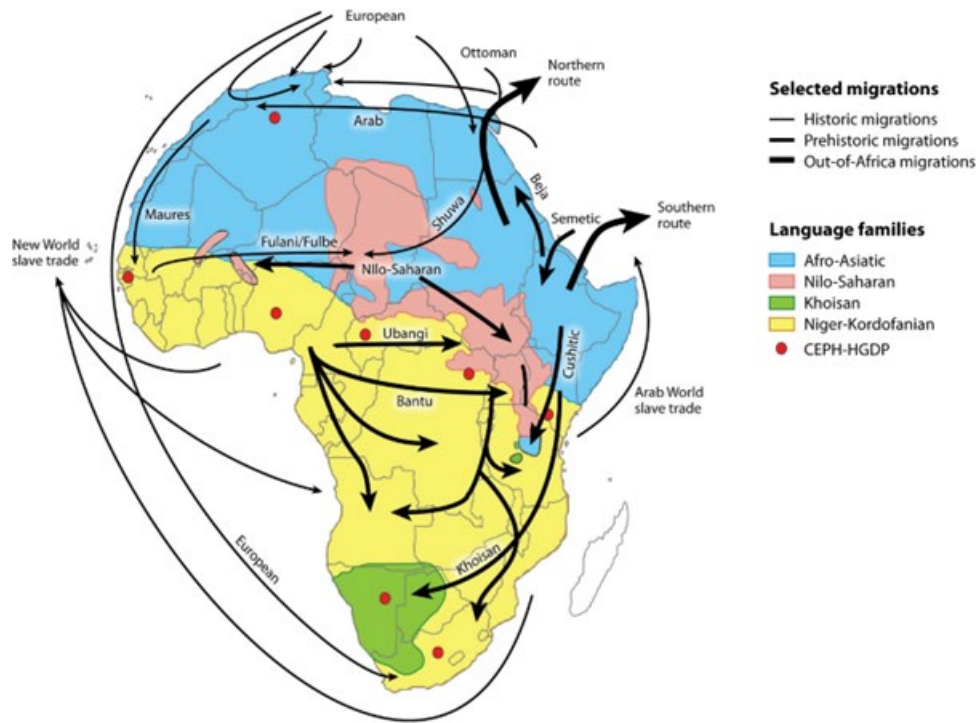


**Figure 1.1. Graphical representation of African language sub-families and patterns of migration.**

The four main language subfamilies, Afro-Asiatic, Nilo-Saharan, Niger-Kordofanian, Khoisan, and the areas where they are spoken are represented. The main hypothesized migration patterns are indicated. The origin of sites of sample sites included in the Center d'Etude du Polymorphisme Humain (CEPH) Human Genome Diversity Panel (HGDP). Reproduced with permission S. Tishkoff (Appendix B).

Similarly to the great diversity in language, is massive genetic diversity among African populations. Indeed, Africa is the most genetically diverse region in the world (17,18) and as such, genomic studies to reflect the very diverse African populations are needed (19). A review showed that only 2.4% of genome-wide association studies (GWAS) included Africans and/or African Americans (18), and there is much less information available on cancer genomics in African populations (19,20). A meta-analysis by Rotimi et al. (19) found that between Jan 1990 - Dec 2019, only 0.329% of the papers produced globally were on cancer in Africa, and only a fifth of these relate to molecular biology.

While African populations are more genetically diverse than European or Asian populations (21,22), most genetic studies in breast cancer have been conducted in women of European descent. These populations are more likely to have estrogen receptor (ER) positive disease. Women of African ancestry have been underrepresented in studies of breast cancer (23), despite the fact that breast cancer accounts for the highest incidence of cancer in African women and sub-Saharan African women with breast cancer estimated to be 27.3% of all cancers in this group (24–26). South African women follow global trends, with the National Cancer registry (NCR) reporting breast cancer accounting for the highest proportion of cancer in women at 23.22% in 2019 (8).

### **1.3. BREAST CANCER SUBTYPING**

Breast cancer is a heterogeneous disease: breast tumors differ greatly in gene expression patterns, growth rates, responses to treatment and clinical outcomes. Breast cancer subtypes have historically been classified according to tumor morphology. Some tumors are restricted by the basement membrane, and are referred to as carcinomas *in situ*, while the majority of tumors have invaded into the stromal tissue, and are referred to as invasive carcinomas. These invasive carcinomas are further categorized as invasive ductal carcinomas (IDCs) or invasive lobular carcinomas (ILCs) (27).

The South African public health care system uses immunohistochemistry (IHC) of four biomarkers, to distinguish between different subtypes. High income countries may also use multigene tests to subtype breast carcinomas.

#### **1.3.1. Histopathological Subtyping**

Breast tumors are classified according to their immunohistochemical profiles, using the estrogen receptor (ER), progesterone receptor (PR), epidermal growth factor receptor 2 (HER2) and the proliferation marker, Ki67. These distinguish between ER-positive luminal A-like [ER and/or PR positive; HER2 negative; Ki67 low], ER-positive luminal B-like [ER and/or PR positive; HER2

negative; Ki67 high; or ER and/or PR positive; HER2 positive, any Ki67]; HER2-enriched [ER and PR negative; HER2 positive; any Ki67]; and triple negative [ER and PR negative; HER2 negative; any Ki67] breast tumors (27–29) ([table 1.1](#)).

**Table 1.1. Molecular and IHC subtypes of breast cancer (modified from Dai et al. (30))**

| Intrinsic Subtype | IHC Classification | IHC Receptor expression         | Grade | Outcome      |
|-------------------|--------------------|---------------------------------|-------|--------------|
| Luminal-A         | A-like             | ER+ and/or PR+, HER2-, Ki67     | 1/2   | Good         |
| Luminal-B         | B-like             | ER+ and/or PR+, HER2-, Ki67+    | 2/3   | Intermediate |
| Luminal-B/HER2    | B-like/HER2        | ER+ and/or PR+, HER2+           | 2/3   | Poor         |
| HER2-enriched     | HER2               | ER-, PR-, HER2+                 | 2/3   | Poor         |
| Basal-like        | TNBC               | ER-, PR-, HER2-, basal marker+* | 3     | Poor         |

\*Other markers, e.g., epidermal growth factor receptor 1 (EGFR) or Cytokeratin 5/6 (CK5/6) may be used for IHC classification. However, these markers have not yet been standardized and their use is not yet widely accepted (28).

Diagnostically, ER and PR are considered negative (assigned as 0) if <1% nuclei stained, (1) weak for 1-10%, (2) moderate for 10-33% and (3) strongly positive if nuclei staining occurred  $\geq 33\%$  (28,31). HER2 is considered negative for assigned values of 0,1,2 and positive for assigned value of 3 (31). The proliferation marker Ki67 has been used to distinguish between the luminal subtypes (32) and been adopted as a marker by the St Gallen International Consensus on Breast Cancer (28). Ki67, which was introduced as a proliferation marker in South Africa in 2013, is considered positive if localization is  $\geq 14\%$  (28,32) ([table 1.1](#)). However, the optimal Ki67 cut off value to distinguish luminal A-like tumors from luminal B-like tumors remains controversial due to uncertainty about how to classify tumors with intermediate (10-30%) Ki67 levels (33). Indeed, the 2015 St Gallen's recommended that a cutoff of 20-29% be used to distinguish A-like and B-like subtypes (34). In addition, immunohistochemistry for Ki67 analysis lacks reproducibility across laboratories and therefore cannot be consistently interpreted (35). Other concerns with immunohistochemical stratification of breast cancers are that the results are affected by the duration of fixation, type of fixative used, speed of assay and completeness of dehydration (6). In addition, the assessment is subjective to the histopathologist, and is not quantifiable. In South

Africa, there is ongoing debate as to the best cutoff of Ki67 to distinguish luminal B-like from luminal A-like with both 14% and 20% currently used at different centers (36).

### **1.3.2. Intrinsic Subtyping**

Due to advances in genetics and molecular biology, the last decade has seen the development of many commercialized multigene tests to guide treatment and provide prognostic information for patients with breast cancer.

The 21-gene OncotypeDx assay (Genome Health Inc, Redwood City, CA) can be used to stratify early stage ER-positive breast cancers, but is not useful for ER-negative cancers (30). The information obtained through differential expression of the 21- genes by quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) provides a Recurrence Score (between 0-100) that stratifies patients in 3 groups, predictive of 10 year risk of distant recurrence (37). Patients are considered at low (0-17); intermediate (18-30) and high (31-100) risk of recurrence, depending on their recurrence scores (37).

MammaPrint (Agendia, Huntington Beach, CA), a molecular assay of 70 genes, has clinical significance in ER-positive and ER-negative early-stage node negative breast cancers (38). The test stratifies patients into either a low or high clinical score based on the risk of disease recurrence at 5 and 10 years (39).

Endopredict (Myriad genetics Inc, Salt Lake City, US) has clinical utility in predicting the risk of recurrence in postmenopausal women with early stage breast cancer, ER positive/HER2 negative, node-negative disease, who had undergone adjuvant hormonal therapy. The 8 gene classifier, based on qRT-PCR, provides a risk score for stratification of patients into low or high risk of distant recurrence (40).

The 50 gene Prediction Analysis of Microarray (PAM50) has a signature of 50 genes (Appendix G) that group tumours into the intrinsic subtypes (luminal A, luminal B, HER2-enriched, basal like)

and can be used on formalin fixed, paraffin embedded (FFPE) biopsy tissue ([table 1.1](#)) (41,42). The PAM50 classifier was adapted by Nanostring to be run on their nCounter Analysis system and incorporates, in addition to 50 cancer- and proliferation- rated genes, a group of 8 reference genes. A list of the genes, their function and grouping are outlined in Appendix G. The commercial version of the PAM50 assay, Prosigna®, has received the CE mark and FDA clearance for clinical use (42). An added benefit of the PAM50 assay is the inclusion of a recurrence score that risk stratifies patients for distant and loco-regional recurrence.

The PAM50 assay is more reliable than the IHC-based techniques because it is less subject to interlab and intralab variability. However, it is much more expensive and labour intensive than the IHC techniques. In South African public hospitals, IHC continues to be used for clinical subtyping because of its lower cost.

It is important to note that all these intrinsic subtyping methods have been developed using populations of European Ancestry, and the efficacy of these methods has not been determined in the more genetically diverse African populations.

### **1.3.3. Clinical Implications of Subtyping**

A recent study at Chris Hani Baragwanath Academic Hospital in Soweto, South Africa found that 64.9% of patients were diagnosed by IHC as luminal B-like, followed by 15.3% as triple negative breast cancer TNBC, 13.8% as luminal A-like, and 6.0% as HER2-enriched (43). Studies investigating receptor status, but not subtyping, found that Black South African women had lower prevalence of hormone receptor (ER/PR) positive tumors and more triple negative tumors than women of European, South Asian or mixed heritage (6,44).

Estrogen positive/luminal breast cancers are the most common subtypes (6,44,45). ER positive tumors are associated with the most favourable outcomes; and have lower recurrence rates than ER negative tumors. It is recommended that both pre-menopausal and post-menopausal patients with ER+ tumors are treated with ER modulators (e.g. Tamoxifen), or ovarian suppression or

aromatase inhibitors, (e.g. Anastrozole) (28,46). These tumours can be luminal-A, luminal-B or luminal-B/HER2. There is mounting evidence that the luminal-A subtype can be further stratified into tumors of ultralow risk, which are of smaller tumor size, lower grade, HER2 negative, low Ki67 and have PR>20% (47). Treatment with tamoxifen reduced the long-term risk of metastatic disease in patients with luminal-A tumors (48). Patients with luminal-B tumors are more likely to develop distant metastatic disease, and the benefits of tamoxifen in these patients is reduced over time (48).

Approximately 25% of black South African women with breast cancer are HER2 positive (44) which can be further divided into 15.8% luminal-B/HER2 and 9.1% HER2-enriched (44). Patients with HER2 positive cancers are treated with the HER2 monoclonal antibody (mAB), trastuzumab, as well as chemotherapeutics, such as paclitaxel or docetaxel (49). While not currently available in the South African public health system, the mABs pertuzumab and margetuximab have been granted FDA approval in limited settings (49,50). Additionally, HER2 is a tyrosine kinase, and there is interest in tyrosine kinase inhibitors such as lapatinib and neratinib (49). HER2 is associated with aggressive tumors, increased risk of recurrence and cancer progression without treatment (28,46).

Triple negative breast cancers are more prevalent in black women compared to white women with breast cancer (51). For example, the frequency of TNBC in black South Africans with breast carcinomas is 25%, compared to 16.8% TNBC in white South Africans and 9.1% in Asian South Africans (44). The definition of triple negative cancers [ER-/PR-/HER2-] and basal-like cancers (express genes found in the basal or myoepithelial cells) are often found in the same tumors (29,52). Since TNBCs lack receptor sites, women with this cancer do not benefit from targeted therapies such as anti-ER or anti-HER2 agents. They are usually treated with cytotoxic chemotherapy and have a more aggressive molecular profile, larger tumor size, high nuclear grade and lower four year survival rate (53,54). TNBCs have high levels of tumor infiltrating lymphocytes in the microenvironment, and there is interest in developing immune checkpoint inhibitors in order to activate anti-tumor responses (51).

Immunohistochemistry against ER, PR, HER2 and Ki67 have been fundamental in the diagnosis and choice of treatment for breast cancer. There are, however other genes that are associated with breast cancer.

#### **1.4. GENES ASSOCIATED WITH BREAST CANCER**

In 1990, Hall et al. (55) discovered Breast Cancer 1 (*BRCA1*) gene, and in 1994 Wooster et al. (56) reported a second breast cancer gene, *BRCA2*. These two hereditary genes were found to substantially increase a woman's risk of breast cancer (57). Subsequently, other rare, but highly penetrative genes including phosphatase and tensin homolog (*PTEN*), tumor protein P53 (*TP53*), epithelial cadherin (*CDH1*) and serine/threonine kinase (*STK11*) have been shown to increase the risk breast cancer (57,58). Together, these highly penetrative genes account for, at most, 25% of hereditary breast cancer cases (57)

Moderate penetrance genes include checkpoint kinase 2 (*CHEK2*), *BRCA1* interacting helicase 1 (*BRIP1*), ataxia-telangiectasia mutated (*ATM*) or partner and localizer of *BRCA2* (*PALB2*) (57,58). These genes account for approximately 3% of familial breast cancers (57). Studies in South Africa found 1.2% *BRCA1*, 0.6% *BRCA2*; 0.6% *ATM*, 0.6% *CHEK2*, 0.6% *PALB2* deleterious variants associated with breast cancer in a self-identified black population, (with a sample size of 165) (59). Similarly, Francies et al. (60) found 3.8% *BRCA1* and 3.8% *BRCA2* pathogenic mutations, but no deleterious mutations in *PALB2* or *CHEK2* in 78 black South African patients. Van der Merwe et al. (61,62) have identified larger rearrangements of the *BRCA1* and *BRCA2* genes that are specific to the black South African population.

Women in Uganda and Cameroon with breast cancer have approximately 5.6% *BRCA1* and 5.6% *BRCA2* mutations, and Nigerian women with breast cancer have 7.1% and 3.9% of *BRCA1* and *BRCA2* respectively (63). A cohort of breast cancer patients of African descent in the Bahamas, Cayman Islands, Barbados, Dominica, Jamaica, Haiti, Trinidad and Tobago have mutations in

either *BRCA1* or *BRCA2* (combined 12.8%), with other mutations in *PARB2*, *RADIC*, *CHEK2*, *ATM*, *STK11*, which is similar to those mutations seen in South Africans (64). A GWAS study in Nigerian patients showed mutations in the introns of *KCNK2*, *C5orf56*, *SCAMP2*, and *SIN3A*, and the other SNPs located close to *GSTM4*, *AMPD2*, *CASTOR2*, and *RP11-168G16.2* (65).

Deleterious mutations in high and medium penetrance genes do not explain the vast majority of breast cancers, and genome-wide association studies (GWAS) (66–69) have led to the detection of over 200 loci associated with breast cancer. These studies allow identification of moderate risk alleles in individuals who do not have high penetrance risk alleles (such as *BRCA1* or *BRCA2*) and yet still develop breast cancer. One of the strongest associations to come out of GWAS shows variant haplotypes in the fibroblast growth factor receptor 2 (FGFR2) (66,67).

### **1.5. FIBROBLAST GROWTH FACTOR RECEPTOR 2**

The fibroblast growth factor receptors (FGFRs) are a family of receptor tyrosine kinases that are involved in signalling pathways that play roles in a variety of biological processes, including cell growth, survival, differentiation, angiogenesis, tumorigenesis (70), and epithelial to mesenchymal transition (71). In normal (non-pathological) circumstances, FGFR signalling is involved in cell proliferation, migration, survival and apoptosis during embryonic development, and homeostasis in adult tissues. Similarly, in cancer, it is involved in mitogenesis, migration and cell survival, and has anti-apoptotic responses in carcinomas (72,73).

FGFR2 is a tyrosine kinase that is located on chromosome 10q26. FGFR2 has been implicated in various breast cancers and gene amplification is associated with 4% of basal-like (triple negative) cancers (73). Signalling of the FGFRs is mediated by fibroblast growth factors (FGFs), which causes the FGFRs to dimerize and undergo phosphorylation of the kinase domain, initiating downstream effects (74). Dysregulation of the FGFR pathways may result in receptors that have a reduced dependence on ligands for activation; alternative splicing of FGFRs that alter ligand specificity; and upregulation of the FGF ligand in cancer cells resulting in paracrine/autocrine activation of

the pathway (73).

Structurally, the FGFRs includes three immunoglobulin-like groups (Ig I, Ig II, Ig III) for binding to the FGF ligands and an acid box on the N-terminus; a transmembrane domain; and a tyrosine kinase domain on the C-terminus (figure 1.2A). The first immunoglobulin-like group (Ig I) and the acid box play a role in receptor auto-inhibition. Ig II and Ig III are involved in ligand binding to the FGFs. The FGFs also bind heparan sulphate glycosaminoglycans (HSGAGs) which facilitate dimerization of FGFRs. The gene encoding for Ig III undergoes alternative splicing, which allows resulting receptor binding to different FGF ligands. Dimerization of the FGFR2 allows the tyrosine kinases to encounter each other, and they can transphosphorylate each other. This allows for a cascade of downstream signals leading to differentiation and mitogenesis (figure 1.2B) (70). Pathways activated by phosphorylation of FGF/FGFR include Ras/Raf-MEK-MAPK (mitogen activated protein kinases), phosphatidylinositol-3-kinase/protein kinase b (PI3K/AKT), phospholipase C $\gamma$  (PLC $\gamma$ ), and Janus kinase/signal transducer and activator of transcription (JAK/STAT) (75).

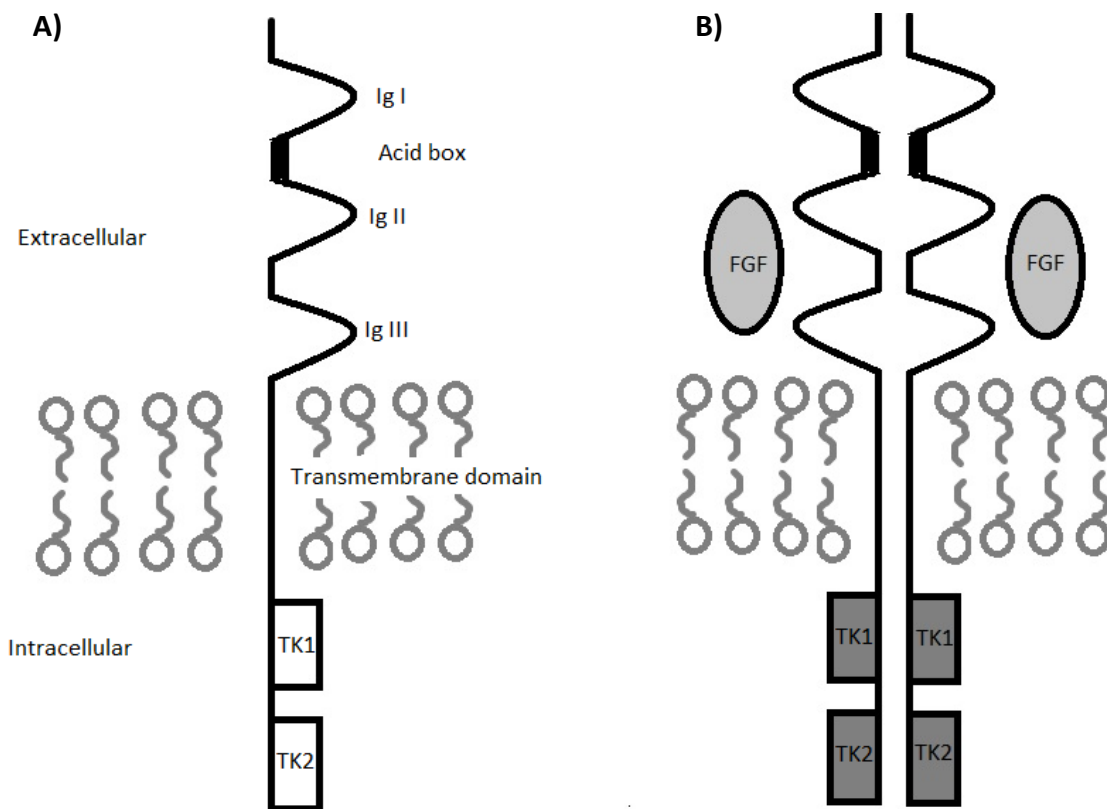


**Figure 1.2. (A) Structure and (B) Activation of FGFR**

A) Structure of The FGFR contains three immunoglobulin-like structure, Ig I, Ig II, Ig III, and an acid box in the extracellular domain. A transmembrane domain links the extracellular to intracellular components, and the tyrosine kinase (TK) exists in the intracellular domain.

B) The acid box and Ig I play a role in receptor auto-inhibition. Ig II and Ig III compose the ligand binding site. Binding causes dimerization of the receptors, which activates the tyrosine kinase to phosphorylate intracellular proteins and elicit a response. Figure created with Microsoft paint and inspired by Miki et al (76).

Alterations of FGFRs in cancer include gene fusions, copy number amplifications, single nucleotide variants and expression of alternative splice variants (77). The effects of *FGFR2* variants and *FGFR2* IIIb and IIIc isoforms will be examined below.

## 1.6. *FGFR2* POLYMORPHISMS

Single nucleotide polymorphisms (SNPs) associated with breast cancer have mainly been found in a 25 kb linkage disequilibrium block in intron 2 of the *FGFR2* gene located on chromosome 10

(66,78). Variants in *FGFR2* have been examined in European; European-American; Asian and African Americans (66,67,78,79). To date, the expression of *FGFR2* variants in southern African women has not been explicitly examined.

Using this information, we looked specifically at SNPs in the intronic region of *FGFR2*. While SNPs usually have small effects individually, polygenic models indicate that an accumulation of these small mutations may contribute to a higher cancer risk (80). Some of the *FGFR2* SNPs found associated with breast cancer include rs2981582, rs35054928, rs2981578 and rs11200014. The minor allele frequencies (MAF) for these SNPs are shown in table 1.2. This includes information from Soweto, South Africa (81); African and Global frequencies available from the 1000 genomes project (82). The African group comprised African Caribbean in Barbados; African Ancestry in Southwest USA; Esan in Nigeria; Yoruba in Ibadan, Nigeria; Gambian in Western division, the Gambia; Mende in Sierra Leone; and Luhya in Webuye, Kenya (83).

**Table 1.2. Allele frequency distribution in South African, African and Global populations**

| SNP        | Position        | Allele | MAF *Soweto   | MAF **African | MAF **Global |
|------------|-----------------|--------|---------------|---------------|--------------|
| rs2981582  | Chr10:121592803 | G/A    | A=0.452       | A=0.495       | A=0.404      |
| rs35054928 | Chr10:12150918  | C-/CC  | Not available | C-=0.331      | CC=0.491     |
| rs2981578  | Chr10:121580797 | C/T    | T=0.106       | T=0.078       | T=0.372      |
| rs11200014 | Chr10:121575416 | G/A    | A=0.112       | A=0.190       | A=0.313      |

SNP, single nucleotide polymorphism; MAF, minor allele frequency; Equilibrium p-value. Allele frequencies from the \*Soweto data (if available) looking at genetic diversity in the black population from Soweto, South Africa (81); \*\*African and \*\*Global frequencies available from the 1000 genomes project (82). The African group comprised African Caribbean in Barbados; African Ancestry in Southwest USA; Esan in Nigeria; Yoruba in Ibadan, Nigeria; Gambian in Western division, the Gambia; Mende in Sierra Leone; and Luhya in Webuye, Kenya.

The *FGFR2* SNP most commonly associated with breast cancer in women of European ancestry, rs2981582 (66), was recently also associated with increased risk of breast cancer in Saudi women (84). Rs29181582 was associated with luminal-A breast cancer in Han Chinese women (85) and Korean women (86) and the expression of the minor allele was associated with early onset breast cancer in Indonesian women (87). A South American study involving women from Argentina and Uruguay did not find an association between rs2981582 and breast cancer, possibly because both

populations are likely to come from mixed populations (88). Likewise, in a study of postmenopausal Turkish women, breast cancer was not associated with rs2981582 (89).

The SNP, rs35054928, seems to be part of the response element, which is a sequence within the promoter region of DNA that regulates the transcription of genes. The rs35054928 risk allele seems to bind the transcription factor, E2F1 (90). E2F1 is a key transcription factor in the control of proliferation, differentiation and apoptosis; and the risk allele of rs35054928 may increase chromatin accessibility and recruitment of RNA polymerase to this site (90). Rs2981578 also seems to be part of a response element. The presence of the risk allele for rs2981578 for example, substantially increases the binding to FOXA1 in MCF7 cells, an epithelial, hormone receptor positive cell line, increasing chromatin accessibility and allowing access to transcriptional repressors such as YY1 (Yin Yang 1), SIN3A and HDAC2 (histone deacetylase) (91). The DNA binding protein YY1, co-repressor SIN3A and histone modifying HDAC form a complex that can inhibit promoter activity (92). The SNPs rs35054928 and rs2981578 are located next to an organic cation transporter (OCT)-binding site. The risk allele of rs2981578 also creates a potential binding site for runt-related transcription factor (*RUNX2*) (71,90). Runx2 is mainly known for its role in osteoclast development, but it is also a regulator for mammary development and breast cancer (93). The rs2981578 variants have equal affinity for OCT1, but the high risk allele has a much higher affinity for *RUNX2* (94,95) possibly due to difference in histone acetylation of the SNP sites (96). OCT1 promotes cell proliferation in ER positive breast cancer cells (97). Martini et al found that rs2981578 was not significantly associated with breast cancer in Ghanaians (98).

The SNP, rs11200014 was associated with breast cancer in African American women(67,99). The heritage of African Americans is often a mixture of African and European ancestry, with a small proportion of Native American ancestry (100,101) and the African ancestry is mainly from West or central West Africa (102). South African populations are genetically different from West and central Africans, and even more different from African Americans (101). We wished to see if rs11200014 was associated with breast cancer in southern African women.

### 1.7. **FGFR2 IIIb AND IIIc ISOFORMS**

The alternative expression of the immunoglobulin loop III of the FGFR ligand binding sites (figure 1.2) fundamentally affects which ligands that can bind to the receptor. In turn, this affects the downstream signalling pathways that are phosphorylated by the FGFR tyrosine kinase (103).

In *FGFR2*, splicing of the third extracellular Ig domain (Ig-III) results in the first half of the loop being formed by exon 7, and the second half of the loop being formed by either exon 8 (*FGFR2* IIIb isoform) or exon 9 (*FGFR2* IIIc isoform) (figure 1.3). These splicing sites are mutually exclusive, with FGFR IIIb expressed in epithelial cells, and FGFR2 IIIc expressed in mesenchymal cells. Moreover, the mesenchymally expressed ligands, FGF7 or FGF10, bind and activate FGFR IIIc; while the epithelial FGFR IIIb is activated by epithelial FGF2, FGF6 and FGF9 (104). In cancer, the switch from IIIb to IIIc is one of the criteria of epithelial-mesenchymal transition (EMT) and can cause a metastatic phenotype in carcinomas (74,103,105). Alternative splicing of two mutually exclusive exons of the *FGFR2* gene, namely exon 8 and exon 9, are expressed in epithelial specific- and mesenchymal specific- isoforms, respectively (106–108).

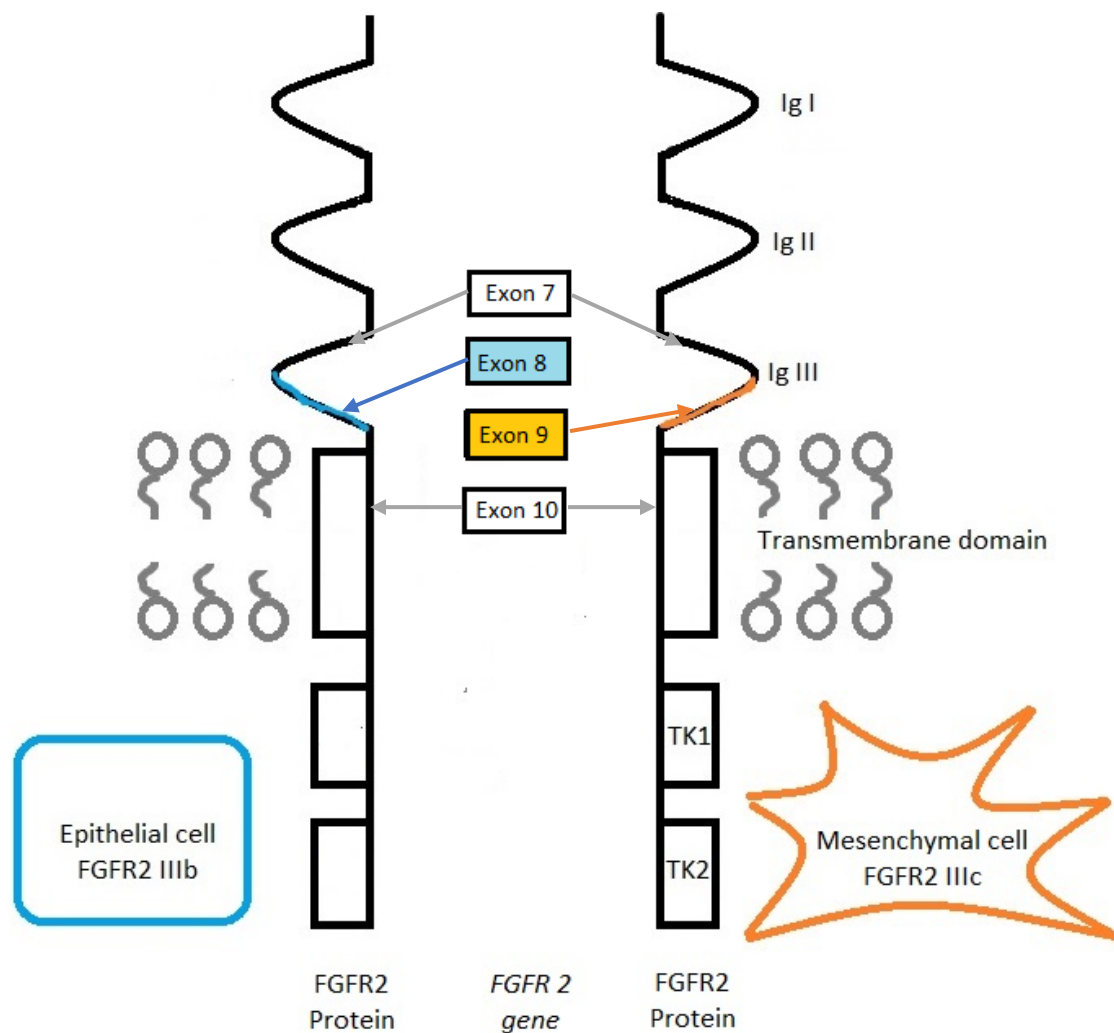


**Figure 1.3. Scheme showing *FGFR2* IIIb and IIIc splice variants.**

The extracellular domain of the FGFR2 protein contains 3 Ig-like loops, a transmembrane domain, and intracellular tyrosine kinases. The Ig-III loop is coded for by exons 7, 8 and 9. *FGFR2* expresses exons 7 and 8 but omits exon 9 in the Ig-III domain to form *FGFR2* IIIb in epithelial cells. By contrast, expression of exons 7 and 9 (omitting exon 8), results in *FGFR2* IIIc in the mesenchymal cells. This splice variant is regarded as a hallmark of EMT, with more aggressive carcinomas associated with the IIIc splice variant. Cartoon created with Microsoft paint and inspired by Matsuda et al (109).

FGF/FGFR signalling pathways are involved in normal development as well as tumorigenesis (110). Knowledge of binding sites, in the cells and their downstream effects, would assist in the development of therapies against carcinomas (110).

## **1.8. RATIONALE AND OBJECTIVES**

### **1.8.1. Rationale:**

Most genetic studies of breast cancer have been conducted in women of European ancestry. Women of African descent are more genetically diverse and are more likely to have ER-negative breast cancers. This study will assess for the first time the gene expression patterns in Southern African women using the PAM50 gene signature. It will attempt to link mRNA *FGFR2* splice variants (*IIIb* and *IIIc*) to the molecular subtypes, as well as identify inherited *FGFR2* risk variants specific to sub-Saharan African women. This may aid in identification and management of the aggressive phenotypes more commonly found in African women.

### **1.8.2. Objectives:**

1. To determine the differential expression pattern of 50 genes in invasive breast carcinomas of black South African women using the PAM50 genomic assay.
2. To determine the extent of concordance between molecular profiling and immunohistochemistry.
3. To identify *FGFR2* SNPs associated with Breast Cancer in a South African population.
4. To correlate cancer phenotypes to *FGFR2 IIIb* and *FGFR2 IIIc* splice variants and *FGFR2* expression.

## **CHAPTER 2: MATERIALS AND METHODS**

### **2.1. PARTICIPANT SELECTION AND DATA COLLECTION**

Ethics for this study was granted from the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (Ethics number M161116) (Appendix A). Participants for this study were drawn from the South African Breast Cancer and HIV Outcome (SABCHO) study, a cohort of breast cancer patients diagnosed and treated at five hospitals in Gauteng and KwaZulu Natal, South Africa (111). In this study, participants with breast cancer were diagnosed at the Charlotte Maxeke Johannesburg Academic Hospital (CMJAH) Surgical Breast Unit and the Batho Pele Breast Unit of the Chris Hani Baragwanath Academic Hospital (CHBAH). The CMJAH Surgical Breast Unit is based in central Johannesburg and identifies about 250 new breast cancer cases yearly. The Batho Pele Breast Unit serves patients from Soweto and surrounding areas and diagnoses about 350 patients with breast cancer yearly (111). Control participants without breast cancer, were recruited from the CMJAH breast unit, and the Batho Pele Breast unit after confirmation that they did not have breast cancer and were considered healthy. Additional control participants, undergoing routine assessment at other clinics at CMJAH, not related to cancer, were also invited to participate in this study. All women signed informed consent (Appendix B and C) and data were collected. The data collected included age, self-assessed African origin, living in South Africa, age, menstrual status, number of pregnancies, history of hormonal contraception, hormone replacement therapy, chronic comorbidities, family history of breast cancer and personal history of any type of cancer (Appendix D). Inclusion criteria for breast cancer patients were self-identified black, southern African women over 18 years with histologically confirmed breast cancer. Exclusion criteria were ductal carcinoma *in situ* (DCIS) or lobular carcinoma *in situ* (LCIS) without invasive breast cancer. For the control participants, inclusion criteria were self-identified black, southern African women over 18 years with no history of breast or ovarian cancers. In addition, control participants were not enrolled if they were pregnant or breastfeeding.

Intrinsic molecular subtyping, using the 50 gene Prediction Analysis of Microarray (PAM50) assay, was performed on RNA from 384 breast cancer tumors. The data obtained investigated concordance between intrinsic subtyping and immunohistochemistry ([chapter 3](#)). In addition, genomic DNA was extracted from blood drawn from 1001 patients and 1006 controls. These samples were used to investigate FGFR2 single nucleotide polymorphisms ([chapter 4](#)). Finally, data from The Cancer Genome Atlas (TCGA) was used to compare IHC data with FGFR2 IIIb and IIIc isoforms ([chapter 5](#)).

The methods to correlate concordance between PAM50 intrinsic subtypes and immunohistochemistry is detailed in section 2.2.; fibroblast growth factor receptor 2 genotyping is found in section 2.3; section 2.4. correlates tumor subtyping with the FGFR2 variants, and an overview of the statistical data analysis is shown in section 2.5. An overview of the study population and methods used is shown in figure 2.1. to indicate their position in the thesis.

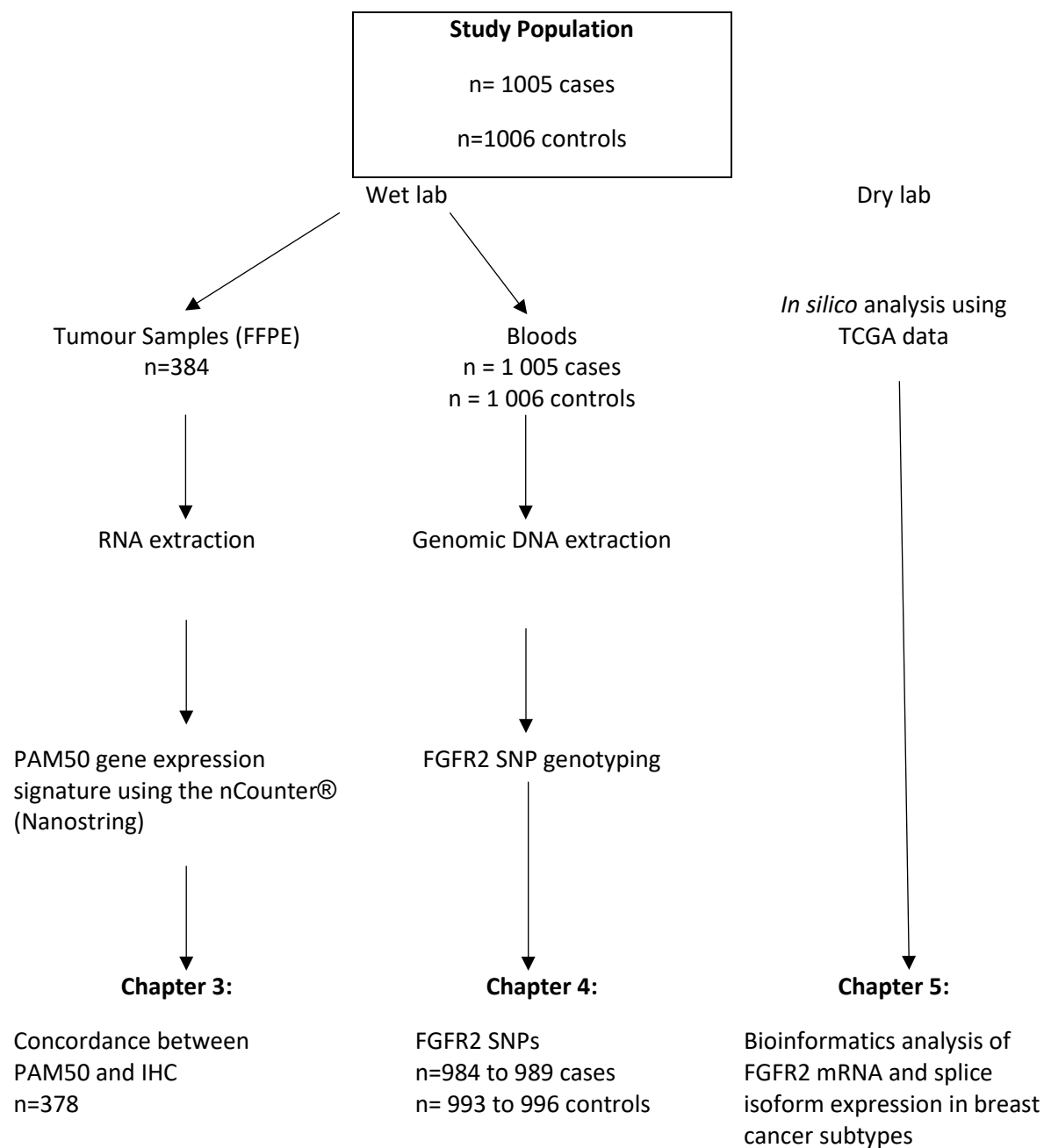


**Figure 2.1. General overview of the assays conducted for the thesis and the chapters where they have been analyzed.**

FGFR2, fibroblast growth factor receptor 2; FFPE, formalin-fixed, paraffin-embedded; PAM50 50 gene prediction analysis of microarray, SNP, single nucleotide polymorphism; TCGA, The Cancer Genome Atlas.

## **2.2. IMMUNOHISTOCHEMISTRY (IHC) AND PAM50 SUBTYPING**

### **2.2.1. Immunohistochemical classification of tumors**

Histopathology, routinely carried out at the National Health Laboratory Services (NHLS), included the histological diagnosis, tumor subtype, grade, and immunolocalization of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and Ki67. The American Joint Committee on Cancer (AJCC) system was used to inform clinical staging of breast cancer (112) and categorized into early (Stage I/II) and advanced (Stage III/IV) stage disease using the TNM (tumor, nodal, metastasis) staging system. The TNM system was developed by the union for international cancer control (UICC); the system assigns a value of T1-T4 based on increasing size of the tumor; nodal involvement is indicated by no involvement (N0) to nodal involvement with increasing distance from the primary tumor (N1-N3). M0 indicates no metastasis, while evidence of distant metastasis is indicated as M1 (113). In addition, the Modified Bloom & Richardson grading system was used, which is based on tubule formation, mitosis, and nuclear pleomorphism; a score of  $\leq 5$  denoted a Grade 1 (well differentiated) tumor, 6-7 a Grade 2 (moderately differentiated) tumor, and 8-9 a Grade 3 (poorly differentiated) tumor (114).

The Allred score was used to determine ER and PR status; with a score of 0-2 regarded as hormone receptor negative and 3-8 regarded as hormone receptor positive (Table 2.1) (115). Tumors were considered HER2 positive if they scored 3+ by IHC, or 2+ by IHC if fluorescent *in situ* hybridization (FISH) showed them to be positive (Table 2.2) (115–117). Specimens in which  $\leq 14\%$  of cells expressed Ki67 were categorized as having low expression, as per the St Gallen 2011 guidelines and South African 2018 guidelines (28,118) and percentage of Ki67 proliferation was determined by visual estimation by an experienced pathologist (36). However, there is ongoing controversy of the most appropriate cut off of the proliferation marker, Ki67 (33). The breast tumors were categorized based on IHC 4 subtype into: Clinical (Clin)-A-like [ER and/or PR positive, HER2 negative, Ki67 < 14%]; Clin-B-like [ER and/or PR positive and HER2 negative with Ki67%  $\geq$

14% or HER2 positive with any Ki67]; Clin-HER2 positive subtype [ER and PR negative, HER2 positive]; and triple negative breast cancer (TNBC) [ER, PR, and HER2 negative] (Table 2.3) (28). The multidisciplinary team at CHBAH use a Ki67 score of 20% in conjunction with the Allred score, grade and age of patient as a cut off to determine if chemotherapeutic treatment in hormone receptor (HR) positive breast cancers is appropriate. In this study, we explored Ki67 cutoffs of 10%, 10-30%, and over 30% for the luminal-like IHC subtypes (Table 2.3). We additionally explored cutoffs of 10%, 15%, 20%, 25% and 30%, because of the uncertainty surrounding those values for clinical decision making (33).

**Table 2.1. Allred score: Estrogen and progesterone receptor analysis**

| Percentage (%) of positive cells | Proportion score | Intensity    | Intensity score |
|----------------------------------|------------------|--------------|-----------------|
| 0                                | 0                | None         | 0               |
| <1                               | 1                | Weak         | 1               |
| 1-10                             | 2                | Intermediate | 2               |
| 11-33                            | 3                | Strong       | 3               |
| 34-66                            | 4                |              |                 |
| >=67                             | 5                |              |                 |

The Allred score is calculated by combining the proportion score with the intensity score. A value of 0-2 is considered negative, 3-4 is weakly positive, 5-6 is intermediate positive and 7-8 is positive for hormone receptor staining (115).

**Table 2.2. HER2 results by IHC**

| Results | Call       | Proportion                                                                                                               |
|---------|------------|--------------------------------------------------------------------------------------------------------------------------|
| 0       | Negative   | No staining, or faint cell membrane staining <10% cancer cells                                                           |
| 1+      | Negative   | Minimal membrane staining > 10% cancer cells                                                                             |
| 2+      | Equivocal* | Faint circumferential membrane staining >10% cancer cells<br>Intense circumferential membrane staining <10% cancer cells |
| 3+      | Positive   | Complete, intense membrane staining in > 10% cancer cells                                                                |

\*If Equivocal, further HER2 testing by *In Situ* Hybridization (FISH or SISH) – negative has less than 4 copies; equivocal has 4 – 6 copies and further testing needed, positive has more than 6 copies.

**Table 2.3. Immunohistochemical classifications of breast cancer**

| IHC Clinical |                                                | IHC Analysis |                           |
|--------------|------------------------------------------------|--------------|---------------------------|
| Clin-A       | ER/PR+/HER2-/Ki67<14%                          | A-like       | ER/PR+/HER2-/Ki67≤10%     |
|              |                                                | A-or B- like | ER/PR+/HER2-/10%<Ki67≤30% |
| Clin-B       | ER/PR+/HER2-/Ki67≥14%<br>ER/PR+/HER2+/Ki67 any | B-like       | ER/PR+/HER2-/Ki67>30%     |
|              |                                                | B/HER2-like  | ER/PR+/HER2+/Ki67 any     |
| Clin-HER2    | ER/PR-/HER2+/Ki67 any                          | HER2-like    | ER/PR-/HER2+/Ki67 any     |
| Clin-TNC     | ER/PR-/HER2-/Ki67 any                          | TNC-like     | ER/PR-/HER2-/Ki67 any     |

ER, estrogen receptor; PR, progesterone receptor, HER2, human epidermal growth factor 2; Ki67, marker of mitosis; + positive; - negative. A) IHC used in the clinical setting, shows the immunohistochemical markers used by the multidisciplinary teams to distinguish the breast cancers for subsequent treatment. B) IHC non clinical, are the groupings that we used for subsequent analyses for this study.

### 2.2.2. Intrinsic subtyping classification

Formalin fixed paraffin embedded (FFPE) specimens (n=384) of histologically confirmed invasive carcinoma were selected from the cohort. For patients who had primary surgery, an excision specimen (mastectomy/ wide local excision specimens) was used; but for patients who had neoadjuvant chemotherapy, a core biopsy specimen was used. The FFPE breast tissue blocks were retrieved from the archives of the NHL at CHBAH. The pathologist examined a haematoxylin & eosin (H&E) stained slide, marked the area of invasive breast cancer and cut 5 µM serial sections (Nanostring Technologies advised 10 µM sections, and the number of sections used for RNA extractions was doubled (Table 2.4)). The cut sections were sent to the Department of Internal Medicine Research Laboratory at the University of Witwatersrand, where the molecular work was undertaken. Since this was part of the SABCHO study, samples were age matched for HIV positive against HIV negative controls within a 5 year age band.

**Table 2.4. Number of required unmarked slides as per the tumor surface area (Prosigna)**  
(Appendix G)

| Measured tumor Surface Area on H&E stained slide (mm <sup>2</sup> ) | Number of unmarked slides* (10 µm) | Number of unmarked slides** (5 µm) |
|---------------------------------------------------------------------|------------------------------------|------------------------------------|
| 4-19                                                                | 6                                  | All slides provided                |
| 20-49                                                               | 3                                  | 6                                  |
| 50-99                                                               | 3                                  | 5                                  |
| ≥100                                                                | 1                                  | 3                                  |

\*The Prosigna manual (42) recommends cutting 10 µm sections, and choosing number of sections based on the area of the tumor. The NHLS, as a diagnostic lab, cuts \*\*5 µm sections, and we doubled the number of sections as we had half the depth.

The slides were deparaffinized with 2 changes of xylene (2 mins each), and 2 changes of absolute ethanol for 2 mins each. The sections were air-dried for 5-10 mins. Using the marked H&E section from the histopathologist, an outline was made of the tumor on the air-dried sections. The sections were rehydrated in 3% glycerol and excess glycerol removed with tissue paper (42). The non-cancerous tissue surrounding the tumor was scraped off using a No 11 Sterile surgical blade (Huaiyin Medical Instruments Co. Ltd, Huaian, China) and deposited in a 1.5 ml microcentrifuge tube. The tumor tissue was then scraped off with a No 11 surgical blade at a 45° angle and deposited in a sterile 1.5 ml microcentrifuge tube and stored at -80°C until processed for nucleic acid extraction (42).

### 2.2.3. RNA Extraction

RNA was extracted from the FFPE sections using the All Prep® DNA/RNA FFPE kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions (Appendix F).

The concentration of the purified RNA was calculated using the optical density at 260 nm using the Nanodrop 2000™ spectrophotometer (Thermo Fisher Scientific, Waltham, MA). For the PAM50 assay, the concentration of RNA used was recommended to be greater than 12.5 ng/µl and the A260/280 ratio between 1.7 and 2.3 (42). RNA was stored at -80°C until further use.

#### **2.2.4. PAM50 assay**

The extracted RNA was processed on the Nanostring nCounter® Analysis System (Nanostring Technologies, Seattle, WA). This system measures the relative abundance of each mRNA transcript of interest using a multiplexed hybridization reaction and a digital readout of fluorescent barcodes that become hybridized to each transcript. The PAM50 gene expression was measured on the nCounter SPRINT™ (Nanostring Technologies), as per the Prosigna® Breast Cancer Prognostic Gene signature assay Package insert (42). The system is based on the principle that gene-specific probe pairs hybridize directly with the mRNA. A list of the 50 genes of interest and the 8 internal reference genes are included in appendix G. This list also shows the type of gene, whether it is associated with a luminal, HER2, proliferative, basal subtype or is considered a reference gene. The 50 gene expression algorithm assigns the genes into centroids, that profiles the tumors into luminal-A, lumina- B, HER2-enriched and basal-like (figure 2.2). An unknown sample was compared with these centroids and assigned a molecular subtype (42).

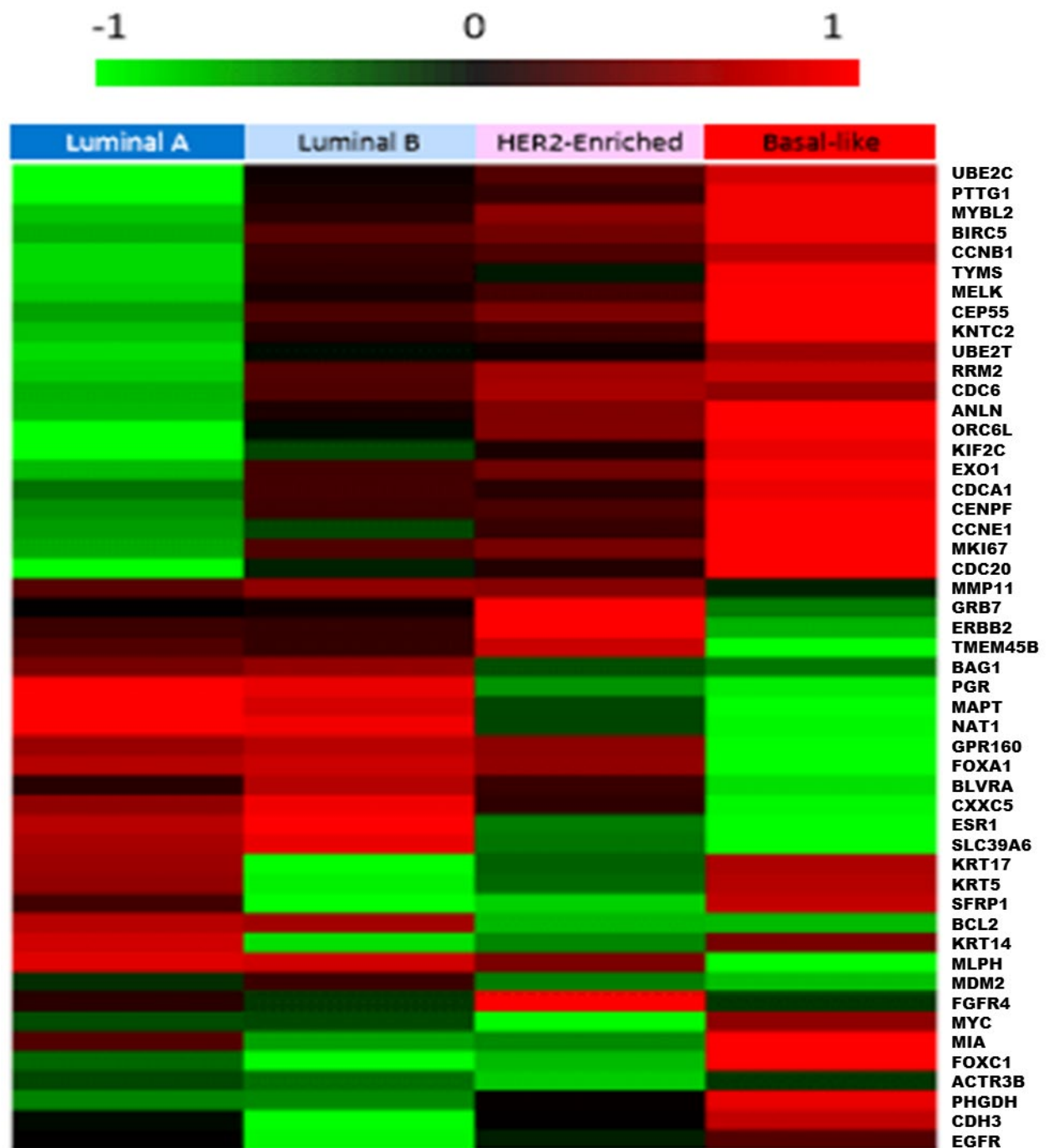


**Figure 2.2. Heatmap of PAM50 genes by subtype.**

Hierarchical clustering of 50 genes analyzed by the PAM50 Nanostring assay. Vertical axis: 50 genes analyzed. Horizontal axis: Breast cancer subtype. Each colour represents the measured expression ratio, with the colour intensity being directly proportional to the magnitude of the gene expressed, as was initially described by Perou and colleagues (29).

Unknown samples would be compared to these centroids and assigned a molecular subtype according to the subtype algorithm. This representative heatmap from Nanostring indicates higher expression with red and lower expression of gene is shown in green (from Nanostring <https://www.nanostring.com/products/ncounter-assays-panels/oncology/ruo-pam50-codeset/>, downloaded 22 Oct 2021)

The RNA samples were thawed on ice, briefly centrifuged and placed back on ice. If appropriate, the RNA was diluted to 50 ng/μl. Ten microlitres of RNA were added to a 200 μl PCR tube to a final amount of 250 ng RNA. The reporter codeset (GXA-PICS, Nanostring Technologies) and capture probeset (GXA-PICP, Nanostring Technologies) were thawed at room temperature. Reporter master mix was prepared by adding 130 μl hybridization buffer to 65 μl reporter codeset, mixed by pipetting, centrifuged briefly and left at room temperature. Fifteen microliters of the reporter master mix was added to the RNA, followed by 5 μl of capture probeset, mixed and briefly centrifuged. The tubes were incubated for exactly 18 hours at 65°C in an MJ Mini Biorad Thermocycler (Hercules, CA).

Half an hour before the hybridization finished, the cartridges were equilibrated to room temperature. After 18 hours, the hybridized RNA was removed from the thermocycler and briefly centrifuged. Water (2 μl) was added to the hybridized sample and centrifuged. Thereafter, 30 μl of sample was loaded onto the cartridge. Excess liquid was blotted from the top of the cartridge, a seal was applied over the sample leading ports, the reagent ports on the cartridge were unsealed, and the cartridge was loaded onto the Nanostring SPRINT™.

**2.2.5. Analysis of the PAM50 Assay**

The SPRINT™ instrument generates reporter code counts (RCC) for each gene and it determines quality control for each sample on each cartridge, by determining the binding density (i.e. the barcodes seen by the instrument) and image quality (how much field of view seen by the instrument). Each sample meeting the QC threshold is reported as one of the four molecular

subtypes: luminal-A, luminal-B, HER2-enriched and basal-like. Subtypes were determined using PAM50 classification algorithm, which simultaneously measures the expression level of 50 genes, 8 housekeeping genes used for signal normalization, 6 positive control and 8 negative controls. The expression of the 50 genes to distinguish the different molecular subtypes is shown in figure 2.2. Typically, luminal-A is ER+ and has lower proliferation markers and lowered risk of recurrence after primary surgery. Luminal-B is also usually ER+ although it has higher expression of proliferation markers and higher risk of recurrence (ROR) after primary treatment. HER2-enriched tumors are mainly HER2+, and can be ER+ or ER- and the ROR is higher. Basal-like is usually ER-/HER2- with high proliferation markers (42). The Risk of recurrence score is a prediction of the recurrence in ER + breast cancer after surgery and 5 years treatment with endocrine therapy (41,119). The ROR depends on the intrinsic subtypes, the size of the tumor, nodal involvement and proliferation score. To calculate ROR, we use:

$$ROR = aR_{LumA} + bR_{LumB} + cR_{Her2} + dR_{Basal} + eP + fT$$

Where R is the Pearson correlation coefficient of the PAM50 expression profile for the tumor compared to the centroids for each of the intrinsic subtypes, P is the proliferation score, and T is the tumor stage (41,119,120).

### **2.3. FIBROBLAST GROWTH FACTOR RECEPTOR 2 SNP GENOTYPING (rs2981582, rs2981578, rs35054928, rs11200014)**

#### **2.3.1. Study Population**

For this part of the study, 1001 participants with breast cancer who were diagnosed at the Surgical Breast Unit at CMJAH and the Batho Pele Breast unit at CHBAH were included. Control participants (n=1006) were recruited from these sites after confirmation that they did not have breast cancer, as well as at other clinics at CMJAH where they were undergoing routine assessment unrelated to cancer. All participants had a sample of peripheral blood drawn and collected in ethylenediaminetetraacetic acid (EDTA) vacutainer tubes. Of the cases, 3 patients were found to have DCIS and 1 patient had an ambiguous diagnosis, so these samples were excluded from the analysis. The DNA samples were sent for genotyping at LGC Diagnostics.

### **2.3.2. Genomic DNA extraction**

DNA was extracted from the whole blood using a modified salting out method (121,122). This method involves disrupting the cell membranes and digesting the cell contents using a detergent (sodium dodecyl sulphate) – proteinase K mix. An extremely high salt concentration (saturated NaCl) was added to the solution to denature the proteins. After centrifugation, the proteins were located in the pellet, while nucleic acids were in the supernatant. The DNA was precipitated in ethanol and resuspended in Tris-EDTA buffer. (The salting out method is shown in Appendix G.) The Nanodrop 2000™ spectrophotometer (Thermo Scientific, Waltham, MA) was used to determine DNA concentration and A260/280 ratios. We used DNA with an A260/280 ratio of between 1.7-2.0. The DNA was stored at -20°C until further analysis.

### **2.3.3. Genotype rs2981582, rs2981578, rs35054928, rs11200014**

The DNA samples underwent serial dilution in low TE buffer (10 mM Tris HCl; 0.1 mM EDTA, pH 8.3) to a final concentration of 25 ng/μl DNA in a volume of 20 μl. The samples were shipped to LGC Genomics (Hoddesdon, UK) for genotyping using KASP (Kompetitive allele specific PCR) technology. This trademarked method has 2 allele-specific forward primers and a common reverse primer. The forward primers each have a unique tail sequence that corresponds with a fluorescent resonant energy transfer cassette; one is labeled with FAM™ dye and the other with HEX™ dye. One allele binds the forward primer with FAM™ and the second allele binds the HEX™ labeled primer. Primers for the SNPs are shown in table 2.5. The results with the corresponding alleles were returned to us, and we have subsequently analyzed the data. The SNPs rs2981582, rs2981578, rs35054928, rs11200014 were successfully genotyped, while rs7895676 was not successfully genotyped. These SNPs were chosen because of their high association with breast cancer in GWAS studies (rs2981582), or breast carcinoma in African Americans (rs2981578, rs35054928, rs11200014) (Table 2.5).

**Table 2.5. *FGFR2* Variants associated with breast cancer**

| SNP        | Allele | Risk allele | Population       | Reference | HWE p |
|------------|--------|-------------|------------------|-----------|-------|
| rs2981582  | G      | A           | European/Asian   | (66)      | 0.09  |
| rs35054928 | -C     | CC          | African American | (78)      | 0.11  |
| Rs2981578  | A      | G           | African American | (78)      | 0.39  |
| rs11200014 | A      | G           | African American | (99)      | 0.50  |

SNP, Single nucleotide polymorphism; HWE, Hardy-Weinberg equilibrium p value. The table shows the risk alleles associated with breast cancer, the populations that were identified to have risk of breast cancer with the risk alleles and HWE found in our South African population.

#### **2.3.4. Analysis of *FGFR2* Variants**

Pathologists at the NHLS provided data on histological diagnosis, tumor subtype, grade, and immunolocalization of ER, PR, HER2 and Ki67, as detailed in section 2.2.1. The data was interrogated with the SNPs.

Continuous variables were assessed for normality using the Shapiro-Wilks test. Normally distributed variables were presented as means and standard deviations, and non-normally distributed data as medians and interquartile ranges. When two groups were compared, normally distributed continuous variables were compared using a t-test, while non-parametric variables were compared using a Mann-Whitney test. Non-parametric data were compared for more than 2 groups using a Kruskal Wallis test, followed by *post hoc* analysis using Dunn's Pairwise Comparison test if the Kruskal Wallis test showed significant differences. Categorical variables were expressed as frequencies and percentages and compared using a Pearson's  $\chi^2$  test. All statistical analyses were done using STATA v14.2 (College Station, Texas).

Using the genotyping data provided by LGC genomics, we calculated both genotype and allele frequencies for each SNP of interest. The frequencies of each genotype (homozygous major, heterozygous and homozygous minor) were compared between cases and controls for the SNPs of interest using a Pearson's chi square test. Genotype frequencies were also examined under four different models of genetic disease risk (125). Results were analyzed using dominant, recessive, multiplicative, additive, and homozygous models. A Bonferroni correction for multiple

comparison testing was applied. A value of  $p=0.008$  ( $0.05/6$ ) was considered statistically significant.

## **2.4. *FGFR2* mRNA EXPRESSION AND *FGFR2* IIIB and IIIC SPLICE ISOFORMS ASSOCIATED WITH BREAST CANCER**

### **2.4.1. *FGFR2* gene expression levels in breast cancer**

Using cBioPortal for cancer genomics (<https://www.cbioportal.org/>), 8993 samples from 14 breast cancer studies, including both general and invasive breast cancer cases, were selected for interrogating *FGFR2* expression. The studies used were: Breast Cancer (MSK, Cancer Cell 2018) (126); Breast Cancer (MSK, Nature Cancer 2020) (127); Breast Cancer (HTAN, 2022); Breast Cancer (METABRIC, Nature 2012 and Nat Commun 2016) (128–130); Breast Cancer (MSK, Clinical Cancer Res 2020) (131); Breast Cancer (MSK, NPJ Breast Cancer 2019) (132); Breast Cancer (SMC, 2018) (133); Breast Invasive Carcinoma (British Columbia, Nature 2012) (134); Breast Invasive Carcinoma (Broad, Nature 2012) (135); Breast Invasive Carcinoma (Sanger, Nature 2012) (136); Breast Invasive Carcinoma (TCGA, Cell 2015) (137); Breast Invasive Carcinoma (TCGA, Firehose Legacy); Breast Invasive Carcinoma (TCGA, Nature 2012) (138); Breast Invasive Carcinoma (TCGA, PanCancer Atlas) (139–148).

Plot data of log transformed *FGFR2* gene expression levels were downloaded in .csv format for PAM50 subtypes (luminal-A, luminal-B, HER2-enriched and basal-like); immunohistochemistry subtypes (ER+, ER+HER2+, HER2+ and TNBC); HER2 immunohistochemistry score and histological subtype (invasive ductal and invasive lobular carcinomas). *FGFR2* mRNA expression data was downloadable for 117 samples with a PAM50 subtype (5 samples with normal-like subtype excluded), 163 samples with an immunohistochemistry subtype (5 samples with unknown subtype excluded), 275 samples with a HER2 score, and 303 sample with a histological subtype.

Data were imported into STATA v14.2 (StataCorp, College Station, TX, USA). A Shapiro-Wilk normality test was performed and data subsequently analyzed non-parametrically. *FGFR2*

expression levels were represented as box and whisker plots with outliers excluded. Boxes showed medians, upper and lower quartiles while whiskers showed non-outlying maximum and minimum values. A Kruskal-Wallis equality-of-populations rank test was used to compare expression levels, followed by Dunn's test of multiple comparisons to compare expression levels between each group. A Bonferroni correction for multiple comparison testing was applied and p-values of less than 0.01 considered significant.

#### **2.4.2. FGFR2 transcript variants**

Splice variants of the *Homo Sapiens* FGFR2 gene were downloaded from Ensembl Genome Browser 108 (<https://www.ensembl.org/index.html>) along with the reference genome sequence (GRCh38:CM000672.2). *FGFR2* has 41 transcript variants of which 25 are protein coding, 1 is a non-stop decay transcript, 2 undergo nonsense mediated decay, 5 have non-defined protein coding sequences (CDS) and 8 have retained introns (supplementary table S5.1). Transcript variants were aligned to the GRCh38 reference genome using the multiple sequence alignment tool MUSCLE (<https://www.ebi.ac.uk/Tools/msa/muscle/>). Aligned sequences were visualized using GeneDoc v2.7, a multiple sequence alignment editor (<https://github.com/karlricholas/GeneDoc>) (149).

#### **2.4.3. FGFR2 splice variants**

Splicing patterns of the *FGFR2* gene in breast cancer were interrogated using the TCGA SpliceSeq database (<https://bioinformatics.mdanderson.org/TCGASpliceSeq/singlegene.jsp>). Percent spliced in (PSI) values were downloaded for all types of splicing events with filters removed. Information from the Exon Table was used to compare the sequence and location of the TCGA SpliceSeq exons with the transcript variants downloaded from ensembl.org and thus correlate the TCGA SpliceSeq exon numbers with those reported in the literature. TCGA SpliceSeq exon 14 correlated to GRCh38 exon 7, TCGA SpliceSeq exon 15 to GRCh38 exon 8, TCGA SpliceSeq exon 16 to GRCh38 exon 9 and TCGA SpliceSeq exons 17.1 and 17.2 correlated to GRCh38 exon 10 (figure 2.3).

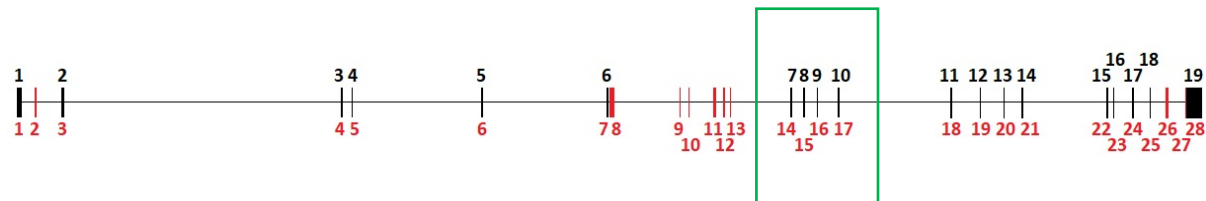


**Figure 2.3. Representative FGFR2 gene map**

The gene map shows all known splicing events derived from TCGA SpliceSeq data correlated to the GRCh38 sequence. Exons are shown in **black**, consistent with the GRCh38 numbering. Alternatively spliced-in-regions are shown in **red**, as is the TCGA SpliceSeq numbering. The splice variants of interest were exons 7, 8, 9, 10 (GRCh38) which correlated to 14, 15, 16, 17.1, 17.2 (TCGA SpliceSeq), highlighted in the **green** box.

Using the TCGA SpliceSeq database, PSI data were transposed and imported into STATA v14.2. Five exon skipping splicing events (alternative splicing identification (as\_id) 13303: from exon 15 to exon 17.1, skipping exon 16; as\_id 13305: from exon 14 to exon 16, skipping exon 15; as\_id 126925: from exon 14 to exon 17.1, skipping exons 15 and 16; as\_id 126926: from exon 14 to exon 17.1, skipping exon 16 and as\_id 126927: from exon 14 to exon 17.1, skipping exon 15) were identified as of interest. Analysis was restricted to samples from tumor tissue only. A Shapiro-Wilk normality test showed data for these splicing events to be non-normally distributed and so results were presented non-parametrically as medians and interquartile ranges.

The TCGA Splicing Variants Database (<http://tsvdb.com/plot.html>) (150) was used to download data on the expression levels of the FGFR2 gene and FGFR2 exons in breast cancer patients along with their clinical estrogen, progesterone and HER2 receptor status. Samples included were restricted to those of female gender. Size and location of each exon was compared to that of the transcript variants downloaded from ensembl.org to correlate the TSVdb exon numbers to those reported in the literature. Data was imported into STATA v14.2. For each sample, expression levels of individual exons were normalized to the overall expression of the FGFR2 gene. A Shapiro-Wilk normality test found these normalized expression levels of the exons to be non-normally distributed and so results were presented non-parametrically as box and whisker plots. Boxes

showed medians, upper and lower quartiles while whiskers showed non-outlying maximum and minimum values. Expression levels between the two exons of interest were compared using a Wilcoxon rank-sum test. A Bonferroni correction for multiple comparison testing was applied and p-values of less than 0.005 taken as significant.

### **FORWARD TO CHAPTER 3**

There is a paucity of data on Black, African women with breast cancer. We investigated how the more modern, molecular PAM50 subtyping compared with immunohistochemical techniques that are used for clinical subtyping in the South African public health sector. This data is important as it is used to inform clinical treatments, such as the choice between hormone receptor (HR) antagonists or epidermal growth factor (HER2) antibodies, in which case it is crucial know the status of these markers. This information also indicates whether a tumor is highly proliferative (as indicated by Ki67 levels) and chemotherapy would be beneficial to the patient.

The paper was submitted to Breast Cancer Research and Treatment. The authors affiliations and contributions are below.

## Discordance between PAM50 intrinsic subtyping and Immunohistochemistry in South African Women with Breast Cancer

### **Authors:**

Thérèse Dix-Peek<sup>1</sup>, Boitumelo P. Phakathi<sup>2,3</sup>, Eunice J. van den Berg<sup>4,5</sup>, Caroline Dickens<sup>1</sup>, Tanya N. Augustine<sup>6</sup>, Herbert Cubasch<sup>7,8</sup>, Alfred I. Neugut<sup>9,10</sup>, Judith S. Jacobson<sup>9,10</sup>, Maureen Joffe<sup>8,11</sup>, Paul Ruff<sup>1,8</sup>, Raquel A.B. Duarte<sup>1</sup>

### **Affiliations:**

<sup>1</sup>Department of Medicine, University of the Witwatersrand, Faculty of Health Sciences, 7 York Road, Johannesburg, 2193, South Africa

<sup>2</sup>Charlotte Maxeke Surgical Breast Unit, Charlotte Maxeke Johannesburg Academic Hospital, Jubilee Road, Johannesburg, 2196, South Africa

<sup>3</sup>Department of Surgery, University of the Witwatersrand, Faculty of Health Sciences, 7 York Road, Johannesburg, 2193, South Africa

<sup>4</sup>Department of Histopathology, National Health Laboratory Service, Chris Hani Baragwanath Hospital, 26 Chris Hani Road, Diepkloof, Johannesburg 1864, South Africa

<sup>5</sup>Department of Anatomical Pathology, University of the Witwatersrand, 7 York Road, Johannesburg, 2193, South Africa

<sup>6</sup> School of Anatomical Sciences, Faculty of Health Sciences, University of the Witwatersrand, 7 York Road, Parktown, 2193 Johannesburg, South Africa

<sup>7</sup>Batho Pele Breast Unit, Chris Hani Baragwanath Academic Hospital, 26 Chris Hani Road, Diepkloof, Soweto, 1860, South Africa

<sup>8</sup>Non-Communicable Diseases Research Division, Wits Health Consortium (PTY) Ltd, 31 Princess of Wales building, Johannesburg, 2193, South Africa

<sup>9</sup>Herbert Irving Comprehensive Cancer Centre, Vagelos College of Physicians and Surgeons, Columbia University, New York 10032, United States of America

<sup>10</sup>Department of Epidemiology, Mailman School of Public Health, Columbia University, New York 10032, United States of America

<sup>11</sup>MRC Developmental Pathways to Health Research Unit, Department of Paediatrics, University of the Witwatersrand, Faculty of Health Sciences, 7 York Road Johannesburg, 2193. South Africa

### **Authors' Contributions**

Concept design RD, TDP, BP; Collecting patient data MJ, HC; Histology EVB; Analysis and interpretation of the IHC and PAM50 data TDP, BP, TA, RD, CD; Clinical perspective MJ, HC, PR, AN, JJ, BP; Manuscript primarily written by TDP, RD, BP, CD, TA with input from all other authors.

## **CHAPTER 3: DISCORDANCE BETWEEN PAM50 INTRINSIC SUBTYPING AND IMMUNOHISTOCHEMISTRY IN SOUTH AFRICAN WOMEN WITH BREAST CANCER**

### **3.1. ABSTRACT**

#### **PURPOSE:**

Breast cancer is a heterogeneous disease with different gene expression profiles, treatment options and outcomes. In South Africa, tumors are classified using immunohistochemistry. In high-income countries multiparameter genomic assays are being utilized with implications for tumor classification and treatment.

#### **METHODS:**

In a cohort of 378 breast cancer patients from the SABCHO study, we investigated the concordance between tumor samples classified by IHC and the PAM50 gene assay.

#### **RESULTS:**

IHC classified patients as ER-positive (77.5%), PR-positive (70.6%), and HER2-positive (32.3%). These results, together with Ki67, were used as surrogates for intrinsic subtyping, and showed 6.9% IHC-A-clinical, 72.7% IHC-B-clinical, 5.3% IHC-HER2-clinical and 15.1% triple negative cancer (TNC). Typing using the PAM50 gave 19.3% luminal-A, 32.5% luminal-B, 23.5% HER2-enriched and 24.6% basal-like. The basal-like and TNC had the highest concordance, while the luminal-A and IHC-A group had the lowest concordance. By altering the cutoff for Ki67, and realigning the HER2/ER/PR-positive patients to IHC-HER2, we improved concordance with the intrinsic subtypes.

#### **CONCLUSION:**

We suggest that the Ki67 be changed to a cutoff of 20-25% in our population to better reflect the luminal subtype classifications. This change would inform treatment options for breast cancer patients in settings where genomic assays are unaffordable.

### **3.2. INTRODUCTION**

Breast cancer is the most commonly diagnosed cancer among South African women accounting for 27.1% of all cancers diagnosed in these women (25). Breast cancer diagnoses on the African continent have been steadily increasing over the past decades, attributed to longer lifespans and changes in lifestyle associated with westernization. In Africa, mortality rates are higher than in Europe and the United States, largely due to late stage at diagnosis and fewer treatment options (2,3). Breast cancer is a heterogeneous disease, differing in gene expression patterns, growth rates, responses to treatment and clinical outcomes.

Breast tumors can be subtyped by immunohistochemistry (IHC) which investigates the expression of four biomarkers: estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2/neu) and a marker of proliferation, Ki67. These markers distinguish ER-positive A-like, ER-positive B-like; HER2-enriched and triple negative (TNC) tumors (Table 1) (27–29). Analysis of expression of the hormone receptors (ER and PR) is a semi-quantitative method based on the Allred Score (116,151,152).

The proliferation marker Ki67 is used to distinguish between the luminal subtypes (32,153) and was adopted as a marker by the St Gallen International Consensus on Breast Cancer (28). Ki67 was introduced as a diagnostic marker in South Africa in 2013 and is indicative of proliferation if Ki67 expression is  $\geq 14\%$  (28,32). However, the optimal Ki67 cut off value to distinguish luminal A-like tumors from luminal B-like tumors remains controversial due to uncertainty about how to classify tumors with intermediate (10-30%) Ki67 levels (33). The 2015 St Gallen's suggested a cutoff of 20-29% be used to distinguish A-like and B-like subtypes, along with clinical validation (34). In addition, IHC for Ki67 analysis lacks reproducibility across laboratories (35). Immunohistochemical results can be affected by the duration of fixation, type of fixative used, speed of assay and completeness of dehydration (6). Moreover, the assessment is subject to interpretation by the histopathologist. In South Africa, the Department of Health 2018 recommendations are to use a Ki67 cutoff of 14 % (118), although there is ongoing debate to the

best cutoff of Ki67 to distinguish between luminal subtypes (36). Ki67 cutoffs of both 14% and 20% are currently used at different centers.

The last decade has seen the development of many commercialized multigene tests to guide treatment and provide prognostic information for patients with breast cancer. The PAM50/Prosigna assay has a 50-gene signature that groups tumors into intrinsic molecular subtypes luminal-A, luminal-B, HER2-enriched and basal-like (42). The PAM50 assay is less subjective than the IHC-based techniques, but is much more expensive and labor intensive than IHC. In South African public hospitals, IHC continues to be used for clinical subtyping because of its lower cost.

A recent study from South Africa found that 64.9% of patients were diagnosed by IHC4 as B-like, 15.3% as TNC, 13.8% as A-like, and 6.0% as HER2-enriched (43). An earlier country-wide study, found that Black South African women had higher levels of ER-negative and PR-negative tumors than women of European, South Asian or admixture heritage, but did not have significantly different HER2 levels (6). More recently, a study showed that white South African women had similar IHC profiles to European women and white American women, with more aggressive subtypes predominant in young women and less aggressive subtypes in older women, whereas black South African women did not have substantial profile changes according to age (44).

This study examines the concordance between PAM50 molecular subtyping assigned and the IHC results currently used for the management of breast cancer diagnosed within the South African Public Health System, focusing on varying Ki67 cutoffs. The data generated should help to inform cutoff values for IHC and may lead to better management of breast cancer in South Africa and other settings where genomic subtyping is unaffordable.

### 3.3. **METHODS**

#### **3.3.1. Study participants**

The South African Breast Cancer and HIV Outcomes (SABCHO) cohort (111) studied patients recruited at the breast clinic of Chris Hani Baragwanath Academic Hospital (CHBAH), Soweto, South Africa. Participants were consenting women with biopsy-confirmed breast cancer who self-identified as Black African. Exclusion criteria were age <18 years or current pregnancy. Clinical staging was according to the American Joint Committee on Cancer (AJCC) system (154). The study was approved by the Human Research Ethics Committee (Medical) at the University of the Witwatersrand (M161116).

#### **3.3.2. IHC classification of tumors**

Histopathological characteristics for 384 patients, obtained from the National Health Laboratory Service (NHLS), included histological type, tumor grade, ER, PR, HER2 scoring and Ki67. All tissues for this study were processed at CHBAH NHLS Laboratory, following College of American Pathologist guidelines. Immunostaining was performed on the Benchmark XT automatic platform. The tumors were classified according to the St Gallen's Guidelines (28,34,155). The Allred score was used to determine ER/PR status, with a value of 0-2 considered negative, and 3-8 considered positive (115–117). Tumors were HER2 positive if they scored 3+ by IHC, or 2+ by IHC with fluorescent *in situ* hybridization (FISH) confirmation. The Ki67 antibody used was 30-9 (Roche diagnostic, Ventana, USA), and multiple scorers at the same laboratory assessed the Ki67 stains. Percentage of proliferation was determined by visual estimation (36). The cut-off for the proliferation marker Ki67 is unresolved. The multidisciplinary team at CHBAH uses a Ki67 score of 20% in conjunction with the Allred score, grade and age of patient as a cut off for chemotherapeutic treatment in HR positive breast cancers. We additionally explored cutoffs of 10%, 15%, 20%, 25% and 30%, because of the uncertainty surrounding those values for clinical decision making (33).

We assigned IHC used for clinical decision making as follows: Clin-A (HR+/HER2-/Ki67≤14%); Clin-B (HR+/HER2-/Ki67>14%, or HR+/HER2+/Ki67 any); Clin-HER2 (HR-/HER2+/Ki67 any); and Clin-TNC (HR-/HER2-/Ki67 any). The IHC subtyping surrogates were assigned as: A-like (HR+/HER2-/Ki67≤10%); A- or B-like (HR+/HER2-/10%<Ki67≤30%); B-like (HR+/HER2-/Ki67>30%); B/HER2-like (HR+/HER2+/Ki67 any); HER2-like (HR-/HER2+/Ki67 any); TNC (HR-/HER2-/Ki67 any). Both the clinical IHC subtypes as well as the IHC subtyping surrogates were compared with the PAM50 Intrinsic subtypes: luminal-A; luminal-B; HER2-enriched and basal-like (Table 1).

**Table 1. Immunohistochemical surrogates for expected PAM50 Intrinsic subtypes, showing IHC used for clinical decisions and Ki67 cutoff values for concordance analysis**

| PAM50             | IHC Clinical* |                        | IHC Analysis  |                            |
|-------------------|---------------|------------------------|---------------|----------------------------|
| Intrinsic Subtype | Designation   | IHC expression         | Designation   | IHC expression             |
| Luminal-A         | Clin-A        | HR+ / HER2- / Ki67≤14% | A-like        | HR+ / HER2- / Ki67≤10%     |
|                   |               |                        | A- or B- like | HR+ / HER2- / 10%<Ki67≤30% |
| Luminal-B         | Clin-B        | HR+ / HER2- / Ki67>14% | B-like        | HR+ / HER2- / Ki67>30%)    |
|                   |               | HR+ / HER2+ / Ki67 any |               |                            |
| HER2-enriched     |               |                        | B/HER2-like   | HR+ / HER2+ / Ki67 any     |
|                   | Clin-HER2     | HR- / HER2+ / Ki67 any | HER2-like     | HR- / HER2+ / Ki67 any     |
| Basal-like        | Clin-TNC      | HR- / HER2- / Ki67 any | TNC           | HR- / HER2- / Ki67 any     |

HR, hormone receptor, indicates either estrogen receptor and/or progesterone receptor; HER2, human epidermal growth factor receptor 2; TNC, triple negative cancer; +, positive; -, negative.

\*Current multidisciplinary tumor board guidance algorithms for treatment decision making.

### 3.3.3. PAM50 intrinsic subtyping

Formalin-fixed, paraffin embedded blocks were cut into 5 µm serial sections; the area of tumor was identified and marked on a hematoxylin and eosin (H&E) section. If available, primary surgery

blocks were preferentially chosen. If the surgery section was unavailable, or if the patient received neoadjuvant chemotherapy or radiation therapy prior to surgery, a biopsy section was used.

RNA was purified from the FFPE sections using the All Prep® DNA/RNA FFPE kit (Qiagen, Hilden, Germany). The RNA concentration was calculated using the optical density at 260 nm on the Nanodrop 2000™ spectrophotometer (Thermo Fisher Scientific, Waltham, MA). The extract was deemed suitable for further analysis if the concentration of RNA was greater than 12.5 ng/μl and the A260/280 ratio was 1.7-2.3. Following RNA extraction, 384 samples were of sufficient quantity and quality for molecular typing.

The PAM50 gene expression was measured on the nCounter SPRINT™ (Nanostring Technologies, Seattle, WA), as per the Prosigna® Breast Cancer Prognostic Gene signature assay Package insert (42). (The 50 genes and 8 housekeeping genes are shown in supplementary Table S1 and an example of the resultant heat map are shown in supplementary figure S1.) nSOLVER 4.0 was used to retrieve the rcf files and perform QC analysis, background subtraction and normalization. Of the 384 samples, 378 passed QC and underwent further analysis; classification of intrinsic subtype was done at NanoString (Seattle, WA). Quality control (QC) of the data was performed by NanoString Technologies, Inc. using their proprietary software, nSolver. For mRNA samples, as used in this study, QC is performed at a number of stages. Imaging QC flags samples if less than 75% of the imaging surface can be read. Binding density QC calculates the barcodes/micron<sup>2</sup>, samples with binding densities between 0.05 and 1.8 are usable with optimal binding densities being around 1.4 barcodes/micron<sup>2</sup>. The PAM50 panel includes both positive and negative controls which are assessed by geometric mean. Positive controls are synthetic RNA targets, spiked in at known concentrations, that are used to ensure proper hybridization and lack of RNase contamination in the samples and to establish limits of detection (the 0.5 fM positive control must be more than 2 standard deviations above the mean of the negative controls to pass QC). Positive controls are also used in normalization QC by generating scaling factors that

must be between 0.3 and 3 to pass QC. Negative controls are probes for which no known target exists in biological samples and are used to establish background levels of detection.

### **3.3.4. Statistical analysis**

Continuous variables were assessed for normality using the Shapiro-Wilks test. The data were described by mean  $\pm$  standard deviation for normally distributed variables and median (interquartile range) for non-normally distributed variables. Categorical variables were described as frequencies and percentages. Statistical analyses were done using STATA v14.2 (College Station, Texas). Significance between the groups was determined using Pearson's  $\chi^2$  test or the Kruskal Wallis rank test, and *post hoc* analysis using Dunn's Pairwise Comparison test. A p-value  $<0.05$  was considered significant. Agreement in subtype call between the IHC and PAM50 subtyping methods was assessed using the kappa statistic. To allow for comparable groups with this method, the IHC results were classified as follows: Clin-A (HR+/HER2-/Ki67 $\leq$ 14%), Clin-B (HR+/HER2-/Ki67 $>$ 14%), Clin-HER2 (HR any/HER2+/Ki67 any) and TNC (HR-/HER2-/Ki67 any).

## **3.4. RESULTS**

### **3.4.1. Characteristics of the study cohort**

The clinicopathological characteristics are shown in table 2. The mean age of study participants was 49.7 years. Most patients had stage II or III cancers, and were more likely to have grade-2 or -3 tumors between 20 mm- 50 mm (AJCC T2), with some nodal involvement.

The intrinsic subtyping distribution by the PAM50 assay was 19.3% luminal-A (n=73), 32.5% luminal-B (n=123), 23.5% HER2-enriched (n=89) and 25.6% basal-like (n=93) (figure 1a, Table 2). When classified by IHC, most patients (79.6%) were HR positive (with, or without, HER2) (figure 1b). Although the intrinsic subtypes (figure 1a) show roughly equal numbers of luminal-A (19.3%), luminal-B (32.5%), HER2-enriched (23.5%) and basal-like (24.6%) subtypes, the clinical IHC results show a massive predominance of Clin-B subtype (72.7%), and only 6.9% Clin-A, 5.3% Clin-HER2

and 15.1% Clin-TNC. High grade (3) Clin-A subtype, treated as Clin-B by the multidisciplinary team, only accounted for 0.53% (Table 2) of the total cohort, and did not meaningfully affect the concordance with the molecular subtypes.

**Table 2. Clinicopathological characteristics of the study cohort**

|                                                            | <b>n=378</b>           | <b>%</b> |
|------------------------------------------------------------|------------------------|----------|
| <b>Age mean (<math>\pm</math>SD)</b>                       | 49.71yr ( $\pm$ 11.02) |          |
| <b>Stage at diagnosis</b>                                  |                        |          |
| Stage I                                                    | 13                     | 3.44     |
| Stage II                                                   | 152                    | 40.21    |
| Stage III                                                  | 173                    | 45.77    |
| Stage IV                                                   | 40                     | 10.58    |
| <b>Grade (n=374)</b>                                       |                        |          |
| Grade 1                                                    | 24                     | 6.42     |
| Grade 2                                                    | 207                    | 55.35    |
| Grade 3                                                    | 143                    | 38.24    |
| <b>Tumor stage* (n=378)</b>                                |                        |          |
| T1                                                         | 35                     | 9.26     |
| T2                                                         | 168                    | 44.44    |
| T3                                                         | 77                     | 20.37    |
| T4                                                         | 98                     | 25.93    |
| <b>Histology Diagnosis or Histological subtype (n=372)</b> |                        |          |
| Invasive Ductal                                            | 329                    | 88.4     |
| Invasive Lobular                                           | 15                     | 4.0      |
| Invasive Mucinous                                          | 10                     | 2.7      |
| Other**                                                    | 18                     | 4.8      |
| <b>Nodal Involvement (n=375)</b>                           |                        |          |
| Absent (0 nodes)                                           | 82                     | 21.87    |
| 1-3 nodes                                                  | 175                    | 46.67    |
| 4-9 nodes                                                  | 88                     | 23.47    |
| 10+ nodes                                                  | 30                     | 8.00     |
| <b>ER (n=378)</b>                                          |                        |          |
| Positive                                                   | 293                    | 77.51    |
| Negative                                                   | 85                     | 22.49    |
| <b>PR (n=378)</b>                                          |                        |          |
| Positive                                                   | 267                    | 70.63    |
| Negative                                                   | 111                    | 29.37    |
| <b>HER2 (n=378)</b>                                        |                        |          |
| Positive                                                   | 122                    | 32.28    |
| Negative                                                   | 256                    | 67.72    |

|                                             |     |       |
|---------------------------------------------|-----|-------|
| <b>Allred score (n=369)</b>                 |     |       |
| Negative (Allred = 0,1,2)                   | 66  | 17.87 |
| Low (Allred = 3,4)                          | 30  | 8.13  |
| Intermediate (Allred= 5,6)                  | 25  | 6.78  |
| High (Allred=7,8)                           | 248 | 67.21 |
| <b>Ki67 (n=375)</b>                         |     |       |
| 0%<Ki67≤10%                                 | 33  | 8.80  |
| 10%<Ki67≤15%                                | 12  | 3.20  |
| 15%<Ki67≤20%                                | 30  | 8.00  |
| 20%<K67≤25%                                 | 10  | 2.67  |
| 25%<Ki67≤30%                                | 40  | 10.67 |
| 30-100%                                     | 250 | 66.67 |
| <b>PAM50 (n=378)</b>                        |     |       |
| Luminal-A                                   | 73  | 19.31 |
| Luminal-B                                   | 123 | 32.54 |
| HER2-enriched                               | 89  | 23.54 |
| Basal-like                                  | 93  | 24.60 |
| <b>IHC Clinical (n=377)</b>                 |     |       |
| Clin-A (HR+/HER2-/Ki67≤14%)                 | 26  | 6.90  |
| Clin-B (HR+/HER2-/Ki67>14% or HR+/HER2+)    | 274 | 72.68 |
| Clin-HER2 (HR-/HER2+)                       | 20  | 5.31  |
| Clin-TNC (HR-/HER2-)                        | 57  | 15.11 |
| <b>IHC Clinical with grading*** (n=374)</b> |     |       |
| Low grade Clin-A                            | 23  | 6.15  |
| High grade Clin-A                           | 2   | 0.53  |
| Low grade Clin-B                            | 178 | 47.59 |
| High grade Clin-B                           | 94  | 25.13 |
| Clin-HER2                                   | 20  | 5.35  |
| Clin-TNC                                    | 57  | 15.24 |
| <b>IHC Analysis (n=377)</b>                 |     |       |
| A-like (HR+/HER2-/Ki67≤10%)                 | 26  | 6.90  |
| A-or B- like (HR+/HER2-/10%<Ki67≤30%)       | 55  | 14.59 |
| B-like (HR+/HER2-/ Ki67>30%)                | 117 | 31.03 |
| B/HER2-like (HR+/HER2+/Ki67 any)            | 102 | 27.06 |
| HER2-like (HR-/HER2+/Ki67 any)              | 20  | 5.31  |
| TNC (HR-/HER2-/Ki67 any)                    | 57  | 15.11 |

ER, estrogen receptor; PR, progesterone receptor; IHC, immunohistochemistry; HR, hormone receptor; HER2, human epidermal growth factor receptor 2; TNC, triple negative (breast) cancer. \*Tumor staging was based on clinical staging. \*\* “Other” histological subtypes were 1 apocrine, 3 medullary, 1 metaplastic, 1 neuroendocrine, 5 papillary, 3 squamous, and 4 tubular invasive carcinomas. \*\*\* IHC clinical with grading – the clinical immunotypes were subanalyzed with low (1 or 2) or high (3) grade.

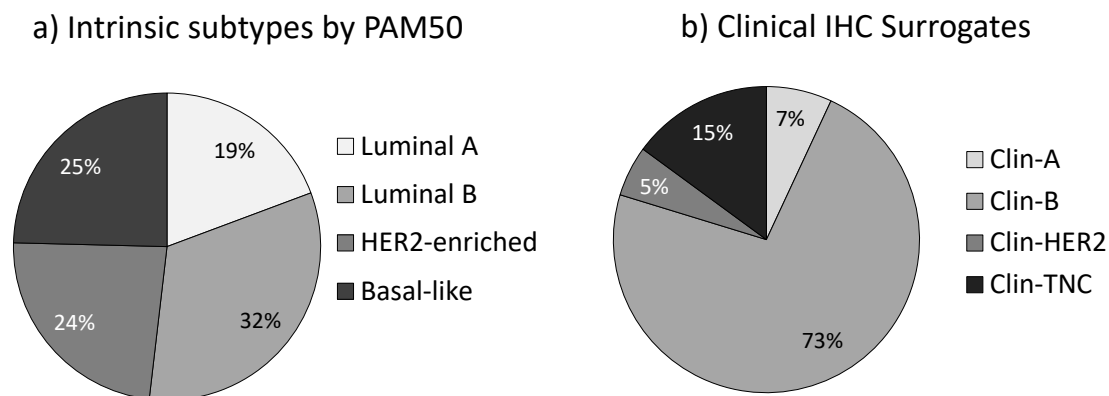


**Figure 1. Cohort subtypes by (a) Intrinsic subtyping by the PAM50 assay and (b) Clinical IHC subtypes**

(a) The percentage distribution of intrinsic subtypes was luminal-B (32.5%, n=123), basal-like (24.6%, n=93), HER2-enriched (23.6%, n=89), and luminal-A (19.3%, n=73).

However, using immunohistochemistry (b) as a surrogate for intrinsic subtypes, only 6.9% (n=26) of the samples tested were categorized as Clin-A (ER/PR pos, HER2 neg and Ki67 $\geq$ 14%). Most of the samples (72.7%, n=274) were Clin-B, i.e. ER/PR pos, HER2 neg, Ki67 $>$ 14% or ER/PR pos, HER2 pos, any Ki67. A small minority (5.3%, n=20) were Clin-HER2 (ER/PR neg, HER2 pos). The TNC (ER/PR neg, HER2 neg, Ki67 any) accounted for the remaining 15.1% (n=57).

### 3.4.2. Comparison of immunohistochemistry and intrinsic subtypes

The luminal-B intrinsic subtype and the IHC B-like (figure 2a) were highly concordant. The intrinsic HER2-enriched showed the best concordance with the IHC B/HER2-like and the HR-/HER2-like (62.9% and 19.1%, respectively), while the intrinsic basal-like was most concordant with the IHC TNC (53.8%). Immunohistochemistry currently classifies the B/HER2-like as B-like tumors because they are HR positive but it may be more appropriate to classify these B/HER2-like tumors as HER2 positive tumors and to treat them accordingly. The intrinsic luminal-A subtype was not strongly associated with any one IHC subtype, raising questions about appropriate Ki67 cutoff values.

By comparison, the IHC-like groups were well reflected by the intrinsic subtypes (figure 2b). The A-like group was mainly composed of luminal-A intrinsic subtypes; A- or B- like was primarily distributed between luminal-A (38.2%) and luminal-B (56.4%) intrinsic subtypes, and the IHC B-like was mainly comprised of luminal-B. The HR positive/HER2+ (B/HER2-like) group consisted mainly of the intrinsic HER2-enriched subtype, followed by the luminal-B subtype. The HR negative / HER2 positive (HER2-like) group was predominantly HER2-enriched, and the TNC group mainly basal-like, as expected.

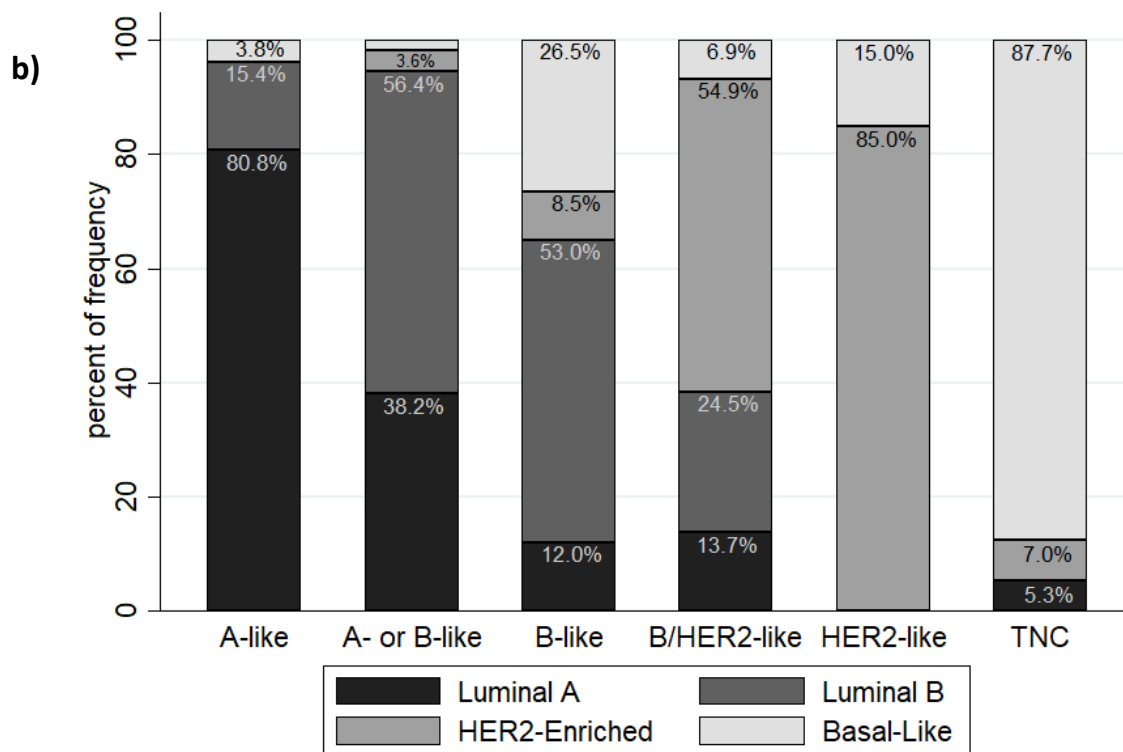
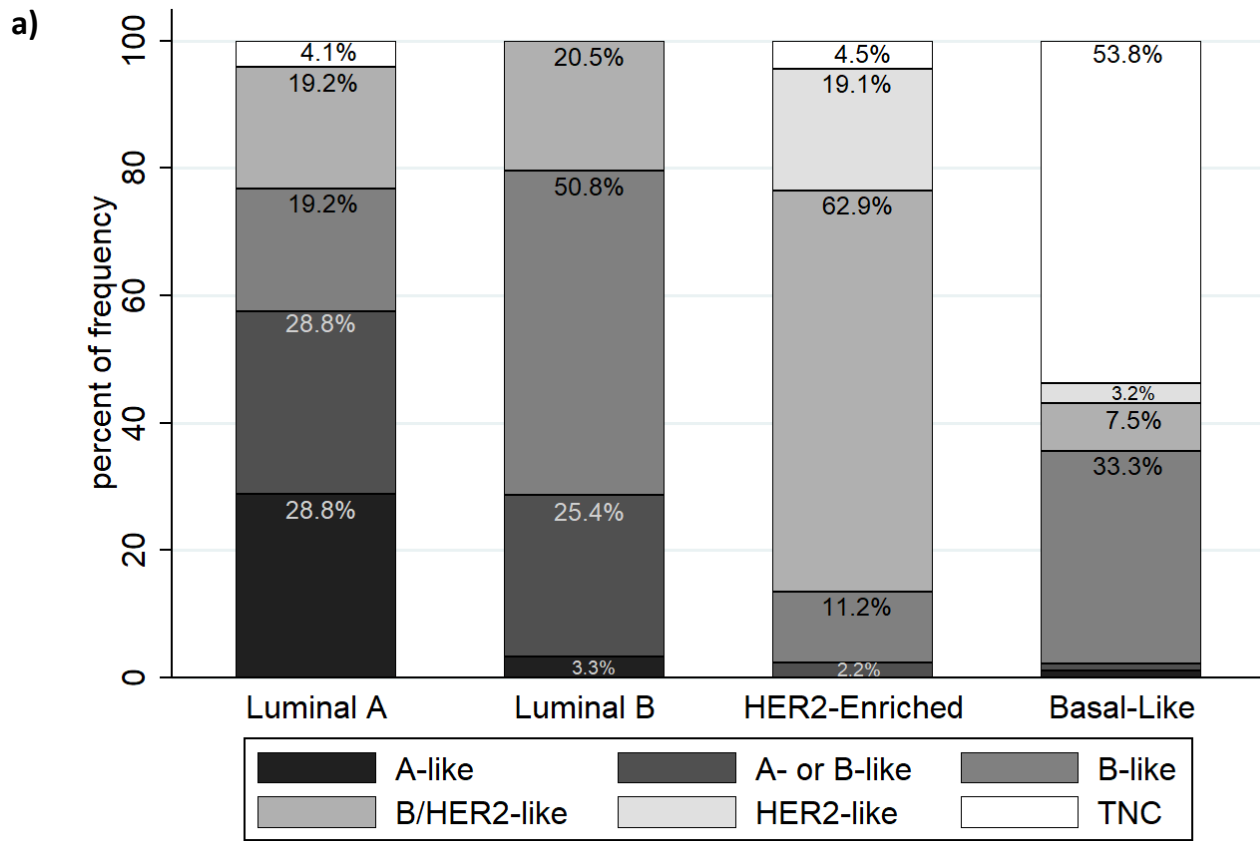


**Figure 2. Comparison between immunohistochemical subtypes and molecular subtypes.**

**a) Frequency of immunohistochemical analysis subtypes (y-axis) within the PAM50 molecular defined tumor subtypes (x-axis).**

The luminal-A intrinsic subtype had the greatest spread of IHC types, with 28.8% IHC A-like, 28.8% IHC A- or B-like, 19.2% IHC B-like, 19.2% IHC B/HER2, 0% HER2-like (HR-/HER2+) and 4.1% TNC. The luminal-B intrinsic subtype was mainly B-like (50.8%), with 25.4% A- or B- like, and 20.5% B/HER2-like. The HER2-enriched intrinsic subtype was mainly B/HER2-like (62.9%), followed by HER2-like (HR-/HER2+, 19.1%), and the basal-like was mainly TNC (53.8%) or B-like (33.3%).

**b) Intrinsic subtype frequency (y-axis) concordance with each IHC-Analysis subgroup classification (x-axis)**

Most of the A-like (HR+/HER2-/Ki67 $\leq$ 10%) group had the luminal-A subtype (80.8%). The A- or B-like group (HR+/HER2-/10%<Ki67 $\leq$ 30%) had slightly more luminal-B (56.4%) than luminal-A (38.2%) subtypes. The B-like group (HR+/HER2-/Ki67<30%) had mainly luminal-B subtypes (53%). B/HER2-like (HR+/HER2+) tumors were mainly HER2-enriched (54.9%), as was expected, while the HER2 group (HR-/HER2+) was predominantly HER2-enriched (85%). The TNC group had overwhelming basal-like subtypes (87.7%).

**3.4.3. Characteristics by intrinsic subtype**

Expression of the proliferation marker, Ki67, was lowest in luminal-A tumors [20% (10-32.5%)], highest in the basal-like subtype [70% (50-80%)], and intermediate in the luminal-B [40% (30-55%)] and HER2-enriched [50% (40-62.5%)] subtypes, as expected (Table 3). Categorical analysis of Ki67 expression showed that the luminal-A tumors had the greatest spread, while close to 80% of the luminal-B tumors had Ki67 levels >30%. The HER2-enriched and basal-like tumors expressed Ki67 at high values (over 30%), as expected. The Allred scores in luminal-A and luminal-B subtypes were predominantly high (scores of 7,8), while HER2-enriched subtypes had a greater spread of HR expression scores and basal-like subtypes were mainly negative or low scoring (Table 3).

Luminal-A (69.9%) and luminal-B (61.8%) subtypes were more likely to have lower T stages (T1 or T2), compared to HER2-enriched (36.0%) and basal-like (47.3%) subtypes. All intrinsic subtypes had T4 tumors, indicative of the late stage at presentation in this setting (Table 3). Tumors with a luminal subtype were more likely to be of lower grade (grade 1 or 2) than basal-like subtypes (75.0% grade 3). Histologically, only the luminal-A subtypes had a significant proportion of

invasive lobular carcinomas (11.1%) and invasive mucinous carcinomas (6.9%). Age and nodal involvement were not associated with intrinsic subtype in this cohort (Table 3).

**Table 3. Tumor features by intrinsic subtype**

|                                     | Luminal-A    | Luminal-B    | HER2-Enriched | Basal-Like   | P-value <sup>#</sup> |
|-------------------------------------|--------------|--------------|---------------|--------------|----------------------|
| <b>Ki67 (n=375)</b>                 |              |              |               |              |                      |
| % Ki67 median (IQR)                 | 20 (10-32.5) | 40 (30-55)   | 50 (40-62.5)  | 70 (50-80)   | <0.001               |
| Ki67 Categories                     |              |              |               |              |                      |
| <10; n=18                           | 14 (19.44%)  | 1 (0.82%)    | 3 (3.41%)     | 0 (0%)       |                      |
| 10-15; n=15                         | 10 (13.89%)  | 3 (2.46%)    | 1 (1.14%)     | 1 (1.08%)    |                      |
| 15-20; n=14                         | 8 (11.11%)   | 5 (4.1%)     | 1 (1.14%)     | 0 (0%)       |                      |
| 20-30; n=38                         | 15 (20.83%)  | 16 (13.11%)  | 6 (6.82%)     | 1 (1.08%)    |                      |
| >30; n=290                          | 25 (34.72%)  | 97 (79.51%)  | 77 (87.5%)    | 91 (97.85%)  |                      |
| <b>Allred (n=369)</b>               |              |              |               |              | <0.001               |
| Negative (0,1,2)                    | 2 (2.78%)    | 0            | 16 (18.60%)   | 48 (54.55%)  |                      |
| Low (3,4)                           | 2 (2.78%)    | 0            | 9 (10.47%)    | 19 (21.59%)  |                      |
| Intermediate (5,6)                  | 3 (4.17%)    | 3 (2.44%)    | 14 (6.88%)    | 5 (5.68%)    |                      |
| High (7,8)                          | 65 (90.28%)  | 120 (97.6%)  | 47 (54.65%)   | 16 (18.18%)  |                      |
| <b>T Staging (n=378)</b>            |              |              |               |              | <0.001               |
| T1, n=35                            | 13 (17.81%)  | 8 (6.50%)    | 7 (7.87%)     | 7 (7.53%)    |                      |
| T2, n=168                           | 38 (52.05%)  | 68 (55.28%)  | 25 (28.09%)   | 37 (39.78%)  |                      |
| T3, n=77                            | 8 (10.96%)   | 19 (15.45%)  | 24 (26.97%)   | 26 (27.96%)  |                      |
| T4, n=98                            | 14 (19.18%)  | 28 (22.76%)  | 33 (37.08%)   | 23 (24.73%)  |                      |
| <b>Age at diagnosis</b>             | 48 (42 – 59) | 49 (42 – 60) | 46 (41 – 52)  | 48 (42 – 56) | 0.297                |
| <b>Age below and above 50 years</b> |              |              |               |              | 0.679                |
| <50 years                           | 39 (18.48%)  | 65 (30.81%)  | 54 (25.59%)   | 53 (25.12%)  |                      |
| ≥50 years                           | 34 (20.36%)  | 58 (34.73%)  | 35 (20.96%)   | 40 (23.95%)  |                      |
| <b>Stage at diagnosis</b>           |              |              |               |              | 0.004                |
| Early Stage                         | 43 (58.9%)   | 55 (44.7%)   | 27 (30.3%)    | 40 (43.0%)   |                      |
| Late Stage                          | 30 (41.1%)   | 68 (55.3%)   | 62 (69.7%)    | 53 (57.0%)   |                      |
| <b>Grade (n=374)</b>                |              |              |               |              | <0.001               |
| Grade 1, n=24                       | 17 (23.6%)   | 4 (3.3%)     | 3 (3.4%)      | 0            |                      |
| Grade 2, n=207                      | 54 (75.0%)   | 81 (66.4%)   | 45 (55.7%)    | 23 (25.0%)   |                      |
| Grade 3, n=143                      | 1 (1.4%)     | 37 (30.3%)   | 36 (40.9%)    | 69 (75.0%)   |                      |

|                                       | Luminal-A  | Luminal-B   | HER2-Enriched | Basal-Like | P-value <sup>#</sup> |
|---------------------------------------|------------|-------------|---------------|------------|----------------------|
| <b>Nodal involvement</b>              | 56 (76.7%) | 98 (79.7%)  | 74 (83.2%)    | 68 (73.1%) | 0.400                |
| <b>Histological Diagnosis (n=372)</b> |            |             |               |            | <0.001               |
| Invasive Ductal                       | 55 (76.4%) | 111 (91.7%) | 79 (90.8%)    | 84 (91.3%) |                      |
| Invasive Lobular                      | 8 (11.1%)  | 5 (4.1%)    | 2 (2.3%)      | 0          |                      |
| Invasive Mucinous                     | 5 (6.9%)   | 4 (3.3%)    | 1 (1.2%)      | 0          |                      |
| Other                                 | 4 (5.6%)   | 1 (0.8%)    | 5 (5.8%)      | 8 (8.7%)   |                      |

<sup>#</sup> Median (IQR) values compared using the Kruskal-Wallis equality-of-populations rank test; frequencies compared using Pearson's  $\chi^2$  test or Fisher's exact test if any groups had frequencies of 5 or less.

### 3.4.4. Comparisons of Ki67 cutoff levels

The kappa test was used to compare the classification of the luminal subtype using IHC and PAM50 based on Ki67 levels (Supplementary Table S2). The IHC groups were split into luminal-A and luminal-B subtypes using Ki67 cutoffs of 10%, 15%, 20%, 25% and 30% and the kappa statistic was used to compare these classifications to the subtypes assigned by the PAM50 analysis. The agreement between the methods ranged from 43% to 49%. The best concordance of the IHC and intrinsic subtypes, was when the cutoff was at 25% Ki67 ( $\kappa=0.128$ ,  $p=0.003$ ) and the worst at a cutoff of 10% ( $\kappa=0.079$ ,  $p=0.033$ ) (Supplementary Table S2). Thus, a Ki67 cutoff of 25% appears best for separating the luminal-A and -B subtypes in our setting. Using the 25% cutoff results in 15.5% IHC A-like and 37.3% B-like (figure 3b), closer in value to the intrinsic subtype proportions of luminal-A (19%) and luminal-B (32%) (figure 1a) than the current clinical cutoff of 14% (figure 1b). Moreover, when IHC HR+/HER2+ samples are separated from the Clin-B (figure 1b) into the B/HER2-like group (figure 3), the B-like group becomes smaller, but the B/HER2-like group (26.9%) and HER2-like group (5.3%) together, are more reflective of the HER2-enriched intrinsic subtype (figure 1b).

**Table 4. Comparisons for PAM50 and IHC luminal subtypes based on Ki67 cutoff levels**

| Ki67 cutoff | Agreement | Expected agreement | Kappa statistic | p-value      |
|-------------|-----------|--------------------|-----------------|--------------|
| 10%         | 47.50%    | 43.01%             | 0.079           | 0.033        |
| 15%         | 49.17%    | 42.33%             | 0.119           | 0.004        |
| 20%         | 47.92%    | 40.97%             | 0.118           | 0.006        |
| 25%         | 47.92%    | 40.28%             | <b>0.128</b>    | <b>0.003</b> |
| 30%         | 43.33%    | 38.33%             | 0.081           | 0.044        |

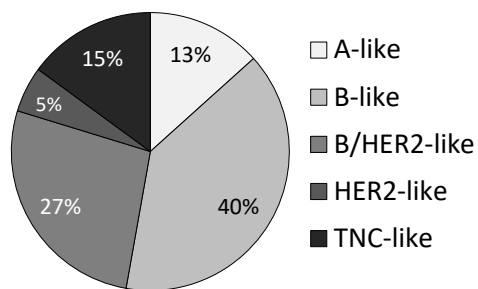
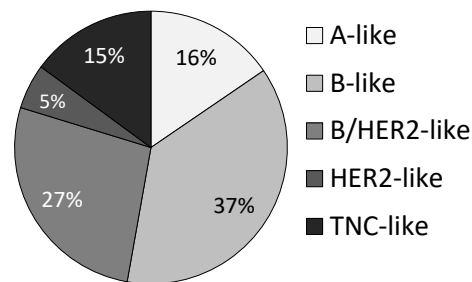
**a) IHC Surrogates with Ki67 at 20%****b) IHC Surrogates with Ki67 at 25%**

**Figure 3. IHC Analysis subtypes with different Ki67 cutoff values to distinguish between IHC A-like and IHC B-like groups. IHC A-like with Ki67  $\geq$  20% and IHC B-like with Ki67  $<$  20% is shown in (a); while IHC A-like with Ki67  $\geq$  25% and IHC B-like with Ki67  $<$  25% is shown in (b).**

When the cutoff for Ki67 between IHC A-like and B-like is at 20% (a), 13.3% (n=50) of the ER/PR pos, HER2 neg cohort is A-like and 39.5% (n=148) is B-like. When the cutoff is at 25% (b), 15.5% (n=58) of the ER/PR pos, HER2 neg cohort is A-like and 37.3% (n=140) is B-like.

When the ER/PR pos, HER2 pos (B/HER2-like) was classified separately from B-like group, it comprised 26.9% (n=101) of the cohort. The HER2 (ER/PR neg, HER2 pos) group made up 5.3% (n=20) of the cohort, and 14.9% (n=56) of the cohort was triple negative (ER/PR neg, HER2 neg).

### **3.5. DISCUSSION AND CONCLUSIONS**

The ability to diagnose breast cancer subtypes accurately and appropriately, fundamentally affects cancer treatment decisions. PAM50 is widely used for molecular diagnosis of breast cancer subtypes in high income countries (HICs) (156) because its results are reproducible and unaffected by inter- and intra-laboratory variability (157). Within resource-constrained settings, IHC is used as a proxy for intrinsic subtypes because it is less expensive, the infrastructure to run

IHC assays is widespread, and it requires less “hands-on” technical expertise than the PAM50 assay. We thus need accurate and population-specific information to assign proxies that optimize concordance calibration with the PAM50 intrinsic subtyping findings.

We found that the luminal-A intrinsic subtype had the greatest spread of IHC-analysis subgroups; the A-like IHC group was mainly composed of luminal-A subtype. This observation suggests that the currently used 14-20% Ki67 cutoff in South Africa may be too low. If the Ki67 cutoff were increased to 20-25%, the IHC A-like and B-like distribution would more accurately reflect the intrinsic subtypes. Subtyping strongly affects treatment options. Patients with luminal-A subtypes are likely to benefit from primary endocrine therapies in place of chemotherapy as first choice systemic treatment, whereas the benefits of chemotherapy to patients with luminal-B subtypes may offset chemotherapy side effects (158). The ambiguity in the Ki67 cutoff is not unique to the South African public health care system. German guidelines state that primary invasive tumors that are HR+, HER2- are considered low risk if  $Ki67 \leq 10\%$ , high risk if  $\geq 25\%$ , and intermediate risk if 10-25% as Ki67 does not differentiate risk groups accurately in this range (33). By contrast, the 14% cutoff was the best to distinguish between luminal-A and luminal-B in Spanish and Italian patients using Prosigna™ assays (159). These results reinforced the original PCR findings of Cheang et al (32) that the 14% cutoff was optimal. However, like Noske et al (33), we observed better concordance at higher Ki67 levels.

In HICs, where most breast cancers are diagnosed in early stages, the ASCO recommendations (35) suggested that PAM50 could be used to inform chemotherapy decisions, much better than IHC in node negative luminal subtypes. Pu et al. (160) found that survival rates were consistently worse in the luminal-B subtype, irrespective of menopausal status. The 2019 St Gallen report recommended that patients with ER  $\geq 1\%$  receive endocrine therapy, although it might have limited benefits (158). This recommendation is in line with the South African policy, which regards ER or PR  $\geq 1\%$  as hormone receptor positive (161). The Allred score shows ER and/or PR expression is high in luminal-A and luminal-B subtypes, as expected. The PAM50 basal-like subtype was predominantly negative for the Allred score, but also had a portion of low Allred scores (3,4). This

second finding is interesting, as it may suggest that the Allred cutoff to distinguish between A-like, B-like IHC subtypes and TNC subtypes could be increased to an Allred score  $\leq 4$ , as seen in table 3. A larger study is needed to confirm this.

Most tumors of the HER2-enriched luminal subtype are assigned to the B/HER-like IHC-analysis group. This finding is obvious when looking at the Allred score, where most of the HER2-enriched luminal subtypes had high HR positivity. While the multidisciplinary teams follow the St Gallen's recommendations and treat HR+/HER2+ as Clin-B, the PAM50 intrinsic subtypes do not make this subtle distinction. In South African public health care, patients in this group received adjuvant endocrine therapy until 2019, when anti-HER2 therapies were introduced. A mere 19% of the HER2 enriched subtype would be HR negative and would not benefit from endocrine therapy.

Patient subtyping should be interpreted cautiously. Mistaking luminal-A patients for luminal-B may result in overtreatment with chemotherapy. Confusion of HER2 with luminal-B subtypes may result in undertreatment by HER2 targeted therapy (e.g., trastuzumab) and/or overtreatment endocrine therapy (162). Trastuzumab is expensive and inconsistently available in the South African public sector (163), so the option of using endocrine therapies if trastuzumab is unavailable would be an advantage for HER2 positive patients.

A Swedish cohort, (164) found 81-85% concordance between molecular luminal-A and IHC-A subtypes. However, 35%-52% of their luminal-B intrinsic subtypes were classified as IHC-A. Ki67 distinguished between good and bad prognostic groups with node negative cancer, but its use is controversial, for example cytology and histology Ki67 values from the same tumor were highly variable (164). Lundgren et al (165) found that concordance with luminal subtypes improved when histological grade was included. Well differentiated tumors (grade 1) tended to have low Ki67 levels (33). Intermediate (grade 2) and poorly differentiated tumors (grade 3) had higher Ki67 levels and a wider range of Ki67 values (33). In our study histological grades were generally high, so including grade with clinical IHC subtype had a negligible effect on concordance.

Previously, women of African ancestry were thought to have fewer hormone receptor positive breast cancers than women of European ancestry (166). West African women and African American women appear more likely to have TNC cancers (167–172). However, research has shown that most sub-Saharan Africans (South African, Kenyan, Sudanese) (6,31,111,173–175) have HR positive cancers. In our cohort, 79.5% were HR positive, and more likely to be B-like (i.e., HR positive, high Ki67), even when the cutoff of Ki67 is 30%. Such cancers are more aggressive and have a poorer prognosis than those classified as luminal-A or IHC A-like.

Our results, with exclusively black participants, did not show differences in between the distribution of subtypes with age, which is consistent with the results found by Achilonu et al (44). Because our study was part of a HIV outcome study, HIV positive and HIV negative cases were age matched within a 5 year band. HIV prevalence amongst black South women is highest among women between 30-50 year (176), while breast cancer is associated with older age, and the highest prevalence in South African women is between 40-60 years(44). Our study participants were therefore younger (49.9 years  $\pm$ 11 years) than South African women with breast cancer on average, which was 56.2  $\pm$ 14.3 years among the SABCHO cohort (177). Younger patients are thought to have more clinically aggressive disease and poorer outcomes. Korean breast cancer patients are much more likely to be premenopausal than other nationalities (178), and this younger population shows poorer outcomes. Sub-Saharan Africa shows huge disparities in IHC subtyping (179). In Uganda, breast cancer patients had mean age of 45, with IHC of 38% A or B; 5% B/HER2; 22% HER2 and 34% TNC (180). Two separate Nigerian groups found very different IHC expression: a study in Ibadan, found 77.6% A or B; 2.6% B/HER2; 4% HER2 and 15.8% TNC (181); while a different study in Lagos found 38% HR pos; 18.3% HER2 pos and 47.4% TNC (182). Patients in Mozambique (183), had IHC of 51% A or B; 24% HER2 pos and 25% TNC; and Angola reported 25.7% A-like; 19.3% B-like; 7.9% B/HER2; 15.7% HER2-like and 31.4% TNC (184); while in Zimbabwe, the IHC was 68% HR positive and 17% TNC (185). Work on 985 participants in South Africa showed 13.8% A-like; 43.9% B-like, 19.0% B/HER2; 6.0% HER2-like and 15.3% TNC, although this work included individuals of different ethnicities (43). Recent work in South Africa (44) found that black South Africans had expression of about 49-53% HR+/HER2- (A- or B-like),

13-18% HR+/HER2+ (B/HER2-like), 7-12% HR-/HER2+ and 23-27% TNC, regardless of age. By comparison, South African whites had 30-65% HR+/HER2- (A- or B-like), 9-29% HR+/HER2+, 4-13% HR-/HER2+ and 14-29% TNC. White women under 40 had higher expression of the more aggressive TNC and HER2 tumors, while women over 60 had more A-like and B-like tumors.

Limitations of this study include the small sample size and lower age of participants. This may have artificially increased the proportion of HER2 positive tumors. However, these limitations may have impacted the main focus of this study; which was to determine the concordance between PAM50 intrinsic subtyping and IHC surrogates.

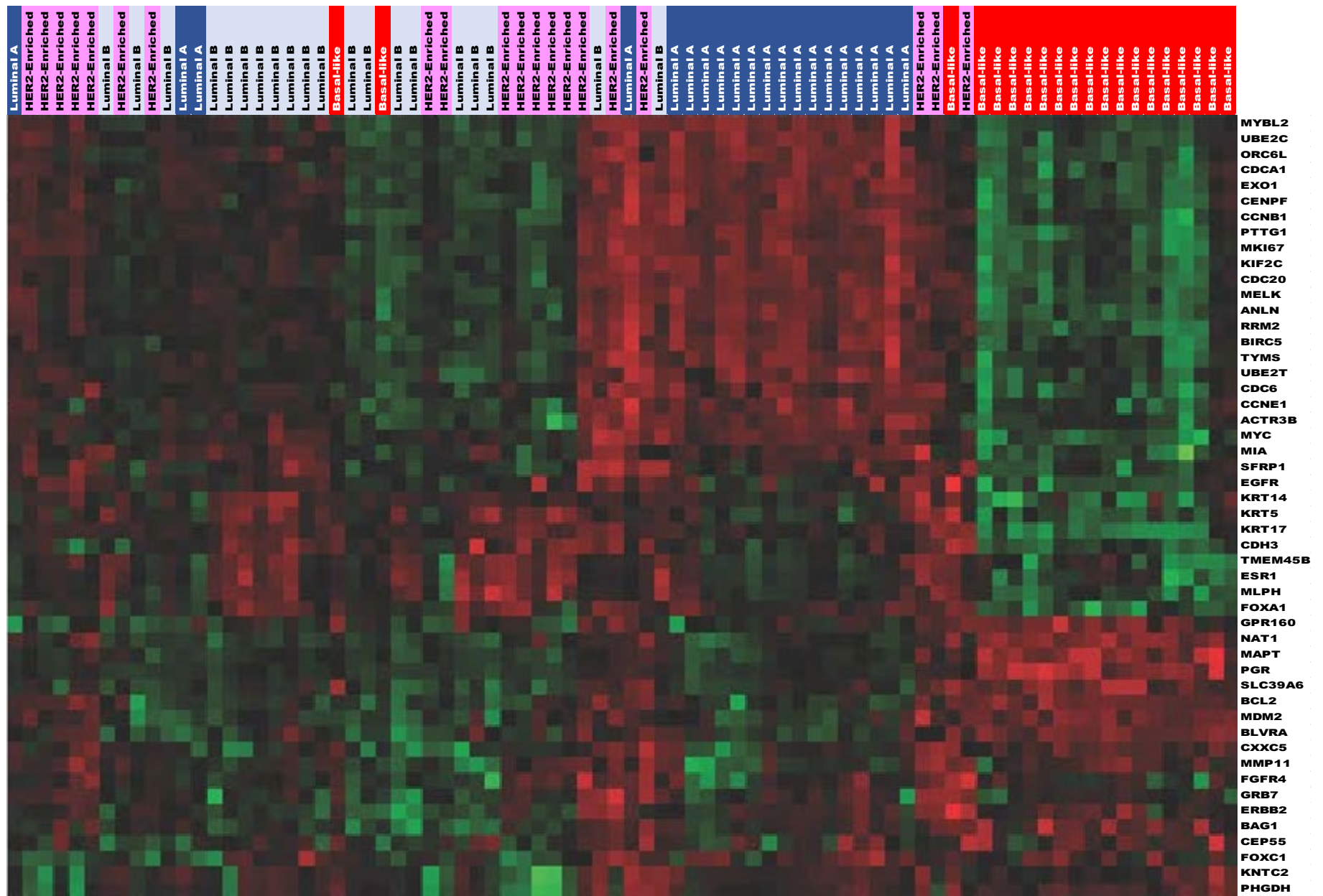
Our study is, as far as we know, the first to compare IHC with PAM50 in black southern African women. Most of our study participants had hormone receptor positive breast cancer, and even tumors with the HER2-enriched subtype were more likely to be HR positive than HR negative. PAM50 is widely used for breast cancer subtyping, with IHC often used in resource constrained settings. The cost and labor of the PAM50 method make it prohibitive for the South African public health care sector and its inability to distinguish between HER2-positive B-subtypes and HR negative/HER2 positive subtypes must also give pause. We found the lowest concordance between molecular and IHC subtyping for the luminal-A group and recommend raising the cutoff for Ki67 to 20-25% to distinguish between A-like and B-like tumors, to better reflect the luminal subtypes when using IHC subtyping.

### **3.6. ADDITIONAL INFORMATION**

#### **Additional Files**

**Figure S1:** Representative heat map of intrinsic subtypes

**Table S1:** PAM50 gene list



**Supplementary Figure S1. Representative heatmap of intrinsic subtypes**

Twenty randomly chosen luminal-A (dark blue), luminal-B (light blue), HER2-enriched (pink) and basal-like (red) tumors are represented. The luminal-A, luminal-B, basal-like cluster with each other. The HER2-enriched subtypes have 2 clusters, one where they cluster with the luminal-subtypes (indicative of HR-positive/HER2 positive subtypes), and a second cluster indicative of HR-negative, HER2-positive subtypes. Upregulation of genes is represented by green and downregulation represented by red.

**Supplementary Table S1. PAM50 gene list (186,187)**

| <b>Type</b> | <b>Gene Card</b> | <b>Accession number</b> | <b>Gene Name</b>                                                                                               |
|-------------|------------------|-------------------------|----------------------------------------------------------------------------------------------------------------|
| Endogenous  | BAG1             | NM_004323.3             | BCL2-associated athanogene                                                                                     |
| Endogenous  | BCL2             | NM_000633.2             | B-cell CLL/lymphoma 2, nuclear gene encoding mitochondrial protein                                             |
| Endogenous  | BLVRA            | NM_000712.3             | Biliverdin reductase A                                                                                         |
| Endogenous  | CXXC5            | NM_016463.5             | CXXC finger 5                                                                                                  |
| Endogenous  | ESR1             | NM_000125.2             | Estrogen receptor 1                                                                                            |
| Endogenous  | FOXA1            | NM_004496.2             | Forkhead box A1                                                                                                |
| Endogenous  | GPR160           | NM_014373.1             | G protein-coupled receptor 160                                                                                 |
| Endogenous  | MAPT             | NM_016835.3             | Microtubule-associated protein tau                                                                             |
| Endogenous  | MDM2             | NM_006878.2             | Mdm2, transformed 3T3 cell double minute 2, p53 binding protein (mouse)                                        |
| Endogenous  | MLPH             | NM_024101.4             | Melanophilin                                                                                                   |
| Endogenous  | MMP11            | NM_005940.3             | Matrix metalloproteinase 11 (stromelysin 3)                                                                    |
| Endogenous  | MYC              | NM_002467.3             | v-myc myelocytomatosis viral oncogene homolog (avian)                                                          |
| Endogenous  | NAT1             | NM_000662.4             | N-acetyltransferase 1 (arylamine N-acetyltransferase)                                                          |
| Endogenous  | PGR              | NM_000926.2             | Progesterone receptor                                                                                          |
| Endogenous  | SLC39A6          | NM_012319.2             | Solute carrier family 39 (zinc transporter), member 6                                                          |
| Endogenous  | ERBB2            | NM_004448.2             | V-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian) |
| Endogenous  | FGFR4            | NM_002011.3             | Fibroblast growth factor receptor 4                                                                            |
| Endogenous  | GRB7             | NM_005310.2             | Growth factor receptor-bound protein 7                                                                         |
| Endogenous  | TMEM45B          | NM_138788.3             | Transmembrane protein 45B                                                                                      |
| Endogenous  | ACTR3B           | NM_001040135.1          | ARP3 actin-related protein 3 homolog B (yeast)                                                                 |
| Endogenous  | CDH3             | NM_001793.3             | Cadherin 3, type 1, P-cadherin (placental)                                                                     |
| Endogenous  | EGFR             | NM_005228.3             | Epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)             |
| Endogenous  | FOXC1            | NM_001453.1             | Forkhead box C1                                                                                                |
| Endogenous  | KRT14            | NM_000526.3             | Keratin 14 (epidermolysis bullosa simplex, Dowling-Meara, Koebner)                                             |
| Endogenous  | KRT17            | NM_000422.1             | Keratin 17                                                                                                     |
| Endogenous  | KRT5             | NM_000424.2             | Keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Koebner/Weber-Cockayne types)                          |
| Endogenous  | MIA              | NM_006533.1             | Melanoma inhibitory activity                                                                                   |

| Type       | Gene Card | Accession number | Gene Name                                                   |
|------------|-----------|------------------|-------------------------------------------------------------|
| Endogenous | PHGDH     | NM_006623.2      | Phosphoglycerate dehydrogenase                              |
| Endogenous | SFRP1     | NM_003012.3      | Secreted frizzled-related protein 1                         |
| Endogenous | ANLN      | NM_018685.2      | Anillin, actin binding protein (scraps homolog, Drosophila) |
| Endogenous | BIRC5     | NM_001168.2      | Baculoviral IAP repeat-containing 5 (survivin)              |
| Endogenous | CCNB1     | NM_031966.2      | Cyclin B1                                                   |
| Endogenous | CCNE1     | NM_001238.1      | Cyclin E1                                                   |
| Endogenous | CDC20     | NM_001255.1      | CDC20 cell division cycle 20 homolog (S. cerevisiae)        |
| Endogenous | CDC6      | NM_001254.3      | CDC6 cell division cycle 6 homolog (S. cerevisiae)          |
| Endogenous | CDCA1     | NM_145697.1      | Cell division cycle associated 1                            |
| Endogenous | CENPF     | NM_016343.3      | Centromere protein F, 350/400ka (mitosin)                   |
| Endogenous | CEP55     | NM_018131.3      | Centrosomal protein 55kDa                                   |
| Endogenous | EXO1      | NM_006027.3      | Exonuclease 1                                               |
| Endogenous | KIF2C     | NM_006845.2      | Kinesin family member 2C                                    |
| Endogenous | KNTC2     | NM_006101.1      | Kinetochore associated 2                                    |
| Endogenous | MELK      | NM_014791.2      | Maternal embryonic leucine zipper kinase                    |
| Endogenous | MKI67     | NM_002417.2      | Antigen identified by monoclonal antibody Ki-67             |
| Endogenous | MYBL2     | NM_002466.2      | v-myb myeloblastosis viral oncogene homolog (avian)-like 2  |
| Endogenous | ORC6L     | NM_014321.2      | Origin recognition complex, subunit 6 like (yeast)          |
| Endogenous | PTTG1     | NM_004219.2      | Pituitary tumor-transforming 1                              |
| Endogenous | RRM2      | NM_001034.1      | Ribonucleotide reductase M2 polypeptide                     |
| Endogenous | TYMS      | NM_001071.1      | Thymidylate synthetase                                      |
| Endogenous | UBE2C     | NM_007019.2      | Ubiquitin-conjugating enzyme E2C                            |
| Endogenous | UBE2T     | NM_014176.1      | Ubiquitin-conjugating enzyme E2T (putative)                 |
| Reference  | ACTB      | NM_001101.2      | Actin, beta                                                 |
| Reference  | GUSB      | NM_000181.1      | Glucuronidase, beta                                         |
| Reference  | MRPL19    | NM_014763.3      | Mitochondrial ribosomal protein L19                         |
| Reference  | PSMC4     | NM_006503.2      | Proteasome (prosome, macropain) 26S subunit, ATPase, 4      |
| Reference  | PUM1      | NM_001020658.1   | Pumilio RNA-binding family member 1                         |
| Reference  | RPLP0     | NM_001002.3      | Ribosomal protein, large, P0                                |
| Reference  | SF3A1     | NM_005877.4      | Splicing factor 3a, subunit 1, 120kDa                       |
| Reference  | TFRC      | NM_003234.1      | Transferrin receptor (p90, CD71)                            |

#### **FORWARD TO CHAPTER 4**

Results in the previous chapter have shown that the immunohistochemical results are discordant with the intrinsic subtyping, and that these differences may be reduced by changing the proportion of Ki67 to 25% to better distinguish between IHC A-like and IHC-B-like.

Furthermore, the chapter showed that the subtypes in black South African women were the more aggressive phenotypes (luminal-B, HER2-enriched and basal-like), whereas published data shows that the less aggressive luminal subtypes are most common in white women.

We therefore wish to investigate possible reasons for the more aggressive subtypes in black South African women, and chose the *fibroblast growth factor receptor 2 (FGFR2)* as associative with breast cancer. The *FGFR2* variants rs2981582, rs35054928, rs2981578 and rs11200014 have been associated with breast cancer using genome wide association studies. We interrogated whether these polymorphisms confirmed the GWAS studies in sub-Saharan African women.

The paper was submitted to Molecular Medicine Reports. The authors affiliations and contributions are below.

## FGFR2 genetic variants in women with breast cancer

THÉRÈSE DIX-PEEK<sup>1</sup>, CAROLINE DICKENS<sup>1</sup>, TANYA N. AUGUSTINE<sup>2</sup>, BOITUMELO P. PHAKATHI<sup>3</sup>, EUNICE J. VAN DEN BERG<sup>4,5</sup>, MAUREEN JOFFE<sup>6,7,8</sup>, OLUWATOSIN A. AYENI<sup>6,7,8,9</sup>, HERBERT CUBASCH<sup>7,8,10,11</sup>, SARAH NIETZ<sup>7,11</sup>, CHRISTOPHER G. MATHEW<sup>12,13,14</sup>, MAHTAAB HAYAT<sup>12,14</sup>, ALFRED I. NEUGUT<sup>15,16</sup>, JUDITH S. JACOBSON<sup>15,16</sup>, PAUL RUFF<sup>1,7,8</sup> and RAQUEL A.B. DUARTE<sup>1</sup>

<sup>1</sup>Department of Internal Medicine, School of Clinical Medicine; <sup>2</sup>School of Anatomical Sciences, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193; <sup>3</sup>Department of Surgery, School of Clinical Medicine, Faculty of Health Sciences, University of Kwa-Zulu Natal, Durban 4001; <sup>4</sup>Department of Histopathology, National Health Laboratory Services, Chris Hani Baragwanath Hospital, Johannesburg 1864; Departments of <sup>5</sup>Anatomical Pathology; <sup>6</sup>Strengthening Oncology Services Research Unit; <sup>7</sup>SA MRC Common Epithelial Cancer Research Centre, Faculty of Health Sciences, University of the Witwatersrand; and <sup>8</sup>Non-Communicable Diseases Research Division, Wits Health Consortium (PTY) Ltd Johannesburg 2193; <sup>9</sup>Division of Radiation Oncology, Department of Radiation Sciences, Faculty of Health Sciences, University of the Witwatersrand; <sup>10</sup>Batho Pele Breast Unit, Chris Hani Baragwanath Academic Hospital, Soweto 1860; <sup>11</sup>Department of Surgery; <sup>12</sup>Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193, South Africa; <sup>13</sup>Department of Medical and Molecular Genetics, Faculty of Life Sciences and Medicine, King's College London, London, WC2R 2LS, United Kingdom <sup>14</sup>Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 2193, South Africa; <sup>15</sup>Herbert Irving Comprehensive Cancer Center, Vagelos College of Physicians and Surgeons; <sup>16</sup>Department of Epidemiology, Mailman School of Public Health, Columbia University, New York 10032, United States of America

### Authors' contributions

TDP, CD, CM, MH and RD designed the study. TDP, CD, BP, EvdB, MJ, OA, HC, SN and RD collected and checked patient data. TDP, CD, MJ, OA and RD collected and checked control participant data. TDP, CD, TA and RD performed the experiments. TDP, CD and RD collected and cross referenced the SNP data to patients and controls. TDP, TA, CD, CM and RD analyzed the data. AN, JJ, PR were responsible for clinical input and interpretation of data from a clinical perspective. TDP, CD and RD confirm the authenticity of all the raw data. TDP, CD, TA, CM, MH, and RD wrote the manuscript. All authors edited the manuscript.

## **CHAPTER 4: FIBROBLAST GROWTH FACTOR RECEPTOR 2 GENETIC VARIANTS IN SOUTHERN AFRICAN WOMEN WITH BREAST CANCER**

### **4.1. ABSTRACT**

Black African populations are more genetically diverse than others, but genetic variants have been studied mainly in European populations. We examined the association of these SNPs with breast cancer in Black, southern African women. Genomic DNA was extracted from whole blood of 1001 patients with breast cancer and 1006 controls (without breast cancer). The rs2981582, rs35054928, rs2981578 and rs11200014 polymorphisms were analyzed using allele-specific PCR KASP™ technologies, and the  $\chi^2$  test or Fisher's exact test were used to evaluate the genotype frequencies. We found no association of those SNPs with breast cancer in our population overall, but we did find an association of the C/C homozygote genotype for rs2981578 with invasive lobular carcinoma. These results show that genetic biomarkers of breast cancer risk in European populations cannot necessarily be extrapolated to determine risk in sub-Saharan African populations. As African populations are more heterogeneous than other populations, it is important that larger, representative populations are studied to determine the genetic risk of breast cancer in this understudied population.

### **4.2. INTRODUCTION**

African populations are the more genetically diverse than Asian or European populations (18,101), but few genomic studies have been conducted in African populations (19). Only 2.4 % of genome-wide association studies (GWAS) have included Africans or African Americans (18), and very little information is available about cancer genomics in African populations (19,20). A meta-analysis by Rotimi et al. (19) found that between Jan 1990 and Dec 2019, only 0.329% of cancer-related publications globally focused on African populations, and only 0.016% related to cancer genetics or genomics in Africa. Breast cancer is the most common cancer in women, including African women. However, it is much less common in African than in European populations. The 2018 Globocan report showed an estimated age standardized incidence rate (ASIR) of 37.9/100 000 (compared to 113.2/100 000 in Belgium) and a lifetime risk (LR) of 1 in 25

for women under 75 years (26). Similarly, the South African national cancer registry reported an ASIR of 20.4/100 000 and an LR of 1 in 47 among black south African women (8).

Since the 1990s, the breast cancer 1 (*BRCA1*) (55) and *BRCA2* (56) genes have been associated with hereditary breast cancer. Other rare, but highly penetrant genes include phosphatase and tensin homolog (*PTEN*), tumor protein P53 (*TP53*), epithelial cadherin (*CDH1*) and serine/threonine kinase (*STK11*) (57,58). Moderate penetrance genes include checkpoint kinase 2 (*CHEK2*), *BRCA1* interacting helicase 1 (*BRIP1*), ataxia-telangiectasia mutated (*ATM*) or partner and localizer of *BRCA2* (*PALB2*) (57,58). Some studies have been done in South Africa to examine some of these genes, particularly examining the effects of *BRCA1* and *BRCA2* in various ethnic populations. In the self-identified black population, (with a sample size of 165), Eygelaar et al. (59) found 1.2% *BRCA1*, 0.6% *BRCA2*; 0.6% *ATM*, 0.6% *CHEK2*, 0.6% *PALB2* deleterious variants associated with breast cancer. Moreover, in 78 black patients, Francies et al. (60) found 3.8% *BRCA1* and 3.8% *BRCA2* pathogenic mutations, but no deleterious mutations in *PALB2* or *CHEK2* with this group. Van der Merwe et al. (61,62) have identified larger rearrangements of the *BRCA1* and *BRCA2* genes that are specific to the black South African population. It is noteworthy that deleterious mutations in high and medium penetrance genes do not explain the vast majority of breast cancers in the black population. Over the past 15 years, genome-wide association studies have led to the detection of over 200 loci associated with breast cancer (66,67,69,188). Among the top hits for these loci are variants in the fibroblast growth factor receptor 2 (*FGFR2*) gene, a low penetrance gene. The FGFRs are receptor tyrosine kinases involved in signaling pathways that catalyze a variety of biological processes, including cell growth, survival, differentiation, angiogenesis, tumorigenesis (70), and epithelial to mesenchymal transition (71). Variants in intron 2 of *FGFR2* have been found to be highly associated with breast cancer (66,67,78,188–190). Although most single nucleotide polymorphisms (SNPs) have small effects individually, polygenic models indicate that an accumulation of small mutations may increase cancer risk (80). Included in the top hits of *FGFR2* SNPs associated with breast cancer are rs2981582, rs35054928 and rs2981578 in women of European ancestry, and rs11200014 in African American women (66,67,191).

The *FGFR2* SNP that is most commonly associated with breast cancer in women of European ancestry, rs2981582 (66), was recently also linked to increased risk of breast cancer in Saudi women (84), and with luminal-A breast cancer in Han Chinese women (85) and Korean women (86). The expression of the minor allele was associated with early onset breast cancer in Indonesian women (87). However, a study of women from Argentina and Uruguay did not find an association between rs2981582 and breast cancer, possibly because the populations of those countries include subpopulations of varied ethnicity (admixed populations) (88). Likewise, among postmenopausal Turkish women, rs2981582 was not associated with breast cancer (89). Admixed populations, such as that of Turkey(192), may have different allele frequencies than European populations and, given similar sample size, less power to detect associations.

The risk allele for rs35054928 appears to bind the transcription factor, E2F1 (90). Both rs35054928 and rs2981578 also are reported to be part of a response element, a sequence within the promoter of a gene that regulates transcription. The presence of the risk allele for rs2981578 for example, substantially increases the binding to FOXA1 in MCF7 epithelial, hormone receptor-positive cells, increasing chromatin accessibility and allowing access to transcriptional repressors such as YY1 (Yin Yang 1), SIN3A, and HDAC (histone deacetylase)(91). The DNA binding protein YY1, co-repressor SIN3A and histone-modifying HDAC form a complex that can inhibit promoter activity (92). The SNPs rs35054928 and rs2981578 are located next to an organic cation transporter (OCT)-binding site. The risk allele of rs2981578 also creates a potential binding site for runt-related transcription factor 2 (*RUNX2*) (71,94). Runx2 is mainly known for its role in osteoclast development, but it is also a regulator of mammary development and breast cancer (93).The rs2981578 variants have equal affinity for OCT1, but the high risk allele has a much higher affinity for RUNX2 (91,94), possibly because the SNP sites differ in histone acetylation (96). OCT1 promotes cell proliferation in ER (estrogen receptor) positive breast cancer cells (97).

The SNP rs11200014 has been associated with breast cancer in African American women (67,99). African Americans generally have admixed African and European ancestry, and a small

proportion of Native American ancestry (100,101), and their African ancestry is mainly from West or central West Africa (102). South African populations are genetically different from West and central Africans, and even more different from African Americans (193).

We tested the *FGFR2* SNPs for association with breast cancer and explored their association with hormone receptor subtypes of breast cancer in an urban South African Black female population.

### **4.3. MATERIALS AND METHODS**

#### **4.3.1. Study population**

Participants in this study were drawn from the South African Breast Cancer and HIV Outcome (SABCHO) study, a cohort of breast cancer patients diagnosed and treated at five hospitals in Gauteng and KwaZulu Natal, South Africa (111). For this study, we selected as cases women diagnosed with breast cancer at the Charlotte Maxeke Johannesburg Academic Hospital (CMJAH) Surgical Breast Unit or the Batho Pele Breast Unit of the Chris Hani Baragwanath Academic Hospital (CHBAH). The CMJAH Surgical Breast Unit is located in central Johannesburg and identifies about 250 new breast cancer cases each year. The Batho Pele Breast Unit serves patients from Soweto and surrounding areas and diagnoses about 350 patients with breast cancer yearly (111). We recruited as controls women seen in the CMJAH breast unit or the Batho Pele Breast unit who were found not to have breast cancer, and patients undergoing routine assessment at other clinics at CMJAH not related to cancer. Eligible cases were self-identified black, southern African women >18 years of age with histologically confirmed invasive breast cancer; exclusion criteria were ductal carcinoma in situ (DCIS) or lobular carcinoma in situ (LCIS). Eligible controls were self-identified black, southern African women >18 years of age with no history of breast or ovarian cancer; and neither pregnant nor breastfeeding. All participants had a sample of peripheral blood drawn and collected into EDTA vacutainer tubes (Becton, Dickson and Company, Sandton, South Africa). Ethics clearance for this study was granted from the

Human Research Ethics Committee (Medical) of the University of the Witwatersrand (Ethics number M161116).

#### **4.3.2. Genomic DNA extraction, SNP selection and analysis**

DNA was extracted from whole blood using a modified salting out method (121), and diluted to 25 ng/μl in low TE buffer (10mM Tris HCl; 0.1mM EDTA, pH 8.3). The Nanodrop 2000™ spectrophotometer (Thermo Scientific, Waltham, MA) was used to determine DNA concentrations and A260/280 ratios; ratios between 1.7 and 2.0 indicated adequate DNA purity for this genotype analysis.

We selected 4 SNPs for genotyping, specifically, rs2981582, rs35054928, rs2981578 and rs11200014, which are located in intron 2 of the *FGFR2* gene. Power for this study was assessed using the University of Michigan School of Public Health Genetic Association Study (GAS) Power Calculator ([https://csg.sph.umich.edu/abecasis/cats/gas\\_power\\_calculator/index.html](https://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html)). A sample size of 1000 cases, 1000 controls was chosen along with a significance level of 0.0125 ( $\alpha=0.05/4$ ). Assuming a dominant pattern of inheritance, odds ratios of 1.5 (for rs2981582) and 1.4 (for rs35054928, rs2981578 and rs11200014) could be detected with 80% power; and reflect odds ratios reported in the literature (78,123,124). Table 1 shows the minor allele frequencies (MAFs) for each SNP and compares them with African and global allele frequencies in other studies. Information about allele frequencies from the African and global population was obtained from the 1000 genomes project (82). Investigations of ancestry informative markers on a similar cohort sourced from this region showed limited evidence of population substructure (194).

We genotyped the *FGFR2* SNP polymorphisms using Kompetitive allele specific PCR (KASP™) technology at LGC Genomics LTD (Hoddesdon, UK). This trademarked method has 2 allele-specific forward primers and a common reverse primer. The forward primers each have a unique tail sequence that corresponds with a fluorescent resonant energy transfer cassette; one is labelled with FAM™ dye and the other with HEX™ dye. One allele binds the forward primer

with FAM™ and the second allele binds the HEX™ labelled primer (Table 1). For each SNP, we determined deviation of genotype frequencies from Hardy-Weinberg equilibrium (HWE) in controls using a  $\chi^2$  test (Table 2).

**Table 1: Primer sequences for the FGFR2 SNPs**

| SNP        | Primer ID       | Primer Sequence 5'-3'             |
|------------|-----------------|-----------------------------------|
| rs2981582  | 2981582_A       | GGCACCAGGTGGACTCTCCA -FAM         |
|            | 2981582_G       | GCACCAGGTGGACTCTCCG – HEX         |
|            | 2981582_Common  | TAAAACGGCAGATCCCAGCACTCAT         |
| rs35054928 | 35054928_CC     | CTCAGAAGGGCTGTGCGCC – FAM         |
|            | 35054928_C      | TCTCAGAAGGGCTGTGCGCG - HEX        |
|            | 35054928_Common | GCCCTGTCCCAGAAAGCCTACAT           |
| rs2981578  | 2981578_T       | TAACCTTTCTTCCCTGCTCCAACT – FAM    |
|            | 2981578_C       | CCTTTCTTCCCTGCTCCAAACC - HEX      |
|            | 2981578_Common  | GTTTTCTTGAAGCTTTTACCTCTATGCAAA    |
| rs11200014 | 11200014_A      | CTCCAAAAAAGATGCACAGAGGGAAGA - FAM |
|            | 11200014_G      | CAAAAAAAGATGCACAGAGGGAAGG - HEX   |
|            | 11200014_Common | ACACGTGTTGGGACCAGAGAGAAAA         |

Each SNP has 3 unique primers. There is a common reverse primer, and 2 forward primers that have the specific allele for the polymorphism; one primer is labelled with FAM™ and the other with HEX™.

**Table 2: Allele frequency distribution in South African, African and Global populations**

| SNP        | Position        | Allele | MAF Present study | MAF <sup>a</sup> Soweto | MAF <sup>b</sup> African | MAF <sup>b</sup> Global | HWEP |
|------------|-----------------|--------|-------------------|-------------------------|--------------------------|-------------------------|------|
| rs2981582  | Chr10:121592803 | G/A    | A=0.465           | A=0.452                 | A=0.495                  | A=0.404                 | 0.09 |
| rs35054928 | Chr10:12150918  | C-/CC  | C=0.426           | Not available           | C=0.331                  | CC=0.491                | 0.11 |
| rs2981578  | Chr10:121580797 | C/T    | T=0.061           | T=0.106                 | T=0.078                  | T=0.372                 | 0.39 |
| rs11200014 | Chr10:121575416 | G/A    | A=0.135           | A=0.112                 | A=0.190                  | A=0.313                 | 0.50 |

SNP, single nucleotide polymorphism; MAF, minor allele frequencies; HWEP, Hardy-Weinberg Equilibrium p-value. Allele frequencies from the <sup>a</sup>Soweto data (if available) looking at genetic diversity in the black population from Soweto, South Africa (81); <sup>b</sup>African and <sup>b</sup>Global frequencies available from the 1000 genomes project (82). The African group comprised African Caribbean in Barbados; African Ancestry in Southwest USA; Esan in Nigeria; Yoruba in Ibadan, Nigeria; Gambian in Western division, the Gambia; Mende in Sierra Leone; and Luhya in Webuye, Kenya.

#### 4.3.3. Classification of tumors

Pathologists at the National Health Laboratory Services (NHLS) provided data on histological diagnosis, tumor subtype, grade, and immunolocalization of estrogen receptor (ER),

progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and Ki67. Immunohistochemistry followed College of American Pathologists (CAP) guidelines.

We determined ER/PR status by Allred score: 0-2 was negative and 3-8 was positive (115). HER2 was regarded as positive when the test showed 3+; negative when it was 0 or 1+; and equivocal when it was 2+; positive results were confirmed by *in situ* hybridization (195). Specimens in which  $\leq 14\%$  of cells expressed Ki67 were categorized as having low expression, as per the St Gallen 2011 guidelines and current South African Guidelines (28,118). Based on IHC subtyping, the breast tumors were further categorized as: A-like [ER and/or PR positive, HER2 negative, Ki 67  $\leq 14\%$ ]; B-like [ER and/or PR positive and HER2 negative with Ki 67%  $> 14\%$ ]; B/HER2-like [ER and/or PR positive, HER2 positive with any Ki 67]; HER2 positive subtype [ER and PR negative, HER2 positive]; and triple negative breast cancer (TNBC) [ER, PR, and HER2 negative] (28).

#### 4.3.4. Statistics

Continuous variables were assessed for normality using the Shapiro-Wilks test. Normally distributed variables were presented as means and standard deviations, and non-normally distributed data as medians and interquartile ranges. Normally distributed continuous variables were compared using a t-test, while non-normally distributed variables were compared using a Mann-Whitney test. Non-parametric data were compared using a Kruskal Wallis test, followed by *post hoc* analysis using Dunn's Pairwise Comparison test if the Kruskal Wallis test showed significant differences. Categorical variables were expressed as frequencies and percentages and compared using a Pearson's  $\chi^2$  test. All statistical analyses were done using STATA v14.2 (College Station, Texas).

Using the genotyping data provided by LGC genomics, we calculated both genotype and allele frequencies for each SNP of interest. The frequencies of each genotype (homozygous major, heterozygous and homozygous minor) were compared between cases and controls for the SNPs of interest using a Pearson's chi square test. Genotype frequencies were also examined under four different models of genetic disease risk (125). Results were analyzed using dominant, recessive, multiplicative, additive, and homozygous models. A Bonferroni correction for multiple

comparison testing was applied. A value of  $p=0.008$  ( $0.05/6$ ) was considered statistically significant.

#### **4.5. RESULTS**

Overall, the 1001 women with breast cancer were older than the 1006 women without breast cancer; the medians [IQRs] were 53 [44-64] and 51 [40-62], respectively (Table 3). Most patients with cancer (57.4%) had been diagnosed in Stage III or IV. Most tumors (82.5%) were invasive ductal carcinomas (IDC) and 75.2% were estrogen receptor-positive (76.7%); 63.8% were progesterone receptor positive; 26.9% were HER2-positive, irrespective of hormone receptor status, and 14.4% were triple negative TNBC.

**Table 3: Characteristics of the study participants**

|                                       |             | Case, n(%)  | Control, n(%)          | p-value   |
|---------------------------------------|-------------|-------------|------------------------|-----------|
| Age                                   |             | n=1001      | n=1004                 |           |
| Overall age in years (median [IQR])   |             | 53 [44-64]  | 51 [40-62]             | <0.001    |
| 18-39 years                           |             | 142 (14.2%) | 241 (24.0%)            |           |
| 40-49 years                           |             | 259 (25.9%) | 232 (23.1%)            |           |
| 50-59 years                           |             | 239 (23.9%) | 223 (22.2%)            |           |
| 60+ years                             |             | 358 (35.8%) | 308 (30.7%)            |           |
| Missing                               |             | 3 (0.3%)    | 0                      |           |
| Cancer stage                          |             | n=1001      |                        |           |
| Stage I                               |             | 38 (3.8%)   |                        |           |
| Stage II                              |             | 378 (37.8%) |                        |           |
| Stage III                             |             | 439 (43.9%) |                        |           |
| Stage IV                              |             | 139 (13.9%) |                        |           |
| Stage unavailable                     |             | 7 (0.7%)    |                        |           |
| Tumor type                            |             | n=1001      |                        |           |
| Invasive ductal carcinoma             |             | 826 (82.5%) |                        |           |
| Invasive lobular carcinoma            |             | 14 (1.4%)   |                        |           |
| Other <sup>a</sup>                    |             | 71 (7.1%)   |                        |           |
| Tumor type unavailable                |             | 90 (9.0%)   |                        |           |
| IHC Phenotype                         |             | n=1001      |                        |           |
| A-like (ER and/or PR+/HER2-/Ki67≤14%) |             | 108 (10.8%) |                        |           |
| B-like (ER and/or PR+/HER2-/Ki67>14%) |             | 447 (44.7%) |                        |           |
| B/HER2-like (ER and/or PR+/HER2+)     |             | 211 (21.1%) |                        |           |
| HER2 (ER-/PR-/HER2+)                  |             | 58 (5.8%)   |                        |           |
| TNBC (ER-/PR-/HER2-)                  |             | 144 (14.4%) |                        |           |
| IHC unavailable                       |             | 10 (1.0%)   |                        |           |
| IHC equivocal <sup>b</sup>            |             | 23 (2.3%)   |                        |           |
| Receptor Status                       |             |             |                        |           |
|                                       | Positive    | Negative    | Equivocal <sup>b</sup> | Unknown   |
| ER status                             | 753 (75.2%) | 237 (23.7%) |                        | 11 (1.1%) |
| PR status                             | 639 (63.8%) | 347 (34.7%) |                        | 15 (1.5%) |
| HER2 status                           | 269 (26.9%) | 704 (70.3%) | 18 (18.0%)             | 10 (1.0%) |

<sup>a</sup>Other tumor types included apocrine, cribriform, medullary, mesenchymal, metaplastic, mucinous, neuroendocrine, papillary, pleomorphic, squamous, tubular, anaplastic, micropapillary carcinomas. ER, estrogen receptor; PR, progesterone receptor, HER2, human epidermal growth factor receptor 2; Ki67, proliferation marker.

<sup>b</sup>IHC Equivocal: HER2 had a value of 2+, but no FISH was done to confirm, or the Ki67 values were unavailable or ambiguous.

The present study showed all 4 SNPs were in HWE (Table 2). The MAFs of the SNPs in this study were more similar to those of the African sample group than those of the global population,

as expected. For example, the global MAF of rs2981578 (C/T) was T=0.372, while it was T=0.078 in the African population and T=0.061 in our study cohort (Table 2).

Cases and controls did not differ in allele or genotype frequencies of the four *FGFR2* SNPs (rs2981582, rs35054928, rs2981578, or rs11200014) (Table 4).

**Table 4: Genotypic frequencies of *FGFR2* variants in case and control populations**

|                        | Case, n (freq.) | Control, n (freq.) | p-value             | *OR (95% CI)        |
|------------------------|-----------------|--------------------|---------------------|---------------------|
| <b>rs2981582</b>       | <b>984</b>      | <b>996</b>         | 0.347               |                     |
| GG                     | 286 (0.291)     | 311 (0.312)        |                     | 1 (Ref)             |
| GA                     | 458 (0.465)     | 467 (0.469)        |                     | 1.07 (0.87-- 1.31)  |
| AA                     | 240 (0.244)     | 218 (0.219)        |                     | 1.22 (0.96-- 1.57)  |
| P trend                |                 |                    | 0.156               | 1.10 (0.98 -- 1.25) |
| GA+AA (vs GG)          | 698 (0.709)     | 685 (0.688)        | 0.295               | 1.11 (0.92-- 1.35)  |
| AA (vs GG+GA)          | 240 (0.244)     | 218 (0.219)        | 0.187               | 1.18 (0.95-- 1.45)  |
| AA (vs GG)             | 240 (0.456)     | 218 (0.412)        | 0.148               | 1.23 (0.96-- 1.57)  |
| G                      | 1030 (0.523)    | 1089 (0.547)       |                     | 1 (Ref)             |
| A                      | 938 (0.477)     | 903 (0.453)        | 0.141               | 1.11 (0.98-- 1.26)  |
| <b>rs35054928</b>      | <b>989</b>      | <b>996</b>         | 0.848 <sup>†</sup>  |                     |
| CC/CC                  | 340 (0.344)     | 335 (0.336)        |                     | 1 (Ref)             |
| CC/C-                  | 463 (0.468)     | 464 (0.466)        |                     | 0.96 (0.78-- 1.17)  |
| C-/C-                  | 186 (0.188)     | 197 (0.198)        |                     | 0.91 (0.70-- 1.17)  |
| P trend                |                 |                    | 0.604               | 0.95 (0.84 -- 1.08) |
| CC/C-+C-/C- (vs CC/CC) | 649 (0.656)     | 661 (0.664)        | 0.727               | 0.94 (0.78-- 1.13)  |
| C-/C- (vs CC/CC+CC/C-) | 186 (0.188)     | 197 (0.198)        | 0.583               | 0.93 (0.74-- 1.17)  |
| C-/C- (vs CC/CC)       | 186 (0.354)     | 197 (0.370)        | 0.572               | 0.90 (0.69-- 1.16)  |
| CC                     | 1143 (0.578)    | 1134 (0.569)       |                     |                     |
| C-                     | 835 (0.422)     | 858 (0.431)        | 0.585               | 0.95 (0.84-- 1.08)  |
| <b>rs2981578</b>       | <b>989</b>      | <b>995</b>         | 0.813 <sup>††</sup> |                     |
| CC                     | 867 (0.877)     | 879 (0.883)        |                     | 1 (Ref)             |
| CT                     | 119 (0.120)     | 114 (0.115)        |                     | 1.06 (0.80-- 1.40)  |
| TT                     | 3 (0.003)       | 2 (0.002)          |                     | 1.52 (0.25-- 9.15)  |
| P trend                |                 |                    | 0.637               | 1.07 (0.82-- 1.40)  |
| CT+TT (vs CC)          | 122 (0.123)     | 116 (0.117)        | 0.642               | 1.07 (0.81-- 1.40)  |
| TT (vs CC+CT)          | 3 (0.003)       | 2 (0.002)          |                     | 1.50 (0.25-- 9.08)  |
| TT (vs CC)             | 3 (0.0034)      | 2 (0.0023)         | 0.685 <sup>††</sup> | 1.52 (0.25-- 9.14)  |
| C                      | 1853 (0.937)    | 1872 (0.941)       |                     | 1 (Ref)             |
| T                      | 125 (0.063)     | 118 (0.059)        | 0.609               | 1.07 (0.82-- 1.39)  |
| <b>rs11200014</b>      | <b>988</b>      | <b>994</b>         | 0.903 <sup>†</sup>  |                     |

|               | Case, n (freq.) | Control, n (freq.) | p-value | *OR (95% CI)       |
|---------------|-----------------|--------------------|---------|--------------------|
| GG            | 737 (0.746)     | 750 (0.755)        |         | 1 (Ref)            |
| GA            | 231 (0.234)     | 224 (0.225)        |         | 1.06 (0.85-- 1.31) |
| AA            | 20 (0.020)      | 20 (0.020)         |         | 1.01 (0.54-- 1.90) |
| P trend       |                 |                    | 0.666   | 1.04 (0.87-- 1.25) |
| GA+AA (vs GG) | 251 (0.254)     | 244 (0.246)        | 0.659   | 1.05 (0.86-- 1.29) |
| AA (vs GG+GA) | 20 (0.020)      | 20 (0.020)         | 0.985   | 1.00 (0.53-- 1.87) |
| AA (vs GG)    | 20 (0.026)      | 20 (0.026)         | 0.956   | 1.01 (0.54-- 1.90) |
| G             | 1705 (0.863)    | 1724 (0.867)       |         | 1 (Ref)            |
| A             | 271 (0.137)     | 264 (0.133)        | 0.689   | 1.04 (0.87-- 1.25) |

\*Odds ratio (OR) adjusted for age. Statistical Test used <sup>†</sup>Pearson  $\chi^2$  or <sup>††</sup>Fishers exact test.

Although in rs11200014, the recessive genotype (AA) was associated with an increased risk of HER2-positive breast cancer compared with the wild type (GG) or heterozygote (AG) (Table 5), the associations were not statistically significant given Bonferroni corrections.

The different immunohistochemical subtypes also were not associated with the *FGFR2* SNPs (Table 6). In addition, ductal carcinoma, lobular carcinoma, and other breast cancer types were not associated with the *FGFR2* SNPs, except that invasive lobular cancers exclusively had the protective allele for rs2981578, whereas the more aggressive invasive ductal cancers had the risk allele (T) (Table 7).

**Table 5: Associations between genotypic frequencies of *FGFR2* variants and immunohistochemical subtyping of patients with Breast Cancer**

| Genotype or allele   | HR+ve (freq.) | HR-ve (freq.) | p-value | OR (95% CI)        | HER2+ve (freq.) | HER2-ve (freq.) | p-value | OR (95% CI)        | HER2+ve HR+ve (freq.) | HER2+ve HR-ve (freq.) | p-value | OR (95% CI)        |
|----------------------|---------------|---------------|---------|--------------------|-----------------|-----------------|---------|--------------------|-----------------------|-----------------------|---------|--------------------|
| <b>rs2981582</b>     | <b>n=776</b>  | <b>n=197</b>  |         |                    | <b>n=266</b>    | <b>n=691</b>    |         |                    | <b>n=678</b>          | <b>n=55</b>           |         |                    |
| GG                   | 229 (0.295)   | 54 (0.274)    |         | 1 (Ref)            | 70 (0.263)      | 209 (0.303)     |         | 1 (Ref)            | 195 (0.288)           | 16 (0.291)            |         | 1 (Ref)            |
| GA                   | 363 (0.468)   | 91 (0.462)    |         | 0.94 (0.64 - 1.36) | 129 (0.485)     | 318 (0.460)     |         | 1.19 (0.85 - 1.68) | 320 (0.472)           | 23 (0.418)            |         | 1.13 (0.58 - 2.20) |
| AA                   | 184 (0.237)   | 52 (0.264)    | 0.699   | 0.83 (0.54 - 1.27) | 67 (0.252)      | 164 (0.237)     | 0.487   | 1.18 (0.79 - 1.75) | 163 (0.240)           | 16 (0.291)            | 0.655   | 0.83 (0.40 - 1.71) |
| P trend              |               |               | 0.414   |                    |                 |                 | 0.299   |                    |                       |                       | 0.656   |                    |
| GA+AA (vs GG)        | 547 (0.705)   | 143 (0.726)   | 0.562   | 0.90 (0.63 - 1.27) | 196 (0.737)     | 482 (0.698)     | 0.231   | 1.19 (0.86 - 1.64) | 483 (0.712)           | 39 (0.709)            | 0.959   | 1.01 (0.55 - 1.85) |
| AA (vs GG+GA)        | 184 (0.237)   | 52 (0.264)    | 0.432   | 0.86 (0.60 - 1.23) | 67 (0.252)      | 164 (0.237)     | 0.638   | 1.06 (0.76 - 1.47) | 163 (0.240)           | 16 (0.291)            | 0.402   | 0.77 (0.42 - 1.41) |
| AA (vs GG)           | 184 (0.446)   | 52 (0.491)    | 0.406   | 0.82 (0.54 - 1.27) | 67 (0.489)      | 164 (0.440)     | 0.321   | 1.19 (0.80 - 1.76) | 163 (0.455)           | 16 (0.500)            | 0.627   | 0.83 (0.40 - 1.72) |
| G                    | 821 (0.529)   | 199 (0.505)   |         | 1 (Ref)            | 269 (0.506)     | 736 (0.533)     |         | 1 (Ref)            | 710 (0.524)           | 55 (0.500)            |         | 1 (Ref)            |
| A                    | 731 (0.471)   | 195 (0.495)   | 0.396   | 0.90 (0.72 - 1.13) | 263 (0.494)     | 646 (0.467)     | 0.291   | 1.09 (0.89 - 1.34) | 646 (0.476)           | 55 (0.500)            | 0.634   | 0.91 (0.61 - 1.34) |
| <b>rs35054928</b>    | <b>n=781</b>  | <b>n=197</b>  |         |                    | <b>n=267</b>    | <b>n=695</b>    |         |                    | <b>n=681</b>          | <b>n=54</b>           |         |                    |
| CC/CC                | 268 (0.343)   | 68 (0.345)    |         | 1 (Ref)            | 92 (0.345)      | 238 (0.342)     |         | 1 (Ref)            | 234 (0.344)           | 22 (0.407)            |         | 1 (Ref)            |
| CC/C                 | 367 (0.470)   | 92 (0.467)    |         | 1.02 (0.72 - 1.45) | 126 (0.472)     | 325 (0.468)     |         | 1.02 (0.74 - 1.40) | 322 (0.473)           | 19 (0.352)            |         | 1.60 (0.85 - 3.03) |
| C/C                  | 146 (0.187)   | 37 (0.188)    | 0.997   | 1.01 (0.65 - 1.59) | 49 (0.184)      | 132 (0.190)     | 0.974   | 1.00 (0.66 - 1.51) | 125 (0.184)           | 13 (0.241)            | 0.219   | 0.91 (0.44 - 1.87) |
| P trend              |               |               | 0.98    |                    |                 |                 | 0.877   |                    |                       |                       | 0.847   |                    |
| CC/C+CC/C (vs CC/CC) | 513 (0.657)   | 129 (0.655)   | 0.957   | 1.02 (0.73 - 1.41) | 175 (0.655)     | 457 (0.658)     | 0.95    | 1.01 (0.75 - 1.37) | 447 (0.656)           | 32 (0.593)            | 0.344   | 1.32 (0.75 - 2.33) |
| C/C (vs CC/CC+CC/C)  | 146 (0.187)   | 37 (0.188)    | 0.977   | 1.00 (0.67 - 1.50) | 49 (0.184)      | 132 (0.190)     | 0.82    | 0.99 (0.68 - 1.43) | 125 (0.184)           | 13 (0.241)            | 0.3     | 0.71 (0.37 - 1.37) |
| C/C (vs CC/CC)       | 146 (0.353)   | 37 (0.352)    | 0.996   | 1.02 (0.65 - 1.59) | 49 (0.348)      | 132 (0.357)     | 0.845   | 0.99 (0.66 - 1.50) | 125 (0.348)           | 13 (0.371)            | 0.783   | 0.91 (0.44 - 1.87) |
| CC                   | 903 (0.578)   | 228 (0.579)   |         |                    | 310 (0.581)     | 801 (0.576)     |         | 1 (Ref)            | 790 (0.580)           | 63 (0.583)            |         | 1 (Ref)            |
| C                    | 659 (0.422)   | 166 (0.421)   | 0.984   | 1.01 (0.81 - 1.26) | 224 (0.420)     | 589 (0.424)     | 0.865   | 1.00 (0.82 - 1.23) | 572 (0.404)           | 45 (0.417)            | 0.947   | 1.02 (0.68 - 1.52) |

| Genotype or allele | HR+ve (freq.) | HR-ve (freq.) | p-value | OR (95% CI)               | HER2+ve (freq.) | HER2-ve (freq.) | p-value      | OR (95% CI)               | HER2+ve HR+ve (freq.) | HER2+ve HR-ve (freq.) | p-value | OR (95% CI)        |
|--------------------|---------------|---------------|---------|---------------------------|-----------------|-----------------|--------------|---------------------------|-----------------------|-----------------------|---------|--------------------|
| <b>rs2981578</b>   | <b>n=780</b>  | <b>n=198</b>  |         |                           | <b>n=267</b>    | <b>n=695</b>    |              |                           | <b>n=680</b>          | <b>n=55</b>           |         |                    |
| CC                 | 692 (0.887)   | 166 (0.838)   |         | 1 (Ref)                   | 242 (0.906)     | 604 (0.869)     |              | 1 (Ref)                   | 609 (0.896)           | 47 (0.855)            |         | 1 (Ref)            |
| CT                 | 85 (0.109)    | 32 (0.162)    |         | <b>0.64 (0.41 - 0.99)</b> | 24 (0.090)      | 89 (0.128)      |              | 0.67 (0.41 - 1.08)        | 68 (0.100)            | 8 (0.146)             |         | 0.65 (0.30 - 1.44) |
| TT                 | 3 (0.0038)    | 0             | 0.089   | -                         | 1 (0.004)       | 2 (0.003)       | 0.254        | 1.25 (0.11 - 14.10)       | 3 (0.004)             | 0                     | 0.508   | -                  |
| P trend            |               |               | 0.065   |                           |                 |                 | 0.114        |                           |                       |                       | 0.353   |                    |
| CT+TT (vs CC)      | 88 (0.113)    | 32 (0.16)     | 0.062   | 0.66 (0.43 - 1.02)        | 25 (0.094)      | 91 (0.131)      | 0.112        | 0.68 (0.43 - 1.09)        | 71 (0.104)            | 8 (0.146)             | 0.345   | 0.68 (0.31 - 1.50) |
| TT (vs CC+CT)      | 3 (0.0038)    | 0             | 0.382   | -                         | 1 (0.004)       | 2 (0.0036)      | 0.829        | 1.31 (0.12 - 14.7)        | 3 (0.004)             | 0                     | 0.622   | -                  |
| TT (vs CC)         | 3 (0.0043)    | 0             | 0.396   | -                         | 1 (0.004)       | 2 (0.003)       | 0.856        | 1.25 (0.11 - 14.08)       | 3 (0.005)             | 0                     | 0.63    | -                  |
| C                  | 1469 (0.942)  | 364 (0.919)   |         | 1 (Ref)                   | 508 (0.951)     | 1297 (0.933)    |              | 1 (Ref)                   | 1286 (0.946)          | 102 (0.927)           |         | 1 (Ref)            |
| T                  | 91 (0.058)    | 32 (0.081)    | 0.1     | 0.70 (0.46 - 1.07)        | 26 (0.049)      | 93 (0.067)      | 0.137        | 0.71 (0.45 - 1.11)        | 74 (0.054)            | 8 (0.073)             | 0.421   | 0.73 (0.34 - 1.56) |
| <b>rs11200014</b>  | <b>n=781</b>  | <b>n=196</b>  |         |                           | <b>n=267</b>    | <b>n=694</b>    |              |                           | <b>n=681</b>          | <b>n=54</b>           |         |                    |
| GG                 | 586 (0.750)   | 145 (0.740)   |         | 1 (Ref)                   | 200 (0.749)     | 519 (0.748)     |              | 1 (Ref)                   | 516 (0.758)           | 35 (0.648)            |         | 1 (Ref)            |
| GA                 | 178 (0.228)   | 48 (0.245)    |         | 0.91 (0.63 - 1.31)        | 57 (0.214)      | 165 (0.238)     |              | 0.86 (0.61 - 1.22)        | 150 (0.220)           | 17 (0.315)            |         | 0.59 (0.32 - 1.08) |
| AA                 | 17 (0.022)    | 3 (0.015)     | 0.764   | 1.41 (0.41 - 4.87)        | 10 (0.038)      | 10 (0.014)      | <b>0.067</b> | <b>2.70 (1.10 - 6.63)</b> | 15 (0.022)            | 2 (0.037)             | 0.196   | 0.50 (0.11 - 2.29) |
| P trend            |               |               | 0.803   |                           |                 |                 | 0.883        |                           |                       |                       | 0.072   |                    |
| GA+AA (vs GG)      | 195 (0.250)   | 51 (0.260)    | 0.762   | 0.94 (0.65 - 1.34)        | 67 (0.251)      | 175 (0.252)     | 0.969        | 0.96 (0.69 - 1.34)        | 165 (0.242)           | 19 (0.352)            | 0.074   | 0.58 (0.32 - 1.05) |
| AA (vs GG+GA)      | 17 (0.022)    | 3 (0.015)     | 0.568   | 1.44 (0.42 - 4.97)        | 10 (0.038)      | 10 (0.014)      | <b>0.025</b> | <b>2.79 (1.14 - 6.83)</b> | 15 (0.022)            | 2 (0.037)             | 0.48    | 0.58 (0.13 - 2.62) |
| AA (vs GG)         | 17 (0.028)    | 3 (0.020)     | 0.592   | 1.41 (0.41 - 4.87)        | 10 (0.048)      | 10 (0.019)      | <b>0.03</b>  | <b>2.71 (1.10 - 6.66)</b> | 15 (0.028)            | 2 (0.054)             | 0.373   | 0.50 (0.11 - 2.28) |
| G                  | 1350 (0.864)  | 338 (0.862)   |         | 1 (Ref)                   | 457 (.856)      | 1203 (0.867)    |              | 1 (Ref)                   | 1182 (0.868)          | 87 (0.806)            |         | 1 (Ref)            |
| A                  | 212 (0.136)   | 54 (0.138)    | 0.916   | 0.98 (0.71 - 1.35)        | 77 (0.144)      | 185 (0.133)     | 0.532        | 1.07 (0.80 - 1.43)        | 180 (0.132)           | 21 (0.194)            | 0.07    | 0.62 (0.38 - 1.03) |

Odds ratio adjusted for age. The statistical test used was Pearson  $\chi^2$  test, with the exception of the P trend, which indicates the non-parametric test for trend. HR, hormone receptor, HER2, human epidermal growth factor receptor 2, freq., frequency, OR, odds ratio, CI, confidence interval

**Table 6: Associations between genotypic frequencies of *FGFR2* variants and immunohistochemical subtyping of patients with breast cancer**

|                   | A-like, n(freq.) | B-like, n(freq.) | B/HER2-like, n(freq.) | HER2-like, n(freq.) | TNBC, n(freq.) | P value             |
|-------------------|------------------|------------------|-----------------------|---------------------|----------------|---------------------|
| <b>rs2981582</b>  | <b>108</b>       | <b>439</b>       | <b>208</b>            | <b>58</b>           | <b>139</b>     | 0.525 <sup>†</sup>  |
| GG                | 36 (0.333)       | 133 (0.303)      | 54 (0.260)            | 16 (0.276)          | 37 (0.266)     |                     |
| GA                | 42 (0.389)       | 209 (0.476)      | 105 (0.505)           | 24 (0.414)          | 66 (0.475)     |                     |
| AA                | 30 (0.278)       | 97 (0.221)       | 49 (0.236)            | 18 (0.310)          | 36 (0.259)     |                     |
| <b>rs35054928</b> | <b>108</b>       | <b>442</b>       | <b>210</b>            | <b>57</b>           | <b>140</b>     | 0.833 <sup>†</sup>  |
| CC/CC             | 39 (0.361)       | 152 (0.344)      | 69 (0.329)            | 23 (0.404)          | 46 (0.329)     |                     |
| CC/C-             | 47 (0.435)       | 207 (0.468)      | 105 (0.500)           | 21 (0.368)          | 70 (0.500)     |                     |
| C-/C-             | 22 (0.204)       | 83 (0.188)       | 36 (0.171)            | 13 (0.228)          | 24 (0.171)     |                     |
| <b>rs2981578</b>  | <b>108</b>       | <b>442</b>       | <b>209</b>            | <b>58</b>           | <b>140</b>     | 0.177 <sup>††</sup> |
| CC                | 92 (0.852)       | 391 (0.885)      | 192 (0.919)           | 50 (0.862)          | 117 (0.836)    |                     |
| CT                | 15 (0.139)       | 50 (0.113)       | 16 (0.077)            | 8 (0.138)           | 23 (0.164)     |                     |
| TT                | 1 (0.009)        | 1 (0.002)        | 1 (0.005)             | 0                   | 0              |                     |
| <b>rs11200014</b> | <b>108</b>       | <b>442</b>       | <b>210</b>            | <b>57</b>           | <b>139</b>     | 0.139 <sup>†</sup>  |
| GG                | 76 (0.704)       | 330 (0.747)      | 163 (0.776)           | 37 (0.649)          | 108 (0.777)    |                     |
| GA                | 31 (0.287)       | 104 (0.235)      | 39 (0.186)            | 18 (0.316)          | 30 (0.216)     |                     |
| AA                | 1 (0.009)        | 8 (0.018)        | 8 (0.038)             | 2 (0.035)           | 1 (0.007)      |                     |

Statistical Test used <sup>†</sup>Pearson  $\chi^2$  or <sup>††</sup>Fishers exact.

A-like, Luminal-A; B-like, Luminal B/HER2-; B/HER2-like, Luminal B/HER2+; TNBC, triple negative breast cancer

**Table 7: Genotypic frequencies of *FGFR2* variants with histological diagnosis**

|                   | Invasive Ductal,<br>n(freq.) | Invasive Lobular,<br>n(freq.) | Other,<br>n(freq.) | P value       |
|-------------------|------------------------------|-------------------------------|--------------------|---------------|
| <b>rs2981582</b>  | <b>814</b>                   | <b>14</b>                     | <b>68</b>          | 0.119         |
| GG                | 230 (0.283)                  | 4 (0.286)                     | 21 (0.309)         |               |
| GA                | 389 (0.478)                  | 3 (0.214)                     | 27 (0.397)         |               |
| AA                | 195 (0.240)                  | 7 (0.500)                     | 20 (0.294)         |               |
| <b>rs35054928</b> | <b>819</b>                   | <b>14</b>                     | <b>68</b>          | 0.270         |
| CC/CC             | 279 (0.341)                  | 7 (0.500)                     | 25 (0.368)         |               |
| CC/C-             | 395 (0.482)                  | 6 (0.429)                     | 26 (0.382)         |               |
| C-/C-             | 145 (0.177)                  | 1 (0.071)                     | 17 (0.250)         |               |
| <b>rs2981578</b>  | <b>819</b>                   | <b>14</b>                     | <b>68</b>          | <b>*0.016</b> |
| CC                | 721 (0.880)                  | 14 (1.0)                      | 55 (0.809)         |               |
| CT                | 97 (0.118)                   | 0                             | 11 (0.162)         |               |
| TT                | 1 (0.001)                    | 0                             | 2 (0.029)          |               |
| <b>rs11200014</b> | <b>819</b>                   | <b>14</b>                     | <b>68</b>          | 0.979         |
| GG                | 606 (0.740)                  | 10 (0.714)                    | 51 (0.750)         |               |
| GA                | 194 (0.237)                  | 4 (0.286)                     | 16 (0.235)         |               |
| AA                | 19 (0.023)                   | 0                             | 1 (0.015)          |               |

Statistical Test used Fishers exact, \*p<0.05

#### 4.6. DISCUSSION

In this study, we evaluated the potential association of *FGFR2* intronic SNPs with breast cancer in black southern African women. *FGFR2* belongs to a tyrosine kinase receptor family that catalyzes multiple processes, including pro-survival signals, anti-apoptotic signals, cell proliferation and cell migration (70). In GWAS studies, the rs2981582 SNP was strongly associated with breast cancer risk (66,67), and particularly with ER-positive cancers. Our study, however found no association between breast cancer and rs2981582. In further analysis, ER, PR, HER2, and immunohistochemical tumor types luminal A, luminal B, HER2 enriched or TNBC, also were not associated with rs2981582.

Intron 2 of *FGFR2* contains putative transcription factor binding sites. Meyer et al (90) showed that the risk-associated allele (C) of rs2981578 preferentially bound FOXA1, and was able to recruit ER $\alpha$  to this site, and that rs35054928 preferentially bound E2F1. E2F1 is important in the regulation of the proliferative response of breast cancer cells to estrogen. The expression of E2F1 increases with later stage of the breast tumors (196). Breast cancer was not associated with rs35054928 (table 4). However, rs2981578 was associated with invasive lobular carcinoma (ILC) (table 7). Classic ILCs are typically of low histologic grade, express ER and PR, and rarely show HER2 protein overexpression or amplification. In our study, all ILCs were homozygous for the major allele, C/C. A defining feature of ILC is a lack of E-cadherin (CDH1) protein expression. Ciriello et al (137) found ILC had mutations in *PTEN*, *TBX3*, and *FOXA1*. These suggest that ILC has mutations in *FOXA1* leading to increased ER recruitment as well as increased FOXA1 binding to *FGFR2*. These results, however, must be taken cautiously, as only 14 ILC samples, out of 989 breast cancer cases, had data on rs2981578. Nevertheless, this association may be interesting to investigate in a larger sample of ILC cases.

Barnholtz-Sloan et al (99) found rs11200014 to be associated with breast cancer in African American women. However, we found no association of rs11200014 with breast cancer in black southern African women. This difference may be caused by the genetic heterogeneity of the two populations; many African American women have both European and African ancestry, and their African ancestry is predominantly West or West-Central African. Additionally, the *FGFR2* intron 2 block is in strong linkage disequilibrium (LD) among

European populations while this LD is weaker among the African populations and thus the selected SNPs in this replication study may not be in LD with the causal variant.

Among our cases, the homozygote recessive genotype AA was more prevalent in HER2-positive (n=10, 3.8%) than in HER2-negative (n=10, 1.4%) tumors; however, given the Bonferroni correction, association was not statistically significant. Fernandez-Noguiera et al (197) found that activation of FGFR2 increased resistance to HER2 therapy. Conversely, when FGFR2 was inactivated, HER2 activity decreased and therapy against resistant HER2 breast cancer cells improved. Hanker et al (198) suggest that resistance to HER2 therapy may be caused by a change from an ER/HER2 signaling pathway to an FGFR2 signaling pathway. About 25% of breast cancers in the South African population are positive for HER2 expression (43,199).

In conclusion, GWAS studies in other populations have highlighted intron 2 of *FGFR2* as a region of interest in breast cancers. We investigated rs2981582, rs35054928, rs2981578 and rs11200014 SNPs in samples from black South African women, but found no significant associations with breast cancer. The black South African population may have weak LD with the causal allele, and the true causal variant may not yet be defined. Some interesting findings, albeit with low numbers, are that rs2981578 is associated with invasive lobular cancer, possibly through the FOXA1 pathway, and that the recessive homozygote of rs11200014 is associated with HER2-positive breast cancer.

## **FORWARD TO CHAPTER 5**

The previous chapter has shown that there were no associations with FGFR2 polymorphisms and breast cancer in black South African women. This was surprising as the literature suggests strong associations of breast cancer with these variants in women in the global north, including African Americans. This finding reinforces the idea that research needs to encompass populations in the south for a better understanding of the genetics around a disease.

Despite the absence of strong associations with the SNPs and breast cancer, we found some suggestive results with rs2981578 and invasive lobular carcinoma, and rs11200014 and HER2 positive breast cancer.

With this as a basis, we wanted to see if these results were also seen in other population groups, and so we investigated the SNPs in publicly available datasets. This work is shown in chapter 5.

This chapter has been formatted in the style of “Frontiers in Genetics” and has not yet been submitted for publication

## **CHAPTER 5: FGFR2 EXPRESSION AND FGFR2 SPLICE ISOFORMS ASSOCIATED WITH BREAST CANCER**

### **5.1. ABSTRACT**

Fibroblast growth factor receptor 2 (FGFR2) has been associated with breast cancer. We performed *in silico* analyses to investigate *FGFR2* mRNA expression and splice variants associated with breast cancer subtypes. Online databases, including cBioPortal, TCGA Splice Seq and TSVdb, were used to examine the association between FGFR2 expression and splice variants with breast cancer subtypes. Statistics were evaluated using STATA v14.2. Higher *FGFR2* mRNA expression was significantly associated with luminal, estrogen receptor (ER) positive breast cancers and invasive lobular carcinomas, whereas lower *FGFR2* was associated with basal, epidermal growth factor receptor 2 (HER2) positive, or triple negative breast cancer and the invasive ductal carcinomas. *FGFR2 IIIb* was significantly found in ER+ breast cancer, and *FGFR2 IIIc* was significantly prevalent in HER2+ cancer. Increased levels of *FGFR2* and *IIIb* splice isoforms are associated with less aggressive breast cancer phenotypes, while decreased levels of *FGFR2* and *IIIc* splice isoforms are associated with more aggressive phenotypes.

### **5.2. INTRODUCTION**

Breast Cancer is the most common malignancy amongst women and causes the most cancer related deaths amongst women worldwide (1). Breast cancer is a heterogeneous disease where the tumors differ greatly in gene expression patterns, growth rates, responses to treatment and clinical outcomes (27,200). Since 2000, gene expression profiles have identified 5 intrinsic molecular subtypes (luminal-A; luminal-B; HER2-enriched; basal-like; normal-like) (29), all of which have different clinical outcomes and treatment options. Immunohistochemistry (IHC), using four biomarkers to distinguish different subtypes, is used as a surrogate for molecular subtyping. The markers are the estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and the proliferation marker, Ki67. These distinguish between ER+ luminal A-like [ER and/or PR positive; HER2-; Ki67 low], ER+ luminal B-like [ER and/or PR positive; HER2-; Ki67 high or ER

and/or PR positive; HER2+; any Ki67]; HER2-enriched [ER and PR neg; HER2+; any Ki67]; and triple negative [ER-/PR-; HER2-; any Ki67] breast tumors (27–29).

Over the last 15 years, genome-wide association studies (GWAS) (66,67,69,188) have led to the detection of over 200 loci associated with breast cancer. These studies allow identification of moderate risk alleles in individuals who do not have high penetrance risk alleles (such as breast cancer 1 gene (*BRCA1*) or 2 (*BRCA2*) and yet still develop breast cancer. One of the strongest associations to come out of GWAS was variant haplotypes in the fibroblast growth factor receptor 2 (*FGFR2*) (66,67). The FGFRs are a family of receptor tyrosine kinases that are involved in signalling pathways with roles in a variety of biological processes, including cell growth, survival, differentiation, angiogenesis, tumorigenesis (70) and epithelial-mesenchymal transition (EMT) (71).

During EMT, epithelial cells lose their characteristic features, such as cell-cell junctions, cobblestone morphology and apical-basal polarity, and become more spindle-like with increased capacity for migration. The cells can degrade the extracellular matrix, invade surrounding tissues and increase resistance to apoptosis (201). EMT is regulated at many levels, including alternative splicing, the epithelial splicing regulatory proteins (ESRPs) and RNA binding Fox-1 homolog-1 (*RBFOX1*) (106,202). Alternative splicing of two mutually exclusive exons of the *FGFR2* gene, namely exon 8 and exon 9, are expressed in epithelial specific- and mesenchymal specific- isoforms, respectively (106–108).

The FGFR structure consists of an extracellular domain with three immunoglobulin (Ig)-like loops (Ig I-III), a transmembrane domain and two intracellular tyrosine kinase (TK) domains (figure 1) (107,108,203,204). The Ig loop III is at the core of the FGFR ligand binding sites, and as such, rearrangements of this domain affect the ligands that can bind to the receptor (103). In *FGFR2*, splicing Ig III results in the first half of the loop being formed by exon 7, and the second half of the loop being formed by either exon 8 (*FGFR2* IIIb isoform) or exon 9 (*FGFR2* IIIc isoform) (figure 1). These splicing sites are mutually exclusive, with *FGFR* IIIb expressed in epithelial cells, and *FGFR2* IIIc expressed in mesenchymal cells. In cancer, the switch from IIIb to IIIc is one of the criteria of EMT and can cause a metastatic phenotype in carcinomas (74,103,105,201,205).

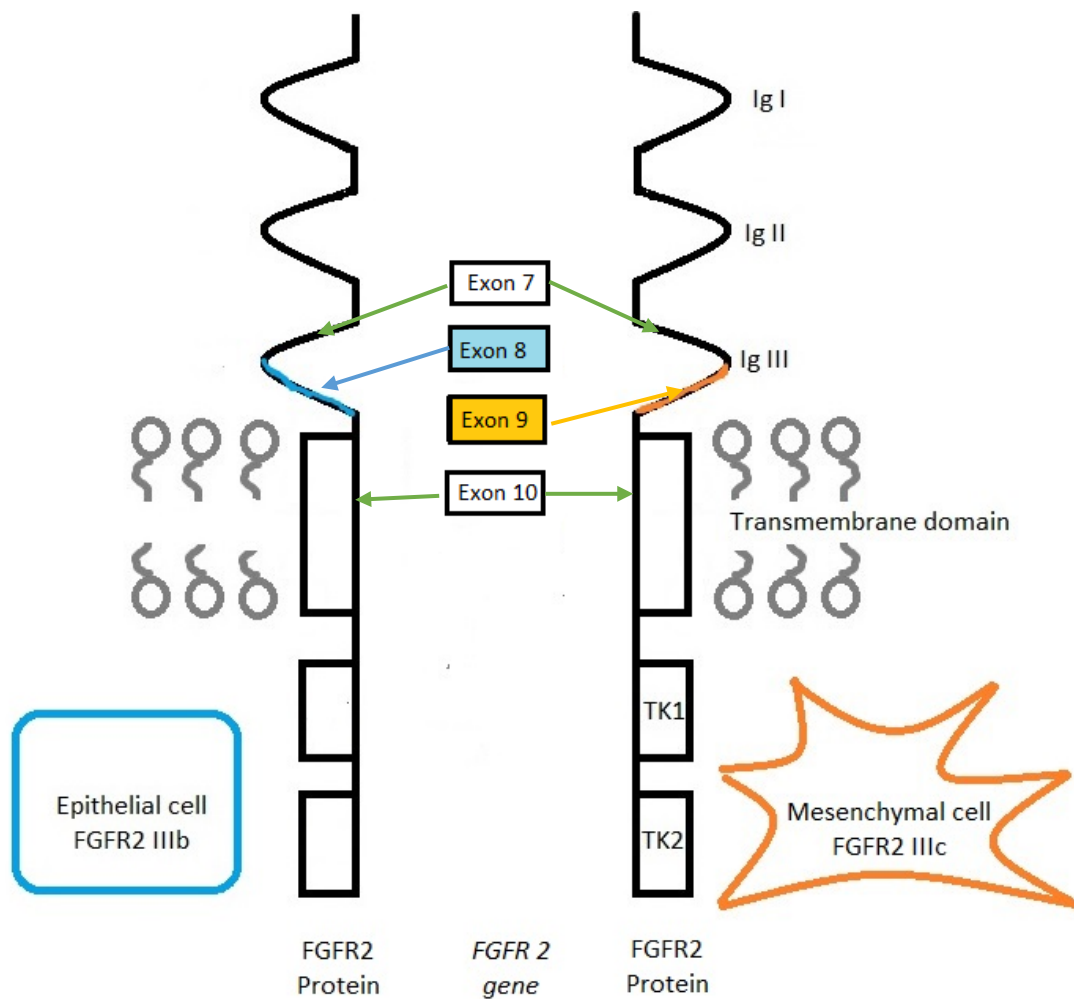


**Figure 1. Scheme showing *FGFR2* IIIb and IIIc splice variants**

The extracellular domain of the FGFR2 protein contains three Ig-like loops, a transmembrane domain, and intracellular tyrosine kinases. The Ig-III loop is coded for by exons 7, 8 and 9. *FGFR2* expresses exons 7 and 8 but excludes exon 9 in the Ig-III domain to form *FGFR2* IIIb in epithelial cells. By contrast, expression of exons 7 and 9 (excluding exon 8), results in *FGFR2* IIIc in the mesenchymal cells. This splice variant is regarded as a hallmark of EMT, with more aggressive carcinomas associated with the IIIc splice variant (figure created with Microsoft Paint software and inspired by Matsuda et al (109)) .

Using data from The Cancer Genome Atlas (TCGA) (<https://www.cancer.gov/tcga>, accessed November 2022), we investigated the associations between breast cancer subtypes and receptor expression with *FGFR2* mRNA expression and *FGFR2* isoforms.

### 5.3. MATERIALS AND METHODS

#### 5.3.1. FGFR2 gene expression levels in breast cancer

Using cBioPortal for cancer genomics (<https://www.cbioportal.org/>), 8993 samples from 14 breast cancer studies, including both general and invasive breast cancer cases, were selected for interrogating *FGFR2* expression. The studies used were: Breast Cancer (MSK, Cancer Cell 2018) (126); Breast Cancer (MSK, Nature Cancer 2020) (127); Breast Cancer (HTAN, 2022); Breast Cancer (METABRIC, Nature 2012 and Nat Commun 2016) (128–130); Breast Cancer (MSK, Clinical Cancer Res 2020) (131); Breast Cancer (MSK, NPJ Breast Cancer 2019) (132); Breast Cancer (SMC, 2018) (133); Breast Invasive Carcinoma (British Columbia, Nature 2012) (134); Breast Invasive Carcinoma (Broad, Nature 2012) (135); Breast Invasive Carcinoma (Sanger, Nature 2012) (136); Breast Invasive Carcinoma (TCGA, Cell 2015) (137); Breast Invasive Carcinoma (TCGA, Firehouse Legacy); Breast Invasive Carcinoma (TCGA, Nature 2012) (138); Breast Invasive Carcinoma (TCGA, PanCancer Atlas) (139–148).

Plot data of log transformed *FGFR2* gene expression levels were downloaded in .csv format for PAM50 subtypes (luminal-A, luminal-B, HER2-enriched and basal-like); immunohistochemistry subtypes (ER+, ER+HER2+, HER2+ and TNBC); HER2 immunohistochemistry score and histological subtype (invasive ductal and invasive lobular carcinomas). *FGFR2* mRNA expression data was downloadable for 117 samples with a PAM50 subtype (5 samples with normal-like subtype were excluded), 163 samples with an immunohistochemistry subtype (5 samples with unknown subtype were excluded), 275 samples with a HER2 score, and 303 sample with a histological subtype.

Data were imported into STATA v14.2 (StataCorp, College Station, TX, USA). A Shapiro-Wilk normality test was performed and data subsequently analyzed non-parametrically. *FGFR2* expression levels were represented as box and whisker plots with outliers excluded. Boxes showed medians, upper and lower quartiles while whiskers showed non-outlying maximum and minimum values. A Kruskal-Wallis equality-of-populations rank test was used to compare expression levels followed by Dunn's test of multiple comparisons to compare expression levels between each group. A Bonferroni correction for multiple comparison testing was applied and p-values of less than 0.01 considered significant.

### 5.3.2. FGFR2 transcript variants

Splice variants of the *Homo Sapiens* FGFR2 gene were downloaded from Ensembl Genome Browser 108 (<https://www.ensembl.org/index.html>) along with the reference genome sequence (GRCh38:CM000672.2). FGFR2 has 41 transcript variants of which 25 are protein coding, 1 is a non-stop decay transcript, 2 undergo nonsense mediated decay, 5 have non-defined protein coding sequences (CDS) and 8 have retained introns (supplementary table S5.1). Transcript variants were aligned to the GRCh38 reference genome using the multiple sequence alignment tool MUSCLE (<https://www.ebi.ac.uk/Tools/msa/muscle/>). Aligned sequences were visualized using GeneDoc v2.7, a multiple sequence alignment editor (<https://github.com/karlricholas/GeneDoc>) (149).

### 5.3.3. FGFR2 splice variants

Splicing patterns of the FGFR2 gene in breast cancer were interrogated using the TCGA SpliceSeq database (<https://bioinformatics.mdanderson.org/TCGASpliceSeq/singlegene.jsp>). Percent spliced in (PSI) values were downloaded for all types of splicing events with filters removed. Information from the Exon Table was used to compare the sequence and location of the TCGA SpliceSeq exons with the transcript variants downloaded from ensembl.org and thus correlate the TCGA SpliceSeq exon numbers with those reported in the literature. TCGA SpliceSeq exon 14 correlated to GRCh38 exon 7, TCGA SpliceSeq exon 15 to GRCh38 exon 8, TCGA SpliceSeq exon 16 to GRCh38 exon 9 and TCGA SpliceSeq exons 17.1 and 17.2 correlated to GRCh38 exon 10 (figure 2).

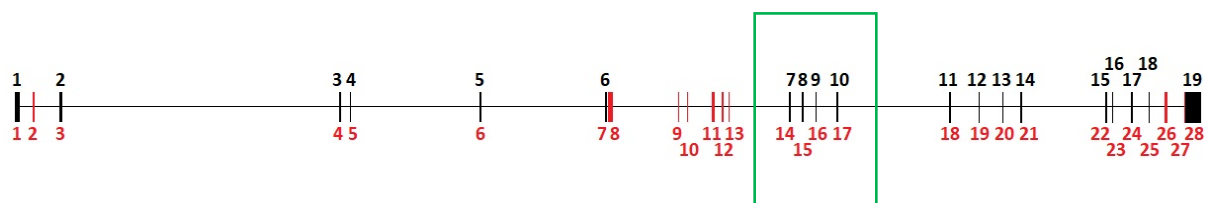


**Figure 2. Representative FGFR2 gene map**

The gene map shows all known splicing events derived from TCGA SpliceSeq data correlated to the GRCh38 sequence. Exons are shown in **black**, consistent with the GRCh38 numbering. Alternatively spliced-in-regions are shown in **red**, as is the TCGA SpliceSeq numbering. The splice variants of interest were exons 7,8,9,10 (GRCh38) which correlated to 14,15,16,17.1, 17.2 (TCGA SpliceSeq), highlighted in the **green** box.

Using the TCGA SpliceSeq database, PSI data were transposed and imported into STATA v14.2. Five exon skipping splicing events (alternative splicing identification (as\_id) 13303: from exon 15 to exon 17.1, skipping exon 16; as\_id 13305: from exon 14 to exon 16, skipping exon 15; as\_id 126925: from exon 14 to exon 17.1, skipping exons 15 and 16; as\_id 126926: from exon 14 to exon 17.1, skipping exon 16 and as\_id 126927: from exon 14 to exon 17.1, skipping exon 15) were identified as of interest. Analysis was restricted to samples from tumor tissue only. A Shapiro-Wilk normality test showed data for these splicing events to be non-normally distributed and so results were presented non-parametrically as medians and interquartile ranges.

The TCGA Splicing Variants Database (<http://tsvdb.com/plot.html>) (150)cn was used to download data on the expression levels of the FGFR2 gene and FGFR2 exons in breast cancer patients along with their clinical estrogen, progesterone and HER2 receptor status. Samples included were restricted to those of female gender. Size and location of each exon was compared to that of the transcript variants downloaded from ensembl.org to correlate the TSVdb exon numbers to those reported in the literature. Data was imported into STATA v14.2. For each sample, expression levels of individual exons were normalized to the overall expression of the FGFR2 gene. A Shapiro-Wilk normality test found these normalized expression levels of the exons to be non-normally distributed and so results were presented non-parametrically as box and whisker plots. Boxes showed medians, upper and lower quartiles while whiskers showed non-outlying maximum and minimum values. Expression levels between the two exons of interest were compared using a Wilcoxon rank-sum test. A Bonferroni correction for multiple comparison testing was applied and p-values of less than 0.005 taken as significant.

## **5.4. RESULTS**

### **5.4.1. FGFR2 mRNA expression and subtypes**

We compared expression of the *FGFR2* gene in the breast cancer subtypes. FGFR2 mRNA expression was highest in luminal-A and luminal-B, slightly lower in basal-like and significantly

lower in HER2-enriched subtypes (figure 3a). Similarly, using IHC, expression levels in HER2 positive samples were significantly lower than ER+ or TNBC samples (figure 3b). This finding was reinforced when examining association between HER2 positive and negative tumors (figure 4). When the tumors were negative (HER2 0 or 1+), or equivocal (HER2 2+), the *FGFR2* expression levels were similar, but levels were significantly lower in HER2 positive tumors (figure 4). On the whole, *FGFR2* expression was higher in ER+ breast carcinomas than ER-cancers (figure 5). There was also an interesting finding with tumor type, with the more aggressive IDC having significantly lower *FGFR2* expression than ILC, and patients who had both ILC/IDC components to their carcinomas had significantly higher *FGFR2* than IDC, but slightly lower levels than ILC alone (figure 6). The median and IQR values of *FGFR2* expression extracted from cBioPortal data plots are shown in supplementary table S5.2.

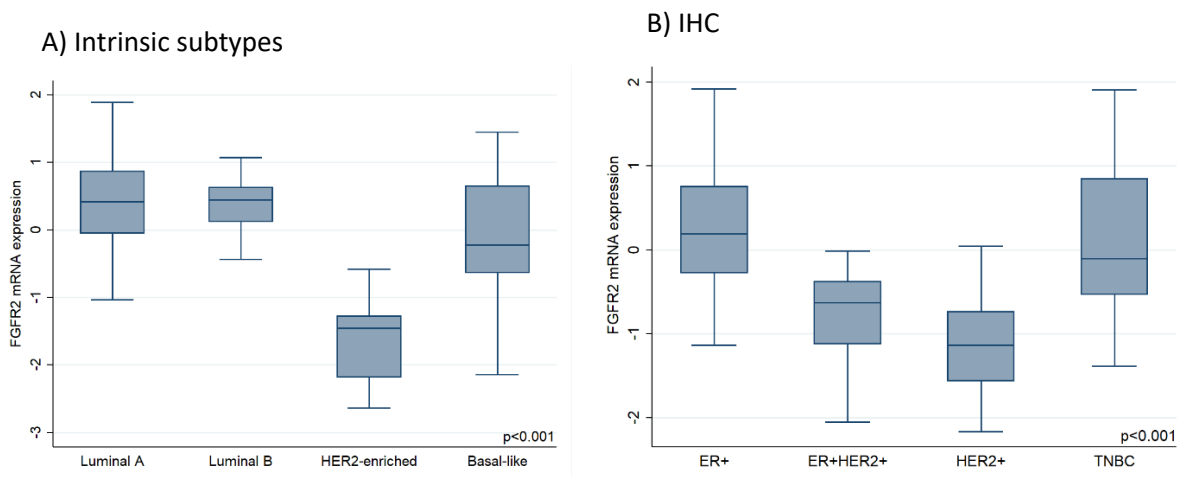


**Figure 3. *FGFR2* mRNA expression in (A) Intrinsic subtypes and (B) IHC subtypes**

(3A) *FGFR2* expression was upregulated in luminal-A and luminal-B subtypes, which were not significantly different from each other, while the basal-like subtype has lower *FGFR2* expression than the luminal subtypes, and the HER2 enriched subtype has much lower expression than the other 3 subtypes.

(3B) These results are recapitulated with IHC classifications. The ER+ subtypes had the highest mRNA expression, followed by TNBC, while HER2 expression is associated with lowered *FGFR2* gene expression. *FGFR* mRNA expression was log<sub>2</sub> transformed.

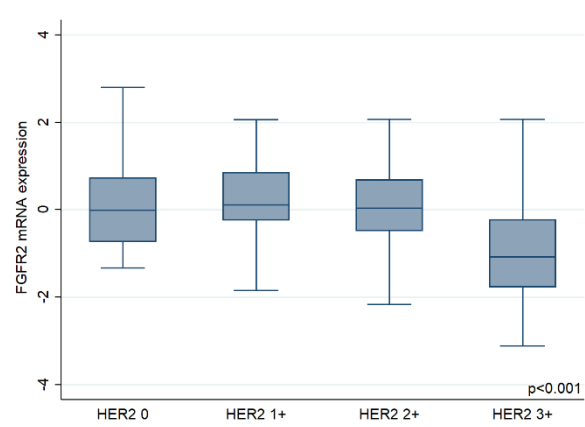


**Figure 4. *FGFR2* expression in HER2 Expression**

HER2 expression is considered negative when there is a value of 0 or 1+, positive at a value of 3+, and equivocal at 2+. The data here shows that HER2 of 0, 1+ and 2+ are not significantly different from each other. However, *FGFR2* is significantly downregulated when HER2 has a 3+ value (i.e. HER2 positive) when compared to HER2 0, 1+, 2+. *FGFR* mRNA expression was log<sub>2</sub> transformed.

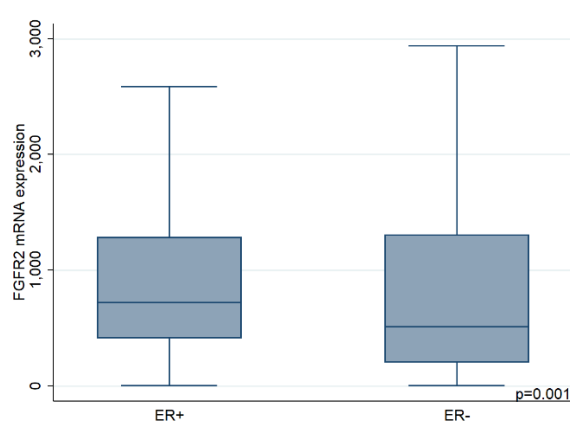


**Figure 5. *FGFR2* expression in ER+ and ER- breast cancers**

*FGFR2* mRNA expression is higher in ER+ than ER- breast cancers.

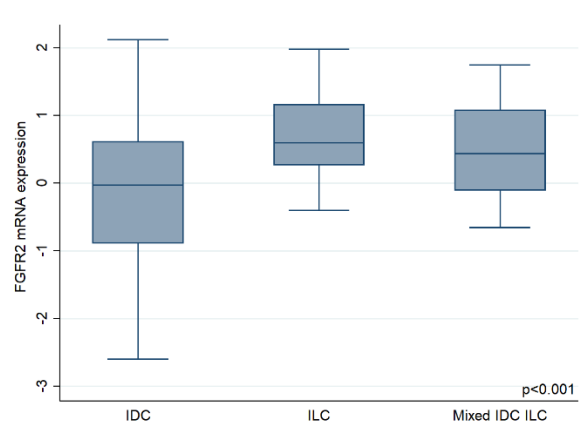


**Figure 6. *FGFR2* expression in invasive ductal and invasive lobular carcinomas**

*FGFR2* mRNA expression was higher in ILC than IDC tumor types. As expected, people who had both lobular and ductal carcinomas had *FGFR2* expression between ILC and IDC.

*FGFR* mRNA in invasive carcinomas presented as log<sub>2</sub> transformed.

#### 5.4.2. Percent spliced in (PSI) of *FGFR2* IIIb and IIIc Isoforms

The TCGA SpliceSeq data was used to investigate PSI values. The proportion of isoforms that included exons 7-9-10 (IIIc) was 44.8% compared with all exons included (7-8-9-10; at 55.2%) (figure 7). The proportion of isoforms that included exons 7-8-10 (IIIb) was 99.22% compared with all exons included (7-8-9-10; at 0.88%). If we compared inclusion of exon 8 (7-8-10) to its exclusion (7-10 only), the PSI has a median of 95.1%, and the inclusion of exon 9 (7-9-10) to its exclusion (exons 7-10 only) is 53.8%. This indicates that there is substantially higher expression of the IIIb than the IIIc variants.

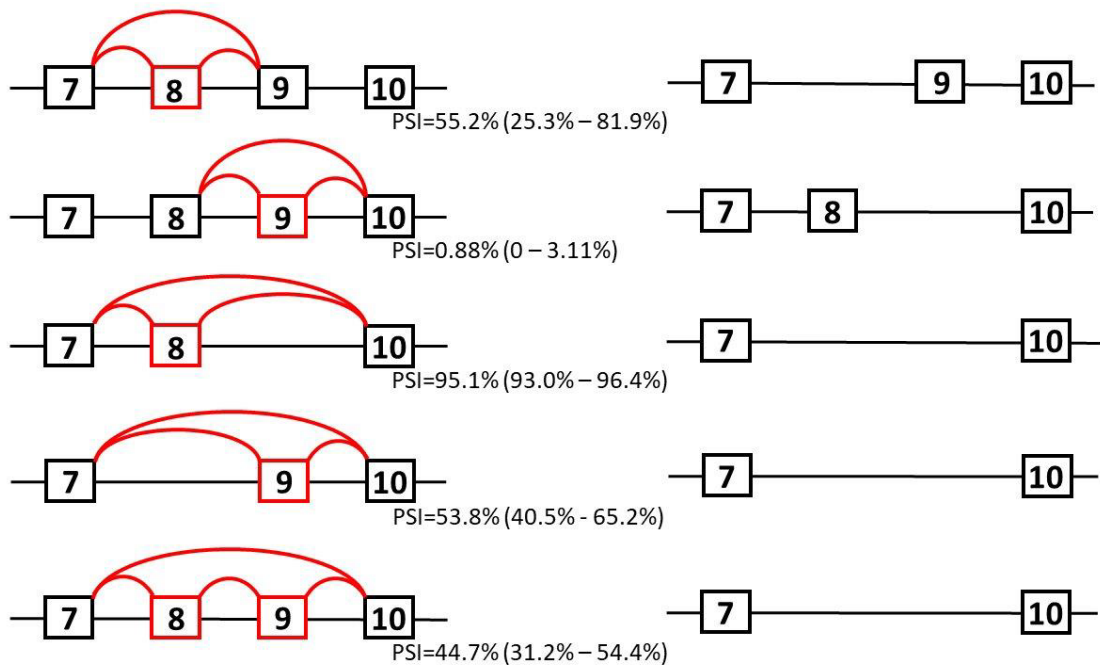


**Figure 7. PSI of *FGFR2* exons 7, 8, 9, 10**

Exclusion of exon 8 from 7-8-9-10, for a 7-9-10 transcript, had a median PSI of 44.8 %, whereas exon 9 excluded from exons 7-8-9-10, for a 7-8-10 transcript, had a median PSI of 99.22%. In the majority of transcript variants, exon 8 is included far more than exon 9. Similarly, including exon 8 with exons 7-10, for a 7-8-10 transcript, had median [IQR] PSI of 95.2% [93.0%-96.4%] and including exon 9 in exons 7-10, with a 7-9-10 transcript had median [IQR] PSI of 53.8% [40.5%-65.2%].

#### 5.4.3. Expression of *FGFR2* exons (exon 8 and exon 9) with IHC

Using the TSVdb database, expression levels of *FGFR2* exon 8 (indicative of *FGFR2* IIIb expression), was higher in ER+ than ER- tumors (figure 8). The more aggressive HER2+ tumors had higher levels of exon 9 (indicative of the *FGFR2* IIIc isoform) (figure 9), while TNBC had lower exon 8 and exon 9 expression levels compared to ER/PR positive cancers (figure 10). Relative *FGFR2* exon 8 and exon 9 expression data (proxies for the IIIb and IIIc isoforms) is shown in supplementary table S5.3.

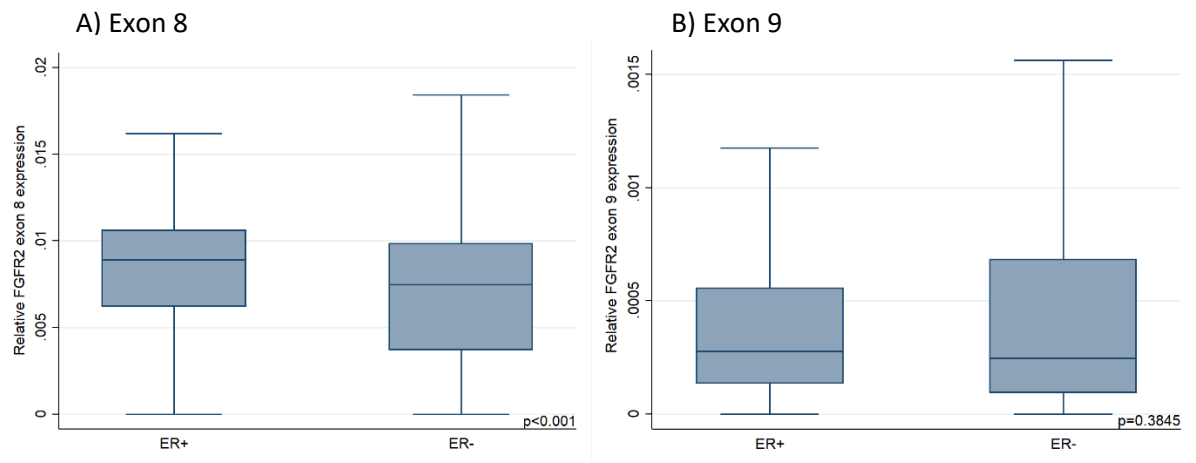


**Figure 8. Expression of FGFR2 Exon 8 (A) and Exon 9 (B) in ER+ and ER- tumors**

There is higher expression of exon 8 (IIIb) in ER+ than ER- tumors, although there is no significant difference in exon 9 (IIIc) expression.

Exon 8 and exon 9 mRNA expression levels presented as relative to overall *FGFR2* expression levels.

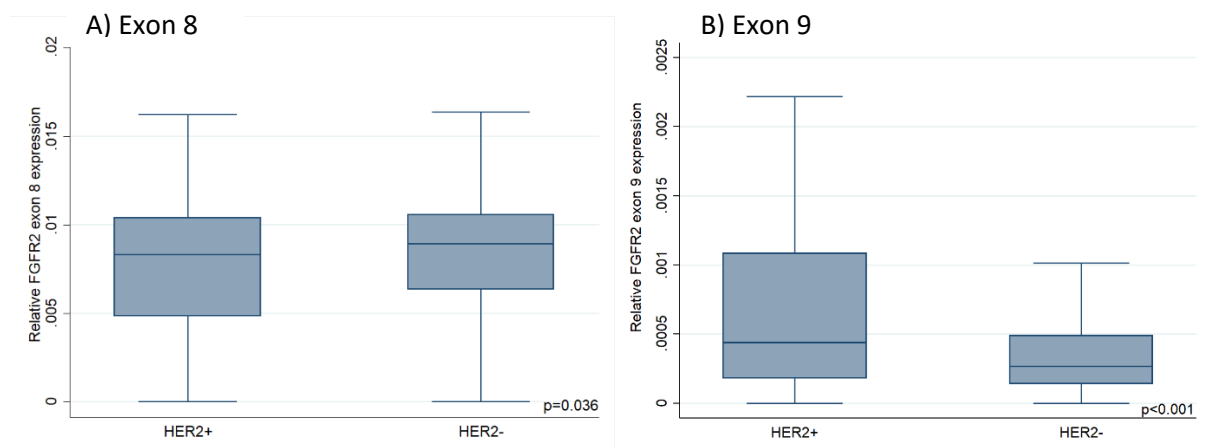


**Figure 9. Expression of FGFR2 Exon 8(A) and Exon 9(B) in HER2+ and HER2- tumors**

There is no significant difference in the expression of exon 8 (IIIb) in HER2+ and HER2- tumors.

The IIIc isoform (exon 9) is higher in HER2 positive than HER2 negative tumors.

Exon 8 and exon 9 mRNA expression levels presented as relative to overall *FGFR2* expression levels.

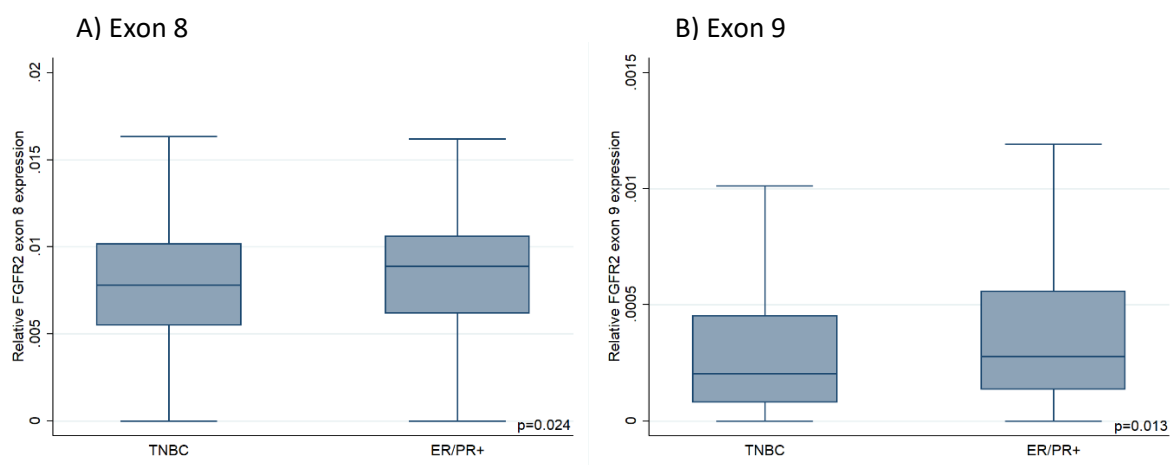


**Figure 10. Expression of FGFR2 Exon 8(A) and Exon 9(B) in TNBC and ER/PR+**

The TNBC has lower IIIb (exon 8) and lower IIIc (exon 9) expression than HR+ tumors.

Exon 8 and exon 9 mRNA expression levels presented as relative to overall *FGFR2* expression levels.

## 5.5. DISCUSSION AND CONCLUSIONS

Aberrant FGFR expression has been reported in many cancers, including cholangiocarcinoma (206), sarcomas (70), keratinocytes (207), endometrial (208), prostate (106), lung (205), colorectal (109) and breast (75) carcinomas. Classic FGF/FGFR pathways include Ras/Raf-MEK-MAPK (mitogen activated protein kinases), phosphatidylinositol-3-kinase/protein kinase b ((PI3K/AKT), phospholipase Cy (PLCy), and signal transducer and activator of transcription (STAT). These pathways are affected by ligand-receptor specificity, expression, and alternative splicing (204). Our study investigated *in silico* the associations between *FGFR2* expression in breast cancer subtypes using publicly available TCGA data, and further interrogated the *FGFR2* exon 8 and exon 9 expression levels, as proxies of the *FGFR2* IIIb and IIIc isoforms, with clinical subtypes.

Our results pointed to higher *FGFR2* expression in luminal tumors, compared to basal-like tumors and lowest expression levels in HER2-enriched tumors. Similarly, the highest levels of FGFR2 were in ER+/HER2- subtypes, lower in TNBC subtypes and lowest in HER2 (ER+ or ER-) subtypes using IHC.

Expression of *FGFR2* mRNA was higher in ER+ than ER- breast cancers. This agrees with other published data where *FGFR2* expression was higher in breast tumors that are ER+PR+ and ER+PR- compared to ER-PR- tumors (209,210).

Estrogen and progesterone are involved in the development of the mammary gland, where estrogen is involved in ductal formation and progesterone promotes lobular growth (211). Progesterone and PR work with estrogen and ER for expansion of glandular structures during breast development (210). The *FGFR2*/*FGF10* control early stages of hormone-dependent development of the mammary ducts; and survival and proliferation of postnatal luminal and basal epithelial cells (210). Low *FGFR2* was associated with a higher grade, higher Ki67 proliferation index and worse overall survival (209). Activation of the *FGFR* pathway leads to resistance of ER-directed therapy against ER+ breast cancers, that can be overcome by *FGFR* pathway inhibitors (91,95,210,212). The change of PR activity by *FGF7*/*FGFR*/*JunB* also changes the response of ER+ breast cancer cells to anti-ER therapies (213).

We found that, overall, *HER2* expression was associated with lower *FGFR2* mRNA expression, whether measured by IHC or intrinsic subtyping. Braun et al. (209) found that low *FGFR2* was associated with poorer overall and disease free survival compared with high *FGFR2* levels. Activation of *FGFR2* decreased the efficacy of *HER2* therapy (197), possibly caused by a change from ER/*HER2* signalling pathway to the *FGFR2* signalling pathway (198). *FGFR2* levels have been found to be amplified and overexpressed in lapatinib-resistant *HER2* positive breast cancer cells (214,215). *HER2*+ breast cancer cells that are lapatinib resistant, have been found to overexpress *FGFR2* levels (214). Moreover, mouse mammary tumor models have shown that combination therapies against *FGF* and *ErbB2* (*HER2*) lead to increased anticancer effects (75,216)

*FGFR2* has been reported in the progression of ER- luminal ductal carcinomas, which are more aggressive and less responsive to treatment, suggesting that *FGFR2* may have an effect promoting IDC progression to an ER independent basal like phenotype (95,210). TNBC cells with *FGFR2* overexpression have a high susceptibility to apoptosis by the *FGFR* inhibitor, PD173074 (217). Zhao et al. (204,218) found the *FGFR* inhibitor, AZD4547, impaired ductal branching and stem-like characteristics in mammary epithelial and tumor cells in mice. The monoclonal antibody, Bemarituzumab (FPA144), is specific against the *FGFR2* IIIb isoform,

and has been found to inhibit cancerous growth in rats, monkeys and humans, particularly in gastroesophageal adenocarcinoma (204,219).

FGFR2 expression was higher in ILC, and lower in the more aggressive IDC. The activation of the FGF7/FGFR2 axis may cause degradation of ER and thus prevent tamoxifen from inhibiting ER+ breast cancer (220). The FGF10/FGFR2 axis was reported to counter estrogen triggered ER+ signalling, thus reducing the effects of anti-ER therapies (91,210). The FGF/FGFR axis was also involved in PR degradation in breast cancer cells (210). Patients with IDC have an inverse relationship between FGFR2 and ER (215,220).

Functional FGFR2 mRNAs include exon 7 and exon 10 (figure 1) and splice in either exon 8, to produce 7-8-9 (IIIb) transcripts; or splice in exon 9 to produce 7-9-10 (IIIc) transcripts (203,221), in a tissue specific manner. Two other transcripts have been described: a skipped product between exons 7 and 10 (7-10); and a double inclusion product that has exons 7, 8, 9 and 10 (7-8-9-10). Both of these mRNAs encode non-functional receptors that are degraded by nonsense mediated decay (NMD). A limitation of this study is that we were unable to directly compare 7-8-10 (IIIb) with 7-9-10 (IIIc). However, if we compared 7-8-10 to 7-8-9-10, and 7-9-10 to 7-8-9-10, the percent spliced in of the IIIb isoform was consistently higher than the IIIc isoform. Similarly, the 7-8-10 to 7-10 had much higher PSI values than 7-9-10 compared to 7-10 transcripts. Together, these show that the IIIb isoform is more highly expressed than IIIc in breast cancer tissues.

The FGFR2-IIIb isoform is expressed in the epithelium, while the FGFR2 IIIc isoform is found in mesenchymal cells (201,222,223). These splice isoforms differ in their ligand binding preferences (223), for example, FGF7, FGF10 and FGF22 bind exclusively to FGFR2 IIIb (107,223), with FGFR2 IIIb/FGF10 highly associated with breast cancer development (210,224). FGFR2 IIIc-positive tumors can be positive for FGFR2IIIb in the same tumor, but both are not positive in the same cancer cells (222,225). Overall, the IIIb isoform was more frequently expressed than the IIIc isoform.

ER+ tumors had higher exon 8 (IIIb) expression than ER- tumors. FGFR2 IIIb is associated with epithelial cells, which have better survival rates. More research had been done in treatment

of FGFR2 IIIb cancers (91,95,210,212,213). FGFR2 IIIc has not been extensively reported in normal human tissue, but it has been associated with various cancers, including prostate, ovarian, gastric and breast (222). Loss of FGFR2 IIIb expression is associated with activation of FGFR2 IIIc expression and a shift to more aggressive behaviour (222). Our results do not show an increase of exon 8 (IIIb) in HER2 positive compared with HER2 negative carcinomas. However, there is a significant increase in expression of exon 9 (IIIc), suggesting an increase in mesenchymal cells in HER2 positive cancers.

The TNBC had lower *FGFR2* mRNA. This recapitulates the work of Lei et al. (226) working in a mouse model that found FGF/FGFR2 signaling drives the development of TNBC as well as epithelial to mesenchymal transition through FGFR2-STAT3 signalling. In humans, FGFR mediated interactions between luminal IDC and its microenvironment can lead to progression of hormone receptor independence (210).

There is interest in FGFR2 as a possible target for therapeutics in various cancers, including breast cancer, increasing the need to understand the effects of FGFR2 on sustaining, and progression, of cancer. Mining data from available sources, we determined that FGFR2 mRNA levels are higher in less aggressive tumors, ER positive tumors and ILC. By contrast, the levels are lower in TNBC, lowest in HER2 expressing tumors, and are lower in IDC. Similarly, the more aggressive splice isoform (IIIc) is found in HER2 positive cancers and TNBC; whereas the splice isoform (IIIb) is found in ER positive cancers. This suggests increased FGFR2 expression is related to maintaining epithelial expression of breast cancer, whereas lower FGFR2 is expressed in more aggressive, basal forms of breast cancer.

## 5.6. SUPPLEMENTARY

**Supplementary Table S5.1. *FGFR2* transcripts**

| Transcript ID      | Name      | bp   | Protein | Translation ID     | Biotype        | CCDS      | UniProt Match |
|--------------------|-----------|------|---------|--------------------|----------------|-----------|---------------|
| ENST00000358487.10 | FGFR2-206 | 4624 | 821aa   | ENSP00000351276.6  | Protein coding | CCDS31298 | P21802        |
| ENST00000683211.1  | FGFR2-235 | 6964 | 819aa   | ENSP00000508257.1  | Protein coding | CCDS81515 | P21802-5      |
| ENST00000351936.11 | FGFR2-203 | 4761 | 819aa   | ENSP00000309878.10 | Protein coding | CCDS81515 | P21802-5      |
| ENST00000613048.4  | FGFR2-224 | 4369 | 732aa   | ENSP00000484154.1  | Protein coding | CCDS73210 | D2CGD1        |
| ENST00000684153.1  | FGFR2-240 | 4251 | 670aa   | ENSP00000506937.1  | Protein coding |           | A0A804HI76    |
| ENST00000478859.5  | FGFR2-218 | 3990 | 593aa   | ENSP00000474011.1  | Protein coding | CCDS81514 | S4R381        |
| ENST00000682550.1  | FGFR2-230 | 3887 | 704aa   | ENSP00000507633.1  | Protein coding | CCDS44485 | P21802-20     |
| ENST00000369061.8  | FGFR2-213 | 3803 | 709aa   | ENSP00000358057.4  | Protein coding | CCDS44486 | P21802-23     |
| ENST00000357555.9  | FGFR2-205 | 3690 | 707aa   | ENSP00000350166.5  | Protein coding | CCDS53584 | P21802-21     |
| ENST00000682772.1  | FGFR2-231 | 3638 | 429aa   | ENSP00000506848.1  | Protein coding |           | A0A1W2PQT9    |
| ENST00000683035.1  | FGFR2-234 | 3615 | 153aa   | ENSP00000507074.1  | Protein coding |           | A0A804HIH8    |
| ENST00000356226.8  | FGFR2-204 | 3547 | 704aa   | ENSP00000348559.4  | Protein coding | CCDS44485 | P21802-20     |
| ENST00000638709.2  | FGFR2-227 | 3290 | 429aa   | ENSP00000491912.2  | Protein coding |           | A0A1W2PQT9    |
| ENST00000369060.8  | FGFR2-212 | 3244 | 705aa   | ENSP00000358056.4  | Protein coding | CCDS44487 | P21802-15     |
| ENST00000457416.6  | FGFR2-215 | 3061 | 822aa   | ENSP00000410294.2  | Protein coding | CCDS7620  | P21802-3      |
| ENST00000369059.5  | FGFR2-211 | 3003 | 707aa   | ENSP00000358055.1  | Protein coding |           | E7EVR7        |
| ENST00000360144.7  | FGFR2-208 | 3001 | 680aa   | ENSP00000353262.3  | Protein coding | CCDS44488 | P21802-22     |
| ENST00000369058.7  | FGFR2-210 | 2727 | 768aa   | ENSP00000358054.3  | Protein coding |           | A0A5S6RJB7    |
| ENST00000369056.5  | FGFR2-209 | 2650 | 769aa   | ENSP00000358052.1  | Protein coding | CCDS44489 | P21802-17     |
| ENST00000336553.10 | FGFR2-201 | 2593 | 669aa   | ENSP00000337665.6  | Protein coding |           | H7BXU9        |
| ENST00000346997.6  | FGFR2-202 | 2472 | 819aa   | ENSP00000263451.5  | Protein coding | CCDS81515 | P21802-5      |
| ENST00000359354.6  | FGFR2-207 | 1680 | 254aa   | ENSP00000352309.2  | Protein coding |           | P21802-14     |
| ENST00000429361.5  | FGFR2-214 | 1138 | 371aa   | ENSP00000404219.1  | Protein coding |           | H7C265        |
| ENST00000613324.4  | FGFR2-225 | 579  | 136aa   | ENSP00000481464.1  | Protein coding |           | A0A087WY21    |
| ENST00000611527.1  | FGFR2-223 | 435  | 64aa    | ENSP00000484892.1  | Protein coding |           | A0A087X2D1    |

| Transcript ID     | Name      | bp   | Protein    | Translation ID    | Biotype                        | CCDS | UniProt Match |
|-------------------|-----------|------|------------|-------------------|--------------------------------|------|---------------|
| ENST00000636922.1 | FGFR2-226 | 587  | 144aa      | ENSP00000490905.1 | Non stop decay                 |      | A0A1B0GWF4    |
| ENST00000683250.1 | FGFR2-236 | 5108 | 137aa      | ENSP00000506847.1 | Nonsense mediated decay        |      | A0A804HI10    |
| ENST00000604236.5 | FGFR2-222 | 3501 | 251aa      | ENSP00000474109.1 | Nonsense mediated decay        |      | S4R3B2        |
| ENST00000682904.1 | FGFR2-232 | 2796 | No protein | -                 | Protein coding CDS not defined |      | -             |
| ENST00000490349.5 | FGFR2-219 | 1478 | No protein | -                 | Protein coding CDS not defined |      | -             |
| ENST00000491111.1 | FGFR2-220 | 574  | No protein | -                 | Protein coding CDS not defined |      | -             |
| ENST00000467584.1 | FGFR2-217 | 497  | No protein | -                 | Protein coding CDS not defined |      | -             |
| ENST00000463870.5 | FGFR2-216 | 386  | No protein | -                 | Protein coding CDS not defined |      | -             |
| ENST00000683418.1 | FGFR2-237 | 6317 | No protein | -                 | Retained intron                |      | -             |
| ENST00000684516.1 | FGFR2-241 | 4989 | No protein | -                 | Retained intron                |      | -             |
| ENST00000683029.1 | FGFR2-233 | 3969 | No protein | -                 | Retained intron                |      | -             |
| ENST00000682296.1 | FGFR2-228 | 3318 | No protein | -                 | Retained intron                |      | -             |
| ENST00000683885.1 | FGFR2-239 | 1838 | No protein | -                 | Retained intron                |      | -             |
| ENST00000683678.1 | FGFR2-238 | 1191 | No protein | -                 | Retained intron                |      | -             |
| ENST00000682400.1 | FGFR2-229 | 846  | No protein | -                 | Retained intron                |      | -             |
| ENST00000491475.1 | FGFR2-221 | 730  | No protein | -                 | Retained intron                |      | -             |

List of human *FGFR2* transcripts downloaded from Ensembl, showing the transcript ID, gene size (bp), size of protein, translation ID, biotype, consensus coding sequence (CCDS) and UniProt identification, to link to transcript associated protein information. The first 25 transcripts shown were protein coding; 1 sample shows that it undergoes non stop decay; 2 undergo nonsense mediated decay; 5 do not have defined coding sequences; and 8 have retained introns.

**Supplementary Table S5.2. Median (IQR) values of *FGFR2* expression levels from cBioPortal with intrinsic subtypes, IHC subtype; HER2 score, ER, tumor subtype**

|                          | <b>N</b>    | <b>Median (IQR)</b>         |
|--------------------------|-------------|-----------------------------|
| <b>Intrinsic subtype</b> | <b>117</b>  |                             |
| Luminal-A                | 57          | 0.4171 (-0.0524; 0.8725)    |
| Luminal-B                | 17          | 0.4426 (0.1236; 0.6321)     |
| HER2-enriched            | 14          | -1.4534 (-2.1825; -1.2758)  |
| Basal-like               | 29          | -0.2207 (-0.6318; 0.6496)   |
| <b>IHC subtype</b>       | <b>163</b>  |                             |
| ER+                      | 90          | 0.1890 (-0.2725; 0.7562)    |
| ER+HER2+                 | 26          | -0.6265 (-1.1214; -0.3790)  |
| HER2+                    | 13          | -1.1338 (-1.5612; -0.7347)  |
| TNBC                     | 34          | -0.1068 (-0.5344; 0.8496)   |
| <b>HER2</b>              | <b>275</b>  |                             |
| HER2 0                   | 36          | -0.0114 (-0.7366; 0.7344)   |
| HER2 1+                  | 116         | 0.1180 (-0.2324; 0.8576)    |
| HER2 2+                  | 81          | 0.0426 (-0.4744; 0.6837)    |
| HER2 3+                  | 42          | -1.0760 (-1.7669; -0.2252)  |
| <b>ER</b>                | <b>1136</b> |                             |
| ER-                      | 259         | 513.825 (203.799; 1302.645) |
| ER+                      | 877         | 722.572 (413.927; 1283.639) |
| <b>Tumor type</b>        | <b>303</b>  |                             |
| IDC                      | 234         | -0.0327 (-0.8895; 0.6076)   |
| ILC                      | 38          | 0.5927 (0.2624; 1.1592)     |
| Mixed IDC/ILC            | 31          | 0.4312 (-0.1056; 1.0745)    |

ER, estrogen receptor; HER2, human epidermal growth factor receptor 2, TNBC, triple negative breast cancer; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma

**Supplementary Table S5.3. Expression of *FGFR2* exon 8 and exon 9 expression levels relative to overall *FGFR2* expression levels with respect to ER, HER2 and TNBC expression.**

|             | <b>N</b>    | <b>Exon 8 Median (IQR)</b>       | <b>Exon 9 Median (IQR)</b>       |
|-------------|-------------|----------------------------------|----------------------------------|
| <b>ER</b>   | <b>1136</b> |                                  |                                  |
| ER-         | 259         | 0.0074912 (0.003694; 0.0098429)  | 0.0002455 (0.000093; 0.0006832)  |
| ER+         | 877         | 0.0089183 (0.0062117; 0.010622)  | 0.0002769 (0.0001346; 0.0005534) |
| <b>HER2</b> | <b>796</b>  |                                  |                                  |
| HER2-       | 613         | 0.0089153 (0.0063422; 0.0105659) | 0.0002659 (0.0001387; 0.0004894) |
| HER2+       | 183         | 0.0082967 (0.0048052; 0.0103987) | 0.0004398 (0.0001803; 0.0010876) |
| <b>TNBC</b> | <b>1023</b> |                                  |                                  |
| ER/PR+      | 896         | 0.0088923 (0.0061799; 0.0106153) | 0.000277 (0.0001352; 0.0005582)  |
| TNBC+       | 127         | 0.0078086 (0.0054925; 0.0101662) | 0.0002042 (0.0000796; 0.0004543) |

ER, estrogen receptor; HER2, human epidermal growth factor receptor 2, TNBC, triple negative breast cancer

## **CHAPTER 6: DISCUSSION AND CONCLUSIONS**

### **6.1. DISCUSSION**

The burden of cancer in Africa is increasing and biological studies of risk factors amongst African populations are important to identify innate risk factors amongst African populations. Cancer is a polygenic disease and techniques to investigate multiple genes are being used to understand the genetic risks of breast cancer in Africa (19). African populations have higher genetic variation than Asian or European populations (18).

Breast cancer in African women has a low age-standardized incident rate (36.2 / 100 000 women per year) but has the highest age-standardized mortality rate (17.3 /100 000) associated with breast cancer, and over time, the comparatively low incidence rate of breast cancer in Africa has been rising (227,228). These higher mortality rates are different among different countries, and survival can be improved with early diagnosis and treatment (229–231).

FGFRs have been shown to be drivers of cancer development and are thus attractive as therapeutic targets (232). FGF/FGFR signalling (particularly FGF10/FGFR2) controls the early stage of ductal development in the developing mammary gland, and in postnatal mammary tissue, the survival and proliferation of luminal and basal cells (210).

In this work, we compared the multigene test for breast cancer, PAM50, with immunohistochemistry in black African women; we examined the effects of FGFR2 SNPs in breast cancer and we studied associations between breast cancer and FGFR2 mRNA expression and splice isoforms.

The ability to diagnose breast cancer subtypes accurately and appropriately, fundamentally affects cancer treatment decisions. PAM50 is widely used for molecular diagnosis of breast cancer subtypes in high income countries (HICs) (156) because its results are reproducible and unaffected by inter- and intra-laboratory variability (157). However, the PAM50 techniques are labor-intensive and expensive. Within resource-constrained settings, it makes more sense

to use IHC as a proxy for intrinsic subtypes because it is less expensive, the infrastructure to run IHC assays is widespread, and it requires less “hands-on” technical expertise than the PAM50 assay. We therefore need accurate and population-specific information to assign proxies that optimize concordance calibration with the PAM50 intrinsic subtyping findings.

Our study is, as far as we know, the first to compare IHC with PAM50 in black southern African women. We found the lowest concordance between molecular and IHC subtyping for the luminal-A group and we recommend raising the cutoff for Ki67 from 14% to 20-25% to distinguish between A-like and B-like tumors, and better reflect the luminal subtypes. Most of our study participants had hormone receptor positive breast cancer, and even tumors with the HER2-enriched subtype were more likely to be HR positive than HR negative. PAM50 is acknowledged to be the gold standard in breast cancer subtyping. However, the cost and labor of the method make it prohibitive for the South African public health care sector as well as other resource limited settings. Its inability to distinguish between HER2+ B-subtypes and HR negative/HER2 positive subtypes is also a limitation. Trastuzumab is expensive and inconsistently available in the South African public sector(163), and the option of using endocrine therapies if trastuzumab is unavailable would be an advantage for HER2 positive patients.

We found that the luminal-A intrinsic subtype had the greatest spread of IHC-analysis subgroups; the A-like IHC group was mainly composed of luminal-A subtype. This observation suggests that the currently used 14-20% Ki67 cutoff in South Africa may be too low. If the Ki67 cutoff were increased to 20-25%, the IHC A-like and B-like distribution would more accurately reflect the intrinsic subtypes. Subtyping strongly affects treatment options, for example patients with luminal-A are likely to benefit from primary endocrine therapies in place of chemotherapy as first choice systemic treatment, whereas the benefits of chemotherapy to patients with luminal-B subtypes may offset chemotherapy side effects (158). The ambiguity in the Ki67 cutoff is not unique to the South African public health care system. Indeed, the German guidelines state that primary invasive tumors that are HR+, HER2-; are considered low risk if  $Ki67 \leq 10\%$ , high risk if  $\geq 25\%$ , and intermediate risk if 10-25%. In this intermediate range (10-25%), Ki67 does not differentiate risk groups accurately (33). By contrast, Fernandez-Martinez et al. (159) found that the 14% cutoff was the best to distinguish

between luminal-A and luminal-B in Spanish and Italian patients. This cohort was tested using IHC and Prosigna™ assays and reinforce the original PCR findings of Cheang et al. (32) that the 14% cutoff was optimal. The ATAC group and the Austrian breast and colorectal cancer group showed that the Prosigna assay can identify a group of patients who do not need adjuvant chemotherapy due to their low risk of recurrence (ROR) at 10 years with endocrine therapy alone. Nevertheless, like Noske et al. (33), we observed better concordance at higher Ki67 levels.

In HICs, where most breast cancers are diagnosed in early stages, the ASCO recommendations (35) suggested that PAM50 could be used to inform chemotherapy decisions, much better than IHC in node negative luminal subtypes. Pu et al. (160) found that survival rates were consistently worse in the luminal-B subtype, irrespective of menopausal status. The 2019 St Gallen report recommended that patients with ER  $\geq$  1% receive endocrine therapy, although it might have limited benefits (158). This recommendation is in line with the South African policy, which regards ER or PR  $\geq$  1% as hormone receptor positive (161). Recently, ASCO has recommended to distinguish ER low (ER 1%-10%), which can receive endocrine therapy, but may also benefit from chemotherapy (46). The Allred score shows ER and/or PR expression is high in luminal-A and luminal-B subtypes, as expected. The PAM50 basal-like subtype was negative for the Allred score and also had Allred score (Allred=3,4). This second finding is interesting, as it may suggest that the Allred cutoff to distinguish between A-like, B-like IHC subtypes and TNC subtypes could be increased to an Allred score  $\leq$  4. However, the sample size in this study is too small to make a definitive claim.

Most tumors of the HER2-enriched luminal subtype are assigned to the B/HER-like IHC-analysis group. This finding is obvious when looking at the Allred score, where most of the HER2-enriched luminal subtypes had high HR positivity. While the multidisciplinary teams follow the St Gallen's recommendations and treat HR+/HER2+ as Clin-B, the PAM50 intrinsic subtypes do not make this subtle distinction. Patients in this group received adjuvant endocrine therapy until 2019, when anti-HER2 therapies were introduced into the South African public health care system. A mere 19% of the HER2 enriched subtype would be HR negative, and would not benefit from endocrine therapy.

Patient subtyping should be interpreted cautiously. Luminal-A patients mistaken for luminal-B may receive overtreatment with chemotherapy, whereas basal-like to normal-like patients may be deprived of appropriate chemotherapy. Confusion of HER2 with luminal-B may result in undertreatment by HER2 targeted therapy (e.g., trastuzumab) and/or overtreatment with chemotherapy and/or endocrine treatment (162). Jensen et al. (233) found, in Danish patients, that patients with HER2- enriched breast cancer derived benefit from anthracycline chemotherapy, whereas anthracyclines are less effective for patients with luminal subtypes.

In a Swedish cohort, Robertson et al. (164) found 81-85% concordance between molecular luminal-A and IHC-A subtypes. However, 35%-52% of luminal-B intrinsic subtypes were classified as IHC-A. Ki67 can distinguish between good and bad prognostic groups with node negative cancer, but Ki67 use is very controversial (164). Also in Sweden, Lundgren et al. (165) found that concordance with luminal subtypes improved when histological grade was included. However, the cohort used in our study had such high histological grades, that including grade with clinical IHC subtype had a negligible effect on the concordance.

Noske et al. (33) found that well differentiated tumors (grade 1) tended to have low Ki67 levels. Intermediate (grade 2) and poorly differentiated tumors (grade 3) had higher Ki67 levels and a wider range of Ki67 values. Fernandez-Martinez et al. (159) compared PAM50 /ROR to Ki67. Most luminal-A tumors with low ROR had low Ki67 values. ROR medium/high with Ki67 0-10% was 42.7% in patients with tumor size <2cm, and 33.9% with tumors >2cm.

Women of African ancestry used to be thought to be, in general, less likely to have hormone receptor positive breast cancers than women of European ancestry. West African women and African American women appear more likely to have TNBC than other women (167–172). However, research has shown that most sub-Saharan Africans (South African, Kenyan, Sudanese, Mozambican, Botswanan) (6,31,111,173–175,183,234) have hormone receptor (HR) positive cancers. Of the samples examined in the current study, 79.5% were HR positive, while the remaining 20.5% were HR negative. However, our cohort was much more likely to be B-like (i.e., HR positive, high Ki67), even when the cutoff of Ki67 is 30%. Such cancers are more aggressive and have a poorer prognosis than those classified as luminal-A subtype or IHC A-like group.

Because our study was part of a HIV outcome study, HIV positive cases and HIV negative controls were age matched within a 5 year age band. Our study participants were therefore younger (49.9 years  $\pm$ 11 years) than South African women with breast cancer overall. Black, African women are more vulnerable to HIV infection than any other group in South Africa, and young women (15-49 years) are at higher risk than older women (235–237). By contrast, a risk factor of breast is increased age, and so unsurprisingly, the HIV positive cohort were younger than the control population. However, in the current study, we did not interrogate the effects of HIV (the analysis of HIV and breast cancer has recently been published by Phakathi et al (238)). In all populations, younger patients with breast cancer are thought to have more clinically aggressive disease and poorer outcomes. For example, Korean breast cancer patients are much more likely to be premenopausal than others (178), and this younger population shows poorer outcomes. However, recent work in South Africa (44) found that black South Africans had consistent expression of about 49-53% HR+/HER2- (A or B-like), 13-18% HR+/HER2+ (B/HER2-like), 7-12% HR-/HER2+ and 23-27% TNC, regardless of age. By comparison, South African whites had 30-65% HR+/HER2- (A or B-like), 9-29% HR+/HER2+ (B/HER2-like), 4-13% HR-/HER2+ and 14-29% TNC, with higher expression of the more aggressive TNC and HER2 expressing tumors in women under 40, and more A-like and B-like in women over 60. Our results, with exclusively black participants, did not show differences in between the distribution of subtypes with age. Although not examined here, the study included a component investigating the subtypes associated with HIV infection. In order to age match, a bias was introduced with regards to the ages of women with breast cancer (44).

There is a dearth of data on the genetic associations of breast cancer in black South African women. We evaluated the potential association of *FGFR2* intronic SNPs with breast cancer in black southern African women as a possible cause of cancer in this genetically diverse population (45). The SNP rs2981582 was highly associated with breast cancer risk GWAS studies (66,67), and particularly with ER positive cancers. Our study, however found no correlation between rs2981582 and breast cancer. Further analysis with ER, PR, HER2, or immunohistochemical tumor types with luminal-A, luminal-B, HER2 enriched or TNC, also showed no association with rs2981582.

Intron 2 of *FGFR2* contains putative transcription factor binding sites. Meyer et al. (90) showed that the risk-associated allele (C) of rs2981578 preferentially bound FOXA1 and was able to recruit ER $\alpha$  to this site and rs35054928 preferentially bound E2F1. E2F1 is important in the regulation of the proliferative response of breast cancer cells to estrogen. The expression of *E2F1* increases with increased malignancy of the breast tumors (196). There were no associations of rs35054928 with breast cancer. Interestingly, there was an association between rs2981578 and invasive lobular carcinoma (ILC). Classic ILCs are typically of low histologic grade, express ER and PR and rarely show HER2 protein overexpression or amplification. In our study, all of the ILC were homozygous for the risk allele, C:C. A defining feature of ILC is a lack of E-cadherin (*CDH1*) protein expression. Ciriello et al. (137) found ILC had mutations in *PTEN*, *TBX3*, and *FOXA1*. These suggest that ILC has FOXA1 mutations leading to increased ER recruitment as well as increased FOXA1 binding to *FGFR2*. These results, however, must be taken cautiously, as there were only 14 ILC samples out of 989 breast cancer cases with data for rs2981578. Nevertheless, this may be interesting to investigate with larger numbers of ILC cases.

Our results pointed to higher *FGFR2* expression in luminal tumors, lower in basal-like like tumors and lowest in HER2-enriched tumors using intrinsic subtyping, and, similarly, highest in ER+/HER2- subtypes, lower in TNC subtypes and lowest in HER2 (ER+ or ER-) subtypes using IHC. We would expect greater *FGFR2* expression in ER+ compared with ER- breast carcinomas (232,239). Interestingly, and unexpectedly, there was no significant difference in the *FGFR2* mRNA expression when comparing ER+ and ER- breast cancers (during normal development of the mammary gland, estrogen and progesterone are highly involved; where estrogen is involved in ductal formation and progesterone promotes lobular growth (211)). However, ER+ tumors had higher exon 8 (*IIIb*) expression than ER- tumors. *FGFR2 IIIb* is associated with epithelial carcinomas, which have better survival rates.

*FGFR2* expression was found to be higher in breast tumors that are ER+PR+ and ER+PR- compared to ER-PR- tumors. However, low *FGFR2* was associated with a higher grade, higher Ki67 proliferation index and worse overall survival. Braun et al.(209) showed a link between low expression of *FGFR2* and poor survival. They also showed a link between *FGFR2* and transcription of PR and increased resistance to anti-ER therapies (209).

Barnholtz-Sloan et al. (99) found rs11200014 to be more prevalent in African American women with breast cancer. However, we found no association with rs11200014 between black southern African women with breast cancer and control participants. This difference may be caused by the genetic heterogeneity in the two populations, with African American women more likely to have both European and African ancestry; and the African ancestry is more likely to be West or West-Central African, while there have been different environmental and genetic influences on southern Africans. Alternatively, this difference may be caused by the different methodologies. Barnholtz-Sloan et al. (99) investigated these SNPs using GWAS studies, while the current work used a candidate gene model to confirm the results. When examining HER2 expression, we found that the homozygote recessive genotype, A:A, was more prevalent in HER2 positive tumors (n=10, 3.8%) than HER2 negative (n=10, 1.4%) tumors, although these are very small numbers. Fernández-Noguiera et al. (197) found that activation of FGFR2 increased resistance to HER2 therapy. Conversely, when FGFR2 was inactivated, HER2 activity decreased and therapy against resistant HER2 breast cancer cells improved. Additionally Hanker et al. (198) suggest that resistance to HER2 therapy may be caused by a change from ER/HER2 signalling pathway to FGFR2 signalling pathway. The South African population has about 25% HER2 positive expression (43,44).

We found that, overall, HER2 expression was associated with lower *FGFR2* mRNA expression. Braun et al. (209) found that low FGFR2 was associated with poorer overall and disease-free survival. Activation of FGFR2 decreased the efficacy of HER2 therapy (197), possibly caused by a change from ER/HER2 signalling pathway to the FGFR2 signalling pathway (198). FGFR2 levels have been found to be amplified and overexpressed in lapatinib-resistant HER2 positive breast cancer cells (214,215). FGFR2 IIIc has not been extensively reported in normal human tissue, but it has been associated with various cancers, including prostate, ovarian, gastric, breast cancers (222). Loss of FGFR2 IIIb expression is associated with activation of FGFR2 IIIc expression and a shift to more aggressive behaviour (222). Our results do not show an increase of exon 8 (*IIIb*) in HER2 positive compared with HER2 negative carcinomas. However, there is a significant increase in expression of exon 9 (*IIIc*), suggesting an increase in the more aggressive FGFR2 isoform, which is also associated with epithelial to mesenchymal transitions.

HER2+ breast cancer cells that are lapatinib resistant, have been found to overexpress FGFR2 levels (214). Moreover, mouse mammary tumor models have shown that combination therapies against FGF and *ErbB2* (HER2) lead to increased anticancer effects (75,216). TNBC cells with FGFR2 overexpression have a high susceptibility to apoptosis by the FGFR inhibitor, PD173074 (217).

The TNBC had lower *FGFR2* mRNA levels. This recapitulates the work of Lei et al. (226), who, working in a mouse model, found that FGF/FGFR2 signaling drives the development of TNBC as well as epithelial to mesenchymal transition through FGFR2-STAT3 signalling. In humans, FGFR mediated interactions between luminal IDC and its microenvironment can lead to progression of hormone receptor independence (210). The TCGA data allows for *in silico* interrogation of many genes. However, downstream analysis is strongly affected by annotations of data that was downloaded into the database. For example, there was limited information on intrinsic subtype, and we were only able to report on 117 samples with both PAM50 data and FGFR2 data. This work would also be improved by laboratory examination of the isoforms with breast cancer subtypes.

There is increasing interest to use tyrosine kinase inhibitors in the treatment of various carcinomas (240). Better knowledge of the action of the tyrosine kinase, FGFR2, in breast cancer may contribute to new therapies associated with breast cancer.

Breast cancer is a complex disease, affected by many gene-gene interactions and gene-environmental factors. We investigated only a small number of genes and believe genome-wide association studies would be beneficial as they would enable us to assay hundreds of thousands of SNPs and relate them to the carcinomas.

## **6.2. FUTURE STUDIES**

Black African people are the most genetically diverse in the world, and their expression of genes in diseases are severely understudied. It would be worthwhile to do a genome-wide

association study on breast cancer in black southern African women to identify groups of genes associated with breast cancer in this group.

The findings of the present study suggest at associations between HER2 and FGFR2; ILC and FGFR2; and these groups should be examined with higher sample numbers to unambiguously show the associations. In addition, it would be useful to interrogate *FGFR2* mRNA isoforms with breast cancer by RTqPCR, as associations between subtypes and *FGFR2* may have implications for anti-FGFR2 treatment options. Functional and mechanistic studies would improve the understanding between FGFR2 and breast cancer.

### **6.3. CONCLUSIONS**

This study, to the best of our knowledge, is the first to compare IHC with PAM50 in black southern African women. Most of our study participants had hormone receptor positive breast cancer, and even tumors with the HER2-enriched subtype were more likely to be HR positive than HR negative. We found better concordance for luminal subtypes if we increased the Ki67 to 25%. We also noted that the black South African population had much higher expression of HER2 than reported in the literature for white populations. Correctly subtyping breast cancer has implications for subsequent treatment.

There is interest in FGFR2 as a possible target for therapeutics in various cancers, including breast cancer, increasing the need to understand the effects of FGFR2 on sustaining and progression of cancer. We found no significant associations with breast cancer when examining rs2981582, rs35054928, rs2918578 and rs11200014 SNPs in samples from black South African women. These results were unlike reported results in African American or European ancestry populations, reinforcing the importance of examining diseases in African populations, and not relying exclusively on data from the global north. The black South African population may have weak LD with the causal allele, and the true causal variant may not yet be defined. Albeit with low sample numbers, we found that rs2981578 is associated with invasive lobular cancer, possibly through the FOXA1 pathway, and that the recessive homozygote of rs11200014 is associated with HER2-positive breast cancer.

By mining data from available sources, we demonstrated that FGFR2 mRNA levels are higher in less aggressive tumors, ER positive tumors and ILC. By contrast, the levels are lower in TNBC, lowest in HER2 expressing tumors, and are lower in IDC. Similarly, the IIIc isoform (exon 9) is found in more aggressive HER2 positive cancers and TNBC; whereas the splice isoform (IIIb) is found in ER positive cancers.

In conclusion, there is less known about the genetics of black, African populations. It is important to investigate gene interactions in this population, as work done in Europe or North America does not reflect the heterogeneous African groups or the polygenic effects on breast cancer in this group. FGFR2 is an interesting gene as it modulates the tumor microenvironment. Our work showed that rs2981582, rs35054928, rs2918578 and rs11200014 did not affect breast cancer in Black southern African women. However, rs2981578 had an effect on intralobular carcinomas and rs11200014 had an effect on HER2 positive carcinomas.



## **CHAPTER 7: REFERENCES**

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209–49.
2. Anyigba CA, Awandare GA, Paemka L. Breast cancer in sub-Saharan Africa: The current state and uncertain future. *Exp Biol Med*. 2021;246(12):1377–87.
3. Stefan DC. Cancer Care in Africa: An Overview of Resources. *J Glob Oncol*. 2015;1(1):30–6.
4. Sharma R, Aashima, Nanda M, Fronterre C, Sewagudde P, Ssentongo AE, et al. Mapping Cancer in Africa: A Comprehensive and Comparable Characterization of 34 Cancer Types Using Estimates From GLOBOCAN 2020. *Front Public Health*. 2022 Apr 25;10.
5. Adams BG, Van de Vijver FJR, de Bruin GP, Bueno Torres C. Identity in Descriptions of Others Across Ethnic Groups in South Africa. *J Cross Cult Psychol*. 2014;45(9):1411–33.
6. Dickens C, Duarte R, Zietsman A, Cubasch H, Kellett P, Schuz J, et al. Racial comparison of receptor-defined breast cancer in Southern African women: Subtype prevalence and age - Incidence analysis of nationwide cancer registry data. *Cancer Epidemiology Biomarkers and Prevention*. 2014;23(11):2311–21.
7. Statistics South Africa. Community Survey 2016, Statistical release P0301 [Internet]. Pretoria, South Africa: Statistics South Africa; 2016. p. 107. Available from: [www.statssa.gov.za](http://www.statssa.gov.za)
8. National Health Laboratory Services. 2019 National Cancer in South Africa. 2019.
9. Bray F, McCarron P, Parkin DM. The changing global patterns of female breast cancer incidence and mortality. *Breast Cancer Res*. 2004;6(6):229–39.
10. Ajaero CK, Wet-Billings N De, Atama C, Agwu P, Eze EJ. The prevalence and contextual correlates of non-communicable diseases among inter-provincial migrants and non-migrants in South Africa. *BMC Public Health*. 2021 Dec 1;21(1).
11. Mabweazara ZS, Leach LL, Smith M, Tsolekile L, Puoane T. Ellisras Longitudinal Study 2017: Patterns of physical activity in an urban and rural setting among black South African adults (ELS 23). *Cardiovasc J Afr*. 2019;30(5):262–7.
12. Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and complex disease. *Nat Rev Genet*. 2002;3(8):611–21.
13. Tishkoff SA, Verrelli BC. Patterns of Human Genetic Diversity: Implications for Human Evolutionary History and Disease. *Annu Rev Genomics Hum Genet*. 2003;4:293–340.
14. Campbell MC, Tishkoff SA. The Evolution of Human Genetic and Phenotypic Variation in Africa. *Current Biology*. 2010 Feb 23;20(4).

15. Ingman M, Kaessmann H, Pääbo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. *Nature*. 2000 Dec 7;408(6813):708–13.
16. Campbell MC, Tishkoff SA. African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet*. 2008;9:403–33.
17. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature* [Internet]. 2015;517(7534):327–32. Available from: <http://dx.doi.org/10.1038/nature13997>
18. Gurdasani D, Barroso I, Zeggini E, Sandhu MS. Genomics of disease risk in globally diverse populations. *Nat Rev Genet*. 2019;20(9):520–35.
19. Rotimi SO, Rotimi OA, Salhia B. A Review of Cancer Genetics and Genomics Studies in Africa. *Front Oncol*. 2021;10(February):1–24.
20. Hayat M, Chen WC, Brandenburg JT, Babb de Villiers C, Ramsay M, Mathew CG. Genetic Susceptibility to Breast Cancer in Sub-Saharan African Populations. *JCO Glob Oncol*. 2021;(7):1462–71.
21. Garcia-Closas, Montserrat , Chanock S. Genetic susceptibility loci for breast cancer by estrogen receptor (ER) status. *Clin Cancer Res*. 2009;14(24):8000–9.
22. Mavaddat N, Pharoah PDP, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst*. 2015;107(5):1–15.
23. Haddad SA, Lunetta KL, Ruiz-Narvaez EA, Bensen JT, Hong CC, Sucheston-Campbell LE, et al. Hormone-related pathways and risk of breast cancer subtypes in African American women. *Breast Cancer Res Treat*. 2015;154(1):145–54.
24. Bray F, Parkin DM, Gnanngnon F, Tshisimogo G, Peko JF, Adoubi I, et al. Cancer in sub-Saharan Africa in 2020: a review of current estimates of the national burden, data gaps, and future needs. *Lancet Oncol*. 2022 Jun 1;23(6):719–28.
25. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. <https://gco.iarc.fr/today/>. 2020. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer.No Title.
26. Sharma R. Breast cancer burden in Africa: evidence from GLOBOCAN 2018. *J Public Health (Oxf)*. 2021;43(4):763–71.
27. Viale G. The current state of breast cancer classification. *Annals of Oncology*. 2012;23(SUPPL. 10):x207–10.
28. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes-dealing with the diversity of breast cancer: Highlights of the St Gallen international

- expert consensus on the primary therapy of early breast cancer 2011. *Annals of Oncology*. 2011;22(8):1736–47.
29. Perou CM, Sørli T, Eisen MB, van de Rijn M, Jeffrey SS, Rees C a, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–52.
  30. Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, et al. Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res*. 2015;5(10):2929–43.
  31. McCormack V a, Joffe M, van den Berg E, Broeze N, Silva IDS, Romieu I, et al. Breast cancer receptor status and stage at diagnosis in over 1,200 consecutive public hospital patients in Soweto, South Africa: a case series. *Breast Cancer Res*. 2013;15(5):R84.
  32. Cheang MCU, Chia SK, Voduc D, Gao D, Leung S, Snider J, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst*. 2009;101(10):736–50.
  33. Noske A, Anders SI, Ettl J, Hapfelmeier A, Steiger K, Specht K, et al. Risk stratification in luminal-type breast cancer: Comparison of Ki-67 with EndoPredict test results. *Breast*. 2020 Feb;49:101–7.
  34. Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart MJ, et al. Tailoring therapies-improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Annals of Oncology*. 2015;26(8):1533–46.
  35. Harris LN, Ismaila N, McShane LM, Andre F, Collyar DE, Gonzalez-Angulo AM, et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of clinical Oncology clinical practice guideline. *Journal of Clinical Oncology*. 2016;34(10):1134–50.
  36. Van Den Berg EJ, Duarte R, Dickens C, Joffe M, Mohanlal R. Ki67 Immunohistochemistry Quantification in Breast Carcinoma: A Comparison of Visual Estimation, Counting, and ImmunoRatio. *Applied Immunohistochemistry and Molecular Morphology*. 2021;29(2):105–11.
  37. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *New England Journal of Medicine*. 2004 Dec 30;351(27):2817–26.
  38. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and Clinical Utility of a 70-Gene Prognostic Signature for Women With Node-Negative Breast Cancer. *JNCI: Journal of the National Cancer Institute*. 2006 Sep 6;98(17):1183–92.
  39. van de Vijver MJ, He YD, van 't Veer LJ, Dai H, Hart AAM, Voskuil DW, et al. A Gene-Expression Signature as a Predictor of Survival in Breast Cancer. *New England Journal of Medicine*. 2002 Dec 19;347(25):1999–2009.
  40. Filipits M, Rudas M, Jakesz R, Dubsy P, Fitzal F, Singer CF, et al. A New Molecular Predictor of Distant Recurrence in ER-Positive, HER2-Negative Breast Cancer Adds Independent

- Information to Conventional Clinical Risk Factors. *Clinical Cancer Research*. 2011 Sep 15;17(18):6012–20.
41. Parker JS, Mullins M, Cheung MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*. 2009;27(8):1160–7.
  42. Prosigna Insert. Prosigna<sup>®</sup> Breast Cancer Prognostic Gene Signature Assay. 2015. p. 1–18.
  43. Phakathi B, Cubasch H, Nietz S, Dickens C, Dix-Peek T, Joffe M, et al. Clinico-pathological characteristics among South African women with breast cancer receiving anti-retroviral therapy for HIV. *Breast*. 2019 Feb 1;43:123–9.
  44. Achilonu OJ, Singh E, Nimako G, Eijkemans RMJC, Musenge E. Rule-Based Information Extraction from Free-Text Pathology Reports Reveals Trends in South African Female Breast Cancer Molecular Subtypes and Ki67 Expression. *Biomed Res Int*. 2022;1–17.
  45. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clinical Cancer Research*. 2010;16(21):5222–32.
  46. Najjar S, Allison KH. Updates on breast biomarkers. *Virchows Archiv*. 2022 Jan 1;480:163–76.
  47. Johansson A, Yu NY, Iftimi A, Tobin NP, 't Veer L, Nordenskjöld B, et al. Clinical and molecular characteristics of estrogen receptor-positive ultralow risk breast cancer tumors identified by the 70-gene signature. *Int J Cancer*. 2022 Jun 15;150(12):2072–82.
  48. Yu NY, Iftimi A, Yau C, Tobin NP, van 't Veer L, Hoadley KA, et al. Assessment of Long-term Distant Recurrence-Free Survival Associated With Tamoxifen Therapy in Postmenopausal Patients With Luminal A or Luminal B Breast Cancer. *JAMA Oncol*. 2019 Sep 1;5(9):1304.
  49. Loibl S, Gianni L. HER2-positive breast cancer. *The Lancet*. 2017 Jun;389(10087):2415–29.
  50. Swain SM, Shastry M, Hamilton E. Targeting HER2-positive breast cancer: advances and future directions. *Nat Rev Drug Discov*. 2023 Nov 7;22(2):101–26.
  51. Almansour NM. Triple-Negative Breast Cancer: A Brief Review About Epidemiology, Risk Factors, Signaling Pathways, Treatment and Role of Artificial Intelligence. *Front Mol Biosci*. 2022 Jan 25;9.
  52. Foulkes WD, Smith IE, Reis-Filho JS. Triple-Negative Breast Cancer. *New England Journal of Medicine*. 2010 Nov 11;363(20):1938–48.
  53. Almansour NM. Triple-Negative Breast Cancer: A Brief Review About Epidemiology, Risk Factors, Signaling Pathways, Treatment and Role of Artificial Intelligence. Vol. 9, *Frontiers in Molecular Biosciences*. Frontiers Media S.A.; 2022.

54. Bou Zerdan M, Ghorayeb T, Saliba F, Allam S, Bou Zerdan M, Yaghi M, et al. Triple Negative Breast Cancer: Updates on Classification and Treatment in 2021. Vol. 14, *Cancers*. MDPI; 2022.
55. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of Early-Onset Familial Breast Cancer to Chromosome 17q21. *Science* (1979). 1990 Dec 21;250(4988):1684–9.
56. Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science*. 1994 Sep 30;265(5181):2088–90.
57. Shiovitz S, Korde LA. Genetics of breast cancer: A topic in evolution. *Annals of Oncology*. 2015 Jul 1;26(7):1291–9.
58. Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M. Genetic susceptibility to breast cancer. *Mol Oncol*. 2010;4(3):174–91.
59. Eygelaar D, van Rensburg EJ, Joubert F. Germline sequence variants contributing to cancer susceptibility in South African breast cancer patients of African ancestry. *Sci Rep*. 2022 Dec 1;12(1).
60. Francies FZ, Wainstein T, De Leeneer K, Cairns A, Murdoch M, Nietz S, et al. BRCA1, BRCA2 and PALB2 mutations and CHEK2 c.1100delC in different South African ethnic groups diagnosed with premenopausal and/or triple negative breast cancer. *BMC Cancer*. 2015 Nov 17;15(1).
61. Van der Merwe NC, Combrink HME, Ntita KS, Oosthuizen J. Prevalence of Clinically Relevant Germline BRCA Variants in a Large Unselected South African Breast and Ovarian Cancer Cohort: A Public Sector Experience. *Front Genet*. 2022 Apr 8;13.
62. Van Der Merwe NC, Oosthuizen J, Theron M, Chong G, Foulkes WD. The contribution of large genomic rearrangements in BRCA1 and BRCA2 to South African familial breast cancer. *BMC Cancer*. 2020 May 6;20(1).
63. Adedokun B, Zheng Y, Ndom P, Gakwaya A, Makumbi T, Zhou AY, et al. Prevalence of inherited mutations in breast cancer predisposition genes among women in Uganda and Cameroon. *Cancer Epidemiology Biomarkers and Prevention*. 2020 Feb 1;29(2):359–67.
64. George SHL, Donenberg T, Alexis C, Degennaro V, Dyer H, Yin S, et al. Gene sequencing for pathogenic variants among adults with breast and ovarian cancer in the caribbean. *JAMA Netw Open*. 2021 Mar 1;4(3).
65. Adedokun B, Du Z, Gao G, Ahearn TU, Lunetta KL, Zirpoli G, et al. Cross-ancestry GWAS meta-analysis identifies six breast cancer loci in African and European ancestry women. *Nat Commun*. 2021 Dec 1;12(1).

66. Easton DF, Pooley K a, Dunning AM, Pharoah PDP, Ballinger DG, Struewing JP, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007;447(7148):1087–93.
67. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*. 2007;39(7):870–4.
68. Fachal L, Dunning AM. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. *Curr Opin Genet Dev*. 2015;30:32–41.
69. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet*. 2020 Jun 1;52(6):572–81.
70. Zhou WY, Zheng H, Du XL, Yang JL. Characterization of FGFR signaling pathway as therapeutic targets for sarcoma patients. *Cancer Biol Med*. 2016;13(2):260–8.
71. Katoh Y, Katoh M. FGFR2-related pathogenesis and FGFR2-targeted therapeutics (Review). *Int J Mol Med*. 2009;23(3):307–11.
72. André F, Cortés J. Rationale for targeting fibroblast growth factor receptor signaling in breast cancer. *Breast Cancer Res Treat*. 2015;150(1):1–8.
73. Dienstmann R, Rodon J, Prat A, Perez-Garcia J, Adamo B, Felip E, et al. Genomic aberrations in the FGFR pathway: Opportunities for targeted therapies in solid tumors. *Annals of Oncology*. 2014;25(3):552–63.
74. Fearon AE, Gould CR, Grose RP. FGFR signalling in women’s cancers. *International Journal of Biochemistry and Cell Biology*. 2013;45(12):2832–42.
75. Lei H, Deng CX. Fibroblast growth factor receptor 2 signaling in breast cancer. *Int J Biol Sci*. 2017;13(9):1163–71.
76. Miki T, Bottaro DP, Fleming TP, Smith CL, Burgess WH, Chan AML, et al. Determination of ligand-binding specificity by alternative splicing: Two distinct growth factor receptors encoded by a single gene. *Proc Natl Acad Sci USA*. 1992;89:246–50.
77. Krook MA, Reeser JW, Ernst G, Barker H, Wilberding M, Li G, et al. Fibroblast growth factor receptors in cancer: genetic alterations, diagnostics, therapeutic targets and mechanisms of resistance. *Br J Cancer*. 2021 Mar 2;124(5):880–92.
78. Udler MS, Meyer KB, Pooley KA, Karlins E, Struewing JP, Zhang J, et al. FGFR2 variants and breast cancer risk: Fine-scale mapping using African American studies and analysis of chromatin conformation. *Hum Mol Genet*. 2009;18(9):1692–703.
79. Rebbeck TR, Demichele A, Tran T V., Panossian S, Bunin GR, Troxel AB, et al. Hormone-dependent effects of FGFR2 and MAP3K1 in breast cancer susceptibility in a population-based

- sample of post-menopausal African-American and European-American women. *Carcinogenesis*. 2009;30(2):269–74.
80. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet*. 2019 Jan 3;104(1):21–34.
  81. May A, Hazelhurst S, Li Y, Norris S a, Govind N, Tikly M, et al. Genetic diversity in black South Africans from Soweto. *BMC Genomics*. 2013;14:644.
  82. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.
  83. Fairley S, Lowy-Gallego E, Perry E, Flicek P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res*. 2020 Jan 1;48(D1):D941–7.
  84. AlRaddadi RIR, Alamri RJN, Shebli WTY, Fallatah EIY, Alhujaily AS, Mohamed HS, et al. Fibroblast growth factor receptor 2 gene (FGFR2) rs2981582T/C polymorphism and susceptibility to breast cancer in Saudi women. *Saudi J Biol Sci*. 2021;28(11):6112–5.
  85. Liang H, Yang X, Chen L, Li H, Zhu A, Sun M, et al. Heterogeneity of Breast Cancer Associations with Common Genetic Variants in FGFR2 according to the Intrinsic Subtypes in Southern Han Chinese Women. *Biomed Res Int*. 2015;2015:626948.
  86. Han W, Woo JH, Yu JH, Lee MJ, Moon HG, Kang D, et al. Common Genetic Variants Associated with Breast Cancer in Korean Women and Differential Susceptibility According to Intrinsic Subtype. *Cancer Epidemiology Biomarkers & Prevention*. 2011;20(5):793–8.
  87. Purnomosari D, Raharjo C, Kalim AS, Herviastuti R, Yushan M, Fajar RA, et al. P21 Ser31Arg and FGFR2 rs2981582 polymorphisms as risk factors for early onset of breast cancer in Yogyakarta, Indonesia. *Asian Pacific Journal of Cancer Prevention*. 2019;20(11):3305–9.
  88. Brignoni L, Cappetta M, Colistro V, Sans M, Artagaveytia N, Bonilla C, et al. Genomic Diversity in Sporadic Breast Cancer in a Latin American Population. *Genes (Basel)*. 2020;11(11):1272.
  89. Özgöz A, Mutlu İçduygu F, Yükseltürk A, Şamlı H, Hekimler Öztürk K, Başkan Z. Low-penetrance susceptibility variants and postmenopausal oestrogen receptor positive breast cancer. *J Genet*. 2020;99(1).
  90. Meyer KB, O'Reilly M, Michailidou K, Carlebur S, Edwards SL, French JD, et al. Fine-scale mapping of the FGFR2 breast cancer risk locus: Putative functional variants differentially bind FOXA1 and E2F1. *Am J Hum Genet*. 2013;93(6):1046–60.
  91. Campbell TM, Castro MAA, De Santiago I, Fletcher MNC, Halim S, Prathalingam R, et al. FGFR2 risk SNPs confer breast cancer risk by augmenting oestrogen responsiveness. *Carcinogenesis*. 2016;37(8):741–50.

92. Lu P, Hankel IL, Hostager BS, Swartzendruber JA, Friedman AD, Brenton JL, et al. The developmental regulator protein Gon4l associates with protein YY1, co-repressor Sin3a, and histone deacetylase 1 and mediates transcriptional repression. *Journal of Biological Chemistry*. 2011 May 20;286(20):18311–9.
93. Owens TW, Rogers RL, Best SA, Ledger A, Mooney AM, Ferguson A, et al. Runx2 is a novel regulator of mammary epithelial cell fate in development and breast cancer. *Cancer Res*. 2014 Jul 23;74(18):5277–86.
94. Meyer KB, Maia AT, O'Reilly M, Teschendorff AE, Chin SF, Caldas C, et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol*. 2008;6(5):1098–103.
95. Campbell TM, Castro MAA, de Oliveira KG, Ponder BAJ, Meyer KB. Era binding by transcription factors NFIB and YBX1 enables FGFR2 signaling to modulate estrogen responsiveness in breast cancer. *Cancer Res*. 2018;78(2):410–21.
96. Zhu X, Asa SL, Ezzat S. Histone-acetylated control of fibroblast growth factor receptor 2 intron 2 polymorphisms and isoform splicing in breast cancer. *Molecular Endocrinology*. 2009;23(9):1397–405.
97. Ogura T, Azuma K, Sato J, Kinowaki K, Takayama KI, Takeiwa T, et al. OCT1 Is a Poor Prognostic Factor for Breast Cancer Patients and Promotes Cell Proliferation via Inducing NCAPH. *International Journal of Molecular Sciences* . 2021;22:11505.
98. Martini R, Chen Y, Jenkins BD, Elhussin IA, Cheng E, Hoda SA, et al. Investigation of triple-negative breast cancer risk alleles in an International African-enriched cohort. *Sci Rep*. 2021 Dec 1;11(1).
99. Barnholtz-Sloan JS, Shetty PB, Guan X, Nyante SJ, Luo J, Brennan DJ, et al. FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. *Carcinogenesis*. 2010;31(8):1417–23.
100. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of african americans, latinos, and european Americans across the United States. *Am J Hum Genet*. 2015 Jan 8;96(1):37–53.
101. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature*. 2015;517(7534):327–32.
102. Zakharia F, Basu A, Absher D, Assimes TL, Go AS, Hlatky MA, et al. Characterizing the admixed African ancestry of African Americans. *Genome Biol*. 2009 Dec 22;10(12).
103. Holzmann K, Grunt T, Heinzle C, Sampl S, Steinhoff H, Reichmann N, et al. Alternative splicing of fibroblast growth factor receptor IgIII loops in cancer. *J Nucleic Acids*. 2012;2012.
104. Eswarakumar VP, Lax I, Schlessinger J. Cellular signaling by fibroblast growth factor receptors. *Cytokine Growth Factor Rev*. 2005 Apr;16(2):139–49.

105. Kelleher FC, O'Sullivan H, Smyth E, McDermott R, Viterbo A. Fibroblast growth factor receptors, developmental corruption and malignant disease. *Carcinogenesis*. 2013;34(10):2198–205.
106. Li L, Zheng J, Stevens M, Oltean S. A repositioning screen using an FGFR2 splicing reporter reveals compounds that regulate epithelial-mesenchymal transitions and inhibit growth of prostate cancer xenografts. *Mol Ther Methods Clin Dev*. 2022 Jun 9;25:147–57.
107. Ornitz DM, Itoh N. The fibroblast growth factor signaling pathway. *Wiley Interdiscip Rev Dev Biol*. 2015;4(3):215–66.
108. Yan G, Fukabori Y, McBride G, Nikolaropolous S, McKeenan WL, Jones WA. Exon Switching and Activation of Stromal and Embryonic Fibroblast Growth Factor (FGF)-FGF Receptor Genes in Prostate Epithelial Cells Accompany Stromal Independence and Malignancy. *Mol Cell Biol*. 1993;Aug:4513–22.
109. Matsuda Y, Ueda J, Ishiwata T. Fibroblast growth factor receptor 2: Expression, roles, and potential as a novel molecular target for colorectal cancer. *Patholog Res Int*. 2012;2012:574768.
110. Liu G, Chen T, Ding Z, Wang Y, Wei Y, Wei X. Inhibition of FGF-FGFR and VEGF-VEGFR signalling in cancer treatment. *Cell Prolif*. 2021 Apr 2;54(4).
111. Cubasch H, Ruff P, Joffe M, Norris S, Chirwa T, Nietz S, et al. South African breast cancer and HIV outcomes study: Methods and baseline assessment. *J Glob Oncol*. 2017;3(2):114–24.
112. Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington MK, et al., editors. *AJCC Cancer Staging Manual*. 8th ed. Springer International Publishing: American Joint Commission on Cancer; 2017.
113. Rosen RD, Sapra A. *TNM Classification*. Treasure Island (FL): StatPearls Publishing ; 2023.
114. Meyer JS, Alvarez C, Milikowski C, Olson N, Russo I, Russo J, et al. Breast carcinoma malignancy grading by Bloom-Richardson system vs proliferation index: Reproducibility of grade and advantages of proliferation index. *Modern Pathology*. 2005;18(8):1067–78.
115. Fitzgibbons PL, Dillon DA, Alsabeh R, Berman MA, Hayes DF, Hicks DG, et al. Template for reporting results of biomarker testing of specimens from patients with carcinoma of the breast. *Arch Pathol Lab Med*. 2014;138(5):595–601.
116. Allred DC. Issues and updates: Evaluating estrogen receptor- $\alpha$ , progesterone receptor, and HER2 in breast cancer. *Modern Pathology* [Internet]. 2010;23(SUPPL. 2):S52–9. Available from: <http://dx.doi.org/10.1038/modpathol.2010.55>
117. Harvey J, Clark G, Osborne C, Allred D. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol*. 1999;17(5):1474–81.

118. Department of Health Republic of South Africa. Clinical Guidelines for Breast Cancer Control and Management [Internet]. 2018 [cited 2022 Aug 23]. p. 1–123. Available from: <https://www.cansa.org.za/files/2019/08/DOH-Breast-Cancer-Guidelines-Final.pdf> Accessed 31/08/2022
119. Bernard PS, Parker JS, Mullins M, Cheung MCU, Leung S, Voduc D, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*. 2009;27(8):1160–7.
120. Dowsett M, Sestak I, Lopez-Knowles E, Sidhu K, Dunbier AK, Cowens JW, et al. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *Journal of Clinical Oncology*. 2013;31(22):2783–90.
121. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 1988;16(3):1215.
122. Griffiths L, Chacon-Cortes D. Methods for extracting genomic DNA from whole blood samples: current perspectives. *Journal of Biorepository Science for Applied Medicine*. 2014;1.
123. Zhang Y, Zeng X, Liu P, Hong R, Lu H, Ji H, et al. Association between FGFR2 (rs2981582, rs2420946 and rs2981578) polymorphism and breast cancer susceptibility: a meta-analysis. *Oncotarget*. 2017;8(2):3454–70.
124. Zhang Y, Lu H, Ji H, Lu L, Liu P, Hong R, et al. Association between rs11200014, rs2981579, and rs1219648 polymorphism and breast cancer susceptibility: A meta-analysis. *Medicine (United States)*. 2017;96(50):9–13.
125. Lewis CM. Genetic association studies: Design, analysis and interpretation. *Brief Bioinform*. 2002;3(2):146–53.
126. Razavi P, Chang MT, Xu G, Bandlamudi C, Ross DS, Vasan N, et al. The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell*. 2018 Sep 10;34(3):427–438.e6.
127. Razavi P, Dickler MN, Shah PD, Toy W, Brown DN, Won HH, et al. Alterations in PTEN and ESR1 promote clinical resistance to alpelisib plus aromatase inhibitors. *Nat Cancer*. 2020 Apr 1;1(4):382–93.
128. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012 Jun 21;486(7403):346–52.
129. Pereira B, Chin SF, Rueda OM, Vollan HKM, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun*. 2016 May 10;7.
130. Rueda OM, Sammut SJ, Seoane JA, Chin SF, Caswell-Jin JL, Callari M, et al. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature*. 2019 Mar 21;567(7748):399–404.

131. Pareja F, Brown DN, Lee JY, Paula ADC, Selenica P, Bi R, et al. Whole-Exome Sequencing Analysis of the Progression from Non–Low-Grade Ductal Carcinoma in Situ to Invasive Ductal Carcinoma. *Clinical Cancer Research*. 2020 Jul 15;26(14):3682–93.
132. Nixon MJ, Formisano L, Mayer IA, Estrada MV, González-Ericsson PI, Isakoff SJ, et al. PIK3CA and MAP3K1 alterations imply luminal A status and are associated with clinical benefit from pan-PI3K inhibitor buparlisib and letrozole in ER+ metastatic breast cancer. *NPJ Breast Cancer*. 2019 Dec 1;5(1).
133. Kan Z, Ding Y, Kim J, Jung HH, Chung W, Lal S, et al. Multi-omics profiling of younger Asian breast cancers reveals distinctive molecular signatures. *Nat Commun*. 2018 Dec 1;9(1).
134. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012 Jun 21;486(7403):395–9.
135. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*. 2012 Jun 21;486(7403):405–9.
136. Stephens PJ, Tarpey PS, Davies H, van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature*. 2012 Jun 21;486(7403):400–4.
137. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*. 2015;163(2):506–19.
138. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, et al. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012 Oct 4;490(7418):61–70.
139. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell*. 2018 Apr 5;173(2):291–304.
140. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C, et al. Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. *Cell Syst*. 2018 Mar 28;6(3):271–81.
141. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, et al. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell*. 2018 Apr 9;33(4):676–89.
142. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell*. 2018 Apr 5;173(2):321–37.
143. Gao Q, Liang WW, Foltz SM, Mutharasu G, Jayasinghe RG, Cao S, et al. Driver Fusions and Their Implications in the Development and Treatment of Human Cancers. *Cell Rep*. 2018 Apr 3;23(1):227–38.

144. Bhandari V, Hoey C, Liu LY, Lalonde E, Ray J, Livingstone J, et al. Molecular landmarks of tumor hypoxia across cancer types. *Nat Genet.* 2019 Feb 1;51(2):308–18.
145. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature.* 2020 Mar 26;579(7800):567–74.
146. Ding L, Bailey MH, Porta-Pardo E, Thorsson V, Colaprico A, Bertrand D, et al. Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. *Cell.* 2018 Apr 5;173(2):305–20.
147. Bonneville R, Krook MA, Kautto EA, Miya J, Wing MR, Chen HZ, et al. Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precis Oncol.* 2017;Epub 2017 Oct 3.
148. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell.* 2018 Apr 5;173(2):400–16.
149. Nicholas K, Nicholas H, Deerfield D. GeneDoc: analysis and visualization of genetic variation. *Embnew News.* 1997;4:14.
150. Sun W, Duan T, Ye P, Chen K, Zhang G, Lai M, et al. TSVdb: A web-tool for TCGA splicing variants analysis. *BMC Genomics.* 2018 May 29;19(1).
151. Allred D, Harvey J, Berardo M, Clark G. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol.* 1998;11(2):155–68.
152. Shousha S. Oestrogen receptor status of breast carcinoma: Allred/H score conversion table. *Histopathology.* 2008;53(3):346–7.
153. Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in Breast Cancer: Recommendations from the international Ki67 in breast cancer working Group. *J Natl Cancer Inst.* 2011;103(22):1656–64.
154. Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington MK, et al., editors. *AJCC Cancer Staging Manual.* 8th ed. Springer International Publishing: American Joint Commission on Cancer; 2017.
155. Harbeck N, Thomssen C, Gnant M. St. Gallen 2013: Brief preliminary summary of the consensus discussion. *Breast Care.* 2013;8(2):102–9.
156. Bastien RRL, Rodríguez-Lescure Á, Ebbert MTW, Prat A, Munárriz B, Rowe L, et al. PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. *BMC Med Genomics.* 2012;5:44.
157. Nielsen T, Wallden B, Schaper C, Ferree S, Liu S, Gao D, et al. Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens. *BMC Cancer.* 2014;14(1):177.

158. Burstein HJ, Curigliano G, Loibl S, Dubsky P, Gnant M, Poortmans P, et al. Estimating the benefits of therapy for early-stage breast cancer: the St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019. *Annals of Oncology*. 2019;30:1541–57.
159. Fernandez-Martinez A, Pascual T, Perrone G, Morales S, de la Haba J, González-Rivera M, et al. Limitations in predicting PAM50 intrinsic subtype and risk of relapse score with Ki67 in estrogen receptor-positive HER2- negative breast cancer. *Oncotarget*. 2017;8(13):21930–7.
160. Pu M, Messer K, Davies SR, Vickery TL, Pittman E, Parker BA, et al. Research-based PAM50 signature and long-term breast cancer survival. *Breast Cancer Res Treat*. 2020;179(1):197–206.
161. Department of Health Republic of South Africa. Breast Cancer Control Policy [Internet]. Vol. June. 2017 [cited 2022 Jan 4]. Available from: <https://www2.tri-kobe.org/nccn/guideline/breast/english/breast.pdf>
162. Paquet ER, Hallett MT. Absolute assignment of breast cancer intrinsic molecular subtype. *J Natl Cancer Inst*. 2015;107(1):1–9.
163. Wiseman RJ, Riddin J, Jugathpal J, Parrish AG, Ruff P, Blockman M. Adjuvant trastuzumab in early HER2-positive breast cancer: Journeying towards the optimal duration of therapy in South Africa. *South African Medical Journal*. 2020;110(4):271–3.
164. Robertson S, Stålhammar G, Darai-Ramqvist E, Rantalainen M, Tobin NP, Bergh J, et al. Prognostic value of Ki67 analysed by cytology or histology in primary breast cancer. *J Clin Pathol*. 2018;71(9):787–94.
165. Lundgren C, Bendahl PO, Borg Å, Ehinger A, Hegardt C, Larsson C, et al. Agreement between molecular subtyping and surrogate subtype classification: a contemporary population-based study of ER-positive/HER2-negative primary breast cancer. *Breast Cancer Res Treat*. 2019;178(2):459–67.
166. Siddharth S, Sharma D. Racial Disparity and Triple-Negative Breast Cancer in African-American Women: A Multifaceted Affair between Obesity, Biology, and Socioeconomic Determinants. *Cancers (Basel)*. 2018 Dec 14;10(12):514.
167. Huo D, Hu H, Rhie SK, Gamazon ER, Cherniack AD, Liu J, et al. Comparison of Breast Cancer Molecular Features and Survival by African and European Ancestry in The Cancer Genome Atlas. *JAMA Oncol*. 2017 Dec;3(12):1654–62.
168. Parada H, Sun X, Fleming JM, Williams-DeVane CLR, Kirk EL, Olsson LT, et al. Race-associated biological differences among luminal A and basal-like breast cancers in the Carolina Breast Cancer Study. *Breast Cancer Research*. 2017;19(1):1–9.
169. Troester MA, Sun X, Allott EH, Geradts J, Cohen SM, Tse CK, et al. Racial Differences in PAM50 Subtypes in the Carolina Breast Cancer Study. *J Natl Cancer Inst*. 2017;110(2):1–7.

170. Ping J, Guo X, Ye F, Long J, Lipworth L, Cai Q, et al. Differences in gene-expression profiles in breast cancer between African and European-ancestry women. *Carcinogenesis*. 2021;41(7):887–93.
171. Nwagu GC, Bhattarai S, Swahn M, Ahmed S, Aneja R. Prevalence and Mortality of Triple-Negative Breast Cancer in West Africa: Biologic and Sociocultural Factors. *JCO Glob Oncol*. 2021;(7):1129–40.
172. Sung H, DeSantis CE, Fedewa SA, Kantelhardt EJ, Jemal A. Breast cancer subtypes among Eastern-African-born black women and other black women in the United States. *Cancer*. 2019;125(19):3401–11.
173. Awadelkarim KD, Elhaj A, Aceto G, Mariani-Costantini R, Abdalla Eltayeb E. Hereditary Breast Cancer in Sub-Saharan Africa. *Curr Womens Health Rev*. 2012;8(1):44–54.
174. Sayed S, Moloo Z, Wasike R, Bird P, Oigara R, Govender D, et al. Is breast cancer from Sub Saharan Africa truly receptor poor? Prevalence of ER/PR/HER2 in breast cancer from Kenya. *Breast*. 2014;23(5):591–6.
175. Popli P, Gutterman EM, Omene C, Ganesan S, Mills D, Marlink R. Receptor-Defined Breast Cancer in Five East African Countries and Its Implications for Treatment: Systematic Review and Meta-Analysis. *JCO Glob Oncol*. 2021;(7):289–301.
176. Zuma K, Simbayi L, Zungu N, Moyo S, Marinda E, Jooste S, et al. The HIV Epidemic in South Africa: Key Findings from 2017 National Population-Based Survey. *Int J Environ Res Public Health*. 2022 Jul 1;19(13):8125.
177. Ayeni OA, Norris SA, Joffe M, Cubasch H, Nietz S, Buccimazza I, et al. The multimorbidity profile of South African women newly diagnosed with breast cancer. *Int J Cancer*. 2020 Jul 15;147(2):361–74.
178. Kim HK, Park KH, Kim Y, Park SE, Lee HS, Lim SW, et al. Discordance of the PAM50 intrinsic subtypes compared with immunohistochemistry-based surrogate in breast cancer patients: Potential implication of genomic alterations of discordance. *Cancer Res Treat*. 2019;51(2):737–47.
179. Vanderpuye V, Grover S, Hammad N, Prabhakar P, Simonds H, Olopade F, et al. An update on the management of breast cancer in Africa. *Infect Agent Cancer*. 2017;12(1):1–12.
180. Galukande M, Wabinga H, Mirembe F, Karamagi C, Asea A. Molecular breast cancer subtypes prevalence in an indigenous Sub Saharan African population. *Pan Afr Med J*. 2014;17:249.
181. Adebamowo CA, Famooto A, Ogundiran TO, Aniagwu T, Nkwodimmah C, Akang EE. Immunohistochemical and molecular subtypes of breast cancer in Nigeria. *Breast Cancer Res Treat*. 2008;110(1):183–8.
182. Adeniji AA, Dawodu OO, Habeebu MY, Oyekan AO, Bashir MA, Martin MG, et al. Distribution of Breast Cancer Subtypes Among Nigerian Women and Correlation to the Risk Factors and Clinicopathological Characteristics. *World J Oncol*. 2020;11(4):165–72.

183. Brandão M, Guisseve A, Bata G, Alberto M, Ferro J, Garcia C, et al. Breast cancer subtypes: Implications for the treatment and survival of patients in Africa - A prospective cohort study from Mozambique. *ESMO Open*. 2020;5(5):1–13.
184. Miguel F, Lopes LV, Ferreira E, Ribas E, Pelaez AF, Leal C, et al. Breast cancer in Angola, molecular subtypes: A first glance. *Ecancer Medicalscience*. 2017;11:1–10.
185. Elmore SNC, Mushonga M, Iyer HS, Kanda C, Chibonda S, Chipidza F, et al. Breast cancer in Zimbabwe: patterns of care and correlates of adherence in a national referral hospital radiotherapy center cohort from 2014 to 2018. *Cancer Med*. 2021;10(11):3489–98.
186. Wallden B, Storhoff J, Nielsen T, Dowidar N, Schaper C, Ferree S, et al. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med Genomics*. 2015;8(1):1–14.
187. Ochoa S, de Anda-Jáuregui G, Hernández-Lemus E. Multi-Omic Regulation of the PAM50 Gene Signature in Breast Cancer Molecular Subtypes. *Front Oncol*. 2020;10(May):1–14.
188. Fachal L, Aschard H, Beesley J, Barnes DR, Allen J, Kar S, et al. Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nat Genet*. 2020 Jan 1;52(1):56–73.
189. Agarwal D, Pineda S, Michailidou K, Herranz J, Pita G, Moreno LT, et al. FGF receptor genes and breast cancer susceptibility: Results from the Breast Cancer Association Consortium. *Br J Cancer*. 2014;110(4):1088–100.
190. Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A, et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat Genet*. 2021 Nov 1;53(11):1527–33.
191. Udler MS, Meyer KB, Pooley KA, Karlins E, Struewing JP, Zhang J, et al. FGFR2 variants and breast cancer risk: Fine-scale mapping using African American studies and analysis of chromatin conformation. 2009;1:Supplementary.
192. Hodoğlugil U, Mahley RW. Turkish Population Structure and Genetic Ancestry Reveal Relatedness among Eurasian Populations. *Ann Hum Genet*. 2012 Mar;76(2):128–41.
193. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, et al. The genetic structure and history of Africans and African Americans. *Science* (1979). 2009 May 22;324(5930):1035–44.
194. Kasembeli AN, Duarte R, Ramsay M, Mosiane P, Dickens C, Dix-Peek T, et al. APOL1 Risk Variants Are Strongly Associated with HIV-Associated Nephropathy in Black South Africans. *J Am Soc Nephrol [Internet]*. 2015;2882–90. Available from: <http://jasn.asnjournals.org.beckerproxy.wustl.edu/content/early/2015/03/17/ASN.2014050469.abstract>
195. Fitzgibbons PL, Dillon DA, Alsabeh R, Berman MA, Hayes DF, Hicks DG, et al. Template for reporting results of biomarker testing of specimens from patients with carcinoma of the breast. *Arch Pathol Lab Med*. 2014;138(5):595–601.

196. Stender JD, Frasor J, Komm B, Chang KCN, Kraus WL, Katzenellenbogen BS. Estrogen-regulated gene networks in human breast cancer cells: Involvement of E2F1 in the regulation of cell proliferation. *Molecular Endocrinology*. 2007 Sep;21(9):2112–23.
197. Fernández-Nogueira P, Mancino M, Fuster G, López-Plana A, Jauregui P, Almendro V, et al. Tumor-associated fibroblasts promote HER2-targeted therapy resistance through FGFR2 activation. *Clinical Cancer Research*. 2020;26(6):1432–48.
198. Hanker AB, Garrett JT, Estrada MV, Moore PD, Ericsson PG, Koch JP, et al. HER2-overexpressing breast cancers amplify FGFR signaling upon acquisition of resistance to dual therapeutic blockade of HER2. *Clinical Cancer Research*. 2017 Aug 1;23(15):4323–34.
199. Achilonu OJ, Singh E, Nimako G, Eijkemans RMJC, Musenge E. Rule-Based Information Extraction from Free-Text Pathology Reports Reveals Trends in South African Female Breast Cancer Molecular Subtypes and Ki67 Expression. *Biomed Res Int*. 2022;2022:1–17.
200. Tsang JYS, Tse GM. Molecular Classification of Breast Cancer. *Adv Anat Pathol*. 2020;27(1):27–35.
201. Lyu J, Cheng C. Regulation of Alternative Splicing during Epithelial-Mesenchymal Transition. *Cells Tissues Organs*. 2022 Mar 1;211(2):238–51.
202. de Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer*. 2013;13(2).
203. Wagner EJ, Curtis ML, Robson ND, Baraniak AP, Eis PS, Garcia-Blanco MA. Quantification of alternatively spliced FGFR2 RNAs using the RNA invasive cleavage assay. *RNA*. 2003 Dec;9(12):1552–61.
204. Xie Y, Su N, Yang J, Tan Q, Huang S, Jin M, et al. FGF/FGFR signaling in health and disease. *Signal Transduct Target Ther*. 2020;5(1).
205. Ferguson HR, Smith MP, Francavilla C. Fibroblast growth factor receptors (Fgfrs) and noncanonical partners in cancer signaling. *Cells*. 2021;10(5):1–35.
206. Zugman M, Botrus G, Pestana RC, Uson Junior PLS. Precision Medicine Targeting FGFR2 Genomic Alterations in Advanced Cholangiocarcinoma: Current State and Future Perspectives. *Front Oncol*. 2022 Apr 4;12:860453.
207. Ranieri D, Rosato B, Nanni M, Magenta A, Belleudi F, Torrì MR. Expression of the FGFR2 mesenchymal splicing variant in epithelial cells drives epithelial-mesenchymal transition. *Oncotarget*. 2015;7(5):5440–60.
208. Adamczyk-Gruszka O, Horecka-Lewitowicz A, Gruszka J, Wawszczak-Kasza M, Strzelecka A, Lewitowicz P. FGFR-2 and Epithelial–Mesenchymal Transition in Endometrial Cancer. *J Clin Med*. 2022 Sep 1;11(18).

209. Braun M, Piasecka D, Tomasik B, Mieczkowski K, Stawiski K, Zielinska A, et al. Hormonal receptor status determines prognostic significance of FGFR2 in invasive breast carcinoma. *Cancers (Basel)*. 2020;12(9):1–18.
210. Piasecka D, Braun M, Kitowska K, Mieczkowski K, Kordek R, Sadej R, et al. FGFs/FGFRs-dependent signalling in regulation of steroid hormone receptors - Implications for therapy of luminal breast cancer. *Journal of Experimental and Clinical Cancer Research*. 2019 May 29;38:230.
211. Russo J, Russo IH. Development of the human breast. *Maturitas*. 2004 Sep 24;49(1):2–15.
212. Mao P, Cohen O, Kowalski KJ, Kusiel JG, Buendia-Buendia JE, Cuoco MS, et al. Acquired FGFR and FGF alterations confer resistance to Estrogen Receptor (ER) targeted therapy in ER $\beta$  metastatic breast cancer. *Clinical Cancer Research*. 2020 Nov 15;26(22):5974–89.
213. Mieczkowski K, Kitowska K, Braun M, Galikowska-Bogut B, Gorska-Arcisz M, Piasecka D, et al. FGF7/FGFR2–JunB signalling counteracts the effect of progesterone in luminal breast cancer. *Mol Oncol*. 2022 Aug 1;16(15):2823–42.
214. Azuma K, Tsurutani J, Sakai K, Kaneda H, Fujisaka Y, Takeda M, et al. Switching addictions between HER2 and FGFR2 in HER2-positive breast tumor cells: FGFR2 as a potential target for salvage after lapatinib failure. *Biochem Biophys Res Commun*. 2011 Apr 1;407(1):219–24.
215. Santolla MF, Maggiolini M. The FGF/FGFR system in breast cancer: Oncogenic features and therapeutic perspectives. *Cancers (Basel)*. 2020 Oct 1;12(10):1–20.
216. Issa A, Gill JW, Heideman MR, Sahin O, Wiemann S, Dey JH, et al. Combinatorial targeting of FGF and ErbB receptors blocks growth and metastatic spread of breast cancer models. *Breast Cancer Research*. 2013 Jan 23;15(1).
217. Turner N, Lambros MB, Horlings HM, Pearson A, Sharpe R, Natrajan R, et al. Integrative molecular profiling of triple negative breast cancers identifies amplicon drivers and potential therapeutic targets. *Oncogene*. 2010 Apr;29(14):2013–23.
218. Zhao Q, Parris AB, Howard EW, Zhao M, Ma Z, Guo Z, et al. FGFR inhibitor, AZD4547, impedes the stemness of mammary epithelial cells in the premalignant tissues of MMTV-ErbB2 transgenic mice. *Sci Rep*. 2017 Dec 1;7(1).
219. Xiang H, Chan AG, Ahene A, Bellovin DI, Deng R, Hsu AW, et al. Preclinical characterization of bemarituzumab, an anti-FGFR2b antibody for the treatment of cancer. *MAbs*. 2021 Jan 1;13(1).
220. Turczyk L, Kitowska K, Mieszkowska M, Mieczkowski K, Czaplinska D, Piasecka D, et al. FGFR2-Driven Signaling Counteracts Tamoxifen Effect on ER $\alpha$ -Positive Breast Cancer Cells. *Neoplasia*. 2017;19(10):791–804.
221. Jones RB, Wang F, Luo Y, Yu C, Jin C, Suzuki T, et al. The Nonsense-mediated Decay Pathway and Mutually Exclusive Expression of Alternatively Spliced FGFR2IIIb and -IIIc mRNAs. *Journal of Biological Chemistry*. 2001 Feb 9;276(6):4158–67.

222. Ishiwata T. Role of fibroblast growth factor receptor-2 splicing in normal and cancer cells. *Frontiers In Bioscience, Landmark*. 2018;626–39.
223. Yeh BK, Igarashi M, Eliseenkova A V, Plotnikov AN, Sher I, Ron D, et al. Structural basis by which alternative splicing confers specificity in fibroblast growth factor receptors. *PNAS*. 2003;100(5):2266–71.
224. Reintjes N, Li Y, Becker A, Rohmann E, Schmutzler R, Wollnik B. Activating Somatic FGFR2 Mutations in Breast Cancer. *PLoS One*. 2013;8(3).
225. Yashiro M, Kuroda K, Masuda G, Okuno T, Miki Y, Yamamoto Y, et al. Clinical difference between fibroblast growth factor receptor 2 subclass, type IIIb and type IIIc, in gastric cancer. *Sci Rep*. 2021 Dec 1;11(1).
226. Lei JH, Lee MH, Miao K, Huang Z, Yao Z, Zhang A, et al. Activation of FGFR2 Signaling Suppresses BRCA1 and Drives Triple-Negative Mammary Tumorigenesis That is Sensitive to Immunotherapy. *Advanced Science*. 2021;8(21):1–17.
227. Azubuike SO, Muirhead C, Hayes L, McNally R. Rising global burden of breast cancer: The case of sub-Saharan Africa (with emphasis on Nigeria) and implications for regional development: A review. *World J Surg Oncol*. 2018 Mar 22;16:63.
228. Saberian M, Mehrabani K, Shahraki HR. Clustering time trends of breast cancer incidence in Africa: A 27-year longitudinal study in 53 countries. *Afr Health Sci*. 2021;21(1):47–53.
229. Foerster M, Anderson BO, McKenzie F, Galukande M, Anele A, Adisa C, et al. Inequities in breast cancer treatment in sub-Saharan Africa: Findings from a prospective multi-country observational study. *Breast Cancer Research*. 2019;21(1):1–11.
230. Foerster M, McKenzie F, Zietsman A, Galukande M, Anele A, Adisa C, et al. Dissecting the journey to breast cancer diagnosis in sub-Saharan Africa: Findings from the multicountry ABC-DO cohort study. *Int J Cancer*. 2021 Jan 15;148(2):340–51.
231. McCormack V, McKenzie F, Foerster M, Zietsman A, Galukande M, Adisa C, et al. Breast cancer survival and survival gap apportionment in sub-Saharan Africa (ABC-DO): a prospective cohort study. *Lancet Glob Health*. 2020;8(9):e1203–12.
232. Sun Y, Li G, Zhu W, He Q, Liu Y, Chen X, et al. A comprehensive pan-cancer study of fibroblast growth factor receptor aberrations in Chinese cancer patients. *Ann Transl Med*. 2020;8(20):1290–1290.
233. Jensen MB, Lænkholm AV, Balslev E, Buckingham W, Ferree S, Glavicic V, et al. The Prosigna 50-gene profile and responsiveness to adjuvant anthracycline-based chemotherapy in high-risk breast cancer patients. *NPJ Breast Cancer*. 2020;6(1).
234. Sinha S, Bhatia R, Narasimamurthy M, Rayne S, Grover S. Epidemiology of Breast Cancer Presentation in Botswana, South Africa, and the United States. *Journal of Surgical Research*. 2022 Nov 1;279:533–9.

235. Mabaso M, Makola L, Naidoo I, Mlangeni LL, Jooste S, Simbayi L. HIV prevalence in South Africa through gender and racial lenses: results from the 2012 population-based national household survey. *Int J Equity Health*. 2019 Dec 30;18(1):167.
236. Shisana O, Zungu N, Evans M, Risher K, Rehle T, Clementano D. The case for expanding the definition of 'key populations' to include high-risk groups in the general population to improve targeted HIV prevention efforts. *South African Medical Journal*. 2015 Sep 22;105(8):664.
237. UNAIDS. HIV and AIDS Estimates South Africa 2022 [Internet]. [cited 2023 Oct 31]. Available from: <https://www.unaids.org/en/regionscountries/countries/southafrica>
238. Phakathi B, Dix-Peek T, Van Den Berg E, Dickens C, Nietz S, Cubasch H, et al. PAM50 intrinsic subtypes, risk of recurrence score and breast cancer survival in HIV-positive and HIV-negative patients—a South African cohort study. *Breast Cancer Res Treat*. 2023 Aug 2;200(3):337–46.
239. Chang J, Liu X, Wang S, Zhang Z, Wu Z, Zhang X, et al. Prognostic value of FGFR gene amplification in patients with different types of cancer: A systematic review and meta-analysis. *PLoS One*. 2014;9(8).
240. Mongre RK, Mishra CB, Shukla AK, Prakash A, Jung S, Ashraf-Uz-zaman M, et al. Emerging importance of tyrosine kinase inhibitors against cancer: Quo vadis to cure? *Int J Mol Sci*. 2021 Nov 1;22:11659.

APPENDICES

**Appendix A: Coauthors' agreement for papers to be used in thesis****Declaration: Student's contribution to article(s) and agreement of co-author(s)**

I, Thérèse Dix-Peek, student number 9603960G, declare that this Thesis is my own work and that I contributed adequately towards research findings published in the article(s) stated below which are included in my Thesis.

Signature of Student ...  ..... Date.....05 November 2023.....

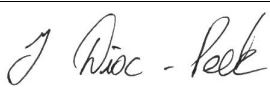
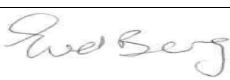
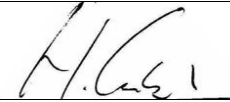
Name of Primary Supervisor..... Raquel Duarte

Signature of Primary Supervisor .....  ..... Date.....15/11/23

**Agreement by co-authors:** By signing this declaration, the co-authors listed below agree to the use of the article(s) by the student as part of his/her Thesis/Dissertation/Research Report. In cases where the student is not the 1<sup>st</sup> author of a published article, the primary supervisor must explain (under comments) why the student is entitled to use the paper for his/her degree purposes.

**Article 1: Discordance between PAM50 intrinsic subtyping and immunohistochemistry in South African women with breast cancer**

Breast Cancer Research and Treatment (2023) May;199(1):1-12

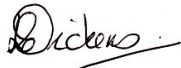
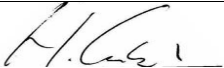
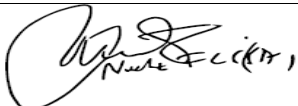
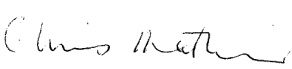
| Authors                | Name                   | Signature                                                                            | Date        |
|------------------------|------------------------|--------------------------------------------------------------------------------------|-------------|
| 1 <sup>st</sup> author | Thérèse Dix-Peek       |  | 11 Oct 23   |
| 2 <sup>nd</sup> author | Boitumelo P. Phakathi  |                                                                                      |             |
| 3 <sup>rd</sup> author | Eunice J. van den Berg |  | 11 Oct 23   |
| 4 <sup>th</sup> author | Caroline Dickens       |   | 23 Oct 23   |
| 5 <sup>th</sup> author | Tanya N. Augustine     |  | 15 Nov 2023 |
| 6 <sup>th</sup> author | Herbert Cubasch        |  | 11 Oct 23   |
| 7 <sup>th</sup> author | Alfred I. Neugut       |                                                                                      |             |
| 8 <sup>th</sup> author | Judith S. Jacobson     |                                                                                      |             |
| 9 <sup>th</sup> author | Maureen Joffe          |  | 11 Oct 23   |

|                         |                    |                                                                                    |           |
|-------------------------|--------------------|------------------------------------------------------------------------------------|-----------|
| 10 <sup>th</sup> author | Paul Ruff          |  | 11 Oct 23 |
| 11 <sup>th</sup> author | Raquel A.B. Duarte |  | 15/11/23  |

**Comments by primary supervisor:**

.....  
 ...

**Article 2: FGFR2 genetic variants in women with breast cancer****Molecular Medicine Reports (2023) 28: 22**

| Authors                 | Name                   | Signature                                                                            | Date      |
|-------------------------|------------------------|--------------------------------------------------------------------------------------|-----------|
| 1 <sup>st</sup> author  | Thérèse Dix-Peek       |    | 11 Oct 23 |
| 2 <sup>nd</sup> author  | Caroline Dickens       |     | 23 Oct 23 |
| 3 <sup>rd</sup> author  | Tanya N. Augustine     |    | 15 Nov 23 |
| 4 <sup>th</sup> author  | Boitumelo P. Phakathi  |                                                                                      |           |
| 5 <sup>th</sup> author  | Eunice J. van den Berg |    | 11 Oct 23 |
| 6 <sup>th</sup> author  | Maureen Joffe          |    | 11 Oct 23 |
| 7 <sup>th</sup> author  | Oluwatosin A. Ayeni    |    | 11 Oct 23 |
| 8 <sup>th</sup> author  | Herbert Cubasch        |    | 11 oct 23 |
| 9 <sup>th</sup> author  | Sarah Nietz            |   | 11 Oct 23 |
| 10 <sup>th</sup> author | Christopher G. Mathew  |  |           |
| 11 <sup>th</sup> author | Mahtaab Hayat          |                                                                                      |           |
| 12 <sup>th</sup> author | Alfred I. Neugut       |                                                                                      |           |
| 13 <sup>th</sup> author | Judith S. Jacobson     |                                                                                      |           |
| 14 <sup>th</sup> author | Paul Ruff              |  | 11 Oct 23 |
| 15 <sup>th</sup> author | Raquel A.B. Duarte     |  | 15/11/23  |

**Comments by primary supervisor:**

.....

.....

.....

.....

**Declaration: Student's contribution to article(s) and agreement of co-author(s)**

I, Thérèse Dix-Peek, student number 9603960G, declare that this Thesis is my own work and that I contributed adequately towards research findings published in the article(s) stated below which are included in my Thesis.

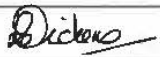
Signature of Student ..... Date.....

Name of Primary Supervisor.....

Signature of Primary Supervisor ..... Date.....

**Agreement by co-authors:** By signing this declaration, the co-authors listed below agree to the use of the article(s) by the student as part of his/her Thesis/Dissertation/Research Report. In cases where the student is not the 1<sup>st</sup> author of a published article, the primary supervisor must explain (under comments) why the student is entitled to use the paper for his/her degree purposes.

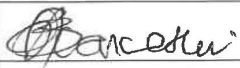
**Article 1: Discordance between PAM50 intrinsic subtyping and immunohistochemistry in South African women with breast cancer**  
**Breast Cancer Research and Treatment (2023) May;199(1):1-12**

| Authors                 | Name                   | Signature                                                                            | Date        |
|-------------------------|------------------------|--------------------------------------------------------------------------------------|-------------|
| 1 <sup>st</sup> author  | Thérèse Dix-Peek       |   | 11 Oct 23   |
| 2 <sup>nd</sup> author  | Boitumelo P. Phakathi  |  | 25/10/23    |
| 3 <sup>rd</sup> author  | Eunice J. van den Berg |                                                                                      |             |
| 4 <sup>th</sup> author  | Caroline Dickens       |  | 23 Oct 23   |
| 5 <sup>th</sup> author  | Tanya N. Augustine     |  | 15 Nov 2023 |
| 6 <sup>th</sup> author  | Herbert Cubasch        |                                                                                      |             |
| 7 <sup>th</sup> author  | Alfred I. Neugut       |                                                                                      |             |
| 8 <sup>th</sup> author  | Judith S. Jacobson     |                                                                                      |             |
| 9 <sup>th</sup> author  | Maureen Joffe          |                                                                                      |             |
| 10 <sup>th</sup> author | Paul Ruff              |                                                                                      |             |
| 11 <sup>th</sup> author | Raquel A.B. Duarte     |                                                                                      | 15/11/23    |

**Comments by primary supervisor:**

.....  
 .....

**Article 2: FGFR2 genetic variants in women with breast cancer**  
**Molecular Medicine Reports (2023) 28: 22**

| Authors                 | Name                   | Signature                                                                          | Date      |
|-------------------------|------------------------|------------------------------------------------------------------------------------|-----------|
| 1 <sup>st</sup> author  | Thérèse Dix-Peek       |                                                                                    | 11 Oct 23 |
| 2 <sup>nd</sup> author  | Caroline Dickens       |                                                                                    |           |
| 3 <sup>rd</sup> author  | Tanya N. Augustine     |  | 15 Nov 23 |
| 4 <sup>th</sup> author  | Boitumelo P. Phakathi  |  | 25/11/23  |
| 5 <sup>th</sup> author  | Eunice J. van den Berg |                                                                                    |           |
| 6 <sup>th</sup> author  | Maureen Joffe          |                                                                                    |           |
| 7 <sup>th</sup> author  | Oluwatosin A. Ayeni    |                                                                                    |           |
| 8 <sup>th</sup> author  | Herbert Cubasch        |                                                                                    |           |
| 9 <sup>th</sup> author  | Sarah Nietz            |                                                                                    |           |
| 10 <sup>th</sup> author | Christopher G. Mathew  |                                                                                    |           |
| 11 <sup>th</sup> author | Mahtaab Hayat          |                                                                                    |           |
| 12 <sup>th</sup> author | Alfred I. Neugut       |                                                                                    |           |
| 13 <sup>th</sup> author | Judith S. Jacobson     |                                                                                    |           |
| 14 <sup>th</sup> author | Paul Ruff              |                                                                                    |           |
| 15 <sup>th</sup> author | Raquel A.B. Duarte     |                                                                                    |           |

**Comments by primary supervisor:**

.....

.....



**Declaration: Student's contribution to article(s) and agreement of co-author(s)**

I, Thérèse Dix-Peek, student number 9603960G, declare that this Thesis is my own work and that I contributed adequately towards research findings published in the article(s) stated below which are included in my Thesis.

Signature of Student ..... Date.....

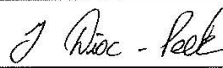
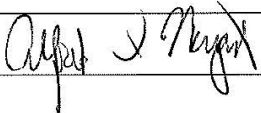
Name of Primary Supervisor.....

Signature of Primary Supervisor ..... Date.....

**Agreement by co-authors:** By signing this declaration, the co-authors listed below agree to the use of the article(s) by the student as part of his/her Thesis/Dissertation/Research Report. In cases where the student is not the 1<sup>st</sup> author of a published article, the primary supervisor must explain (under comments) why the student is entitled to use the paper for his/her degree purposes.

**Article 1: Discordance between PAM50 intrinsic subtyping and immunohistochemistry in South African women with breast cancer**

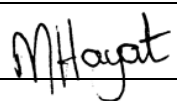
Breast Cancer Research and Treatment (2023) May;199(1):1-12

| Authors                 | Name                   | Signature                                                                             | Date       |
|-------------------------|------------------------|---------------------------------------------------------------------------------------|------------|
| 1 <sup>st</sup> author  | Thérèse Dix-Peek       |     | 11 Oct 23  |
| 2 <sup>nd</sup> author  | Boitumelo P. Phakathi  |                                                                                       |            |
| 3 <sup>rd</sup> author  | Eunice J. van den Berg |                                                                                       |            |
| 4 <sup>th</sup> author  | Caroline Dickens       |                                                                                       |            |
| 5 <sup>th</sup> author  | Tanya N. Augustine     |   | 15 Nov 23  |
| 6 <sup>th</sup> author  | Herbert Cubasch        |                                                                                       |            |
| 7 <sup>th</sup> author  | Alfred J. Neugut       |   | 11/10/2023 |
| 8 <sup>th</sup> author  | Judith S. Jacobson     |                                                                                       |            |
| 9 <sup>th</sup> author  | Maureen Joffe          |                                                                                       |            |
| 10 <sup>th</sup> author | Paul Ruff              |                                                                                       |            |
| 11 <sup>th</sup> author | Raquel A.B. Duarte     |  | 15/11/23   |

**Comments by primary supervisor:**

.....  
 .....

**Article 2: FGFR2 genetic variants in women with breast cancer****Molecular Medicine Reports (2023) 28: 22**

| Authors                 | Name                   | Signature                                                                            | Date      |
|-------------------------|------------------------|--------------------------------------------------------------------------------------|-----------|
| 1 <sup>st</sup> author  | Thérèse Dix-Peek       |    | 11 Oct 23 |
| 2 <sup>nd</sup> author  | Caroline Dickens       |                                                                                      |           |
| 3 <sup>rd</sup> author  | Tanya N. Augustine     |    | 15 Nov 23 |
| 4 <sup>th</sup> author  | Boitumelo P. Phakathi  |                                                                                      |           |
| 5 <sup>th</sup> author  | Eunice J. van den Berg |                                                                                      |           |
| 6 <sup>th</sup> author  | Maureen Joffe          |                                                                                      |           |
| 7 <sup>th</sup> author  | Oluwatosin A. Ayeni    |                                                                                      |           |
| 8 <sup>th</sup> author  | Herbert Cubasch        |                                                                                      |           |
| 9 <sup>th</sup> author  | Sarah Nietz            |                                                                                      |           |
| 10 <sup>th</sup> author | Christopher G. Mathew  |                                                                                      |           |
| 11 <sup>th</sup> author | Mahtaab Hayat          |   | 11 Oct 23 |
| 12 <sup>th</sup> author | Alfred I. Neugut       |                                                                                      |           |
| 13 <sup>th</sup> author | Judith S. Jacobson     |                                                                                      |           |
| 14 <sup>th</sup> author | Paul Ruff              |                                                                                      |           |
| 15 <sup>th</sup> author | Raquel A.B. Duarte     |  |           |

**Comments by primary supervisor:**

.....

.....

.....

.....

## Appendix B: Permission from Prof S Tishkoff to use Fig 1.1

**Sent:** Monday, 20 March 2023 22:30  
**To:** Therese Dix-Peek <Therese.Dix-Peek@wits.ac.za>  
**Subject:** RE: Permission to use a published figure for a PhD Dissertation

Hi Therese,

That's fine with me as long as you cite it.

Best,

Sarah Tishkoff

Sarah Tishkoff, Ph.D.  
David and Lyn Silfen University Professor  
Departments of Genetics and Biology  
University of Pennsylvania  
Tel: 215-746-2670  
[tishkoff@pennmedicine.upenn.edu](mailto:tishkoff@pennmedicine.upenn.edu)  
<http://www.med.upenn.edu/tishkoff/>

Director, Center for Global Genomics & Health Equity

<https://globalgenomics.med.upenn.edu/index.html>



**From:** Therese Dix-Peek <[Therese.Dix-Peek@wits.ac.za](mailto:Therese.Dix-Peek@wits.ac.za)>  
**Sent:** Monday, March 20, 2023 7:49 AM  
**Subject:** [External] Permission to use a published figure for a PhD Dissertation

Dear Prof Tishkoff

My name is Therese Dix-Peek and I am writing to ask if I can use a figure that you published for your paper:

Reed FA, Tishkoff SA. 2006. African human diversity, origins and migrations. *Curr. Opin. Genet. Dev.* 16: 597–6

I am specifically interested in Figure 2, which shows the movement of people from Western Africa to Eastern and Southern Africa, as well as the movement of people to the New World.

The work will be properly cited, both in text and in the references. This figure will be shown in the introduction of the dissertation, and will be available online as a dissertation, but will not be otherwise republished.

The copyright clearance center, which was recommended by the journal returns ownership of the figures to the author and journal (attached below).

Please may I reproduce your article?

Many thanks

Therese

---



Participant Number

\_\_\_\_ BC \_\_\_\_

**Storage of specimens in the study entitled :** Association of Genetic Variants with Breast Cancer Intrinsic Subtypes and Splice Variants of the Fibroblast Growth Factor Receptor 2 in a South African Population

I \_\_\_\_\_, hereby confirm that I have been informed by the study interviewer about the nature, conduct, benefits and risks of the study.

- I understand that my samples will be stored for further studies into breast cancer.
- I give permission for further tests (such as gene sequencing, protein expression) is done on my samples with permission from the ethics committee.

**PARTICIPANT**

\_\_\_\_\_  
Printed name

\_\_\_\_\_  
Signature/mark/thumbprint

\_\_\_\_\_  
Date

**INTERVIEWER**

I herewith confirm that the above participant has been fully informed about the nature, conduct and risks of the above study

\_\_\_\_\_  
Printed Name

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

**WITNESS STATEMENT:**

As an impartial third party, I witnessed the entire consent discussion and the participant's signature on the form. I attest that this entire form was read to the participant named above. This person had enough time to consider this information, had an opportunity to ask questions, and voluntarily agreed to be in this study.

\_\_\_\_\_  
Printed Name

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

## Appendix C: Ethics Clearance Certificate: M161116



R14/49 Ms Thérèse Dix-Peek et al

**HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)****CLEARANCE CERTIFICATE NO. M161116**

**NAME:** Ms Thérèse Dix-Peek et al  
**(Principal Investigator)**  
**DEPARTMENT:** Internal Medicine  
**PROJECT TITLE:** Association of Genetic Variants with Breast Cancer  
 Intrinsic Subtypes and Splice Variants of the  
 Fibroblast Growth Factor Receptor 2 in a  
 South African Population  
**DATE CONSIDERED:** 25/11/2016  
**DECISION:** Approved unconditionally  
**CONDITIONS:**  
**SUPERVISOR:** Dr Raquel Duarte and Dr Tanya Augustine  
**APPROVED BY:**   
 Professor A Dhai, Co-Chairperson, HREC (Medical)  
**DATE OF APPROVAL:** 20/12/2016

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

**DECLARATION OF INVESTIGATORS**

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report.** The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in November and will therefore be due in the month of November each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

UNIVERSITY OF THE  
WITWATERSRAND,  
JOHANNESBURG



HUMAN RESEARCH ETHICS COMMITTEE  
(MEDICAL)

17 April 2019

**Ms Therese Dix-Peek**

Department of Internal Medicine  
University of the Witwatersrand  
Medical School  
Parktown

Sent by email to: [therese.dix-peek@wits.ac.za](mailto:therese.dix-peek@wits.ac.za)

Dear Ms Dix-Peek

**Re: Protocol Ref no: M161116**

**Protocol Title:** Association of Genetic Variants with Breast Cancer Intrinsic Subtypes and Splice Variants of the Fibroblast Growth Factor Receptor 2 in a South African Population

**Principal Investigator(s):** Ms Therese Dix-Peek

**Protocol Amendment:** Request to collect blood from the remaining 541 controls, pilot protein extraction, and immunohistochemistry studies

This letter serves to confirm that the Chairman of the Human Research Ethics Committee (Medical) has approved the amendments for the abovementioned protocol, as detailed in your letter, dated 15 February 2019.

The following documents were received:

- Summary Letter
- HREC (Medical) Application Form
- Clearance Certificates for M161116, M150351, M150347 and M140980
- Study Information Sheet – Controls

Thank you for keeping us informed and updated.

Yours Sincerely,

**Miss Mapula Ramaila**  
**Administrative Officer**  
**Human Research Ethics Committee (Medical)**



**Appendix D: Informed Consent**

Participant Number

\_\_\_\_ BC \_\_\_\_

**Informed Consent for participation in the study entitled :** Association of Genetic Variants with Breast Cancer Intrinsic Subtypes and Splice Variants of the Fibroblast Growth Factor Receptor 2 in a South African Population

I \_\_\_\_\_, hereby confirm that I have been informed by the study interviewer about the nature, conduct, benefits and risks of the study.

- I have received, read and understood the above written information (Participant Information Leaflet and Informed Consent) regarding the study.
- I am aware that the results of the study, including personal details regarding my sex, age, date of birth, initials and diagnosis will be anonymously processed into any research reports.
- I agree for samples of my blood to be used in this study.
- In view of the requirements of research, I agree that the data collected during this study can be processed in a computerised system by the study team.
- I may, at any stage, without prejudice, withdraw my consent and participation in the study.
- I have had sufficient opportunity to ask questions and (of my own free will) declare myself prepared to participate in the study.
- I am aware that the results of the study, including my personal details and diagnosis will be kept confidential.
- I may, at any stage, without prejudice to myself and my future medical treatment, withdraw my consent to participate in this study. I have had sufficient opportunity to ask questions and hereby accept to participate in the study.
- I agree to all of the procedures and restrictions included in this Information Sheet.

**PARTICIPANT**\_\_\_\_\_  
Printed name\_\_\_\_\_  
Signature/mark/thumbprint\_\_\_\_\_  
Date**INTERVIEWER**

I herewith confirm that the above participant has been fully informed about the nature, conduct and risks of the above study

\_\_\_\_\_  
Printed Name\_\_\_\_\_  
Signature\_\_\_\_\_  
Date**WITNESS STATEMENT:**

As an impartial third party, I witnessed the entire consent discussion and the participant's signature on the form. I attest that this entire form was read to the participant named above. This person had enough time to consider this information, had an opportunity to ask questions, and voluntarily agreed to be in this study.

\_\_\_\_\_  
Printed Name\_\_\_\_\_  
Signature\_\_\_\_\_  
Date



Participant Number

\_\_\_\_ BC \_\_\_\_

**Storage of specimens in the study entitled :** Association of Genetic Variants with Breast Cancer Intrinsic Subtypes and Splice Variants of the Fibroblast Growth Factor Receptor 2 in a South African Population

I \_\_\_\_\_, hereby confirm that I have been informed by the study interviewer about the nature, conduct, benefits and risks of the study.

- I understand that my samples will be stored for further studies into breast cancer.
- I give permission for further tests (such as gene sequencing, protein expression) is done on my samples with permission from the ethics committee.

## PARTICIPANT

|              |                           |      |
|--------------|---------------------------|------|
|              |                           |      |
| Printed name | Signature/mark/thumbprint | Date |

## INTERVIEWER

I herewith confirm that the above participant has been fully informed about the nature, conduct and risks of the above study

|              |           |      |
|--------------|-----------|------|
|              |           |      |
| Printed Name | Signature | Date |

## WITNESS STATEMENT:

As an impartial third party, I witnessed the entire consent discussion and the participant's signature on the form. I attest that this entire form was read to the participant named above. This person had enough time to consider this information, had an opportunity to ask questions, and voluntarily agreed to be in this study.

|              |           |      |
|--------------|-----------|------|
|              |           |      |
| Printed Name | Signature | Date |

**Appendix E: Information Sheet for Participants**

Participant Number

\_\_\_\_ BC \_\_\_\_

**Study Number:** M161116**Study Title:** Association of Genetic Variants with Breast Cancer Intrinsic Subtypes and Splice Variants of the Fibroblast Growth Factor Receptor 2 in a South African Population.**Principal Investigator:** Ms Therese Dix-Peek**Institution:** University of the Witwatersrand**STUDY INFORMATION SHEET - CONTROLS**

Good day, my name is \_\_\_\_\_ and I am a study interviewer.

I would like to invite you to consider taking part in a research study entitled “**Association of Genetic Variants with Breast Cancer Intrinsic Subtypes and Splice Variants of the Fibroblast Growth Factor Receptor 2 in a South African Population**”.

You are being asked to be involved in this study because you do not have breast cancer, and will act as a control for breast cancer studies. This is further explained under the heading, “What is a control group”.

Before you decide, it is important that you read and understand why the research is being done and what it will involve for you. Please take time to read the following information carefully and discuss with others if you wish. Your research team will explain if there is anything that is not clear, or if you would like more information. Take time to decide whether or not you wish to take part.

**Purpose of study**

We are doing this study because little is known about risk factors for breast cancer in a Black African population. We want to investigate gene changes linked to increased risk (likelihood) of developing inherited and non-inherited types of breast cancer and to compare our findings with those previously reported in other parts of the world. For genetic predisposition characterization on your DNA extracts (genetic material), we plan to examine small genetic material called single nucleotide polymorphisms (SNPs) that are known to be related to breast cancer using known breast cancer genes FGFR2, CHEK2, MAP3K1, TNRC9/TOX3, MRPS30 and LSP1, among others. We may also wish to identify in your biological samples, hormones and small substances called biomarkers. We will compare our results against what is known mainly about white patients from other parts of the world. The results of this study may help us to develop interventions that can help to decrease breast cancer in our communities.

**What is a control group**

For any research study to work well, there is always a need for a control group. For this study, the results from the control group will be compared to the results from people with breast cancer. This can show where there are differences between the two groups and potentially lead to new therapies. The control group needed for this study is black women with no history of cancer. This will help us understand breast cancer in black South

African women better. You are being asked to be a control in this study, because you do not have breast cancer.

### **Study Procedures**

If you agree to take part in this study, you will first be asked questions to see if you qualify for this study. These questions will take about 20 minutes. We ask permission to take approximately 8 ml (about 1-2 teaspoons) of your blood. These blood samples will be used for genetic studies.

We would like to keep your samples for future testing and will keep these samples for a maximum of 15 years. Before we use any of your stored samples we will get permission from the University of the Witwatersrand Human Research Ethics Committee (Medical) to do the research and we will make sure that all your information is kept confidential.

### **Benefits**

There is no direct benefit to participating, but you might help us learn about risk factors for breast cancer.

### **Possible Risks**

Drawing blood is normally done as part of routine medical care. It will cause some pain at the site and some physical discomfort.

### **Confidentiality**

All the information we will get during this study will be kept strictly confidential. Your samples will be given a unique study number, so that no result can be traced back to you. The information and results from the study will be made available to the public through scientific publications, reports, websites or publications, but you will not be identified personally in these.

### **Participation**

Your participation in this study is entirely voluntary and you can withdraw at any stage, without giving an explanation. At that time, you can decide if you are happy for us to use samples collected, or you would like these to be destroyed.

If our experiments show that you might have breast cancer, we will contact you and refer you to a breast cancer clinic so that they can check you for the disease, and give you appropriate treatment.

### **Additional Information**

If you have any other questions, please contact Therese Dix-Peek at 084-736-3155.

## Appendix F: Inclusion/Exclusion Criteria

## Appendix F

Participant Number:

\_\_\_ B C \_\_\_

### Eligibility Verification Form Controls

| A.                                                                                                            | Inclusion Criteria                                               |                              |                             |
|---------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|------------------------------|-----------------------------|
| 1.                                                                                                            | Are you $\geq 18$ years of age?                                  | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| 2.                                                                                                            | Are you of self-assessed African ethnic origin?                  | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| 3.                                                                                                            | Have you lived in the Greater Gauteng area for the past 3 years? | <input type="checkbox"/> Yes | <input type="checkbox"/> No |
| <p><i>If she responds "No" to any of the former questions, <b>she is INELIGIBLE</b>. Go to Section C.</i></p> |                                                                  |                              |                             |

| B.                                                                                                             | Exclusion Criteria                                                                                                                                                                                                                                |                              |                             |                              |
|----------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|-----------------------------|------------------------------|
| 4.                                                                                                             | <b>Only CASES, controls=N/A: (in situ and invasive cases will be included)</b><br>Have you received tumor treatment, such as radiotherapy, chemotherapy or anti-estrogens (for example, tamoxifen)?                                               | <input type="checkbox"/> N/A | <input type="checkbox"/> No | <input type="checkbox"/> Yes |
| 5.                                                                                                             | Are you currently taking or have you taken any of the following medications in the past 6 months: tamoxifen, Evista (raloxifene), Fareston (toremifene), Aromasin (exemestane), Femara (letrozole), Arimidez (anastrozole) or Megace (megestrol)? |                              | <input type="checkbox"/> No | <input type="checkbox"/> Yes |
| 6.                                                                                                             | Are you suffering from Stage IV or V chronic kidney disease?                                                                                                                                                                                      |                              | <input type="checkbox"/> No | <input type="checkbox"/> Yes |
| 7.                                                                                                             | Are you pregnant or currently nursing?                                                                                                                                                                                                            |                              | <input type="checkbox"/> No | <input type="checkbox"/> Yes |
| 8.                                                                                                             | Have you been previously diagnosed with a breast cancer?                                                                                                                                                                                          |                              | <input type="checkbox"/> No | <input type="checkbox"/> Yes |
| 9.                                                                                                             | Have you been previously diagnosed with any other cancers?                                                                                                                                                                                        |                              | <input type="checkbox"/> No | <input type="checkbox"/> Yes |
| 10.                                                                                                            | Have you previously been diagnosed with ductal carcinoma in situ (DCIS) or lobular carcinoma in situ (LCIS)?                                                                                                                                      |                              | <input type="checkbox"/> No | <input type="checkbox"/> Yes |
| 11.                                                                                                            | Any other reason to exclude? Explain: _____                                                                                                                                                                                                       |                              | <input type="checkbox"/> No | <input type="checkbox"/> Yes |
| 12.                                                                                                            | <b>INTERVIEWER ASSESSMENT:</b> Does the women you are interviewing have a pathology that will hinder adequate communication?                                                                                                                      |                              | <input type="checkbox"/> No | <input type="checkbox"/> Yes |
| <p><i>If she responds "Yes" to any of the former questions, <b>she is ineligible</b>. Go to Section C.</i></p> |                                                                                                                                                                                                                                                   |                              |                             |                              |

| C.                              | Final Eligibility Status for Enrollment                                                                                                                                                                                                                 |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                 | <input type="checkbox"/> Eligible (all responses to Section A="Yes" or "N/A" and Section B="No")<br><input type="checkbox"/> Ineligible (any response in Section A="No" or Section B="Yes")<br><input type="checkbox"/> Declined, specify reason _____. |
| Interviewer's Name and Surname: | Completion Date:                                                                                                                                                                                                                                        |
| _____                           | ___ / ___ / ____<br>dd mm yyyy                                                                                                                                                                                                                          |
|                                 | Interviewers Signature:                                                                                                                                                                                                                                 |
|                                 | _____                                                                                                                                                                                                                                                   |

Participant Number

\_\_\_\_ BC \_\_\_\_

Date: \_\_\_\_\_

| Demographics                                                                                                                                                      |                                                                                                                                                                                |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| First Name(s):                                                                                                                                                    | Surname:                                                                                                                                                                       |
| Title: <input type="checkbox"/> Mrs <input type="checkbox"/> Miss <input type="checkbox"/> Ms <input type="checkbox"/> Dr<br><input type="checkbox"/> Other _____ | Race: <input type="checkbox"/> Black <input type="checkbox"/> White <input type="checkbox"/> Mixed race<br><input type="checkbox"/> Asian <input type="checkbox"/> Other _____ |
| Date Of Birth:                                                                                                                                                    | Age:                                                                                                                                                                           |
| ID No.:                                                                                                                                                           | Home Language:                                                                                                                                                                 |
| Home number: _____<br>Cell number: _____<br>Other contact: _____                                                                                                  | Address: _____<br>_____<br>_____                                                                                                                                               |
| Hospital                                                                                                                                                          |                                                                                                                                                                                |
| Hospital:                                                                                                                                                         | Clinic:                                                                                                                                                                        |
| Hospital #:                                                                                                                                                       |                                                                                                                                                                                |

| Patient History and Risk factors                                                                                                                                                                                                                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. What is your weight (in kg) _____ What is your height (in meters) _____                                                                                                                                                                           |
| 2. When did menarche start?<br><input type="checkbox"/> Age _____<br><input type="checkbox"/> Do not remember.                                                                                                                                       |
| 3. Current menstrual status:<br><input type="checkbox"/> Fertile / Premenopausal<br><input type="checkbox"/> Perimenopausal<br><input type="checkbox"/> Postmenopausal - How old were you _____<br><input type="checkbox"/> Not applicable / Unknown |

|                                                                                                                                                                                |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 4. Number of full term pregnancies<br><input type="checkbox"/> None/Not applicable<br><input type="checkbox"/> Had children - How many full term pregnancies? _____            |
| 5. Breastfeeding<br><input type="checkbox"/> None/Not applicable<br><input type="checkbox"/> How many months did you spend breastfeeding each child? _____                     |
| 6. Have you ever been on oral contraceptives?<br><input type="checkbox"/> No <input type="checkbox"/> Yes - How long? _____                                                    |
| 7. Have you ever used injectable contraceptives?<br><input type="checkbox"/> No <input type="checkbox"/> Yes - How long? _____                                                 |
| 8. Have you had a hysterectomy?<br><input type="checkbox"/> No <input type="checkbox"/> Yes – What year? _____                                                                 |
| 9. Ever had a BSO (Bilateral salphingo-oophorectomy)<br><input type="checkbox"/> No <input type="checkbox"/> Yes – What year? _____ <input type="checkbox"/> Unsure            |
| 10. Have you ever been / are you on hormone replacement therapy<br><input type="checkbox"/> No <input type="checkbox"/> Yes – What year? _____ <input type="checkbox"/> Unsure |
| 11. Have you ever been diagnosed with diabetes?<br><input type="checkbox"/> No <input type="checkbox"/> Yes – What year? _____                                                 |
| 12. Have you ever been diagnosed with hypertension?<br><input type="checkbox"/> No <input type="checkbox"/> Yes – What year? _____                                             |
| 13. Have you ever been diagnosed with heart disease?<br><input type="checkbox"/> No <input type="checkbox"/> Yes – What year? _____                                            |
| 14. Have you ever had a stroke<br><input type="checkbox"/> No <input type="checkbox"/> Yes – What year? _____                                                                  |
| 15. Have you ever had TB?<br><input type="checkbox"/> No <input type="checkbox"/> Yes – What year? _____                                                                       |
| 16. Have you ever been diagnosed with arthritis?<br><input type="checkbox"/> No <input type="checkbox"/> Yes – What year? _____                                                |
| 17. Have you ever been diagnosed with asthma/COPD?                                                                                                                             |

|                                                                                                                                                                                                              |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/> No <input type="checkbox"/> Yes – What year? _____                                                                                                                                  |
| 18. Have you been diagnosed with any other chronic conditions?<br><input type="checkbox"/> No <input type="checkbox"/> Yes. If yes, what illnesses and when were they were diagnosed?                        |
| 19. Has anyone in your family been diagnosed with breast cancer?<br><input type="checkbox"/> No <input type="checkbox"/> Yes. If yes, who had breast cancer, and how old were they when they were diagnosed? |
| 20. Do you have/had any other cancer?<br><input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> Unsure                                                                            |
|                                                                                                                                                                                                              |

## Appendix G: Prosigna® Package Insert

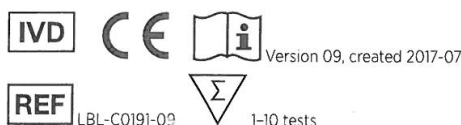
nanoString

www.nanostring.com



## Package Insert

## Prosigna® Breast Cancer Prognostic Gene Signature Assay



## Storage conditions

|  |                                                                    |
|--|--------------------------------------------------------------------|
|  | Prosigna Breast Cancer Prognostic Gene Signature Assay Cartridges  |
|  | Prosigna Breast Cancer Prognostic Gene Signature Assay CodeSet     |
|  | Prosigna Breast Cancer Prognostic Gene Signature Assay Prep Pack   |
|  | Prosigna Breast Cancer Prognostic Gene Signature Assay Prep Plates |

## TABLE OF CONTENTS

|      |                                                                       |    |
|------|-----------------------------------------------------------------------|----|
| 1    | INTENDED USE / PURPOSE .....                                          | 1  |
| 2    | SUMMARY OF THE TEST SYSTEM .....                                      | 1  |
| 2.1  | Principles of the nCounter Dx Analysis System .....                   | 2  |
| 2.2  | Principles of the Prosigna algorithm for output calculation .....     | 2  |
| 3    | REAGENTS AND EQUIPMENT PROVIDED .....                                 | 2  |
| 3.1  | Prosigna Kit Overview .....                                           | 2  |
| 3.2  | Prosigna Kit Contents for a 1, 2, 3, 4, or 10 test Prosigna Kit ..... | 3  |
| 4    | WARNINGS AND PRECAUTIONS .....                                        | 3  |
| 5    | GENERAL ASSAY CONSIDERATIONS .....                                    | 3  |
| 5.1  | Tissue Processing .....                                               | 3  |
| 5.2  | Performing the Prosigna Assay .....                                   | 4  |
| 6    | TRAINING INFORMATION .....                                            | 4  |
| 7    | WASTE HANDLING .....                                                  | 4  |
| 8    | STORAGE AND HANDLING (REAGENTS) .....                                 | 4  |
| 9    | INSTRUMENTS REQUIRED FOR PROSIGNA .....                               | 4  |
| 10   | REAGENTS AND EQUIPMENT REQUIRED BUT NOT PROVIDED .....                | 4  |
| 10.1 | Materials .....                                                       | 4  |
| 10.2 | Equipment .....                                                       | 4  |
| 10.3 | Equipment Specifications .....                                        | 4  |
| 11   | SPECIMEN COLLECTION AND PROCESSING .....                              | 5  |
| 11.1 | Tissue Specimen Requirements and Pathology Review .....               | 5  |
| 11.2 | Specimen Collection and Storage .....                                 | 5  |
| 11.3 | Slide Preparation .....                                               | 5  |
| 11.4 | Slide Processing .....                                                | 5  |
| 11.5 | Isolation of RNA .....                                                | 6  |
| 11.6 | Measure the RNA Concentration and Quality .....                       | 6  |
| 11.7 | Assay Procedure .....                                                 | 7  |
| 12   | TROUBLESHOOTING AND TEST FAILURES .....                               | 9  |
| 13   | RESULTS OF THE ASSAY .....                                            | 9  |
| 13.1 | Intrinsic Subtypes .....                                              | 9  |
| 13.2 | ROR Score .....                                                       | 10 |
| 13.3 | Probability of 10-Year Distant Recurrence .....                       | 10 |
| 13.4 | Risk Classification .....                                             | 10 |
| 13.5 | Quality Control .....                                                 | 10 |
| 14   | LIMITATIONS OF THE PROCEDURES .....                                   | 10 |
| 15   | EXPECTED VALUES .....                                                 | 11 |
| 15.1 | ROR range by Subtype .....                                            | 11 |
| 15.2 | Frequency of ROR score by nodal status .....                          | 11 |
| 15.3 | Distant Recurrence-Free Survival by Risk Categorization .....         | 11 |

|      |                                                |    |
|------|------------------------------------------------|----|
| 16   | PERFORMANCE CHARACTERISTICS .....              | 12 |
| 16.1 | Analytical Precision and Reproducibility ..... | 12 |
| 16.2 | Sensitivity / RNA input .....                  | 13 |
| 16.3 | Interference testing .....                     | 13 |
| 16.4 | Clinical Performance .....                     | 14 |
| 17   | BIBLIOGRAPHY .....                             | 25 |
| 18   | SYMBOLS AND DEFINITIONS .....                  | 26 |
| 19   | CONTACT INFORMATION .....                      | 26 |

## 1 INTENDED USE / PURPOSE

The Prosigna® Breast Cancer Prognostic Gene Signature Assay is an *in vitro* diagnostic assay which uses the gene expression profile of cells found in breast cancer tissue to assess a patient's risk of distant recurrence. The assay measures the gene expression profile using RNA extracted from Formalin-Fixed, Paraffin-Embedded (FFPE) breast tumor tissue. The gene expression data are weighted together with clinical variables to generate both a subtype (luminal A, luminal B, HER2-enriched, or basal-like) and a score indicative of the probability of distant recurrence of disease. The assay is performed on the NanoString nCounter® Dx Analysis System using FFPE breast tumor tissue previously diagnosed as invasive breast carcinoma.

The Prosigna Breast Cancer Prognostic Gene Signature Assay is indicated in female breast cancer patients who have undergone either mastectomy or breast-conserving therapy in conjunction with locoregional treatment consistent with standard of care, either as:

- A prognostic indicator for distant recurrence-free survival at 10 years in post-menopausal women with Hormone Receptor-Positive (HR+), lymph node-negative, Stage I or II breast cancer to be treated with adjuvant endocrine therapy alone, when used in conjunction with other clinicopathological factors.
- A prognostic indicator for distant recurrence-free survival at 10 years in post-menopausal women with Hormone Receptor-Positive (HR+), lymph node-positive (1-3 positive nodes, or 4 or more positive nodes), Stage II or IIIA breast cancer to be treated with adjuvant endocrine therapy alone, when used in conjunction with other clinicopathological factors.

## 2 SUMMARY OF THE TEST SYSTEM

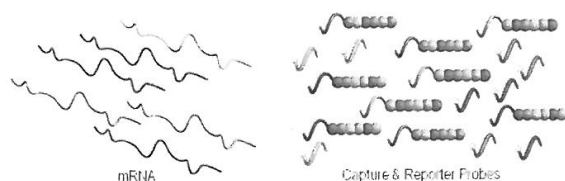
The NanoString nCounter Dx Analysis System delivers direct, multiplexed measurements of gene expression through digital readouts of the relative abundance of mRNA transcripts using the following steps: 1) hybridization of the RNA to fluorescent Reporter Probes and Capture Probes, 2) purification of the target/probe complexes using nCounter Prep Plates containing reagents necessary for post-hybridization processing and immobilization onto the nCounter Cartridge on the nCounter Prep Station, and 3) analysis of the nCounter Cartridge on the nCounter Digital Analyzer to provide a test result. Both the Capture and Reporter Probes contain unique DNA probe sequences for target hybridization and purification. The Capture and Reporter Probes are combined with the positive and negative controls to form the CodeSet. Prosigna simultaneously measures the expression levels of 50 genes used for the intrinsic subtype classification algorithm<sup>2</sup>, 8 housekeeping genes used for signal normalization, 6 positive controls, and 8 negative controls in a single hybridization reaction using nucleic acid probes designed specifically to those genes. Also included in the Prosigna kit is a Reference Sample consisting of *in vitro* transcribed RNA targets for each of the 58 genes. The Reference Sample is tested with each batch of patient RNA samples to qualify the run and normalize the signal from each gene.

The Prosigna Assay is performed on RNA isolated from FFPE breast tumor tissue. A pathologist examines a hematoxylin and eosin (H&E) stained slide and identifies (and marks) the area of invasive breast carcinoma suitable for the test. The pathologist also measures the tumor surface area, which determines the number of unstained slides required for the test, and the tumor cellularity to ensure the presence of sufficient tumor tissue for the test. A trained technologist macrodissects the area on the unstained slides corresponding to the marked tumor area on the H&E stained slide and isolates RNA from the tissue. The isolated RNA is then tested on the NanoString nCounter Dx Analysis System to provide test results, including the intrinsic subtype, Risk of Recurrence (ROR) score, and risk category.

## 2.1 Principles of the nCounter Dx Analysis System

The nCounter Dx Analysis System uses gene-specific probe pairs (Figure 1) that hybridize directly to the mRNA sample in solution, eliminating enzymatic reactions that might introduce bias in the results. In the first step of the assay the DNA probes are hybridized directly to a 70–100 base pair region of the RNA sample in solution. The fluorescent Reporter Probe consists of a 35–50 base probe sequence that is complementary to the mRNA target and a unique backbone DNA sequence that hybridizes to six RNA segments labeled with one of four fluorescent dyes; red (R), yellow (Y), blue (B), or green (G). The fluorescent segments create a six-position/four-color fluorescent “color code” that is unique to each target. A separate Capture Probe consists of a 35–50 base probe sequence that is complementary to the mRNA target and biotin, which is used for immobilization onto a streptavidin-coated slide.

Figure 1: Hybridize CodeSet to mRNA



After hybridization, all of the sample purification steps are automated on the nCounter Prep Station. First, excess Capture and Reporter Probes are removed (Figure 2) using successive magnetic bead capture steps followed by binding of the probe-target complexes to random locations on the surface of the nCounter Cartridge via a streptavidin-biotin linkage (Figure 3). Finally, probe/target complexes are aligned and immobilized (Figure 4) in the nCounter Cartridge.

Figure 2: Remove Excess Reporters

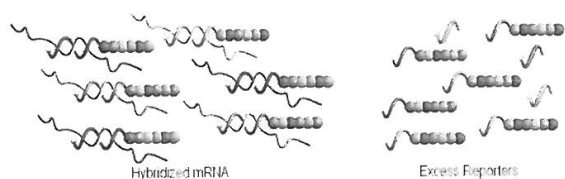


Figure 3: Bind Hybridized Reporters to Surface of Cartridge

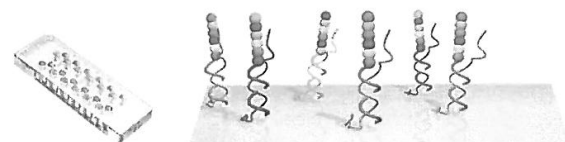


Figure 4: Align and Immobilize Hybridized Reporters



After sample processing has completed, the Cartridge is placed in the nCounter Digital Analyzer for data collection. Each target molecule of interest is identified by the “color code” generated by six ordered fluorescent spots present on its associated Reporter Probe. The Reporter Probes on the surface of the Cartridge are then counted and tabulated for each target molecule and processed with the algorithm (Figure 5).

Figure 5: Data Collection

| Code   | Gene | Count |
|--------|------|-------|
| •••••• | x    | 3     |
| •••••• | y    | 1     |
| •••••• | z    | 2     |

## 2.2 Principles of the Prosigna algorithm for output calculation

The test is based on the reported 50-gene classifier algorithm originally named PAM50<sup>2</sup> and is performed on the nCounter Dx Analysis system using RNA extracted from formalin fixed, paraffin embedded (FFPE) breast tumor tissue samples. The algorithm uses a 50-gene expression profile to assign breast cancer to one of four molecular classes, or intrinsic subtypes: Luminal A, Luminal B, HER2-enriched, or Basal-like<sup>2</sup>. The prototypical gene expression profiles (e.g., centroid) of the four intrinsic subtypes were retrained on the nCounter Analysis System using FFPE breast tumor samples collected from multiple clinical sites in North America. After performing the assay on a patient test sample, a computational algorithm based on a Pearson's correlation compares the normalized 50-gene expression profile of the patient test sample to the prototypical expression profiles of the four breast cancer intrinsic subtypes. The patient test sample is assigned the subtype with the highest Pearson's correlation.

The algorithm further reports a Risk of Recurrence (ROR) score on a 0–100 scale<sup>3</sup>, which is correlated with the probability of distant recurrence at ten years for post-menopausal women with hormone receptor-positive, early stage breast cancer<sup>4</sup>. The report also provides a risk category (low, intermediate, or high). The ROR score is calculated using coefficients from a Cox model that includes the Pearson correlation of a 46-gene subset of the 50 genes to each intrinsic subtype centroid, a proliferation score, and gross tumor size. The test variables are multiplied by the corresponding coefficients from the Cox model to generate the score, which is then adjusted to a 0–100 scale based on coefficients generated from the training set of FFPE breast tumor samples. Risk categories are also reported based on cut-offs for ROR determined in a clinical validation study.

## 3 REAGENTS AND EQUIPMENT PROVIDED

### 3.1 Prosigna Kit Overview

The Prosigna kit contains reagents sufficient to process 1, 2, 3, 4, or 10 patient samples depending on the product ordered. See below for ordering information. The Prosigna kit contains a CodeSet, one Reference Sample tube for every set of one to ten tests, and consumable components, which are tested together for performance prior to release.

| Catalogue Number | Number of tests in Kit | Reference Sample Tubes included |
|------------------|------------------------|---------------------------------|
| PROSIGNA-001     | 1                      | 2                               |
| PROSIGNA-002     | 2                      | 2                               |
| PROSIGNA-003     | 3                      | 2                               |
| PROSIGNA-004     | 4                      | 2                               |
| PROSIGNA-010     | 10                     | 2                               |

Recommended for use in conjunction with the Roche FFPE RNA Isolation Kit (Roche-FFPET-025 available only through NanoString Technologies)

### 3.2 Prosigna Kit Contents for a 1, 2, 3, 4, or 10 test Prosigna Kit

| Number of Tests                   | 1          | 2          | 3          | 4          | 10         |
|-----------------------------------|------------|------------|------------|------------|------------|
| <b>Prosigna CodeSet Box</b>       |            |            |            |            |            |
| Prosigna Reporter CodeSet         | 1 x 65 µL  | 1 x 65 µL  | 1 x 65 µL  | 1 x 65 µL  | 1 x 65 µL  |
| Prosigna Capture ProbeSet         | 1 x 70 µL  | 1 x 70 µL  | 1 x 70 µL  | 1 x 70 µL  | 1 x 70 µL  |
| Prosigna RNA Reference Sample     | 1 x 30 µL  | 1 x 30 µL  | 1 x 30 µL  | 1 x 30 µL  | 1 x 30 µL  |
| CodeSet Barcode Sticker           | 1          | 1          | 1          | 1          | 1          |
| Test Configuration Code           | 1          | 1          | 1          | 1          | 1          |
| <b>Prosigna Prep Plate Box</b>    |            |            |            |            |            |
| Prep Plates                       | 1          | 1          | 1          | 1          | 2          |
| <b>Prosigna Cartridge Box</b>     |            |            |            |            |            |
| nCounter Cartridges               | 1          | 1          | 1          | 1          | 1          |
| <b>Prosigna Prep Pack Box</b>     |            |            |            |            |            |
| nCounter Prep Station Tips        | 1          | 1          | 1          | 1          | 1          |
| nCounter Cartridge Adhesive Cover | 2          | 2          | 2          | 2          | 2          |
| nCounter Tip Sheaths              | 2          | 2          | 2          | 2          | 2          |
| nCounter Hybridization Buffer     | 1 x 580 µL | 1 x 580 µL | 1 x 580 µL | 1 x 580 µL | 1 x 580 µL |
| 12-Well Notched Strip Tubes       | 4          | 4          | 4          | 4          | 4          |
| 12-Well Notched Strip Tube Lids   | 4          | 4          | 4          | 4          | 4          |

#### Contents Description

##### Prosigna CodeSet

|                               |                                             |
|-------------------------------|---------------------------------------------|
| Prosigna Reporter CodeSet     | buffer, nucleic acids with fluorescent dyes |
| Prosigna Capture ProbeSet     | buffer, nucleic acids                       |
| Prosigna RNA Reference Sample | buffer, nucleic acids                       |
| CodeSet Barcode Sticker       | sticker sheet                               |
| Test Configuration Code       | card with sticker                           |

##### Prosigna Prep Plates

|             |                                                                                                         |
|-------------|---------------------------------------------------------------------------------------------------------|
| Prep Plates | superparamagnetic beads, buffer, salts, oligonucleotides, polystyrene beads containing fluorescent dyes |
|-------------|---------------------------------------------------------------------------------------------------------|

##### Prosigna Cartridges

|                       |                     |
|-----------------------|---------------------|
| nCounter Cartridge(s) | sample cartridge(s) |
|-----------------------|---------------------|

##### Prosigna Prep Pack

|                                   |                                          |
|-----------------------------------|------------------------------------------|
| nCounter Hybridization Buffer     | buffer, salts                            |
| 12-Well Notched Strip Tubes       | plastic strips                           |
| 12-Well Notched Strip Tube Lids   | plastic lids                             |
| nCounter Prep Station Tips        | 2 racks of 90 tips + 6 piercers nCounter |
| nCounter Cartridge Adhesive Cover | adhesive films                           |
| nCounter Tip Sheaths              | 6-well tip holders                       |

## 4 WARNINGS AND PRECAUTIONS

1. For *In Vitro* Diagnostic Use.
2. This assay is intended to be run by operators trained in highly complex molecular biology techniques, based on local regulations.
3. Do not mix components of kits across Prosigna lots. Functionality can only be assured for Prosigna kit lots as provided, as they are qualified in this manner during manufacture.
4. Any remnant reagents should not be reused in the Prosigna Assay.
5. Discard any reactions with compromised hybridization times or temperatures.
6. It is important to maintain the integrity of the sample chain of custody (tissue to RNA and RNA to assay) to ensure the patient sample ID is associated with the correct test result.
7. Failure to store reagents under the conditions stated on the label could adversely affect assay performance.
8. Always wear gloves while handling reagents and samples.
9. Avoid RNase contamination, which may negatively affect the quality of the results.
10. All biological specimens and materials should be handled as if the potential exists for transmitting infectious agents and should be disposed of with proper precautions in accordance with federal, state, and local regulations.

11. Never pipet by mouth.
12. Avoid reagent contact with eyes, skin and mucous membranes.
13. Use molecular laboratory best practices to prevent cross-contamination between test samples or with high concentration nucleic acid targets (synthetic or PCR-amplified) which may negatively affect results quality.
14. Very low levels of sodium azide (< 0.1%) are contained post-process within the Prosigna Prep Plates and nCounter Cartridges, therefore it is recommended that plastic (not metal) waste receptacles are used for disposal. While extremely unlikely for Prosigna, sodium azide accumulation on metal is known to create an explosive hazard.
15. Additional instrument-specific disposal information can be found in the nCounter Dx Analysis System User Manual and Service Manuals for the Prep Station and Digital Analyzer.
16. Material Safety Data Sheet information for Reporter CodeSet, Capture ProbeSet, Hybridization Buffer, and Prep Plates can be found at [www.prosigna.com](http://www.prosigna.com).
17. All hazardous materials should be disposed of according to your institution's guidelines for hazardous disposal.
18. Any unused CodeSet should be discarded.
19. If a patient's tumor size category is entered into the software incorrectly, the ROR score and risk classification may be adversely affected (e.g., shifted ROR Score and/or mis-classification).
20. If a patient's nodal status is entered into the software incorrectly, the patient's test results may be reported incorrectly (e.g., incorrect risk classification).
21. Do not use RNA of insufficient quality or quantity or tumor samples with insufficient tumor surface area or cellularity in the Prosigna Assay. The Prosigna Assay may be unable to give a valid result and instead will report as assay failure.

## 5 GENERAL ASSAY CONSIDERATIONS

1. The assay is intended for use only on formalin-fixed, paraffin embedded (FFPE) breast cancer tissue specimens from surgical resection; it is not intended for use on fresh, frozen or non-breast cancer tissue.
2. The gross size of a patient's primary tumor and nodal status are required to perform the assay.
3. Use sterile disposable micropipette tips to avoid microbial and nuclease contamination of reagents or samples during processing.
4. Maintain the isolated RNA samples on wet ice when not actively being manipulated.
5. Calibrated thermometers are required for heat blocks.
6. Do not use kit components in the event that they arrive damaged.
7. It is recommended that clinical controls (e.g., for risk category) are developed and utilized by labs running Prosigna in order to ensure accuracy of results over time as part of standard laboratory quality control procedures.

### 5.1 Tissue Processing

1. Failure to properly remove surrounding non-tumor/normal tissue by macrodissection during tissue processing could result in an underestimate of risk due to a lower ROR score reported to the physician.
2. Failure to properly remove human genomic DNA during RNA isolation could result in a higher failure rate due to lower assay signal, or an overestimate of risk due to a higher ROR score reported to the physician.
3. All unstained tissue sections should be mounted onto positively-charged microscope slides to avoid detachment during tissue processing.
4. For specimens that require multiple slides, all slides must be processed together.
5. Slide mounted tissue sections may degrade if stored for longer than 9 months in a desiccated environment.
6. Replace 3% glycerol working solution each week, or if solution becomes turbid, to avoid contamination.
7. Change the first D-Limonene wash contents after processing 4 slide sets, and the Ethanol (EtOH) and second D-Limonene staining dish contents after processing 8 slide sets to avoid compromising tissue quality.
8. Use caution when outlining the tumor area on the unstained slide and removing non-tumor tissue to ensure the tumor tissue is not disturbed.

9. Handle sharps with care during macrodissection.
10. Use a fresh razor blade for each tissue sample processed.
11. 10% SDS frequently precipitates at room temperature and should be warmed at 37°C until precipitates have dissolved.
12. New lots/batches of RNA isolation kits should be tested against the isolation kit specifications to qualify the new kit lots for patient testing (see section 11.5 for details).

## 5.2 Performing the Prosigna Assay

1. Ensure the patient's categorical primary gross tumor size is entered correctly into the software.
2. Ensure the patient's categorical nodal status is entered correctly into the software.
3. Verify that the heat block with heated lid required for hybridization meets the specifications and is routinely calibrated.
4. Use only those consumables that were provided with the Prosigna kit. They are designed specifically to work with the nCounter Prep Station and nCounter Digital Analyzer.
5. If Hybridization Buffer has been stored at cold temperatures and a precipitate is observed, warm tubes at 37°C until salts have dissolved.
6. Do not vortex assay components vigorously to mix, as it may damage the reagents. Mixing should be performed using a pipette.
7. Do not centrifuge Reporter CodeSet faster than 3,000 × g for more than 10 seconds. Do not use the "pulse" option to centrifuge. Doing so may precipitate the CodeSet.
8. Maintain hybridization reactions at 65°C until they are ready to be transferred to the Prep Station. Setting the heat block to ramp down to 4°C or placing samples on ice at the end of the hybridization could result in cross-hybridization, which may compromise assay results.
9. Failure to place the strip tubes at 65°C within 15 minutes of adding the Capture ProbeSet could result in cross-hybridization, which may compromise assay results.
10. Failure to initiate processing of the Prep Station within 15 minutes of removing the samples from 65°C could result in cross-hybridization, which may compromise assay results.
11. Ensure the strip tube caps are firmly sealed prior to hybridization in the heat block to prevent evaporation, which may compromise assay results.

## 6 TRAINING INFORMATION

This assay is intended to be run by professional operators trained in highly complex molecular biology techniques, based on local regulations. Please contact NanoString for training information specific to running the Prosigna Assay.

## 7 WASTE HANDLING

See the nCounter Dx Analysis System User Manual for waste handling details specific to the reagents and instruments for use in IVD applications.

## 8 STORAGE AND HANDLING (REAGENTS)

The expiration date for all assay kit components is listed on the barcode label supplied with the CodeSet box as well as on the outer box labeling for all Prosigna components.

- The Prosigna CodeSet box components (Prosigna Reporter CodeSet, Prosigna Capture ProbeSet, and Prosigna RNA Reference Sample) must be stored at -80°C or below.
- The nCounter Cartridges must be stored at or below -20°C.
- The nCounter Prep Plates must be stored at 4°C (2-8°C).
- The nCounter Prep Pack components must be stored at room temperature 15°C-25°C.

## 9 INSTRUMENTS REQUIRED FOR PROSIGNA

- nCounter Dx Analysis System (Catalog number NCT-SYST-DX) (includes both instruments below)
  - nCounter Prep Station 5s (Catalog number NCT-PREP-STATION-FLEX)
  - nCounter Digital Analyzer 5s (Catalog number NCT-DIGITAL-ANALYZER-FLEX)

Refer to the nCounter Dx Analysis System User Manual for additional information.

## 10 REAGENTS AND EQUIPMENT REQUIRED BUT NOT PROVIDED

### 10.1 Materials

1. FFPE RNA Isolation kit (see Section 11.5 for isolation kit requirements if not using Roche FFPE RNA isolation kit purchased through NanoString Technologies)
2. Hematoxylin and Eosin (H&E)
3. Positively charged glass microscope slides
4. D-Limonene clearing agent (histology grade)
5. 100% Ethanol (Absolute), ACS grade or equivalent (not less than 99.5%)
6. Glycerol, molecular biology grade
7. Nuclease-free water, molecular biology grade
8. 10% SDS, molecular biology grade
9. Razor Blades (or disposable scalpels)
10. Disposable Microtome Blades
11. 0.5 mL screw-cap and 1.5 or 1.7 mL nonstick RNase-free microcentrifuge tubes
12. RNase-free Micropipette tips with aerosol barrier

### 10.2 Equipment

1. Microtome
2. Water Bath (40°C)
3. Slide Warmer (45°C)
4. Microscope slide drying rack
5. Micropipettes; 2 µL, 20 µL, 200 µL and 1000 µL
6. Mini-centrifuge with a 0.2 mL strip tube rotor and standard 1.5/2.0 mL microcentrifuge tube rotor
7. Standard bench top micro-centrifuge with a fixed angle rotor that fits 1.5 mL centrifuge tubes
8. Rectangular glass staining dishes with covers (approximate interior dimensions of 3.6 × 2.8 × 2.4" (91 × 71 × 60 mm); quantity of 3 required)
9. Slide rack (holds up to ten 3" × 1" (75 × 25 mm) glass slides)
10. Dry heat block, stationary
11. Bench top vortexer for microcentrifuge tubes
12. Graduated cylinder (suggested size: 100-250 mL)
13. Dissecting Needle or cover glass forceps (angled, non-serrated)
14. Calibrated thermometers (covering the 55°C through 80°C range)
15. Micro-volume UV/Vis spectrophotometer (see specifications below)
16. Heat block with heated lid (see specifications below)
17. Centrifuge with plate microplate adapter (see specifications below)
18. Coplin jar

### 10.3 Equipment Specifications

Table 1: Full Spectrum Micro-volume UV/Vis spectrophotometer for Nucleic Acid Quantitation

| Design Feature                                   | Specifications             |
|--------------------------------------------------|----------------------------|
| Sample Volume Range                              | 1-2 µL                     |
| Path Length                                      | 1 mm                       |
| Wavelength Range                                 | 260-280 nm                 |
| Wavelength Accuracy or Error                     | ± 1 nm                     |
| Spectral Resolution or bandwidth                 | Less than or equal to 4 nm |
| Absorbance Precision or random photometric error | 0.003 (1 mm path)          |
| Detection Limit                                  | 5 ng/µL RNA                |
| Maximum Concentration                            | ≥ 1000 ng/µL RNA           |

Table 2: Microvolume Photodiode UV-Vis Spectrophotometer for Nucleic Acid Quantitation

| Design Feature        | Specifications             |
|-----------------------|----------------------------|
| Sample Volume Range   | 1-2 µL                     |
| Path Length           | 0.5 mm                     |
| Wavelength Range      | 260 and 280 nm             |
| Spectral Resolution   | Less than or equal to 8 nm |
| Absorbance Accuracy   | 3% (at 1.05 Abs at 260 nm) |
| Detection Limit       | 4 ng/µL RNA                |
| Maximum Concentration | ≥ 1000 ng/µL RNA           |

Table 3: Heat Block with Heated Lid for Assay Hybridization

| Design Feature    | Specifications                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Heat Block Design | <ul style="list-style-type: none"> <li>Must fit regular profile, 12 × 0.2 mL keyed strip tubes that are provided as part of the nCounter Prep Pack.</li> <li>Heat Blocks designed for Low Profile (LP) and High Profile (HP) tubes are not compatible (also referred to as "rapid" blocks for thermocycling).</li> <li>Heat Blocks designed for other types of tubes (e.g., 0.1 mL tubes, 1.5 mL tubes) are not compatible.</li> <li>Must be programmable to hold at a temperature of 65°C.</li> <li>Must hold temperature within ± 1°C of 65°C.</li> </ul> |
| Heated Lid Design | <ul style="list-style-type: none"> <li>Fixed or adjustable height lids are acceptable.</li> <li>Lid must be programmable to 70°C.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                |

Table 4: Centrifuge with microplate carrier to spin nCounter Prep Plates

| Design Feature       | Specifications                                                                                                           |
|----------------------|--------------------------------------------------------------------------------------------------------------------------|
| Centrifugation Speed | Minimum of 2000 × g                                                                                                      |
| Rotors               | 4 × 750 mL swinging bucket rotors with microplate carriers (or equivalent) to accommodate SBS-format 96-well microplates |
| Modes                | Acceleration/Deceleration modes                                                                                          |

## 11 SPECIMEN COLLECTION AND PROCESSING

### 11.1 Tissue Specimen Requirements and Pathology Review

- The Prosigna Breast Cancer Prognostic Gene Signature Assay should be performed on a formalin-fixed, paraffin-embedded (FFPE) hormone-receptor positive breast tumor tissue specimen that is further specified by a pathologist as one of the following types of invasive breast carcinoma:
  - Invasive ductal carcinoma
  - Invasive lobular carcinoma
  - Invasive carcinoma with ductal and lobular features ("mixed type carcinoma")
  - No special type (NST) or not otherwise specified (NOS)
- A pathologist should select the FFPE tumor block with the greatest area of viable invasive breast carcinoma for this test.
- The test requires unstained slide mounted tissue sections for processing and a corresponding H&E stained slide from the FFPE tumor block.
- It is recommended that tissue sections for assay processing are cut contiguous to the tissue section cut for H&E staining to ensure that the tumor area identified on the H&E stained slide is representative of the tumor area on the unstained slides.
- A pathologist must circle the region of viable invasive breast carcinoma on the H&E slide, excluding surrounding non-tumor tissue.
- A pathologist or trained laboratory technician must estimate the tumor cellularity and tumor surface area within the circled area of the H&E stained slide.
  - The tumor cellularity percentage on the H&E stained slide must be ≥ 10%.
  - The circled tumor surface area on the H&E stained slide must be ≥ 4 mm<sup>2</sup>.

\*Note that tumor cellularity percentage refers to the percentage of viable tumor cells within the circled tumor area.
- A total tumor surface area of greater than 100 mm<sup>2</sup> is recommended as an input for the test. The following table illustrates the number of slides recommended based on the measured tumor surface area on the H&E stained slide.
- If the tissue review process shows that the tumor block has insufficient tumor area or insufficient tumor cellularity, then a different block from the same tumor may be assessed. If there are no FFPE blocks which contain sufficient tumor tissue, then the Prosigna Assay should not be run. Please note that for tumors with less than 20 mm<sup>2</sup> surface area, it is more likely that RNA input requirements will not be met.

Table 5: Recommended slide requirements based on tumor surface area

| Measured Tumor Surface Area on H&E Stained Slide (mm <sup>2</sup> ) | Number of unstained slides |
|---------------------------------------------------------------------|----------------------------|
| 4-19                                                                | 6                          |
| 20-99                                                               | 5                          |
| ≥ 100                                                               | 1                          |

### 11.2 Specimen Collection and Storage

- The following may be performed according to the laboratory's standard operating procedures: tissue collection and formalin fixation, FFPE tumor block handling and storage, and shipping of slide mounted FFPE tissue.
- Slide mounted FFPE tissue sections must be stored according to the laboratory's standard operating procedures. If storing for longer periods of time (> 30 days), the slides must be stored in a desiccated environment and processed within 9 months to ensure quality of the test results.

### 11.3 Slide Preparation

- Using a microtome, cut a 4-5 µm thick section for H&E staining.
- Using a microtome, cut 10 µm thick sections for use in the Prosigna Assay.
- Float the sections in a water bath at 40°C.
- Mount the sections onto positively charged glass microscope slides.
- Allow slides to air dry.
- Bake the slides overnight at 45°C.

### 11.4 Slide Processing

- Prepare a 3% glycerol working solution by mixing 1.5 mL of glycerol stock with 48.5 mL of molecular grade, nuclease-free water; scale as appropriate. Pour the solution into a Coplin jar for processing slides.
  - Pour approximately 200-250 mL of D-Limonene clearing agent into two staining dishes, ensuring the slides in the slide rack will be completely submerged.
  - Pour approximately 200-250 mL absolute Ethanol (EtOH) into a third staining dish.
  - Place the unstained slide-mounted tissue section(s) into a slide rack.
  - Place the slide rack into the first D-Limonene staining dish and gently agitate the slide rack back and forth for 10-15 seconds. Leave the rack in the first D-Limonene staining dish for a total time of 2 minutes.
  - Move the slide rack from the first D-Limonene staining dish into the second D-Limonene dish. Gently agitate the slide rack back and forth for 10-15 seconds. Leave the slide rack in the second D-Limonene staining dish for a total time of 2 minutes. Be sure all paraffin is removed; otherwise, leave the rack in the second D-Limonene staining dish for about 1 minute longer.
  - Move the slide rack from the second D-Limonene staining dish to the EtOH wash. Gently agitate the slide rack back and forth for 10-15 seconds and remove after 2 minutes.
  - Let slides air dry for 5-10 minutes or until completely dry, and the tissue appears white (this may take longer depending on the size of the tissue).
  - Outline the tumor area on the backside of the unstained slide(s) by aligning it with the corresponding H&E stained slide and transposing the outlined area.
  - Working with one slide at a time, rehydrate tissue on the outlined unstained slide by dipping the slide in the 3% glycerol solution.
  - Remove any excess glycerol from the slide with a lab tissue.
  - When processing multiple slides, the user may allow the slides to dry on a drying rack while rehydrating the other slides.
  - Scrape away any non-tumor tissue surrounding the outlined tumor area with a razor blade or scalpel and discard.
  - Holding one end of the slide and resting the other end on a solid surface at a 45 degree angle, collect the macro-dissected tumor tissue onto the edge of a razor blade. The tissue should easily "curl" onto the razor blade as it is being collected.
  - Repeat previous step for each slide from the same specimen.
- Note:** Multiple unstained slides from a single FFPE specimen can be collected onto the same razor blade.

16. Gently slide the tissue sections from the same specimen into a labeled 1.5 mL microcentrifuge tube.
17. If used, clean the dissecting needle or forceps by dipping in D-Limonene for a few seconds and drying between tissue samples.

### 11.5 Isolation of RNA

NanoString recommends use of the Roche FFPE RNA Isolation Kit, which has been validated specifically for use with Prosigna.

Other RNA isolation kits may be used to prepare samples for Prosigna if they yield RNA from slide mounted FFPE breast tumor tissue sections which meets the following specifications:

Table 6: RNA Isolation Kit specifications

| Metric                      | Test or Measurement                                                             | Specification                                                                   |
|-----------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| RNA concentration           | Optical Density at 260 nm                                                       | $\geq 12.5$ ng/ $\mu$ L                                                         |
| RNA total volume ( $\mu$ L) | Total eluted volume                                                             | $\geq 12$ $\mu$ L                                                               |
| RNA purity                  | Ratio of optical density at 260 nm to optical density at 280 nm (OD 260/280 nm) | 1.7–2.3                                                                         |
| DNA Contamination           | Genomic DNA content of eluted RNA sample                                        | $\leq 1$ ng/ $\mu$ L                                                            |
| RNA Integrity               | Size distribution of the isolated RNA fragments                                 | $\geq 90\%$ of the isolated RNA fragments must be $> 100$ nucleotides in length |

#### RNA Isolation Procedure:

Isolate RNA from the macro-dissected tumor tissue according to the manufacturer's instructions if using a kit other than the Roche FFPE RNA Isolation Kit.

The following instructions pertain to the Roche FFPE RNA Isolation Kit recommended for use with Prosigna. Please reference the Roche FFPE RNA Isolation Kit Package Insert for contents information and storage, safety, and handling instructions.

#### Preparation of Working Solutions

1. Prepare working solutions before proceeding with tissue digests and RNA isolation:

| Preparation of Working Solutions |                                                                                                                                                             |
|----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reagent                          | Procedure                                                                                                                                                   |
| Wash Buffer I (WB1)*             | Add 15 mL 100% Ethanol, store prepared WB1 at +15 to +25°C.                                                                                                 |
| Wash Buffer II (WB2)*            | Add 80 mL 100% Ethanol, store prepared WB2 at +15 to +25°C.                                                                                                 |
| Proteinase K (PK)                | Dissolve the lyophilized Proteinase K in 4.5 mL Reagent Preparation Buffer (RPB). Prepare 600 $\mu$ L aliquots, label and store aliquots at -15 to -25°C.   |
| DNase I (DNase I)*               | Dissolve the lyophilized DNase I in 740 $\mu$ L of Reagent Preparation Buffer (RPB). Prepare 50 $\mu$ L aliquots, label and store aliquots at -15 to -25°C. |

**Note:** Items marked with an asterisk (\*) are not required until after tissue has been digested.

#### Tissue Digestion

1. Thaw a sufficient number of Proteinase K (PK) aliquots. Once thawed, store the aliquots on ice until ready to use.  
**Note:** One 600  $\mu$ L PK aliquot contains enough Proteinase K for 4 tissue digests.
2. Add 100  $\mu$ L of Tissue Lysis Buffer (TLB), 16  $\mu$ L of 10% SDS, and 120  $\mu$ L of Proteinase K working solution to sample tubes with tissue.
3. Vortex sample tubes with tissue for several seconds and spin down briefly.
4. Incubate at 55°C overnight (12 hours – 23 hours).

#### RNA Isolation

1. Visually inspect the sample tubes to determine if the tissue is completely digested. A complete digest will appear clear and have little to no tissue in solution.
  - a. If digestion is complete, proceed to next step.

- b. If digest is incomplete, add 20  $\mu$ L of additional Proteinase K and incubate for an additional hour.
2. Briefly spin down all samples ( $< 30$  seconds in a mini centrifuge).
3. Incubate the samples at 80°C for 15 minutes.
4. Prepare or thaw a sufficient number of DNase aliquots. Once thawed, store the aliquots on ice until ready to use.  
**Note:** One 50  $\mu$ L DNase aliquot contains enough DNase for 4 RNA isolations.
5. Dilute the DNase in DNase Incubation Buffer (DIB). The required amount for N (number of) RNA isolations is:  $(N+1) \times 90$   $\mu$ L DNase Incubation Buffer +  $(N+1) \times 10$   $\mu$ L DNase. Store the diluted DNase solution on ice until ready to use.
6. Add 325  $\mu$ L Paraffin Binding Buffer (PBB) and 325  $\mu$ L absolute Ethanol to each sample and mix by pipetting.
7. Spin down briefly ( $< 30$  seconds in a mini centrifuge).
8. Place a High Pure Filter Tube onto a High Pure Collection Tube.
9. Pipette the sample into the upper reservoir of the filter tube, avoiding any residual tissue.
10. Centrifuge for 30 seconds at 6,000  $\times g$ .
11. Place High Pure Filter Tube onto a new High Pure Collection tube (discard former collection tube containing the flow-through).
12. Centrifuge for 2 minutes at 16,000  $\times g$  to completely dry the filter fleece.
13. Place High Pure Filter Tube onto new High Pure Collection Tube (discard former collection tube containing the flow-through).
14. Add 100  $\mu$ L diluted DNase solution onto the High Pure Filter Tube filter fleece and incubate for 15 minutes at +15 to +25°C.
15. Add 500  $\mu$ L Wash Buffer I working solution (WB1) to High Pure filter fleece, centrifuge for 20–30 seconds at 6,000  $\times g$ , discard flow through.
16. Add 500  $\mu$ L Wash Buffer II working solution (WB2) to High Pure filter fleece, centrifuge for 20–30 seconds at 6,000  $\times g$ , discard flow through.
17. Add 500  $\mu$ L Wash Buffer II working solution (WB2) to High Pure filter fleece, centrifuge for 20–30 seconds at 6,000  $\times g$ , discard flow through.
18. Centrifuge for 2 minutes at 16,000  $\times g$  to dry filter fleece completely.
19. Place the High Pure Filter Tube into a labeled RNase-free 1.5 or 1.7 mL microcentrifuge tube.
20. Add 30  $\mu$ L Elution Buffer (EB) to the center of the High Pure filter fleece.
21. Incubate for 1 minute at +15 to +25°C.
22. Centrifuge for 1 minute at 6,000  $\times g$  to elute RNA from column. Remove and discard the High Pure filter tube.
23. Centrifuge the eluted RNA in the microcentrifuge tube for 2 minutes at maximum speed.
24. Transfer the supernatant to a 0.5 mL screw-cap tube without disturbing any glass fibers that might have washed off of the filter fleece at the bottom of the original tube.
25. Measure the concentration of the isolated RNA within the same working day (store at +2 to +8°C) or freeze at -70°C or below until use.

### 11.6 Measure the RNA Concentration and Quality

1. Measure the optical density (OD) at 260 and 280 nm of 2  $\mu$ L of the isolated RNA using a spectrophotometer that meets the defined specifications indicated in 10.3 Equipment Specifications. Avoid pipetting the 2  $\mu$ L volume from the bottom of the source tube in case any glass fibers remain, which will interfere with the optical density reading.
2. Follow the spectrophotometer manufacturer's instructions for measuring RNA.
3. If any sample fails to meet the minimum RNA purity or concentration metrics (Table 6), centrifuge the sample tube for 1 minute at maximum speed ( $> 10,000 \times g$ ), place the tube on ice and repeat the measurement process. If the sample continues to fail either the purity or concentration metric, the RNA sample is not suitable for analysis under the Prosigna Assay procedure. Do not use RNA of insufficient quality or quantity in the Prosigna Assay.
4. The RNA extraction may be repeated if the minimum concentration or minimum purity specifications are not met (Table 6). Users can choose to isolate additional slides from the same FFPE block or choose a separate block from the same patient.

5. If the RNA concentration exceeds 250 ng/ $\mu$ L it must be diluted with molecular grade RNase- and DNase-free water to a target concentration of 200 ng/ $\mu$ L prior to performing the downstream hybridization assay. Use the recorded OD 260/280 ratio result from the undiluted sample to determine if the diluted sample meets the minimum RNA purity of 1.7.
6. Freeze the RNA at  $-70^{\circ}\text{C}$  or below if the Prosigna Assay cannot be completed within the same work day.

### 11.7 Assay Procedure

This assay procedure describes the steps necessary to perform the Prosigna Assay using the NanoString nCounter Dx Analysis System. These steps can be summarized into the following categories on two consecutive days:

#### Day One

- Setup Run Set Identification (RSID) record on web application
- Setup of RNA hybridization with Prosigna CodeSet (30 minute setup, 15–21 hour hybridization)

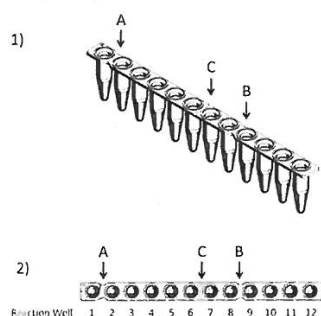
#### Day Two

- Setup and running of Prep Station (20 minutes of setup, 2–3 hours per run, depending on the number of samples run)
- Setup and scanning of Cartridge on the Digital Analyzer (5 minutes for setup, 2.5–4.5 hours for each Cartridge, depending on the number of samples run)
- Retrieving Report (30 minutes)

#### Patient Sample Selection and Batch Set Up

1. Determine the patient samples that will be part of the test run. Up to 10 samples can be included in a single batch
  - a. Each sample within the batch will be assigned to a unique position within a 12-well strip tube used for hybridization, which is registered as part of the Run Set ID on the instrument (Run Set ID performed through web application software). Note that positions 1 and 2 are reserved for the Reference Sample, and positions 3–12 are for tumor RNA samples.
  - b. The illustration below shows side 1) and top views 2) of the strip tube. The strip tubes are asymmetrically keyed between reaction wells 1 and 2 (A) and 8 and 9 (B) to help maintain sample order during processing. The strip tubes are also notched between reaction wells 6 and 7 (C) to facilitate cutting the strip tube if necessary to accommodate standard centrifuge adapters.

Figure 6: Illustration of keyed strip tubes



2. Calculate the amount of RNA and water (when required) to be added to the hybridization reaction for each sample within the batch.
  - a. The recommended RNA input is 250 ng for the assay. The acceptable RNA input range for hybridization is 125–500 ng.
  - b. Calculate the volume (in microliters) of RNA sample to add to the hybridization reaction by dividing the desired sample input (e.g., 250 ng) by the measured concentration.
  - c. If the calculated concentration of the sample is between 12.5 ng/ $\mu$ L and 25 ng/ $\mu$ L, add the maximum volume of 10  $\mu$ L.

- d. For samples that require less than 10  $\mu$ L, calculate the volume of water required to generate 10  $\mu$ L total sample volume.

**Example:** For a sample with measured RNA concentration of 85 ng/ $\mu$ L, 2.9  $\mu$ L of sample is needed for a total mass of 250 ng and 7.1  $\mu$ L water is required to bring the volume to 10  $\mu$ L before adding the remaining reagents. In equation:  $250 \text{ ng} \div 85 \text{ ng}/\mu\text{L} = 2.9 \mu\text{L}$

#### Sample Registration and Processing

The user will build a unique Run Set ID for each batch of samples associating the sample IDs with strip tube location (positions 3–12) using the nCounter Dx Analysis System Services web application. The user may refer to the User Manual for instructions on using the nCounter Dx Analysis System Services web application.

1. If the RNA was frozen prior to use, perform the following steps before proceeding:
    - a. Completely thaw RNA samples and store on ice.
    - b. Centrifuge the thawed sample tube for 1 minute at maximum speed ( $> 10,000 \times g$ ) and place back on ice.
  2. Choose the appropriate Prosigna test kit size based on the number of patient samples being tested (1, 2, 3, 4 or 10). Remove a tube of each of the following CodeSet kit reagents from the  $-80^{\circ}\text{C}$  freezer to thaw. Store the reagents on ice if not proceeding immediately with the subsequent steps.
    - a. Prosigna Reporter CodeSet (Green sticker on cap)
    - b. Prosigna Capture ProbeSet (Grey sticker on cap)
    - c. Prosigna Reference Sample (No sticker on cap)
  3. Remove the CodeSet lot barcode sticker and a test configuration code from the CodeSet box.
  4. Using a web browser, log into the IVD nCounter Dx Analysis System web application and select Prosigna as the assay type to begin setting up the digital registration forms.
  5. On the Main Page select "Create New Run Set".
  6. The first required field in defining a Prosigna run is the Run Set ID. Enter a unique identifier in the Run Set ID field to identify the batch of samples.
  7. Scan or manually enter the Test Configuration Code into the web application. Once this is scanned or entered, it can be discarded.
  8. Scan or manually enter the CodeSet Kit Number into the web application.
  9. Next, enter the unique sample ID for the sample that will be located in the third position/well of the strip tube into the corresponding sample ID field.
    - a. Enter patient RNA Sample IDs using a barcode scanner or manually by entering sample IDs using a keyboard.
    - b. After entering in each sample ID, tab over to fill in the required drop down fields (gross tumor size and node status) for the sample before entering the next sample.
      - i. Use the number of positive nodes established during the patient's pathological assessment to select the appropriate nodal category for the test (zero, 1–3,  $\geq 4$ ).
      - ii. Use the measured gross tumor size or stage established during the patient's pathological assessment to select the appropriate gross tumor size category for the test ( $\leq 2 \text{ cm}$  or  $> 2 \text{ cm}$ ).
    - c. Any comments may be entered into the optional Memo field for each sample.
- Note:** If any strip tube wells/positions are not required, leave the remaining fields blank. If additional fields are required for more samples, use a different test configuration that accommodates more samples.
10. After completing sample entry, specify which users receive the following:
    - a. Status updates for the Prep Station and Digital Analyzer runs.
    - b. The notification that the final report is available.
  11. Save the completed Run Set.
    - a. The Run Set Worksheet may be printed and used for sample traceability and verification purposes.

### Hybridization Reaction Procedure

**Note:** The following steps assume ten (10) patient samples and two (2) reference samples.

**Note:** Do not spin Reporter CodeSet faster than  $3,000 \times g$  or for more than 10 seconds and do not "pulse" it to spin. Doing so will cause the centrifuge to reach maximum speed and may spin the CodeSet out of solution.

1. Program the heat block using 30  $\mu\text{L}$  volume, calculated block and lid temperature, and "forever" time setting (or equivalent hold time setting). Set the temperature of the heat block to  $65^\circ\text{C}$  and set the heated lid at  $70^\circ\text{C}$ .

**Note:** For the following steps, it is critical to maintain the order with which the samples are added to the strip tube, ensuring they match the order on the Run Set ID.

2. Label the provided keyed, 12-well strip tube to distinguish positions 1-6 from positions 7-12 (see strip tube illustration).
3. If necessary, cut the strip tube in half so it will fit into a mini-centrifuge with a strip tube adapter.
4. Pipette 10  $\mu\text{L}$  of Reference Sample into positions 1 and 2 of the keyed strip tube.
5. Pipette the calculated volume of water required for each sample into the respective positions in the keyed strip tube.
6. Pipette the calculated volume of RNA required for each sample into the appropriate position in the keyed strip tube, using a fresh pipette tip for each sample.
7. Once the patient sample has been added to the strip tube, it is recommended to place the sample tube in a sample tube rack, maintaining the order with which the sample was added to the strip tube. This is for verification that samples were added in the intended order after all samples are added to the strip tube.
8. Once all samples have been added to the strip tube, verify that the sample order was maintained in the final layout (refer to nCounter Dx Analysis System User Manual for instructions on editing an existing Run Set ID).

- a. If needed, edit the Run Set ID using the Web Application software to reflect the sample order in the final layout (refer to nCounter Dx Analysis System User Manual for instructions on editing an existing Run Set ID).

9. After the sample order has been verified, place the individual RNA sample tubes back on ice.

10. Create a master mix containing 130  $\mu\text{L}$  of the Hybridization Buffer and 65  $\mu\text{L}$  of the Reporter CodeSet.

**Note:** If the Reporter CodeSet was stored on ice, allow it to equilibrate to room temperature for 1 minute before adding Hybridization Buffer.

11. Mix by pipetting and briefly spin down the master mix.

**Note:** Do not add master mix on ice.

12. Pipette 15  $\mu\text{L}$  of master mix into each of the 12 wells. Use a fresh pipette tip for each well.

**Note:** After completion of the next step, the strip tube must be placed in the heat block at  $65^\circ\text{C}$  within 15 minutes.

13. Add 5  $\mu\text{L}$  Capture ProbeSet to each well, using a new pipette tip for each well.

14. Cap the strip tube wells and mix the reagents by inverting the strip tube several times and flicking with a finger to ensure complete mixing.

15. Briefly spin down the samples in the strip tube in a picofuge or mini-centrifuge (at  $< 3000 \times g$ ).

**Note:** Use a picofuge that can accommodate a 12-well strip tube, or if necessary a mini-centrifuge that can accommodate cut strip tubes.

16. Place the strip tube(s) in a  $65^\circ\text{C}$  heat block with a heated lid. Incubate hybridization assays at  $65^\circ\text{C}$  for 15–21 hours. Hybridizations should be left at  $65^\circ\text{C}$  until ready for processing on the Prep Station.

**Note:** Discard any unused CodeSet.

### Processing Samples on the nCounter Prep Station

1. Locate the Prep Station associated with the Digital Analyzer.
2. Remove nCounter Cartridge from  $-20^\circ\text{C}$  storage and let equilibrate to room temperature for 10–15 minutes in the foil pouch.

**Note:** Ensure that components from the same kit lot are used together.

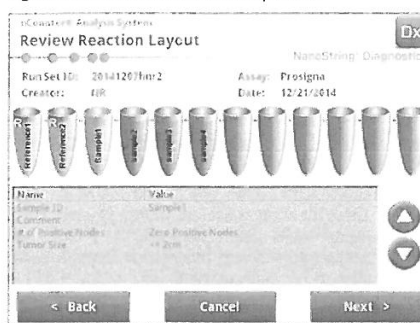
3. When the Cartridge has reached room temperature, remove it from the foil pouch prior to loading the Cartridge onto the Prep Station deck.
4. Remove the nCounter Prep Plate(s) from  $4^\circ\text{C}$  storage and let

equilibrate to room temperature for 10–15 minutes.

**Note:** Only one Prep Plate is required for runs performed using a 1, 2, 3, or 4-test Prosigna kit.

5. Centrifuge the Prep Plates at  $2000 \times g$  for 2 minutes to collect liquids at the bottom of the wells prior to loading the Prep Plates onto the Prep Station deck.
6. While the Cartridges and Prep Plate(s) come to room temperature, prepare the Prep Station with the nCounter Prep Pack consumables.
7. Using the nCounter Prep Station touch screen interface select the "Diagnostics" button for your assay.
8. Under the Main Menu screen, select the "Process Samples" button on the touch screen interface.
9. Browse the list of available Run Set IDs (RSID) illustrated on the screen to confirm the RSID for the samples currently being processed.
10. Select the RSID by touching the screen and select "Next" on the touch screen interface.
11. On the touch screen interface verify the appropriate RSID was selected by looking at each tube on the screen and cross-referencing the sample information.
  - a. Run Set Worksheet may be used here for sample traceability and verification.
  - b. If the incorrect RSID was selected, touch the "Back" button and select the correct RSID.
  - c. If the RSID was correct but sample entry errors exist, touch the "Back" button and go to a computer work station and edit the RSID through the web application.

Figure 7: Process a Run screen on Prep Station



12. In the following several screens you will be prompted to scan the requested Reagent Barcode IDs in the open fields or confirm the deck placement of the required consumables. After performing each task, select "Next" on the touch screen interface to move to the next prompt.

**Note:** Only one Prep Plate and one empty heater strip tube are required for runs performed using a 1, 2, 3, or 4-test Prosigna kit. For 1, 2, 3, or 4-test kit runs, load the Prep Plate and empty heater strip tube into their respective front positions (nearest the user) on the Prep Station deck.

13. Remove samples from the heat block.

**Note:** Initiate the Prep Station run within 15 minutes of removing the samples from the heat block.

14. Place the strip tube(s) in a picofuge or mini-centrifuge and spin down briefly (at  $< 3000 \times g$ ).

15. Gently remove tube caps from the strip tube(s).

16. The notches on the strip tube and guides on the Prep Station should maintain correct order and orientation for the samples.

17. Place the strip tube(s) with the wells in order 1-12 from left-to-right on the nCounter Prep Station deck. If performing a run using a 1, 2, 3, or 4-test kit, only the first half of the strip tube (wells 1-6) needs to be placed in the left side of the sample tube holder on the deck, when applicable. Please note that it is important to use only wells 1-6; the second half of the strip tube (wells 7-12) will not fit into the left half of the holder due to the notched design of the tube.

18. Confirm the strip tubes are seated firmly on the Prep Station and close the metal lid.

19. If the lid does not close properly, you will be prompted to close it

- during the validation of the deck layout.
20. Select "Next" on the touch screen interface.
  21. Close the instrument door when prompted and select "Next" to start the deck layout validation.
  22. If an error occurs, follow the directions associated with the specific error to continue with the deck layout validation.

Figure 8: Post-Hybridization Deck Layout Validation Prep Station



23. After the deck layout has been validated, select "Start Processing" on the touch screen interface.
- Note:** If you encounter problems starting the Prep Station, return your hybridized samples to the heat block, but do not exceed the maximum time of 21 hours.
24. Follow prompts on the Prep Station once the run is complete.
25. Once the Prep Station has completed the run, carefully remove the Cartridge from the Prep Station and seal the wells of the Cartridge with the provided Cartridge well adhesive cover.
- Note:** Do not let the Cartridge remain unsealed on the Prep Station overnight.
26. If you will not be scanning the samples within the same day, store the Cartridge at 4°C in an opaque box for up to 1 week.

#### Scanning Cartridge on the nCounter Digital Analyzer

1. Locate the Digital Analyzer that is linked to the Prep Station that processed the samples. Load the Cartridge onto the nCounter Digital Analyzer for scanning.
  - a. Open the door to the Digital Analyzer.
  - b. Place the Cartridge to be added in an empty slot.
  - c. Close door and refer to the touch screen display.
2. The Digital Analyzer touch screen interface has several different graphics to help quickly identify the positional status:
  - a. Empty Location: This slot is empty and ready to be loaded with a new Cartridge.
  - b. Complete Blue Cartridge: Completed scan.

**DO NOT REMOVE THE FOLLOWING CARTRIDGES:**

  - a. White Cartridge: This slot contains a Cartridge that is registered but un-scanned.
  - b. Partial Blue Cartridge: This slot contains a Cartridge that is in the process of scanning
3. Cartridges that have finished scanning may be removed from the Digital Analyzer.
4. If this is the first Cartridge loaded on the Digital Analyzer touch the "Diagnostics" button, then select "Main Menu" to log in on the Digital Analyzer. If the Digital Analyzer is already scanning Cartridges, proceed to step 9 below.
5. Carefully place the Cartridge into a vacant slot (see positional status guide above) on the Digital Analyzer. The slot and Cartridge are keyed to help ensure the correct orientation. The barcode will be facing up.
6. Lower the slot cover and press on the Cartridge through the opening in the slot lid to ensure that the Cartridge is seated properly.
7. Touch the "Start Counting" button and wait for the scanner to begin the scanning process. You will hear a series of small rhythmic clicking sounds as the Digital Analyzer begins to scan the Cartridge.
8. Confirm that a blue bar appears in the Cartridge position on the screen (within five minutes of starting the scan) indicating that the scan has started.
9. To add a Cartridge to a Digital Analyzer that is already scanning

Cartridges, touch "Pause" on the 'Counting Cartridges' screen and wait for the Digital Analyzer to pause the current scan.

10. Open the door to the Digital Analyzer.
11. Place the Cartridge to be added in an empty slot (see positional status guide above).
12. Close the door and touch "Resume".
13. When the scan is complete the software will send the report to the previously specified user email addresses.
14. Upon receiving the email notification, remove the completed Cartridge and dispose according to your institution's guidelines.
- Note:** Reports will be generated for successfully completed runs as well as runs with errors associated with data Quality Control (QC). Reports will not be generated in the event of an error that is not associated with data QC. Contact NanoString Customer Service for assistance if this occurs.
15. Using the link attached to that email, open the web application and download all the test reports associated with the RSID currently being processed.
16. Following Errors: Follow the recommendation outlined in the test report for individual sample or System error.

**Note:** Individual sample failures are not considered System errors.

## 12 TROUBLESHOOTING AND TEST FAILURES

Table 7: Error Codes for Test Repeat

| Error Code | Failure Description | Recommended Action                                |
|------------|---------------------|---------------------------------------------------|
| 5          | Scan Failure        | Re-run sample with 250 ng of RNA                  |
| 7          | High Signal         | Re-spec sample and re-run test with 125 ng of RNA |
| 6          | Low Signal          | Re-spec sample and re-run test with 500 ng of RNA |
| 30         | Low Signal          | Re-spec sample and re-run test with 500 ng of RNA |
| 31         | Low RNA Signal      | Re-spec sample and re-run test with 500 ng of RNA |

#### Reasons to Repeat the Assay:

1. The assay report will identify failing samples and no assay results will be reported. The assay results will be reported in the case of passing samples.
2. The assay report will identify the type of failure and recommended action in the case of a assay failure. The RNA concentration of failing samples may be remeasured and the samples may be re-run (as part of a new batch/RSID) depending on the type of failure and amount of RNA mass that remains in order to obtain an assay result.

## 13 RESULTS OF THE ASSAY

The Prosigna Assay includes a series of quality control metrics that are automatically applied to each sample during analysis. These metrics evaluate the performance of the assay to determine whether the results fall within expected values. Upon successful analysis of these quality control metrics, the Prosigna Assay gives the following results:

Table 8: Results and output of the Prosigna Assay

| Result                                                                       | Output Values                                           |
|------------------------------------------------------------------------------|---------------------------------------------------------|
| The Intrinsic Subtype of the Breast Cancer Specimen                          | Luminal A,<br>Luminal B,<br>HER2-Enriched<br>Basal-Like |
| Individual Estimate of the Probability of Distant Recurrence within 10 years | 0-100%                                                  |
| Risk of Recurrence (ROR) Score                                               | Integer value on a 0-100 scale                          |
| Risk Category                                                                | Low, Intermediate, High                                 |

### 13.1 Intrinsic Subtypes

The Intrinsic Subtype of a breast cancer tumor has been shown to be related to prognosis in Early Stage Breast Cancer. On average, patients with a Luminal A tumor have significantly better outcomes than patients with Luminal B, HER2-Enriched, or Basal-like tumors<sup>25</sup>.

The Intrinsic Subtype is identified by comparing the gene expression profile of 50 genes in an unknown sample with the expected expression profiles for the four intrinsic subtypes. The subtype with the most similar profile is assigned to the unknown sample.

The most common subtypes of breast cancer are the luminal subtypes,

Luminal A (LumA) and Luminal B (LumB). Prior studies suggest that Luminal A comprises approximately 30% to 40% and Luminal B approximately 20% of breast cancers<sup>5</sup>. However, greater than 90% of hormone-receptor positive patients have luminal tumors. The gene expression pattern of these subtypes resembles the luminal epithelial component of the breast tissue<sup>5</sup>. These tumors are characterized by high expression of estrogen receptor (ER), progesterone receptor (PR), and genes associated with ER activation, such as *LIV1*, *GATA3*, and *cyclin D1*, as well as expression of luminal cytokeratins 8 and 18. Luminal A breast cancers exhibit lower expression of genes associated with cell cycle activation when compared to Luminal B breast cancers resulting in a better prognosis.

Prior studies suggest that the HER2-Enriched subtype (HER2-E) comprises approximately 20% of breast cancers<sup>5</sup>. However, HER2-Enriched tumors are generally ER-negative, so only 5% of the tested ER-positive patient population was found to have HER2-Enriched breast cancer. Regardless of ER-status, HER2-Enriched tumors are HER2-positive in the majority of cases with high expression of the ERBB2 cluster, including *ERBB2* and *GRB7*. Genes associated with cell cycle activation are also highly expressed.

Published data suggest that the Basal-like subtype comprises approximately 20% of breast cancers<sup>5</sup>. However, Basal-like tumors are generally ER-negative, so only 1% of hormone receptor-positive patients have Basal-like breast cancer. The Basal-like subtype is almost always clinically HER2-negative and expresses a suite of "basal" biomarkers including the basal epithelial cytokeratins (CK) and Epidermal Growth Factor Receptor (EGFR). Genes associated with cell cycle activation are highly expressed.

### 13.2 ROR Score

The ROR score is an integer value on a 0–100 scale that is related to an individual patient's probability of distant recurrence within 10 years for the defined intended use population. The ROR score is calculated by comparing the expression profile of 46 genes in an unknown sample with the expected profiles for the four intrinsic subtypes, as described above, to calculate four different correlation values. These correlation values are then combined with a proliferation score and the gross tumor size to calculate the ROR score.

### 13.3 Probability of 10-Year Distant Recurrence

The ROR scores for 2 cohorts of post-menopausal women with hormone receptor-positive early stage breast cancer were compared to distant recurrence-free survival following surgery and treatment with 5 years of adjuvant endocrine therapy followed by 5 years of observation (see Clinical Performance Section 16.4 for details). These two studies resulted in a model relating the ROR score to the probability of distant recurrence in this tested patient population including a 95% confidence interval.

### 13.4 Risk Classification

Risk classification is also provided to allow interpretation of the ROR score by using cutoffs related to clinical outcome in tested patient populations.

Table 9: Risk classification by ROR range and nodal status

| Nodal Status              | ROR Range | Risk Classification |
|---------------------------|-----------|---------------------|
| Node-Negative             | 0–40      | Low                 |
|                           | 41–60     | Intermediate        |
|                           | 61–100    | High                |
| Node-Positive (1–3 nodes) | 0–15      | Low                 |
|                           | 16–40     | Intermediate        |
|                           | 41–100    | High                |
| Node-Positive (≥ 4 nodes) | 0–100     | High                |

### 13.5 Quality Control

Each lot of the Prosigna Assay components is tested using predetermined specifications. All kit-level items are lot tracked, and the critical components contained within each kit are tested together and released as a Prosigna kit lot.

The Prosigna Assay kit includes a series of internal controls that are used to assess the quality of each run set as a whole and each sample individually. These controls are listed below.

#### Batch Control Set: *In vitro* transcribed RNA Reference Sample

A synthetic RNA Reference Sample is included as a control within the NanoString Prosigna Assay kit. The Reference Sample is comprised of *in-vitro* transcribed RNA targets from the 50 algorithm and 8 housekeeping genes. The Reference Sample is processed in duplicate in each Prosigna Assay run along with a set of up to 10 unknown breast tumor RNA samples in a 12 reaction strip

tube. The signal from the Reference Sample is analyzed against pre-defined thresholds to qualify the run.

The signal from each of the 50 algorithm genes of the breast tumor RNA sample is normalized to the corresponding genes of the Reference Sample.

#### Positive Control Set: *in vitro* transcribed RNA targets and corresponding Capture and Reporter Probes

Synthetic RNA targets are used as positive controls (PCs) for the Prosigna Assay. The PC target sequences are derived from the External RNA Control Consortium (ERCC) DNA sequence library<sup>6</sup>. The RNA targets are *in-vitro* transcribed from DNA plasmids. Six RNA targets are included within the assay kit in a 4-fold titration series (128–0.125 fM final concentration in the hybridization reaction) along with the corresponding Capture and Reporter Probes. The PCs are added to each breast tumor RNA sample and Reference Sample tested with the Prosigna Assay. A sample will be disqualified from further analysis if the signal intensities from the PCs do not meet pre-defined thresholds.

#### Negative control set: exogenous probes without targets

Negative control target sequences are derived from the ERCC DNA sequence library<sup>6</sup>. The probes designed to detect these target sequences are included as part of the assay kit without the corresponding target sequence. The negative controls (NCs) are added to each breast tumor RNA sample and Reference Sample tested with the Prosigna Assay as a quality control measure. The sample will be disqualified from further analysis if the signal intensities from the NCs do not meet pre-defined thresholds.

#### RNA Integrity Control Set: Housekeeping genes

Capture and Reporter Probes designed to detect 8 housekeeping genes and 50 algorithm genes are included as part of the Prosigna kit. The expression levels of the 8 housekeeping genes are analyzed to determine the quality of RNA extracted from the FFPE tissue sample and input into the Prosigna Assay. The sample will be disqualified from further analysis if the expression levels of the housekeeping genes falls below pre-defined thresholds.

The housekeeping genes are also used to normalize for any differences in the intact RNA amount in a sample prior to Reference Sample normalization.

## 14 LIMITATIONS OF THE PROCEDURES

1. The Prosigna Assay has been optimized to identify the intrinsic subtype of a breast cancer tumor and the patient's 10-year risk of distant recurrence as an ROR score and risk category, using purified RNA extracted from formalin-fixed, paraffin embedded human breast tissue. Other types of specimens or fixatives have not been tested and should not be used.
2. The performance of the Prosigna Assay was validated using the procedures provided in this package insert only. Modifications to these procedures may alter the performance of the test.
3. Performance characteristics of the Prosigna Assay have been established for postmenopausal women with hormone receptor-positive early stage breast cancer treated with 5 years of adjuvant endocrine therapy. Performance with other treatment regimens or in other patient populations has not been established.
4. If RNA of insufficient quality or quantity is added to the assay, then the Prosigna Assay may be unable to give a valid result and instead will report an assay failure.
5. The interpretation of Prosigna Assay results (intrinsic subtype, ROR score, risk category) should be evaluated within the context of other clinicopathological factors, the patient's medical history and any other laboratory test results.
6. The performance of the Prosigna Assay has been established with RNA meeting the specifications defined in reference to the procedure above. Performance with isolated RNA that does not meet these specifications has not been established.
7. Known interfering substances to the Prosigna Assay include genomic DNA and non-tumor tissue (e.g., normal tissue). Please refer to general assay considerations before starting the procedure. The area of viable invasive carcinoma must be clearly identified by a pathologist prior to running the procedure. Additionally, each RNA sample must be treated with DNase. Prior to proceeding with patient test samples, each new lot of DNase should be tested and qualified per the provided specification when using an isolation kit other than the Roche FFPE RNA isolation kit.

## Appendix H: Purification of Genomic DNA and Total RNA from FFPE Tissue

# Protocol: Purification of Genomic DNA and Total RNA (Including Small RNAs) from FFPE Tissue Sections

### Important points before starting

- If using the AllPrep DNA/RNA FFPE Kit for the first time, read “Important Notes” (page 13).
- Buffer RLT, Buffer FRN, Buffer AL, and Buffer AW1 contain a guanidine salt and are therefore not compatible with disinfecting reagents containing bleach. See page 6 for safety information.
- Unless otherwise indicated, perform all steps of the procedure at room temperature (15-25°C). During the procedure, work quickly.
- Perform all centrifugation steps using a microcentrifuge placed at 15-25°C. If using a refrigerated microcentrifuge, set the temperature to 20-25°C, otherwise significant cooling below 15°C may occur.
- In the procedure below, ▲ indicates specific steps for purification of total RNA that does not include small RNAs, and • indicates specific steps for purification of total RNA that does include small RNAs.

### Things to do before starting

- If using Buffer FRN, Buffer RPE, Buffer AW1, Buffer AW2, and RNase-Free DNase I for the first time, reconstitute them as described in “Preparation of buffers” (page 15).
- If necessary, warm and gently agitate Buffer RLT, Buffer ATL, and Buffer AL to redissolve any precipitates that may have formed.
- Equilibrate all buffers to room temperature (15-25°C). Mix reconstituted Buffer FRN, Buffer RPE, Buffer AW1, and Buffer AW2 by shaking.
- Set a thermal mixer or heated orbital incubator to 56°C for use in step 6. Incubation at 56°C (steps 6 and 26) and at 80°C (step 10) can be done with or without agitation. Incubation at 90°C (step 27) must be done without agitation. If a thermal mixer or heated orbital incubator is not available, a heating block or water bath can be used instead.

### Procedure

1. Using a scalpel, trim excess paraffin off the sample block.
2. Cut sections 10-20 µm thick.

Do not use more than four 10 µm sections of 150 mm<sup>2</sup> surface area or two 20 µm sections of 150 mm<sup>2</sup> surface area.

If the sample surface has been exposed to air, discard the first 2–3 sections.

3. Immediately place the sections in a 1.5 ml Safe-Lock microcentrifuge tube (not supplied), and close the lid.
4. Remove the paraffin according to step 4a, 4b, or 4c.
- 4a. Deparaffinization using Deparaffinization Solution:
  - Add Deparaffinization Solution: for two 10  $\mu\text{m}$  sections or one 20  $\mu\text{m}$  section, add 320  $\mu\text{l}$  Deparaffinization Solution; for more sample material, add 640  $\mu\text{l}$  Deparaffinization Solution.
  - Vortex vigorously for 10 s, and centrifuge briefly to bring the sample to the bottom of the tube.
  - Incubate at 56 °C for 3 min, then allow to cool at room temperature (15-25 °C) and centrifuge at full speed for 2 min.
  - Carefully remove the supernatant by pipetting without disturbing the pellet. Carefully remove any residual Deparaffinization Solution using a fine pipet tip.
  - Keep the lid open, and incubate for 10 min at 37 °C to dry the pellet. Proceed to step 5.
- 4b. Deparaffinization using heptane (may provide more compact sample pellets in step 8):
  - Add 500  $\mu\text{l}$  heptane, vortex vigorously for 10 s, and incubate for 10 min at room temperature (15-25 °C).
  - Add 25  $\mu\text{l}$  methanol, vortex vigorously for 10 s, and centrifuge for 2 min at 9000 x *g*.
  - Carefully remove the supernatant by pipetting without disturbing the pellet. Carefully remove any residual heptane/methanol using a fine pipet tip.
  - Add 1 ml ethanol (96-100%) to the pellet, mix by vortexing, and centrifuge at full speed for 2 min.
  - Carefully remove the supernatant by pipetting without disturbing the pellet. Carefully remove any residual ethanol using a fine pipet tip.
  - Keep the lid open, and incubate at room temperature (15-25 °C) or at up to 37 °C for 10 min or until all residual ethanol has evaporated. Proceed to step 5.
- 4c. Deparaffinization using xylene:
  - Add 1 ml xylene, vortex vigorously for 10 s, and centrifuge at full speed for 2 min.
  - Carefully remove the supernatant by pipetting without disturbing the pellet.
  - Add 1 ml ethanol (96-100%) to the pellet, mix by vortexing, and centrifuge at full speed for 2 min.

The ethanol extracts residual xylene from the sample.

- **Carefully remove the supernatant by pipetting without disturbing the pellet. Carefully remove any residual ethanol using a fine pipet tip.**
  - **Keep the lid open, and incubate at room temperature (15-25 °C) or at up to 37 °C. Incubate for 10 min or until all residual ethanol has evaporated. Proceed to step 5.**
5. **Resuspend the pellet by adding 150  $\mu$ l Buffer PKD and flicking the tube to loosen the pellet. Add 10  $\mu$ l proteinase K, and mix by vortexing.**
  6. **Incubate at 56 °C for 15 min.**  
Depending on the sample material, the sample may not be completely lysed. This does not affect the procedure. Proceed to step 7.
  7. **Incubate on ice for 3 min.**  
Complete cooling is important for efficient precipitation in step 8.
  8. **Centrifuge for 15 min at 20,000 x g.**
  9. **Carefully transfer the supernatant, without disturbing the pellet, to a new ▲ 1.5 ml or • 2 ml Safe-Lock microcentrifuge tube for RNA purification in steps 10-24. Keep the pellet for DNA purification in steps 25-35.**

**Note:** Depending on the amount and nature of the FFPE sample, the pellet may be very small or difficult to see. If the pellet is aspirated with the supernatant, allow the pellet to drop slowly to the bottom of the tube and use the pipet tip to reattach the pellet to the tube. Alternatively, centrifuge the supernatant again.

The DNA-containing pellet can be stored for 2 h at room temperature (15–25°C), for up to 1 day at 2–8°C, or for longer periods at –20°C.

## Purification of total RNA

10. **Incubate the supernatant from step 9 at 80 °C for 15 min.**  
This incubation step partially reverses formaldehyde modification of nucleic acids. Longer incubation times or higher incubation temperatures may result in more fragmented RNA.  
If using only one heating block, keep the supernatant at room temperature until the heating block has reached 80°C. To ensure maximum RNA yields, the supernatant must be incubated at 80°C only for exactly 15 min.
11. **Briefly centrifuge the tube to remove drops from the inside of the lid.**
12. **Add 320  $\mu$ l Buffer RLT to adjust binding conditions, and mix by vortexing or pipetting.**
13. **Add ▲ 720  $\mu$ l or • 1120  $\mu$ l ethanol (96-100%), and mix well by vortexing or pipetting. Proceed immediately to step 14.**

Precipitates may be visible after addition of ethanol. This does not affect the procedure.

14. Transfer 700  $\mu$ l of the sample, including any precipitate that may have formed, to an RNeasy MinElute spin column placed in a 2 ml collection tube (supplied). Close the lid gently, and centrifuge for 15 s at  $\geq 8000 \times g$  ( $\geq 10,000$  rpm). Discard the flow-through.\*  
Reuse the collection tube in step 15.
15. Repeat step 14 until the entire sample has passed through the RNeasy MinElute spin column.  
Reuse the collection tube in step 16.
16. Add 350  $\mu$ l Buffer FRN to the RNeasy MinElute spin column. Close the lid gently, and centrifuge for 15 s at  $\geq 8000 \times g$  ( $\geq 10,000$  rpm). Discard the flow-through.†  
**Note:** Buffer FRN is supplied as a concentrate. Ensure that isopropanol is added before use as described in "Preparing Buffer FRN" (page 16).  
Reuse the collection tube in step 17.
17. Add 10  $\mu$ l DNase I stock solution to 70  $\mu$ l Buffer RDD. Mix by gently inverting the tube, and centrifuge briefly to collect residual liquid from the sides of the tube.  
**Note:** DNase I is supplied lyophilized and should be reconstituted as described in "Preparing DNase I stock solution" (page 16).

\* Flow-through contains Buffer RLT and is therefore not compatible with bleach. See page 6 for safety information.

† Flow-through contains Buffer FRN and is therefore not compatible with bleach. See page 6 for safety information.

**Note:** DNase I is especially sensitive to physical denaturation. Mixing should only be carried out by gently inverting the tube. Do not vortex.

18. Add the DNase I incubation mix (80  $\mu$ l) directly to the RNeasy MinElute spin column membrane, and place on the benchtop (20– 30 °C) for 15 min.  
Note: Be sure to add the DNase I incubation mix directly to the membrane. DNase digestion will be incomplete if part of the mix sticks to the walls or the O-ring of the spin column.
19. Add 500  $\mu$ l Buffer FRN to the RNeasy MinElute spin column. Close the lid gently, and centrifuge for 15 s at  $\geq 8000 \times g$  ( $\geq 10,000$  rpm). Save the flow-through for use in step 20.  
Do not discard the flow-through, as it contains RNA including small RNAs.
20. Place the RNeasy MinElute spin column in a new 2 ml collection tube (supplied). Apply the flow-through from step 19 to the spin column. Close

the lid gently, and centrifuge for 15 s at  $\geq 8000 \times g$  ( $\geq 10,000$  rpm). Discard the flow-through.\*

Reuse the collection tube in step 21.

21. Add 500  $\mu\text{l}$  Buffer RPE to the RNeasy MinElute spin column. Close the lid gently, and centrifuge for 15 s at  $\geq 8000 \times g$  ( $\geq 10,000$  rpm) to wash the spin column membrane. Discard the flow-through.

Reuse the collection tube in step 22.

**Note:** Buffer RPE is supplied as a concentrate. Ensure that ethanol is added before use as described in "Preparing Buffer RPE" (page 16).

22. Add 500  $\mu\text{l}$  Buffer RPE to the RNeasy MinElute spin column. Close the lid gently, and centrifuge for 15 s at  $\geq 8000 \times g$  ( $\geq 10,000$  rpm) to wash the spin column membrane.

Discard the collection tube with the flow-through.

23. Place the RNeasy MinElute spin column in a new 2 ml collection tube (supplied). Open the lid, and centrifuge at full speed for 5 min. Discard the collection tube with the flow-through.

To avoid damage to their lids, place the spin columns into the centrifuge with at least one empty position between columns. Orient the lids so that they point in a direction opposite to the rotation of the rotor (e.g., if the rotor rotates clockwise, orient the lids counterclockwise).

It is important to dry the spin column membrane, since residual ethanol may interfere with downstream reactions. Centrifugation with the lids open ensures that no ethanol is carried over during RNA elution.

\* Flowthrough contains Buffer FRN and is therefore not compatible with bleach. See page 6 for safety information.

24. Place the RNeasy MinElute spin column in a new 1.5 ml collection tube (supplied). Add 14–30  $\mu\text{l}$  RNase-free water directly to the spin column membrane. Close the lid gently, and incubate for 1 min at room temperature (15–25 °C). Centrifuge at full speed for 1 min to elute the RNA.

Elution with smaller volumes of RNase-free water leads to higher total RNA concentrations, but lower RNA yields.

The dead volume of the RNeasy MinElute spin column is 2  $\mu\text{l}$ : elution with 14  $\mu\text{l}$  RNase-free water results in a 12  $\mu\text{l}$  eluate.

## Purification of genomic DNA

25. Resuspend the pellet from step 9 in 180  $\mu\text{l}$  Buffer ATL, add 40  $\mu\text{l}$  proteinase K, and mix by vortexing.

The pellet should be equilibrated to room temperature (15–25°C) prior to resuspension.

**26. Incubate at 56 °C for 1 h.**

**27. Incubate at 90 °C for 2 h without agitation.**

This incubation step partially reverses formaldehyde modification of nucleic acids. Longer incubation times or higher incubation temperatures may result in more fragmented DNA.

If using only one heating block, keep the sample at room temperature until the heating block has reached 90°C.

**Note:** Agitation during this incubation step leads to lower DNA yields.

**28. Briefly centrifuge the microcentrifuge tube to remove drops from the inside of the lid.**

**Optional:** In contrast to commonly used DNA preparation methods, DNA purified using the AllPrep DNA/RNA FFPE procedure contains only trace amounts of RNA, since RNA is efficiently separated from DNA during the procedure. If RNA-free genomic DNA is required, allow the sample to cool to room temperature (15–25°C) and then add 4  $\mu$ l RNase A (100 mg/ml). Incubate for 2 min at room temperature before proceeding to step 29.

**29. Add 200  $\mu$ l Buffer AL, and mix thoroughly by vortexing. Then add 200  $\mu$ l ethanol (96–100%), and mix thoroughly again by vortexing or pipetting.**

It is essential that the sample, Buffer AL, and ethanol are mixed immediately and thoroughly by vortexing or pipetting to yield a homogeneous solution. Buffer AL and ethanol can be premixed and added together in one step to save time when processing multiple samples.

A white precipitate may form on addition of Buffer AL and ethanol. This precipitate does not interfere with the AllPrep procedure.

**30. Transfer the entire sample to a QIAamp MinElute spin column placed in a 2 ml collection tube (supplied). Close the lid gently, and centrifuge for 1 min at  $\geq 8000 \times g$  ( $\geq 10,000$  rpm). Discard the collection tube with the flow-through.\***

If the sample has not completely passed through the membrane after centrifugation, centrifuge again at a higher speed until the QIAamp MinElute spin column is empty.

**31. Place the QIAamp MinElute spin column in a new 2 ml collection tube (supplied). Add 700  $\mu$ l Buffer AW1 to the spin column. Close the lid gently, and centrifuge for 15 sec at  $\geq 8000 \times g$  ( $\geq 10,000$  rpm) to wash the spin column membrane. Discard the flow-through.\***

Reuse the collection tube in step 32.

**Note:** Buffer AW1 is supplied as a concentrate. Ensure that ethanol is added before use as described in "Preparing Buffer AW1" (page 17).

**32. Add 700  $\mu$ l Buffer AW2 to the QIAamp MinElute spin column. Close the lid gently, and centrifuge for 15 sec at  $\geq 8000 \times g$  ( $\geq 10,000$  rpm) to wash the spin column membrane. Discard the flow-through.**

Reuse the collection tube in step 33.

Note: Buffer AW2 is supplied as a concentrate. Ensure that ethanol is added before use as described in "Preparing Buffer AW2" (page 17).

- 33. Add 700  $\mu$ l ethanol (96-100%) to the QIAamp MinElute spin column. Close the lid gently, and centrifuge for 15 sec at  $\geq 8000 \times g$  ( $\geq 10,000$  rpm). Discard the collection tube with the flow-through.**
- 34. Place the QIAamp MinElute spin column in a new 2 ml collection tube (supplied). Open the lid of the spin column, and centrifuge at full speed for 5 min. Discard the collection tube with the flow-through.**

To avoid damage to their lids, place the spin columns into the centrifuge with at least one empty position between columns. Orient the lids so that they point in a direction opposite to the rotation of the rotor (e.g., if the rotor rotates clockwise, orient the lids counterclockwise).

It is important to dry the spin column membrane, since residual ethanol may interfere with downstream reactions. Centrifugation with the lids open ensures that no ethanol is carried over during DNA elution,

- 35. Place the QIAamp MinElute spin column in a new 1.5 ml collection tube (supplied). Add 30-100  $\mu$ l Buffer ATE directly to the spin column membrane. Close the lid gently, and incubate for 1 min at room temperature (15-25°C). Centrifuge at full speed for 1 min to elute the DNA.**

**Important:** Ensure that Buffer ATE is equilibrated to room temperature. If using small elution volumes ( $< 50 \mu$ l), pipet Buffer ATE onto the center of the membrane to ensure complete elution of bound DNA.

QIAamp MinElute spin columns provide flexibility in the choice of elution volume. Choose a volume according to the requirements of the downstream application. The volume of eluate will be up to 5  $\mu$ l less than the volume of Buffer ATE applied to the column.

Incubating the QIAamp MinElute spin column loaded with Buffer ATE for 5 min at room temperature before centrifugation can increase DNA yield.

\* Flow-through contains Buffer AL or Buffer AW1 and is therefore not compatible with bleach. See page 6 for safety information.

**Appendix I: PAM50 Gene list**

Table modified from (172,173)

| Type    | Gene Card | Accession number | Gene Name                                                                                                      |
|---------|-----------|------------------|----------------------------------------------------------------------------------------------------------------|
| Luminal | BAG1      | NM_004323.3      | BCL2-associated athanogene                                                                                     |
| Luminal | BCL2      | NM_000633.2      | B-cell CLL/lymphoma 2, nuclear gene encoding mitochondrial protein                                             |
| Luminal | BLVRA     | NM_000712.3      | Biliverdin reductase A                                                                                         |
| Luminal | CXXC5     | NM_016463.5      | CXXC finger 5                                                                                                  |
| Luminal | ESR1      | NM_000125.2      | Estrogen receptor 1                                                                                            |
| Luminal | FOXA1     | NM_004496.2      | Forkhead box A1                                                                                                |
| Luminal | GPR160    | NM_014373.1      | G protein-coupled receptor 160                                                                                 |
| Luminal | MAPT      | NM_016835.3      | Microtubule-associated protein tau                                                                             |
| Luminal | MDM2      | NM_006878.2      | Mdm2, transformed 3T3 cell double minute 2, p53 binding protein (mouse)                                        |
| Luminal | MLPH      | NM_024101.4      | Melanophilin                                                                                                   |
| Luminal | MMP11     | NM_005940.3      | Matrix metalloproteinase 11 (stromelysin 3)                                                                    |
| Luminal | MYC       | NM_002467.3      | v-myc myelocytomatosis viral oncogene homolog (avian)                                                          |
| Luminal | NAT1      | NM_000662.4      | N-acetyltransferase 1 (arylamine N-acetyltransferase)                                                          |
| Luminal | PGR       | NM_000926.2      | Progesterone receptor                                                                                          |
| Luminal | SLC39A6   | NM_012319.2      | Solute carrier family 39 (zinc transporter), member 6                                                          |
| HER2    | ERBB2     | NM_004448.2      | V-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian) |
| HER2    | FGFR4     | NM_002011.3      | Fibroblast growth factor receptor 4                                                                            |
| HER2    | GRB7      | NM_005310.2      | Growth factor receptor-bound protein 7                                                                         |
| HER2    | TMEM45B   | NM_138788.3      | Transmembrane protein 45B                                                                                      |
| Basal   | ACTR3B    | NM_001040135.1   | ARP3 actin-related protein 3 homolog B (yeast)                                                                 |
| Basal   | CDH3      | NM_001793.3      | Cadherin 3, type 1, P-cadherin (placental)                                                                     |
| Basal   | EGFR      | NM_005228.3      | Epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)             |
| Basal   | FOXC1     | NM_001453.1      | Forkhead box C1                                                                                                |
| Basal   | KRT14     | NM_000526.3      | Keratin 14 (epidermolysis bullosa simplex, Dowling-Meara, Koebner)                                             |
| Basal   | KRT17     | NM_000422.1      | Keratin 17                                                                                                     |
| Basal   | KRT5      | NM_000424.2      | Keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Koebner/Weber-Cockayne types)                          |
| Basal   | MIA       | NM_006533.1      | Melanoma inhibitory activity                                                                                   |
| Basal   | PHGDH     | NM_006623.2      | Phosphoglycerate dehydrogenase                                                                                 |
| Basal   | SFRP1     | NM_003012.3      | Secreted frizzled-related protein 1                                                                            |

| Type          | Gene Card | Accession number | Gene Name                                                   |
|---------------|-----------|------------------|-------------------------------------------------------------|
| Proliferative | ANLN      | NM_018685.2      | Anillin, actin binding protein (scraps homolog, Drosophila) |
| Proliferative | BIRC5     | NM_001168.2      | Baculoviral IAP repeat-containing 5 (survivin)              |
| Proliferative | CCNB1     | NM_031966.2      | Cyclin B1                                                   |
| Proliferative | CCNE1     | NM_001238.1      | Cyclin E1                                                   |
| Proliferative | CDC20     | NM_001255.1      | CDC20 cell division cycle 20 homolog (S. cerevisiae)        |
| Proliferative | CDC6      | NM_001254.3      | CDC6 cell division cycle 6 homolog (S. cerevisiae)          |
| Proliferative | CDCA1     | NM_145697.1      | Cell division cycle associated 1                            |
| Proliferative | CENPF     | NM_016343.3      | Centromere protein F, 350/400ka (mitosin)                   |
| Proliferative | CEP55     | NM_018131.3      | Centrosomal protein 55kDa                                   |
| Proliferative | EXO1      | NM_006027.3      | Exonuclease 1                                               |
| Proliferative | KIF2C     | NM_006845.2      | Kinesin family member 2C                                    |
| Proliferative | KNTC2     | NM_006101.1      | Kinetochore associated 2                                    |
| Proliferative | MELK      | NM_014791.2      | Maternal embryonic leucine zipper kinase                    |
| Proliferative | MKI67     | NM_002417.2      | Antigen identified by monoclonal antibody Ki-67             |
| Proliferative | MYBL2     | NM_002466.2      | v-myb myeloblastosis viral oncogene homolog (avian)-like 2  |
| Proliferative | ORC6L     | NM_014321.2      | Origin recognition complex, subunit 6 like (yeast)          |
| Proliferative | PTTG1     | NM_004219.2      | Pituitary tumor-transforming 1                              |
| Proliferative | RRM2      | NM_001034.1      | Ribonucleotide reductase M2 polypeptide                     |
| Proliferative | TYMS      | NM_001071.1      | Thymidylate synthetase                                      |
| Proliferative | UBE2C     | NM_007019.2      | Ubiquitin-conjugating enzyme E2C                            |
| Proliferative | UBE2T     | NM_014176.1      | Ubiquitin-conjugating enzyme E2T (putative)                 |
| Reference     | ACTB      | NM_001101.2      | Actin, beta                                                 |
| Reference     | GUSB      | NM_000181.1      | Glucuronidase, beta                                         |
| Reference     | MRPL19    | NM_014763.3      | Mitochondrial ribosomal protein L19                         |
| Reference     | PSMC4     | NM_006503.2      | Proteasome (prosome, macropain) 26S subunit, ATPase, 4      |
| Reference     | PUM1      | NM_001020658.1   | Pumilio RNA-binding family member 1                         |
| Reference     | RPLP0     | NM_001002.3      | Ribosomal protein, large, P0                                |
| Reference     | SF3A1     | NM_005877.4      | Splicing factor 3a, subunit 1, 120kDa                       |
| Reference     | TFRC      | NM_003234.1      | Transferrin receptor (p90, CD71)                            |

## Appendix J: Modified Salting Out Procedure for DNA Extraction

### Salting out Procedure for extracting DNA from Human Nucleated Cells in Whole Blood.

#### Solutions Needed:

##### 1. Sucrose-Triton-X Lysis Buffer

|                      | <u>1 000 ml</u> | <u>500 ml</u> | <u>100 ml</u> |
|----------------------|-----------------|---------------|---------------|
| 1M Tris-HCL pH 8.0   | 10 ml           | 5 ml          | 1 ml          |
| 1M MgCl <sub>2</sub> | 5 ml            | 2.5 ml        | 0.5 ml        |
| Triton X-100         | 10 ml           | 5 ml          | 1 ml          |
| distilled water      | 1 litre         | 500 ml        | 100 ml        |

Autoclave and store at 4°C.

Sucrose - *just before use*      109.5g.      54.75 g      10.95 g

Once the sucrose has been added, the solution does not keep, so only make sufficient each batch. Need  $\pm$  70ml/ 10ml sample. Work on ice.

[Sucrose: 7.7g sucrose per 70ml (i.e. one sample)]

##### 2. T 20 E 5

20 mM Tris-HCL

5mM EDTA

pH 8.0

Autoclave and store at room temperature.

##### 3. Saturated NaCl.

Slowly add 40g NaCl to 100ml sterile water until completely saturated.

Filter sterilize. Store at room temperature.

##### 4. Proteinase K Mix.

Stock solutions;

10% SDS made in sterile H<sub>2</sub>O

0.5M EDTA made in sterile H<sub>2</sub>O

20mg/ml Proteinase K

| Stock Solution           | [Final]                | 500µl<br>(1 sample) | 8.5 mL<br>(16 samples+<br>1 extra) | 10.5 mL<br>(20 samples+<br>1 extra) |  |
|--------------------------|------------------------|---------------------|------------------------------------|-------------------------------------|--|
| 10% SDS                  | 1% SDS                 | 50 µl               | 850 µl                             | 1 050 µl                            |  |
| 0.5M EDTA                | 2mM EDTA               | 2 µl                | 34 µl                              | 42 µl                               |  |
| 20 mg/mL<br>Proteinase K | 2mg/mL<br>Proteinase K | 50 µl               | 850 µl                             | 1 050 µl                            |  |
| H <sub>2</sub> O         | H <sub>2</sub> O       | 400 µl              | 6 800 µl                           | 8 400 µl                            |  |

Make fresh each time because you need 500 µl /blood sample. (make up in 15 ml tube)

- 5. TBE Stock Solution (10x).**
- |                    |        |            |          |
|--------------------|--------|------------|----------|
| 108g Tris Base     | 54 g   | Tris       | (Powder) |
| 55g Boric Acid     | 27.5 g | Boric acid | (Powder) |
| 40ml 0.5 EDTA pH 8 | 4.65 g | EDTA       | (Powder) |
- Make up to 1 litre with H<sub>2</sub>O
- 10X TBE**
- Make up to 500 ml with distilled Water
- 6. TE Buffer (10 mM Tris-Cl, 1 mM EDTA)**
- 10 ml 1M Tris-Cl, pH 7.5
- 2 ml 0.5M EDTA, pH 8.0
- Make up to 1 litre with H<sub>2</sub>O
- 7. Plasticware:**
- 2-3X 50 ml tubes per sample
- 2X 15 ml tubes per run
- 6X Blue tips per sample
- 4X yellow tips per sample
- 1-2X white tips per sample
- 1-2X Eppendorfs/sample
- 8. Other things needed**
- Autoclaved beaker for measuring out sucrose
- Sterile magnetic flea for mixing sucrose in lysis buffer.
- Autoclaved beaker for pouring lysis buffer

## METHODS

### Day 1

1. Pre-cool the centrifuge to 4°C.
2. Blood must be collected into EDTA or ACD tubes.
3. Use **±10ml blood**, place in a 50ml tube and mix with **±40ml lysing solution**.  
**NB: Lysing solution must be kept cold during the procedure.** Mix well. Work on ice.
4. Spin for **10 min** at **± 2300 rpm** at 4°C.
5. Decant supernatant and **re-suspend pellet** in **± 20ml lysing** buffer. Work on ice.
6. Place on **ice** or at -20°C for **5 min**.
7. Spin for a further **10 min** at **2300 rpm** at 4°C.
8. Decant supernatant and add **3 ml T20E5** to each tube. Mix well.
9. Add **200µl of 10% SDS** to each and mix.  
 Then, add **500µl proteinase K mix**, mix and place samples at **42°C -50°C overnight**.

Day 2

10. Add **1 ml saturated NaCl** solution to each tube. Mix well and place on **ice for 5 min.**
11. Spin at **± 2300 rpm for 30 min, 10°C.**  
If the supernatant is cloudy, decant into a clean tube and spin for a further 15- 20 min. The pellet consists of proteins precipitated by the salt and the DNA should be present in the supernatant.
12. **Transfer the supernatant** to a clean 50ml tube and add **20ml absolute ethanol** at room temperature.
13. **DNA** should be visible as a **white aggregate.**  
If none is visible, place samples at **\*– 20°C** overnight, or at **-70°C** for 3 hours.  
(See 13b below).
14. Otherwise, fish or spool the DNA out using a P200 gilson, wash gently in **ice cold 70% ethanol (approx. 200ul).**
15. Place in a clean eppendorf tube and allow ethanol to evaporate off at room temperature. Do not allow to over dry.
16. Dissolve DNA in an appropriate volume (**~100ul**) of **TE buffer** (or water) to a maximum of 1 ml. Put at **37°C** to dissolve completely (can leave at **37°C/room** temperature overnight to ensure DNA has dissolved).
17. Check DNA concentration either by spectrophotometry or on a gel.

\*Day 3 (if DNA samples did not aggregate)

- 13b. Centrifuge samples in absolute ethanol at **2000 rpm, 4°C, 15 mins**
- 13c. Pour off absolute alcohol. Add 1 ml of ice-cold **70% ethanol.**
- 13d. Centrifuge at **2000 rpm, 4°C, 15 mins.**
- 13e. Remove 70% Ethanol. (Blot excess ethanol on tissue paper). Allow pellet to dry (but not over dry).
- 13f. Add **100 µl TE buffer.** Close lid. Allow DNA pellet to dissolve in TE overnight, RT°, shaker.
- 13g. Transfer to a labelled 1.5ml Eppendorf.
- 13f. Measure DNA concentration.