Aus dem Deutschen Krebsforschungszentrum Heidelberg
(Vorstand: Prof. Dr. Michael Baumann)
in Zusammenarbeit mit
dem Institut für Medizinische Biometrie und Informatik der Universität Heidelberg
(Geschäftsführender Direktor: Prof. Dr. Meinhard Kieser)

INTERACTION BETWEEN GENETIC ANCESTRY AND COMMON SUSCEPTIBILITY VARIANTS IN COLOMBIAN BREAST CANCER PATIENTS

Inauguraldissertation
zur Erlangung des Doctor scientiarum humanarum (Dr.sc.hum.)
an der
Medizinischen Fakultät Heidelberg
der
Ruprecht-Karls-Universität

vorgelegt von
Karen Garcia Mesa

aus
La Habana, Cuba
2019

Dekan: Frau Prof. Dr. med. Andreas Draguhn

Doktormutter: Frau Prof. Dr. rer. nat. Ute Hamann

Table of Contents

List of Abbreviations

AIMs	Ancestry informative markers
AUC	Area under the Receiver Operating Characteristic curve
BC	Breast cancer
BMI	Body mass index
CI	Confidence interval
Col-BCCC	Colombian Breast Cancer Case-Control study
DNA	Deoxyribonucleic acid
ER	Estrogen receptor
ER+	Estrogen receptor positive
ER-	Estrogen receptor negative
GWAS	Genome-wide association studies
HDI	Human development index
HER-2	Human epidermal growth factor receptor 2
HER-2+	Human epidermal growth factor receptor 2 positive
HER-2-	Human epidermal growth factor receptor 2 negative
HWE	Hardy-Weinberg equilibrium
K	Cumulative BC risk by age 80 years
MAR	Missing at random
MCAR	Missing completely at random
MCMC	Markov chain Monte Carlo
MIR	Mortality to incidence ratio
MNAR	Missing not at random
MRS	Multifactorial risk score
OR	Odds ratio
P_A	Frequency of the high risk allele
PCA	Principal component analysis
PR	Progesterone receptor
PR+	Progesterone receptor positive
PR-	Progesterone receptor negative
PROC MI	MI procedure
P-value	Probability value
RR_1	Relative risks for homozygous genotypes
RR_2	Relative risks for heterozygous genotypes
SAP	Shrimp alkaline phosphatase
SNP	Single nucleotide polymorphism
TCGA	The Cancer Genome Atlas
TNBC	Triple negative breast cancer
WGA	Whole Genome Amplified

List of Figures

List of Tables

1. Introduction

1.1. Breast cancer incidence and mortality rates

1.1.1. Rates worldwide

Breast cancer (BC) is the most common cancer in women worldwide, accounting for 11.6% of all cancers (Bray *et al.*, 2018). According to GLOBOCAN (2018), it was estimated that 2.09 million new BC cases were diagnosed. Among the cases, 508,420 were attributed to countries with low and medium human development index (HDI), while 1,579,200 of them were attributed to countries with high and very high HDI. Age-standardized incidence rates in women vary nearly four-fold across the world regions, with rates ranging from 25.9 per 100,000 in South-Central Asia to 94.2 per 100,000 in Australia and New Zealand (Figure 1).

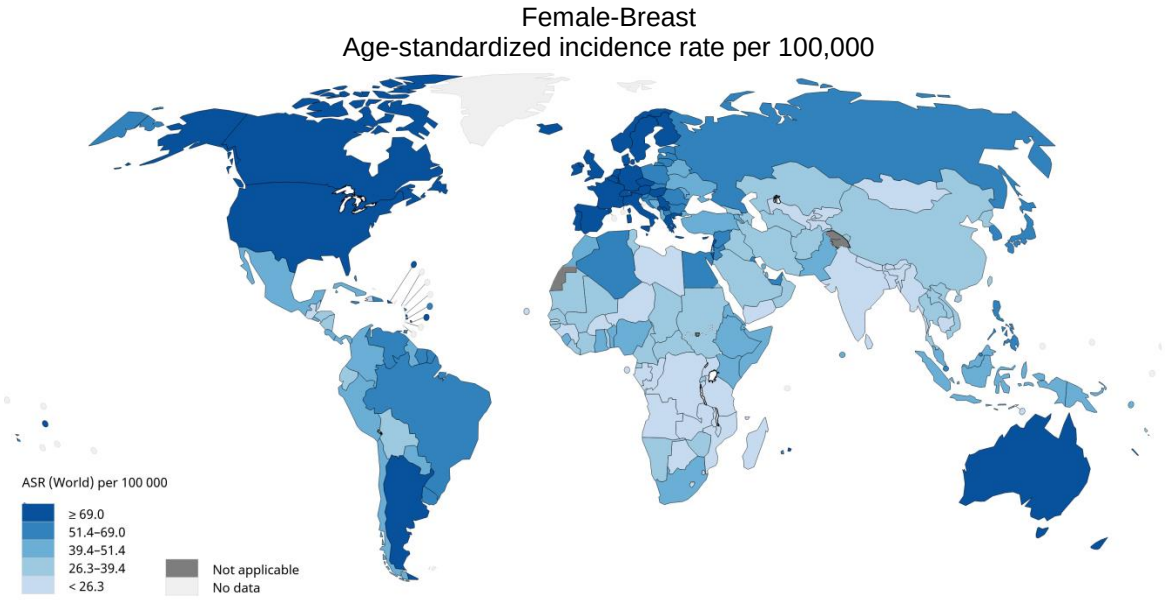

Figure 1: Estimated age-standardized breast cancer incidence rates worldwide in 2018 (Extracted from Globocan 2018).

GLOBOCAN (2018) reported that 626,679 deaths were due to BC in the world in 2018, ranking BC as the fifth cause of death from cancer overall. The variation in mortality rates between world regions is less than that for incidence because of the

more favorable survival of BC in (high-incidence) developed regions (Ghoncheh *et al.*, 2016), with rates ranging from 8.6 per 100,000 in Eastern Asia to 25.5 per 100,000 in Melanesia (Figure 2).

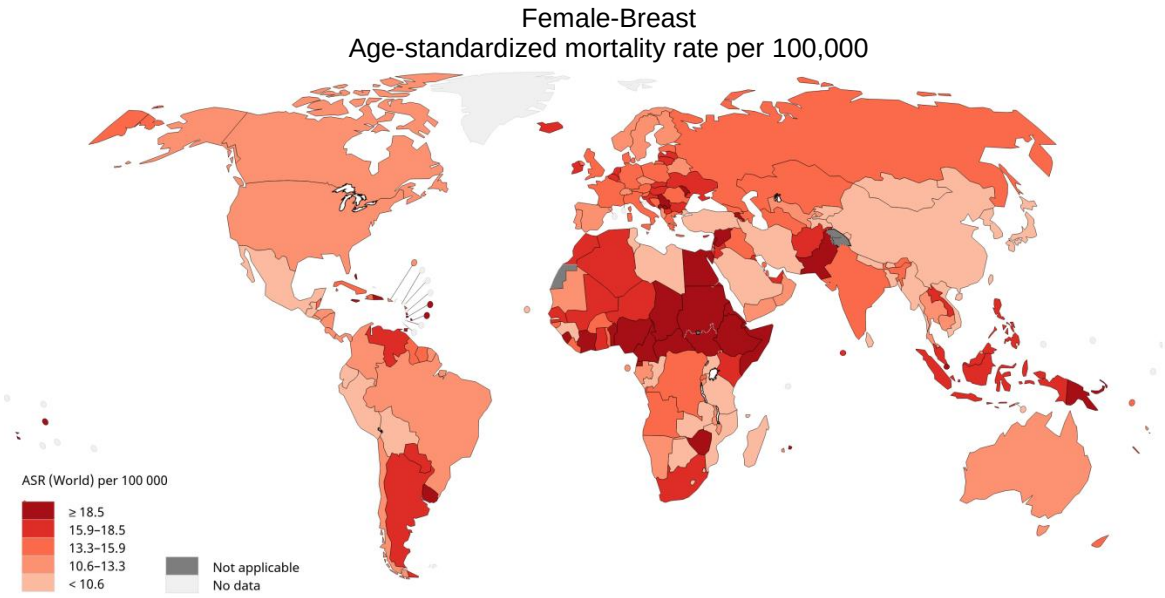

Figure 2: Estimated age-standardized breast cancer mortality rates worldwide in 2018 (Extracted from Globocan 2018).

The BC mortality to incidence ratio (MIR) computed by dividing the mortality rate by the incidence rate is another means to assess the burden of a diseases (Choi *et al.*, 2017). Less developed countries tend to have higher MIRs of BC compared to developed countries. For example, in Colombia the MIR is 0.27, while the MIR in United States is 0.15 and those in Western Europe countries are in the range of 0.14 and 0.19. High MIR indicates low survival. This could be due to late stage at diagnosis, lack of treatment modalities, and deficient strategies for BC prevention.

1.1.2. Rates in Latin America

BC is a major public health problem and the main cause of cancer related death among Latin Americans and Caribbean women, with an estimated 199,734 women diagnosed in 2018, representing 9.6% of all BC cases worldwide (Bray *et al.*, 2018).

Across Latin American and the Caribbean countries, a heterogeneous distribution of the estimated incidence rates, with values ranging from 23.6 to 78.3 per 100,000 women, is observed. The highest incidence rates in these countries were found in Argentina and Puerto Rico, with 78.3 and 73.0 per 100,000 women, respectively. Intermediate rates have been observed in Colombia with 44.1 per 100,000 women, whereas the lowest incidence rates were found in Haiti, Guatemala, and Bolivia (23.6, 26.2, and 26.5 per 100,000 women, respectively) (Bray et al., 2018).

Similar to the incidence rates, the estimated mortality rates also vary considerably according to country, with values from 7.3 to 33.1 per 100,000 women. The highest mortality rates are observed in Dominican Republic with 26.1 per 100,000 women. In Colombia, the mortality rate is 11.9 per 100,000 women (Bray *et al.*, 2018).

The BC MIR in the Latin American and Caribbean countries are also variable. For every 100 women diagnosed with BC, 25 will die of the disease in Mexico, whereas 31 and 61 will die in Guyana and Haiti, respectively.

Figure 3 shows the cumulative probability of BC incidence and mortality rate between 1980 and 2010 for individuals aged 15-79 years in some Latin American countries. A very high cumulative incidence of BC (more than 8% of cumulative probability) is observed in Uruguay and Argentina. By contrast, El Salvador shows a cumulative risk of less than 3%.

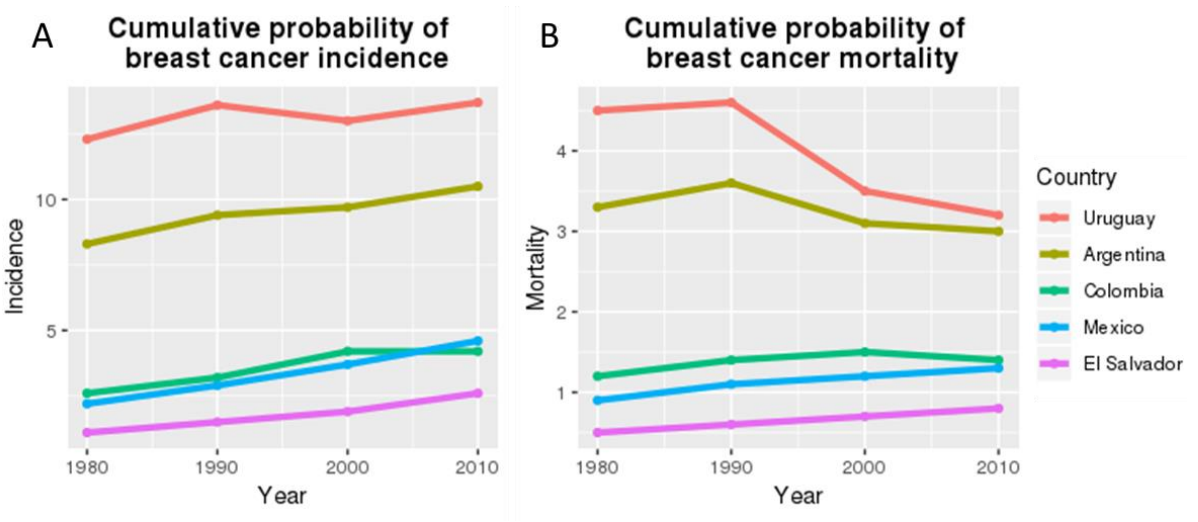

Figure 3: Cumulative probability of breast cancer incidence and mortality in Latin American countries. (A) Cumulative probability of breast cancer incidence increase over time. (B) Cumulative probability of breast cancer mortality varies across countries. Data extracted from Table 2 and Table 4 (Forouzanfar *et al.*, 2011).

An increase in the incidence and mortality rates was observed in some Latin American countries (Forouzanfar *et al.*, 2011). For example, in Colombia the cumulative probability of BC for women was 2.6 (2.3–3.8) in 1980 and 4.2 (3.6–5.0) in 2010 (Figure 3). A similar increase was observed in the mortality rate, which was 1.2 (1.1–1.7) in 1980 and 1.4 (1.2–1.6) in 2010.

The absolute number of cases and deaths has risen among Latin American and Hispanic women. This is probably due to the rising population numbers of women around the time of the menopause, together with the late stage at diagnosis (Pineros *et al.*, 2004), and limitations existing BC prevention programs for Latin American and Hispanic/Latino populations.

1.2. Classification of breast cancer

BC is a heterogeneous disease due to its diverse pathological features, variable clinical outcome, and response to treatments (Viale, 2012).

The histopathological classification of breast tumors is based on several morphological tumor features. It can be classified into several categories, ranging from benign, carcinoma in situ to invasive carcinoma, and sarcoma (Sinn and Kreipe, 2013). Breast carcinoma in situ, characterized by the proliferation of cancer cells within the epithelial tissue without invasion of the surrounding tissue can be further classified as either ductal or lobular (Malhotra *et al.*, 2010). Invasive breast tumors are characterized by invasion of tumor cells in adjacent tissues and can be further separated in several categories based on their histological morphology, location, and structure. Invasive ductal carcinoma is the most common subtype representing 75% of all BCs, followed by invasive lobular carcinoma, which represents about 10% of all BCs (Li *et al.*, 2005; Weigelt *et al.*, 2008). The remaining BCs distribute in various other rare subtypes, such as medullary, tubular, mucinous, apocrine, neuroendocrine, inflammatory, metaplastic, comedo, adenoid cystic, and macropapillary subtypes (Dieci *et al.*, 2014; Weigelt *et al.*, 2008).

In clinical practice, breast tumors are routinely classified in at least three main subtypes on the basis of immunohistochemistry: estrogen receptor (ER)/progesterone receptor (PR) positive (ER+/PR+), epidermal growth factor receptor 2 (HER-2) positive (HER-2+), and triple negative (ER-/PR-/HER-2-).

Breast tumors can also be classified by gene expression profiling analyzed by cDNA microarrays and hierarchical clustering. This led to the identification of five intrinsic molecular subtypes: luminal A (ER+, PR+, Ki-67 low), luminal B (ER+, PR+, Ki-67 high), HER-2+, normal-like, and basal-like (Prat and Perou, 2011). These BC

subtypes associate with different clinical prognosis. Luminal A and ER+/PR+ subtypes generally associate with a better prognosis, while basal, ER-/PR-, and *BRCA1*-associated tumors are linked with a worse prognosis (Dai *et al.*, 2015). The determination of the molecular subtype is of importance for treatment decisions.

Detailed genetic and clinical response of BC to hormonal therapy has not been systematically taken into account in BC studies in Latin America and among Hispanic populations. A study conducted among 301 patients from Colombia reported a prevalence of 37.2% for luminal B BCs, 26.3% for luminal A BCs, 11.6% for non-basal triple negative BCs (TNBCs), 9% for basal like tumors, 8.6% for HER-2+ tumors, and 69.1% for ER+ tumors (Serrano-Gomez *et al.*, 2016). In Europeans countries a prevalence of 6%-12.5% is observed for luminal B BCs, 68.7%-75% for luminal A BCs, 11.8%-12% for TNBCs, and 3%-8% for HER-2+ tumors (Awadelkarim *et al.*, 2008; Elidrissi Errahhali *et al.*, 2017; Puig-Vives *et al.*, 2013; Yang *et al.*, 2007).

1.3. Risk factors associated with breast cancer

BC is a complex disease with many non-genetic and genetic risk factors contributing to its development.

1.3.1. Non-genetic risk factors

Non-genetic risk factors of BC are conditioned by lifestyle, age or long-term medical intervention such as use of oral hormonal contraceptives. As many other cancers, age is one of the strongest risk factors for BC. It has been reported that BC incidence for women doubles about every 10 years, reaching an incidence of 500 cases per 100,000 woman-years at age 70 (Baselga and Norton, 2002). Age of diagnosis is variable worldwide and Latin American women have a high risk of developing BC at younger ages compared to their European counterparts. It is estimated that 20.2% of

Colombian BC cases occur in women younger than 45 years, in contrast to 12% in higher-income countries (Franco-Marina *et al.*, 2015).

Reproductive factors are associated with BC risk and are partly responsible for the diverse distribution of BC. First full-term pregnancy after 30 years of age, never having children, and lack of breastfeeding increase BC risk. On the other hand, multiparous women with more than five children had a reduced risk of BC (Collaborative Group on Hormonal Factors in Breast, 2002; Kelsey et al., 1993). In a hospital case-control study carried out in Bogota, parous women who had breastfed their first child had a reduced BC risk (P-value = 0.001). Compared with women with more than three children, nulliparous women had an OR of 3.35 (95% CI 1.40-8.0, P-value = 0.001) (Olaya-Contreras *et al.*, 1999).

Early menarche and late menopause are known to increase women´s risk of developing BC (Collaborative Group on Hormonal Factors in Breast, 2012). A study among 3,993 BC cases and 11,783 controls demonstrated that a delay of two years in the age at menarche corresponds to a 10% reduction in BC risk (Hsieh *et al.*, 1990).

Passive and active smoking habits have also been linked to an increased risk of BC in younger, primarily premenopausal women (Bottorff *et al.*, 2010; Johnson, 2005). There is some evidence for a link between heavy second-hand smoke exposure and BC risk in postmenopausal women (Johnson, 2005).

The use of oral contraceptives and menopausal hormonal therapy, especially combined estrogen and progestin therapy, may also increase the risk of BC (Collaborative Group on Hormonal Factors in Breast, 1996; Tavassoli FA, 2003). There is evidence for a small increase in BC risk associated with postmenopausal

estrogen therapy with longer duration of use in current and recent users (Tavassoli FA, 2003).

Higher body mass index (BMI) increases BC risk among postmenopausal women and the association appears to increase with advancing age after menopause (Tavassoli FA, 2003). Several studies have shown that obese pre- and postmenopausal women with BC have a poor survival and tend to have tumors of larger size (Cuello-Lopez *et al.*, 2017; Eichholzer *et al.*, 2012; Ioannides *et al.*, 2014; Kann *et al.*, 2014; Niraula *et al.*, 2012; Renehan *et al.*, 2008). In Colombian BC patients, 34.28% are overweight and 28.15% suffer from obesity. Higher BMI among premenopausal women was associated with hormone receptor negative tumors (i.e ER-/PR-) and greater lymphovascular invasion (Cuello-Lopez *et al.*, 2017).

1.3.2. Genetic risk factors

1.3.2.1. Genetic ancestry and family history of breast cancer

Differences in race and ethnicity may contribute to variation of BC incidence across populations. Caucasian women have higher age adjusted BC incidence rates than most African, Asian, and Latin American women.

Although incidence of BC is higher in Caucasian women, sub-Saharan African and African-American women are more likely to be diagnosed with biologically aggressive phenotypes like TNBC and have higher BC mortality rates (Jiagge *et al.*, 2018; Jiagge *et al.*, 2016; Newman and Kaljee, 2017).

In Asia, BC is the most common cancer. The incidence rate is low compared to other regions, but it is increasing rapidly. Asian women tend to have children at younger ages than in the west and the age-specific incidence decreases for women over 50 years old (Agarwal *et al.*, 2007; Yip, 2009).

The variation in incidence can be noticed not only worldwide, but also between ethnic groups within countries. In a previous study among Latina women in the San Francisco Bay Area it was shown that Latinas have incidence rates that are 35% lower than the rates of Caucasian women (Ziv *et al.*, 2006).

The term Hispanic and Latino describe a heterogeneous population. Compared to others populations they have a more complex genetic background representing a mix of Native American, European, and African ancestries (Fejerman *et al.*, 2010). Among US Latina women, those with a high proportion of Native American ancestry are at a lower risk of developing BC and those with higher European ancestry have an increased risk of BC (Fejerman *et al.*, 2012; Fejerman *et al.*, 2010; Ziv *et al.*, 2006).

Hereditary BC comprises about 5% to 10% of all BC diagnosis worldwide (Society, 2014). Along with genetic ancestry, the family history of BC contributes to the risk of developing the disease (Jiang *et al.*, 2012; Phipps *et al.*, 2011). In a meta-analysis published in 1997, Pharoah et al. reported that the magnitude of the risk of BC depends on the type of relative affected, the age at which the relative developed cancer, and the number of affected members (Pharoah *et al.*, 1997). Compared to individuals with no family history of BC, the risk of BC doubles for women with first degree relatives (mother, sister, or daughter) who have been diagnosed with BC at 50 years of age or older, and increased if the first-degree relative developed BC before 50 years of age. The relative risk associated with having a second-degree relative with BC was 1.5 (95% CI 1.4-1.6). Increasing numbers of affected first degree family members also increased the BC risk. Recently, first degree family history of BC was associated with an increased risk of ER-, PR-, and TNBC

(Collaborative Group on Hormonal Factors in Breast, 2001; Jiang *et al.*, 2012; Pharoah *et al.*, 1997; Phipps *et al.*, 2011; Singletary, 2003; Society, 2017).

1.3.2.2. Susceptibility genes and loci for breast cancer

Genome-wide association studies (GWASs) have led to the identification of several BC susceptibility genes and loci that contribute to the familial risk. Up to 20% of the familial relative risk is due to rare mutations in high-penetrance genes including *BRCA1*, *BRCA2*, *PTEN*, *TP53*, *CDH,* and *STK11.* Mutations in these genes have a less than 0.1% risk allele frequency in the general population and confer a relative risk of up to 8 per risk allele (Figure 4). The high-penetrance genes confer a lifetime risks of BC up to 50% (Ghoussaini *et al.*, 2013).

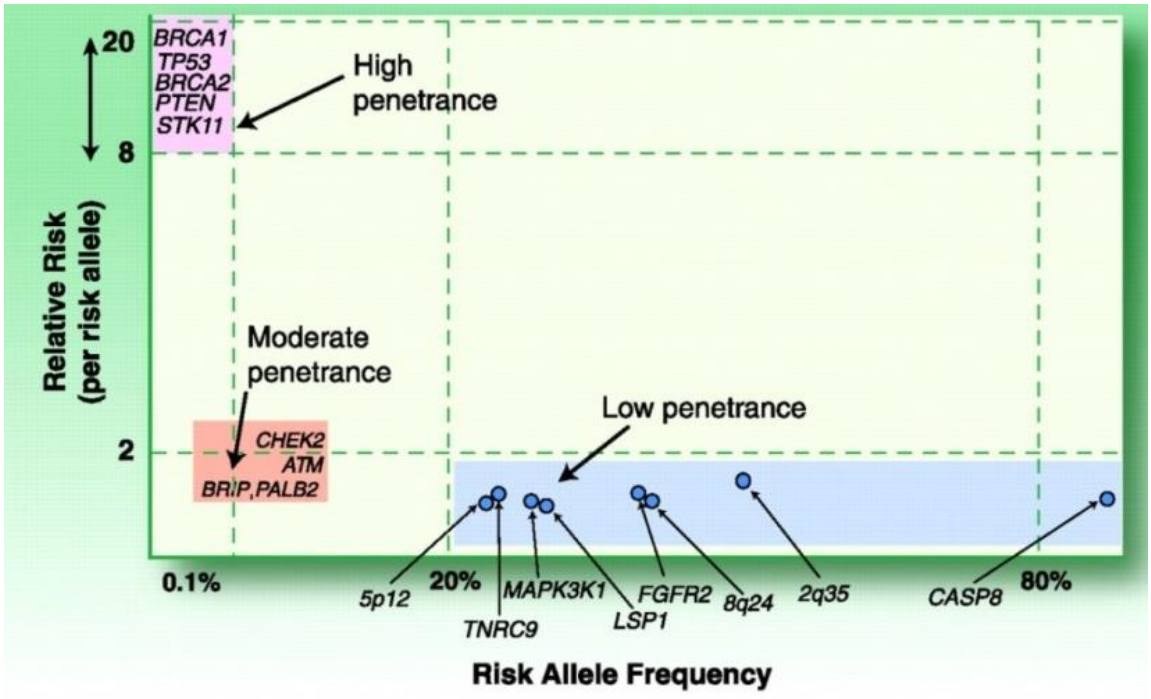

Figure 4: Breast cancer susceptibility loci and their relative risk per risk allele and frequency (Extracted from (Garcia-Closas and Chanock, 2008)).

Between 2% to 3% of cases are due to a mutation in moderate-penetrance genes, such as *CHEK2*, *BRIP1*, *ATM*, and *PALB2*, each associated with a two-fold increase in risk and lifetime risk up to 20%. Common low-penetrance susceptibility loci are

associated with small increases in risk with relative risk less than 2 and a frequency of the risk allele higher than 20%. Low-penetrance alleles confer a lifetime risk in the range of 10% to 20% (Beggs and Hodgson, 2009; Garcia-Closas and Chanock, 2008; Ghoussaini *et al.*, 2013; Shiovitz and Korde, 2015).

High-penetrance mutations

High-penetrance mutations in *TP53* are associated with the Li–Fraumeni syndrome, mutations in *PTEN* with the Cowden syndrome, and in *STK11* with the Peutz–Jeghers syndrome. *CDH1* predispose to familial diffuse gastric cancer and *BRCA1* and *BRCA2* mutations to breast and ovarian cancer (Beggs and Hodgson, 2009; Campeau *et al.*, 2008; Garcia-Closas and Chanock, 2008).

The most common cause of hereditary BC is an inherited mutation in the *BRCA1* or *BRCA2* gene. *BRCA1* gene is located on chromosome 17 (locus: 17q12-q21). Its tumor suppressor role makes it crucial for genomic stability, DNA repair, DNA damage response, cell cycle checkpoint control, chromatin remodeling, transcriptional regulation, and protein ubiquitination (Hall *et al.*, 1990; Narod and Foulkes, 2004). The *BRCA2* gene was cloned in 1995. It is located on chromosome 13q12-q13 (Wooster *et al.*, 1995). *BRCA2* plays an important role in homologous recombination, both in meiosis and repair of double-strand breaks (Thorslund and West, 2007). Mutations in *BRCA2* are associated with an increased risk of ductal and lobular cancers, while *BRCA1* mutations are frequently associated with invasive ductal carcinoma (Dossus and Benusiglio, 2015). Up to 15% of TNBCs have been associated with *BRCA1* and *BRCA2*. TNBC accounts for 70% of breast tumors arising in *BRCA1* mutation carriers and 16-23% of breast tumors in *BRCA2* carriers (Ling *et al.*, 2016; Stevens *et al.*, 2013). A recent comprehensive analysis of The

Cancer Genome Atlas (TCGA) program showed that around twenty percent of basal-like breast tumors have an inherited or somatic *BRCA1* or *BRCA2* mutation (Ling *et al.*, 2016).

Different studies on the *BRCA1* and *BRCA2* genes have been conducted in Colombia. These studies showed that *BRCA1* and *BRCA2* mutations affect between 5% and 8% of unselected Colombian BC patients (Torres *et al.*, 2017; Torres *et al.*, 2007). Among unselected BC cases, four founder mutations were identified: A1708E and 3450del4 in *BRCA1* and 3034del4 and 1991del4 in *BRCA2* (Torres *et al.*, 2017). The cumulative risk of BC by age 70 years for Colombian women carrying a mutation in the *BRCA1* gene was 14% (95% CI 5–38) and 3% for the general population (relative risk of BC 4.05) (Torres *et al.*, 2017). In another report, Hernández et al. (Hernandez *et al.*, 2014) assessed the prevalence of *BRCA1/2* mutations in BC cases unselected for family history from Medellin. Among 244 patients, three (1.2%) deleterious mutations were identified: two in *BRCA1,* the same *BRCA1* founder mutations as identified in the Colombian study (Torres *et al.*, 2017) and one in *BRCA2.*

Moderate-penetrance variants

The variants in *ATM*, *CHEK2*, *BRIP1*, *BARD1*, and *PALB2* have a contribution to familial risk less than 3%, due to the modest relative risk and relatively low frequency of the risk allele. These genes act in the same DNA repair pathways as *BRCA1* and *BRCA2*, however, they confer less risk of breast and ovarian cancer (Figure 4) (Beggs and Hodgson, 2009; Garcia-Closas and Chanock, 2008; Pharoah *et al.*, 2008).

Low-penetrance variants

Mutations in high-risk BC genes such as *BRCA1 and BRCA2* affect only small numbers of women, whereas variation in lower-impact, common susceptibility loci or single nucleotide polymorphisms (SNPs) can be responsible for a larger percentage of cancers in the population (Howell *et al.*, 2014). The relative risks of common variants are small, but in combination can be associated with substantial increases or decreases in risk. The combination of risks of common variants can be measured through a polygenic risk score. A polygenic risk score gives a risk percentile of individual genetic predisposition to specific diseases and can be used to stratify women in the general population at different levels of risk of developing BC (Torkamani *et al.*, 2018).

Recently, about 170 common low-penetrance variants that cumulatively explain 18% of the familial relative risk have been identified in collaborative GWASs for individuals of European descent (Amos *et al.*, 2017; Michailidou *et al.*, 2015; Michailidou *et al.*, 2013; Michailidou *et al.*, 2017).

Despite the enormous progress in the discovery of novel BC susceptibility genes in the European population, only a few of the associations have been subsequently replicated in Latin American and Hispanic women (Fejerman *et al.*, 2014; Fejerman *et al.*, 2013; Slattery *et al.*, 2011).

2. Objectives and Aims

BC is the main cause of cancer-related death among Colombian women. Its incidence is increasing and early diagnosis and strategies for cancer prevention remain unresolved problems. Only a few studies have been conducted in Latinas and Hispanic populations outside of the United States, and most strategies for BC prevention are based on women of European descent.

The objective of this dissertation was to develop effective BC risk prediction models considering the characteristics of the Colombian population, taking advantage of individual genetic ancestry, information on established susceptibility factors, and recently identified common variants. The relevance of 78 BC susceptibility variants previously shown to be associated with BC risk in large-scale genetic association in European women to Colombian BC was examined and possible interactions with genetic ancestry were assessed. Study participants were women participating in the Colombian BC case-control study (Col-BCCC).

The specific objectives of the research project were:

- To quantify ancestry proportions in each woman and assess the relationship between ancestry and BC risk
- To determine associations between genetic risk variants and BC risk, subtype specific risk and quantify the variance in BC liability due to susceptibility variants
- To assess interactions between common variants and genetic ancestry on BC susceptibility
- To build a multifactorial risk score (MRS) for BC risk prediction

3. Materials and Methods

3.1. Study populations

The study population included 1,022 BC patients unselected for family history and age at BC diagnosis, and 1,023 healthy controls who participated in the Col-BCCC study. The Ethics Committee of the Pontificia Universidad Javeriana in Bogota approved the research protocol. All study participants signed the informed consent.

The majority of cases were recruited in the period of 03/2007 to 02/2011 from hospitals located in the center of Colombia (Bogota, Neiva and Villavicencio). BC cases were diagnosed after January 1st, 2004. Controls were healthy and unrelated women recruited during the period 06/2007 to 06/2011. Controls reported no family history of any type of cancer in two generations and participated in the Colombian National Pap Smear Program (Pineros *et al.*, 2007). Cases and controls were of Hispanic origin and matched by 2-year age classes. They were eligible if they resided in the study region.

Table 1 provides information on established and potential BC risk factors that was collected from all study participants: age of diagnosis for cases and age at interview for controls, family history of BC in first-degree female relatives (yes / no), oral contraceptive use (yes / no), menopausal status and hormone therapy (premenopausal and postmenopausal never-hormone therapy, postmenopausal ever-hormone therapy), BMI (<25 / 25 to 29.9 / >=30 kg/m2), current smoking (never / former / current), parity (nulliparous,1 to 2, ≥3 children), age at first full-term pregnancy (<20, 20 to 29, ≥30 years), age at menarche (≥14 / 12 to 13 / <12 years), and breastfeeding (Never / ≤12 / >12 months).

Table1: Baseline characteristics of the study population

Variable	Level	Cases N (%)		Controls N (%)		P-value[1]
Age (years)		1,021	(99.90)	1,023	(100.00)	0.96
Family history of breast cancer in first-degree female relatives	No	770	(75.34)	921	(90.03)	**<0.0001**
	Yes	246	(24.07)	79	(7.72)	
Oral contraceptive use	No	687	(67.22)	742	(72.53)	**0.03**
	Yes	310	(30.33)	270	(26.39)	
Menopausal status and postmenopausal hormone therapy use	Premenopausal and postmenopausal who never used hormone therapy	894	(87.48)	972	(95.01)	**<0.0001**
	Postmenopausal women who ever used hormone therapy	104	(10.18)	43	(4.20)	
Body mass index (kg/m2)	<25	494	(48.34)	444	(43.40)	0.13
	25-29.9	366	(35.81)	388	(37.93)	
	≥30	113	(11.06)	127	(12.41)	
Smoking status	Never	684	(66.93)	775	(75.76)	**0.0004**
	Former	251	(24.56)	186	(18.18)	
	Current	65	(6.36)	57	(5.57)	
Parity	Nulliparous	158	(15.46)	118	(11.53)	**0.002**
	1-2	478	(46.77)	439	(42.91)	
	≥3	377	(36.89)	440	(43.01)	
Age at first full-term pregnancy (years)	<20	176	(17.22)	208	(20.33)	**0.0002**
	20-29	490	(47.95)	545	(53.27)	
	≥30	178	(17.42)	119	(11.63)	
Age at menarche (years)	≥14	379	(37.08)	406	(39.69)	0.45
	12-13	463	(45.30)	447	(43.70)	
	<12	155	(15.17)	144	(14.08)	
Breastfeeding (months)	Never	219	(21.43)	158	(15.44)	**0.0002**
	≤12	296	(28.96)	264	(25.81)	
	>12	468	(45.79)	546	(53.37)	

[1]P-values from two-sided chi-square test and smaller than 0.05 are marked in bold.

3.2. Selection of genetic variants and genotyping analysis

Thirty markers highly informative for continental ancestry information (Ruiz-Linares *et al.*, 2014) were selected for genotyping analysis. The ancestry informative markers (AIMs) were genotyped using the KASP allelic discrimination method (LGC Genomics). The sample set consisted of 1,757 native DNA samples and 288 Whole Genome Amplified (WGA) DNA samples. Cluster plots and genotypes were evaluated by the SNPviewer2 software version 4.0.0 (LGC Genomics). In addition, the clustering of intensity plots for AIMs with a P-value for Hardy-Weinberg equilibrium (HWE) departure among controls under 0.005 was visually inspected. The genotype concordance between WGA amplified DNA and native DNA was >99%. AIM call rates were ≥98.9% and the concordance rate of 122 DNA duplicates (6.0%) was 100%. Eight AIMs deviated from HWE in controls but corresponding intensity plots revealed clear genotype clusters.

Seventy eight common BC susceptibility variants were selected for genotyping analysis (Table 2). They included 60 single nucleotide polymorphisms (SNP, 38 novel and 22 previously known) associated with BC risk at genome-wide significance in a large European GWAS (Michailidou *et al.*, 2013), 10 SNPs identified in fine-mapping studies (Bojesen *et al.*, 2013a; French *et al.*, 2013a; Ghoussaini *et al.*, 2014; Meyer *et al.*, 2013; Orr *et al.*, 2015a; Udler *et al.*, 2010a), 4 SNPs from a GWAS on ER- BC (Garcia-Closas *et al.*, 2013a), 2 SNPs from a GWAS on postmenopausal BC (Hunter *et al.*, 2007b), 1 SNP associated with BC risk in a postmenopausal case-control study but not in a GWAS study (Easton *et al.*, 2007a; Prentice *et al.*, 2009b), and 1 SNP from a GWAS on *BRCA1*-associated BC (Antoniou *et al.*, 2010a). DNA samples from 1,022 BC patients and 1,023 healthy controls were genotyped using three different methods. Seventy-five SNPs were

analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) (Sequenom, San Diego, CA, USA). SNP rs4808801 was analyzed using KASP genotyping chemistry. Cluster plots and genotypes were evaluated by the SNPviewer2 software version 4.0.0 (LGC Genomics). Two SNPs - rs11552449_C>G and rs2736108_C>T were genotyped by TaqMan allelic discrimination using 20x TaqMan® SNP Genotyping Assays (assay ID C_25619066_10 and C_26414910_20, respectively) (ThermoFisher Scientific Inc., USA). SNP call rates were between 97.5% and 100%, the concordance rate of 119 duplicates (5.8% of total samples) was >99% and the distribution of all genotypes met HWE in the control group. Genotyped variants, their location, gene/nearest gene(s), and genotyping methods are listed in Table 2.

Table 2: Genotyped variants, their location, gene/nearest gene(s), and genotyping methods

#	SNP ID	Locus	Gene/nearest gene(s)*	Genotyping method	Reference (Initial study)
1	rs11249433	1p11.2	*EMBP1*	MALDI-TOF MS	(Thomas *et al.*, 2009)
2	rs11552449	1p13.2	*DCLRE1B*	TaqMan assay	(Michailidou *et al.*, 2013)
3	rs616488	1p36.22	*PEX14*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
4	rs6678914	1q32.1	*LGR6*	MALDI-TOF MS	(Garcia-Closas *et al.*, 2013b)
5	rs4245739	1q32.1	*MDM4*	MALDI-TOF MS	(Garcia-Closas *et al.*, 2013b)
6	rs12710696	2p24.1	*OSR1*	MALDI-TOF MS	(Garcia-Closas *et al.*, 2013b)
7	rs4849887	2q14.1	*INHBB*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
8	rs1550623	2q31.1	*CDCA7*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
9	rs2016394	2q31.1	*DLX2*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
10	rs16857609	2q35	*DIRC3*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
11	rs4442975	2q35	*IGFBP5*	MALDI-TOF MS	(Ghoussaini *et al.*, 2014)
12	rs4973768	3p24.1	*SLC4A7*	MALDI-TOF MS	(Ahmed *et al.*, 2009)
13	rs12493607	3p24.1	*TGFBR2*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
14	rs6762644	3p26.1	*EGOT/ITPR1*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
15	rs9790517	4q24	*TET2*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
16	rs6828523	4q34.1	*ADAM29*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
17	rs10941679	5p12	*MRPS30*	MALDI-TOF MS	(Ghoussaini *et al.*, 2016)
18	rs10069690	5p15.33	*TERT*	MALDI-TOF MS	(Haiman *et al.*, 2011)
19	rs2736108	5p15.33	*TERT*	TaqMan assay	(Bojesen *et al.*, 2013b)
20	rs3215401	5p15.33	*TERT*	MALDI-TOF MS	(Bojesen *et al.*, 2013b)
21	rs2242652	5p15.33	*TERT*	MALDI-TOF MS	(Bojesen *et al.*, 2013b)
22	rs889312	5q11.2	*MAP3K1*	MALDI-TOF MS	(Easton *et al.*, 2007b)
23	rs1353747	5q11.2	*PDE4D*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
24	rs10472076	5q11.2	*RAB3C*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
25	rs1432679	5q33.3	*EBF1*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
26	rs204247	6p23	*RANBP9*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
27	rs11242675	6p25.3	*FOXQ1*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
28	rs17530068	6q14.1	*FAM46A*	MALDI-TOF MS	(Siddiq *et al.*, 2012)
29	rs2046210	6q25.1	*ESR1/CCDC170*	MALDI-TOF MS	(Zheng *et al.*, 2009)
30	rs720475	7q35	*ARHGEF5/NOBOX*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
31	rs9693444	8p12	*DUSP4*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
32	rs6472903	8q21.11	*HNF4G*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
33	rs13281615	8q24.21	*POU5F1B*	MALDI-TOF MS	(Easton *et al.*, 2007b)
34	rs11780156	8q24.21	*PVT1/MYC*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
35	rs1011970	9p21.3	*CDKN2A/B*	MALDI-TOF MS	(Turnbull *et al.*, 2010)
36	rs10816625	9q31.2	*KLF4*	MALDI-TOF MS	(Orr *et al.*, 2015b)
37	rs676256	9q31.2	*KLF4*	MALDI-TOF MS	(Orr *et al.*, 2015b)
38	rs865686	9q31.2	*KLF4*	MALDI-TOF MS	(Fletcher *et al.*, 2011)
39	rs10759243	9q31.2	*KLF4*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
40	rs11814448	10p12.31	*DNAJC1*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
41	rs7072776	10p12.31	*MLLT10/DNAJC1*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
42	rs10995190	10q21	*ZNF365*	MALDI-TOF MS	(Turnbull *et al.*, 2010)
43	rs704010	10q22.3	*ZMZ1*	MALDI-TOF MS	(Turnbull *et al.*, 2010)
44	rs7904519	10q25.2	*TCF7L2*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
45	rs11199914	10q26.12	*FGFR2*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
46	rs35054928	10q26.13	*FGFR2*	MALDI-TOF MS	(Meyer *et al.*, 2013)
47	rs2981579	10q26.13	*FGFR2*	MALDI-TOF MS	(Hunter *et al.*, 2007a)
48	rs2981578	10q26.13	*FGFR2*	MALDI-TOF MS	(Meyer *et al.*, 2013)
49	rs3750817	10q26.13	*FGFR2*	MALDI-TOF MS	(Easton *et al.*, 2007b)
50	rs11200014	10q26.13	*FGFR2*	MALDI-TOF MS	(Hunter *et al.*, 2007a)
51	rs2420946	10q26.13	*FGFR2*	MALDI-TOF MS	(Hunter *et al.*, 2007a)
52	rs3817198	11p15.5	*LSP1*	MALDI-TOF MS	(Easton *et al.*, 2007b)
53	rs614367	11q13.3	*MYEOV/CCND1*	MALDI-TOF MS	(Turnbull *et al.*, 2010)
54	rs3903072	11q13.1	*CFL1/OVOL1*	MALDI-TOF MS	(Michailidou *et al.*, 2013)

55	rs554219	11q13.3	*CCND1*	MALDI-TOF MS	(French *et al.*, 2013b)
56	rs11820646	11q24.3	*BARX2*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
57	rs12422552	12p13.1	*ATF7IP*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
58	rs17356907	12q22	*NTN4*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
59	rs1292011	12q24.21	*MED13L/TBX3*	MALDI-TOF MS	(Ghoussaini *et al.*, 2012)
60	rs2236007	14q13.3	*PAX9*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
61	rs999737	14q24.1	*RAD51B*	MALDI-TOF MS	(Thomas *et al.*, 2009)
62	rs2588809	14q24.1	*RAD51B*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
63	rs941764	14q32.11	*CCDC88C*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
64	rs17817449	16q12.2	*FTO*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
65	rs11075995	16q12.2	*FTO*	MALDI-TOF MS	(Garcia-Closas *et al.*, 2013b)
66	rs3803662	16q12.1	*TOX3*	MALDI-TOF MS	(Stacey *et al.*, 2007)
67	rs4784227	16q12.1	*TOX3*	MALDI-TOF MS	(Udler *et al.*, 2010b)
68	rs13329835	16q23.2	*CDYL2*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
69	rs6504950	17q22	*STXBP4*	MALDI-TOF MS	(Ahmed *et al.*, 2009)
70	rs1436904	18q11.2	*CHST9*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
71	rs527616	18q11.2	*AQP4*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
72	rs8170	19p13.11	*BABAM1=MERIT40*	MALDI-TOF MS	(Antoniou *et al.*, 2010b)
73	rs2363956	19p13.11	*ANKLE1*	MALDI-TOF MS	(Antoniou *et al.*, 2010b)
74	rs4808801	19p13.11	*ELL*	KASP assay	(Michailidou *et al.*, 2013)
75	rs3760982	19q13.31	*KCNN4/LYPD5*	MALDI-TOF MS	(Michailidou *et al.*, 2013)
76	rs2284378	20q11.22	*RALY*	MALDI-TOF MS	(Siddiq *et al.*, 2012)
77	rs2823093	21q21.1	*NRIP1*	MALDI-TOF MS	(Ghoussaini *et al.*, 2012)
78	rs6001930	22q13.1	*MKL1*	MALDI-TOF MS	(Michailidou *et al.*, 2013)

MALDI-TOF MS: matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

*Genome assembly GRCh37.

3.2.1. MALDI-TOF and TaqMan genotyping analyses

Genomic DNA was extracted from 9 ml of peripheral blood collected in an EDTA tube using the salting out extraction method (Laitinen *et al.*, 1994). Homogenous MassEXTEND and iPLEX GOLD methodology were carried out according to the manufacturer´s instructions (http://www.sequenom.com). Multiplex PCR assays were designed using the online tool Assay Design Suite (http://agenabio.com/products/massarraysystem/software/). In brief, 6.25 ng of genomic DNA was amplified by PCR in a final volume of 6.2 µl containing locus-specific primers at 100 nM final concentrations and 0.5 units HotStarTaq DNA Polymerase (Qiagen, Hilden, Germany). PCR conditions were 94°C for 15 minutes for Hot start, followed by 45 cycles of denaturation at 94°C for 20 seconds, annealing at 56°C for 30 seconds, and extension for 1 minute at 72°C, finally followed by incubation at 72°C for 3 minutes. PCR products were treated with 0.6 units of shrimp alkaline phosphatase (SAP) enzyme (iPLEX Gold Reagent Kit, Sequenom) for 40 minutes at 37°C to remove 5'- and 3'-phosphate groups from unincorporated dNTPs followed by 5 minutes at 85°C to inactivate SAP. After adjusting the concentrations of the extension primers according to their masses the base extension (homogenous MassEXTEND (hME^{TM}), Sequenom) reactions were carried out in a final volume of 10 µl containing 0.6 units polymerase (iPLEX enzyme, iPLEX Gold Reagent Kit, Sequenom). Base extension reaction conditions were 94°C for 30 seconds, followed by 40 cycles of 94°C for 5 seconds, 52°C for 5 seconds, and 80°C for 5 seconds. Within each of the 40 cycles the 52°C and 80°C steps were repeated 5 times. The final incubation was carried out at 72°C for 3 minutes. All reactions including PCR amplification, SAP treatment, and base extension were carried out in a DNA Engine DYAD PTC 220 PCR thermal cycler (MJ Research, USA). For a final volume of 26 µl, 16 µl of DNase/RNase free water (Life Technologies, USA) was added to each base

extension reaction. The final base extension products were treated with SpectroCLEAN resin (Sequenom) to remove salts from the reaction buffer. Following a quick centrifugation (2,000 rpm, 3 minutes in an Eppendorf Centrifuge 5810, Hamburg, Germany), 10 nl of reaction solution was dispensed onto a 384 format SpectroCHIP microarray (Sequenom) prespotted with a matrix of 3-hydroxypicolinic acid (3-HPA) by using a SpectroPointnanodispenser (Sequenom). A modified Bruker Biflex MALDI-TOF MS was used for data acquisitions from the SpectroCHIP. Genotype calls were made in real time with MassARRAY RT software vers. 3.0.0.4 (Sequenom). Post run cluster analysis was performed with MassARRAYTyper vers.4.0 (Sequenom). All primers were synthesized by Sigma-Aldrich Chemie GmbH, Steinheim, Germany. Primer sequences are shown in Supporting Information, Table S1.

Figure 5 shows exemplary genotype cluster plots and MALDI-TOF MS spectra representing homozygous and heterozygous genotypes for rs2981578 (*FGFR2*), rs941764 (*CCDC88C*), rs4784227 (*TOX3*), and rs2823093 (*NRIP1*).

For genotyping by TaqMan allelic discrimination analysis, 6.25 ng of genomic DNA was amplified in a final volume of 5 µl containing DNase/RNase free water, TaqMan® Genotyping Master Mix (Applied Biosystems, USA), and 20x SNP Genotyping Assay according to the manufacturer's instructions. PCR conditions were 95°C for 10 minutes for hot start, followed by 60 cycles of denaturation at 95°C for 15 seconds, annealing at 60°C for 1 minute, and a final cooling step at 40°C for 30 seconds. Real time PCR was performed with the LightCycler 480 Real-Time PCR System (Roche Diagnostics International AG, Switzerland) using its Endpoint Genotyping application module.

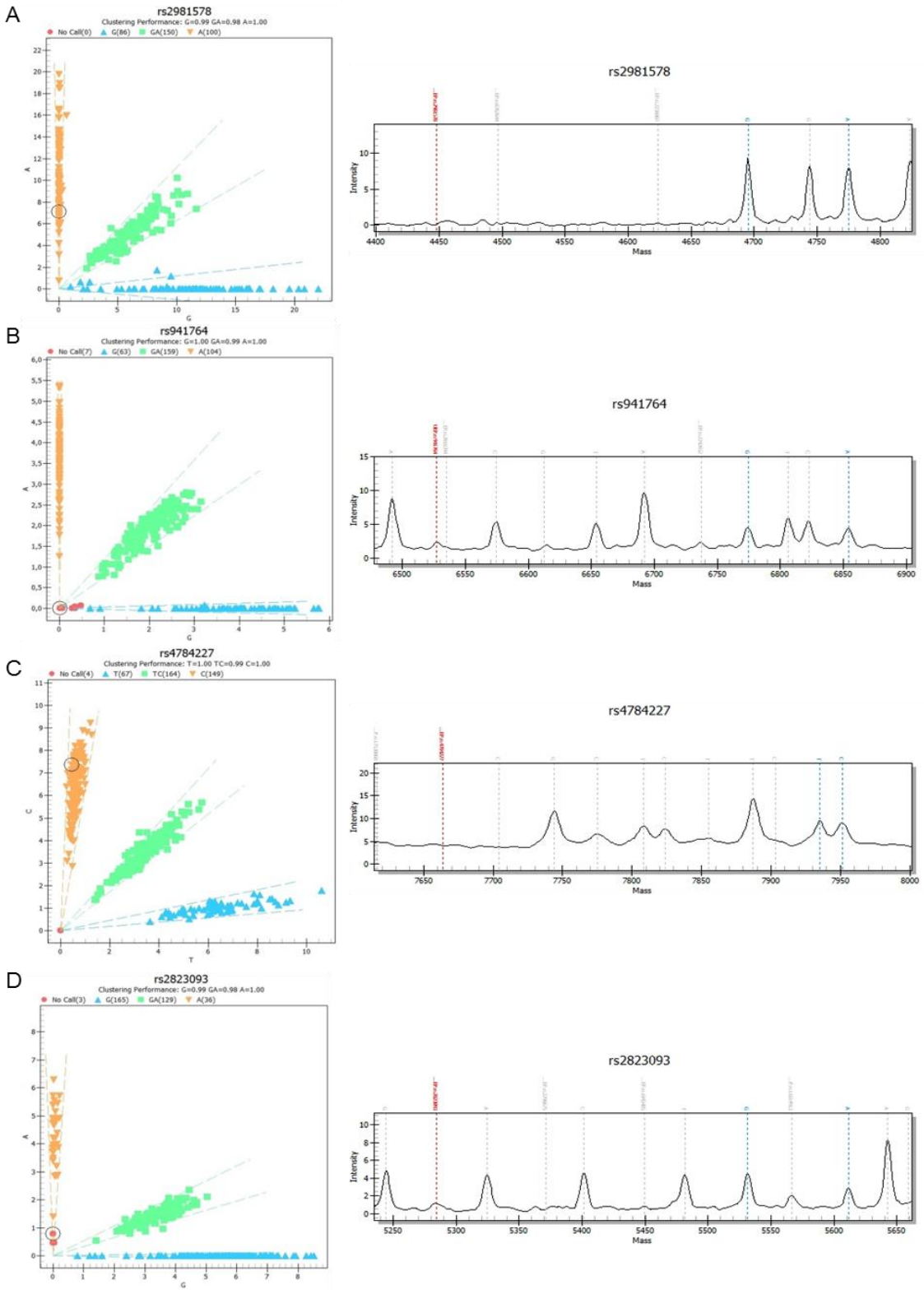

Figure 5: MALDI-TOF MS genotyping analysis. Exemplary genotype cluster plots (left panels) and MS spectra (right panels) for rs2981578 (A>G in FGFR2) (A), rs941764 (A>G in CCDC88C) (B), rs4784227 (C>T in TOX3) (C), and rs2823093 (G>A in NRIP1) (D). Green squares represent heterozygous genotypes, yellow and blue triangles represent homozygous genotypes, and red circles represent negative controls and drop outs. Masses of the expected unextended primers are marked by the red dotted line and both extension products (allele 1 and allele 2) by the blue dotted lines. Masses are given in Dalton.

3.3. Statistical analyses

3.3.1. Imputation of missing ancestry informative markers

Genotype imputation is the process of estimating and replacing missing values in untyped genotypes in genetic studies (Li *et al.*, 2009). Standard genotype imputation methods, like IMPUTE2, rely on linkage disequilibrium patterns and are not appropriate for a set of unlinked variants. Therefore, alternative statistical techniques for multiple imputations were applied.

3.3.1.1. Mechanism, patterns and types of variables for missing data

The generated missing values follow a process that is described by the missing data mechanism. Three types of missingness mechanisms are described in the literature (He, 2010):

- Missing completely at random (MCAR): The probability a variable is missing does not depend on any other observed variables. The missing values have no systematic differences with the observed values.
- Missing at random (MAR): The probability of missingness depends only on observed variables. In other words, the missingness on a given variable can be predicted using other variables in the dataset.
- Missing not at random (MNAR): The value of the unobserved variable itself predicts missingness.

A correct assumption about the missing data mechanism is essential for accurate statistical inferences. MAR mechanism cannot be tested considering the observed data alone and MNAR assumption claim something that is not accessible from the observed data. However, it is possible to check if the data follow an MCAR

mechanism, comparing if the dataset with and without missing data didn't have significant differences (He, 2010).

A Monotone

Group	Y1	Y2	Y3
1	X	X	X
2	X	X	.
3	X	.	.

B Non-monotone

Group	Y1	Y2	Y3
1	X	X	X
2	X	.	X
3	X	X	.
4	.	.	X

Figure 6: Patterns of missing data.

The missing data patterns can be classified in monotone or not. A dataset has a monotone missing pattern when V_j is missing then all the subsequent variables V_k, k>j are also missing (Figure 6) (Dong and Peng, 2013).

Another important consideration in the quality of an imputation is the proportion of missing data. Schafer considered that a missing rate less or equal than 5% is irrelevant (Schafer, 1999). On the other hand, if more than 10% of data are missing, there is a risk of biased statistical analysis (Bennett, 2001).

The type of variables (numeric or character) that either will contribute to the imputation process or is required for the imputation method should be taken into account in planning an imputation. Special consideration to the variable type will help ensure that the imputation is well conducted (Hayati Rezvan *et al.*, 2015).

3.3.1.2. Multiple imputation

The multiple imputation approach assumes MAR condition, creates multiple completed datasets, generating possible replacements for each missing value (He, 2010) (Figure 7).

There is not a consensus of how many imputations should be done. Schafer (1999) stated that “unless rates of missing information are unusually high, there tends to be little or no practical benefit to using more than five to ten imputations” (Schafer, 1999). Bodner (2008) recommends having as many imputations as the percentage of missing data (Bodner, 2008). One parameter used to select the number of imputations is the relative efficiency (Pan *et al.*, 2014):

$$RE = (1 + \frac{\lambda}{m})^{-1/2}$$

where m is the number of imputations and λ the proportion of missing data. The number of imputations is directly proportional to the amount of missing information. Once the complete datasets are generated, the second step is to fit statistical methods to each imputed dataset. Estimations in each of the imputed datasets will differ because of the variation present in the multiple imputations. The next step is to average the estimations in order to obtain an overall result (Sterne *et al.*, 2009).

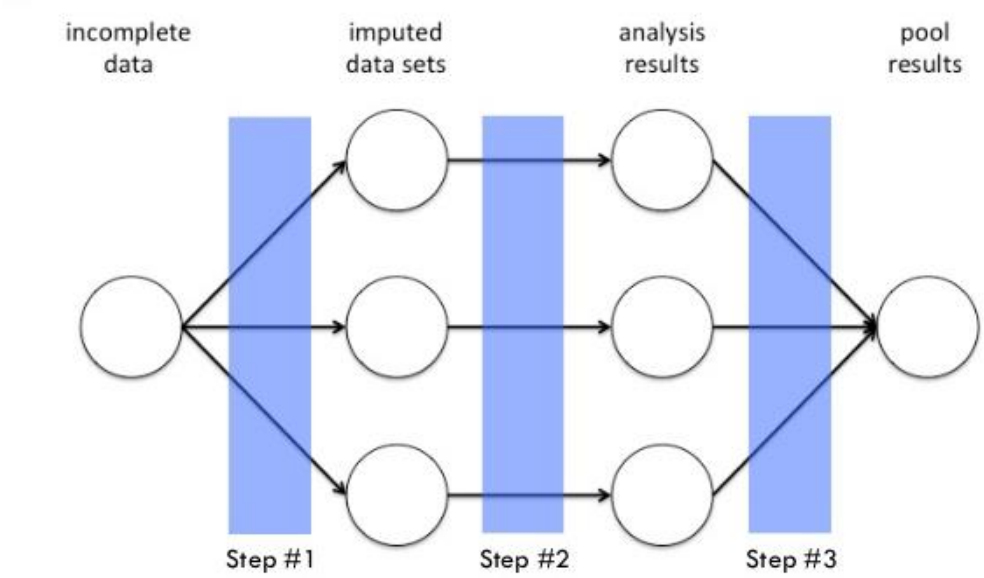

Figure 7: Steps in multiple imputation. The process starts with an incomplete dataset (on the left side), which is imputed *n* times (*n*=3) thus creating *n* completed datasets. Each complete dataset is analyzed, resulting in *n* analysis results. Finally, these *n* results are pooled into one final result (Adapted from https://www.slideshare.net/otrec/kim-slides-01232015).

3.3.1.3. The MI procedure in SAS

The SAS PROC MI package includes multiple imputation procedures, depending on the pattern of missingness and the type of the imputed variables, as summarized in Table 3 (Inc, 2012).

There are three methods available for datasets with monotone missing patterns and continuous variables: a regression method, a predictive mean-matching method, or a propensity score method.

The logistic regression method or the discriminant function method can be used to impute missing values in datasets with classification variables and monotone missing patterns.

Markov chain Monte Carlo (MCMC) methods assume multivariate normality and can be applied in datasets with arbitrary missing patterns and continuous variables.

Table 3: Imputation methods in SAS Procedure MI

Pattern of missingness	Type of imputed variables	Type of covariates	Available methods
Monotone	Continuous	Arbitrary	Monotone regression
			Monotone predicted mean matching
			Monotone propensity score
Monotone	Classification (ordinal)	Arbitrary	Monotone logistic regression
Monotone	Classification (nominal)	Arbitrary	Monotone discriminant function
Arbitrary	Continuous	Continuous	MCMC full-data imputation
			MCMC monotone-data imputation
Arbitrary	Continuous	Arbitrary	FCS regression
			FCS predicted mean matching
Arbitrary	Classification (ordinal)	Arbitrary	FCS logistic regression
Arbitrary	Classification (nominal)	Arbitrary	FCS discriminant function

*Adapted from Table 56.5 Imputation Methods in PROC MI. SAS Procedures Guide. Version 9.3, Cary, NC: SAS Institute Inc.

FCS methods can be effectively applied to impute missing values for continuous and classification variables in datasets with arbitrary missing patterns.

3.3.1.4. Missing data problem in this study

Of the 2,045 study participants (1,022 cases and 1,023 controls), 1,344 (722 cases and 622 controls) had complete data available for all 30 AIMs. Complete non-missing data on all 30 AIMs is required in order to estimate the proportion of genetic ancestry. For this reason, missing marker genotypes were imputed for participants with at least 27 non-missing AIMs (1,842 samples, 989 cases and 853 controls). Table 4 shows the name of each AIM, the number, and percent of missing data. The data show moderate missing values for each AIM.

A multiple imputation analysis was carried out using the PROC MI procedure in SAS. A discriminant function method was chosen and data were imputed using seven iterations to achieve relative efficiencies of 99%.

Table 4: Missing values and percentages for each AIM

	Missing values	
AIM	**N**	**%**
rs10037656	16	0.87
rs10510511	10	0.54
rs10935320	18	0.98
rs11725412	12	0.65
rs1197062	13	0.71
rs12347078	20	1.09
rs1243370	140	7.60
rs12662498	26	1.41
rs1544450	5	0.27
rs1559163	20	1.09
rs17086231	8	0.43
rs174570	16	0.87
rs1834619	32	1.74
rs1849384	14	0.76
rs2042314	73	3.96
rs2052386	3	0.16
rs2176046	19	1.03
rs2426552	6	0.33
rs260690	11	0.60
rs2719921	11	0.60
rs356652	6	0.33
rs3870336	63	3.42
rs4145160	51	2.77
rs4769128	27	1.47
rs6119879	14	0.76
rs6464749	4	0.22
rs7018273	5	0.27
rs7134749	39	2.12
rs717225	10	0.54
rs734241	12	0.65

3.3.2. Estimation of ancestry proportions

Genetic ancestry estimation has been applied in genetic association studies to analyze population stratification. It explains the genetic basis for ethnic differences in disease susceptibility (Liu *et al.*, 2013).

Computer programs like STRUCTURE and ADMIXTURE are extensively used for estimating global genetic ancestry. STRUCTURE is based on a clustering approach which utilizes multi-locus genotype data to infer admixture proportions. It assigns

individuals to populations and estimates ancestral population allele frequencies in admixed populations (Pritchard *et al.*, 2000). The ADMIXTURE software uses the same statistical model as STRUCTURE with faster estimation time (Alexander *et al.*, 2009).

In this thesis, the ADMIXTURE program was used for supervised estimation of individual European, African, and Native American ancestry proportions. Reference individuals rely on 80 Utah residents with Northern and Western European ancestry and 87 Yorubas in Ibadan, Nigeria from the 1000 Genome Project as surrogates of European and African ancestry, respectively, and on 64 samples from the Americas in the Human Genome Diversity Project as Native American (Alexander *et al.*, 2009; Cavalli-Sforza, 2005; Genomes Project *et al.*, 2015; Price *et al.*, 2006).

3.3.3. Genetic principal component analysis for adjustment of population stratification

The EIGENSTRAT method for genetic principal component analysis (PCA) was conducted to reduce the high-dimensionality of genetic data and to stratify the set of individuals into genetically similar groups (Price *et al.*, 2006).

The method consists of three steps. Suppose there are N individuals genotyped at M SNPs. First, EIGENSTRAT centered and normalized the genotypes for each sample (Price *et al.*, 2006). For each individual i, let X_{ij} be the resulting genotype data at SNP j, and let Y_i be the phenotype information. The empirical covariance matrix Ψ between the samples $(i,j)th$ are calculated

$$\psi_{ij} = \frac{1}{M}\sum_{s=1}^{M}\frac{(X_{is}-2p_s)(X_{js}-2p_s)}{2p_s(1-p_s)}$$

where p_s is an allele frequency estimate at marker s. Intuitively, the top eigenvectors of a covariance matrix between samples describe the variability of the axes. Then, EIGENSTRAT adjusts for genotypes and phenotypes using the top principal components obtained from Ψ as covariates in a multi-linear regression

$$Y_i = \alpha + \beta_1 v_{i1} + \beta_2 v_{i2} + \cdots + \beta_K v_{iK} + \gamma X_{is} + \epsilon_i$$

Here, v_{iK} is the ancestry of individual i along the k^{th} axis of variation, which equals the i^{th} coordinate of the k^{th} eigenvector, and K is the number of axes of variation used to adjust for ancestry.

After ancestry adjustment, the method computes association statistics using ancestry-adjusted genotypes and phenotypes which a generalization of the Armitage trend χ^2 statistic. The objective is to test for correlations between two projected vectors into a space of reduced dimension. The test statistic is defined as

$$EG = (N - K - 1)Corr^2(X_s^{adj}, Y^{adj})$$

where X_s^{adj} is the adjusted genotype at marker s, defined as the residuals after regressing genotypes on the top K principal component. The adjusted Y^{adj} is similarly defined (Price *et al.*, 2006; Wu *et al.*, 2011). The top principal components reflect the major ancestral dimensions of the study.

3.3.4. Association and interaction analyses for common breast cancer susceptibility variants

Possible interactions with genetic ancestry and associations with newly identified common BC susceptibility variants were tested using logistic regression with BC as response variable. For individual genotypes an additive model was assumed. Separate statistical analyses were performed for any type of BC, ER+, and ER- BC.

3.3.5. Variance in liability explained by identified susceptibility variants

The heritability or variance in BC liability explained by associated risk variants was calculated based on a multifactorial liability threshold model (So *et al.*, 2011). This model proposes a latent continuous liability, which is assumed to follow a normal distribution with mean 0 and variance 1. The mean liability is assumed to differ between genotypes, denoting risk allele A and protective allele a. Calculations relied on the frequency of the high risk allele (P_A), the relative risks for homozygote (RR_1) and heterozygote genotypes (RR_2), and the cumulative BC risk by age 80 years (K) in Colombian (0.052) and European (0.087) women (Boyle and Ferlay, 2005; Hemminki and Bermejo, 2007).

The genotype frequencies of the three genotypes P_{aa}, P_{Aa}, and P_{AA} were calculated assuming Hardy-Weinberg disequilibrium. The penetrance for each genotype was calculated based on the relative risk and the overall probability of disease in the population, which is given by

$$f_{aa} = \frac{K}{(P_{aa} + P_{Aa} * RR_1 + P_{AA} * RR_2)}$$

$$f_{Aa} = RR_1 * f_{aa}$$

$$f_{Aa} = RR_2 * f_{aa}$$

Assuming that the residual variance of each genotype is 1, the genotype-specific mean liabilities and the overall mean liability are given by

$$\mu_{aa} = T - \phi^{-1}(1 - f_{aa})$$

$$\mu_{Aa} = T - \phi^{-1}(1 - f_{Aa})$$

$$\mu_{AA} = T - \phi^{-1}(1 - f_{AA})$$

$$\mu_{all} = P_{Aa} * \mu_{Aa} + P_{AA} * \mu_{AA}$$

where T is the liability threshold and ϕ^{-1} represents the quantile of the normal distribution.

The three genotypic classes have the same residual variance with different mean liabilities. The variance explained by the locus can be written as:

$$V^* = P_A^2(\mu_{AA} - \mu_{all})^2 + 2P_A P_a(\mu_{Aa} - \mu_{all})^2 + P_a^2(\mu_{aa} - \mu_{all})^2$$

To allow for comparisons with heritability estimates, a standardized variance in liability was calculated as:

$$V_g = \frac{V^*}{(1 + V^*)}$$

A scatter plot was used to compare the contribution of each risk variant to BC heritability in Colombian and European women.

3.3.6. Area under the Receiver Operating Characteristic curve

Logistic regression models were also fitted to study data to estimate the area under Receiver Operating Characteristic curve (AUC) as measure of discrimination ability for the established risk factors age, family history of BC in first-degree female relatives, oral contraceptive use, menopausal status and hormone therapy, BMI, current smoking, parity, age at first full-term pregnancy, age at menarche, and breastfeeding, alone and in combination with common variants associated with BC risk in Colombia and individual ancestry proportions.

3.3.7. Development of the multifactorial risk score

In order to further investigate the association between BC risk and the combined effects of established BC risk factors, common variants associated with BC risk in Colombia, and individual ancestry proportions, a MRS was built using the formula:

$$MRS = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \cdots + \beta_n x_n \qquad (1)$$

where β_k represents the estimated log ORs from a multiple logistic regression model. The model included the ten risk factors listed in Table 1, the estimated individual Native American proportions, and variants associated with risk in Colombia assuming an additive model taking into account risk estimates from the literature and from Colombian data (Lange *et al.*, 2013; Lange *et al.*, 2005).

4. Results

In this work, genetic polymorphisms and ancestry have been investigated for their role in BC risk in Colombian women. The results of the project are divided into six parts as listed below:

- Description of the study participants: baseline characteristics, geographical distribution, genetic ancestry, and BC incidence

- Associations between 78 recently identified common risk variants and BC risk in Colombian women

- Risk associations between 78 recently identified common risk variants and ER status in Colombian women

- Variance in BC liability for 78 common variants in Europe and Colombia

- Interactions between common variants and genetic ancestry on BC susceptibility

- Discrimination ability of established risk factors, genetic ancestry, and common BC susceptibility variants, and estimated BC risks based on a MRS

4.1. Description of the study participants: baseline characteristics, geographical distribution, genetic ancestry, and breast cancer incidence

The Col-BCCC study population included 1,022 BC patients unselected for family history and age at BC diagnosis, and 1,023 healthy controls for which detailed information and DNA samples were available. Their baseline characteristics are given in Table 1. Age, BMI, and age at menarche showed similar distributions in cases and controls. Different distributions were noticed for smoking status, parity, age at first full-term pregnancy, and breastfeeding. Study individuals affected by BC had larger proportions of family history of BC in first-degree female relatives and oral contraceptive use than non-affected controls. The use of hormone therapy was higher in postmenopausal women affected by BC than in non-affected controls.

Figure 8 shows the BC incidence rates in Colombia, geographical distribution of the study participants, results from a genetic PCA, and the estimated ancestry proportions. Panel 8A shows the regional BC incidence in Colombia using previously published data from the National Cancer Institute of Colombia corresponding to the period 2007-2011 (Pardo and Cendales, 2009) (http://www.cancer.gov.co/files/libros/archivos/incidencia1.pdf; Table 83). The geographical distribution of study participants is shown in Panel 8B (BC cases) and Panel 8C (controls). The majority of study women were recruited in the central part of Colombia (Cundinamarca and Huila regions). Panel 8D shows results from a genetic PCA of study individuals. The first principal component separated non-African from African ancestry components and explained 24.1% of genetic variability. The second principal component separated the Native American and European ancestry components and explained 10.8% of variability. The European proportions of study

women have the highest average with an amount of 55.3% (5th-95th percentiles: 28.2%-80.7%). Followed by the average of Native American proportions 37.8% (5th-95th percentiles: 11.7%-63.0%) and the average of African proportions was 5.3% (5th-95th percentiles: 0.0%-22.5%).

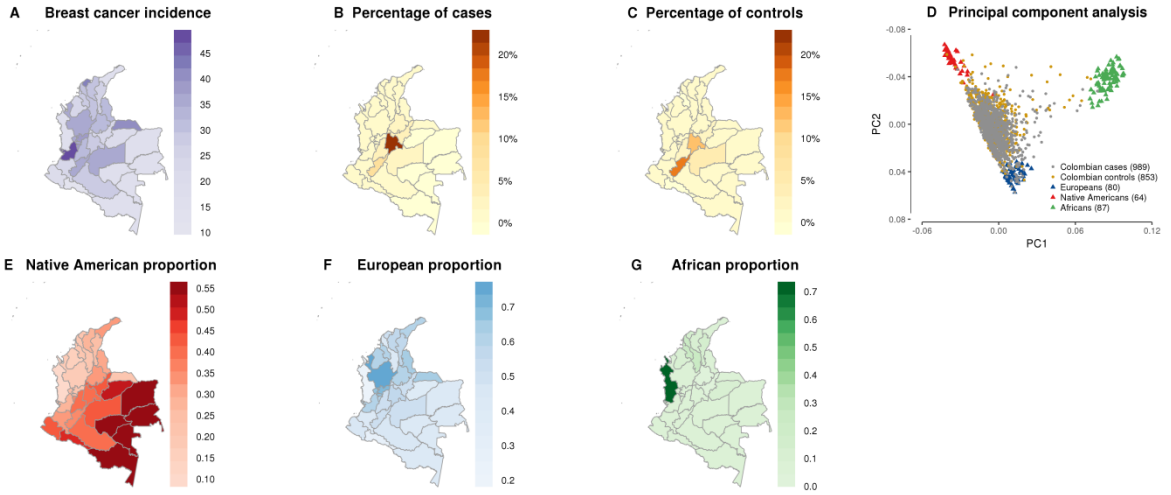

Figure 8: Breast cancer incidence rates in Colombia and description of the study participants. (A) Regional breast cancer incidence in Colombia. (B) Geographical distribution of the study cases. (C) Geographical distribution of the study controls. (D) First two principal components reflecting genetic variation among the study participants and three populations from the 1000 Genomes Project (Utah residents with Northern and Western European ancestry from the CEPH collection; Yoruba in Ibadan, Nigeria and the Human Genome Diversity Project (samples from the Americas). (E-G) Regional estimates of Native American, European, and African ancestry proportions.

Native American ancestry proportions were lower in Colombian BC patients than unaffected controls (P-value=5.2x10^{-16}). Per each 1% increase in the Native American proportion the unadjusted risk of BC decreased by 2.6% (95% CI: 2.0-3.2). After adjustment for established risk factors, the risk reduction was 2.2% (95% CI: 1.4-2.9). Contrarily, per each 1% increase in the European proportion the adjusted risk of BC increased by 1.6% (95% CI: 0.9-2.3).The association between BC risk and African ancestry proportions did not reach statistical significance (unadjusted P-value=0.32, adjusted P-value=0.08). Panels 8E-G show regional estimates of European, Native American, and African ancestry proportions based on study

individuals, confirming the inverse association between increasing Native American proportions and decreasing BC risk (Figure 8a, e).

4.2. Associations between 78 recently identified common risk variants and breast cancer risk in Colombian women

Out of 78 common variants robustly associated with BC risk in women of European origin, 13 showed nominal associations with BC risk in Colombia. Two variants rs3803662 (*TOX3*) and rs11199914 (*FGFR2*) revealed probability values smaller than 0.05 after potential cofounder and multiplicity adjustment using multiple permutation (Table 5). Complete results for all 78 investigated variants are presented in Supporting Information, Table S2. As expected, the allele frequencies of associated variants were partially different in Europeans and Colombians. For example, the major G allele for SNP rs4442975 (gene/nearest gene *IGFBP5*) among Europeans was the less frequent allele in Colombians, with a minor allele frequency equal to 0.35 (95% CI: 0.32-0.38, non-overlapping CIs). In general, adjustment for potential confounders barely affected the estimated ORs. Risk effects estimated for associated variants were also partially different in Europeans and Colombians. Estimated allele frequencies and risk effects for women of Asian and Colombian origin are compared in Supporting Information, Table S3.

Table 5: Common variants associated with breast cancer susceptibility in European and Colombian women

			EUROPEANS			COLOMBIANS										Ref.
SNP ID	**Locus**	**Gene/nearest gene(s)***	**Alleles Major/Minor**	**MAF**	**OR**	**AF[1]**	**95% CI**		**OR[1]**	**95% CI**		**P-value[2]**	**$OR_{adj.}$[3]**	**95% CI**		**P-value[2]**	
rs4442975	2q35	*IGFBP5*	G/T	0.49	0.87	**0.65**	0.62	0.68	0.82	0.65	0.96	**0.01**	0.82	0.62	0.99	**0.03**	(Ghoussaini *et al.*, 2014)
rs2046210	6q25.1	*ESR1/CCDC170*	G/A	0.34	1.08	**0.26**	0.24	0.29	1.18	1.03	1.36	**0.02**	1.30	1.11	1.54	**0.002**	(Michailidou *et al.*, 2013)
rs10759243	9q31.2	*KLF4*	C/A	0.39	1.06	**0.43**	0.40	0.46	1.15	1.01	1.30	**0.03**	1.18	1.01	1.37	**0.03**	(Michailidou *et al.*, 2013)
rs11199914	10q26.12	*FGFR2*	C/T	0.32	0.95	**0.43**	0.40	0.46	**0.79**	0.69	0.89	**0.0002**	0.77	0.66	0.89	**0.0007**	(Michailidou *et al.*, 2013)
rs35054928	10q26.13	*FGFR2*	-/C	0.44	1.27	**0.39**	0.36	0.42	1.20	1.06	1.36	**0.005**	1.17	1.00	1.35	**0.05**	(Meyer *et al.*, 2013)
rs2981579	10q26.13	*FGFR2*	G/A	0.40	1.27	0.39	0.36	0.42	1.21	1.06	1.37	**0.003**	1.17	1.00	1.35	**0.05**	(Michailidou *et al.*, 2013)
rs2981578	10q26.13	*FGFR2*	T/C	0.49	1.24	0.46	0.43	0.49	1.24	1.10	1.41	**0.0005**	1.20	1.03	1.38	**0.02**	(Meyer *et al.*, 2013)
rs2588809	14q24.1	*RAD51B*	C/T	0.16	1.08	0.16	0.14	0.18	**1.28**	1.09	1.51	**0.003**	1.31	1.07	1.59	**0.008**	(Michailidou *et al.*, 2013)
rs941764	14q32.11	*CCDC88C*	A/G	0.34	1.06	**0.42**	0.39	0.45	**1.21**	1.06	1.37	**0.004**	1.26	1.08	1.46	**0.003**	(Michailidou *et al.*, 2013)
rs3803662	16q12.1	*TOX3*	G/A	0.26	1.24	**0.43**	0.40	0.46	1.21	1.07	1.37	**0.003**	1.30	1.12	1.51	**0.0007**	(Michailidou *et al.*, 2013)
rs4784227	16q12.1	*TOX3*	C/T	0.25	1.19	**0.35**	0.32	0.38	1.21	1.07	1.38	**0.003**	1.20	1.03	1.40	**0.02**	(Long *et al.*, 2010)
rs13329835	16q23.2	*CDYL2*	A/G	0.22	1.08	**0.17**	0.15	0.20	1.22	1.04	1.43	**0.02**	1.23	1.02	1.49	**0.03**	(Michailidou *et al.*, 2013)
rs2823093	21q21.1	*NRIP1*	G/A	0.27	0.92	**0.32**	0.29	0.34	**0.77**	0.67	0.89	**0.0003**	0.84	0.71	0.99	**0.04**	(Michailidou *et al.*, 2013)

Abbreviations: MAF, minor allele frequency; AF, allele frequency; OR, odds ratio; adj., adjusted; CI, confidence interval.

*Genome assembly GRCh37.

[1]AFs and ORs different in Europe and Colombia (non-overlapping confidence intervals) are marked in bold.

[2]P-values smaller than 0.05 are marked in bold, P-values smaller than 0.05 after multiplicity correction using multiple permutation are underlined.

[3]OR adjusted for age, family history of breast cancer in first-degree female relatives, oral contraceptive use, menopausal status combined with postmenopausal hormone therapy use, body mass index, smoking status, parity, age at first full-term pregnancy, age at menarche, and breastfeeding.

4.3. Risk associations by estrogen receptor status

Risk associations by ER status of diagnosed breast tumors are shown in Table 6. One variant was specifically associated with ER+ disease and three variants with ER- disease in Colombian women. Out of 13 variants associated with BC risk in Colombia, three variants did not show an association with either ER+ or ER- disease, one variant was associated with ER- , seven with ER+, and three with ER+ and ER- BC.

However, the estimated risk differences for ER+ and ER- disease did not reach statistical significance. For example, SNP rs2823093 (*NRIP1*) showed an OR of 0.73 (95% CI: 0.60-0.89) for ER+ BC, compared to an OR of 0.94 (95% CI: 0.70-1.26) for ER- BC (overlapping CIs). Risk associations by ER status for all 78 investigated variants are presented in Supporting Information, Table S4.

4.4. Variance in breast cancer liability

Estimated risk allele frequencies were combined with genotype relative risks to assess the variance in BC liability due to susceptibility variants in European and Colombian women. Results are represented in Figure 9. Four out of 13 variants associated with BC risk in Colombia explained a larger proportion of heritability in Europe than in Colombia, with SNP rs35054928 (*FGR2*) explaining the largest variance proportion in Europe (0.0071) as previously reported. By contrast, nine associated variants showed a larger attributable heritability in Colombia than in Europe. The cumulative variance explained by all 13 SNPs was 0.04 (95% CI 0.03-0.06).

Table 6: Associations with breast cancer susceptibility by estrogen receptor status

SNP ID	Locus	Gene/nearest gene(s)*	ER-positive (N=592)				ER-negative (N=170)				Associated breast cancer type
			$OR_{adj.}$[1]	95% CI		P-value[2]	$OR_{adj.}$[1]	95% CI		P-value[2]	
rs889312	*5q11.2*	*MAP3K1*	1.25	1.05	1.49	**0.01**	0.85	0.65	1.11	0.24	ER+
rs6762644	3p26.1	*EGOT/ITPR1*	1.03	0.86	1.24	0.75	1.48	1.12	1.94	**0.005**	ER-
rs1011970	*9p21.3*	*CDKN2A/B*	1.00	0.84	1.19	0.99	1.38	1.07	1.79	**0.01**	ER-
rs6504950	*17q22*	*STXBP4*	0.99	0.80	1.23	0.95	1.46	1.08	1.98	**0.01**	ER-
rs2981579	*10q26.13*	*FGFR2*	1.16	0.97	1.38	0.10	1.33	1.02	1.74	**0.04**	overall, ER-
rs35054928	*10q26.13*	*FGFR2*	1.20	1.01	1.43	**0.04**	1.23	0.94	1.61	0.13	overall, ER+
rs2981578	*10q26.13*	*FGFR2*	1.24	1.05	1.48	**0.01**	1.27	0.98	1.65	0.07	overall, ER+
rs2588809	*14q24.1*	*RAD51B*	1.28	1.01	1.61	**0.04**	1.16	0.81	1.64	0.40	overall, ER+
rs3803662	*16q12.1*	*TOX3*	1.33	1.12	1.58	**0.001**	1.06	0.80	1.38	0.70	overall, ER+
rs4784227	*16q12.1*	*TOX3*	1.25	1.05	1.49	**0.01**	1.06	0.81	1.40	0.67	overall, ER+
rs2823093	*21q21.1*	*NRIP1*	0.73	0.60	0.89	**0.002**	0.94	0.70	1.26	0.69	overall, ER+
rs2046210	*6q25.1*	*ESR1/CCDC170*	1.33	1.10	1.61	**0.003**	1.34	1.00	1.78	**0.05**	overall, ER+, ER-
rs11199914	*10q26.12*	*FGFR2*	0.74	0.62	0.89	**0.001**	0.75	0.57	0.99	**0.04**	overall, ER+, ER-
rs941764	14q32.11	*CCDC88C*	1.30	1.09	1.55	**0.004**	1.42	1.08	1.86	**0.01**	overall, ER+, ER-

Abbreviations: ER+, estrogen receptor positive; ER-, estrogen receptor negative; OR, odds ratio; adj., adjusted; CI, confidence interval.

*Genome assembly GRCh37.

[1]OR adjusted for age, family history of breast cancer in first-degree female relatives, oral contraceptive use, menopausal status combined with postmenopausal hormone therapy use, body mass index, smoking status, parity, age at first full-term pregnancy, age at menarche, and breastfeeding.

[2]P-values smaller than 0.05 are marked in bold.

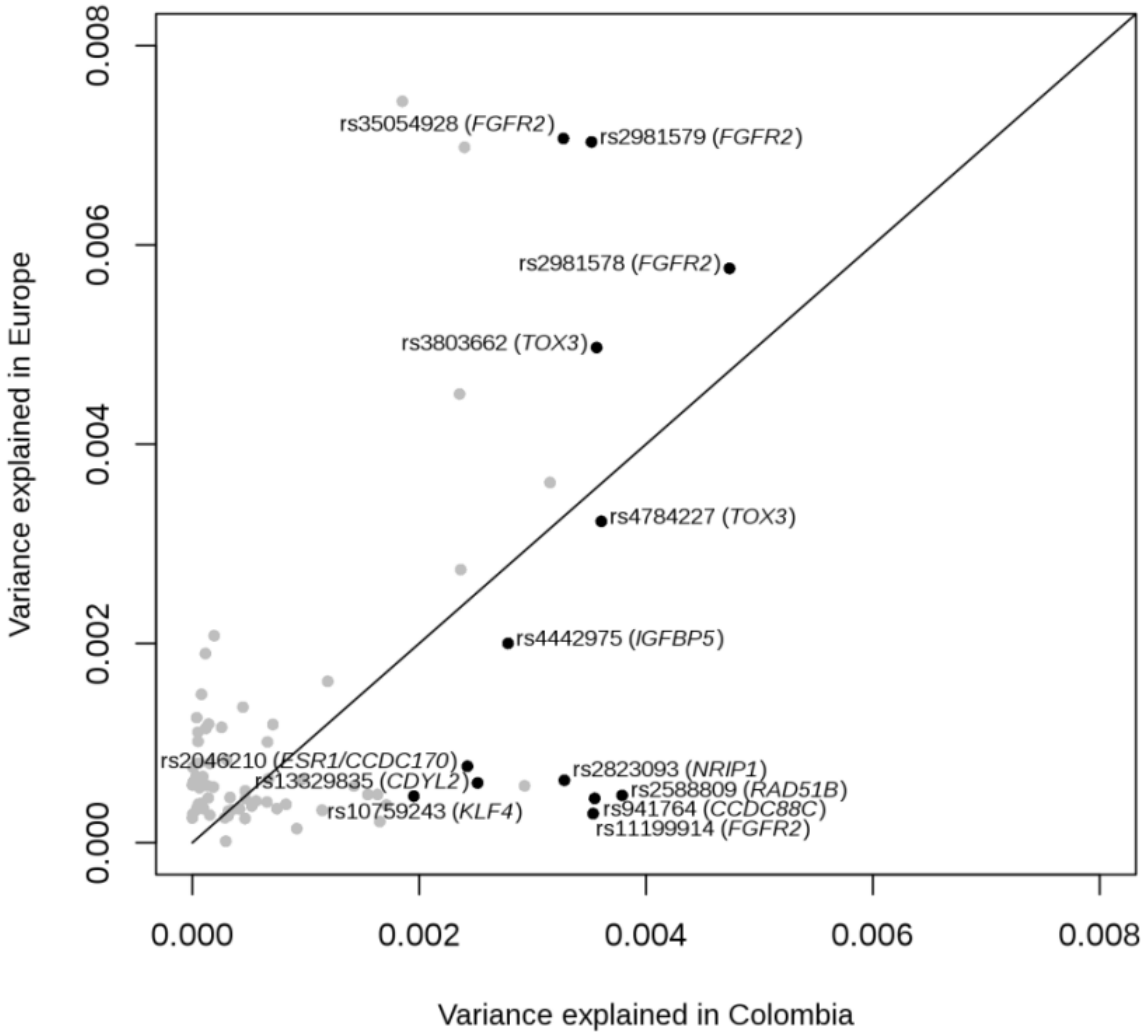

Figure 9: Fractions of explained variance for the 78 investigated common variants in Europe and Colombia. Black dots indicate the 13 variants with associated P-values smaller than 0.05; grey dots indicate the 65 variants with P-values higher than 0.05 in Colombia.

4.5. Interactions between common variants and genetic ancestry on breast cancer susceptibility

Results from logistic regression analyses revealed statistically significant interactions between two common BC susceptibility variants and genetic ancestry. Table 7 shows detailed results on the identified interactions with European proportions. When study women were grouped into four categories according to the quartiles of European proportions (1st quartile 45%, 2nd quartile 55%, 3rd quartile 66%), the strength of the association between SNP rs3803662 (*TOX3*) and BC risk decreased with increasing

European proportions from OR=1.32 for equal or less than 45% European ancestry to OR=1.07 for over 66% European ancestry. The opposite moderation of genetic risk by European ancestry was noticed for SNP rs941764 (*CCDC88C*). The OR increased from 0.91 for equal or less than 45% European ancestry to 1.76 for over 66% European ancestry. Interactions between the 13 risk SNPs and European and Native American proportions are shown in Supporting Information, Tables S5 and S6, respectively.

4.6. Discrimination ability of established risk factors, genetic ancestry and common breast cancer susceptibility variants, and estimated breast cancer risks based on a multifactorial risk score

The ability of established BC risk factors, Native American proportions, and variants associated with BC risk in Colombia to separate affected cases from unaffected healthy controls was investigated. The central part of Table 8 shows estimates based on own genotype and risk factor data combined with information from the literature on European genetic risk estimates and risk estimates for established BC risk factors in Hispanic women. The AUCs on the right of the table were based on own Colombian data only. Native American ancestry proportions resulted in the highest AUC (0.61), followed by the established risk factors “family history of BC in first-degree female relatives” (AUC=0.58), and “breastfeeding” (AUC=0.55). With the exception of the risk factor “parity”, AUC estimates based on the literature and on the own Colombian data were similar.

The combination of all 13 variants associated with BC risk in Colombia resulted in an AUC of 0.57, similar to the discrimination ability of Native American proportions. The true discrimination ability of the combined established risk factors, Native American proportions, and variants associated with BC risk in Colombia probably lies between

the rather conservative AUCs that integrate European and Hispanic risk estimates from the literature (AUC=0.65, 95% CI: 0.62-0.68) and the over-optimistic results exclusively based on Colombian data (AUC=0.71, 95% CI: 0.68-0.73).

Table 7: Assessment of potential interactions between two SNPs and genetic ancestry in the Colombian population

SNP ID	Locus	Gene/nearest gene(s)*	Interaction P-value[1]	Group of women according to the quartiles of European ancestry (ancestry range, N cases, N controls)											
				0-45.0%, 218, 243			45.1%-55.0%, 225, 235			55.1%-66.0%, 231, 229			66.1%-100%, 315, 146		
				$OR_{adj.}$[2]	95% CI		$OR_{adj.}$[2]	95% CI		$OR_{adj.}$[2]	95% CI		$OR_{adj.}$[2]	95% CI	
rs941764	14q32.11	*CCDC88C*	**0.02**	0.91	0.65	1.27	1.29	0.94	1.78	1.63	1.18	2.27	1.76	1.23	2.56
rs3803662	16q12.1	*TOX3*	**0.03**	1.32	0.96	1.84	1.46	1.06	2.04	1.35	0.99	1.85	1.07	0.74	1.56

Abbreviations: OR, odds ratio; adj., adjusted; CI, confidence interval.

*Genome assembly GRCh37.

[1]P-values smaller than 0.05 are marked in bold.

[2]OR adjusted for age, family history of breast cancer in first-degree female relatives, oral contraceptive use, menopausal status combined with postmenopausal hormone therapy use, body mass index, smoking status, parity, age at first full-term pregnancy, age at menarche, and breastfeeding.

Table 8: Estimated AUCs with their corresponding 95% confidence intervals (CIs) for established risk factors, genetic ancestry, and common breast cancer susceptibility variants based on risk estimates from the literature and own Colombian data

Risk factor		**Based on risk estimates from literature**							**Exclusively based on own Colombian data**					
Established risk factors		**OR**	**95%**	**CI**	**Ref.**	**AUC**	**95%**	**CI**	**OR**	**95%**	**CI**	**AUC**	**95%**	**CI**
Age (years)		1.043	1.041	1.046	(Ferlay *et al.*, 2015)	0.53	0.51	0.56	1.01	1.00	1.02	0.53	0.51	0.56
Family history of breast cancer in first-degree female relatives	No	Ref.			(Banegas *et al.*, 2017)				Ref.					
	Yes	2.48	1.67	3.68		0.58	0.56	0.60	3.04	2.30	4.07	0.58	0.56	0.60
Oral contraceptive use	No	Ref.			(Hines *et al.*, 2010)					Ref.				
	Yes	1.36	0.96	1.92		0.53	0.51	0.55	1.19	0.96	1.47	0.53	0.51	0.55
Menopausal status and hormone therapy use	Premenopausal and postmenopausal who never used hormone therapy	Ref.			(Amadou *et al.*, 2013)				Ref.					
	Postmenopausal women who ever used hormone therapy	1.78	1.32	2.41		0.54	0.52	0.55	2.11	1.41	3.20	0.54	0.52	0.55
Body mass index (kg/m^2)	<25	1.23	0.81	1.87					1.19	0.87	1.63			
	25-29.9	1.10	0.72	1.67	(Hines *et al.*, 2010)	0.53	0.51	0.56	1.00	0.73	1.37	0.53	0.51	0.55
	≥30	Ref.							Ref.					
Smoking status	Never	Ref.							Ref.					
	Former	1.43	1.04	1.96	(Connor *et al.*, 2016)	0.53	0.51	0.55	1.36	1.08	1.73	0.54	0.52	0.56
	Current	0.82	0.62	1.08					1.14	0.76	1.74			
Parity	Nulliparous	0.95	0.50	1.83					1.13	0.64	1.97			
	1-2	0.96	0.64	1.44	(Hines *et al.*, 2010)	0.46	0.44	0.49	1.29	1.02	1.64	0.53	0.51	0.56
	≥3	Ref.							Ref.					
Age at first full-term pregnancy (years)	<20	Ref.							Ref.					
	20-29	1.60	1.35	1.88	(Banegas *et al.*, 2017)	0.54	0.51	0.56	0.93	0.72	1.21	0.53	0.50	0.55
	>30	2.54	1.84	3.53					1.59	1.13	2.26			
Age at menarche (years)	≥14	Ref.							Ref.					
	12-13	1.30	1.12	1.50	(Banegas *et al.*, 2017)	0.51	0.49	0.54	1.12	0.90	1.38	0.52	0.49	0.54
	<12	1.68	1.26	2.25					1.07	0.80	1.44			
Breastfeeding (months)	Never	1.40	0.87	2.27					1.58	0.99	2.54			
	≤12	1.09	0.70	1.69	(Hines *et al.*, 2010)	0.55	0.52	0.57	1.10	0.86	1.40	0.55	0.52	0.57
	>12	Ref.							Ref.					
Established risk factors combined						0.61	0.58	0.64				0.64	0.61	0.67

Genetic ancestry														
Native American ancestry[1]		0.31	0.18	0.54	(Fejerman *et al.*, 2014)	0.61	0.58	0.63	0.07	0.04	0.13	0.61	0.58	0.63
Established risk factors and genetic ancestry combined						0.63	0.60	0.65				0.67	0.64	0.70
Breast cancer susceptibility variants														
rs4442975		0.87	0.86	0.89	(Ghoussaini *et al.*, 2014)	0.46	0.44	0.49	0.82	0.62	0.99	0.46	0.44	0.49
rs2046210		1.08	1.06	1.10	(Michailidou *et al.*, 2013)	0.52	0.50	0.54	1.30	1.11	1.54	0.52	0.50	0.54
rs10759243		1.06	1.04	1.08	(Michailidou *et al.*, 2013)	0.53	0.51	0.55	1.18	1.01	1.37	0.53	0.51	0.55
rs11199914		0.95	0.93	0.97	(Michailidou *et al.*, 2013)	0.54	0.52	0.56	0.77	0.66	0.89	0.54	0.52	0.56
rs35054928		1.27	1.24	1.29	(Meyer *et al.*, 2013)	0.53	0.51	0.56	1.17	1.00	1.35	0.53	0.51	0.56
rs2981579		1.27	1.24	1.29	(Michailidou *et al.*, 2013)	0.53	0.51	0.56	1.17	1.00	1.35	0.53	0.51	0.56
rs2981578		1.24	1.21	1.26	(Meyer *et al.*, 2013)	0.54	0.52	0.57	1.20	1.03	1.38	0.54	0.52	0.57
rs2588809		1.08	1.05	1.11	(Michailidou *et al.*, 2013)	0.53	0.51	0.55	1.31	1.07	1.59	0.53	0.51	0.55
rs941764		1.06	1.04	1.09	(Michailidou *et al.*, 2013)	0.54	0.51	0.56	1.26	1.08	1.46	0.54	0.51	0.56
rs3803662		1.24	1.21	1.27	(Michailidou *et al.*, 2013)	0.53	0.50	0.55	1.30	1.12	1.51	0.53	0.50	0.55
rs4784227		1.19	1.09	1.31	(Long *et al.*, 2010)	0.53	0.50	0.55	1.20	1.03	1.40	0.53	0.50	0.55
rs13329835		1.08	1.05	1.10	(Michailidou *et al.*, 2013)	0.53	0.51	0.55	1.23	1.02	1.49	0.53	0.51	0.55
rs2823093		0.92	0.90	0.94	(Michailidou *et al.*, 2013)	0.55	0.52	0.57	0.84	0.71	0.99	0.55	0.52	0.57
Breast cancer susceptibility variants combined						0.57	0.54	0.59				0.59	0.57	0.62
Established risk factors and breast cancer susceptibility variants combined						0.63	0.60	0.66				0.68	0.65	0.71
Breast cancer susceptibility variants and Native American proportion combined						0.60	0.57	0.63				0.64	0.62	0.67
Grand total						**0.65**	**0.62**	**0.68**				**0.71**	**0.68**	**0.73**

Abbreviations: OR, odds ratio; Ref., reference.

[1]For consistency with the publication by Fejerman et al. (Fejerman *et al.*, 2014), was used here a 0-1 scale for Native American ancestry instead of the 1% proportion scale used in the rest of the manuscript.

5. Discussion

Colombia is composed mainly of three racial groups: Europeans, Native Americans, and Africans. When the Spanish began the colonization of Colombia in 1525, they found several communities of Native Americans belonging to different tribes and linguistic groups. In the late fifteenth century, the population became more diverse with the arrival of Africans, who were brought to serve as labor source. The genetic characteristics of the Colombian population became more complex when the three ethnic groups began to mix, evolving to the modern Colombians (Ossa *et al.*, 2016).

This study examined for the first time the separate and combined contributions of established risk factors, genetic ancestry, and susceptibility variants to BC risk in this admixed population. The results showed that the risk effects conferred by common susceptibility variants are modified by Native American proportions. Also, Native American proportions can be used to predict BC risk as precisely as family history of BC. A new MRS was developed, showing that risk differences between Colombian women at high and low risk were as relevant as for women with European ancestry.

From a total of 78 SNPs previously found to be associated with overall BC risk in Europeans, thirteen common variants were found be also associated with BC risk in Colombian women, and explained 4% of BC heritability in Colombia. Three variants were identified to be specifically associated with ER+ and ER- BC risk. Fejerman and colleagues performed a GWAS including 1,497 US Hispanic BC cases and 3,213 controls and reported five identified associations at P-value<0.05 (rs10759243 (*KLF4*), rs2981579 (*FGFR2*), rs941764 (*CCDC88C*), rs3803662 and rs4784227 (*TOX3)* (Fejerman *et al.*, 2014). On the contrary, five variants associated with BC in this Colombian study (rs2046210 (*ESR1/CCDC170*), rs11199914 (*FGFR2*),

rs2588809 (*RAD51B*), rs13329835 (*CDYL2*), and rs2823093 (*NRIP1*)) were investigated in the same GWAS, but no associations were found. The distinct results may be due to differences between the study designs and samples sizes, as well as to dissimilarities between US Hispanics and Colombians in genetic ancestry and exposure to external factors.

High European ancestry proportions have been associated with increased BC risk among Mexican women and US women of Latin American origin (Fejerman *et al.*, 2008; Fejerman *et al.*, 2010; Slattery *et al.*, 2012). These results are consistent with the present study, where the risk of developing BC decreased with the increasing Native American ancestry proportions. By local admixture mapping, two genetic regions associated with BC risk have been identified: a 6q25 region near to the *ESR1* gene and a region on 11p15 (Fejerman *et al.*, 2012). Equivalently with the conclusions obtained for global ancestry, the magnitude of the association signals in the two identified regions decreased with increasing Native American proportions.

A study that included 2,107 US Hispanic and Mexican women diagnosed with BC and 2,590 healthy controls had reported statistical interactions between Native American ancestry proportions and genetic variants associated with BC (Fejerman *et al.*, 2013). The risk allele for three interacting variants rs13387042 (2q35), rs17157903 (*RELN*), and rs7696175 (*TLR1*) showed the strongest association in the group of women with the higher Native American proportions. Two statistical interactions were identified in the present Colombian study between global ancestry and associated variants. For rs941764 (*CCDC88C*), BC relative risks tended to increase with decreasing European ancestry proportions, while for rs3803662 (*TOX3*) a decrease was observed.

For the first time, a MRS was developed taking into account Native American proportions, well-established risk factors, and the 13 genetic variants associated with BC risk in Colombia. The MRS resulted in a good discriminative accuracy with an AUC between 0.65 and 0.71. An important finding was that the discriminative ability of established BC susceptibility factors such as family history of BC in first-degree female relatives was similar to the capacity of Native American ancestry proportions to discern Colombian women who developed BC from non-affected controls. A similar discriminative ability was also shown for the 13 genetic variants associated with BC risk in Colombia and Native American ancestry proportions. These results emphasize the importance of individual estimates of Native American ancestry proportions on the prediction of BC risk in Colombians and other Hispanic populations. Improvement in the quality of BC risk prediction model in Latino women may be achieved through more precise assessments of environmental exposures and by integrating newly identified susceptibility variants in future risk prediction models.

Recent reports on known BC risk factors together with genomic profiles suggested insubstantial improvements in BC risk discrimination approach (Husing *et al.*, 2012; Wacholder *et al.*, 2010). More recently, polygenic risk scores based on large sets of susceptibility variants have been shown to be useful for BC risk discrimination in women of European ancestry from the general population, in males, and in female carriers of *BRCA1* and *BRCA2* mutations (Kuchenbaecker *et al.*, 2017; Lecarpentier *et al.*, 2017; Mavaddat *et al.*, 2015). In the current study, a MRS was developed based on risk estimates from the literature combined with own genotype and risk factor data. The MRS combined established risk factors, 13 susceptibility variants, and Native American proportions. The discrimination ability of established risk factors

(AUC=0.61) increased to AUC=0.63 by considering Native American proportions, and further to AUC=0.65 by integrating recently identified BC susceptibility variants.

The Colombian study has several weaknesses. Due to the limited size of the study including 1,022 BC patients and 1,023 healthy controls, only common BC susceptibility variants identified in the first large genome-wide association scans and fine-mapping studies in Europe were investigated. Many additional risk variants have been newly detected (Michailidou *et al.*, 2017) and ongoing studies will identify other novel variants that should be investigated in future Colombian studies. A further limitation of this study was the presented missing genotypes for susceptibility variants and AIMs. Nevertheless, the ancestry proportion estimations were obtained with high accuracy (at least 99% correlation between estimates based on complete and imputed data). The number of participants was small to test the associations of 78 variants. However the main goal of this study was to describe the associated risk effects, frequency, and potential interactions of these variants with genetic ancestry in the Colombian population. The majority of the study participants were recruited in the central area of Colombia, the Cundinamarca and Huila regions, and their representativeness for the whole country is questionable. One of the strengths of this study was the access to appropriate genotype data to investigate a timely hypothesis, which resulted in relevant findings for BC prevention in Latinas.

Finally, the findings demonstrate that for BC risk prediction in Colombian women a MRS could be useful. A risk prediction model for Colombian women that combined established BC risk factors, newly identified susceptibility variants, and Native American proportions may constitute a powerful tool for clinical applications in stratified BC prevention programs in Latin American and Hispanic populations.

6. Summary

The etiology of BC involves both non-genetic and genetic factors. Environmental and lifestyle factors such as age, use of menopausal hormone therapy, smoking, and BMI have been associated with the risk of developing BC. Genetic susceptibility determined mainly by family history and ethnic background have an important role in the risk of developing this disease (Dossus and Benusiglio, 2015). In recent years, novel variants robustly associated with BC risk have been identified in large-scale genetic association studies in women of European and Asian origin. However, few studies directed towards the identification of BC susceptibility variants have been conducted among Latin American and Hispanic populations.

This thesis examined the contributions of genetic ancestry, established risk factors, and newly identified susceptibility variants to BC risk in Colombia. A total of 2,045 participants from the Col-BCCC study were included in this analysis: 1,022 BC patients, and 1,023 healthy controls. BC patients were unselected for family history and age at BC diagnosis. European, Native American, and African ancestry proportion were quantified in each woman based on 30 AIMs, aiming to obtain the relationship between ancestry and BC risk. Seventy-eight previously identified common BC susceptibility variants were genotyped and associations of these variants with BC risk in the Colombian population were determined. To assess the interactions between the variants and ancestry proportions logistic regression models were applied.

Native American proportions were lower in Colombian BC patients than in unaffected controls (P-value=5.2x10-16). This difference translated into an unadjusted decreased BC risk of 2.6% per each 1% increase in the Native American proportion

(95% CI: 2.0-3.2). Associations with BC risk in Colombian women were obtained for thirteen variants, which in comparison with European women have partially different risk effects and allele frequencies. The risk effects of rs941764 (*CCDC88C*) and rs3803662 (*TOX3*) was controlled for ancestry proportions. One variant was associated with ER-, seven with ER+, and three with ER+ and ER- disease. The variance in BC liability due to susceptibility variants in European and Colombian women was estimated. Out of 13 variants associated with BC risk in Colombia, four explained a larger attributable heritability in Europe than in Colombia and nine revealed larger attributable heritability in Colombia than in Europe.

AUCs with their corresponding 95% CIs were estimated for established risk factors, genetic ancestry, and common BC susceptibility variants based on risk estimates from the literature and own Colombian data. The discriminative ability to separate Colombian cases and controls of family history of BC in first-degree female relatives (AUC=0.58) and the combination of all 13 associated risk variants (AUC=0.57) were similar to the discriminative ability of Native American proportions (AUC=0.61).

The findings demonstrate that individual ancestry proportions predict BC risk in Colombia as accurately as established BC risk factors. Combining Native American proportions, established risk factors, and newly identified genetic susceptibility variants could translate in promising clinical strategies on BC prevention in Latin American and Hispanic women.

7. Zusammenfassung

Die Ätiologie von Brustkrebs (BC) beinhaltet sowohl nicht-genetische als auch genetische Faktoren. Umwelt- und Lebensstilfaktoren wie Alter, Anwendung der Hormontherapie in den Wechseljahren, Rauchen und Body-Mass-Index (BMI) wurden mit dem Risiko der Entwicklung von BC in Verbindung gebracht. Die genetische Anfälligkeit, die hauptsächlich durch die Familiengeschichte und den ethnischen Hintergrund bestimmt wird, spielt eine wichtige Rolle für das Risiko, an dieser Krankheit zu erkranken (Dossus and Benusiglio, 2015). In den letzten Jahren wurden in groß angelegten genetischen Assoziationsstudien bei Frauen europäischer und asiatischer Herkunft neuartige Varianten identifiziert, die stark mit dem BC-Risiko verbunden sind. Es wurden jedoch nur wenige Studien zur Identifizierung von BC-Anfälligskeitsvarianten unter lateinamerikanischen und hispanischen Bevölkerungen durchgeführt.

Diese Arbeit untersuchte die Beiträge genetischer Abstammung, etablierte Risikofaktoren und neu identifizierte Anfälligkeitsvarianten für das BC-Risiko in Kolumbien. Insgesamt wurden 2.045 Teilnehmer einer kolumbianischen Brustkrebs Kontroll-Studie (Col-BCCC-Studie) in diese Analyse einbezogen: 1.022 BC-Patienten und 1.023 gesunde Kontrollpersonen. BC-Patienten wurden zum Zeitpunkt der BC-Diagnose nicht nach Familienanamnese und Alter ausgewählt. Der Anteil europäischer, indianischer und afrikanischer Vorfahren wurde bei jeder Frau anhand von 30 abstammungs-informativen Markern (AIMs) quantifiziert, um die Beziehung zwischen Abstammung und BC-Risiko zu ermitteln. Achtundsiebzig zuvor identifizierte häufige BC-Anfälligkeitsvarianten wurden genotypisiert und Assoziationen dieser Varianten mit dem BC-Risiko in der kolumbianischen Bevölkerung bestimmt. Um die Wechselwirkungen zwischen den Varianten und den

Abstammungsverhältnissen zu bewerten, wurden logistische Regressionsmodelle angewendet.

Die Anteile der amerikanischen Ureinwohner waren bei kolumbianischen BC-Patienten niedriger als bei nicht betroffenen Kontrollen (P-Wert = 5,2 × 10 –16). Dieser Unterschied führte zu einem nicht angepassten verringerten BC-Risiko von 2,6% pro 1% Anstieg des Anteils der amerikanischen Ureinwohner (95% CI: 2,0-3,2). Assoziationen mit dem BC-Risiko bei kolumbianischen Frauen wurden für dreizehn Varianten erhalten, die im Vergleich zu europäischen Frauen teilweise unterschiedliche Risikoeffekte und Allel-Frequenzen aufweisen. Die Risikoeffekte von rs941764 (CCDC88C) und rs3803662 (TOX3) wurden hinsichtlich der Abstammungsverhältnisse kontrolliert. Eine Variante war mit Estrogenrezeptor negativ (ER-) assoziiert, sieben mit ER + und drei mit ER + und ER-. Die Varianz der BC-Haftung aufgrund von Anfälligkeitsvarianten bei europäischen und kolumbianischen Frauen wurde geschätzt. Von 13 Varianten, die mit dem BC-Risiko in Kolumbien assoziiert sind, erklärten vier eine größere zurechenbare Erblichkeit in Europa als in Kolumbien und neun zeigten eine größere zurechenbare Erblichkeit in Kolumbien als in Europa auf.

Die Fläche unter der Kurve (AUC) mit ihren entsprechenden 95% CIs wurden auf der Grundlage von Risikoschätzungen aus der Literatur und eigenen kolumbianischen Daten für etablierte Risikofaktoren, genetischer Abstammung und häufiger BC-Anfälligkeitsvarianten geschätzt. Die Unterscheidungsfähigkeit, kolumbianische Fälle und Kontrollen der Familiengeschichte von BC bei weiblichen Verwandten ersten Grades (AUC = 0,58) zu trennen, und die Kombination aller 13 assoziierten Risikovarianten (AUC = 0,57) waren ähnlich der Unterscheidungsfähigkeit der Anteile der amerikanischen Ureinwohner (AUC = 0,61).

Die Ergebnisse zeigen, dass die individuellen Abstammungsanteile das BC-Risiko in Kolumbien genauso akkurat vorhersagen wie die etablierten BC-Risikofaktoren. Die Kombination von Anteilen der amerikanischen Ureinwohner, etablierten Risikofaktoren und neu identifizierten Varianten der genetischen Anfälligkeit könnte zu vielversprechenden klinischen Strategien zur BC-Prävention bei lateinamerikanischen und hispanischen Frauen führen.

8. References

Agarwal, G., Pradeep, P. V., Aggarwal, V., Yip, C. H. and Cheung, P. S. (2007). **Spectrum of breast cancer in Asian women**. World J Surg *31*, 1031-1040, doi: 10.1007/s00268-005-0585-9.

Ahmed, S., Thomas, G., Ghoussaini, M., Healey, C. S., Humphreys, M. K., Platte, R., Morrison, J., Maranian, M., Pooley, K. A., Luben, R., Eccles, D., Evans, D. G., Fletcher, O., Johnson, N., dos, S. S., I, Peto, J., Stratton, M. R., Rahman, N., Jacobs, K., Prentice, R., Anderson, G. L., Rajkovic, A., Curb, J. D., Ziegler, R. G., Berg, C. D., Buys, S. S., McCarty, C. A., Feigelson, H. S., Calle, E. E., Thun, M. J., Diver, W. R., Bojesen, S., Nordestgaard, B. G., Flyger, H., Dork, T., Schurmann, P., Hillemanns, P., Karstens, J. H., Bogdanova, N. V., Antonenkova, N. N., Zalutsky, I. V., Bermisheva, M., Fedorova, S., Khusnutdinova, E., Kang, D., Yoo, K. Y., Noh, D. Y., Ahn, S. H., Devilee, P., van Asperen, C. J., Tollenaar, R. A., Seynaeve, C., Garcia-Closas, M., Lissowska, J., Brinton, L., Peplonska, B., Nevanlinna, H., Heikkinen, T., Aittomaki, K., Blomqvist, C., Hopper, J. L., Southey, M. C., Smith, L., Spurdle, A. B., Schmidt, M. K., Broeks, A., van Hien, R. R., Cornelissen, S., Milne, R. L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Schmutzler, R. K., Burwinkel, B., Bartram, C. R., Meindl, A., Brauch, H., Justenhoven, C., Hamann, U., Chang-Claude, J., Hein, R., Wang-Gohrke, S., Lindblom, A., Margolin, S., Mannermaa, A., Kosma, V. M., Kataja, V., Olson, J. E., Wang, X., Fredericksen, Z., Giles, G. G., Severi, G., Baglietto, L., English, D. R., Hankinson, S. E., Cox, D. G., Kraft, P., Vatten, L. J., Hveem, K., Kumle, M., Sigurdson, A., Doody, M., Bhatti, P., Alexander, B. H., Hooning, M. J., van den Ouweland, A. M., Oldenburg, R. A., Schutte, M., Hall, P., Czene, K., Liu, J., Li, Y., Cox, A., Elliott, G., Brock, I., Reed, M. W., Shen, C. Y., Yu, J. C., Hsu, G. C., Chen, S. T., Anton-Culver, H., Ziogas, A., Andrulis, I. L., Knight, J. A., Beesley, J., Goode, E. L., Couch, F., Chenevix-Trench, G., Hoover, R. N., Ponder, B. A., Hunter, D. J., Pharoah, P. D., Dunning, A. M., Chanock, S. J. and Easton, D. F. (2009). **Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2**. Nat. Genet *41*, 585-590, doi: ng.354 [pii];10.1038/ng.354 [doi].

Alexander, D. H., Novembre, J. and Lange, K. (2009). **Fast model-based estimation of ancestry in unrelated individuals**. Genome Res *19*, 1655-1664, doi: 10.1101/gr.094052.109.

Amadou, A., Fabre, A., Torres-Mejia, G., Ortega-Olvera, C., Angeles-Llerenas, A., McKenzie, F., Biessy, C., Hainaut, P. and Romieu, I. (2013). **Hormonal therapy and risk of breast cancer in mexican women**. PLoS One *8*, e79695, doi: 10.1371/journal.pone.0079695.

Amos, C. I., Dennis, J., Wang, Z., Byun, J., Schumacher, F. R., Gayther, S. A., Casey, G., Hunter, D. J., Sellers, T. A., Gruber, S. B., Dunning, A. M., Michailidou, K., Fachal, L., Doheny, K., Spurdle, A. B., Li, Y., Xiao, X., Romm, J., Pugh, E., Coetzee, G. A., Hazelett, D. J., Bojesen, S. E., Caga-Anan, C., Haiman, C. A., Kamal, A., Luccarini, C., Tessier, D., Vincent, D., Bacot, F., Van Den Berg, D. J., Nelson, S., Demetriades, S., Goldgar, D. E., Couch, F. J., Forman, J. L., Giles, G. G., Conti, D. V., Bickeboller, H., Risch, A., Waldenberger, M., Bruske-Hohlfeld, I., Hicks, B. D., Ling, H., McGuffog, L., Lee, A., Kuchenbaecker, K., Soucy, P., Manz, J., Cunningham, J. M., Butterbach, K., Kote-Jarai, Z., Kraft, P., FitzGerald, L., Lindstrom, S., Adams, M., McKay, J. D., Phelan, C. M., Benlloch, S., Kelemen, L. E., Brennan, P., Riggan, M., O'Mara, T. A., Shen, H., Shi, Y., Thompson, D. J., Goodman, M. T., Nielsen, S. F., Berchuck, A., Laboissiere, S., Schmit, S. L., Shelford, T., Edlund, C. K., Taylor, J. A., Field, J. K., Park, S. K., Offit, K., Thomassen, M., Schmutzler, R., Ottini, L., Hung, R. J., Marchini, J., Amin Al Olama, A., Peters, U., Eeles, R. A., Seldin, M. F., Gillanders, E., Seminara, D., Antoniou, A. C., Pharoah, P. D., Chenevix-Trench,

G., Chanock, S. J., Simard, J. and Easton, D. F. (2017). **The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers**. Cancer Epidemiol Biomarkers Prev *26*, 126-135, doi: 10.1158/1055-9965.EPI-16-0106.

Antoniou, A. C., Wang, X., Fredericksen, Z. S., McGuffog, L., Tarrell, R., Sinilnikova, O. M., Healey, S., Morrison, J., Kartsonaki, C., Lesnick, T., Ghoussaini, M., Barrowdale, D., Embrace, Peock, S., Cook, M., Oliver, C., Frost, D., Eccles, D., Evans, D. G., Eeles, R., Izatt, L., Chu, C., Douglas, F., Paterson, J., Stoppa-Lyonnet, D., Houdayer, C., Mazoyer, S., Giraud, S., Lasset, C., Remenieras, A., Caron, O., Hardouin, A., Berthet, P., Collaborators, G. S., Hogervorst, F. B., Rookus, M. A., Jager, A., van den Ouweland, A., Hoogerbrugge, N., van der Luijt, R. B., Meijers-Heijboer, H., Gomez Garcia, E. B., Hebon, Devilee, P., Vreeswijk, M. P., Lubinski, J., Jakubowska, A., Gronwald, J., Huzarski, T., Byrski, T., Gorski, B., Cybulski, C., Spurdle, A. B., Holland, H., kConFab, Goldgar, D. E., John, E. M., Hopper, J. L., Southey, M., Buys, S. S., Daly, M. B., Terry, M. B., Schmutzler, R. K., Wappenschmidt, B., Engel, C., Meindl, A., Preisler-Adams, S., Arnold, N., Niederacher, D., Sutter, C., Domchek, S. M., Nathanson, K. L., Rebbeck, T., Blum, J. L., Piedmonte, M., Rodriguez, G. C., Wakeley, K., Boggess, J. F., Basil, J., Blank, S. V., Friedman, E., Kaufman, B., Laitman, Y., Milgrom, R., Andrulis, I. L., Glendon, G., Ozcelik, H., Kirchhoff, T., Vijai, J., Gaudet, M. M., Altshuler, D., Guiducci, C., Swe, B., Loman, N., Harbst, K., Rantala, J., Ehrencrona, H., Gerdes, A. M., Thomassen, M., Sunde, L., Peterlongo, P., Manoukian, S., Bonanni, B., Viel, A., Radice, P., Caldes, T., de la Hoya, M., Singer, C. F., Fink-Retter, A., Greene, M. H., Mai, P. L., Loud, J. T., Guidugli, L., Lindor, N. M., Hansen, T. V., Nielsen, F. C., Blanco, I., Lazaro, C., Garber, J., Ramus, S. J., Gayther, S. A., Phelan, C., Narod, S., Szabo, C. I., Mod, S., Benitez, J., Osorio, A., Nevanlinna, H., Heikkinen, T., Caligo, M. A., Beattie, M. S., Hamann, U., Godwin, A. K., Montagna, M., Casella, C., Neuhausen, S. L., Karlan, B. Y., Tung, N., Toland, A. E., Weitzel, J., Olopade, O., Simard, J., Soucy, P., Rubinstein, W. S., Arason, A., Rennert, G., Martin, N. G., Montgomery, G. W., Chang-Claude, J., Flesch-Janys, D., Brauch, H., Genica, Severi, G., Baglietto, L., Cox, A., Cross, S. S., Miron, P., Gerty, S. M., Tapper, W., Yannoukakos, D., Fountzilas, G., Fasching, P. A., Beckmann, M. W., Dos Santos Silva, I., Peto, J., Lambrechts, D., Paridaens, R., Rudiger, T., Forsti, A., Winqvist, R., Pylkas, K., Diasio, R. B., Lee, A. M., Eckel-Passow, J., Vachon, C., Blows, F., Driver, K., Dunning, A., Pharoah, P. P., Offit, K., Pankratz, V. S., Hakonarson, H., Chenevix-Trench, G., Easton, D. F. and Couch, F. J. (2010a). **A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population**. Nat Genet *42*, 885-892, doi: 10.1038/ng.669.

Antoniou, A. C., Wang, X., Fredericksen, Z. S., McGuffog, L., Tarrell, R., Sinilnikova, O. M., Healey, S., Morrison, J., Kartsonaki, C., Lesnick, T., Ghoussaini, M., Barrowdale, D., Peock, S., Cook, M., Oliver, C., Frost, D., Eccles, D., Evans, D. G., Eeles, R., Izatt, L., Chu, C., Douglas, F., Paterson, J., Stoppa-Lyonnet, D., Houdayer, C., Mazoyer, S., Giraud, S., Lasset, C., Remenieras, A., Caron, O., Hardouin, A., Berthet, P., Hogervorst, F. B., Rookus, M. A., Jager, A., van den Ouweland, A., Hoogerbrugge, N., van der Luijt, R. B., Meijers-Heijboer, H., Gomez Garcia, E. B., Devilee, P., Vreeswijk, M. P., Lubinski, J., Jakubowska, A., Gronwald, J., Huzarski, T., Byrski, T., Gorski, B., Cybulski, C., Spurdle, A. B., Holland, H., Goldgar, D. E., John, E. M., Hopper, J. L., Southey, M., Buys, S. S., Daly, M. B., Terry, M. B., Schmutzler, R. K., Wappenschmidt, B., Engel, C., Meindl, A., Preisler-Adams, S., Arnold, N., Niederacher, D., Sutter, C., Domchek, S. M., Nathanson, K. L., Rebbeck, T., Blum, J. L., Piedmonte, M., Rodriguez, G. C., Wakeley, K., Boggess, J. F., Basil, J., Blank, S. V., Friedman, E., Kaufman, B., Laitman, Y., Milgrom, R., Andrulis, I. L., Glendon, G., Ozcelik, H., Kirchhoff, T., Vijai, J., Gaudet, M. M., Altshuler, D., Guiducci, C., Loman, N., Harbst, K., Rantala, J., Ehrencrona, H., Gerdes, A. M., Thomassen, M., Sunde, L., Peterlongo, P., Manoukian, S., Bonanni, B., Viel, A., Radice, P., Caldes, T., de la Hoya, M., Singer, C. F., Fink-Retter, A., Greene, M. H., Mai, P. L., Loud, J. T., Guidugli, L.,

Lindor, N. M., Hansen, T. V., Nielsen, F. C., Blanco, I., Lazaro, C., Garber, J., Ramus, S. J., Gayther, S. A., Phelan, C., Narod, S., Szabo, C. I., Benitez, J., Osorio, A., Nevanlinna, H., Heikkinen, T., Caligo, M. A., Beattie, M. S., Hamann, U., Godwin, A. K., Montagna, M., Casella, C., Neuhausen, S. L., Karlan, B. Y., Tung, N., Toland, A. E., Weitzel, J., Olopade, O., Simard, J., Soucy, P., Rubinstein, W. S., Arason, A., Rennert, G., Martin, N. G., Montgomery, G. W., Chang-Claude, J., Flesch-Janys, D., Brauch, H., Severi, G., Baglietto, L., Cox, A., Cross, S. S., Miron, P., Gerty, S. M., Tapper, W., Yannoukakos, D., Fountzilas, G., Fasching, P. A., Beckmann, M. W., dos, S. S., I, Peto, J., Lambrechts, D., Paridaens, R., Rudiger, T., Forsti, A., Winqvist, R., Pylkas, K., Diasio, R. B., Lee, A. M., Eckel-Passow, J., Vachon, C., Blows, F., Driver, K., Dunning, A., Pharoah, P. P., Offit, K., Pankratz, V. S., Hakonarson, H., Chenevix-Trench, G., Easton, D. F. and Couch, F. J. (2010b). **A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population**. Nat. Genet *42*, 885-892, doi: ng.669 [pii];10.1038/ng.669 [doi].

Awadelkarim, K. D., Arizzi, C., Elamin, E. O., Hamad, H. M., De Blasio, P., Mekki, S. O., Osman, I., Biunno, I., Elwali, N. E., Mariani-Costantini, R. and Barberis, M. C. (2008). **Pathological, clinical and prognostic characteristics of breast cancer in Central Sudan versus Northern Italy: implications for breast cancer in Africa**. Histopathology *52*, 445-456, doi: 10.1111/j.1365-2559.2008.02966.x.

Banegas, M. P., John, E. M., Slattery, M. L., Gomez, S. L., Yu, M., LaCroix, A. Z., Pee, D., Chlebowski, R. T., Hines, L. M., Thompson, C. A. and Gail, M. H. (2017). **Projecting Individualized Absolute Invasive Breast Cancer Risk in US Hispanic Women**. J Natl Cancer Inst *109*, doi: 10.1093/jnci/djw215.

Barnholtz-Sloan, J. S., Shetty, P. B., Guan, X., Nyante, S. J., Luo, J., Brennan, D. J. and Millikan, R. C. (2010). **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, doi: bgq128 [pii];10.1093/carcin/bgq128 [doi].

Baselga, J. and Norton, L. (2002). **Focus on breast cancer**. Cancer Cell *1*, 319-322.

Beggs, A. D. and Hodgson, S. V. (2009). **Genomics and breast cancer: the different levels of inherited susceptibility**. Eur J Hum Genet *17*, 855-856, doi: 10.1038/ejhg.2008.235.

Bennett, D. A. (2001). **How can I deal with missing data in my study?** Aust N Z J Public Health *25*, 464-469.

Bodner, T. E. (2008). **What Improves with Increased Missing Data Imputations?** Structural Equation Modeling *15*, 651–675, doi: 10.1080/10705510802339072.

Bojesen, S. E., Pooley, K. A., Johnatty, S. E., Beesley, J., Michailidou, K., Tyrer, J. P., Edwards, S. L., Pickett, H. A., Shen, H. C., Smart, C. E., Hillman, K. M., Mai, P. L., Lawrenson, K., Stutz, M. D., Lu, Y., Karevan, R., Woods, N., Johnston, R. L., French, J. D., Chen, X., Weischer, M., Nielsen, S. F., Maranian, M. J., Ghoussaini, M., Ahmed, S., Baynes, C., Bolla, M. K., Wang, Q., Dennis, J., McGuffog, L., Barrowdale, D., Lee, A., Healey, S., Lush, M., Tessier, D. C., Vincent, D., Bacot,

F., Australian Cancer, S., Australian Ovarian Cancer, S., Kathleen Cuningham Foundation Consortium for Research into Familial Breast, C., Gene Environment, I., Breast, C., Swedish Breast Cancer, S., Hereditary, B., Ovarian Cancer Research Group, N., Epidemiological study of, B., Carriers, B. M., Genetic Modifiers of Cancer Risk in, B. M. C., Vergote, I., Lambrechts, S., Despierre, E., Risch, H. A., Gonzalez-Neira, A., Rossing, M. A., Pita, G., Doherty, J. A., Alvarez, N., Larson, M. C., Fridley, B. L., Schoof, N., Chang-Claude, J., Cicek, M. S., Peto, J., Kalli, K. R., Broeks, A., Armasu, S. M., Schmidt, M. K., Braaf, L. M., Winterhoff, B., Nevanlinna, H., Konecny, G. E., Lambrechts, D., Rogmann, L., Guenel, P., Teoman, A., Milne, R. L., Garcia, J. J., Cox, A., Shridhar, V., Burwinkel, B., Marme, F., Hein, R., Sawyer, E. J., Haiman, C. A., Wang-Gohrke, S., Andrulis, I. L., Moysich, K. B., Hopper, J. L., Odunsi, K., Lindblom, A., Giles, G. G., Brenner, H., Simard, J., Lurie, G., Fasching, P. A., Carney, M. E., Radice, P., Wilkens, L. R., Swerdlow, A., Goodman, M. T., Brauch, H., Garcia-Closas, M., Hillemanns, P., Winqvist, R., Durst, M., Devilee, P., Runnebaum, I., Jakubowska, A., Lubinski, J., Mannermaa, A., Butzow, R., Bogdanova, N. V., Dork, T., Pelttari, L. M., Zheng, W., Leminen, A., Anton-Culver, H., Bunker, C. H., Kristensen, V., Ness, R. B., Muir, K., Edwards, R., Meindl, A., Heitz, F., Matsuo, K., du Bois, A., Wu, A. H., Harter, P., Teo, S. H., Schwaab, I., Shu, X. O., Blot, W., Hosono, S., Kang, D., Nakanishi, T., Hartman, M., Yatabe, Y., Hamann, U., Karlan, B. Y., Sangrajrang, S., Kjaer, S. K., Gaborieau, V., Jensen, A., Eccles, D., Hogdall, E., Shen, C. Y., Brown, J., Woo, Y. L., Shah, M., Azmi, M. A., Luben, R., Omar, S. Z., Czene, K., Vierkant, R. A., Nordestgaard, B. G., Flyger, H., Vachon, C., Olson, J. E., Wang, X., Levine, D. A., Rudolph, A., Weber, R. P., Flesch-Janys, D., Iversen, E., Nickels, S., Schildkraut, J. M., Silva Idos, S., Cramer, D. W., Gibson, L., Terry, K. L., Fletcher, O., Vitonis, A. F., van der Schoot, C. E., Poole, E. M., Hogervorst, F. B., Tworoger, S. S., Liu, J., Bandera, E. V., Li, J., Olson, S. H., Humphreys, K., Orlow, I., Blomqvist, C., Rodriguez-Rodriguez, L., Aittomaki, K., Salvesen, H. B., Muranen, T. A., Wik, E., Brouwers, B., Krakstad, C., Wauters, E., Halle, M. K., Wildiers, H., Kiemeney, L. A., Mulot, C., Aben, K. K., Laurent-Puig, P., Altena, A. M., Truong, T., Massuger, L. F., Benitez, J., Pejovic, T., Perez, J. I., Hoatlin, M., Zamora, M. P., Cook, L. S., Balasubramanian, S. P., Kelemen, L. E., Schneeweiss, A., Le, N. D., Sohn, C., Brooks-Wilson, A., Tomlinson, I., Kerin, M. J., Miller, N., Cybulski, C., Henderson, B. E., Menkiszak, J., Schumacher, F., Wentzensen, N., Le Marchand, L., Yang, H. P., Mulligan, A. M., Glendon, G., Engelholm, S. A., Knight, J. A., Hogdall, C. K., Apicella, C., Gore, M., Tsimiklis, H., Song, H., Southey, M. C., Jager, A., den Ouweland, A. M., Brown, R., Martens, J. W., Flanagan, J. M., Kriege, M., Paul, J., Margolin, S., Siddiqui, N., Severi, G., Whittemore, A. S., Baglietto, L., McGuire, V., Stegmaier, C., Sieh, W., Muller, H., Arndt, V., Labreche, F., Gao, Y. T., Goldberg, M. S., Yang, G., Dumont, M., McLaughlin, J. R., Hartmann, A., Ekici, A. B., Beckmann, M. W., Phelan, C. M., Lux, M. P., Permuth-Wey, J., Peissel, B., Sellers, T. A., Ficarazzi, F., Barile, M., Ziogas, A., Ashworth, A., Gentry-Maharaj, A., Jones, M., Ramus, S. J., Orr, N., Menon, U., Pearce, C. L., Bruning, T., Pike, M. C., Ko, Y. D., Lissowska, J., Figueroa, J., Kupryjanczyk, J., Chanock, S. J., Dansonka-Mieszkowska, A., Jukkola-Vuorinen, A., Rzepecka, I. K., Pylkas, K., Bidzinski, M., Kauppila, S., Hollestelle, A., Seynaeve, C., Tollenaar, R. A., Durda, K., Jaworska, K., Hartikainen, J. M., Kosma, V. M., Kataja, V., Antonenkova, N. N., Long, J., Shrubsole, M., Deming-Halverson, S., Lophatananon, A., Siriwanarangsan, P., Stewart-Brown, S., Ditsch, N., Lichtner, P., Schmutzler, R. K., Ito, H., Iwata, H., Tajima, K., Tseng, C. C., Stram, D. O., van den Berg, D., Yip, C. H., Ikram, M. K., Teh, Y. C., Cai, H., Lu, W., Signorello, L. B., Cai, Q., Noh, D. Y., Yoo, K. Y., Miao, H., Iau, P. T., Teo, Y. Y., McKay, J., Shapiro, C., Ademuyiwa, F., Fountzilas, G., Hsiung, C. N., Yu, J. C., Hou, M. F., Healey, C. S., Luccarini, C., Peock, S., Stoppa-Lyonnet, D., Peterlongo, P., Rebbeck, T. R., Piedmonte, M., Singer, C. F., Friedman, E., Thomassen, M., Offit, K., Hansen, T. V., Neuhausen, S. L., Szabo, C. I., Blanco, I., Garber, J., Narod, S. A., Weitzel, J. N., Montagna, M., Olah, E., Godwin, A. K., Yannoukakos, D., Goldgar, D. E., Caldes, T., Imyanitov, E. N., Tihomirova, L., Arun, B. K., Campbell, I., Mensenkamp, A. R., van Asperen, C. J., van Roozendaal, K. E., Meijers-Heijboer, H., Collee, J. M., Oosterwijk, J. C., Hooning, M. J., Rookus, M. A., van der Luijt, R. B., Os, T. A., Evans, D. G., Frost, D., Fineberg, E., Barwell, J., Walker, L., Kennedy, M. J., Platte, R.,

Davidson, R., Ellis, S. D., Cole, T., Bressac-de Paillerets, B., Buecher, B., Damiola, F., Faivre, L., Frenay, M., Sinilnikova, O. M., Caron, O., Giraud, S., Mazoyer, S., Bonadona, V., Caux-Moncoutier, V., Toloczko-Grabarek, A., Gronwald, J., Byrski, T., Spurdle, A. B., Bonanni, B., Zaffaroni, D., Giannini, G., Bernard, L., Dolcetti, R., Manoukian, S., Arnold, N., Engel, C., Deissler, H., Rhiem, K., Niederacher, D., Plendl, H., Sutter, C., Wappenschmidt, B., Borg, A., Melin, B., Rantala, J., Soller, M., Nathanson, K. L., Domchek, S. M., Rodriguez, G. C., Salani, R., Kaulich, D. G., Tea, M. K., Paluch, S. S., Laitman, Y., Skytte, A. B., Kruse, T. A., Jensen, U. B., Robson, M., Gerdes, A. M., Ejlertsen, B., Foretova, L., Savage, S. A., Lester, J., Soucy, P., Kuchenbaecker, K. B., Olswold, C., Cunningham, J. M., Slager, S., Pankratz, V. S., Dicks, E., Lakhani, S. R., Couch, F. J., Hall, P., Monteiro, A. N., Gayther, S. A., Pharoah, P. D., Reddel, R. R., Goode, E. L., Greene, M. H., Easton, D. F., Berchuck, A., Antoniou, A. C., Chenevix-Trench, G. and Dunning, A. M. (2013a). **Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer**. Nat Genet *45*, 371-384, 384e371-372, doi: 10.1038/ng.2566.

Bojesen, S. E., Pooley, K. A., Johnatty, S. E., Beesley, J., Michailidou, K., Tyrer, J. P., Edwards, S. L., Pickett, H. A., Shen, H. C., Smart, C. E., Hillman, K. M., Mai, P. L., Lawrenson, K., Stutz, M. D., Lu, Y., Karevan, R., Woods, N., Johnston, R. L., French, J. D., Chen, X., Weischer, M., Nielsen, S. F., Maranian, M. J., Ghoussaini, M., Ahmed, S., Baynes, C., Bolla, M. K., Wang, Q., Dennis, J., McGuffog, L., Barrowdale, D., Lee, A., Healey, S., Lush, M., Tessier, D. C., Vincent, D., Bacot, F., Vergote, I., Lambrechts, S., Despierre, E., Risch, H. A., Gonzalez-Neira, A., Rossing, M. A., Pita, G., Doherty, J. A., Alvarez, N., Larson, M. C., Fridley, B. L., Schoof, N., Chang-Claude, J., Cicek, M. S., Peto, J., Kalli, K. R., Broeks, A., Armasu, S. M., Schmidt, M. K., Braaf, L. M., Winterhoff, B., Nevanlinna, H., Konecny, G. E., Lambrechts, D., Rogmann, L., Guenel, P., Teoman, A., Milne, R. L., Garcia, J. J., Cox, A., Shridhar, V., Burwinkel, B., Marme, F., Hein, R., Sawyer, E. J., Haiman, C. A., Wang-Gohrke, S., Andrulis, I. L., Moysich, K. B., Hopper, J. L., Odunsi, K., Lindblom, A., Giles, G. G., Brenner, H., Simard, J., Lurie, G., Fasching, P. A., Carney, M. E., Radice, P., Wilkens, L. R., Swerdlow, A., Goodman, M. T., Brauch, H., Garcia-Closas, M., Hillemanns, P., Winqvist, R., Durst, M., Devilee, P., Runnebaum, I., Jakubowska, A., Lubinski, J., Mannermaa, A., Butzow, R., Bogdanova, N. V., Dork, T., Pelttari, L. M., Zheng, W., Leminen, A., Anton-Culver, H., Bunker, C. H., Kristensen, V., Ness, R. B., Muir, K., Edwards, R., Meindl, A., Heitz, F., Matsuo, K., du, B. A., Wu, A. H., Harter, P., Teo, S. H., Schwaab, I., Shu, X. O., Blot, W., Hosono, S., Kang, D., Nakanishi, T., Hartman, M., Yatabe, Y., Hamann, U., Karlan, B. Y., Sangrajrang, S., Kjaer, S. K., Gaborieau, V., Jensen, A., Eccles, D., Hogdall, E., Shen, C. Y., Brown, J., Woo, Y. L., Shah, M., Azmi, M. A., Luben, R., Omar, S. Z., Czene, K., Vierkant, R. A., Nordestgaard, B. G., Flyger, H., Vachon, C., Olson, J. E., Wang, X., Levine, D. A., Rudolph, A., Weber, R. P., Flesch-Janys, D., Iversen, E., Nickels, S., Schildkraut, J. M., Silva, I. S., Cramer, D. W., Gibson, L., Terry, K. L., Fletcher, O., Vitonis, A. F., van der Schoot, C. E., Poole, E. M., Hogervorst, F. B., Tworoger, S. S., Liu, J., Bandera, E. V., Li, J., Olson, S. H., Humphreys, K., Orlow, I., Blomqvist, C., Rodriguez-Rodriguez, L., Aittomaki, K., Salvesen, H. B., Muranen, T. A., Wik, E., Brouwers, B., Krakstad, C., Wauters, E., Halle, M. K., Wildiers, H., Kiemeney, L. A., Mulot, C., Aben, K. K., Laurent-Puig, P., Altena, A. M., Truong, T., Massuger, L. F., Benitez, J., Pejovic, T., Perez, J. I., Hoatlin, M., Zamora, M. P., Cook, L. S., Balasubramanian, S. P., Kelemen, L. E., Schneeweiss, A., Le, N. D., Sohn, C., Brooks-Wilson, A., Tomlinson, I., Kerin, M. J., Miller, N., Cybulski, C., Henderson, B. E., Menkiszak, J., Schumacher, F., Wentzensen, N., Le, M. L., Yang, H. P., Mulligan, A. M., Glendon, G., Engelholm, S. A., Knight, J. A., Hogdall, C. K., Apicella, C., Gore, M., Tsimiklis, H., Song, H., Southey, M. C., Jager, A., den Ouweland, A. M., Brown, R., Martens, J. W., Flanagan, J. M., Kriege, M., Paul, J., Margolin, S., Siddiqui, N., Severi, G., Whittemore, A. S., Baglietto, L., McGuire, V., Stegmaier, C., Sieh, W., Muller, H., Arndt, V., Labreche, F., Gao, Y. T., Goldberg, M. S., Yang, G., Dumont, M., McLaughlin, J. R., Hartmann, A., Ekici, A. B., Beckmann, M. W., Phelan, C. M., Lux, M. P., Permuth-Wey, J., Peissel, B., Sellers, T. A., Ficarazzi, F., Barile, M. and Ziogas, A. (2013b). **Multiple independent**

variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat. Genet *45*, 371-384, doi: ng.2566 [pii];10.1038/ng.2566 [doi].

Bottorff, J. L., McKeown, S. B., Carey, J., Haines, R., Okoli, C., Johnson, K. C., Easley, J., Ferrence, R., Baillie, L. and Ptolemy, E. (2010). **Young women's responses to smoking and breast cancer risk information**. Health Educ Res *25*, 668-677, doi: 10.1093/her/cyp067.

Boyle, P. and Ferlay, J. (2005). **Cancer incidence and mortality in Europe, 2004**. Ann. Oncol *16*, 481-488, doi: mdi098 [pii];10.1093/annonc/mdi098 [doi].

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. and Jemal, A. (2018). **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. CA Cancer J Clin *68*, 394-424, doi: 10.3322/caac.21492.

Campeau, P. M., Foulkes, W. D. and Tischkowitz, M. D. (2008). **Hereditary breast cancer: new genetic developments, new therapeutic avenues**. Hum Genet *124*, 31-42, doi: 10.1007/s00439-008-0529-1.

Cavalli-Sforza, L. L. (2005). **The Human Genome Diversity Project: past, present and future**. Nat Rev Genet *6*, 333-340, doi: 10.1038/nrg1596.

Choi, E., Lee, S., Nhung, B. C., Suh, M., Park, B., Jun, J. K. and Choi, K. S. (2017). **Cancer mortality-to-incidence ratio as an indicator of cancer management outcomes in Organization for Economic Cooperation and Development countries**. Epidemiol Health *39*, e2017006, doi: 10.4178/epih.e2017006.

Collaborative Group on Hormonal Factors in Breast, C. (1996). **Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies**. Lancet *347*, 1713-1727.

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

Collaborative Group on Hormonal Factors in Breast, C. (2002). **Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease**. Lancet *360*, 187-195, doi: 10.1016/S0140-6736(02)09454-0.

Collaborative Group on Hormonal Factors in Breast, C. (2012). **Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies**. Lancet Oncol *13*, 1141-1151, doi: 10.1016/S1470-2045(12)70425-4.

Connor, A. E., Baumgartner, K. B., Baumgartner, R. N., Pinkston, C. M., Boone, S. D., John, E. M., Torres-Mejia, G., Hines, L. M., Giuliano, A. R., Wolff, R. K. and Slattery, M. L. (2016). **Cigarette Smoking and Breast Cancer Risk in Hispanic and Non-Hispanic White Women: The Breast Cancer Health Disparities Study**. J Womens Health (Larchmt) *25*, 299-310, doi: 10.1089/jwh.2015.5502.

Cuello-Lopez, J., Fidalgo-Zapata, A. and Vasquez-Trespalacios, E. (2017). **Obesity and Prognostic Variables in Colombian Breast Cancer Patients: A Cross-Sectional Study**. Int J Breast Cancer *2017*, 9574874, doi: 10.1155/2017/9574874.

Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J. and Shi, B. (2015). **Breast cancer intrinsic subtype classification, clinical use and future trends**. Am J Cancer Res *5*, 2929-2943.

Dieci, M. V., Orvieto, E., Dominici, M., Conte, P. and Guarneri, V. (2014). **Rare breast cancer subtypes: histological, molecular, and clinical peculiarities**. Oncologist *19*, 805-813, doi: 10.1634/theoncologist.2014-0108.

Dong, Y. and Peng, C. Y. (2013). **Principled missing data methods for researchers**. Springerplus *2*, 222, doi: 10.1186/2193-1801-2-222.

Dossus, L. and Benusiglio, P. R. (2015). **Lobular breast cancer: incidence and genetic and non-genetic risk factors**. Breast Cancer Res *17*, 37, doi: 10.1186/s13058-015-0546-7.

Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D., Thompson, D., Ballinger, D. G., Struewing, J. P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C. S., Bowman, R., collaborators, S., Meyer, K. B., Haiman, C. A., Kolonel, L. K., Henderson, B. E., Le Marchand, L., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C. Y., Wu, P. E., Wang, H. C., Eccles, D., Evans, D. G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M. R., Rahman, N., Chenevix-Trench, G., Bojesen, S. E., Nordestgaard, B. G., Axelsson, C. K., Garcia-Closas, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska, B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K. Y., Noh, D. Y., Ahn, S. H., Hunter, D. J., Hankinson, S. E., Cox, D. G., Hall, P., Wedren, S., Liu, J., Low, Y. L., Bogdanova, N., Schurmann, P., Dork, T., Tollenaar, R. A., Jacobi, C. E., Devilee, P., Klijn, J. G., Sigurdson, A. J., Doody, M. M., Alexander, B. H., Zhang, J., Cox, A., Brock, I. W., MacPherson, G., Reed, M. W., Couch, F. J., Goode, E. L., Olson, J. E., Meijers-Heijboer, H., van den Ouweland, A., Uitterlinden, A., Rivadeneira, F., Milne, R. L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J. L., McCredie, M., Southey, M., Giles, G. G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y. D., Spurdle, A. B., Beesley, J., Chen, X., kConFab, Group, A. M., Mannermaa, A., Kosma, V. M., Kataja, V., Hartikainen, J., Day, N. E., Cox, D. R. and Ponder, B. A. (2007a). **Genome-wide association study identifies novel breast cancer susceptibility loci**. Nature *447*, 1087-1093, doi: 10.1038/nature05887.

Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D., Thompson, D., Ballinger, D. G., Struewing, J. P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C. S., Bowman, R., Meyer, K. B., Haiman, C. A., Kolonel, L. K., Henderson, B. E., Le, M. L., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C. Y., Wu, P. E., Wang, H. C., Eccles, D., Evans, D. G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M. R., Rahman, N.,

Chenevix-Trench, G., Bojesen, S. E., Nordestgaard, B. G., Axelsson, C. K., Garcia-Closas, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska, B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K. Y., Noh, D. Y., Ahn, S. H., Hunter, D. J., Hankinson, S. E., Cox, D. G., Hall, P., Wedren, S., Liu, J., Low, Y. L., Bogdanova, N., Schurmann, P., Dork, T., Tollenaar, R. A., Jacobi, C. E., Devilee, P., Klijn, J. G., Sigurdson, A. J., Doody, M. M., Alexander, B. H., Zhang, J., Cox, A., Brock, I. W., MacPherson, G., Reed, M. W., Couch, F. J., Goode, E. L., Olson, J. E., Meijers-Heijboer, H., van den Ouweland, A., Uitterlinden, A., Rivadeneira, F., Milne, R. L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J. L., McCredie, M., Southey, M., Giles, G. G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y. D., Spurdle, A. B., Beesley, J., Chen, X., Mannermaa, A., Kosma, V. M., Kataja, V., Hartikainen, J., Day, N. E., Cox, D. R. and Ponder, B. A. (2007b). **Genome-wide association study identifies novel breast cancer susceptibility loci**. Nature *447*, 1087-1093, doi: nature05887 [pii];10.1038/nature05887 [doi].

Eichholzer, M., Schmid, S. M., Bovey, F., Jordan, P., Rohrmann, S., Huang, D. J., Rochlitz, C. and Guth, U. (2012). **Impact of overweight and obesity on postmenopausal breast cancer: analysis of 20-year data from Switzerland**. Arch Gynecol Obstet *285*, 797-803, doi: 10.1007/s00404-011-2022-7.

Elgazzar, S., Zembutsu, H., Takahashi, A., Kubo, M., Aki, F., Hirata, K., Takatsuka, Y., Okazaki, M., Ohsumi, S., Yamakawa, T., Sasa, M., Katagiri, T., Miki, Y. and Nakamura, Y. (2012). **A genome-wide association study identifies a genetic variant in the SIAH2 locus associated with hormonal receptor-positive breast cancer in Japanese**. J. Hum. Genet *57*, 766-771, doi: jhg2012108 [pii];10.1038/jhg.2012.108 [doi].

Elidrissi Errahhali, M., Elidrissi Errahhali, M., Ouarzane, M., El Harroudi, T., Afqir, S. and Bellaoui, M. (2017). **First report on molecular breast cancer subtypes and their clinico-pathological characteristics in Eastern Morocco: series of 2260 cases**. BMC Womens Health *17*, 3, doi: 10.1186/s12905-016-0361-z.

Fejerman, L., Ahmadiyeh, N., Hu, D., Huntsman, S., Beckman, K. B., Caswell, J. L., Tsung, K., John, E. M., Torres-Mejia, G., Carvajal-Carmona, L., Echeverry, M. M., Tuazon, A. M., Ramirez, C., Gignoux, C. R., Eng, C., Gonzalez-Burchard, E., Henderson, B., Le, M. L., Kooperberg, C., Hou, L., Agalliu, I., Kraft, P., Lindstrom, S., Perez-Stable, E. J., Haiman, C. A. and Ziv, E. (2014). **Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25**. Nat. Commun *5*, 5260, doi: ncomms6260 [pii];10.1038/ncomms6260 [doi].

Fejerman, L., Chen, G. K., Eng, C., Huntsman, S., Hu, D., Williams, A., Pasaniuc, B., John, E. M., Via, M., Gignoux, C., Ingles, S., Monroe, K. R., Kolonel, L. N., Torres-Mejia, G., Perez-Stable, E. J., Burchard, E. G., Henderson, B. E., Haiman, C. A. and Ziv, E. (2012). **Admixture mapping identifies a locus on 6q25 associated with breast cancer risk in US Latinas**. Hum. Mol. Genet *21*, 1907-1917, doi: ddr617 [pii];10.1093/hmg/ddr617 [doi].

Fejerman, L., John, E. M., Huntsman, S., Beckman, K., Choudhry, S., Perez-Stable, E., Burchard, E. G. and Ziv, E. (2008). **Genetic ancestry and risk of breast cancer among U.S. Latinas**. Cancer Res *68*, 9723-9728, doi: 68/23/9723 [pii];10.1158/0008-5472.CAN-08-2039 [doi].

Fejerman, L., Romieu, I., John, E. M., Lazcano-Ponce, E., Huntsman, S., Beckman, K. B., Perez-Stable, E. J., Gonzalez, B. E., Ziv, E. and Torres-Mejia, G. (2010). **European ancestry is positively associated with breast cancer risk in Mexican women**. Cancer Epidemiol. Biomarkers Prev *19*, 1074-1082, doi: 1055-9965.EPI-09-1193 [pii];10.1158/1055-9965.EPI-09-1193 [doi].

Fejerman, L., Stern, M. C., Ziv, E., John, E. M., Torres-Mejia, G., Hines, L. M., Wolff, R., Wang, W., Baumgartner, K. B., Giuliano, A. R. and Slattery, M. L. (2013). **Genetic ancestry modifies the association between genetic risk variants and breast cancer risk among Hispanic and non-Hispanic white women**. Carcinogenesis *34*, 1787-1793, doi: bgt110 [pii];10.1093/carcin/bgt110 [doi].

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D. and Bray, F. (2015). **Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012**. Int J Cancer *136*, E359-386, doi: 10.1002/ijc.29210.

Fletcher, O., Johnson, N., Orr, N., Hosking, F. J., Gibson, L. J., Walker, K., Zelenika, D., Gut, I., Heath, S., Palles, C., Coupland, B., Broderick, P., Schoemaker, M., Jones, M., Williamson, J., Chilcott-Burns, S., Tomczyk, K., Simpson, G., Jacobs, K. B., Chanock, S. J., Hunter, D. J., Tomlinson, I. P., Swerdlow, A., Ashworth, A., Ross, G., dos, S. S., I, Lathrop, M., Houlston, R. S. and Peto, J. (2011). **Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study**. J. Natl. Cancer Inst *103*, 425-435, doi: djq563 [pii];10.1093/jnci/djq563 [doi].

Forouzanfar, M. H., Foreman, K. J., Delossantos, A. M., Lozano, R., Lopez, A. D., Murray, C. J. and Naghavi, M. (2011). **Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis**. Lancet *378*, 1461-1484, doi: 10.1016/S0140-6736(11)61351-2.

Franco-Marina, F., Lopez-Carrillo, L., Keating, N. L., Arreola-Ornelas, H. and Marie Knaul, F. (2015). **Breast cancer age at diagnosis patterns in four Latin American Populations: A comparison with North American countries**. Cancer Epidemiol *39*, 831-837, doi: 10.1016/j.canep.2015.10.004.

French, J. D., Ghoussaini, M., Edwards, S. L., Meyer, K. B., Michailidou, K., Ahmed, S., Khan, S., Maranian, M. J., O'Reilly, M., Hillman, K. M., Betts, J. A., Carroll, T., Bailey, P. J., Dicks, E., Beesley, J., Tyrer, J., Maia, A. T., Beck, A., Knoblauch, N. W., Chen, C., Kraft, P., Barnes, D., Gonzalez-Neira, A., Alonso, M. R., Herrero, D., Tessier, D. C., Vincent, D., Bacot, F., Luccarini, C., Baynes, C., Conroy, D., Dennis, J., Bolla, M. K., Wang, Q., Hopper, J. L., Southey, M. C., Schmidt, M. K., Broeks, A., Verhoef, S., Cornelissen, S., Muir, K., Lophatananon, A., Stewart-Brown, S., Siriwanarangsan, P., Fasching, P. A., Loehberg, C. R., Ekici, A. B., Beckmann, M. W., Peto, J., dos Santos Silva, I., Johnson, N., Aitken, Z., Sawyer, E. J., Tomlinson, I., Kerin, M. J., Miller, N., Marme, F., Schneeweiss, A., Sohn, C., Burwinkel, B., Guenel, P., Truong, T., Laurent-Puig, P., Menegaux, F., Bojesen, S. E., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Milne, R. L., Zamora, M. P., Arias Perez, J. I., Benitez, J., Anton-Culver, H., Brenner, H., Muller, H., Arndt, V., Stegmaier, C., Meindl, A., Lichtner, P., Schmutzler, R. K., Engel, C., Brauch, H., Hamann, U., Justenhoven, C., Network, G., Aaltonen, K., Heikkila, P., Aittomaki, K., Blomqvist, C., Matsuo, K., Ito, H., Iwata, H., Sueta, A., Bogdanova, N. V., Antonenkova, N. N., Dork, T., Lindblom, A., Margolin, S., Mannermaa, A., Kataja, V., Kosma, V. M., Hartikainen, J. M., kConFab, I., Wu, A. H., Tseng, C. C., Van Den Berg, D., Stram, D. O., Lambrechts, D., Peeters,

S., Smeets, A., Floris, G., Chang-Claude, J., Rudolph, A., Nickels, S., Flesch-Janys, D., Radice, P., Peterlongo, P., Bonanni, B., Sardella, D., Couch, F. J., Wang, X., Pankratz, V. S., Lee, A., Giles, G. G., Severi, G., Baglietto, L., Haiman, C. A., Henderson, B. E., Schumacher, F., Le Marchand, L., Simard, J., Goldberg, M. S., Labreche, F., Dumont, M., Teo, S. H., Yip, C. H., Ng, C. H., Vithana, E. N., Kristensen, V., Zheng, W., Deming-Halverson, S., Shrubsole, M., Long, J., Winqvist, R., Pylkas, K., Jukkola-Vuorinen, A., Grip, M., Andrulis, I. L., Knight, J. A., Glendon, G., Mulligan, A. M., Devilee, P., Seynaeve, C., Garcia-Closas, M., Figueroa, J., Chanock, S. J., Lissowska, J., Czene, K., Klevebring, D., Schoof, N., Hooning, M. J., Martens, J. W., Collee, J. M., Tilanus-Linthorst, M., Hall, P., Li, J., Liu, J., Humphreys, K., Shu, X. O., Lu, W., Gao, Y. T., Cai, H., Cox, A., Balasubramanian, S. P., Blot, W., Signorello, L. B., Cai, Q., Pharoah, P. D., Healey, C. S., Shah, M., Pooley, K. A., Kang, D., Yoo, K. Y., Noh, D. Y., Hartman, M., Miao, H., Sng, J. H., Sim, X., Jakubowska, A., Lubinski, J., Jaworska-Bieniek, K., Durda, K., Sangrajrang, S., Gaborieau, V., McKay, J., Toland, A. E., Ambrosone, C. B., Yannoukakos, D., Godwin, A. K., Shen, C. Y., Hsiung, C. N., Wu, P. E., Chen, S. T., Swerdlow, A., Ashworth, A., Orr, N., Schoemaker, M. J., Ponder, B. A., Nevanlinna, H., Brown, M. A., Chenevix-Trench, G., Easton, D. F. and Dunning, A. M. (2013a). **Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers**. Am J Hum Genet *92*, 489-503, doi: 10.1016/j.ajhg.2013.01.002.

French, J. D., Ghoussaini, M., Edwards, S. L., Meyer, K. B., Michailidou, K., Ahmed, S., Khan, S., Maranian, M. J., O'Reilly, M., Hillman, K. M., Betts, J. A., Carroll, T., Bailey, P. J., Dicks, E., Beesley, J., Tyrer, J., Maia, A. T., Beck, A., Knoblauch, N. W., Chen, C., Kraft, P., Barnes, D., Gonzalez-Neira, A., Alonso, M. R., Herrero, D., Tessier, D. C., Vincent, D., Bacot, F., Luccarini, C., Baynes, C., Conroy, D., Dennis, J., Bolla, M. K., Wang, Q., Hopper, J. L., Southey, M. C., Schmidt, M. K., Broeks, A., Verhoef, S., Cornelissen, S., Muir, K., Lophatananon, A., Stewart-Brown, S., Siriwanarangsan, P., Fasching, P. A., Loehberg, C. R., Ekici, A. B., Beckmann, M. W., Peto, J., dos, S. S., I, Johnson, N., Aitken, Z., Sawyer, E. J., Tomlinson, I., Kerin, M. J., Miller, N., Marme, F., Schneeweiss, A., Sohn, C., Burwinkel, B., Guenel, P., Truong, T., Laurent-Puig, P., Menegaux, F., Bojesen, S. E., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Milne, R. L., Zamora, M. P., Arias Perez, J. I., Benitez, J., Anton-Culver, H., Brenner, H., Muller, H., Arndt, V., Stegmaier, C., Meindl, A., Lichtner, P., Schmutzler, R. K., Engel, C., Brauch, H., Hamann, U., Justenhoven, C., Aaltonen, K., Heikkila, P., Aittomaki, K., Blomqvist, C., Matsuo, K., Ito, H., Iwata, H., Sueta, A., Bogdanova, N. V., Antonenkova, N. N., Dork, T., Lindblom, A., Margolin, S., Mannermaa, A., Kataja, V., Kosma, V. M., Hartikainen, J. M., Wu, A. H., Tseng, C. C., Van Den Berg, D., Stram, D. O., Lambrechts, D., Peeters, S., Smeets, A., Floris, G., Chang-Claude, J., Rudolph, A., Nickels, S., Flesch-Janys, D., Radice, P., Peterlongo, P., Bonanni, B., Sardella, D., Couch, F. J., Wang, X., Pankratz, V. S., Lee, A., Giles, G. G., Severi, G., Baglietto, L., Haiman, C. A., Henderson, B. E., Schumacher, F., Le, M. L., Simard, J., Goldberg, M. S., Labreche, F., Dumont, M., Teo, S. H., Yip, C. H., Ng, C. H., Vithana, E. N., Kristensen, V., Zheng, W., Deming-Halverson, S., Shrubsole, M., Long, J., Winqvist, R., Pylkas, K., Jukkola-Vuorinen, A., Grip, M., Andrulis, I. L., Knight, J. A., Glendon, G., Mulligan, A. M., Devilee, P., Seynaeve, C., Garcia-Closas, M., Figueroa, J., Chanock, S. J., Lissowska, J., Czene, K., Klevebring, D., Schoof, N., Hooning, M. J., Martens, J. W., Collee, J. M., Tilanus-Linthorst, M., Hall, P., Li, J., Liu, J., Humphreys, K., Shu, X. O., Lu, W., Gao, Y. T., Cai, H., Cox, A., Balasubramanian, S. P., Blot, W., Signorello, L. B., Cai, Q., Pharoah, P. D., Healey, C. S., Shah, M., Pooley, K. A., Kang, D., Yoo, K. Y., Noh, D. Y., Hartman, M., Miao, H., Sng, J. H., Sim, X., Jakubowska, A., Lubinski, J., Jaworska-Bieniek, K., Durda, K., Sangrajrang, S., Gaborieau, V., McKay, J., Toland, A. E., Ambrosone, C. B., Yannoukakos, D., Godwin, A. K., Shen, C. Y., Hsiung, C. N., Wu, P. E., Chen, S. T., Swerdlow, A., Ashworth, A., Orr, N., Schoemaker, M. J., Ponder, B. A., Nevanlinna, H., Brown, M. A., Chenevix-Trench, G., Easton, D. F. and Dunning, A. M. (2013b). **Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through**

long-range enhancers. Am. J. Hum. Genet *92*, 489-503, doi: S0002-9297(13)00031-1 [pii];10.1016/j.ajhg.2013.01.002 [doi].

Garcia-Closas, M. and Chanock, S. (2008). **Genetic susceptibility loci for breast cancer by estrogen receptor status**. Clin Cancer Res *14*, 8000-8009, doi: 10.1158/1078-0432.CCR-08-0975.

Garcia-Closas, M., Couch, F. J., Lindstrom, S., Michailidou, K., Schmidt, M. K., Brook, M. N., Orr, N., Rhie, S. K., Riboli, E., Feigelson, H. S., Le Marchand, L., Buring, J. E., Eccles, D., Miron, P., Fasching, P. A., Brauch, H., Chang-Claude, J., Carpenter, J., Godwin, A. K., Nevanlinna, H., Giles, G. G., Cox, A., Hopper, J. L., Bolla, M. K., Wang, Q., Dennis, J., Dicks, E., Howat, W. J., Schoof, N., Bojesen, S. E., Lambrechts, D., Broeks, A., Andrulis, I. L., Guenel, P., Burwinkel, B., Sawyer, E. J., Hollestelle, A., Fletcher, O., Winqvist, R., Brenner, H., Mannermaa, A., Hamann, U., Meindl, A., Lindblom, A., Zheng, W., Devillee, P., Goldberg, M. S., Lubinski, J., Kristensen, V., Swerdlow, A., Anton-Culver, H., Dork, T., Muir, K., Matsuo, K., Wu, A. H., Radice, P., Teo, S. H., Shu, X. O., Blot, W., Kang, D., Hartman, M., Sangrajrang, S., Shen, C. Y., Southey, M. C., Park, D. J., Hammet, F., Stone, J., Veer, L. J., Rutgers, E. J., Lophatananon, A., Stewart-Brown, S., Siriwanarangsan, P., Peto, J., Schrauder, M. G., Ekici, A. B., Beckmann, M. W., Dos Santos Silva, I., Johnson, N., Warren, H., Tomlinson, I., Kerin, M. J., Miller, N., Marme, F., Schneeweiss, A., Sohn, C., Truong, T., Laurent-Puig, P., Kerbrat, P., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Milne, R. L., Perez, J. I., Menendez, P., Muller, H., Arndt, V., Stegmaier, C., Lichtner, P., Lochmann, M., Justenhoven, C., Ko, Y. D., Gene, E. I., breast, C. N., Muranen, T. A., Aittomaki, K., Blomqvist, C., Greco, D., Heikkinen, T., Ito, H., Iwata, H., Yatabe, Y., Antonenkova, N. N., Margolin, S., Kataja, V., Kosma, V. M., Hartikainen, J. M., Balleine, R., kConFab, I., Tseng, C. C., Berg, D. V., Stram, D. O., Neven, P., Dieudonne, A. S., Leunen, K., Rudolph, A., Nickels, S., Flesch-Janys, D., Peterlongo, P., Peissel, B., Bernard, L., Olson, J. E., Wang, X., Stevens, K., Severi, G., Baglietto, L., McLean, C., Coetzee, G. A., Feng, Y., Henderson, B. E., Schumacher, F., Bogdanova, N. V., Labreche, F., Dumont, M., Yip, C. H., Taib, N. A., Cheng, C. Y., Shrubsole, M., Long, J., Pylkas, K., Jukkola-Vuorinen, A., Kauppila, S., Knight, J. A., Glendon, G., Mulligan, A. M., Tollenaar, R. A., Seynaeve, C. M., Kriege, M., Hooning, M. J., van den Ouweland, A. M., van Deurzen, C. H., Lu, W., Gao, Y. T., Cai, H., Balasubramanian, S. P., Cross, S. S., Reed, M. W., Signorello, L., Cai, Q., Shah, M., Miao, H., Chan, C. W., Chia, K. S., Jakubowska, A., Jaworska, K., Durda, K., Hsiung, C. N., Wu, P. E., Yu, J. C., Ashworth, A., Jones, M., Tessier, D. C., Gonzalez-Neira, A., Pita, G., Alonso, M. R., Vincent, D., Bacot, F., Ambrosone, C. B., Bandera, E. V., John, E. M., Chen, G. K., Hu, J. J., Rodriguez-Gil, J. L., Bernstein, L., Press, M. F., Ziegler, R. G., Millikan, R. M., Deming-Halverson, S. L., Nyante, S., Ingles, S. A., Waisfisz, Q., Tsimiklis, H., Makalic, E., Schmidt, D., Bui, M., Gibson, L., Muller-Myhsok, B., Schmutzler, R. K., Hein, R., Dahmen, N., Beckmann, L., Aaltonen, K., Czene, K., Irwanto, A., Liu, J., Turnbull, C., Familial Breast Cancer, S., Rahman, N., Meijers-Heijboer, H., Uitterlinden, A. G., Rivadeneira, F., Australian Breast Cancer Tissue Bank, I., Olswold, C., Slager, S., Pilarski, R., Ademuyiwa, F., Konstantopoulou, I., Martin, N. G., Montgomery, G. W., Slamon, D. J., Rauh, C., Lux, M. P., Jud, S. M., Bruning, T., Weaver, J., Sharma, P., Pathak, H., Tapper, W., Gerty, S., Durcan, L., Trichopoulos, D., Tumino, R., Peeters, P. H., Kaaks, R., Campa, D., Canzian, F., Weiderpass, E., Johansson, M., Khaw, K. T., Travis, R., Clavel-Chapelon, F., Kolonel, L. N., Chen, C., Beck, A., Hankinson, S. E., Berg, C. D., Hoover, R. N., Lissowska, J., Figueroa, J. D., Chasman, D. I., Gaudet, M. M., Diver, W. R., Willett, W. C., Hunter, D. J., Simard, J., Benitez, J., Dunning, A. M., Sherman, M. E., Chenevix-Trench, G., Chanock, S. J., Hall, P., Pharoah, P. D., Vachon, C., Easton, D. F., Haiman, C. A. and Kraft, P. (2013a). **Genome-wide association studies identify four ER negative-specific breast cancer risk loci**. Nat Genet *45*, 392-398, doi: 10.1038/ng.2561.

Garcia-Closas, M., Couch, F. J., Lindstrom, S., Michailidou, K., Schmidt, M. K., Brook, M. N., Orr, N., Rhie, S. K., Riboli, E., Feigelson, H. S., Le, M. L., Buring, J. E., Eccles, D., Miron, P., Fasching, P. A., Brauch, H., Chang-Claude, J., Carpenter, J., Godwin, A. K., Nevanlinna, H., Giles, G. G., Cox, A., Hopper, J. L., Bolla, M. K., Wang, Q., Dennis, J., Dicks, E., Howat, W. J., Schoof, N., Bojesen, S. E., Lambrechts, D., Broeks, A., Andrulis, I. L., Guenel, P., Burwinkel, B., Sawyer, E. J., Hollestelle, A., Fletcher, O., Winqvist, R., Brenner, H., Mannermaa, A., Hamann, U., Meindl, A., Lindblom, A., Zheng, W., Devillee, P., Goldberg, M. S., Lubinski, J., Kristensen, V., Swerdlow, A., Anton-Culver, H., Dork, T., Muir, K., Matsuo, K., Wu, A. H., Radice, P., Teo, S. H., Shu, X. O., Blot, W., Kang, D., Hartman, M., Sangrajrang, S., Shen, C. Y., Southey, M. C., Park, D. J., Hammet, F., Stone, J., Veer, L. J., Rutgers, E. J., Lophatananon, A., Stewart-Brown, S., Siriwanarangsan, P., Peto, J., Schrauder, M. G., Ekici, A. B., Beckmann, M. W., dos, S. S., I, Johnson, N., Warren, H., Tomlinson, I., Kerin, M. J., Miller, N., Marme, F., Schneeweiss, A., Sohn, C., Truong, T., Laurent-Puig, P., Kerbrat, P., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Milne, R. L., Perez, J. I., Menendez, P., Muller, H., Arndt, V., Stegmaier, C., Lichtner, P., Lochmann, M., Justenhoven, C., Ko, Y. D., Muranen, T. A., Aittomaki, K., Blomqvist, C., Greco, D., Heikkinen, T., Ito, H., Iwata, H., Yatabe, Y., Antonenkova, N. N., Margolin, S., Kataja, V., Kosma, V. M., Hartikainen, J. M., Balleine, R., Tseng, C. C., Berg, D. V., Stram, D. O., Neven, P., Dieudonne, A. S., Leunen, K., Rudolph, A., Nickels, S., Flesch-Janys, D., Peterlongo, P., Peissel, B., Bernard, L., Olson, J. E., Wang, X., Stevens, K., Severi, G., Baglietto, L., McLean, C., Coetzee, G. A., Feng, Y., Henderson, B. E., Schumacher, F., Bogdanova, N. V., Labreche, F., Dumont, M., Yip, C. H., Taib, N. A., Cheng, C. Y., Shrubsole, M., Long, J., Pylkas, K., Jukkola-Vuorinen, A., Kauppila, S., Knight, J. A., Glendon, G., Mulligan, A. M., Tollenaar, R. A., Seynaeve, C. M., Kriege, M., Hooning, M. J., van den Ouweland, A. M., van Deurzen, C. H., Lu, W., Gao, Y. T., Cai, H., Balasubramanian, S. P., Cross, S. S., Reed, M. W., Signorello, L., Cai, Q., Shah, M., Miao, H., Chan, C. W., Chia, K. S., Jakubowska, A., Jaworska, K., Durda, K., Hsiung, C. N., Wu, P. E., Yu, J. C., Ashworth, A., Jones, M., Tessier, D. C., Gonzalez-Neira, A., Pita, G., Alonso, M. R., Vincent, D., Bacot, F., Ambrosone, C. B., Bandera, E. V., John, E. M., Chen, G. K., Hu, J. J., Rodriguez-Gil, J. L., Bernstein, L., Press, M. F., Ziegler, R. G., Millikan, R. M., Deming-Halverson, S. L., Nyante, S., Ingles, S. A., Waisfisz, Q., Tsimiklis, H., Makalic, E., Schmidt, D., Bui, M., Gibson, L., Muller-Myhsok, B., Schmutzler, R. K., Hein, R., Dahmen, N., Beckmann, L., Aaltonen, K., Czene, K., Irwanto, A., Liu, J., Turnbull, C., Rahman, N., Meijers-Heijboer, H., Uitterlinden, A. G., Rivadeneira, F., Olswold, C., Slager, S., Pilarski, R., Ademuyiwa, F., Konstantopoulou, I., Martin, N. G., Montgomery, G. W., Slamon, D. J., Rauh, C., Lux, M. P., Jud, S. M., Bruning, T., Weaver, J., Sharma, P., Pathak, H., Tapper, W., Gerty, S., Durcan, L., Trichopoulos, D., Tumino, R., Peeters, P. H., Kaaks, R., Campa, D., Canzian, F., Weiderpass, E., Johansson, M., Khaw, K. T., Travis, R., Clavel-Chapelon, F., Kolonel, L. N., Chen, C., Beck, A., Hankinson, S. E., Berg, C. D., Hoover, R. N., Lissowska, J., Figueroa, J. D., Chasman, D. I. and Gaudet, M. M. (2013b). **Genome-wide association studies identify four ER negative-specific breast cancer risk loci**. Nat. Genet *45*, 392-392, doi: ng.2561 [pii];10.1038/ng.2561 [doi].

Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A. and Abecasis, G. R. (2015). **A global reference for human genetic variation**. Nature *526*, 68-74, doi: 10.1038/nature15393.

Ghoncheh, M., Pournamdar, Z. and Salehiniya, H. (2016). **Incidence and Mortality and Epidemiology of Breast Cancer in the World**. Asian Pac J Cancer Prev *17*, 43-46.

Ghoussaini, M., Edwards, S. L., Michailidou, K., Nord, S., Cowper-Sal, L. R., Desai, K., Kar, S., Hillman, K. M., Kaufmann, S., Glubb, D. M., Beesley, J., Dennis, J., Bolla, M. K., Wang, Q., Dicks, E.,

Guo, Q., Schmidt, M. K., Shah, M., Luben, R., Brown, J., Czene, K., Darabi, H., Eriksson, M., Klevebring, D., Bojesen, S. E., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Lambrechts, D., Thienpont, B., Neven, P., Wildiers, H., Broeks, A., Van't Veer, L. J., Th Rutgers, E. J., Couch, F. J., Olson, J. E., Hallberg, E., Vachon, C., Chang-Claude, J., Rudolph, A., Seibold, P., Flesch-Janys, D., Peto, J., dos-Santos-Silva, I., Gibson, L., Nevanlinna, H., Muranen, T. A., Aittomaki, K., Blomqvist, C., Hall, P., Li, J., Liu, J., Humphreys, K., Kang, D., Choi, J. Y., Park, S. K., Noh, D. Y., Matsuo, K., Ito, H., Iwata, H., Yatabe, Y., Guenel, P., Truong, T., Menegaux, F., Sanchez, M., Burwinkel, B., Marme, F., Schneeweiss, A., Sohn, C., Wu, A. H., Tseng, C. C., Van Den Berg, D., Stram, D. O., Benitez, J., Zamora, M. P., Perez, J. I., Menendez, P., Shu, X. O., Lu, W., Gao, Y. T., Cai, Q., Cox, A., Cross, S. S., Reed, M. W., Andrulis, I. L., Knight, J. A., Glendon, G., Tchatchou, S., Sawyer, E. J., Tomlinson, I., Kerin, M. J., Miller, N., Haiman, C. A., Henderson, B. E., Schumacher, F., Le, M. L., Lindblom, A., Margolin, S., Teo, S. H., Yip, C. H., Lee, D. S., Wong, T. Y., Hooning, M. J., Martens, J. W., Collee, J. M., van Deurzen, C. H., Hopper, J. L., Southey, M. C., Tsimiklis, H., Kapuscinski, M. K., Shen, C. Y., Wu, P. E., Yu, J. C., Chen, S. T., Alnaes, G. G., Borresen-Dale, A. L., Giles, G. G., Milne, R. L., McLean, C., Muir, K., Lophatananon, A., Stewart-Brown, S., Siriwanarangsan, P., Hartman, M., Miao, H., Buhari, S. A., Teo, Y. Y., Fasching, P. A., Haeberle, L., Ekici, A. B., Beckmann, M. W., Brenner, H., Dieffenbach, A. K., Arndt, V., Stegmaier, C., Swerdlow, A., Ashworth, A., Orr, N., Schoemaker, M. J., Garcia-Closas, M., Figueroa, J., Chanock, S. J., Lissowska, J., Simard, J., Goldberg, M. S., Labreche, F., Dumont, M., Winqvist, R., Pylkas, K., Jukkola-Vuorinen, A., Brauch, H., Bruning, T., Koto, Y. D., Radice, P., Peterlongo, P., Bonanni, B., Volorio, S., Dork, T., Bogdanova, N. V., Helbig, S., Mannermaa, A., Kataja, V., Kosma, V. M., Hartikainen, J. M., Devilee, P., Tollenaar, R. A., Seynaeve, C., van Asperen, C. J., Jakubowska, A., Lubinski, J., Jaworska-Bieniek, K., Durda, K., Slager, S., Toland, A. E., Ambrosone, C. B., Yannoukakos, D., Sangrajrang, S., Gaborieau, V., Brennan, P., McKay, J., Hamann, U., Torres, D., Zheng, W., Long, J., Anton-Culver, H., Neuhausen, S. L., Luccarini, C., Baynes, C., Ahmed, S., Maranian, M., Healey, C. S., Gonzalez-Neira, A., Pita, G., Alonso, M. R., Alvarez, N., Herrero, D., Tessier, D. C., Vincent, D., Bacot, F., de, S., I, Carroll, J., Caldas, C., Brown, M. A., Lupien, M., Kristensen, V. N., Pharoah, P. D., Chenevix-Trench, G., French, J. D., Easton, D. F. and Dunning, A. M. (2014). **Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation**. Nat. Commun *4*, 4999, doi: ncomms5999 [pii];10.1038/ncomms5999 [doi].

Ghoussaini, M., Fletcher, O., Michailidou, K., Turnbull, C., Schmidt, M. K., Dicks, E., Dennis, J., Wang, Q., Humphreys, M. K., Luccarini, C., Baynes, C., Conroy, D., Maranian, M., Ahmed, S., Driver, K., Johnson, N., Orr, N., dos, S. S., I, Waisfisz, Q., Meijers-Heijboer, H., Uitterlinden, A. G., Rivadeneira, F., Hall, P., Czene, K., Irwanto, A., Liu, J., Nevanlinna, H., Aittomaki, K., Blomqvist, C., Meindl, A., Schmutzler, R. K., Muller-Myhsok, B., Lichtner, P., Chang-Claude, J., Hein, R., Nickels, S., Flesch-Janys, D., Tsimiklis, H., Makalic, E., Schmidt, D., Bui, M., Hopper, J. L., Apicella, C., Park, D. J., Southey, M., Hunter, D. J., Chanock, S. J., Broeks, A., Verhoef, S., Hogervorst, F. B., Fasching, P. A., Lux, M. P., Beckmann, M. W., Ekici, A. B., Sawyer, E., Tomlinson, I., Kerin, M., Marme, F., Schneeweiss, A., Sohn, C., Burwinkel, B., Guenel, P., Truong, T., Cordina-Duverger, E., Menegaux, F., Bojesen, S. E., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Milne, R. L., Alonso, M. R., Gonzalez-Neira, A., Benitez, J., Anton-Culver, H., Ziogas, A., Bernstein, L., Dur, C. C., Brenner, H., Muller, H., Arndt, V., Stegmaier, C., Justenhoven, C., Brauch, H., Bruning, T., Wang-Gohrke, S., Eilber, U., Dork, T., Schurmann, P., Bremer, M., Hillemanns, P., Bogdanova, N. V., Antonenkova, N. N., Rogov, Y. I., Karstens, J. H., Bermisheva, M., Prokofieva, D., Khusnutdinova, E., Lindblom, A., Margolin, S., Mannermaa, A., Kataja, V., Kosma, V. M., Hartikainen, J. M., Lambrechts, D., Yesilyurt, B. T., Floris, G., Leunen, K., Manoukian, S., Bonanni, B., Fortuzzi, S., Peterlongo, P., Couch, F. J., Wang, X., Stevens, K., Lee, A., Giles, G. G., Baglietto, L., Severi, G., McLean, C., Alnaes, G. G., Kristensen, V., Borrensen-Dale, A. L., John, E. M., Miron, A., Winqvist, R., Pylkas, K., Jukkola-Vuorinen, A., Kauppila, S., Andrulis, I. L., Glendon, G., Mulligan, A. M., Devilee, P., van

Asperen, C. J., Tollenaar, R. A., Seynaeve, C., Figueroa, J. D., Garcia-Closas, M., Brinton, L., Lissowska, J., Hooning, M. J., Hollestelle, A., Oldenburg, R. A., van den Ouweland, A. M., Cox, A., Reed, M. W., Shah, M., Jakubowska, A., Lubinski, J., Jaworska, K., Durda, K., Jones, M., Schoemaker, M., Ashworth, A., Swerdlow, A., Beesley, J., Chen, X., Muir, K. R., Lophatananon, A., Rattanamongkongul, S., Chaiwerawattana, A., Kang, D., Yoo, K. Y., Noh, D. Y., Shen, C. Y., Yu, J. C., Wu, P. E., Hsiung, C. N., Perkins, A., Swann, R., Velentzis, L., Eccles, D. M., Tapper, W. J., Gerty, S. M., Graham, N. J., Ponder, B. A., Chenevix-Trench, G., Pharoah, P. D., Lathrop, M., Dunning, A. M., Rahman, N., Peto, J. and Easton, D. F. (2012). **Genome-wide association analysis identifies three new breast cancer susceptibility loci**. Nat. Genet *44*, 312-318, doi: ng.1049 [pii];10.1038/ng.1049 [doi].

Ghoussaini, M., French, J. D., Michailidou, K., Nord, S., Beesley, J., Canisus, S., Hillman, K. M., Kaufmann, S., Sivakumaran, H., Moradi, M. M., Lee, J. S., Dennis, J., Bolla, M. K., Wang, Q., Dicks, E., Milne, R. L., Hopper, J. L., Southey, M. C., Schmidt, M. K., Broeks, A., Muir, K., Lophatananon, A., Fasching, P. A., Beckmann, M. W., Fletcher, O., Johnson, N., Sawyer, E. J., Tomlinson, I., Burwinkel, B., Marme, F., Guenel, P., Truong, T., Bojesen, S. E., Flyger, H., Benitez, J., Gonzalez-Neira, A., Alonso, M. R., Pita, G., Neuhausen, S. L., Anton-Culver, H., Brenner, H., Arndt, V., Meindl, A., Schmutzler, R. K., Brauch, H., Hamann, U., Tessier, D. C., Vincent, D., Nevanlinna, H., Khan, S., Matsuo, K., Ito, H., Dork, T., Bogdanova, N. V., Lindblom, A., Margolin, S., Mannermaa, A., Kosma, V. M., Wu, A. H., Van Den Berg, D., Lambrechts, D., Floris, G., Chang-Claude, J., Rudolph, A., Radice, P., Barile, M., Couch, F. J., Hallberg, E., Giles, G. G., Haiman, C. A., Le, M. L., Goldberg, M. S., Teo, S. H., Yip, C. H., Borresen-Dale, A. L., Zheng, W., Cai, Q., Winqvist, R., Pylkas, K., Andrulis, I. L., Devilee, P., Tollenaar, R. A., Garcia-Closas, M., Figueroa, J., Hall, P., Czene, K., Brand, J. S., Darabi, H., Eriksson, M., Hooning, M. J., Koppert, L. B., Li, J., Shu, X. O., Zheng, Y., Cox, A., Cross, S. S., Shah, M., Rhenius, V., Choi, J. Y., Kang, D., Hartman, M., Chia, K. S., Kabisch, M., Torres, D., Luccarini, C., Conroy, D. M., Jakubowska, A., Lubinski, J., Sangrajrang, S., Brennan, P., Olswold, C., Slager, S., Shen, C. Y., Hou, M. F., Swerdlow, A., Schoemaker, M. J., Simard, J., Pharoah, P. D., Kristensen, V., Chenevix-Trench, G., Easton, D. F., Dunning, A. M. and Edwards, S. L. (2016). **Evidence that the 5p12 Variant rs10941679 Confers Susceptibility to Estrogen-Receptor-Positive Breast Cancer through FGF10 and MRPS30 Regulation**. Am. J. Hum. Genet *99*, 903-911, doi: S0002-9297(16)30291-9 [pii];10.1016/j.ajhg.2016.07.017 [doi].

Ghoussaini, M., Pharoah, P. D. and Easton, D. F. (2013). **Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning?** Am. J. Pathol *183*, 1038-1051, doi: S0002-9440(13)00520-8 [pii];10.1016/j.ajpath.2013.07.003 [doi].

Haiman, C. A., Chen, G. K., Vachon, C. M., Canzian, F., Dunning, A., Millikan, R. C., Wang, X., Ademuyiwa, F., Ahmed, S., Ambrosone, C. B., Baglietto, L., Balleine, R., Bandera, E. V., Beckmann, M. W., Berg, C. D., Bernstein, L., Blomqvist, C., Blot, W. J., Brauch, H., Buring, J. E., Carey, L. A., Carpenter, J. E., Chang-Claude, J., Chanock, S. J., Chasman, D. I., Clarke, C. L., Cox, A., Cross, S. S., Deming, S. L., Diasio, R. B., Dimopoulos, A. M., Driver, W. R., Dunnebier, T., Durcan, L., Eccles, D., Edlund, C. K., Ekici, A. B., Fasching, P. A., Feigelson, H. S., Flesch-Janys, D., Fostira, F., Forsti, A., Fountzilas, G., Gerty, S. M., Giles, G. G., Godwin, A. K., Goodfellow, P., Graham, N., Greco, D., Hamann, U., Hankinson, S. E., Hartmann, A., Hein, R., Heinz, J., Holbrook, A., Hoover, R. N., Hu, J. J., Hunter, D. J., Ingles, S. A., Irwanto, A., Ivanovich, J., John, E. M., Johnson, N., Jukkola-Vuorinen, A., Kaaks, R., Ko, Y. D., Kolonel, L. N., Konstantopoulou, I., Kosma, V. M., Kulkarni, S., Lambrechts, D., Lee, A. M., Marchand, L. L., Lesnick, T., Liu, J., Lindstrom, S., Mannermaa, A., Margolin, S., Martin, N. G., Miron, P., Montgomery, G. W., Nevanlinna, H., Nickels, S., Nyante, S., Olswold, C., Palmer, J., Pathak, H., Pectasides, D.,

Perou, C. M., Peto, J., Pharoah, P. D., Pooler, L. C., Press, M. F., Pylkas, K., Rebbeck, T. R., Rodriguez-Gil, J. L., Rosenberg, L., Ross, E., Rudiger, T., Silva, I. S., Sawyer, E., Schmidt, M. K., Schulz-Wendtland, R., Schumacher, F., Severi, G., Sheng, X., Signorello, L. B., Sinn, H. P., Stevens, K. N., Southey, M. C., Tapper, W. J., Tomlinson, I., Hogervorst, F. B., Wauters, E., Weaver, J., Wildiers, H., Winqvist, R., Van Den Berg, D., Wan, P., Xia, L. Y., Yannoukakos, D., Zheng, W., Ziegler, R. G., Siddiq, A., Slager, S. L., Stram, D. O., Easton, D., Kraft, P., Henderson, B. E. and Couch, F. J. (2011). **A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer**. Nat. Genet *43*, 1210-1214, doi: ng.985 [pii];10.1038/ng.985 [doi].

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B. and King, M. C. (1990). **Linkage of early-onset familial breast cancer to chromosome 17q21**. Science *250*, 1684-1689.

Hayati Rezvan, P., Lee, K. J. and Simpson, J. A. (2015). **The rise of multiple imputation: a review of the reporting and implementation of the method in medical research**. BMC Med Res Methodol *15*, 30, doi: 10.1186/s12874-015-0022-1.

He, Y. (2010). **Missing data analysis using multiple imputation: getting to the heart of the matter**. Circ Cardiovasc Qual Outcomes *3*, 98-105, doi: 10.1161/CIRCOUTCOMES.109.875658.

Hemminki, K. and Bermejo, J. L. (2007). **Constraints for genetic association studies imposed by attributable fraction and familial risk**. Carcinogenesis *28*, 648-656, doi: bgl182 [pii];10.1093/carcin/bgl182 [doi].

Hernandez, J. E., Llacuachaqui, M., Palacio, G. V., Figueroa, J. D., Madrid, J., Lema, M., Royer, R., Li, S., Larson, G., Weitzel, J. N. and Narod, S. A. (2014). **Prevalence of BRCA1 and BRCA2 mutations in unselected breast cancer patients from medellin, Colombia**. Hered Cancer Clin Pract *12*, 11, doi: 10.1186/1897-4287-12-11.

Hines, L. M., Risendal, B., Slattery, M. L., Baumgartner, K. B., Giuliano, A. R., Sweeney, C., Rollison, D. E. and Byers, T. (2010). **Comparative analysis of breast cancer risk factors among Hispanic and non-Hispanic white women**. Cancer *116*, 3215-3223, doi: 10.1002/cncr.25154.

Howell, A., Anderson, A. S., Clarke, R. B., Duffy, S. W., Evans, D. G., Garcia-Closas, M., Gescher, A. J., Key, T. J., Saxton, J. M. and Harvie, M. N. (2014). **Risk determination and prevention of breast cancer**. Breast Cancer Res *16*, 446, doi: 10.1186/s13058-014-0446-2.

Hsieh, C. C., Trichopoulos, D., Katsouyanni, K. and Yuasa, S. (1990). **Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case-control study**. Int J Cancer *46*, 796-800.

Hunter, D. J., Kraft, P., Jacobs, K. B., Cox, D. G., Yeager, M., Hankinson, S. E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., Wang, J., Yu, K., Chatterjee, N., Orr, N., Willett, W. C., Colditz, G. A., Ziegler, R. G., Berg, C. D., Buys, S. S., McCarty, C. A., Feigelson, H. S., Calle, E. E., Thun, M. J., Hayes, R. B., Tucker, M., Gerhard, D. S., Fraumeni, J. F., Jr., Hoover, R. N., Thomas, G. and

Chanock, S. J. (2007a). **A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer**. Nat. Genet *39*, 870-874, doi: ng2075 [pii];10.1038/ng2075 [doi].

Hunter, D. J., Kraft, P., Jacobs, K. B., Cox, D. G., Yeager, M., Hankinson, S. E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., Wang, J., Yu, K., Chatterjee, N., Orr, N., Willett, W. C., Colditz, G. A., Ziegler, R. G., Berg, C. D., Buys, S. S., McCarty, C. A., Feigelson, H. S., Calle, E. E., Thun, M. J., Hayes, R. B., Tucker, M., Gerhard, D. S., Fraumeni, J. F., Jr., Hoover, R. N., Thomas, G. and Chanock, S. J. (2007b). **A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer**. Nat Genet *39*, 870-874, doi: 10.1038/ng2075.

Husing, A., Canzian, F., Beckmann, L., Garcia-Closas, M., Diver, W. R., Thun, M. J., Berg, C. D., Hoover, R. N., Ziegler, R. G., Figueroa, J. D., Isaacs, C., Olsen, A., Viallon, V., Boeing, H., Masala, G., Trichopoulos, D., Peeters, P. H., Lund, E., Ardanaz, E., Khaw, K. T., Lenner, P., Kolonel, L. N., Stram, D. O., Le Marchand, L., McCarty, C. A., Buring, J. E., Lee, I. M., Zhang, S., Lindstrom, S., Hankinson, S. E., Riboli, E., Hunter, D. J., Henderson, B. E., Chanock, S. J., Haiman, C. A., Kraft, P., Kaaks, R. and Bpc (2012). **Prediction of breast cancer risk by genetic risk factors, overall and by hormone receptor status**. J Med Genet *49*, 601-608, doi: 10.1136/jmedgenet-2011-100716.

Inc, S. I. (2012). **PROC MI. SAS Procedures Guide**. Version 9.4 *Cary, NC: SAS Institute Inc*.

Ioannides, S. J., Barlow, P. L., Elwood, J. M. and Porter, D. (2014). **Effect of obesity on aromatase inhibitor efficacy in postmenopausal, hormone receptor-positive breast cancer: a systematic review**. Breast Cancer Res Treat *147*, 237-248, doi: 10.1007/s10549-014-3091-7.

Jiagge, E., Chitale, D. and Newman, L. A. (2018). **Triple-Negative Breast Cancer, Stem Cells, and African Ancestry**. Am J Pathol *188*, 271-279, doi: 10.1016/j.ajpath.2017.06.020.

Jiagge, E., Oppong, J. K., Bensenhaver, J., Aitpillah, F., Gyan, K., Kyei, I., Osei-Bonsu, E., Adjei, E., Ohene-Yeboah, M., Toy, K., Jackson, K. E., Akpaloo, M., Acheampong, D., Antwi, B., Agyeman, F. O., Alhassan, Z., Fondjo, L. A., Owusu-Afriyie, O., Brewer, R. N., Gyamfuah, A., Salem, B., Johnson, T., Wicha, M., Merajver, S., Kleer, C., Pang, J., Amankwaa-Frempong, E., Stark, A., Abantanga, F., Newman, L. and Awuah, B. (2016). **Breast Cancer and African Ancestry: Lessons Learned at the 10-Year Anniversary of the Ghana-Michigan Research Partnership and International Breast Registry**. J Glob Oncol *2*, 302-310, doi: 10.1200/JGO.2015.002881.

Jiang, X., Castelao, J. E., Chavez-Uribe, E., Fernandez Rodriguez, B., Celeiro Munoz, C., Redondo, C. M., Pena Fernandez, M., Novo Dominguez, A., Pereira, C. D., Martinez, M. E., Garcia-Caballero, T., Fraga Rodriguez, M., Antunez, J., Carracedo, A., Forteza-Vila, J. and Gago-Dominguez, M. (2012). **Family history and breast cancer hormone receptor status in a Spanish cohort**. PLoS One *7*, e29459, doi: 10.1371/journal.pone.0029459.

Johnson, K. C. (2005). **Accumulating evidence on passive and active smoking and breast cancer risk**. Int J Cancer *117*, 619-628, doi: 10.1002/ijc.21150.

Kann, S., Schmid, S. M., Eichholzer, M., Huang, D. J., Amann, E. and Guth, U. (2014). **The impact of overweight and obesity on breast cancer: data from Switzerland, so far a country little affected by the current global obesity epidemic**. Gland Surg *3*, 181-197, doi: 10.3978/j.issn.2227-684X.2013.12.01.

Kelsey, J. L., Gammon, M. D. and John, E. M. (1993). **Reproductive factors and breast cancer**. Epidemiol. Rev *15*, 36-47.

Kuchenbaecker, K. B., McGuffog, L., Barrowdale, D., Lee, A., Soucy, P., Dennis, J., Domchek, S. M., Robson, M., Spurdle, A. B., Ramus, S. J., Mavaddat, N., Terry, M. B., Neuhausen, S. L., Schmutzler, R. K., Simard, J., Pharoah, P. D. P., Offit, K., Couch, F. J., Chenevix-Trench, G., Easton, D. F. and Antoniou, A. C. (2017). **Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers**. J Natl Cancer Inst *109*, doi: 10.1093/jnci/djw302.

Laitinen, J., Samarut, J. and Holtta, E. (1994). **A nontoxic and versatile protein salting-out method for isolation of DNA**. Biotechniques *17*, 316, 318, 320-312.

Lange, K., Papp, J. C., Sinsheimer, J. S., Sripracha, R., Zhou, H. and Sobel, E. M. (2013). **Mendel: the Swiss army knife of genetic analysis programs**. Bioinformatics *29*, 1568-1570, doi: 10.1093/bioinformatics/btt187.

Lange, K., Sinsheimer, J. S. and Sobel, E. (2005). **Association testing with Mendel**. Genet Epidemiol *29*, 36-50, doi: 10.1002/gepi.20073.

Lecarpentier, J., Silvestri, V., Kuchenbaecker, K. B., Barrowdale, D., Dennis, J., McGuffog, L., Soucy, P., Leslie, G., Rizzolo, P., Navazio, A. S., Valentini, V., Zelli, V., Lee, A., Amin Al Olama, A., Tyrer, J. P., Southey, M., John, E. M., Conner, T. A., Goldgar, D. E., Buys, S. S., Janavicius, R., Steele, L., Ding, Y. C., Neuhausen, S. L., Hansen, T. V. O., Osorio, A., Weitzel, J. N., Toss, A., Medici, V., Cortesi, L., Zanna, I., Palli, D., Radice, P., Manoukian, S., Peissel, B., Azzollini, J., Viel, A., Cini, G., Damante, G., Tommasi, S., Peterlongo, P., Fostira, F., Hamann, U., Evans, D. G., Henderson, A., Brewer, C., Eccles, D., Cook, J., Ong, K. R., Walker, L., Side, L. E., Porteous, M. E., Davidson, R., Hodgson, S., Frost, D., Adlard, J., Izatt, L., Eeles, R., Ellis, S., Tischkowitz, M., Embrace, Godwin, A. K., Meindl, A., Gehrig, A., Dworniczak, B., Sutter, C., Engel, C., Niederacher, D., Steinemann, D., Hahnen, E., Hauke, J., Rhiem, K., Kast, K., Arnold, N., Ditsch, N., Wang-Gohrke, S., Wappenschmidt, B., Wand, D., Lasset, C., Stoppa-Lyonnet, D., Belotti, M., Damiola, F., Barjhoux, L., Mazoyer, S., Collaborators, G. S., Van Heetvelde, M., Poppe, B., De Leeneer, K., Claes, K. B. M., de la Hoya, M., Garcia-Barberan, V., Caldes, T., Perez Segura, P., Kiiski, J. I., Aittomaki, K., Khan, S., Nevanlinna, H., van Asperen, C. J., Hebon, Vaszko, T., Kasler, M., Olah, E., Balmana, J., Gutierrez-Enriquez, S., Diez, O., Teule, A., Izquierdo, A., Darder, E., Brunet, J., Del Valle, J., Feliubadalo, L., Pujana, M. A., Lazaro, C., Arason, A., Agnarsson, B. A., Johannsson, O. T., Barkardottir, R. B., Alducci, E., Tognazzo, S., Montagna, M., Teixeira, M. R., Pinto, P., Spurdle, A. B., Holland, H., Investigators, K. C., Lee, J. W., Lee, M. H., Lee, J., Kim, S. W., Kang, E., Kim, Z., Sharma, P., Rebbeck, T. R., Vijai, J., Robson, M., Lincoln, A., Musinsky, J., Gaddam, P., Tan, Y. Y., Berger, A., Singer, C. F., Loud, J. T., Greene, M. H., Mulligan, A. M., Glendon, G., Andrulis, I. L., Toland, A. E., Senter, L., Bojesen, A., Nielsen, H. R., Skytte, A. B., Sunde, L., Jensen, U. B., Pedersen, I. S., Krogh, L., Kruse, T. A.,

Caligo, M. A., Yoon, S. Y., Teo, S. H., von Wachenfeldt, A., Huo, D., Nielsen, S. M., Olopade, O. I., Nathanson, K. L., Domchek, S. M., Lorenchick, C., Jankowitz, R. C., Campbell, I., James, P., Mitchell, G., Orr, N., Park, S. K., Thomassen, M., Offit, K., Couch, F. J., Simard, J., Easton, D. F., Chenevix-Trench, G., Schmutzler, R. K., Antoniou, A. C. and Ottini, L. (2017). **Prediction of Breast and Prostate Cancer Risks in Male BRCA1 and BRCA2 Mutation Carriers Using Polygenic Risk Scores**. J Clin Oncol *35*, 2240-2250, doi: 10.1200/JCO.2016.69.4935.

Li, C. I., Uribe, D. J. and Daling, J. R. (2005). **Clinical characteristics of different histologic types of breast cancer**. Br J Cancer *93*, 1046-1052, doi: 10.1038/sj.bjc.6602787.

Li, Y., Willer, C., Sanna, S. and Abecasis, G. (2009). **Genotype imputation**. Annu Rev Genomics Hum Genet *10*, 387-406, doi: 10.1146/annurev.genom.9.081307.164242.

Ling, H., Li, S., Wu, Y., Zheng, Y. Z., Qiao, F., Yao, L., Cao, Z. G., Ye, F. G., Wu, J., Hu, X., Wang, B. and Shao, Z. M. (2016). **Genetic evaluation of BRCA1 associated a complex genes with triple-negative breast cancer susceptibility in Chinese women**. Oncotarget *7*, 9759-9772, doi: 10.18632/oncotarget.7112.

Liu, Y., Nyunoya, T., Leng, S., Belinsky, S. A., Tesfaigzi, Y. and Bruse, S. (2013). **Softwares and methods for estimating genetic ancestry in human populations**. Hum Genomics *7*, 1, doi: 10.1186/1479-7364-7-1.

Long, J., Cai, Q., Shu, X. O., Qu, S., Li, C., Zheng, Y., Gu, K., Wang, W., Xiang, Y. B., Cheng, J., Chen, K., Zhang, L., Zheng, H., Shen, C. Y., Huang, C. S., Hou, M. F., Shen, H., Hu, Z., Wang, F., Deming, S. L., Kelley, M. C., Shrubsole, M. J., Khoo, U. S., Chan, K. Y., Chan, S. Y., Haiman, C. A., Henderson, B. E., Le, M. L., Iwasaki, M., Kasuga, Y., Tsugane, S., Matsuo, K., Tajima, K., Iwata, H., Huang, B., Shi, J., Li, G., Wen, W., Gao, Y. T., Lu, W. and Zheng, W. (2010). **Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium**. PLoS. Genet *6*, e1001002, doi: 10.1371/journal.pgen.1001002 [doi].

Malhotra, G. K., Zhao, X., Band, H. and Band, V. (2010). **Histological, molecular and functional subtypes of breast cancers**. Cancer Biol Ther *10*, 955-960.

Mavaddat, N., Pharoah, P. D., Michailidou, K., Tyrer, J., Brook, M. N., Bolla, M. K., Wang, Q., Dennis, J., Dunning, A. M., Shah, M., Luben, R., Brown, J., Bojesen, S. E., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Czene, K., Darabi, H., Eriksson, M., Peto, J., dos-Santos-Silva, I., Dudbridge, F., Johnson, N., Schmidt, M. K., Broeks, A., Verhoef, S., Rutgers, E. J., Swerdlow, A., Ashworth, A., Orr, N., Schoemaker, M. J., Figueroa, J., Chanock, S. J., Brinton, L., Lissowska, J., Couch, F. J., Olson, J. E., Vachon, C., Pankratz, V. S., Lambrechts, D., Wildiers, H., Van, O. C., Van, L. E., Kristensen, V., Grenaker, A. G., Nord, S., Borresen-Dale, A. L., Nevanlinna, H., Muranen, T. A., Aittomaki, K., Blomqvist, C., Chang-Claude, J., Rudolph, A., Seibold, P., Flesch-Janys, D., Fasching, P. A., Haeberle, L., Ekici, A. B., Beckmann, M. W., Burwinkel, B., Marme, F., Schneeweiss, A., Sohn, C., Trentham-Dietz, A., Newcomb, P., Titus, L., Egan, K. M., Hunter, D. J., Lindstrom, S., Tamimi, R. M., Kraft, P., Rahman, N., Turnbull, C., Renwick, A., Seal, S., Li, J., Liu, J., Humphreys, K., Benitez, J., Pilar, Z. M., Arias Perez, J. I., Menendez, P., Jakubowska, A., Lubinski, J., Jaworska-Bieniek, K., Durda, K., Bogdanova, N. V., Antonenkova, N. N., Dork, T., Anton-Culver, H., Neuhausen, S. L., Ziogas, A., Bernstein, L., Devilee, P., Tollenaar, R. A.,

Seynaeve, C., van Asperen, C. J., Cox, A., Cross, S. S., Reed, M. W., Khusnutdinova, E., Bermisheva, M., Prokofyeva, D., Takhirova, Z., Meindl, A., Schmutzler, R. K., Sutter, C., Yang, R., Schurmann, P., Bremer, M., Christiansen, H., Park-Simon, T. W., Hillemanns, P., Guenel, P., Truong, T., Menegaux, F., Sanchez, M., Radice, P., Peterlongo, P., Manoukian, S., Pensotti, V., Hopper, J. L., Tsimiklis, H., Apicella, C., Southey, M. C., Brauch, H., Bruning, T., Ko, Y. D., Sigurdson, A. J., Doody, M. M., Hamann, U., Torres, D., Ulmer, H. U., Forsti, A., Sawyer, E. J., Tomlinson, I., Kerin, M. J., Miller, N., Andrulis, I. L., Knight, J. A., Glendon, G., Marie, M. A., Chenevix-Trench, G., Balleine, R., Giles, G. G., Milne, R. L., McLean, C., Lindblom, A., Margolin, S., Haiman, C. A., Henderson, B. E., Schumacher, F., Le, M. L., Eilber, U., Wang-Gohrke, S., Hooning, M. J., Hollestelle, A., van den Ouweland, A. M., Koppert, L. B., Carpenter, J., Clarke, C., Scott, R., Mannermaa, A., Kataja, V., Kosma, V. M., Hartikainen, J. M., Brenner, H., Arndt, V., Stegmaier, C., Karina, D. A., Winqvist, R., Pylkas, K., Jukkola-Vuorinen, A., Grip, M., Offit, K., Vijai, J., Robson, M., Rau-Murthy, R., Dwek, M., Swann, R., Annie, P. K., Goldberg, M. S., Labreche, F., Dumont, M., Eccles, D. M., Tapper, W. J., Rafiq, S., John, E. M., Whittemore, A. S., Slager, S., Yannoukakos, D., Toland, A. E., Yao, S., Zheng, W., Halverson, S. L., Gonzalez-Neira, A., Pita, G., Rosario, A. M., Alvarez, N., Herrero, D., Tessier, D. C., Vincent, D., Bacot, F., Luccarini, C., Baynes, C., Ahmed, S., Maranian, M., Healey, C. S., Simard, J., Hall, P., Easton, D. F. and Garcia-Closas, M. (2015). **Prediction of breast cancer risk based on profiling with common genetic variants**. J. Natl. Cancer Inst *107*, doi: djv036 [pii];10.1093/jnci/djv036 [doi].

Meyer, K. B., O'Reilly, M., Michailidou, K., Carlebur, S., Edwards, S. L., French, J. D., Prathalingham, R., Dennis, J., Bolla, M. K., Wang, Q., de, S., I, Hopper, J. L., Tsimiklis, H., Apicella, C., Southey, M. C., Schmidt, M. K., Broeks, A., Van 't Veer, L. J., Hogervorst, F. B., Muir, K., Lophatananon, A., Stewart-Brown, S., Siriwanarangsan, P., Fasching, P. A., Lux, M. P., Ekici, A. B., Beckmann, M. W., Peto, J., dos, S. S., I, Fletcher, O., Johnson, N., Sawyer, E. J., Tomlinson, I., Kerin, M. J., Miller, N., Marme, F., Schneeweiss, A., Sohn, C., Burwinkel, B., Guenel, P., Truong, T., Laurent-Puig, P., Menegaux, F., Bojesen, S. E., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Milne, R. L., Zamora, M. P., Arias, J. I., Benitez, J., Neuhausen, S., Anton-Culver, H., Ziogas, A., Dur, C. C., Brenner, H., Muller, H., Arndt, V., Stegmaier, C., Meindl, A., Schmutzler, R. K., Engel, C., Ditsch, N., Brauch, H., Bruning, T., Ko, Y. D., Nevanlinna, H., Muranen, T. A., Aittomaki, K., Blomqvist, C., Matsuo, K., Ito, H., Iwata, H., Yatabe, Y., Dork, T., Helbig, S., Bogdanova, N. V., Lindblom, A., Margolin, S., Mannermaa, A., Kataja, V., Kosma, V. M., Hartikainen, J. M., Chenevix-Trench, G., Wu, A. H., Tseng, C. C., Van Den Berg, D., Stram, D. O., Lambrechts, D., Thienpont, B., Christiaens, M. R., Smeets, A., Chang-Claude, J., Rudolph, A., Seibold, P., Flesch-Janys, D., Radice, P., Peterlongo, P., Bonanni, B., Bernard, L., Couch, F. J., Olson, J. E., Wang, X., Purrington, K., Giles, G. G., Severi, G., Baglietto, L., McLean, C., Haiman, C. A., Henderson, B. E., Schumacher, F., Le, M. L., Simard, J., Goldberg, M. S., Labreche, F., Dumont, M., Teo, S. H., Yip, C. H., Phuah, S. Y., Kristensen, V., Grenaker, A. G., Borresen-Dale, A. L., Zheng, W., Deming-Halverson, S., Shrubsole, M., Long, J., Winqvist, R., Pylkas, K., Jukkola-Vuorinen, A., Kauppila, S., Andrulis, I. L., Knight, J. A., Glendon, G., Tchatchou, S., Devilee, P., Tollenaar, R. A., Seynaeve, C. M., Garcia-Closas, M., Figueroa, J., Chanock, S. J., Lissowska, J., Czene, K., Darabi, H., Eriksson, K., Hooning, M. J., Martens, J. W., van den Ouweland, A. M., van Deurzen, C. H., Hall, P., Li, J., Liu, J., Humphreys, K., Shu, X. O., Lu, W., Gao, Y. T., Cai, H., Cox, A., Reed, M. W., Blot, W., Signorello, L. B., Cai, Q., Pharoah, P. D., Ghoussaini, M., Harrington, P., Tyrer, J., Kang, D., Choi, J. Y., Park, S. K., Noh, D. Y., Hartman, M., Hui, M., Lim, W. Y., Buhari, S. A., Hamann, U., Forsti, A., Rudiger, T., Ulmer, H. U., Jakubowska, A., Lubinski, J., Jaworska, K., Durda, K., Sangrajrang, S., Gaborieau, V., Brennan, P., McKay, J., Vachon, C., Slager, S., Fostira, F., Pilarski, R., Shen, C. Y., Hsiung, C. N., Wu, P. E., Hou, M. F., Swerdlow, A., Ashworth, A., Orr, N., Schoemaker, M. J., Ponder, B. A., Dunning, A. M. and Easton, D. F. (2013). **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, doi: S0002-9297(13)00483-7 [pii];10.1016/j.ajhg.2013.10.026 [doi].

Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M. J., Maranian, M. J., Bolla, M. K., Wang, Q., Shah, M., Perkins, B. J., Czene, K., Eriksson, M., Darabi, H., Brand, J. S., Bojesen, S. E., Nordestgaard, B. G., Flyger, H., Nielsen, S. F., Rahman, N., Turnbull, C., Fletcher, O., Peto, J., Gibson, L., dos-Santos-Silva, I., Chang-Claude, J., Flesch-Janys, D., Rudolph, A., Eilber, U., Behrens, S., Nevanlinna, H., Muranen, T. A., Aittomaki, K., Blomqvist, C., Khan, S., Aaltonen, K., Ahsan, H., Kibriya, M. G., Whittemore, A. S., John, E. M., Malone, K. E., Gammon, M. D., Santella, R. M., Ursin, G., Makalic, E., Schmidt, D. F., Casey, G., Hunter, D. J., Gapstur, S. M., Gaudet, M. M., Diver, W. R., Haiman, C. A., Schumacher, F., Henderson, B. E., Le, M. L., Berg, C. D., Chanock, S. J., Figueroa, J., Hoover, R. N., Lambrechts, D., Neven, P., Wildiers, H., Van, L. E., Schmidt, M. K., Broeks, A., Verhoef, S., Cornelissen, S., Couch, F. J., Olson, J. E., Hallberg, E., Vachon, C., Waisfisz, Q., Meijers-Heijboer, H., Adank, M. A., van der Luijt, R. B., Li, J., Liu, J., Humphreys, K., Kang, D., Choi, J. Y., Park, S. K., Yoo, K. Y., Matsuo, K., Ito, H., Iwata, H., Tajima, K., Guenel, P., Truong, T., Mulot, C., Sanchez, M., Burwinkel, B., Marme, F., Surowy, H., Sohn, C., Wu, A. H., Tseng, C. C., Van Den Berg, D., Stram, D. O., Gonzalez-Neira, A., Benitez, J., Zamora, M. P., Perez, J. I., Shu, X. O., Lu, W., Gao, Y. T., Cai, H., Cox, A., Cross, S. S., Reed, M. W., Andrulis, I. L., Knight, J. A., Glendon, G., Mulligan, A. M., Sawyer, E. J., Tomlinson, I., Kerin, M. J., Miller, N., Lindblom, A., Margolin, S., Teo, S. H., Yip, C. H., Taib, N. A., Tan, G. H., Hooning, M. J., Hollestelle, A., Martens, J. W., Collee, J. M., Blot, W., Signorello, L. B., Cai, Q., Hopper, J. L., Southey, M. C., Tsimiklis, H., Apicella, C., Shen, C. Y., Hsiung, C. N., Wu, P. E., Hou, M. F., Kristensen, V. N., Nord, S., Alnaes, G. I., Giles, G. G., Milne, R. L., McLean, C., Canzian, F., Trichopoulos, D., Peeters, P., Lund, E., Sund, M., Khaw, K. T., Gunter, M. J., Palli, D., Mortensen, L. M., Dossus, L., Huerta, J. M., Meindl, A., Schmutzler, R. K., Sutter, C., Yang, R., Muir, K., Lophatananon, A., Stewart-Brown, S., Siriwanarangsan, P., Hartman, M., Miao, H., Chia, K. S., Chan, C. W., Fasching, P. A., Hein, A., Beckmann, M. W., Haeberle, L., Brenner, H., Dieffenbach, A. K., Arndt, V., Stegmaier, C., Ashworth, A., Orr, N., Schoemaker, M. J., Swerdlow, A. J., Brinton, L., Garcia-Closas, M., Zheng, W., Halverson, S. L., Shrubsole, M., Long, J., Goldberg, M. S., Labreche, F., Dumont, M., Winqvist, R., Pylkas, K., Jukkola-Vuorinen, A., Grip, M., Brauch, H., Hamann, U., Bruning, T., Radice, P., Peterlongo, P., Manoukian, S., Bernard, L., Bogdanova, N. V., Dork, T., Mannermaa, A., Kataja, V., Kosma, V. M., Hartikainen, J. M., Devilee, P., Tollenaar, R. A., Seynaeve, C., van Asperen, C. J., Jakubowska, A., Lubinski, J., Jaworska, K., Huzarski, T., Sangrajrang, S., Gaborieau, V., Brennan, P., McKay, J., Slager, S., Toland, A. E., Ambrosone, C. B., Yannoukakos, D., Kabisch, M., Torres, D., Neuhausen, S. L., Anton-Culver, H., Luccarini, C., Baynes, C., Ahmed, S., Healey, C. S., Tessier, D. C., Vincent, D., Bacot, F., Pita, G., Alonso, M. R., Alvarez, N., Herrero, D., Simard, J., Pharoah, P. P., Kraft, P., Dunning, A. M., Chenevix-Trench, G., Hall, P. and Easton, D. F. (2015). **Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer**. Nat. Genet *47*, 373-380, doi: ng.3242 [pii];10.1038/ng.3242 [doi].

Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R. L., Schmidt, M. K., Chang-Claude, J., Bojesen, S. E., Bolla, M. K., Wang, Q., Dicks, E., Lee, A., Turnbull, C., Rahman, N., Fletcher, O., Peto, J., Gibson, L., dos, S. S., I, Nevanlinna, H., Muranen, T. A., Aittomaki, K., Blomqvist, C., Czene, K., Irwanto, A., Liu, J., Waisfisz, Q., Meijers-Heijboer, H., Adank, M., van der Luijt, R. B., Hein, R., Dahmen, N., Beckman, L., Meindl, A., Schmutzler, R. K., Muller-Myhsok, B., Lichtner, P., Hopper, J. L., Southey, M. C., Makalic, E., Schmidt, D. F., Uitterlinden, A. G., Hofman, A., Hunter, D. J., Chanock, S. J., Vincent, D., Bacot, F., Tessier, D. C., Canisius, S., Wessels, L. F., Haiman, C. A., Shah, M., Luben, R., Brown, J., Luccarini, C., Schoof, N., Humphreys, K., Li, J., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Couch, F. J., Wang, X., Vachon, C., Stevens, K. N., Lambrechts, D., Moisse, M., Paridaens, R., Christiaens, M. R., Rudolph, A., Nickels, S., Flesch-Janys, D., Johnson, N., Aitken, Z., Aaltonen, K., Heikkinen, T., Broeks, A., Veer, L. J., van der Schoot, C. E., Guenel, P., Truong, T., Laurent-

Puig, P., Menegaux, F., Marme, F., Schneeweiss, A., Sohn, C., Burwinkel, B., Zamora, M. P., Perez, J. I., Pita, G., Alonso, M. R., Cox, A., Brock, I. W., Cross, S. S., Reed, M. W., Sawyer, E. J., Tomlinson, I., Kerin, M. J., Miller, N., Henderson, B. E., Schumacher, F., Le, M. L., Andrulis, I. L., Knight, J. A., Glendon, G., Mulligan, A. M., Lindblom, A., Margolin, S., Hooning, M. J., Hollestelle, A., van den Ouweland, A. M., Jager, A., Bui, Q. M., Stone, J., Dite, G. S., Apicella, C., Tsimiklis, H., Giles, G. G., Severi, G., Baglietto, L., Fasching, P. A., Haeberle, L., Ekici, A. B., Beckmann, M. W., Brenner, H., Muller, H., Arndt, V., Stegmaier, C., Swerdlow, A., Ashworth, A., Orr, N., Jones, M., Figueroa, J., Lissowska, J., Brinton, L., Goldberg, M. S., Labreche, F., Dumont, M., Winqvist, R., Pylkas, K., Jukkola-Vuorinen, A., Grip, M., Brauch, H., Hamann, U., Bruning, T., Radice, P., Peterlongo, P., Manoukian, S., Bonanni, B., Devilee, P., Tollenaar, R. A., Seynaeve, C., van Asperen, C. J., Jakubowska, A., Lubinski, J., Jaworska, K., Durda, K., Mannermaa, A., Kataja, V., Kosma, V. M., Hartikainen, J. M., Bogdanova, N. V., Antonenkova, N. N., Dork, T., Kristensen, V. N., Anton-Culver, H., Slager, S., Toland, A. E., Edge, S., Fostira, F., Kang, D., Yoo, K. Y., Noh, D. Y., Matsuo, K., Ito, H., Iwata, H., Sueta, A., Wu, A. H., Tseng, C. C., Van Den Berg, D., Stram, D. O., Shu, X. O., Lu, W., Gao, Y. T., Cai, H., Teo, S. H., Yip, C. H., Phuah, S. Y., Cornes, B. K., Hartman, M., Miao, H., Lim, W. Y., Sng, J. H., Muir, K., Lophatananon, A., Stewart-Brown, S., Siriwanarangsan, P., Shen, C. Y., Hsiung, C. N., Wu, P. E., Ding, S. L., Sangrajrang, S., Gaborieau, V., Brennan, P., McKay, J., Blot, W. J., Signorello, L. B., Cai, Q., Zheng, W., Deming-Halverson, S., Shrubsole, M., Long, J., Simard, J., Garcia-Closas, M., Pharoah, P. D., Chenevix-Trench, G., Dunning, A. M., Benitez, J. and Easton, D. F. (2013). **Large-scale genotyping identifies 41 new loci associated with breast cancer risk**. Nat. Genet *45*, 353-361, doi: ng.2563 [pii];10.1038/ng.2563 [doi].

Michailidou, K., Lindstrom, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemacon, A., Soucy, P., Glubb, D., Rostamianfar, A., Bolla, M. K., Wang, Q., Tyrer, J., Dicks, E., Lee, A., Wang, Z., Allen, J., Keeman, R., Eilber, U., French, J. D., Qing Chen, X., Fachal, L., McCue, K., McCart Reed, A. E., Ghoussaini, M., Carroll, J. S., Jiang, X., Finucane, H., Adams, M., Adank, M. A., Ahsan, H., Aittomaki, K., Anton-Culver, H., Antonenkova, N. N., Arndt, V., Aronson, K. J., Arun, B., Auer, P. L., Bacot, F., Barrdahl, M., Baynes, C., Beckmann, M. W., Behrens, S., Benitez, J., Bermisheva, M., Bernstein, L., Blomqvist, C., Bogdanova, N. V., Bojesen, S. E., Bonanni, B., Borresen-Dale, A. L., Brand, J. S., Brauch, H., Brennan, P., Brenner, H., Brinton, L., Broberg, P., Brock, I. W., Broeks, A., Brooks-Wilson, A., Brucker, S. Y., Bruning, T., Burwinkel, B., Butterbach, K., Cai, Q., Cai, H., Caldes, T., Canzian, F., Carracedo, A., Carter, B. D., Castelao, J. E., Chan, T. L., David Cheng, T. Y., Seng Chia, K., Choi, J. Y., Christiansen, H., Clarke, C. L., Collaborators, N., Collee, M., Conroy, D. M., Cordina-Duverger, E., Cornelissen, S., Cox, D. G., Cox, A., Cross, S. S., Cunningham, J. M., Czene, K., Daly, M. B., Devilee, P., Doheny, K. F., Dork, T., Dos-Santos-Silva, I., Dumont, M., Durcan, L., Dwek, M., Eccles, D. M., Ekici, A. B., Eliassen, A. H., Ellberg, C., Elvira, M., Engel, C., Eriksson, M., Fasching, P. A., Figueroa, J., Flesch-Janys, D., Fletcher, O., Flyger, H., Fritschi, L., Gaborieau, V., Gabrielson, M., Gago-Dominguez, M., Gao, Y. T., Gapstur, S. M., Garcia-Saenz, J. A., Gaudet, M. M., Georgoulias, V., Giles, G. G., Glendon, G., Goldberg, M. S., Goldgar, D. E., Gonzalez-Neira, A., Grenaker Alnaes, G. I., Grip, M., Gronwald, J., Grundy, A., Guenel, P., Haeberle, L., Hahnen, E., Haiman, C. A., Hakansson, N., Hamann, U., Hamel, N., Hankinson, S., Harrington, P., Hart, S. N., Hartikainen, J. M., Hartman, M., Hein, A., Heyworth, J., Hicks, B., Hillemanns, P., Ho, D. N., Hollestelle, A., Hooning, M. J., Hoover, R. N., Hopper, J. L., Hou, M. F., Hsiung, C. N., Huang, G., Humphreys, K., Ishiguro, J., Ito, H., Iwasaki, M., Iwata, H., Jakubowska, A., Janni, W., John, E. M., Johnson, N., Jones, K., Jones, M., Jukkola-Vuorinen, A., Kaaks, R., Kabisch, M., Kaczmarek, K., Kang, D., Kasuga, Y., Kerin, M. J., Khan, S., Khusnutdinova, E., Kiiski, J. I., Kim, S. W., Knight, J. A., Kosma, V. M., Kristensen, V. N., Kruger, U., Kwong, A., Lambrechts, D., Le Marchand, L., Lee, E., Lee, M. H., Lee, J. W., Neng Lee, C., Lejbkowicz, F., Li, J., Lilyquist, J., Lindblom, A., Lissowska, J., Lo, W. Y., Loibl, S., Long, J., Lophatananon, A., Lubinski, J., Luccarini, C., Lux, M. P., Ma, E. S. K., MacInnis, R. J., Maishman, T., Makalic, E., Malone, K. E., Kostovska, I. M.,

Mannermaa, A., Manoukian, S., Manson, J. E., Margolin, S., Mariapun, S., Martinez, M. E., Matsuo, K., Mavroudis, D., McKay, J., McLean, C., Meijers-Heijboer, H., Meindl, A., Menendez, P., Menon, U., Meyer, J., Miao, H., Miller, N., Taib, N. A. M., Muir, K., Mulligan, A. M., Mulot, C., Neuhausen, S. L., Nevanlinna, H., Neven, P., Nielsen, S. F., Noh, D. Y., Nordestgaard, B. G., Norman, A., Olopade, O. I., Olson, J. E., Olsson, H., Olswold, C., Orr, N., Pankratz, V. S., Park, S. K., Park-Simon, T. W., Lloyd, R., Perez, J. I. A., Peterlongo, P., Peto, J., Phillips, K. A., Pinchev, M., Plaseska-Karanfilska, D., Prentice, R., Presneau, N., Prokofyeva, D., Pugh, E., Pylkas, K., Rack, B., Radice, P., Rahman, N., Rennert, G., Rennert, H. S., Rhenius, V., Romero, A., Romm, J., Ruddy, K. J., Rudiger, T., Rudolph, A., Ruebner, M., Rutgers, E. J. T., Saloustros, E., Sandler, D. P., Sangrajrang, S., Sawyer, E. J., Schmidt, D. F., Schmutzler, R. K., Schneeweiss, A., Schoemaker, M. J., Schumacher, F., Schurmann, P., Scott, R. J., Scott, C., Seal, S., Seynaeve, C., Shah, M., Sharma, P., Shen, C. Y., Sheng, G., Sherman, M. E., Shrubsole, M. J., Shu, X. O., Smeets, A., Sohn, C., Southey, M. C., Spinelli, J. J., Stegmaier, C., Stewart-Brown, S., Stone, J., Stram, D. O., Surowy, H., Swerdlow, A., Tamimi, R., Taylor, J. A., Tengstrom, M., Teo, S. H., Beth Terry, M., Tessier, D. C., Thanasitthichai, S., Thone, K., Tollenaar, R., Tomlinson, I., Tong, L., Torres, D., Truong, T., Tseng, C. C., Tsugane, S., Ulmer, H. U., Ursin, G., Untch, M., Vachon, C., van Asperen, C. J., Van Den Berg, D., van den Ouweland, A. M. W., van der Kolk, L., van der Luijt, R. B., Vincent, D., Vollenweider, J., Waisfisz, Q., Wang-Gohrke, S., Weinberg, C. R., Wendt, C., Whittemore, A. S., Wildiers, H., Willett, W., Winqvist, R., Wolk, A., Wu, A. H., Xia, L., Yamaji, T., Yang, X. R., Har Yip, C., Yoo, K. Y., Yu, J. C., Zheng, W., Zheng, Y., Zhu, B., Ziogas, A., Ziv, E., Investigators, A., ConFab, A. I., Lakhani, S. R., Antoniou, A. C., Droit, A., Andrulis, I. L., Amos, C. I., Couch, F. J., Pharoah, P. D. P., Chang-Claude, J., Hall, P., Hunter, D. J., Milne, R. L., Garcia-Closas, M., Schmidt, M. K., Chanock, S. J., Dunning, A. M., Edwards, S. L., Bader, G. D., Chenevix-Trench, G., Simard, J., Kraft, P. and Easton, D. F. (2017). **Association analysis identifies 65 new breast cancer risk loci**. Nature *551*, 92-94, doi: 10.1038/nature24284.

Narod, S. A. and Foulkes, W. D. (2004). **BRCA1 and BRCA2: 1994 and beyond**. Nat Rev Cancer *4*, 665-676, doi: 10.1038/nrc1431.

Newman, L. A. and Kaljee, L. M. (2017). **Health Disparities and Triple-Negative Breast Cancer in African American Women: A Review**. JAMA Surg *152*, 485-493, doi: 10.1001/jamasurg.2017.0005.

Niraula, S., Ocana, A., Ennis, M. and Goodwin, P. J. (2012). **Body size and breast cancer prognosis in relation to hormone receptor and menopausal status: a meta-analysis**. Breast Cancer Res Treat *134*, 769-781, doi: 10.1007/s10549-012-2073-x.

Olaya-Contreras, P., Pierre, B., Lazcano-Ponce, E., Villamil-Rodriguez, J. and Posso-Valencia, H. J. (1999). **[Reproductive risk factors associated with breast cancer in Colombian women]**. Rev Saude Publica *33*, 237-245.

Orr, N., Dudbridge, F., Dryden, N., Maguire, S., Novo, D., Perrakis, E., Johnson, N., Ghoussaini, M., Hopper, J. L., Southey, M. C., Apicella, C., Stone, J., Schmidt, M. K., Broeks, A., Van't Veer, L. J., Hogervorst, F. B., Fasching, P. A., Haeberle, L., Ekici, A. B., Beckmann, M. W., Gibson, L., Aitken, Z., Warren, H., Sawyer, E., Tomlinson, I., Kerin, M. J., Miller, N., Burwinkel, B., Marme, F., Schneeweiss, A., Sohn, C., Guenel, P., Truong, T., Cordina-Duverger, E., Sanchez, M., Bojesen, S. E., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Benitez, J., Zamora, M. P., Arias Perez, J. I., Menendez, P., Anton-Culver, H., Neuhausen, S. L., Brenner, H., Dieffenbach, A. K.,

Arndt, V., Stegmaier, C., Hamann, U., Brauch, H., Justenhoven, C., Bruning, T., Ko, Y. D., Nevanlinna, H., Aittomaki, K., Blomqvist, C., Khan, S., Bogdanova, N., Dork, T., Lindblom, A., Margolin, S., Mannermaa, A., Kataja, V., Kosma, V. M., Hartikainen, J. M., Chenevix-Trench, G., Beesley, J., Lambrechts, D., Moisse, M., Floris, G., Beuselinck, B., Chang-Claude, J., Rudolph, A., Seibold, P., Flesch-Janys, D., Radice, P., Peterlongo, P., Peissel, B., Pensotti, V., Couch, F. J., Olson, J. E., Slettedahl, S., Vachon, C., Giles, G. G., Milne, R. L., McLean, C., Haiman, C. A., Henderson, B. E., Schumacher, F., Le Marchand, L., Simard, J., Goldberg, M. S., Labreche, F., Dumont, M., Kristensen, V., Alnaes, G. G., Nord, S., Borresen-Dale, A. L., Zheng, W., Deming-Halverson, S., Shrubsole, M., Long, J., Winqvist, R., Pylkas, K., Jukkola-Vuorinen, A., Grip, M., Andrulis, I. L., Knight, J. A., Glendon, G., Tchatchou, S., Devilee, P., Tollenaar, R. A., Seynaeve, C. M., Van Asperen, C. J., Garcia-Closas, M., Figueroa, J., Chanock, S. J., Lissowska, J., Czene, K., Darabi, H., Eriksson, M., Klevebring, D., Hooning, M. J., Hollestelle, A., van Deurzen, C. H., Kriege, M., Hall, P., Li, J., Liu, J., Humphreys, K., Cox, A., Cross, S. S., Reed, M. W., Pharoah, P. D., Dunning, A. M., Shah, M., Perkins, B. J., Jakubowska, A., Lubinski, J., Jaworska-Bieniek, K., Durda, K., Ashworth, A., Swerdlow, A., Jones, M., Schoemaker, M. J., Meindl, A., Schmutzler, R. K., Olswold, C., Slager, S., Toland, A. E., Yannoukakos, D., Muir, K., Lophatananon, A., Stewart-Brown, S., Siriwanarangsan, P., Matsuo, K., Ito, H., Iwata, H., Ishiguro, J., Wu, A. H., Tseng, C. C., Van Den Berg, D., Stram, D. O., Teo, S. H., Yip, C. H., Kang, P., Ikram, M. K., Shu, X. O., Lu, W., Gao, Y. T., Cai, H., Kang, D., Choi, J. Y., Park, S. K., Noh, D. Y., Hartman, M., Miao, H., Lim, W. Y., Lee, S. C., Sangrajrang, S., Gaborieau, V., Brennan, P., McKay, J., Wu, P. E., Hou, M. F., Yu, J. C., Shen, C. Y., Blot, W., Cai, Q., Signorello, L. B., Luccarini, C., Bayes, C., Ahmed, S., Maranian, M., Healey, C. S., Gonzalez-Neira, A., Pita, G., Alonso, M. R., Alvarez, N., Herrero, D., Tessier, D. C., Vincent, D., Bacot, F., Hunter, D. J., Lindstrom, S., Dennis, J., Michailidou, K., Bolla, M. K., Easton, D. F., dos Santos Silva, I., Fletcher, O., Peto, J., Network, G., kConFab, I. and Australian Ovarian Cancer Study, G. (2015a). **Fine-mapping identifies two additional breast cancer susceptibility loci at 9q31.2**. Hum Mol Genet *24*, 2966-2984, doi: 10.1093/hmg/ddv035.

Orr, N., Dudbridge, F., Dryden, N., Maguire, S., Novo, D., Perrakis, E., Johnson, N., Ghoussaini, M., Hopper, J. L., Southey, M. C., Apicella, C., Stone, J., Schmidt, M. K., Broeks, A., Van't Veer, L. J., Hogervorst, F. B., Fasching, P. A., Haeberle, L., Ekici, A. B., Beckmann, M. W., Gibson, L., Aitken, Z., Warren, H., Sawyer, E., Tomlinson, I., Kerin, M. J., Miller, N., Burwinkel, B., Marme, F., Schneeweiss, A., Sohn, C., Guenel, P., Truong, T., Cordina-Duverger, E., Sanchez, M., Bojesen, S. E., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Benitez, J., Zamora, M. P., Arias Perez, J. I., Menendez, P., Anton-Culver, H., Neuhausen, S. L., Brenner, H., Dieffenbach, A. K., Arndt, V., Stegmaier, C., Hamann, U., Brauch, H., Justenhoven, C., Bruning, T., Ko, Y. D., Nevanlinna, H., Aittomaki, K., Blomqvist, C., Khan, S., Bogdanova, N., Dork, T., Lindblom, A., Margolin, S., Mannermaa, A., Kataja, V., Kosma, V. M., Hartikainen, J. M., Chenevix-Trench, G., Beesley, J., Lambrechts, D., Moisse, M., Floris, G., Beuselinck, B., Chang-Claude, J., Rudolph, A., Seibold, P., Flesch-Janys, D., Radice, P., Peterlongo, P., Peissel, B., Pensotti, V., Couch, F. J., Olson, J. E., Slettedahl, S., Vachon, C., Giles, G. G., Milne, R. L., McLean, C., Haiman, C. A., Henderson, B. E., Schumacher, F., Le, M. L., Simard, J., Goldberg, M. S., Labreche, F., Dumont, M., Kristensen, V., Alnaes, G. G., Nord, S., Borresen-Dale, A. L., Zheng, W., Deming-Halverson, S., Shrubsole, M., Long, J., Winqvist, R., Pylkas, K., Jukkola-Vuorinen, A., Grip, M., Andrulis, I. L., Knight, J. A., Glendon, G., Tchatchou, S., Devilee, P., Tollenaar, R. A., Seynaeve, C. M., van Asperen, C. J., Garcia-Closas, M., Figueroa, J., Chanock, S. J., Lissowska, J., Czene, K., Darabi, H., Eriksson, M., Klevebring, D., Hooning, M. J., Hollestelle, A., van Deurzen, C. H., Kriege, M., Hall, P., Li, J., Liu, J., Humphreys, K., Cox, A., Cross, S. S., Reed, M. W., Pharoah, P. D., Dunning, A. M., Shah, M., Perkins, B. J., Jakubowska, A., Lubinski, J., Jaworska-Bieniek, K., Durda, K., Ashworth, A., Swerdlow, A., Jones, M., Schoemaker, M. J., Meindl, A., Schmutzler, R. K., Olswold, C., Slager, S., Toland, A. E., Yannoukakos, D., Muir, K., Lophatananon, A., Stewart-Brown, S., Siriwanarangsan, P., Matsuo, K., Ito, H., Iwata, H.,

Ishiguro, J., Wu, A. H., Tseng, C. C., Van Den Berg, D., Stram, D. O., Teo, S. H., Yip, C. H., Kang, P., Ikram, M. K., Shu, X. O., Lu, W., Gao, Y. T., Cai, H., Kang, D., Choi, J. Y., Park, S. K., Noh, D. Y., Hartman, M., Miao, H., Lim, W. Y., Lee, S. C., Sangrajrang, S., Gaborieau, V., Brennan, P., McKay, J., Wu, P. E., Hou, M. F., Yu, J. C., Shen, C. Y., Blot, W., Cai, Q., Signorello, L. B., Luccarini, C., Bayes, C., Ahmed, S., Maranian, M., Healey, C. S., Gonzalez-Neira, A., Pita, G., Alonso, M. R., Alvarez, N., Herrero, D., Tessier, D. C., Vincent, D., Bacot, F., Hunter, D. J., Lindstrom, S., Dennis, J., Michailidou, K., Bolla, M. K., Easton, D. F., dos, S. S., I, Fletcher, O. and Peto, J. (2015b). **Fine-mapping identifies two additional breast cancer susceptibility loci at 9q31.2**. Hum. Mol. Genet *24*, 2966-2984, doi: ddv035 [pii];10.1093/hmg/ddv035 [doi].

Ossa, H., Aquino, J., Pereira, R., Ibarra, A., Ossa, R. H., Perez, L. A., Granda, J. D., Lattig, M. C., Groot, H., Fagundes de Carvalho, E. and Gusmao, L. (2016). **Outlining the Ancestry Landscape of Colombian Admixed Populations**. PLoS One *11*, e0164414, doi: 10.1371/journal.pone.0164414.

Pan, Q., Wei, R., Shimizu, I. and Jamoom, E. (2014). **Determining Sufficient Number of Imputations Using Variance of Imputation Variances: Data from 2012 NAMCS Physician Workflow Mail Survey**. Appl Math (Irvine) *5*, 3421-3430, doi: 10.4236/am.2014.521319.

Pardo, C. and Cendales, R. (2009). **[Survival analysis of cervical cancer patients]**. Biomedica *29*, 437-447.

Pharoah, P. D., Antoniou, A. C., Easton, D. F. and Ponder, B. A. (2008). **Polygenes, risk prediction, and targeted prevention of breast cancer**. N Engl J Med *358*, 2796-2803, doi: 10.1056/NEJMsa0708739.

Pharoah, P. D., Day, N. E., Duffy, S., Easton, D. F. and Ponder, B. A. (1997). **Family history and the risk of breast cancer: a systematic review and meta-analysis**. Int J Cancer *71*, 800-809.

Phipps, A. I., Buist, D. S., Malone, K. E., Barlow, W. E., Porter, P. L., Kerlikowske, K. and Li, C. I. (2011). **Family history of breast cancer in first-degree relatives and triple-negative breast cancer risk**. Breast Cancer Res Treat *126*, 671-678, doi: 10.1007/s10549-010-1148-9.

Pineros, M., Cendales, R., Murillo, R., Wiesner, C. and Tovar, S. (2007). **[Pap test coverage and related factors in Colombia, 2005]**. Rev. Salud Publica (Bogota.) *9*, 327-341, doi: S0124-00642007000300002 [pii].

Pineros, M., Hernandez, G. and Bray, F. (2004). **Increasing mortality rates of common malignancies in Colombia: an emerging problem**. Cancer *101*, 2285-2292, doi: 10.1002/cncr.20607.

Prat, A. and Perou, C. M. (2011). **Deconstructing the molecular portraits of breast cancer**. Mol Oncol *5*, 5-23, doi: 10.1016/j.molonc.2010.11.003.

Prentice, R. L., Huang, Y., Hinds, D. A., Peters, U., Pettinger, M., Cox, D. R., Beilharz, E., Chlebowski, R. T., Rossouw, J. E., Caan, B. and Ballinger, D. G. (2009a). **Variation in the FGFR2 gene and the**

effects of postmenopausal hormone therapy on invasive breast cancer. Cancer Epidemiol. Biomarkers Prev *18*, 3079-3085, doi: 1055-9965.EPI-09-0611 [pii];10.1158/1055-9965.EPI-09-0611 [doi].

Prentice, R. L., Huang, Y., Hinds, D. A., Peters, U., Pettinger, M., Cox, D. R., Beilharz, E., Chlebowski, R. T., Rossouw, J. E., Caan, B. and Ballinger, D. G. (2009b). **Variation in the FGFR2 gene and the effects of postmenopausal hormone therapy on invasive breast cancer**. Cancer Epidemiol Biomarkers Prev *18*, 3079-3085, doi: 10.1158/1055-9965.EPI-09-0611.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. and Reich, D. (2006). **Principal components analysis corrects for stratification in genome-wide association studies**. Nat. Genet *38*, 904-909, doi: ng1847 [pii];10.1038/ng1847 [doi].

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000). **Inference of population structure using multilocus genotype data**. Genetics *155*, 945-959.

Puig-Vives, M., Sanchez, M. J., Sanchez-Cantalejo, J., Torrella-Ramos, A., Martos, C., Ardanaz, E., Chirlaque, M. D., Perucha, J., Diaz, J. M., Mateos, A., Machon, M. and Marcos-Gragera, R. (2013). **Distribution and prognosis of molecular breast cancer subtypes defined by immunohistochemical biomarkers in a Spanish population-based study**. Gynecol Oncol *130*, 609-614, doi: 10.1016/j.ygyno.2013.05.039.

Renehan, A. G., Tyson, M., Egger, M., Heller, R. F. and Zwahlen, M. (2008). **Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies**. Lancet *371*, 569-578, doi: 10.1016/S0140-6736(08)60269-X.

Ruiz-Linares, A., Adhikari, K., Acuna-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., Fuentes, M., Pizarro, M., Everardo, P., de, A. F., Gomez-Valdes, J., Leon-Mimila, P., Hunemeier, T., Ramallo, V., Silva de Cerqueira, C. C., Burley, M. W., Konca, E., de Oliveira, M. Z., Veronez, M. R., Rubio-Codina, M., Attanasio, O., Gibbon, S., Ray, N., Gallo, C., Poletti, G., Rosique, J., Schuler-Faccini, L., Salzano, F. M., Bortolini, M. C., Canizales-Quinteros, S., Rothhammer, F., Bedoya, G., Balding, D. and Gonzalez-Jose, R. (2014). **Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals**. PLoS. Genet *10*, e1004572, doi: 10.1371/journal.pgen.1004572 [doi];PGENETICS-D-13-01796 [pii].

Schafer, J. L. (1999). **Multiple imputation: a primer**. Stat Methods Med Res *8*, 3-15, doi: 10.1177/096228029900800102.

Serrano-Gomez, S. J., Sanabria-Salas, M. C., Hernandez-Suarez, G., Garcia, O., Silva, C., Romero, A., Mejia, J. C., Miele, L., Fejerman, L. and Zabaleta, J. (2016). **High prevalence of luminal B breast cancer intrinsic subtype in Colombian women**. Carcinogenesis *37*, 669-676, doi: 10.1093/carcin/bgw043.

Shiovitz, S. and Korde, L. A. (2015). **Genetics of breast cancer: a topic in evolution**. Ann Oncol *26*, 1291-1299, doi: 10.1093/annonc/mdv022.

Siddiq, A., Couch, F. J., Chen, G. K., Lindstrom, S., Eccles, D., Millikan, R. C., Michailidou, K., Stram, D. O., Beckmann, L., Rhie, S. K., Ambrosone, C. B., Aittomaki, K., Amiano, P., Apicella, C., Baglietto, L., Bandera, E. V., Beckmann, M. W., Berg, C. D., Bernstein, L., Blomqvist, C., Brauch, H., Brinton, L., Bui, Q. M., Buring, J. E., Buys, S. S., Campa, D., Carpenter, J. E., Chasman, D. I., Chang-Claude, J., Chen, C., Clavel-Chapelon, F., Cox, A., Cross, S. S., Czene, K., Deming, S. L., Diasio, R. B., Diver, W. R., Dunning, A. M., Durcan, L., Ekici, A. B., Fasching, P. A., Feigelson, H. S., Fejerman, L., Figueroa, J. D., Fletcher, O., Flesch-Janys, D., Gaudet, M. M., Gerty, S. M., Rodriguez-Gil, J. L., Giles, G. G., van Gils, C. H., Godwin, A. K., Graham, N., Greco, D., Hall, P., Hankinson, S. E., Hartmann, A., Hein, R., Heinz, J., Hoover, R. N., Hopper, J. L., Hu, J. J., Huntsman, S., Ingles, S. A., Irwanto, A., Isaacs, C., Jacobs, K. B., John, E. M., Justenhoven, C., Kaaks, R., Kolonel, L. N., Coetzee, G. A., Lathrop, M., Le, M. L., Lee, A. M., Lee, I. M., Lesnick, T., Lichtner, P., Liu, J., Lund, E., Makalic, E., Martin, N. G., McLean, C. A., Meijers-Heijboer, H., Meindl, A., Miron, P., Monroe, K. R., Montgomery, G. W., Muller-Myhsok, B., Nickels, S., Nyante, S. J., Olswold, C., Overvad, K., Palli, D., Park, D. J., Palmer, J. R., Pathak, H., Peto, J., Pharoah, P., Rahman, N., Rivadeneira, F., Schmidt, D. F., Schmutzler, R. K., Slager, S., Southey, M. C., Stevens, K. N., Sinn, H. P., Press, M. F., Ross, E., Riboli, E., Ridker, P. M., Schumacher, F. R., Severi, G., dos, S. S., I, Stone, J., Sund, M., Tapper, W. J., Thun, M. J., Travis, R. C., Turnbull, C., Uitterlinden, A. G., Waisfisz, Q., Wang, X., Wang, Z., Weaver, J., Schulz-Wendtland, R., Wilkens, L. R., Van Den Berg, D., Zheng, W., Ziegler, R. G., Ziv, E., Nevanlinna, H., Easton, D. F., Hunter, D. J., Henderson, B. E., Chanock, S. J., Garcia-Closas, M., Kraft, P., Haiman, C. A. and Vachon, C. M. (2012). **A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11**. Hum. Mol. Genet *21*, 5373-5384, doi: dds381 [pii];10.1093/hmg/dds381 [doi].

Singletary, S. E. (2003). **Rating the risk factors for breast cancer**. Ann Surg *237*, 474-482, doi: 10.1097/01.SLA.0000059969.64262.87.

Sinn, H. P. and Kreipe, H. (2013). **A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition**. Breast Care (Basel) *8*, 149-154, doi: 10.1159/000350774.

Slattery, M. L., Baumgartner, K. B., Giuliano, A. R., Byers, T., Herrick, J. S. and Wolff, R. K. (2011). **Replication of five GWAS-identified loci and breast cancer risk among Hispanic and non-Hispanic white women living in the Southwestern United States**. Breast Cancer Res. Treat *129*, 531-539, doi: 10.1007/s10549-011-1498-y [doi].

Slattery, M. L., John, E. M., Torres-Mejia, G., Lundgreen, A., Herrick, J. S., Baumgartner, K. B., Hines, L. M., Stern, M. C. and Wolff, R. K. (2012). **Genetic variation in genes involved in hormones, inflammation and energetic factors and breast cancer risk in an admixed population**. Carcinogenesis *33*, 1512-1521, doi: bgs163 [pii];10.1093/carcin/bgs163 [doi].

So, H. C., Gui, A. H., Cherny, S. S. and Sham, P. C. (2011). **Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases**. Genet Epidemiol *35*, 310-317, doi: 10.1002/gepi.20579.

Society, A. C. (2014). **What are the risk factors for breast cancer?**

Society, A. C. (2017). **Breast Cancer Risk Factors You Cannot Change**. URL: https://www.cancer.org/cancer/breast-cancer/risk-and-prevention/breast-cancer-risk-factors-you-cannot-change.html [as of29.10.2018].

Stacey, S. N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S. A., Masson, G., Jakobsdottir, M., Thorlacius, S., Helgason, A., Aben, K. K., Strobbe, L. J., Albers-Akkers, M. T., Swinkels, D. W., Henderson, B. E., Kolonel, L. N., Le, M. L., Millastre, E., Andres, R., Godino, J., Garcia-Prats, M. D., Polo, E., Tres, A., Mouy, M., Saemundsdottir, J., Backman, V. M., Gudmundsson, L., Kristjansson, K., Bergthorsson, J. T., Kostic, J., Frigge, M. L., Geller, F., Gudbjartsson, D., Sigurdsson, H., Jonsdottir, T., Hrafnkelsson, J., Johannsson, J., Sveinsson, T., Myrdal, G., Grimsson, H. N., Jonsson, T., von, H. S., Werelius, B., Margolin, S., Lindblom, A., Mayordomo, J. I., Haiman, C. A., Kiemeney, L. A., Johannsson, O. T., Gulcher, J. R., Thorsteinsdottir, U., Kong, A. and Stefansson, K. (2007). **Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer**. Nat. Genet *39*, 865-869, doi: ng2064 [pii];10.1038/ng2064 [doi].

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M. and Carpenter, J. R. (2009). **Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls**. BMJ *338*, b2393, doi: 10.1136/bmj.b2393.

Stevens, K. N., Vachon, C. M. and Couch, F. J. (2013). **Genetic susceptibility to triple-negative breast cancer**. Cancer Res *73*, 2025-2030, doi: 0008-5472.CAN-12-1699 [pii];10.1158/0008-5472.CAN-12-1699 [doi].

Tavassoli FA, D. P. (2003). **World Health Organization Classification of Tumors: Pathology and Genetics of Tumors of the Breast and Female Genital Organs.**, IARC Press, Lyon, France.

Thomas, G., Jacobs, K. B., Kraft, P., Yeager, M., Wacholder, S., Cox, D. G., Hankinson, S. E., Hutchinson, A., Wang, Z., Yu, K., Chatterjee, N., Garcia-Closas, M., Gonzalez-Bosquet, J., Prokunina-Olsson, L., Orr, N., Willett, W. C., Colditz, G. A., Ziegler, R. G., Berg, C. D., Buys, S. S., McCarty, C. A., Feigelson, H. S., Calle, E. E., Thun, M. J., Diver, R., Prentice, R., Jackson, R., Kooperberg, C., Chlebowski, R., Lissowska, J., Peplonska, B., Brinton, L. A., Sigurdson, A., Doody, M., Bhatti, P., Alexander, B. H., Buring, J., Lee, I. M., Vatten, L. J., Hveem, K., Kumle, M., Hayes, R. B., Tucker, M., Gerhard, D. S., Fraumeni, J. F., Jr., Hoover, R. N., Chanock, S. J. and Hunter, D. J. (2009). **A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1)**. Nat. Genet *41*, 579-584, doi: ng.353 [pii];10.1038/ng.353 [doi].

Thorslund, T. and West, S. C. (2007). **BRCA2: a universal recombinase regulator**. Oncogene *26*, 7720-7730, doi: 10.1038/sj.onc.1210870.

Torkamani, A., Wineinger, N. E. and Topol, E. J. (2018). **The personal and clinical utility of polygenic risk scores**. Nat Rev Genet *19*, 581-590, doi: 10.1038/s41576-018-0018-x.

Torres, D., Bermejo, J. L., Rashid, M. U., Briceno, I., Gil, F., Beltran, A., Ariza, V. and Hamann, U. (2017). **Prevalence and Penetrance of BRCA1 and BRCA2 Germline Mutations in Colombian Breast Cancer Patients**. Sci Rep *7*, 4713, doi: 10.1038/s41598-017-05056-y.

Torres, D., Rashid, M. U., Gil, F., Umana, A., Ramelli, G., Robledo, J. F., Tawil, M., Torregrosa, L., Briceno, I. and Hamann, U. (2007). **High proportion of BRCA1/2 founder mutations in Hispanic breast/ovarian cancer families from Colombia**. Breast Cancer Res. Treat *103*, 225-232, doi: 10.1007/s10549-006-9370-1 [doi].

Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M., Seal, S., Ghoussaini, M., Hines, S., Healey, C. S., Hughes, D., Warren-Perry, M., Tapper, W., Eccles, D., Evans, D. G., Hooning, M., Schutte, M., van den Ouweland, A., Houlston, R., Ross, G., Langford, C., Pharoah, P. D., Stratton, M. R., Dunning, A. M., Rahman, N. and Easton, D. F. (2010). **Genome-wide association study identifies five new breast cancer susceptibility loci**. Nat. Genet *42*, 504-507, doi: ng.586 [pii];10.1038/ng.586 [doi].

Udler, M. S., Ahmed, S., Healey, C. S., Meyer, K., Struewing, J., Maranian, M., Kwon, E. M., Zhang, J., Tyrer, J., Karlins, E., Platte, R., Kalmyrzaev, B., Dicks, E., Field, H., Maia, A. T., Prathalingam, R., Teschendorff, A., McArthur, S., Doody, D. R., Luben, R., Caldas, C., Bernstein, L., Kolonel, L. K., Henderson, B. E., Wu, A. H., Le Marchand, L., Ursin, G., Press, M. F., Lindblom, A., Margolin, S., Shen, C. Y., Yang, S. L., Hsiung, C. N., Kang, D., Yoo, K. Y., Noh, D. Y., Ahn, S. H., Malone, K. E., Haiman, C. A., Pharoah, P. D., Ponder, B. A., Ostrander, E. A., Easton, D. F. and Dunning, A. M. (2010a). **Fine scale mapping of the breast cancer 16q12 locus**. Hum Mol Genet *19*, 2507-2515, doi: 10.1093/hmg/ddq122.

Udler, M. S., Ahmed, S., Healey, C. S., Meyer, K., Struewing, J., Maranian, M., Kwon, E. M., Zhang, J., Tyrer, J., Karlins, E., Platte, R., Kalmyrzaev, B., Dicks, E., Field, H., Maia, A. T., Prathalingam, R., Teschendorff, A., McArthur, S., Doody, D. R., Luben, R., Caldas, C., Bernstein, L., Kolonel, L. K., Henderson, B. E., Wu, A. H., Le, M. L., Ursin, G., Press, M. F., Lindblom, A., Margolin, S., Shen, C. Y., Yang, S. L., Hsiung, C. N., Kang, D., Yoo, K. Y., Noh, D. Y., Ahn, S. H., Malone, K. E., Haiman, C. A., Pharoah, P. D., Ponder, B. A., Ostrander, E. A., Easton, D. F. and Dunning, A. M. (2010b). **Fine scale mapping of the breast cancer 16q12 locus**. Hum. Mol. Genet *19*, 2507-2515, doi: ddq122 [pii];10.1093/hmg/ddq122 [doi].

Viale, G. (2012). **The current state of breast cancer classification**. Ann Oncol *23 Suppl 10*, x207-210, doi: 10.1093/annonc/mds326.

Wacholder, S., Hartge, P., Prentice, R., Garcia-Closas, M., Feigelson, H. S., Diver, W. R., Thun, M. J., Cox, D. G., Hankinson, S. E., Kraft, P., Rosner, B., Berg, C. D., Brinton, L. A., Lissowska, J., Sherman, M. E., Chlebowski, R., Kooperberg, C., Jackson, R. D., Buckman, D. W., Hui, P., Pfeiffer, R., Jacobs, K. B., Thomas, G. D., Hoover, R. N., Gail, M. H., Chanock, S. J. and Hunter, D. J. (2010). **Performance of common genetic variants in breast-cancer risk models**. N Engl J Med *362*, 986-993, doi: 10.1056/NEJMoa0907727.

Weigelt, B., Horlings, H. M., Kreike, B., Hayes, M. M., Hauptmann, M., Wessels, L. F., de Jong, D., Van de Vijver, M. J., Van't Veer, L. J. and Peterse, J. L. (2008). **Refinement of breast cancer**

classification by molecular characterization of histological special types. J Pathol *216*, 141-150, doi: 10.1002/path.2407.

Wen, W., Shu, X. O., Guo, X., Cai, Q., Long, J., Bolla, M. K., Michailidou, K., Dennis, J., Wang, Q., Gao, Y. T., Zheng, Y., Dunning, A. M., Garcia-Closas, M., Brennan, P., Chen, S. T., Choi, J. Y., Hartman, M., Ito, H., Lophatananon, A., Matsuo, K., Miao, H., Muir, K., Sangrajrang, S., Shen, C. Y., Teo, S. H., Tseng, C. C., Wu, A. H., Yip, C. H., Simard, J., Pharoah, P. D., Hall, P., Kang, D., Xiang, Y., Easton, D. F. and Zheng, W. (2016). **Prediction of breast cancer risk based on common genetic variants in women of East Asian ancestry**. Breast Cancer Res *18*, 124, doi: 10.1186/s13058-016-0786-1 [doi];10.1186/s13058-016-0786-1 [pii].

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C. and Micklem, G. (1995). **Identification of the breast cancer susceptibility gene BRCA2**. Nature *378*, 789-792, doi: 10.1038/378789a0 [doi].

Wu, C., DeWan, A., Hoh, J. and Wang, Z. (2011). **A comparison of association methods correcting for population stratification in case-control studies**. Ann Hum Genet *75*, 418-427, doi: 10.1111/j.1469-1809.2010.00639.x.

Yang, X. R., Sherman, M. E., Rimm, D. L., Lissowska, J., Brinton, L. A., Peplonska, B., Hewitt, S. M., Anderson, W. F., Szeszenia-Dabrowska, N., Bardin-Mikolajczak, A., Zatonski, W., Cartun, R., Mandich, D., Rymkiewicz, G., Ligaj, M., Lukaszek, S., Kordek, R. and Garcia-Closas, M. (2007). **Differences in risk factors for breast cancer molecular subtypes in a population-based study**. Cancer Epidemiol Biomarkers Prev *16*, 439-443, doi: 10.1158/1055-9965.EPI-06-0806.

Yip, C. H. (2009). **Breast cancer in Asia**. Methods Mol Biol *471*, 51-64, doi: 10.1007/978-1-59745-416-2_3.

Zheng, W., Long, J., Gao, Y. T., Li, C., Zheng, Y., Xiang, Y. B., Wen, W., Levy, S., Deming, S. L., Haines, J. L., Gu, K., Fair, A. M., Cai, Q., Lu, W. and Shu, X. O. (2009). **Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1**. Nat. Genet *41*, 324-328, doi: ng.318 [pii];10.1038/ng.318 [doi].

Zheng, W., Zhang, B., Cai, Q., Sung, H., Michailidou, K., Shi, J., Choi, J. Y., Long, J., Dennis, J., Humphreys, M. K., Wang, Q., Lu, W., Gao, Y. T., Li, C., Cai, H., Park, S. K., Yoo, K. Y., Noh, D. Y., Han, W., Dunning, A. M., Benitez, J., Vincent, D., Bacot, F., Tessier, D., Kim, S. W., Lee, M. H., Lee, J. W., Lee, J. Y., Xiang, Y. B., Zheng, Y., Wang, W., Ji, B. T., Matsuo, K., Ito, H., Iwata, H., Tanaka, H., Wu, A. H., Tseng, C. C., Van Den Berg, D., Stram, D. O., Teo, S. H., Yip, C. H., Kang, I. N., Wong, T. Y., Shen, C. Y., Yu, J. C., Huang, C. S., Hou, M. F., Hartman, M., Miao, H., Lee, S. C., Putti, T. C., Muir, K., Lophatananon, A., Stewart-Brown, S., Siriwanarangsan, P., Sangrajrang, S., Shen, H., Chen, K., Wu, P. E., Ren, Z., Haiman, C. A., Sueta, A., Kim, M. K., Khoo, U. S., Iwasaki, M., Pharoah, P. D., Wen, W., Hall, P., Shu, X. O., Easton, D. F. and Kang, D. (2013). **Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls**. Hum. Mol. Genet *22*, 2539-2550, doi: ddt089 [pii];10.1093/hmg/ddt089 [doi].

Ziv, E., John, E. M., Choudhry, S., Kho, J., Lorizio, W., Perez-Stable, E. J. and Burchard, E. G. (2006). **Genetic ancestry and risk factors for breast cancer among Latinas in the San Francisco Bay Area**. Cancer Epidemiol. Biomarkers Prev *15*, 1878-1885, doi: 15/10/1878 [pii];10.1158/1055-9965.EPI-06-0092 [doi].

9. Appendix

Table S1: Primer sequences of the 78 SNPs under investigation

#	SNP-ID	Forward primer 5´---- 3´	Reverse primer 5´---- 3´	Extension primer 5´---- 3´
1	rs11249433	ACGTTGGATGGGAAACATGGAATCCAAAAC	ACGTTGGATGTGTGCTAAGGAGAAGCAACC	ATCAGGGCTGGGTTTAA
2	rs11552449	LC480 // Thermo Fisher assay (C__25619066_10)	LC480 // Thermo Fisher assay (C__25619066_10)	-
3	rs616488	ACGTTGGATGTAGGGACAAACTAGCTCCTG	ACGTTGGATGTCTTGGTGCTCTGAACAAGG	GGGACTGGGGTACAGATGAGTG
4	rs6678914	ACGTTGGATGGTGTCATCTTGTGTGTTTGG	ACGTTGGATGCTCCCCACTTTAGATGGCAC	TTTTAGATGGCACCTTCTTA
5	rs4245739	ACGTTGGATGAGCATAATGGTAGTACGAAC	ACGTTGGATGAACATTCTCTGACAGGTTGG	TTCAAATAATGTGGTAAGTGAAC
6	rs12710696	ACGTTGGATGTCCGTCTACAGGAATGATGC	ACGTTGGATGAAGAAGGCGCCTCACATAGT	GATTAATCATCAGGCCC
7	rs4849887	ACGTTGGATGTTAAACCCTGGGTGTGATGC	ACGTTGGATGTGGTTGAGGGGATTACCTAC	GGGGATTCCAGCTGTGATCTGGCAATG
8	rs1550623	ACGTTGGATGGTTAAAGAGCTTTGAATTCCC	ACGTTGGATGGGAATTCTTTGTGCAAGGTG	AGTTATAGAGAAACACAGAATTCTTA
9	rs2016394	ACGTTGGATGTCTGTTCAGGAATCTTCCCG	ACGTTGGATGTCTCTCCCCAAAGAGAGTCC	TTACAGAGAGTCCTGGAAGGG
10	rs16857609	ACGTTGGATGCTCAAGAAAGGGATGCCAAC	ACGTTGGATGCAGAGACCGGAAGGTATTCT	TGTCTGGGAGACTTC
11	rs4442975	ACGTTGGATGAGTGCAGGTCACTAATATAG	ACGTTGGATGGTATCCTACAGGTATGGAGC	TGATTGCAGGTCACTAATATAGTCTATT
12	rs4973768	ACGTTGGATGCACTGTCTCTCAATGAATGC	ACGTTGGATGACTCCATTTAAGAGCAAAGG	TGAGAGCAAAGGTAACTCATGTTTA
13	rs12493607	ACGTTGGATGATGTATATGACCTCAAGGGC	ACGTTGGATGGACTCCCTATTAGTTGACCC	GGAAGGCTACCTGGAATTGTTAAG
14	rs6762644	ACGTTGGATGGGATGGGATGTAAAAGAGGC	ACGTTGGATGATTCCTGAGGTCCCAAATCG	AAATCGCATCCTGCC
15	rs9790517	ACGTTGGATGTACCATGCTAAGCACTTTCC	ACGTTGGATGTAACCCCATGTCTTGGTTTC	GGGGTCTTCCTTAGTAAAGTGAG
16	rs6828523	ACGTTGGATGGGCATGGATTTAATAACGAGG	ACGTTGGATGTCTCCATACCCACTTTCTCC	CTTCATAACGAGGTGTTAATATTGGTT
17	rs10941679	ACGTTGGATGGCCAGTAAAATGTGGGATGC	ACGTTGGATGCTTCTTTGAGCAATGATCGG	ATTCTGCTGTGTTCTTTCCA
18	rs10069690	ACGTTGGATGCTGTTTGAAACGGGTTCCTG	ACGTTGGATGACCCCGTCATCTGAGGAGA	TCACCCGGGATCCTCATGCCA
19	rs2736108	LC480 // Thermo Fisher assay (C__26414910_20)	LC480 // Thermo Fisher assay (C__26414910_20)	-
20	rs3215401	ACGTTGGATGGCATTTCAGTGTTTGCCGAC	ACGTTGGATGCTCCGGGTTGCTCAAGTTTG	CCGAAGTTTCTCGCCCC
21	rs2242652	ACGTTGGATGAGGCTCTGAGGACCACAAGA	ACGTTGGATGACAGCAGGACACGGATCCAG	TCTTAGGACCACAAGAAGCAGC
22	rs889312	ACGTTGGATGGAGATGATCTCTGAGATGCC	ACGTTGGATGGACACAGATTTATGGGAAGG	GGCCTGCTGGAGAAAGG
23	rs1353747	ACGTTGGATGTTACAATGAAGCTGGGAGTC	ACGTTGGATGACTTGAGGCTCAGGCTTTGA	CCTCGTCCCCTCTGAGGTTCGGC
24	rs10472076	ACGTTGGATGCCTCATGTCCAAGCTCTTTG	ACGTTGGATGAGGTCCAGCCGCCAAGTTAA	AGCAGCTGTCCAGACATC
25	rs1432679	ACGTTGGATGACTTCGCAATTGCCACCAAG	ACGTTGGATGGTGAGGGCTACATCATGCAG	AAAAGTGGGCTGGTCA
26	rs204247	ACGTTGGATGTCTGCTTCAGCTGCAATTAG	ACGTTGGATGCACAGACAATGCTACCTCTC	TTGAAAATGCTACCTCTCTTTAAATAA
27	rs11242675	ACGTTGGATGTCTAGAGGAGACTACAACGC	ACGTTGGATGAGCAAACAGTGTCACCGTCC	TTGCTGAAGGGTCTAGATAC
28	rs17530068	ACGTTGGATGCTTATAACCAGTGTGGAAAG	ACGTTGGATGTCCTGTGTTGTGTGTTGACG	ACTCAGAGAATTTGTTTAGCCCCAT
29	rs2046210	ACGTTGGATGTGAAACCATCAGGGTGCCTC	ACGTTGGATGCCTCACACATACATACAGTC	CCCGCACACATACATACAGTCACATAC
30	rs720475	ACGTTGGATGCAAACAACTACAAGGAACCC	ACGTTGGATGCAGTGAGCAAGTTAACAGCG	AAAGTGAAGTAGGAGAGCC

31	rs9693444	ACGTTGGATGCAAAGGCTATGCTTCTCCTC	ACGTTGGATGGGCTTAGAAGACTTTGCCAG	GCTAGGCTTCTCCTCTTTGCCCA
32	rs6472903	ACGTTGGATGTGAAAAGACAAGCTGCCAGG	ACGTTGGATGGACCTCCAAAAACTGTCTTC	CAATCAGTCAATATTTACTGAATAC
33	rs13281615	ACGTTGGATGCCTGGAATCTAGGGATGTAG	ACGTTGGATGAACCCCCTACTCAGAATATC	GAATATCTGCGTTCTGC
34	rs11780156	ACGTTGGATGACTGGTCCCAGATAAGGTGA	ACGTTGGATGCTCACTTACGAAGGTATGGG	CATGGAACACAGCTAC
35	rs1011970	ACGTTGGATGAGGGAAGATACAGGTGGAAC	ACGTTGGATGCTCTAGAACTGATAGGGAGC	GAGTGTGGAACTGGGCCAGTGTTT
36	rs10816625	ACGTTGGATGAGGTTACATTCACAGAGGTC	ACGTTGGATGCACGTGACATCACTTCAAAC	TTTACAAATAAACGGTCTATTTATTCA
37	rs676256	ACGTTGGATGGGATACAACATGGGTTTGCG	ACGTTGGATGAAGAGCTAGAGCAGCACTTC	ATGGGTTTGCGGGCACA
38	rs865686	ACGTTGGATGGTCTCATTAGCCAGGAAACC	ACGTTGGATGCGAGACTGTGGTATTCTAGG	GACCGGAAACCAAGGTTTCCAAC
39	rs10759243	ACGTTGGATGTGGGTAAACACAGCAAAATG	ACGTTGGATGCCCAGCCCCTTTTTATTTTC	GGGTAGATAATGTGTGACAGTGTATA
40	rs11814448	ACGTTGGATGTTGAGCCCCAGAAGCTAATT	ACGTTGGATGAAGTATGTGAATGTCTCACC	GCTAATTTTCTTTTCAGTTATTATG
41	rs7072776	ACGTTGGATGGCAGTACTCACACTTGGTAG	ACGTTGGATGAGACCACTAAGTAGATTGGG	GGAACACTAAGTAGATTGGGCTTCTC
42	rs10995190	ACGTTGGATGCAATGGTTGTGTCCAAGTGC	ACGTTGGATGTGGCTCATGACTTGAACCAC	GGGTGTTGTGTCCAAGTGCATATTGA
43	rs704010	ACGTTGGATGTACTGCCACGTCTTACAACC	ACGTTGGATGCTGGGCCTTATTATCTCTGC	CTGCGACCTGACCTGAAAATAGC
44	rs7904519	ACGTTGGATGAAGAAGGGTGGTTAGACAGG	ACGTTGGATGGCAGAGCGTGGACATTTTTC	CCTAGCTGAATGCCGACC
45	rs11199914	ACGTTGGATGACAGTGGGTGGAATAAGCAG	ACGTTGGATGGGAGCTGCGTAGATATCAAG	CAATACCACATTCAAGTCCTGT
46	rs35054928	ACGTTGGATGCTCTCACAATCCTCCAAAGC	ACGTTGGATGTCCCAGAAAGCCTACATTCG	GGGAGAAAGCCTACATTCGTGGGAGCCG
47	rs2981579	ACGTTGGATGGGTTGAAGGTCACTTGCTTG	ACGTTGGATGTTCATATTTCCACCTGCCCG	ATGGATACGACCTCTG
48	rs2981578	ACGTTGGATGGTGAGAGGTAGTGACTCTTC	ACGTTGGATGCACAGTTAACCTTTCTTCCC	TTCCCTGCTCCAAAC
49	rs3750817	ACGTTGGATGGGGTTGTCTTGAGACAGTTG	ACGTTGGATGGCTGGTTTGGCAACGTAAAG	GCAAAAGGGTTTGGAAA
50	rs11200014	ACGTTGGATGTGTTTTCGGCTGTTCATGAC	ACGTTGGATGAATACACGTGTTGGGACCAG	GGGACCAGAGAGAAAA
51	rs2420946	ACGTTGGATGCTCATGGAAACTATAAACCC	ACGTTGGATGCTGCTCAACCTGGGATCTGT	GGTGCGGGATCTGTGGATGC
52	rs3817198	ACGTTGGATGTCTCACCTGATACCAGATTC	ACGTTGGATGCTTGAGCTTCCCTAGTGGAG	GGGCTGGCTGACTCTAGTGAAATGAGC
53	rs614367	ACGTTGGATGTCCTTGGCTTTTTCCTCCAG	ACGTTGGATGGGGTTTGACAGGCTTCTTGG	GCTTCTCTGCAACTC
54	rs3903072	ACGTTGGATGCATTCATCCACCATATCCCG	ACGTTGGATGCTTCTTTAGTCCTCTTCCAG	AAGGAGTTCTGTTCTGTGCTG
55	rs554219	ACGTTGGATGAGGTGCTGGGTTGACTGTG	ACGTTGGATGTGTTGTGTGATTCCACTCCC	CGAGAGCAGGGAAATCCTCAC
56	rs11820646	ACGTTGGATGTGCTCCTGGAAAGAAGGTAG	ACGTTGGATGGTCATATGACTCTGGAGGTG	GTGGGACTTGAAATCAAAAGACA
57	rs12422552	ACGTTGGATGTGAGGGTAACATCGCGGTAG	ACGTTGGATGGACACTGATGGAAAATCGTC	AGAGCATCGCGGTAGTCAGTCCA
58	rs17356907	ACGTTGGATGTGCTGCCAGATGATCACTTC	ACGTTGGATGTTGTCAGATCCAGCAGAGAC	TAATGGGGATTAGATGGT
59	rs1292011	ACGTTGGATGACTCCCAAGATCCTAAGTCC	ACGTTGGATGCAAGTACACCGCTTGCCTTC	CATCCTGCTGCCTCTGA
60	rs2236007	ACGTTGGATGAGGCTGGGCGTGTGGATGT	ACGTTGGATGAAAGACAGTGTCCCTGAGGC	GGATGTGCAGCGTCT
61	rs999737	ACGTTGGATGGGTCCCGTTCACATGATATG	ACGTTGGATGCTTTTTCAGGTCCTGGAATG	TCCGGTCACATGATATGAATGGGGC
62	rs2588809	ACGTTGGATGTAAGGTGAACAGTGAGCCAG	ACGTTGGATGCTTGCCTCAACTTGACTTGG	GGAGACTGGTGAATGAGATTAGGT
63	rs941764	ACGTTGGATGTCCAACAGTCTGGCAGCATC	ACGTTGGATGTGTGTGAGACTTCCTTGTGC	GGGGTTCGGAATGAGCCTGAC
64	rs17817449	ACGTTGGATGCCCTTTGTGTTTCAGCTTGG	ACGTTGGATGGAGTGCACCAAATTCAAACC	CTTAAGCTTGGCACACAGAAAC
65	rs11075995	ACGTTGGATGTGCTTTGAGTATACTGGGGC	ACGTTGGATGGCCATGAGCAGCAACTCTAA	CTCTCACTATACTGGTTACTTTCT
66	rs3803662	ACGTTGGATGTTTTCTCTCCTTAATGCCTC	ACGTTGGATGGCTGAAGACCCAGTACTTTC	CAGTACACAGTTTTATTCTTCGCTAAG
67	rs4784227	ACGTTGGATGTTAACAGCCAACTCTTTGGG	ACGTTGGATGGAAATTGGTCATGATGGGAG	TGGGAGTATTTACATCACAATAATC
68	rs13329835	ACGTTGGATGCCTAGCAACACACCAGTGAG	ACGTTGGATGGAGCTTTGGGTATGTCATGG	ATGTCATGGATCTCTCTG
69	rs6504950	ACGTTGGATGCTGAATCACTCCTTGCCAAC	ACGTTGGATGCCAGGGTTTGTCTACCAAAG	TCTACCAAAGGCAGGATAC

70	rs1436904	ACGTTGGATGTGCAAGACTATGTGTGCCTC	ACGTTGGATGAGACAGGAGACAGATTCTGC	CCCTGTGTCTCATTCCTCA
71	rs527616	ACGTTGGATGTTACACGAGACTGAGCCAAC	ACGTTGGATGATGGAAATGCCCCTTAGGAC	AGGACAAGTCTAAACTAGG
72	rs8170	ACGTTGGATGAGTTTCGCCATGCAGTTGTG	ACGTTGGATGTGGATACCAAGGGTACCAG	GATCCCAAGGGTACCAGCTACAA
73	rs2363956	ACGTTGGATGAGCCATGCAGAGGTGACAAC	ACGTTGGATGGTTTGTCCACAGTTTCAAGG	CCTGGGTCAGCCTCC
74	rs4808801	LGC (KASP allelic discrimination)	LGC (KASP allelic discrimination)	-
75	rs3760982	ACGTTGGATGTACACCAGCACACAGTAACG	ACGTTGGATGATGTGGTCGTTTGTGGAAGG	CAAGTAACGCAAAACACATA
76	rs2284378	ACGTTGGATGAGCATCGTTTGGTCAGTGTG	ACGTTGGATGCCCCCCCAGATGATTTTATG	TGGGCTGGACTAAGG
77	rs2823093	ACGTTGGATGTCAAGTCAAGGAACAAAGGG	ACGTTGGATGAAGTGCTGGAGGAAGTCACG	GACGGGTTCAGGAGACA
78	rs6001930	ACGTTGGATGAGCAGAAATTGGAACCCAGC	ACGTTGGATGGAACTGGTGTGTTTTATGAG	GTATAGCCTGAATCTTCAA

Table S2: Associations between 78 previously known risk SNPs in Europeans with breast cancer risk in Colombians

			EUROPEANS			COLOMBIANS										
SNP ID	**Locus**	**Gene/nearest gene(s)***	**Alleles Major/Minor**	**MAF**	**OR**	**AF[1]**	**95% CI**		**OR[1]**	**95% CI**		**P-value[2]**	**$OR_{adj.}$[3]**	**95% CI**		**P-value[2]**	**Reference EUROPEANS**
rs11249433	*1p11.2*	*EMBP1*	A/G	0.40	1.09	**0.27**	0.24	0.29	1.09	0.95	1.25	0.21	1.02	0.87	1.19	0.84	(Michailidou *et al.*, 2013)
rs11552449	*1p13.2*	*DCLRE1B*	C/T	0.17	1.07	**0.39**	0.36	0.42	**0.91**	0.80	1.03	0.13	0.92	0.79	1.07	0.27	(Michailidou *et al.*, 2013)
rs616488	*1p36.22*	*PEX14*	A/G	0.33	0.94	**0.40**	0.37	0.43	0.92	0.81	1.05	0.24	0.95	0.81	1.11	0.55	(Michailidou *et al.*, 2013)
rs6678914	*1q32.1*	*LGR6*	G/A	0.41	0.90[4]	**0.30**	0.27	0.32	**1.09**	0.95	1.25	0.21	1.10	0.93	1.29	0.26	(Garcia-Closas *et al.*, 2013b)
rs4245739	*1q32.1*	*MDM4*	A/C	0.26	1.14[4]	0.25	0.22	0.27	1.04	0.90	1.20	0.62	1.00	0.84	1.19	0.98	(Garcia-Closas *et al.*, 2013b)
rs12710696	*2p24.1*	*OSR1*	G/A	0.36	1.10[4]	**0.32**	0.29	0.35	**0.96**	0.84	1.10	0.56	0.96	0.82	1.12	0.59	(Garcia-Closas *et al.*, 2013b)
rs4849887	*2q14.1*	*INHBB*	C/T	0.10	0.91	0.11	0.09	0.13	0.85	0.69	1.04	0.11	0.88	0.69	1.12	0.29	(Michailidou *et al.*, 2013)
rs1550623	*2q31.1*	*CDC7*	A/G	0.16	0.94	**0.21**	0.19	0.24	1.00	0.86	1.16	0.96	0.99	0.83	1.18	0.90	(Michailidou *et al.*, 2013)
rs2016394	*2q31.1*	*DLX2*	G/A	0.48	0.95	0.47	0.44	0.50	0.97	0.86	1.10	0.62	0.97	0.83	1.12	0.65	(Michailidou *et al.*, 2013)
rs16857609	*2q35*	*DIRC3*	C/T	0.26	1.08	**0.38**	0.35	0.41	0.98	0.86	1.12	0.79	0.92	0.78	1.07	0.28	(Michailidou *et al.*, 2013)
rs4442975	*2q35*	*IGFBP5*	G/T	0.49	0.87	**0.65**	0.62	0.68	0.82	0.65	0.96	**0.01**	0.82	0.62	0.99	**0.03**	(Ghoussaini *et al.*, 2014)
rs4973768	*3p24.1*	*SLC4A7*	C/T	0.47	1.10	**0.62**	0.59	0.65	1.02	0.89	1.14	0.77	1.03	0.86	1.16	0.74	(Michailidou *et al.*, 2013)
rs12493607	*3p24.1*	*TGFBR2*	G/C	0.35	1.06	**0.27**	0.25	0.30	0.96	0.84	1.10	0.56	0.93	0.79	1.09	0.38	(Michailidou *et al.*, 2013)
rs6762644	*3p26.1*	*EGOT/ITPR1*	A/G	0.40	1.07	**0.32**	0.29	0.35	1.10	0.97	1.26	0.14	1.10	0.94	1.29	0.23	(Michailidou *et al.*, 2013)
rs9790517	*4q24*	*TET2*	C/T	0.23	1.05	**0.34**	0.31	0.37	0.94	0.83	1.08	0.39	0.97	0.83	1.14	0.72	(Michailidou *et al.*, 2013)
rs6828523	*4q34.1*	*ADAM29*	C/A	0.13	0.90	**0.18**	0.15	0.20	0.91	0.77	1.07	0.25	0.89	0.73	1.08	0.25	(Michailidou *et al.*, 2013)
rs10941679	*5p12*	*MRPS30*	A/G	0.25	1.13	**0.33**	0.30	0.36	1.12	0.98	1.27	0.10	1.08	0.92	1.26	0.37	(Michailidou *et al.*, 2013)
rs10069690	*5p15.33*	*TERT*	C/T	0.26	1.06	**0.23**	0.20	0.25	1.11	0.96	1.28	0.17	1.03	0.87	1.23	0.73	(Michailidou *et al.*, 2013)
rs2736108	*5p15.33*	*TERT*	C/T	0.29	0.94	**0.21**	0.18	0.23	1.03	0.88	1.19	0.73	1.05	0.87	1.26	0.61	(Bojesen *et al.*, 2013b)
rs3215401	*5p15.33*	*TERT*	-/C	0.30	0.94	**0.22**	0.19	0.24	1.03	0.89	1.20	0.68	1.06	0.89	1.27	0.52	(Bojesen *et al.*, 2013b)
rs2242652	*5p15.33*	*TERT*	G/A	0.20	1.06	**0.14**	0.12	0.16	1.15	0.97	1.37	0.11	1.02	0.83	1.25	0.87	(Bojesen *et al.*, 2013b)
rs889312	*5q11.2*	*MAP3K1*	A/C	0.28	1.12	**0.43**	0.40	0.46	**0.97**	0.86	1.10	0.67	1.13	0.97	1.31	0.12	(Michailidou *et al.*, 2013)
rs1353747	*5q11.2*	*PDE4D*	T/G	0.10	0.92	0.11	0.09	0.13	0.94	0.78	1.13	0.51	1.00	0.80	1.26	0.97	(Michailidou *et al.*, 2013)

rs10472076	*5q11.2*	*RAB3C*	T/C	0.38	1.05	**0.28**	0.26	0.31	1.12	0.98	1.28	0.10	1.06	0.91	1.24	0.47	(Michailidou *et al.*, 2013)
rs1432679	*5q33.3*	*EBF1*	T/C	0.43	1.07	**0.62**	0.59	0.65	1.08	0.96	1.20	0.18	1.12	0.97	1.25	0.10	(Michailidou *et al.*, 2013)
rs204247	*6p23*	*RANBP9*	A/G	0.43	1.05	0.46	0.43	0.49	1.06	0.94	1.20	0.33	1.08	0.93	1.25	0.32	(Michailidou *et al.*, 2013)
rs11242675	*6p25.3*	*FOXQ1*	T/C	0.39	0.94	**0.49**	0.46	0.52	0.94	0.84	1.07	0.35	0.95	0.83	1.10	0.52	(Michailidou *et al.*, 2013)
rs17530068	*6q14.1*	*FAM46A*	A/G	0.22	1.05	**0.25**	0.23	0.28	1.08	0.94	1.24	0.31	1.01	0.85	1.19	0.93	(Michailidou *et al.*, 2013)
rs2046210	*6q25.1*	*ESR1/CCDC170*	G/A	0.34	1.08	**0.26**	0.24	0.29	1.18	1.03	1.36	**0.02**	1.30	1.11	1.54	**0.002**	(Michailidou *et al.*, 2013)
rs720475	*7q35*	*ARHGEF5/NOBOX*	G/A	0.25	0.94	**0.19**	0.16	0.21	0.98	0.83	1.14	0.76	0.96	0.79	1.16	0.67	(Michailidou *et al.*, 2013)
rs9693444	*8p12*	*DUSP4*	C/A	0.32	1.07	**0.28**	0.25	0.31	1.00	0.87	1.15	0.97	1.00	0.84	1.18	0.99	(Michailidou *et al.*, 2013)
rs6472903	*8q21.11*	*HNF4G*	T/G	0.18	0.91	**0.07**	0.06	0.09	**1.23**	0.98	1.56	0.07	1.13	0.86	1.50	0.38	(Michailidou *et al.*, 2013)
rs13281615	*8q24.21*	*POU5F1B*	A/G	0.41	1.09	0.57	0.54	0.60	1.02	0.89	1.14	0.74	0.94	0.76	1.09	0.45	(Michailidou *et al.*, 2013)
rs11780156	*8q24.21*	*PVT1/MYC*	C/T	0.16	1.07	0.18	0.16	0.21	1.09	0.93	1.28	0.28	1.12	0.93	1.36	0.23	(Michailidou *et al.*, 2013)
rs1011970	*9p21.3*	*CDKN2A/B*	G/T	0.17	1.06	**0.37**	0.34	0.40	0.99	0.88	1.13	0.93	1.06	0.92	1.23	0.41	(Michailidou *et al.*, 2013)
rs10816625	*9q31.2*	*KLF4*	A/G	0.06	1.12	**0.15**	0.13	0.17	1.09	0.92	1.29	0.31	1.06	0.87	1.30	0.57	(Orr *et al.*, 2015b)
rs676256	*9q31.2*	*KLF4*	A/G	0.38	0.90	**0.32**	0.29	0.35	0.96	0.84	1.11	0.61	0.88	0.74	1.03	0.12	(Orr *et al.*, 2015b)
rs865686	*9q31.2*	*KLF4*	T/G	0.38	0.89	**0.36**	0.33	0.39	0.93	0.82	1.06	0.27	0.91	0.78	1.06	0.22	(Michailidou *et al.*, 2013)
rs10759243	*9q31.2*	*KLF4*	C/A	0.39	1.06	**0.43**	0.40	0.46	1.15	1.01	1.30	**0.03**	1.18	1.01	1.37	**0.03**	(Michailidou *et al.*, 2013)
rs11814448	*10p12.31*	*DNAJC1*	A/C	0.02	1.26	**0.06**	0.04	0.07	**0.94**	0.72	1.22	0.62	1.26	0.92	1.74	0.16	(Michailidou *et al.*, 2013)
rs7072776	*10p12.31*	*MLLT10/DNAJC1*	G/A	0.29	1.07	**0.37**	0.34	0.40	0.98	0.86	1.11	0.71	1.01	0.87	1.18	0.90	(Michailidou *et al.*, 2013)
rs10995190	*10q21*	*ZNF365*	G/A	0.16	0.86	**0.09**	0.07	0.10	0.96	0.77	1.20	0.73	1.03	0.79	1.36	0.81	(Michailidou *et al.*, 2013)
rs704010	*10q22.3*	*ZMZ1*	C/T	0.38	1.08	**0.44**	0.41	0.47	0.98	0.86	1.11	0.72	1.02	0.88	1.18	0.82	(Michailidou *et al.*, 2013)
rs7904519	*10q25.2*	*TCF7L2*	A/G	0.46	1.06	**0.33**	0.30	0.35	1.14	1.00	1.29	**0.05**	1.12	0.96	1.30	0.16	(Michailidou *et al.*, 2013)
rs11199914	*10q26.12*	*FGFR2*	C/T	0.32	0.95	**0.43**	0.40	0.46	**0.79**	0.69	0.89	**0.0002**	0.77	0.66	0.89	**0.0007**	(Michailidou *et al.*, 2013)
rs35054928	*10q26.13*	*FGFR2*	-/C	0.44	1.27	**0.39**	0.36	0.42	1.20	1.06	1.36	**0.005**	1.17	1.00	1.35	**0.05**	(Meyer *et al.*, 2013)
rs2981579	*10q26.13*	*FGFR2*	G/A	0.40	1.27	0.39	0.36	0.42	1.21	1.06	1.37	**0.003**	1.17	1.00	1.35	**0.05**	(Michailidou *et al.*, 2013)
rs2981578	*10q26.13*	*FGFR2*	T/C	0.49	1.24	0.46	0.43	0.49	1.24	1.10	1.41	**0.0005**	1.20	1.03	1.38	**0.02**	(Meyer *et al.*, 2013)
rs3750817	*10q26.13*	*FGFR2*	C/T	0.39	0.78[5]	0.41	0.38	0.44	0.83	0.73	0.94	**0.005**	0.89	0.76	1.03	0.12	(Prentice *et al.*, 2009a)
rs11200014	*10q26.13*	*FGFR2*	G/A	0.39	1.28	0.36	0.33	0.39	1.14	1.01	1.30	**0.03**	1.06	0.92	1.23	0.43	(Barnholtz-Sloan *et al.*, 2010)
rs2420946	*10q26.13*	*FGFR2*	C/T	0.38	1.27	0.39	0.36	0.42	1.17	1.03	1.32	**0.02**	1.12	0.96	1.30	0.14	(Easton *et al.*, 2007b)

rs3817198	*11p15.5*	*LSP1*	T/C	0.31	1.07	**0.19**	0.16	0.21	1.23	1.06	1.43	**0.008**	1.14	0.95	1.37	0.15	(Michailidou *et al.*, 2013)
rs614367	*11q13.3*	*MYEOV/CCND1*	C/T	0.15	1.21	**0.06**	0.04	0.07	1.34	1.05	1.73	**0.02**	1.15	0.85	1.57	0.36	(Michailidou *et al.*, 2013)
rs3903072	*11q13.1*	*CFL1/OVOL1*	G/T	0.47	0.95	**0.32**	0.29	0.35	1.02	0.90	1.17	0.75	0.91	0.78	1.07	0.26	(Michailidou *et al.*, 2013)
rs554219	*11q13.3*	*CCND1*	C/G	0.12	1.27	**0.06**	0.05	0.08	1.40	1.09	1.79	**0.008**	1.20	0.89	1.62	0.23	(French *et al.*, 2013b)
rs11820646	*11q24.3*	*BARX2*	C/T	0.41	0.95	**0.48**	0.45	0.51	0.94	0.83	1.07	0.35	0.89	0.77	1.03	0.13	(Michailidou *et al.*, 2013)
rs12422552	*12p13.1*	*ATF7IP*	G/C	0.26	1.05	**0.23**	0.20	0.25	1.06	0.92	1.23	0.40	1.14	0.96	1.36	0.13	(Michailidou *et al.*, 2013)
rs17356907	*12q22*	*NTN4*	A/G	0.30	0.91	**0.38**	0.35	0.41	0.94	0.83	1.07	0.38	0.93	0.80	1.09	0.37	(Michailidou *et al.*, 2013)
rs1292011	*12q24.21*	*MED13L/TBX3*	A/G	0.42	0.92	0.41	0.38	0.44	0.96	0.85	1.09	0.55	0.96	0.82	1.11	0.57	(Michailidou *et al.*, 2013)
rs2236007	*14q13.3*	*PAX9*	G/A	0.21	0.93	**0.15**	0.13	0.17	0.88	0.74	1.04	0.13	0.88	0.72	1.08	0.22	(Michailidou *et al.*, 2013)
rs999737	*14q24.1*	*RAD51B*	C/T	0.23	0.92	**0.14**	0.12	0.16	0.94	0.78	1.12	0.49	0.84	0.68	1.04	0.12	(Michailidou *et al.*, 2013)
rs2588809	*14q24.1*	*RAD51B*	C/T	0.16	1.08	0.16	0.14	0.18	**1.28**	1.09	1.51	**0.003**	1.31	1.07	1.59	**0.008**	(Michailidou *et al.*, 2013)
rs941764	*14q32.11*	*CCDC88C*	A/G	0.34	1.06	**0.42**	0.39	0.45	**1.21**	1.06	1.37	**0.004**	1.26	1.08	1.46	**0.003**	(Michailidou *et al.*, 2013)
rs17817449	*16q12.2*	*FTO*	T/G	0.40	0.93	**0.27**	0.24	0.30	0.99	0.86	1.14	0.89	0.93	0.79	1.10	0.41	(Michailidou *et al.*, 2013)
rs11075995	*16q12.2*	*FTO*	T/A	0.24	1.11[4]	**0.36**	0.33	0.39	**0.95**	0.84	1.08	0.41	0.99	0.85	1.15	0.86	(Garcia-Closas *et al.*, 2013b)
rs3803662	*16q12.1*	*TOX3*	G/A	0.26	1.24	**0.43**	0.40	0.46	1.21	1.07	1.37	**0.003**	1.30	1.12	1.51	**0.0007**	(Michailidou *et al.*, 2013)
rs4784227	*16q12.1*	*TOX3*	C/T	0.25	1.19	**0.35**	0.32	0.38	1.21	1.07	1.38	**0.003**	1.20	1.03	1.40	**0.02**	(Long *et al.*, 2010)
rs13329835	*16q23.2*	*CDYL2*	A/G	0.22	1.08	**0.17**	0.15	0.20	1.22	1.04	1.43	**0.02**	1.23	1.02	1.49	**0.03**	(Michailidou *et al.*, 2013)
rs6504950	*17q22*	*STXBP4*	G/A	0.28	0.94	**0.17**	0.15	0.20	**1.17**	1.00	1.37	**0.05**	1.14	0.95	1.37	0.17	(Michailidou *et al.*, 2013)
rs1436904	*18q11.2*	*CHST9*	T/G	0.40	0.96	**0.45**	0.42	0.48	0.87	0.76	0.98	**0.02**	0.90	0.78	1.05	0.18	(Michailidou *et al.*, 2013)
rs527616	*18q11.2*	*AQP4*	G/C	0.38	0.95	**0.74**	0.71	0.77	0.94	0.78	1.08	0.40	0.99	0.80	1.14	0.88	(Michailidou *et al.*, 2013)
rs8170	*19p13.11*	*BABAM1=MERIT40*	G/A	0.19	1.04	**0.09**	0.08	0.11	1.16	0.95	1.43	0.15	1.07	0.84	1.37	0.58	(Michailidou *et al.*, 2013)
rs2363956	*19p13.11*	*ANKLE1*	C/A	0.49	1.01	**0.60**	0.57	0.63	0.95	0.81	1.07	0.42	0.94	0.77	1.09	0.48	(Antoniou *et al.*, 2010b)
rs4808801	*19p13.11*	*ELL*	A/G	0.35	0.93	0.32	0.29	0.35	0.96	0.84	1.10	0.59	0.95	0.81	1.11	0.51	(Michailidou *et al.*, 2013)
rs3760982	*19q13.31*	*KCNN4/LYPD5*	G/A	0.46	1.06	**0.31**	0.28	0.34	1.14	1.00	1.29	0.06	1.09	0.93	1.27	0.29	(Michailidou *et al.*, 2013)
rs2284378	*20q11.22*	*RALY*	C/T	0.31	1.14[4]	0.32	0.29	0.35	1.04	0.92	1.19	0.52	1.00	0.85	1.17	0.99	(Siddiq *et al.*, 2012)
rs2823093	*21q21.1*	*NRIP1*	G/A	0.27	0.92	**0.32**	0.29	0.34	**0.77**	0.67	0.89	**0.0003**	0.84	0.71	0.99	**0.04**	(Michailidou *et al.*, 2013)
rs6001930	*22q13.1*	*MKL1*	T/C	0.11	1.12	0.11	0.09	0.13	0.99	0.81	1.20	0.92	0.96	0.76	1.21	0.73	(Michailidou *et al.*, 2013)

Abbreviations: MAF, minor allele frequency; AF, allele frequency; OR, odds ratio; adj., adjusted; CI, confidence interval.

*Genome assembly GRCh37.
[1]AFs and ORs different in Europe and Colombia (non-overlapping confidence intervals) are marked in bold.
[2]P-values smaller than 0.05 are marked in bold, P-values smaller than 0.05 after multiplicity adjustment using multiple permutation are underlined.
[3]OR adjusted for age, family history of breast cancer in first-degree female relatives, oral contraceptive use, menopausal status combined with postmenopausal hormone therapy use, body mass index, smoking status, parity, age at first full-term pregnancy, age at menarche, and breastfeeding.
[4]ER-negative women.
[5]Postmenopausal women.

Table S3: Associations between 78 previously known risk SNPs in Asians with breast cancer risk in Colombians

			ASIANS			COLOMBIANS										
SNP ID	**Locus**	**Gene/nearest gene(s)***	**Alleles Major/Minor**	**MAF**	**OR**	**AF[1]**	**95% CI**		**OR**	**95% CI**		**P-value[2]**	**$OR_{adj.}$[1,3]**	**95 % CI**		**P-value[2]**	**Reference ASIANS**
rs11249433	*1p11.2*	*EMBP1*	A/G	0.03	1.16	**0.27**	0.24	0.29	1.09	0.95	1.25	0.21	1.02	0.87	1.19	0.84	(Zheng *et al.*, 2013)
rs11552449	*1p13.2*	*DCLRE1B*	T/C	0.39	0.97	**0.61**	0.58	0.64	1.09	0.97	1.20	0.13	1.08	0.93	1.21	0.27	(Zheng *et al.*, 2013)
rs616488	*1p36.22*	*PEX14*	A/G	0.31	0.94	**0.40**	0.37	0.43	0.92	0.81	1.05	0.24	0.95	0.81	1.11	0.55	(Zheng *et al.*, 2013)
rs6678914	*1q32.1*	*LGR6*	G/A	0.24	0.96[4]	**0.30**	0.27	0.32	1.09	0.95	1.25	0.21	1.10	0.93	1.29	0.26	(Garcia-Closas *et al.*, 2013b)
rs4245739	*1q32.1*	*MDM4*	A/C	0.06	0.97[4]	**0.25**	0.22	0.27	1.04	0.90	1.20	0.62	1.00	0.84	1.19	0.98	(Garcia-Closas *et al.*, 2013b)
rs12710696	*2p24.1*	*OSR1*	G/A	0.33	1.05[4]	0.32	0.29	0.35	0.96	0.84	1.10	0.56	0.96	0.82	1.12	0.59	(Garcia-Closas *et al.*, 2013b)
rs4849887	*2q14.1*	*INHBB*	C/T	0.18	0.93	**0.11**	0.09	0.13	0.85	0.69	1.04	0.11	0.88	0.69	1.12	0.29	(Zheng *et al.*, 2013)
rs1550623	*2q31.1*	*CDCA7*	A/G	0.01	0.79	**0.21**	0.19	0.24	1.00	0.86	1.16	0.96	**0.99**	0.83	1.18	0.90	(Zheng *et al.*, 2013)
rs2016394	*2q31.1*	*DLX2*	G/A	0.19	0.99	**0.47**	0.44	0.50	0.97	0.86	1.10	0.62	0.97	0.83	1.12	0.65	(Zheng *et al.*, 2013)
rs16857609	*2q35*	*DIRC3*	T/C	0.41	0.93	0.62	0.59	0.65	1.02	0.88	1.14	0.79	1.08	0.93	1.22	0.28	(Zheng *et al.*, 2013)
rs4442975	*2q35*	*IGFBP5*	G/T	0.13	0.94	**0.65**	0.62	0.68	0.82	0.65	0.96	**0.01**	0.82	0.62	0.99	**0.03**	(Ghoussaini *et al.*, 2014)
rs4973768	*3p24.1*	*SLC4A7*	C/T	0.19	1.11	**0.62**	0.59	0.65	1.02	0.89	1.14	0.77	1.03	0.86	1.16	0.74	(Zheng *et al.*, 2013)
rs12493607	*3p24.1*	*TGFBR2*	C/G	0.33	0.95	**0.73**	0.70	0.75	1.04	0.90	1.16	0.56	1.07	0.91	1.21	0.38	(Zheng *et al.*, 2013)
rs6762644	*3p26.1*	*EGOT/ITPR1*	A/G	0.08	1.03	**0.32**	0.29	0.35	1.10	0.97	1.26	0.14	1.10	0.94	1.29	0.23	(Zheng *et al.*, 2013)
rs9790517	*4q24*	*TET2*	T/C	0.39	0.98	**0.66**	0.63	0.69	1.06	0.92	1.17	0.39	1.03	0.86	1.17	0.72	(Zheng *et al.*, 2013)
rs6828523	*4q34.1*	*ADAM29*	A/C	0.24	0.93	**0.82**	0.80	0.85	1.09	0.93	1.23	0.25	1.11	0.92	1.27	0.25	(Zheng *et al.*, 2013)
rs10941679	*5p12*	*MRPS30*	A/G	0.48	1.08	**0.33**	0.30	0.36	1.12	0.98	1.27	0.10	1.08	0.92	1.26	0.37	(Zheng *et al.*, 2013)
rs10069690	*5p15.33*	*TERT*	C/T	0.22	1.05	0.23	0.20	0.25	1.11	0.96	1.28	0.17	1.03	0.87	1.23	0.73	(Zheng *et al.*, 2013)
rs2736108	*5p15.33*	*TERT*	C/T	0.28	0.98	**0.21**	0.18	0.23	1.03	0.88	1.19	0.73	1.05	0.87	1.26	0.61	(Bojesen *et al.*, 2013b)
rs3215401	*5p15.33*	*TERT*	- /C	0.34	0.97	**0.22**	0.19	0.24	1.03	0.89	1.20	0.68	1.06	0.89	1.27	0.52	(Bojesen *et al.*, 2013b)
rs2242652	*5p15.33*	*TERT*	G/A	0.20	1.04	**0.14**	0.12	0.16	1.15	0.97	1.37	0.11	1.02	0.83	1.25	0.87	(Bojesen *et al.*, 2013b)
rs889312	*5q11.2*	*MAP3K1*	C/A	0.48	0.95	**0.57**	0.54	0.60	1.03	0.90	1.14	0.67	0.87	0.69	1.03	0.12	(Zheng *et al.*, 2013)
rs1353747	*5q11.2*	*PDE4D*	T/G		1.18	**0.11**	0.09	0.13	0.94	0.78	1.13	0.51	1.00	0.80	1.26	0.97	(Zheng *et al.*, 2013)
rs10472076	*5q11.2*	*RAB3C*	T/C	0.27	1.02	0.28	0.26	0.31	1.12	0.98	1.28	0.10	1.06	0.91	1.24	0.47	(Zheng *et al.*, 2013)
rs1432679	*5q33.3*	*EBF1*	C/T	0.34	0.92	0.38	0.35	0.41	0.92	0.80	1.04	0.18	0.88	0.75	1.03	0.10	(Zheng *et al.*, 2013)
rs204247	*6p23*	*RANBP9*	G/A	0.38	0.97	**0.54**	0.51	0.57	0.94	0.80	1.06	0.33	0.92	0.75	1.07	0.32	(Zheng *et al.*, 2013)

rs11242675	*6p25.3*	*FOXQ1*	C/T	0.48	0.99	**0.51**	0.48	0.54	1.06	0.93	1.16	0.35	1.05	0.90	1.17	0.52	(Zheng *et al.*, 2013)
rs17530068	*6q14.1*	*FAM46A*	A/G	0.22		**0.25**	0.23	0.28	1.08	0.94	1.24	0.31	1.01	0.85	1.19	0.93	(Wen *et al.*, 2016)
rs2046210	*6q25.1*	*ESR1/CCDC170*	G/A	0.34	1.27	**0.26**	0.24	0.29	1.18	1.03	1.36	**0.02**	1.30	1.11	1.54	**0.002**	(Zheng *et al.*, 2013)
rs720475	*7q35*	*ARHGEF5/NOBOX*	G/A	0.03	0.98	**0.19**	0.16	0.21	0.98	0.83	1.14	0.76	0.96	0.79	1.16	0.67	(Zheng *et al.*, 2013)
rs9693444	*8p12*	*DUSP4*	C/A	0.29	1.07	0.28	0.25	0.31	1.00	0.87	1.15	0.97	1.00	0.84	1.18	0.99	(Zheng *et al.*, 2013)
rs6472903	*8q21.11*	*HNF4G*	T/G	0.04	0.84	**0.07**	0.06	0.09	1.23	0.98	1.56	0.07	**1.13**	0.86	1.50	0.38	(Zheng *et al.*, 2013)
rs13281615	*8q24.21*	*POU5F1B*	G/A	0.48	0.97	**0.43**	0.40	0.46	0.98	0.86	1.11	0.74	1.06	0.91	1.24	0.45	(Zheng *et al.*, 2013)
rs11780156	*8q24.21*	*PVT1/MYC*	C/T	0.22	1.00	**0.18**	0.16	0.21	1.09	0.93	1.28	0.28	1.12	0.93	1.36	0.23	(Zheng *et al.*, 2013)
rs1011970	*9p21.3*	*CDKN2A/B*	G/T	0.08	1.06	**0.37**	0.34	0.40	0.99	0.88	1.13	0.93	1.06	0.92	1.23	0.41	(Zheng *et al.*, 2013)
rs10816625	*9q31.2*	*KLF4*	A/G	0.38	1.12	**0.15**	0.13	0.17	1.09	0.92	1.29	0.31	1.06	0.87	1.30	0.57	(Orr *et al.*, 2015b)
rs676256	*9q31.2*	*KLF4*	A/G	0.05	0.94	**0.32**	0.29	0.35	0.96	0.84	1.11	0.61	0.88	0.74	1.03	0.12	(Orr *et al.*, 2015b)
rs865686	*9q31.2*	*KLF4*	T/G	0.07	0.96	**0.36**	0.33	0.39	0.93	0.82	1.06	0.27	0.91	0.78	1.06	0.22	(Zheng *et al.*, 2013)
rs10759243	*9q31.2*	*KLF4*	C/A	0.42	1.05	0.43	0.40	0.46	1.15	1.01	1.30	**0.03**	1.18	1.01	1.37	**0.03**	(Zheng *et al.*, 2013)
rs11814448	*10p12.31*	*DNAJC1*	A/C	0.01	1.08	**0.06**	0.04	0.07	0.94	0.72	1.22	0.62	1.26	0.92	1.74	0.16	(Zheng *et al.*, 2013)
rs7072776	*10p12.31*	*MLLT10/DNAJC1*	G/A	0.03	1.04	**0.37**	0.34	0.40	0.98	0.86	1.11	0.71	1.01	0.87	1.18	0.90	(Zheng *et al.*, 2013)
rs10995190	*10q21*	*ZNF365*	G/A	0.02	1.06	**0.09**	0.07	0.10	0.96	0.77	1.20	0.73	1.03	0.79	1.36	0.81	(Zheng *et al.*, 2013)
rs704010	*10q22.3*	*ZMZ1*	C/T	0.29	1.05	**0.44**	0.41	0.47	0.98	0.86	1.11	0.72	1.02	0.88	1.18	0.82	(Zheng *et al.*, 2013)
rs7904519	*10q25.2*	*TCF7L2*	A/G	0.04	1.02	**0.33**	0.30	0.35	1.14	1.00	1.29	**0.05**	1.12	0.96	1.30	0.16	(Zheng *et al.*, 2013)
rs11199914	*10q26.12*	*FGFR2*	C/T	0.40	0.97	0.43	0.40	0.46	0.79	0.69	0.89	**0.0002**	**0.77**	0.66	0.89	**0.0007**	(Zheng *et al.*, 2013)
rs35054928	*10q26.13*	*FGFR2*	-/C	0.44	1.15	**0.39**	0.36	0.42	1.20	1.06	1.36	**0.005**	1.17	1.00	1.35	**0.05**	(Meyer *et al.*, 2013)
rs2981579	*10q26.13*	*FGFR2*	G/A	0.43	1.16	**0.39**	0.36	0.42	1.21	1.06	1.37	**0.003**	1.17	1.00	1.35	**0.05**	(Meyer *et al.*, 2013)
rs2981578	*10q26.13*	*FGFR2*	T/C	0.49	1.19	0.46	0.43	0.49	1.24	1.10	1.41	**0.0005**	1.20	1.03	1.38	**0.02**	(Meyer *et al.*, 2013)
rs3750817	*10q26.13*	*FGFR2*	C/T	0.49	1.22[5]	**0.41**	0.38	0.44	0.83	0.73	0.94	**0.005**	**0.89**	0.76	1.03	0.12	(Elgazzar *et al.*, 2012)
rs11200014	*10q26.13*	*FGFR2*	G/A	0.26	1.16	**0.36**	0.33	0.39	1.14	1.01	1.30	**0.03**	1.06	0.92	1.23	0.43	(Easton *et al.*, 2007b)
rs2420946	*10q26.13*	*FGFR2*	C/T	0.37	1.17	0.39	0.36	0.42	1.17	1.03	1.32	**0.02**	1.12	0.96	1.30	0.14	(Easton *et al.*, 2007b)
rs3817198	*11p15.5*	*LSP1*	T/C	0.13	1.07	**0.19**	0.16	0.21	1.23	1.06	1.43	**0.008**	1.14	0.95	1.37	0.15	(Zheng *et al.*, 2013)
rs614367	*11q13.3*	*MYEOV/CCND1*	C/T	0.01	0.71	**0.06**	0.04	0.07	1.34	1.05	1.73	**0.02**	**1.15**	0.85	1.57	0.36	(Zheng *et al.*, 2013)
rs3903072	*11q13.1*	*CFL1/OVOL1*	G/T	0.18	0.98	**0.32**	0.29	0.35	1.02	0.90	1.17	0.75	0.91	0.78	1.07	0.26	(Zheng *et al.*, 2013)
rs554219	*11q13.3*	*CCND1*	C/G	0.02	1.36	**0.94**	0.92	0.95	0.60	0.21	0.91	**0.008**	**0.80**	0.38	1.11	0.23	(French *et al.*, 2013b)

rs11820646	*11q24.3*	*BARX2*	C/T	0.44	0.95	**0.48**	0.45	0.51	0.94	0.83	1.07	0.35	0.89	0.77	1.03	0.13	(Zheng *et al.*, 2013)
rs12422552	*12p13.1*	*ATF7IP*	G/C	0.28	1.05	**0.23**	0.20	0.25	1.06	0.92	1.23	0.40	1.14	0.96	1.36	0.13	(Zheng *et al.*, 2013)
rs17356907	*12q22*	*NTN4*	A/G	0.23	0.93	**0.38**	0.35	0.41	0.94	0.83	1.07	0.38	0.93	0.80	1.09	0.37	(Zheng *et al.*, 2013)
rs1292011	*12q24.21*	*MED13L/TBX3*	A/G	0.26	0.90	**0.41**	0.38	0.44	0.96	0.85	1.09	0.55	0.96	0.82	1.11	0.57	(Zheng *et al.*, 2013)
rs2236007	*14q13.3*	*PAX9*	G/A	0.25	0.92	**0.15**	0.13	0.17	0.88	0.74	1.04	0.13	0.88	0.72	1.08	0.22	(Zheng *et al.*, 2013)
rs999737	*14q24.1*	*RAD51B*	T/C	0.00	1.08	**0.86**	0.84	0.88	1.06	0.88	1.22	0.49	1.16	0.96	1.32	0.12	(Zheng *et al.*, 2013)
rs2588809	*14q24.1*	*RAD51B*	C/T	0.02	1.06	**0.16**	0.14	0.18	1.28	1.09	1.51	**0.003**	**1.31**	1.07	1.59	**0.008**	(Zheng *et al.*, 2013)
rs941764	*14q32.11*	*CCDC88C*	A/G	0.14	1.05	**0.42**	0.39	0.45	1.21	1.06	1.37	**0.004**	**1.26**	1.08	1.46	**0.003**	(Zheng *et al.*, 2013)
rs17817449	*16q12.2*	*FTO*	T/G	0.12	0.92	**0.27**	0.24	0.30	0.99	0.86	1.14	0.89	0.93	0.79	1.10	0.41	(Zheng *et al.*, 2013)
rs11075995	*16q12.2*	*FTO*	T/A	0.30	1.03[4]	**0.36**	0.33	0.39	0.95	0.84	1.08	0.41	0.99	0.85	1.15	0.86	(Garcia-Closas *et al.*, 2013b)
rs3803662	*16q12.1*	*TOX3*	A/G	0.36	0.87	**0.57**	0.54	0.60	0.79	0.63	0,93	**0.003**	0.70	0.49	0.88	**0.0007**	(Zheng *et al.*, 2013)
rs4784227	*16q12.1*	*TOX3*	C/T	0.24	1.24	**0.35**	0.32	0.38	1.21	1.07	1.38	**0.003**	1.20	1.03	1.40	**0.02**	(Zheng *et al.*, 2013)
rs13329835	*16q23.2*	*CDYL2*	A/G	0.05	1.02	**0.17**	0.15	0.20	1.22	1.04	1.43	**0.02**	1.23	1.02	1.49	**0.03**	(Zheng *et al.*, 2013)
rs6504950	*17q22*	*STXBP4*	G/A	0.07	0.98	**0.17**	0.15	0.20	1.17	1.00	1.37	**0.05**	1.14	0.95	1.37	0.17	(Zheng *et al.*, 2013)
rs1436904	*18q11.2*	*CHST9*	T/G	0.45	0.98	0.45	0.42	0.48	0.87	0.76	0.98	**0.02**	0.90	0.78	1.05	0.18	(Zheng *et al.*, 2013)
rs527616	*18q11.2*	*AQP4*	G/C	0.31	0.97	**0.74**	0.71	0.77	0.94	0.78	1.08	0.40	0.99	0.80	1.14	0.88	(Zheng *et al.*, 2013)
rs8170	*19p13.11*	*BABAM1=MERIT40*	G/A	0.01		**0.09**	0.08	0.11	1.16	0.95	1.43	0.15	1.07	0.84	1.37	0.58	(Zheng *et al.*, 2013)
rs2363956	*19p13.11*	*ANKLE1*	A/C	0.32		**0.40**	0.37	0.43	1.05	0.93	1.19	0.42	1.06	0.91	1.23	0.48	(Wen *et al.*, 2016)
rs4808801	*19p13.11*	*ELL*	A/G	0.25	0.96	**0.32**	0.29	0.35	0.96	0.84	1.10	0.59	0.95	0.81	1.11	0.51	(Zheng *et al.*, 2013)
rs3760982	*19q13.31*	*KCNN4/LYPD5*	G/A	0.13	1.02	**0.31**	0.28	0.34	1.14	1.00	1.29	0.06	1.09	0.93	1.27	0.29	(Zheng *et al.*, 2013)
rs2284378	*20q11.22*	*RALY*	C/T	0.26	1.08	**0.32**	0.29	0.35	1.04	0.92	1.19	0.52	1.00	0.85	1.17	0.99	(Siddiq *et al.*, 2012)
rs2823093	*21q21.1*	*NRIP1*	G/A	0.03	0.93	**0.32**	0.29	0.34	0.77	0.67	0.89	**0.0003**	0.84	0.71	0.99	**0.04**	(Zheng *et al.*, 2013)
rs6001930	*22q13.1*	*MKL1*	T/C	0.28	1.03	**0.11**	0.09	0.13	0.99	0.81	1.20	0.92	0.96	0.76	1.21	0.73	(Zheng *et al.*, 2013)

Abbreviations: MAF, minor allele frequency; AF, allele frequency; OR, odds ratio; adj., adjusted; CI, confidence interval.

* Genome assembly GRCh37.

[1]AFs and ORs different in Asia and Colombia (non-overlapping confidence intervals) are marked in bold.

[2]*P* values smaller than 0.05 are marked in bold, *P* values smaller than 0.05 after multiplicity adjustment using multiple permutation are underlined.

[3]OR adjusted for age, family history of breast cancer in first-degree female relatives, oral contraceptive use, menopausal status combined with postmenopausal hormone therapy use, body mass index, smoking status, parity, age at first full-term pregnancy, age at menarche, and breastfeeding.

[4]ER-negative women.

[5]ER-positive women.

Table S4: Associations between 78 previously known risk SNPs in the Colombian population by estrogen receptor status

			ER-positive (N=592)				ER-negative (N=170)			
SNP ID	**Locus**	**Gene/nearest gene(s)***	**$OR_{adj.}$[1]**	**95% CI**		**P-value[2]**	**$OR_{adj.}$[1]**	**95% CI**		**P-value[2]**
rs11249433	*1p11.2*	*EMBP1*	0.99	0.82	1.19	0.92	0.94	0.70	1.25	0.66
rs11552449	*1p13.2*	*DCLRE1B*	0.89	0.74	1.06	0.20	0.97	0.74	1.27	0.84
rs616488	*1p36.22*	*PEX14*	0.92	0.77	1.10	0.35	0.82	0.62	1.09	0.17
rs6678914	*1q32.1*	*LGR6*	1.15	0.96	1.38	0.13	0.98	0.72	1.32	0.88
rs4245739	*1q32.1*	*MDM4*	1.00	0.82	1.22	0.99	0.96	0.70	1.30	0.79
rs12710696	*2p24.1*	*OSR1*	1.03	0.85	1.24	0.77	0.79	0.58	1.07	0.13
rs4849887	*2q14.1*	*INHBB*	0.81	0.61	1.08	0.16	0.80	0.49	1.24	0.33
rs1550623	*2q31.1*	*CDCA7*	0.97	0.79	1.20	0.79	1.04	0.75	1.43	0.80
rs2016394	*2q31.1*	*DLX2*	0.97	0.82	1.15	0.74	0.80	0.61	1.04	0.09
rs16857609	*2q35*	*DIRC3*	0.91	0.76	1.10	0.34	0.87	0.66	1.15	0.35
rs4442975	*2q35*	*IGFBP5*	1.16	0.98	1.39	0.09	1.17	0.89	1.53	0.26
rs4973768	*3p24.1*	*SLC4A7*	0.95	0.79	1.13	0.54	0.96	0.73	1.26	0.76
rs12493607	*3p24.1*	*TGFBR2*	0.88	0.73	1.07	0.19	0.98	0.73	1.32	0.91
rs6762644	*3p26.1*	*EGOT/ITPR1*	1.03	0.86	1.24	0.75	1.48	1.12	1.94	**0.005**
rs9790517	*4q24*	*TET2*	0.96	0.80	1.15	0.66	1.05	0.79	1.40	0.71
rs6828523	*4q34.1*	*ADAM29*	0.94	0.75	1.18	0.60	0.87	0.60	1.23	0.43
rs10941679	*5p12*	*MRPS30*	1.14	0.94	1.37	0.18	0.99	0.74	1.31	0.93
rs10069690	*5p15.33*	*TERT*	1.01	0.83	1.24	0.91	1.22	0.91	1.63	0.18
rs2736108	*5p15.33*	*TERT*	1.08	0.87	1.33	0.50	0.89	0.63	1.23	0.49
rs3215401	*5p15.33*	*TERT*	1.09	0.88	1.34	0.42	0.93	0.67	1.28	0.68
rs2242652	*5p15.33*	*TERT*	1.12	0.89	1.42	0.33	1.18	0.83	1.67	0.35
rs889312	*5q11.2*	*MAP3K1*	1.25	1.05	1.49	**0.01**	0.85	0.65	1.11	0.24
rs1353747	*5q11.2*	*PDE4D*	0.94	0.72	1.22	0.65	1.13	0.76	1.64	0.53
rs10472076	*5q11.2*	*RAB3C*	1.11	0.92	1.33	0.27	0.83	0.62	1.12	0.24
rs1432679	*5q33.3*	*EBF1*	0.86	0.72	1.03	0.10	1.07	0.81	1.42	0.61
rs204247	*6p23*	*RANBP9*	1.17	0.98	1.39	0.08	0.85	0.65	1.11	0.23
rs11242675	*6p25.3*	*FOXQ1*	0.95	0.80	1.12	0.55	0.83	0.64	1.07	0.16
rs17530068	*6q14.1*	*FAM46A*	0.97	0.80	1.18	0.75	1.27	0.95	1.70	0.10
rs2046210	*6q25.1*	*ESR1/CCDC170*	1.33	1.10	1.61	**0.003**	1.34	1.00	1.78	**0.05**
rs720475	*7q35*	*ARHGEF5/NOBOX*	0.84	0.66	1.06	0.14	1.23	0.88	1.71	0.21
rs9693444	*8p12*	*DUSP4*	1.05	0.87	1.28	0.60	0.91	0.67	1.23	0.55
rs6472903	*8q21.11*	*HNF4G*	1.22	0.89	1.68	0.21	1.06	0.64	1.70	0.80
rs13281615	*8q24.21*	*POU5F1B*	1.02	0.85	1.21	0.86	1.09	0.83	1.43	0.55
rs11780156	*8q24.21*	*PVT1/MYC*	1.15	0.92	1.44	0.20	1.08	0.77	1.49	0.67
rs1011970	*9p21.3*	*CDKN2A/B*	1.00	0.84	1.19	0.99	1.38	1.07	1.79	**0.01**
rs10816625	*9q31.2*	*KLF4*	1.09	0.86	1.37	0.49	1.05	0.73	1.49	0.79
rs676256	*9q31.2*	*KLF4*	0.88	0.73	1.07	0.20	0.82	0.61	1.10	0.19
rs865686	*9q31.2*	*KLF4*	0.89	0.74	1.07	0.21	0.85	0.65	1.11	0.25
rs10759243	*9q31.2*	*KLF4*	1.18	0.99	1.41	0.06	1.01	0.77	1.32	0.94
rs11814448	*10p12.31*	*DNAJC1*	1.11	0.75	1.63	0.61	1.35	0.74	2.33	0.30
rs7072776	*10p12.31*	*MLLT10/DNAJC1*	0.99	0.83	1.18	0.91	1.10	0.84	1.43	0.50
rs10995190	*10q21*	*ZNF365*	1.07	0.78	1.47	0.67	0.88	0.51	1.43	0.61
rs704010	*10q22.3*	*ZMZ1*	1.03	0.87	1.23	0.71	0.95	0.73	1.23	0.68
rs7904519	*10q25.2*	*TCF7L2*	1.03	0.86	1.23	0.78	1.01	0.77	1.33	0.92
rs11199914	*10q26.12*	*FGFR2*	0.74	0.62	0.89	**0.001**	0.75	0.57	0.99	**0.04**
rs35054928	*10q26.13*	*FGFR2*	1.20	1.01	1.43	**0.04**	1.23	0.94	1.61	0.13
rs2981579	*10q26.13*	*FGFR2*	1.16	0.97	1.38	0.10	1.33	1.02	1.74	**0.04**

rs2981578	*10q26.13*	*FGFR2*	1.24	1.05	1.48	**0.01**	1.27	0.98	1.65	0.07
rs3750817	*10q26.13*	*FGFR2*	0.91	0.76	1.08	0.27	0.80	0.61	1.06	0.12
rs11200014	*10q26.13*	*FGFR2*	1.10	0.93	1.31	0.28	1.09	0.84	1.42	0.51
rs2420946	*10q26.13*	*FGFR2*	1.16	0.97	1.38	0.10	1.24	0.95	1.61	0.11
rs3817198	*11p15.5*	*LSP1*	1.19	0.97	1.46	0.09	1.09	0.78	1.50	0.60
rs614367	*11q13.3*	*MYEOV/CCND1*	1.06	0.73	1.51	0.77	1.44	0.85	2.35	0.16
rs3903072	*11q13.1*	*CFL1/OVOL1*	0.86	0.72	1.04	0.12	1.05	0.79	1.39	0.73
rs554219	*11q13.3*	*CCND1*	1.14	0.80	1.61	0.46	1.21	0.71	1.99	0.47
rs11820646	*11q24.3*	*BARX2*	0.90	0.76	1.07	0.23	0.91	0.70	1.18	0.47
rs12422552	*12p13.1*	*ATF7IP*	1.13	0.92	1.39	0.24	0.97	0.69	1.33	0.83
rs17356907	*12q22*	*NTN4*	0.86	0.72	1.02	0.09	1.08	0.83	1.41	0.56
rs1292011	*12q24.21*	*MED13L/TBX3*	1.02	0.86	1.22	0.80	1.03	0.78	1.34	0.85
rs2236007	*14q13.3*	*PAX9*	0.90	0.71	1.13	0.36	1.04	0.73	1.46	0.83
rs999737	*14q24.1*	*RAD51B*	0.79	0.61	1.01	0.07	0.79	0.52	1.16	0.24
rs2588809	*14q24.1*	*RAD51B*	1.28	1.01	1.61	**0.04**	1.16	0.81	1.64	0.40
rs941764	*14q32.11*	*CCDC88C*	1.30	1.09	1.55	**0.004**	1.42	1.08	1.86	**0.01**
rs17817449	*16q12.2*	*FTO*	0.91	0.75	1.11	0.36	0.93	0.68	1.25	0.63
rs11075995	*16q12.2*	*FTO*	0.99	0.82	1.18	0.89	0.94	0.71	1.23	0.63
rs3803662	*16q12.1*	*TOX3*	1.33	1.12	1.58	**0.001**	1.06	0.80	1.38	0.70
rs4784227	*16q12.1*	*TOX3*	1.25	1.05	1.49	**0.01**	1.06	0.81	1.40	0.67
rs13329835	*16q23.2*	*CDYL2*	1.20	0.96	1.49	0.11	1.34	0.98	1.82	0.07
rs6504950	*17q22*	*STXBP4*	0.99	0.80	1.23	0.95	1.46	1.08	1.98	**0.01**
rs1436904	*18q11.2*	*CHST9*	0.92	0.77	1.10	0.38	1.05	0.80	1.36	0.74
rs527616	*18q11.2*	*AQP4*	0.99	0.82	1.21	0.95	1.13	0.84	1.51	0.41
rs8170	*19p13.11*	*BABAM1=MERIT40*	1.02	0.76	1.36	0.91	1.39	0.92	2.06	0.11
rs2363956	*19p13.11*	*ANKLE1*	1.09	0.92	1.30	0.32	0.87	0.66	1.14	0.31
rs4808801	*19p13.11*	*ELL*	1.04	0.86	1.26	0.67	0.80	0.60	1.06	0.13
rs3760982	*19q13.31*	*KCNN4/LYPD5*	1.10	0.92	1.33	0.29	1.15	0.87	1.52	0.33
rs2284378	*20q11.22*	*RALY*	1.06	0.88	1.27	0.55	1.08	0.81	1.41	0.60
rs2823093	*21q21.1*	*NRIP1*	0.73	0.60	0.89	**0.002**	0.94	0.70	1.26	0.69
rs6001930	*22q13.1*	*MKL1*	0.93	0.71	1.22	0.62	0.91	0.58	1.37	0.66

Abbreviations: OR, odds ratio; adj., adjusted; CI, confidence interval.

*Genome assembly GRCh37.

[1]OR adjusted for age, family history of breast cancer in first-degree female relatives, oral contraceptive use, menopausal status combined with postmenopausal hormone therapy use, body mass index, smoking status, parity, age at first full-term pregnancy, age at menarche, and breastfeeding.

[2]P-values smaller than 0.05 are marked in bold.

Table S5: Assessment of potential interaction between 13 significant SNPs in the Colombian population and European ancestry

				Group of women according to quartiles of European ancestry (ancestry ranges, N cases, N controls)											
			Interaction	0-45.0%, 218, 243			45.1%-55.0%, 225, 235			55.1%-66.0%, 231, 229			66.1%-100%, 315, 146		
SNP ID	Locus	Gene/nearest gene(s)*	P-value[1]	$OR_{adj.}$[2]	95% CI		$OR_{adj.}$[2]	95% CI		$OR_{adj.}$[2]	95% CI		$OR_{adj.}$[2]	95% CI	
rs4442975	2q35	*IGFBP5*	0.23	1.34	0.95	1.91	0.91	0.65	1.28	1.42	1.03	1.98	1.02	0.72	1.44
rs2046210	6q25.1	*ESR1/CCDC170*	0.96	1.08	0.75	1.57	1.64	1.13	2.38	1.07	0.75	1.51	1.31	0.90	1.92
rs10759243	*9q31.2*	*KLF4*	0.84	1.13	0.82	1.57	1.34	0.99	1.84	0.96	0.69	1.35	1.43	1.01	2.06
rs11199914	10q26.12	*FGFR2*	0.99	0.71	0.51	0.99	0.81	0.59	1.11	0.84	0.60	1.17	0.79	0.54	1.13
rs35054928	10q26.13	*FGFR2*	0.51	1.46	1.04	2.05	1.14	0.85	1.54	0.98	0.71	1.37	1.13	0.79	1.64
rs2981579	10q26.13	*FGFR2*	0.39	1.43	1.02	2.01	1.18	0.88	1.60	0.98	0.71	1.36	1.06	0.74	1.53
rs2981578	10q26.13	*FGFR2*	0.53	1.44	1.04	2.00	1.19	0.89	1.59	1.11	0.80	1.54	1.12	0.79	1.61
rs2588809	14q24.1	*RAD51B*	0.40	1.15	0.73	1.80	1.58	1.03	2.45	1.42	0.93	2.21	1.18	0.74	1.92
rs941764	14q32.11	*CCDC88C*	**0.02**	0.91	0.65	1.27	1.29	0.94	1.78	1.63	1.18	2.27	1.76	1.23	2.56
rs3803662	16q12.1	*TOX3*	**0.03**	1.32	0.96	1.84	1.46	1.06	2.04	1.35	0.99	1.85	1.07	0.74	1.56
rs4784227	16q12.1	*TOX3*	0.32	1.05	0.76	1.44	1.35	0.98	1.86	1.41	1.03	1.93	0.88	0.59	1.30
rs13329835	16q23.2	*CDYL2*	0.24	1.04	0.71	1.51	0.94	0.63	1.40	1.17	0.73	1.89	1.86	1.16	3.05
rs2823093	21q21.1	*NRIP1*	0.58	0.78	0.54	1.12	0.97	0.68	1.38	0.85	0.60	1.19	0.78	0.51	1.19

Abbreviations: OR, odds ratio; adj., adjusted; CI, confidence interval.

*Genome assembly GRCh37.

[1]P-values smaller than 0.05 are marked in bold.

[2]OR adjusted for age, family history of breast cancer in first-degree female relatives, oral contraceptive use, menopausal status combined with postmenopausal hormone therapy use, body mass index, smoking status, parity, age at first full-term pregnancy, age at menarche, and breastfeeding.

Table S6: Assessment of potential interaction between 13 significant SNPs in the Colombian population and Native American ancestry

				Group of women according to quartiles of Native American ancestry (ancestry ranges, N cases, N controls)											
			Interaction	0-27.0%, 326, 135			27.1%-38.0%, 242, 218			38.1%-47.0%, 207, 253			47.1%-100%,214, 247		
SNP ID	Locus	Gene/nearest gene(s)*	P-value[1]	$OR_{adj.}$[2]	95% CI		$OR_{adj.}$[2]	95% CI		$OR_{adj.}$[2]	95% CI		$OR_{adj.}$[2]	95% CI	
rs4442975	2q35	*IGFBP5*	0.37	1.18	0.83	1.69	1.22	0.88	1.69	1.09	0.79	1.52	1.10	0.77	1.57
rs2046210	6q25.1	*ESR1/CCDC170*	0.66	1.25	0.86	1.84	1.42	1.00	2.03	1.18	0.82	1.69	1.20	0.82	1.77
rs10759243	*9q31.2*	*KLF4*	0.91	1.53	1.06	2.22	1.09	0.78	1.51	1.11	0.79	1.55	1.27	0.92	1.75
rs11199914	10q26.12	*FGFR2*	0.58	0.89	0.61	1.29	0.77	0.55	1.06	0.82	0.60	1.13	0.66	0.47	0.91
rs35054928	10q26.13	*FGFR2*	0.69	1.06	0.73	1.54	1.00	0.72	1.40	1.19	0.86	1.65	1.27	0.92	1.76
rs2981579	10q26.13	*FGFR2*	0.75	0.95	0.66	1.37	1.06	0.76	1.48	1.21	0.87	1.67	1.26	0.91	1.74
rs2981578	10q26.13	*FGFR2*	0.68	0.89	0.62	1.29	1.22	0.88	1.69	1.33	0.97	1.84	1.21	0.88	1.67
rs2588809	14q24.1	*RAD51B*	0.63	1.48	0.91	2.50	1.33	0.87	2.04	1.62	1.05	2.53	1.09	0.70	1.69
rs941764	14q32.11	*CCDC88C*	**0.02**	1.41	0.98	2.03	1.84	1.31	2.60	1.34	0.98	1.82	0.97	0.68	1.38
rs3803662	16q12.1	*TOX3*	**0.02**	1.13	0.77	1.65	1.25	0.91	1.73	1.61	1.17	2.24	1.42	1.02	1.99
rs4784227	16q12.1	*TOX3*	0.25	1.02	0.68	1.53	1.19	0.87	1.65	1.54	1.12	2.14	1.11	0.80	1.53
rs13329835	16q23.2	*CDYL2*	0.98	1.16	0.77	1.77	1.40	0.89	2.22	0.83	0.54	1.25	1.24	0.81	1.91
rs2823093	21q21.1	*NRIP1*	0.68	0.84	0.57	1.25	0.80	0.56	1.14	0.84	0.59	1.19	0.99	0.68	1.43

Abbreviations: OR, odds ratio; adj., adjusted; CI, confidence interval.

*Genome assembly GRCh37.

[1]P-values smaller than 0.05 are marked in bold.

[2]OR adjusted for age, family history of breast cancer in first-degree female relatives, oral contraceptive use, menopausal status combined with postmenopausal hormone therapy use, body mass index, smoking status, parity, age at first full-term pregnancy, age at menarche, and breastfeeding.

10. Publications

Parts of this thesis have previously been published:

Torres, D.†, Lorenzo Bermejo, J.†, **Garcia Mesa, K.†**, Gilbert, M., Briceno, I., Pohl Zeidler, S., Gonzalez Silos, R., Boekstegers, F., Plass, C. and Hamann, U. (2018). Interaction between genetic ancestry and common breast cancer susceptibility variants in Colombian women. Int J Cancer, doi: 10.1002/ijc.32023.

†: Contributed equally to this work

Other publications:

Bermejo, J. L.†, Huang, G.†, Manoochehri, M.†, **Mesa, K. G.**, Schick, M., Silos, R. G., Ko, Y. D., Bruning, T., Brauch, H., Lo, W. Y., Hoheisel, J. D. and Hamann, U. (2018). Long intergenic noncoding RNA 299 methylation in peripheral blood is a biomarker for triple-negative breast cancer. Epigenomics, doi: 10.2217/epi-2018-0121.

†: Contributed equally to this work

Garcia Mesa, K., Sistachs, V. , Marrero A., et al. (2013) Stochastic Model of the propagation of the spinocerebellar ataxia type 2 SCA-2 in: Mathematical modelling of phenomenon of environment and health. Volume 2, 123-130 ISBN: 84-695-6649-0, 123-130.

Conference Abstracts:

Garcia Mesa, K., Torres, D., Lorenzo Bermejo, J., Gilbert, M., Briceno, I., Pohl-Zeidler, S., Gonzalez Silos, R., Boekstegers, F., Plass, C. and Hamann, U. Interaction between genetic ancestry and common susceptibility variants in Colombian breast cancer patients, Deutsches Krebsforschungszentrum PhD Poster Presentation, January 2018, Heidelberg

Garcia Mesa, K., Torres, D., Lorenzo Bermejo, J., Gilbert, M., Briceno, I., Pohl-Zeidler, S., Gonzalez Silos, R., Boekstegers, F., Plass, C. and Hamann, U. Breast cancer susceptibility variants in Colombia, Workshop Biometrische Aspekte der Genomanalyse, March 2017, Heidelberg

Garcia Mesa, K., Torres, D., Lorenzo Bermejo, J., Gilbert, M., Briceno, I., Pohl-Zeidler, S., Gonzalez Silos, R., Boekstegers, F., Plass, C. and Hamann, U. Hereditary Breast Cancer in Colombia, Seminar Functional and Structural Genomics, September 2016, Heidelberg

11. Curriculum Vitae

PERSONAL DETAILS

Contact details	Stauffenbergstr. 2 70806 Kornwestheim, Germany +491764380886 karen.garciamesa@gmail.com
Date of birth and place	13.04.1989, Havana

PROFESSIONAL EXPERIENCE

Since 2016	**German Cancer Research Center DKFZ (Heidelberg, Germany)** PhD student with a scholarship from the Helmholtz Graduate School • Statistical analysis of genetic and epigenetic data for the diagnosis and treatment of breast cancer
07/2015-10/2015	**Institute of Molecular Biology (Mainz, Germany)** Guest researches with scholarship from the German Academic Exchange Service (DAAD) • Inference of gene regulatory networks
09/2011 - 09/2014	**Institute of Neurology und Neurosurgery (Havana, Cuba)** Data Scientist • Leader of the investigation project “Machine learning methods for sleep stage classification”

EDUCATION

01/2016 - 01/2019	**Dr. sc. hum, Ruprecht Karls University of Heidelberg**
09/2011 - 09/2013	**M. Sc. Mathematics, Probabilities and Statistic, University of Havana**
09/2007 - 06/2011	**B. Sc. Mathematics, University of Havana**
09/2004 – 06/2007	**Raul-Cepero-Bonilla-High School, Havana** Major Academic Course: Informatics

LANGUAGE SKILLS

Spanish	Mother tongue
English	Excellent
German	Excellent

IT SKILLS

Programming Languages	R, Matlab, SAS, C/C++, Python, Java
Operating Systems	Linux, Windows, Mac OS X
Microsoft Office Programs	MS Office, Latex, OpenOffice
Others	SQL, Mendel, Maple, STATA,STATISTICA,

AWARDS

12/2016	Scholarship from the Helmholtz Graduate School
07/2015	Scholarship from the German Academic Exchange Service (DAAD)

PUBLICATIONS

2018	Torres, D.†, Lorenzo Bermejo, J.†, **Garcia Mesa, K.**†, Gilbert, M., Briceno, I., Pohl-Zeidler, S., Gonzalez Silos, R., Boekstegers, F., Plass, C. and Hamann, U. (2018). **Interaction between genetic ancestry and common breast cancer susceptibility variants in Colombian women**. Int J Cancer, doi: 10.1002/ijc.32023.
2018	Bermejo, J. L., Huang, G., Manoochehri, M., **Mesa, K. G.**, Schick, M., Silos, R. G., Ko, Y. D., Bruning, T., Brauch, H., Lo, W. Y., Hoheisel, J. D. and Hamann, U. (2018). Long intergenic noncoding RNA 299 methylation in peripheral blood is a biomarker for triple-negative breast cancer. Epigenomics, doi: 10.2217/epi-2018-0121.
2013	**Garcia Mesa, K.**, Sistachs, V., Marrero A., et al.(2013) Stochastic Model of the propagation of the spinocerebellar ataxia type 2 SCA-2 in: Mathematical modelling of phenomenon of environment and health. Volume 2, 123-130 ISBN: 84-695-6649-0, 123-130.

CONFERENCES

01/2018	„Interaction between genetic ancestry and common susceptibility variants in Colombian breast cancer patients“, Deutsches Krebsforschungszentrum DKFZ PhD Poster Presentation, Heidelberg
03/2017	„Breast cancer susceptibility variants in Colombia“, Workshop Biometrische Aspekte der Genomanalyse, Heidelberg
07/2016	„Hereditary Breast Cancer in Colombia“, Seminar Functional and Structural Genomics, Heidelberg

VOLUNTEER EXPERIENCE

Since 2017	Volunteer in the German Federal Agency for Technical Relief (THW)
Since 2011	Founding member and volunteer in Asociación Cubana para el Desarrollo de la Educación Infantil (ACDEI)

Karen Garcia Mesa
Heidelberg, 29 March 2019

12. Acknowledgements

This PhD thesis has been performed within the framework of the Col-BCCC study and conducted at the Deutsches Krebsforschungszentrum (DKFZ) in Heidelberg (Director: Prof. Dr. Michael Baumann). The research protocol was approved by the Ethics Committee of the Pontificia Universidad Javeriana in Bogota, Colombia. The thesis was supervised and coordinated by Prof. Ute Hamann and Prof. Justo Lorenzo Bermejo.

I would like to express my sincere gratitude to my mentors, Prof. Ute Hamann and Prof. Justo Lorenzo Bermejo, who gave me the opportunity to join their team. Their continuous support of my PhD study and related research allowed me to embark on a highly competitive research area and to report the end results in two manuscripts.

I thank my colleagues in the B070 and B072 departments. In particular, I am grateful to Michael Gilbert for providing me with all the necessary information. My sincere thanks also goes to my friends Felix Boekstegers, Rosa Gonzalez, Regina Brinster, Carol Barahona, Heike Deutelmoser und Blanca Flores, for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last three years. Nevertheless, I am also grateful to Ofure Obazae for her very valuable comments on this thesis. Last but not the least; I would like to thank my family: my parents, my sister, and my boyfriend for supporting me spiritually throughout writing this thesis and my life in general.

13. Eidesstattliche Erklärung

EIDESSTATTLICHE ERKLÄRUNG

zum Antrag auf Zulassung zur Promotion gemäß PromO „Dr.sc.hum."

AFFIDAVIT

1. Ich habe an keiner anderen Stelle einen Antrag auf Zulassung zur Promotion gestellt oder bereits einen Doktortitel auf der Grundlage des vorgelegten Studienabschlusses erworben und mich auch nicht einer Doktorprüfung erfolglos unterzogen (dies schließt äquivalente Verfahren bzw. Titel ausserhalb Deutschlands ein). *I have not applied anywhere else for a doctoral degree nor have I obtained a doctoral title on the basis of my presented studies or failed a doctoral examination (this includes similar procedures and titles in countries other than Germany).*

2. Die an der Medizinischen Fakultät der Universität Heidelberg zur Promotion eingereichte Arbeit mit dem Titel: *This thesis, submitted to the Medical Faculty of the University of Heidelberg is entitled:*

Interaction between genetic ancestry and common susceptibility variants in Colombian breast cancer

in der Klinik/am Institut für Deutschen Krebsforschungszentrum (DKFZ)
in the clinic/at the insitute of

unter Anleitung von (Doktorvater/ -mutter) Prof. Dr. Ute Hamann
under the supervision of (name of offical supervisor)

habe ich selbst verfasst und bei der Abfassung der Arbeit keine anderen als die in der Abhandlung aufgeführten Hilfsmittel benutzt. *I have written this thesis independently and I have not used any sources other than indicated in the thesis.*

3. Die Arbeit oder Teile davon habe ich bislang an keiner Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt. *I have not yet presented this thesis or parts thereof to a university as part of an examination or degree.*

4. Die Dissertation wurde ohne Hinzuziehung einer kommerziellen Promotionsberatung erstellt. *This thesis was written without the assistance of any kind of commercial doctoral consulting agency.*

5. Mit der Veröffentlichung meines Lebenslaufes im Rahmen des Promotionsverfahrens (Dissertation) bin ich einverstanden. *I agree to the publication of my CV as part of my thesis.*

6. Ich komme der Veröffentlichungspflicht gemäß § 13 PromO nach und stimme der Veröffentlichung der Zusammenfassung meiner Dissertation im Internet unter Angabe meines Namens und des Studienabschlusses zu. *I will fulfill the publication requirement according to § 13 of the doctoral regulations and respectively agree to the publication of the summary of my thesis on the internet, quoting my name and the type of degree conferred.*

7. Die Bedeutung der eidesstattlichen Erklärung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Erklärung sind mir bekannt. *I am aware of the importance of a sworn affidavit and the prosecution in case of a false or incomplete affidavit.*

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.
I declare in lieu of oath that to the best of my knowledge, all the declarations are true and that I have not concealed anything.

Stuttgart, 01.03.2020
Ort und Datum / *place and date*

Unterschrift / *signature*

The German text is legally binding.