FGFR2 genetic variants in women with breast cancer

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

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

Received December 9, 2022; Accepted May 11, 2023

DOI: 10.3892/mmr.2023.13113

Correspondence to: Ms. Thérèse Dix-Peek or Professor Raquel A.B. Duarte, Department of Internal Medicine, School of Clinical Medicine, Faculty of Health Sciences, University of The Witwatersrand, 7 York Road, Parktown, Johannesburg 2193, South Africa
E-mail: therese.dix-peek@wits.ac.za
E-mail: raquel.duarte@wits.ac.za

Abbreviations: ASIR, age standardized incidence rate; ATM, ataxia-telangiectasia mutated; BRCA1, breast cancer 1; CDH1, epithelial cadherin 1; CHBAH, Chris Hani Baragwanath Academic Hospital; CHEK2, checkpoint kinase 2; CMJAH, Charlotte Maxeke Johannesburg Academic Hospital; ER, estrogen receptor; FGFR2, fibroblast growth factor receptor 2; GWAS, genome-wide association studies; HDAC, histone deacetylase; HER2, human epidermal growth factor receptor 2; HWE, Hardy-Weinberg equilibrium; ILC, invasive lobular carcinoma; KASP, Kompetitive allele-specific PCR; LD, linkage disequilibrium; LR, lifetime risk; MAF, minor allele frequency; NHLS, National Health Laboratory Services; PALB2, partner and localizer of BRCA2; PR, progesterone receptor; PTEN, phosphatase and tensin homolog; RUNX2, runt-related transcription factor 2; SABCHO, South African Breast Cancer and HIV Outcome; SNP, single nucleotide polymorphism; TNBC, triple-negative breast cancer; TP53, tumor protein P53; YY1, Yin Yang 1

Key words: FGFR2, SNP, African, breast cancer, heterogenous, risk

Abstract. Black African populations are more genetically diverse than others, but genetic variants have been studied primarily in European populations. The present study examined the association of four single nucleotide polymorphisms (SNPs) of the fibroblast growth factor receptor 2, associated with breast cancer in non-African populations, with breast cancer in Black, southern African women. Genomic DNA was extracted from whole blood samples of 1,001 patients with breast cancer and 1,006 controls (without breast cancer), and the rs2981582, rs35054928, rs2981578, and rs11200014 polymorphisms were analyzed using allele-specific Kompetitive allele-specific PCR™, and the χ^2 or Fisher's exact tests were used to compare the genotype frequencies. There was no association between those SNPs and breast cancer in the studied cohort, although an association was identified between the C/C homozygote genotype for rs2981578 and invasive lobular carcinoma. These results show that genetic biomarkers of breast cancer risk in European populations are not necessarily associated with risk in sub-Saharan African populations. African populations are more heterogenous than other populations, and the information from this population can help focus genetic risks of cancer in this understudied population.

Introduction

African populations are more genetically diverse than Asian or European populations (1,2), but few genomic studies have

been conducted in African populations (3). Only 2.4% of genome-wide association studies (GWAS) have included Africans or African Americans (2), and very little information is available regarding cancer genomics in African populations (3,4). A meta-analysis by Rotimi *et al* (3) found that between January 1990 and December 2019, only 0.329% of cancer-related publications globally focused on African populations, and only 0.016% were related to cancer genetics or genomics in Africa. Breast cancer is the most common type of cancer in women worldwide, 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 with 113.2/100,000 in Belgium) and a lifetime risk (LR) of 1 in 25 for women under 75 years (5). 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 (6).

Since the 1990s, the breast cancer 1 (*BRCA1*) (7) and *BRCA2* (8) 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 (9,10). Moderate penetrance genes include checkpoint kinase 2 (*CHEK2*), BRCA1 interacting helicase 1, ataxia-telangiectasia mutated (*ATM*), or partner and localizer of BRCA2 (*PALB2*) (8,10). Studies have been performed 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* (11) found 1.2% *BRCA1*, 0.6% *BRCA2*; 0.6% *ATM*, 0.6% *CHEK2*, and 0.6% *PALB* deleterious variants associated with breast cancer. Similarly, in 78 black patients, Francies *et al* (12) found 3.8% *BRCA1* and 3.8% *BRCA2* pathogenic mutations, but no deleterious mutations in *PALB2* or *CHEK2* in this group. Van der Merwe *et al* (13,14) identified larger rearrangements of the *BRCA1* and *BRCA2* genes that are specific to the black South African population. 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, GWAS has led to the detection of over 200 loci associated with breast cancer (15-18). 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 (19), and epithelial-to-mesenchymal transition (20). Variants in intron 2 of *FGFR2* have been found to be highly associated with breast cancer (15,17,18,21-23). Although most single nucleotide polymorphisms (SNPs) have small effects individually, polygenic models indicate that an accumulation of small mutations may increase the risk of cancer (24). 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.

The *FGFR2* SNP that is most commonly associated with breast cancer in women of European ancestry, rs2981582 (17), was recently also associated with an increased risk of breast cancer in Saudi Arabian women (25), and with luminal A breast cancer in Han Chinese women (26) and Korean women (27). The expression of the minor allele was associated with early-onset breast cancer in Indonesian women (28). 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) (29). Likewise, among postmenopausal Turkish women, rs2981582 was not associated with breast cancer (30). Admixed populations, such as that of Turkey (31), may have different allele frequencies than European populations and, given similar sample sizes, less power to detect associations.

The risk allele for rs35054928 appears to bind the transcription factor, E2F1 (32). Both rs35054928 and rs2981578 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 Yin Yang 1 (YY1), SIN3A, and histone deacetylase (HDAC) (33). The DNA binding protein YY1, co-repressor, SIN3A, and histone-modifying HDAC form a complex that can inhibit promoter activity (34). 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) (20,35). Runx2 is primarily known for its role in osteoclast development, but it is also a regulator of mammary development and breast cancer (36). The rs2981578 variants have equal affinity for OCT1, but the high-risk allele has a much higher affinity for RUNX2 (33,35), possibly because the SNP sites differ in histone acetylation (37). OCT1 promotes cell proliferation in estrogen receptor (ER) positive breast cancer cells (38).

The SNP rs11200014 has been associated with breast cancer in African American women (18,39). African Americans generally have admixed African and European ancestry, and a small proportion of Native American ancestry (1,40), and their African ancestry is primarily from West or central West Africa (41). South African populations are genetically different from West and Central Africans, and differ even more from African Americans.

Thus, in the present study, the association between *FGFR2* and breast cancer was explored, and their association with hormone receptor subtypes of breast cancer in an urban South African Black female population was assessed.

Materials and methods

Study population. The median age [interquartile range (IQR)] of the cases was 53 years (44-64 years), and 51 years (40-62 years) in the control participants. Participants in the present 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 (42). For this study, 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) were selected. The CMJAH Surgical Breast Unit is located in central Johannesburg and identifies ~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 (42). 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 were used. Eligible cases were self-identified black, southern African women >18 years of age with histologically confirmed invasive breast cancer; exclusion criteria were patients with ductal carcinoma *in situ* or lobular carcinoma *in situ*. 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 (2-8 ml) and collected into EDTA vacutainer tubes (Becton, Dickson, and Company) between October 2014 to March 2020. Ethics clearance for this study was granted from the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (Ethics numbers M140980, M161116). Written permission was granted by the CEOs of both CHBAH and CMJAH for the study.

Genomic DNA extraction, SNP selection, and analysis. DNA was extracted from whole blood using a modified salting out method. Briefly, whole blood was lysed with 320 mM sucrose in ice cold buffer (10 mM Tris-Cl, 5 mM $MgCl_2$, 1% Triton X; 1 part blood: 4 parts lysis buffer), and was centrifuged at 900 x g for 10 min at 4°°C. The pellet was resuspended in fresh lysis buffer, recentrifuged and the resulting pellet was resuspended in 3 ml T20E5 (20 mM Tris-HCl, 5 mM EDTA), 200 μl 10% SDS and 495 μl proteinase K solution (2 mg/ml proteinase K; 1% SDS, 2mM EDTA) and incubated overnight at 42-50˚C. Subsequently, 1 ml saturated NaCl (40%) was added to the solution, incubated on ice for 5 min and centrifuged at 900 x g for 30 min at 4˚C. The supernatant was transferred to a clean tube where 20 ml absolute ethanol was added, which caused the DNA to precipitate out of solution. The DNA could then be spooled and transferred a clean 1.5-ml microcentrifuge tube. The DNA was airdried and dissolved in low TE buffer (10 mM Tris HCl; 0.1 mM EDTA, pH 8.3) and diluted to a final concentration of 25 ng/μl. This was based on the method of Miller *et al* (43). A Nanodrop 2000™ spectrophotometer (Thermo Fisher Scientific, Inc.) was used to determine DNA concentrations and the A260/280 ratios; a ratio between 1.7 and 2.0 indicated adequate DNA purity for this genotype analysis.

For this confirmatory candidate gene study, 4 SNPs were selected 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 Power Calculator (https://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html). A sample size of 1,000 cases and 1,000 controls was chosen along with a significance level of 0.0125 (α=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 (23,44,45). Table I shows the minor allele frequencies (MAFs) for each SNP and compares them with the African and global allele frequencies in other studies. Information regarding allele frequencies from the African and global populations was obtained from the 1000 genomes project (46). Investigations of ancestry informative markers on a similar cohort sourced from this region showed limited evidence of population substructure (47).

The *FGFR2* SNP polymorphisms were genotyped using Kompetitive allele-specific PCR (KASP™) technology at LGC Genomics Ltd. 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 forward primer. During PCR, the allele-specific forward primer binds the DNA template, and subsequent PCR rounds generate the complement, which unquenches the fluorescent tag. If the genotype at a given SNP is homozygous, only one of two possible fluorescent signals will be generated, while a heterozygous genotype will generate a mixed fluorescent signal (48). The primers for each SNP are listed in Table II. For each SNP, the deviation of genotype frequencies from the Hardy-Weinberg equilibrium (HWE) in controls was determined using a χ^2 test (Table I).

Classification of tumors. Histopathological characteristics, including histological diagnosis, tumor subtype and grade, and immunolocalization of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and Ki67, were determined by histopathologists at the National Health Laboratory Services (NHLS) at CHBAH and CMJAH as part of patient standard of care. Pathological reports were produced for the clinical care of the patients, and selected data from these reports were included in the database for the present study. Data on these reports included tumor type, IHC, stage, in situ component. Immunostaining was performed on the benchmark XT automatic platform (Roche Diagnostics). Immunohistochemistry was performed according to the College of American Pathologists guidelines (49).

The ER/PR status was determined using Allred scoring as described previously (50) and scored as follows: 0-2, negative; and 3-8, positive. HER2 was regarded as positive if the score was 3+; negative when it was 0 or 1+ and equivocal when it was 2+; equivocal HER2 results were confirmed as positive by *in situ* hybridization as described previously (50). Specimens in which ≤14% of cells expressed Ki67 were categorized as having low expression, as per the St Gallen 2011 guidelines which form the basis of the current South African Guidelines (51,52). Immunostaining was performed according to the protocol in Bancroft's theory and practice of histological techniques (53) on an accredited diagnostic instrument as part of routine standard of care and according to the College of American Pathologists guidelines (49), as aforementioned. Based on IHC subtyping, the breast tumors were further categorized as: A-like or luminal A (ER and/or PR positive, HER2 negative, Ki67 ≤14%); B-like or luminal B (ER and/or PR positive and HER2 negative with Ki67% >14%); B/HER2-like or luminal B/HER2+ (ER and/or PR positive, HER2 positive with any

Table I. Allele frequency distribution in South African, African, and global populations.

Single nucleotide polymorphism	Position	Allele	MAF Present study	MAF [a]Soweto	MAF [b]African	MAF [b]Global	Hardy-Weinberg Equilibrium P-value
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

MAF, minor allele frequencies. [a]Allele frequencies looking at genetic diversity in the black population from Soweto, South Africa (69); [b]Frequencies available from the 1000 genomes project (46). The African group consisted of 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.

Table II. Sequences of the FGFR2 primers for detection of SNPs.

SNP	Primer ID	Allele	Primer sequence, 5'-3'
rs2981582	2981582_A	A	GGCACCAGGTGGACTCTCCA-FAM
	2981582_G	G	GCACCAGGTGGACTCTCCG-HEX
	2981582_Common		TAAAACGGCAGATCCCAGCACTCAT
rs35054928	35054928_CC	CC	CTCAGAAGGGCTGTGCGCC-FAM
	35054928_C	C-	TCTCAGAAGGGCTGTGCGCG-HEX
	35054928_Common		GCCCTGTCCCAGAAAGCCTACAT
rs2981578	2981578_T	T	TAACCTTTCTTCCCTGCTCCAAACT-FAM
	2981578_C	C	CCTTTCTTCCCTGCTCCAAACC-HEX
	2981578_Common		GTTTTCTTGAAGCTTTTACCTCTATGCAAA
rs11200014	11200014_A	A	CTCCAAAAAAAGATGCACAGAGGGAAGA-FAM
	11200014_G	G	CAAAAAAAGATGCACAGAGGGAAGG-HEX
	11200014_Common		ACACGTGTTGGGACCAGAGAGAAAA

SNP, single nucleotide polymorphism.

Ki67); HER2 positive subtype (ER and PR negative, HER2 positive); and triple-negative breast cancer (TNBC; ER, PR, and HER2 negative) (51,54).

Statistical analysis. Continuous variables were assessed for normality using the Shapiro-Wilkes test. Normally distributed variables are presented as the mean ± SD, and non-normally distributed data as the median and IQR. Normally distributed continuous variables were compared using a Student's t-test, while non-normally distributed variables were compared using a Mann-Whitney U test. Categorical variables are presented as frequencies and percentages and were compared using a Pearson's χ^2 test or Fisher's exact test if frequencies of >20% of cells were <5. All statistical analyses were performed using STATA version 14.2 (StataCorp LP). P<0.05 was considered to indicate a statistically significant difference.

Using the genotyping data provided by LGC genomics, both genotype and allele frequencies for each SNP of interest were calculated. 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 χ^2 test or Fisher's exact test as appropriate. Genotype frequencies were also examined under four different models of genetic disease risk (55). 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 to indicate a statistically significant difference.

Results

Overall, the 1,001 women with breast cancer were significantly older than the 1,006 women without breast cancer; the medians [IQRs] were 53 [44-64] and 51 [40-62], respectively (Table III). Most patients with cancer (57.4%) had been diagnosed with Stage III or IV cancer at presentation. The majority of tumors (82.5%) were invasive ductal carcinomas; 75.2% were ER-positive (76.7%), 63.8% were PR-positive, 26.9% were HER2-positive, irrespective of hormone receptor status, and 14.4% were TNBC (Tables III and IV).

The present study showed all 4 SNPs were in HWE (Table I). The MAFs of the SNPs in this study were more

Table III. Characteristics of the study participants.

Characteristic	Case	Control	P-value
Age, n (%)	n=1,001	n=1,006	
Overall age in years, median (IQR)	53 (44-64)	51 (40-62)	<0.001[a]
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%)	2 (0.2%)	
Cancer stage, n (%)			
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 (%)			
Invasive ductal carcinoma	826 (82.5%)		
Invasive lobular carcinoma	14 (1.4%)		
Other[b], n (%)	71 (7.1%)		
Tumor type unavailable	90 (9.0%)		
IHC phenotype, n (%)			
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[c]	23 (2.3%)		

[a]P<0.001. [b]Other tumor types included apocrine, cribriform, medullary, mesenchymal, metaplastic, mucinous, neuroendocrine, papillary, pleomorphic, squamous, tubular, anaplastic, micropapillary carcinomas. [c]HER2 had a value of 2+, but no fluorescence *in* situ hybridization was performed to confirm, or the Ki67 values were unavailable or ambiguous. ER, estrogen receptor; PR, progesterone receptor, HER2, human epidermal growth factor receptor 2; IQR, interquartile range; Ki67, proliferation marker; A-like, luminal-A; B-like, luminal B/HER2-; B/HER2-like, luminal B/HER2+; TNBC, triple-negative breast cancer.

Table IV. ER, PR and HER2 expression in the breast tumors.

Receptor status	Positive (%)	Negative (%)	Equivocal[a] (%)	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)

[a]HER2 had a value 2+, but no fluorescence *in situ* hybridization was performed to confirm, or the Ki67 values were unavailable or ambiguous.

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 the present study cohort (Table I).

Cases and controls did not differ in allele or genotype frequencies of the four *FGFR2* SNPs. The odds ratios (ORs) (95% confidence intervals) were: rs2981582, OR=1.10 (0.98-1.25) and P=0.156; rs35054928, OR=0.95 (0.84-1.40) and P=0.604; rs2981578, OR=1.07 (0.82-1.40) and P=0.637; and rs11200014, OR=1.04 (0.87-1.25) and P=0.666 (Table V).

Although in rs11200014, the recessive genotype (AA) was associated with an increased risk of HER2-positive breast cancer (0.038) compared with HER2-negative tumors (0.014)

Table V. Genotypic frequencies of *FGFR2* variants in the case and control populations.

Allele	Case, n (freq.)	Control, n (freq.)	P-value	Odds ratio[a] (95% confidence interval)
rs2981582	984	996	0.347[b]	
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[d]			0.156	1.10 (0.98-1.25)
GA+AA vs. GG	698 (0.709)	685 (0.688)	0.295[b]	1.11 (0.92-1.35)
AA vs. GG+GA	240 (0.244)	218 (0.219)	0.187[b]	1.18 (0.95-1.45)
AA vs. GG	240 (0.456)	218 (0.412)	0.148[b]	1.23 (0.96-1.57)
G	1030 (0.523)	1089 (0.547)		1 (Ref)
A	938 (0.477)	903 (0.453)	0.141[b]	1.11 (0.98-1.26)
rs35054928	989	996	0.848[b]	
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[d]			0.604	0.95 (0.84-1.08)
CC/C-+C-/C- vs. CC/CC)	649 (0.656)	661 (0.664)	0.727[b]	0.94 (0.78-1.13)
C-/C- vs. CC/CC+CC/C-)	186 (0.188)	197 (0.198)	0.583[b]	0.93 (0.74-1.17)
C-/C- vs. CC/CC	186 (0.354)	197 (0.370)	0.572[b]	0.90 (0.69-1.16)
CC	1143 (0.578)	1134 (0.569)		
C-	835 (0.422)	858 (0.431)	0.585[b]	0.95 (0.84-1.08)
rs2981578	989	995	0.813[c]	
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[d]			0.637	1.07 (0.82-1.40)
CT+TT vs. CC	122 (0.123)	116 (0.117)	0.642[b]	1.07 (0.81-1.40)
TT vs. CC+CT	3 (0.003)	2 (0.002)	0.686[c]	1.50 (0.25-9.08)
TT vs. CC	3 (0.0034)	2 (0.0023)	0.685[c]	1.52 (0.25-9.14)
C	1853 (0.937)	1872 (0.941)		1 (Ref)
T	125 (0.063)	118 (0.059)	0.609[b]	1.07 (0.82-1.39)
rs11200014	988	994	0.903[b]	
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[d]			0.666	1.04 (0.87-1.25)
GA+AA vs. GG	251 (0.254)	244 (0.246)	0.659[b]	1.05 (0.86-1.29)
AA vs. GG+GA	20 (0.020)	20 (0.020)	0.985[b]	1.00 (0.53-1.87)
AA vs. GG	20 (0.026)	20 (0.026)	0.956[b]	1.01 (0.54-1.90)
G	1705 (0.863)	1724 (0.867)		1 (Ref)
A	271 (0.137)	264 (0.133)	0.689[b]	1.04 (0.87-1.25)

[a]Adjusted for age. [b]Pearson χ^2, [c]Fisher's exact test. [d]Non-parametric test for trend.

(Table VI), the associations were not statistically significant after Bonferroni correction.

The different immunohistochemical subtypes were not associated with the *FGFR2* SNPs (Table VII). In addition, ductal carcinoma, lobular carcinoma, and other breast cancer types were not associated with the *FGFR2* SNPs, except for invasive lobular cancers exclusively, which had a protective allele in rs2981578 (P=0.016), whereas the more aggressive invasive ductal cancers had the risk allele (T) (Table VIII).

Immunohistochemical images of the various subtypes are shown in Figs. S1-S6. Each figure has a representative hematoxylin and eosin-stained image, and an IHC-stained image

Table VI. Association between genotypic frequencies of *FGFR2* variants and immunohistochemical subtyping of patients with breast cancer.

Genotype or allele	HR +ve (freq.)	HR-ve (freq.)	[c]P-value	[a]OR (95% CI)	HER2 +ve (freq.)	HER2-ve (freq.)	[c]P-value	[a]OR (95% CI)	HER2 +ve HR +ve (freq.)	HER2 +ve HR-ve (freq.)	[c]P-value	[a]OR (95% CI)
rs2981582	n=776	n=197			n=266	n=691			n=678	n=55		
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[b]			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)
rs35054928	n=781	n=197			n=267	n=695			n=681	n=54		
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[b]			0.98				0.877				0.847	
CC/C+C/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 (0.040)	45 (0.417)	0.947	1.02 (0.68-1.52)
rs2981578	n=780	n=198			n=267	n=695			n=680	n=55		
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)		0.64 (0.41-0.99)	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[b]			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	1,469 (0.942)	364 (0.919)		1 (Ref)	508 (0.951)	1,297 (0.933)		1 (Ref)	1,286 (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)

Table VI. Continued.

Genotype or allele	HR +ve (freq.)	HR-ve (freq.)	[c]P-value	[a]OR (95% CI)	HER2 +ve (freq.)	HER2-ve (freq.)	[c]P-value	[a]OR (95% CI)	HER2 +ve HR +ve (freq.)	HER2 +ve HR-ve (freq.)	[c]P-value	[a]OR (95% CI)
rs11200014	n=781	n=196			n=267	n=694			n=681	n=54		
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)	0.067	2.70 (1.10-6.63)	15 (0.022)	2 (0.037)	0.196	0.50 (0.11-2.29)
P-trend[b]			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)	0.025	2.79 (1.14-6.83)	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)	0.03	2.71 (1.10-6.66)	15 (0.028)	2 (0.054)	0.373	0.50 (0.11-2.28)
G	1350 (0.864)	338 (0.862)		1 (Ref)	457 (0.856)	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)

[a]Adjusted for age. [b]Non-parametric test for trend. [c]Pearson χ^2 test. HR, hormone receptor, HER2, human epidermal growth factor receptor 2, freq., frequency, OR, odds ratio, CI, confidence interval.

for ER, PR, HER2 and Ki67 (51). Fig. S1 shows an example of an ILC positive for hormone receptors (HRs) but negative for HER2 and with a Ki67 of 10%. Figs. S2-S6 show IDCs. Fig. S2 is indicative of an A-like IHC, with positive HR status, no HER2 localization and low Ki67 expression. B-like IHC (Fig. S3) shows positive HR, negative HER2 and high Ki67 expression. B/HER2 (Fig. S4) is a sample that is HR-positive and HER2-positive, and HER2-like (Fig. S5) is HR-negative and HER2-positive. Fig. S6 shows a triple negative cancer, where there is no expression of ER, PR or HER2. Images were selected by a senior pathologist at the NHLS.

Discussion

There is very little data on the genetics of cancer in sub-Saharan Africa in general, and here no association between *FGFR2* variants with breast cancer within the Saharan or sub-Saharan populations was shown. However, *FGFR2* variants have been studied in African American populations (23). In the present study, the potential association of *FGFR2* intronic SNPs with breast cancer in black southern African women was assessed. FGFR2 belongs to a tyrosine kinase receptor family that catalyzes multiple processes, including pro-survival signals, anti-apoptotic signals, cell proliferation, and cell migration (19). In GWAS studies, the rs2981582 SNP was strongly associated with breast cancer risk (17,18), particularly with ER-positive cancers [$P=6x10^{-7}$]. The present candidate gene replication study however, found no association between breast cancer and rs2981582, similar to the results of Udler *et al* (23) who also found no significant association. In further analysis, ER, PR, HER2, and immunohistochemical tumor types luminal A, luminal B, HER2 enriched, or TNBC, were also not associated with rs2981582.

Intron 2 of *FGFR2* contains putative transcription factor binding sites. Meyer *et al* (32) showed that the risk-associated allele (C) of rs2981578 preferentially bound FOXA1, and was able to recruit ERα to this site, and that rs35054928 preferentially bound to E2F1. E2F1 is important in the regulation of the proliferative response of breast cancer cells to estrogen. The expression of E2F1 increases with more advanced stages of breast cancer (56). Breast cancer was not associated with rs35054928 in the present study. However, rs2981578 was associated with ILC. Classic ILCs are typically of low histological grade, express ER and PR, and rarely show HER2 protein overexpression or amplification. In the present study, all ILCs were homozygous for the major allele, C/C. A defining feature of ILC is a lack of CDH1 protein expression. Ciriello *et al* (57) 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 cohort of patients with ILC cases.

Udler *et al* (23) showed a significant association between rs2981578 and ER-positive cases compared with the controls in African American women, whereas no significant associations were found in the black southern African women in the present study. Similarly, Barnholtz-Sloan *et al* (39) found rs11200014 to be associated with breast cancer in African American

Table VII. Association between genotypic frequencies of *FGFR2* variants and immunohistochemical subtyping of patients with breast cancer

Allele	A-like, n (freq.)	B-like, n (freq.)	B/HER2-like, n (freq.)	HER2-like, n (freq.)	TNBC, n (freq.)	P-value
rs2981582	108	439	208	58	139	0.525[a]
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)	
rs35054928	108	442	210	57	140	0.833[a]
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)	
rs2981578	108	442	209	58	140	0.177[b]
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	
rs11200014	108	442	210	57	139	0.139[a]
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)	

[a]Pearson χ^2, [b]Fisher's exact test. A-like, Luminal-A; B-like, Luminal B/HER2-; B/HER2-like, Luminal B/HER2+; TNBC, triple-negative breast cancer.

Table VIII. Genotypic frequencies of *FGFR2* variants with histological diagnosis.

Allele	Invasive ductal, n (freq.)	Invasive lobular, n (freq.)	Other, n (freq.)	P-value[b]
rs2981582	814	14	68	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)	
rs35054928	819	14	68	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)	
rs2981578	819	14	68	0.016[a]
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)	
rs11200014	819	14	68	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)	

[a]$P<0.05$. [b]Fisher's exact test.

women, but no association was found between rs11200014 and breast cancer in the present study. This difference may be caused by the genetic heterogeneity of the two populations; several African American women have both European and African ancestry, and their African ancestry is predominantly West or West-Central African (1). Indeed, multiple studies have shown that African Americans are most closely related to the Yoruba or Esan groups of Nigeria (58-60), or to groups from Sierra Leone (61), which are West African states, and to a lesser extent to the people from the Gambia, also a West

African state, or people from Central Africa (58,59). African Americans are less related to people from East African states such as Kenya, and to groups from Southern Africa such as the Xhosa or the San, and people from Northern Africa (41,58,61). Most African Americans are in populations that form a continuum from Europeans to West Africans (40,41,60,62).

Sub-Saharan Africans, including western, eastern and southern Africans as well as African Americans, are genetically diverse groups. Indeed, the human hereditary and health in Africa (1,2,63) consortium and malaria genomic epidemiology network (64) have shown that while there is genetic transfer between different African groups, there are distinct geographic and genomic groups. In sub-Saharan Africa, the movement of people speaking Niger-Congo languages, seems to have been from Nigeria (West Africa), through central Africa to Zambia (East Africa). From Zambia, there was a movement of people north and east (to present-day Uganda, Kenya, and Ethiopia), which makes up the Eastern Africa group. From Zambia, there was movement south to Botswana, South Africa, Namibia, and Eswatini, which make up the southern African group. These groups are genetically different from each other. The southern African group has also interacted with the Khoe and San groups, which are as distinct from people who speak Niger-Congo languages as are people from Europe (63). 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 the cases of the present study, the homozygote recessive genotype AA for rs11200014 was more prevalent in HER2-positive than in HER2-negative tumors; however, given the Bonferroni correction, the association was not statistically significant. Fernández-Noguiera *et al* (65) 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* (66) suggested that resistance to HER2 therapy may be caused by a change from an ER/HER2 signaling pathway to an FGFR2 signaling pathway. In ~25% of breast cancer cases in the South African population, patients are positive for HER2 expression (67,68).

In conclusion, GWAS studies in other populations have highlighted intron 2 of *FGFR2* as a region of interest in breast cancer. In the present study, the rs2981582, rs35054928, rs2981578, and rs11200014 SNPs were investigated in samples from black South African women but found no significant association with breast cancer. Thus, it is surmised that the difference between black southern African women and African American women is caused by the genetic diversity between southern Africans and west Africans, as well as the historical influence of European ancestry in the African American population. A limitation of this study is that the scope did not allow for interrogation of environmental factors that could cause epigenetic or germline mutations, and hence affect breast cancer susceptibility in this population. This study was used to investigate the low penetrance gene, FGFR2, that was highlighted by GWAS, as has been done in other geographic regions. As such, high and medium penetrance genes such as BRCA1/2, CHEK2, or PALB2, were not interrogated in the black South African population. These studies will be performed in the future. 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.

Acknowledgements

The authors would like to thank Mr. Eric Liebenberg (National Health Laboratory Services, Johannesburg, South Africa) for his help with the presentation of the immunohistochemistry micrographs; and Ms. Confidence Makgoro (Internal Medicine, Johannesburg, South Africa); as well as Mr. Victor Shandukani, , Ms. Nontlanta Mkwanazi, Ms. Mokgadi Mawela, Ms. Thandi Mtyapi, Ms. Sihle Sibiya, Ms. Yvonne Chaka and Ms. Olebogeng Mokgadi (Strengthening Oncology Research Unit, University of the Witwatersrand, Johannesburg, South Africa), for help recruiting patients and control participants.

Funding

This study was funded by the National Institutes of Health of the National Cancer Institute (grant nos. 01-CA192627 and P30-CA136696); the University of the Witwatersrand/South African Medical Research Council Common Epithelial Cancer Research Center Grant; a South African National Research Foundation grant award (grant no. 105646); the Cancer Association of South Africa; the National Research Foundation (grant no. NRF 87935); the minor Capex from the University of the Witwatersrand; and an AORTIC/NCI BIG CAT 2 Research grant.

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author on reasonable request.

Authors' contributions

TDP, CD and RABD designed the study with input from CGM and MH on GWAS, SNP selection and power calculations. TDP, CD, BPP, EJvdB, MJ, OAA, HC, SN, AIN, JSJ, PR and RABD collected and checked patient data. TDP, CD, MJ, OA and RABD collected and checked control participant data. TDP, CD, TNA and RABD performed the experiments. TDP, CD and RABD collected and cross referenced the SNP data to patients and controls. TDP, TNA, CD and RABD analyzed the data. TDP, CD and RABD confirm the authenticity of all the raw data. TDP, CD, TNA, CGM, MH, and RABD wrote the manuscript. All authors edited the manuscript. All authors read and approved the final manuscript.

Ethics approval and participant consent

The protocol used in the present study regarding human participants complied with the guidelines described in the Declaration of Helsinki. The present study was approved by the Human Research Ethics Committee (Medical) of the

University of the Witwatersrand (approval no. M140980, M161116). All participants provided signed informed consent prior to being enrolled in the study.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, Karthikeyan S, Iles L, Pollard MO, Choudhury A, *et al*: The African Genome Variation Project shapes medical genetics in Africa. Nature 517: 327-332, 2015.
2. Gurdasani D, Barroso I, Zeggini E and Sandhu MS: Genomics of disease risk in globally diverse populations. Nat Rev Genet 20: 520-535, 2019.
3. Rotimi SO, Rotimi OA and Salhia B: A Review of cancer genetics and genomics studies in Africa. Front Oncol 10: 606400, 2021.
4. Hayat M, Chen WC, Brandenburg JT, Babb de Villiers C, Ramsay M and Mathew CG: Genetic susceptibility to breast cancer in Sub-Saharan African populations. JCO Glob Oncol 7: 1462-1471, 2021.
5. Sharma R: Breast cancer burden in Africa: Evidence from GLOBOCAN 2018. J Public Health (Oxf) 43: 763-771, 2021.
6. South African National Cancer Registry. National Cancer Registry: Cancer in South Africa 2019. https://www.nicd.ac.za/wp-content/uploads/2021/12/NCR_Path_2019_Full_Report_8dec2021.pdf. Accessed May 13, 2022.
7. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B and King MC: Linkage of Early-onset familial breast cancer to chromosome 17q21. Science 250: 1684-1689, 1990.
8. Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T, Averill D, *et al*: Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science 265: 2088-2090, 1994.
9. Shiovitz S and Korde LA: Genetics of breast cancer: A topic in evolution. Ann Oncology 26: 1291-1299, 2015.
10. Mavaddat N, Antoniou AC, Easton DF and Garcia-Closas M: Genetic susceptibility to breast cancer. Mol Oncol 4: 174-191, 2010.
11. Eygelaar D, van Rensburg EJ and Joubert F: Germline sequence variants contributing to cancer susceptibility in South African breast cancer patients of African ancestry. Sci Rep 12: 802, 2022.
12. Francies FZ, Wainstein T, De Leeneer K, Cairns A, Murdoch M, Nietz S, Cubasch H, Poppe B, Van Maerken T, Crombez B, *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 15: 912, 2015.
13. Van der Merwe NC, Combrink HME, Ntaita KS and 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 13: 834265, 2022.
14. Van Der Merwe NC, Oosthuizen J, Theron M, Chong G and Foulkes WD: The contribution of large genomic rearrangements in BRCA1 and BRCA2 to South African familial breast cancer. BMC Cancer 20: 391, 2020.
15. Fachal L, Aschard H, Beesley J, Barnes DR, Allen J, Kar S, Pooley KA, Dennis J, Michailidou K, Turman C, *et al*: Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. Nat Genet 52: 56-73, 2020.
16. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, Jiang X, O'Mara TA, Zhao N, Bolla MK, *et al*: Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nat Genet 52: 572-581, 2020.
17. Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, *et al*: Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447: 1087-1093, 2007.
18. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wang Z, Welch R, Hutchinson A, Wang J, *et al*: A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 39: 870-874, 2007.
19. Zhou WY, Zheng H, Du XL and Yang JL: Characterization of FGFR signaling pathway as therapeutic targets for sarcoma patients. Cancer Biol Med 13: 260-268, 2016.
20. Katoh Y and Katoh M: FGFR2-related pathogenesis and FGFR2-targeted therapeutics (Review). Int J Mol Med 23: 307-311, 2009.
21. Agarwal D, Pineda S, Michailidou K, Herranz J, Pita G, Moreno LT, Alonso MR, Dennis J, Wang Q, Bolla MK, *et al*: FGF receptor genes and breast cancer susceptibility: Results from the Breast Cancer Association Consortium. Br J Cancer 110: 1088-1100, 2014.
22. Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A, Fumis L, Hayhurst J, Buniello A, Karim MA, *et al*: An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. Nat Genet 53: 1527-1533, 2021.
23. Udler MS, Meyer KB, Pooley KA, Karlins E, Struewing JP, Zhang J, Doody DR, MacArthur S, Tyrer J, Pharoah PD, *et al*: FGFR2 variants and breast cancer risk: Fine-scale mapping using African American studies and analysis of chromatin conformation. Hum Mol Genet 18: 1692-1703, 2009.
24. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, Tyrer JP, Chen TH, Wang Q, Bolla MK, *et al*: Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet 104: 21-34, 2019.
25. AlRaddadi RIR, Alamri RJN, Shebli WTY, Fallatah EIY, Alhujaily AS, Mohamed HS and Alotibi MKH: Fibroblast growth factor receptor 2 gene (FGFR2) rs2981582T/C polymorphism and susceptibility to breast cancer in Saudi women. Saudi J Biol Sci 28: 6112-6115, 2021.
26. Liang H, Yang X, Chen L, Li H, Zhu A, Sun M, Wang H and Li M: 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: 626948, 2015.
27. Han W, Woo JH, Yu JH, Lee MJ, Moon HG, Kang D and Noh DY: Common genetic variants associated with breast cancer in korean women and differential susceptibility according to intrinsic subtype. Cancer Epidemiol Biomarkers Prev 20: 793-798, 2011.
28. Purnomosari D, Raharjo C, Kalim AS, Herviastuti R, Yushan M, Fajar RA, Harris KK and Wahyono A: P21 Ser31Arg and FGFR2 rs2981582 polymorphisms as risk factors for early onset of breast cancer in Yogyakarta, Indonesia. Asian Pac J Cancer Prev 20: 3305-3309, 2019.
29. Brignoni L, Cappetta M, Colistro V, Sans M, Artagaveytia N, Bonilla C and Bertoni B: Genomic diversity in sporadic breast cancer in a Latin American population. Genes (Basel) 11: 1272, 2020.
30. Özgöz A, Mutlu İçduygu F, Yükseltürk A, ŞamlI H, HekİmIer Öztürk K and Başkan Z: Low-penetrance susceptibility variants and postmenopausal oestrogen receptor positive breast cancer. J Genet 99: 15, 2020.
31. Hodoğlugil U and Mahley RW: Turkish population structure and genetic ancestry reveal relatedness among eurasian populations. Ann Hum Genet 76: 128-141, 2012.
32. Meyer KB, O'Reilly M, Michailidou K, Carlebur S, Edwards SL, French JD, Prathalingham R, Dennis J, Bolla MK, Wang Q, *et al*: Fine-scale mapping of the FGFR2 breast cancer risk locus: Putative functional variants differentially bind FOXA1 and E2F1. Am J Hum Genet 93: 1046-1060, 2013.
33. Campbell TM, Castro MAA, De Santiago I, Fletcher MNC, Halim S, Prathalingam R, Ponder BAJ and Meyer KB: FGFR2 risk SNPs confer breast cancer risk by augmenting oestrogen responsiveness. Carcinogenesis 37: 741-750, 2016.
34. Lu P, Hankel IL, Hostager BS, Swartzendruber JA, Friedman AD, Brenton JL, Rothman PB and Colgan JD: The developmental regulator protein Gon4l associates with protein YY1, co-repressor Sin3a, and histone deacetylase 1 and mediates transcriptional repression. J Biol Chem 286: 18311-18319, 2011.
35. Meyer KB, Maia AT, O'Reilly M, Teschendorff AE, Chin SF, Caldas C and Ponder BA: Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. PLoS Biol 6: e108, 2008.
36. Owens TW, Rogers RL, Best SA, Ledger A, Mooney AM, Ferguson A, Shore P, Swarbrick A, Ormandy CJ, Simpson PT, *et al*: Runx2 is a novel regulator of mammary epithelial cell fate in development and breast cancer. Cancer Res 74: 5277-5286, 2014.

37. Zhu X, Asa SL and Ezzat S: Histone-acetylated control of fibroblast growth factor receptor 2 intron 2 polymorphisms and isoform splicing in breast cancer. Mol Endocrinol 23: 1397-1405, 2009.
38. Ogura T, Azuma K, Sato J, Kinowaki K, Takayama KI, Takeiwa T, Kawabata H and Inoue S: OCT1 is a poor prognostic factor for breast cancer patients and promotes cell proliferation via inducing NCAPH. Int J Mol Sci 22: 11505, 2021.
39. Barnholtz-Sloan JS, Shetty PB, Guan X, Nyante SJ, Luo J, Brennan DJ and Millikan RC: FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. Carcinogenesis 31: 1417-1423, 2010.
40. Bryc K, Durand EY, Macpherson JM, Reich D and Mountain JL: The genetic ancestry of African Americans, Latinos, and European Americans across the United States. Am J Hum Genet 96: 37-53, 2015.
41. Zakharia F, Basu A, Absher D, Assimes TL, Go AS, Hlatky MA, Iribarren C, Knowles JW, Li J, Narasimhan B, *et al*: Characterizing the admixed African ancestry of African Americans. Genome Biol 10: R141, 2009.
42. Cubasch H, Ruff P, Joffe M, Norris S, Chirwa T, Nietz S, Sharma V, Duarte R, Buccimazza I, Čačala S, *et al*: South African breast cancer and HIV outcomes study: Methods and baseline assessment. J Glob Oncol 3: 114-124, 2017.
43. Miller SA, Dykes DD and Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16: 1215, 1988.
44. Zhang Y, Zeng X, Liu P, Hong R, Lu H, Ji H, Lu L and Li Y: Association between FGFR2 (rs2981582, rs2420946 and rs2981578) polymorphism and breast cancer susceptibility: A meta-analysis. Oncotarget 8: 3454-3470, 2017.
45. Zhang Y, Lu H, Ji H, Lu L, Liu P, Hong R and Li Y: Association between rs11200014, rs2981579, and rs1219648 polymorphism and breast cancer susceptibility: A meta-analysis. Medicine (Baltimore) 96: 9-13, 2017.
46. 1000 Genomes Project Consortium; Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA and Abecasis GR: A global reference for human genetic variation. Nature 526: 68-74, 2015.
47. Kasembeli AN, Duarte R, Ramsay M, Mosiane P, Dickens C, Dix-Peek T, Limou S, Sezgin E, Nelson GW, Fogo AB, *et al*: APOL1 risk variants are strongly associated with HIV-associated nephropathy in black South Africans. J Am Soc Nephrol 26: 2882-2890, 2015.
48. He C, Holme J and Anthony J: SNP genotyping: The KASP assay. Methods Mol Biol 1145: 75-86, 2014.
49. Fitzgibbons PL, Connolly JL, Bose S, Chen YY, de Baca ME, Edgerton M, Hayes DF, Hil KA, Lester SC, Simpson JF, *et al*: Protocol for the examination of resection specimens from patients with invasive carcinoma of the breast. College of American Pathologists. Northfield (IL), 2020.
50. Fitzgibbons PL, Dillon DA, Alsabeh R, Berman MA, Hayes DF, Hicks DG, Hughes KS and Nofech-Mozes S: Template for reporting results of biomarker testing of specimens from patients with carcinoma of the breast. Arch Pathol Lab Med 138: 595-601, 2014.
51. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B and 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. Ann Oncol 22: 1736-1747, 2011.
52. Department of Health Republic of South Africa. Clinical Guidelines for Breast Cancer Control and Management, 1-123, 2018. Available from: https://cansa.org.za/files/2019/08/DOH-Breast-Cancer-Guidelines-Final.pdf. Accessed August 31, 2022.
53. Suvarna SK, Layton C and Bancroft JD: Bancroft's Theory and Practice of Histological Techniques. 8th edition, Elsevier, 2019.
54. Dix-Peek T, Phakathi BP, van den Berg EJ, Dickens C, Augustine TN, Cubasch H, Neugut AI, Jacobson JS, Joffe M, Ruff P and Duarte RAB. Discordance between PAM50 intrinsic subtyping and immunohistochemistry in South African women with breast cancer. Breast Cancer Res Treat 199: 1-12, 2023.
55. Lewis CM: Genetic association studies: Design, analysis and interpretation. Brief Bioinform 3: 146-153, 2002.
56. Stender JD, Frasor J, Komm B, Chang KCN, Kraus WL and Katzenellenbogen BS: Estrogen-regulated gene networks in human breast cancer cells: Involvement of E2F1 in the regulation of cell proliferation. Mol Endocrinol 21: 2112-2123, 2007.
57. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, Zhang H, McLellan M, Yau C, Kandoth C, *et al*: Comprehensive molecular portraits of invasive lobular breast cancer. Cell 163: 506-519, 2015.
58. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, *et al*: The genetic structure and history of Africans and African Americans. Science 324: 1035-1044, 2009.
59. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD and Kenny EE: Human demographic history impacts genetic risk prediction across diverse populations. Am J Hum Genet 100: 635-649, 2017.
60. Dai CL, Vazifeh MM, Yeang CH, Tachet R, Wells RS, Vilar MG, Daly MJ, Ratti C and Martin AR: Population histories of the United States revealed through fine-scale migration and haplotype analysis. Am J Hum Genet 106: 371-388, 2020.
61. Zimmerman KD, Schurr TG, Chen WM, Nayak U, Mychaleckyj JC, Quet Q, Moultrie LH, Divers J, Keene KL, Kamen DL, *et al*: Genetic landscape of Gullah African Americans. Am J Phys Anthropol 175: 905-919, 2021.
62. Sucheston LE, Bensen JT, Xu Z, Singh PK, Preus L, Mohler JL, Su LJ, Fontham ET, Ruiz B, Smith GJ and Taylor JA: Genetic ancestry, self-reported race and ethnicity in African Americans and European Americans in the PCaP cohort. PLoS One 7: e30950, 2012.
63. Choudhury A, Aron S, Botigué LR, Sengupta D, Botha G, Bensellak T, Wells G, Kumuthini J, Shriner D, Fakim YJ, *et al*: High-depth African genomes inform human migration and health. Nature 586: 741-748, 2020.
64. Busby GB, Band G, Si Le Q, Jallow M, Bougama E, Mangano VD, Amenga-Etego LN, Enimil A, Apinjoh T, Ndila CM, *et al*: Admixture into and within sub-Saharan Africa. Elife 5: e15266, 2016.
65. Fernández-Nogueira P, Mancino M, Fuster G, López-Plana A, Jauregui P, Almendro V, Enreig E, Menéndez S, Rojo F, Noguera-Castells A, *et al*: Tumor-associated fibroblasts promote HER2-targeted therapy resistance through FGFR2 activation. Clin Cancer Res 26: 1432-1448, 2020.
66. Hanker AB, Garrett JT, Estrada MV, Moore PD, Ericsson PG, Koch JP, Langley E, Singh S, Kim PS, Frampton GM, *et al*: HER2-overexpressing breast cancers amplify FGFR signaling upon acquisition of resistance to dual therapeutic blockade of HER2. Clin Cancer Res 23: 4323-4334, 2017.
67. Phakathi B, Cubasch H, Nietz S, Dickens C, Dix-Peek T, Joffe M, Neugut AI, Jacobson J, Duarte R and Ruff P: Clinico-pathological characteristics among South African women with breast cancer receiving anti-retroviral therapy for HIV. Breast 43: 123-129, 2019.
68. Achilonu OJ, Singh E, Nimako G, Eijkemans RMJC and 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: 6157861, 2022.
69. May A, Hazelhurst S, Li Y, Norris SA, Govind N, Tikly M, Hon C, Johnson KJ, Hartmann N, Staedtler F and Ramsay M: Genetic diversity in black South Africans from Soweto. BMC Genomics 14: 644, 2013.