HHS Public Access

Author manuscript

Breast Cancer Res Treat. Author manuscript; available in PMC 2015 July 10.

Published in final edited form as:

Breast Cancer Res Treat. 2015 January ; 149(1): 181–190. doi:10.1007/s10549-014-3203-4.

Single-nucleotide polymorphisms in DNA bypass polymerase genes and association with breast cancer and breast cancer subtypes among African Americans and Whites

Leila Family,
Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

Jeannette T. Bensen,
Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA. Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA

Melissa A. Troester,
Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA. Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA

Michael C. Wu,
Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA. Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Carey K. Anders, and
Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA. Division of Hematology-Oncology, Department of Medicine, University of North Carolina, Chapel Hill, NC, USA

Andrew F. Olshan
Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA. Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA

Leila Family: frida2ys@gmail.com

Abstract

DNA damage recognition and repair is a complex system of genes focused on maintaining genomic stability. Recently, there has been a focus on how breast cancer susceptibility relates to genetic variation in the DNA bypass polymerases pathway. Race-stratified and subtype-specific logistic regression models were used to estimate odds ratios (ORs) and 95 % confidence intervals

Disclosures None.

Electronic supplementary material The online version of this article (doi:10.1007/s10549-014-3203-4) contains supplementary material, which is available to authorized users.

(CIs) for the association between 22 single-nucleotide polymorphisms (SNPs) in seven bypass polymerase genes and breast cancer risk in the Carolina Breast Cancer Study, a population-based, case–control study (1,972 cases and 1,776 controls). We used SNP-set kernel association test (SKAT) to evaluate the multi-gene, multi-locus (combined) SNP effects within bypass polymerase genes. We found similar ORs for breast cancer with three *POLQ* SNPs (rs487848 AG/AA vs. GG; OR = 1.31, 95 % CI 1.03–1.68 for Whites and OR = 1.22, 95 % CI 1.00–1.49 for African Americans), (rs532411 CT/TT vs. CC; OR = 1.31, 95 % CI 1.02–1.66 for Whites and OR = 1.22, 95 % CI 1.00–1.48 for African Americans), and (rs3218634 CG/CC vs. GG; OR = 1.29, 95 % CI 1.02–1.65 for Whites). These three SNPs are in high linkage disequilibrium in both races. Tumor subtype analysis showed the same SNPs to be associated with increased risk of Luminal breast cancer. SKAT analysis showed no significant combined SNP effects. These results suggest that variants in the *POLQ* gene may be associated with the risk of Luminal breast cancer.

Keywords

Breast cancer; Single-nucleotide polymorphisms; Breast cancer subtypes; DNA bypass; Polymerases; Translesion synthesis (TLS) pathway; Pathway-based analysis

Introduction

The integrity of DNA is constantly threatened by damage from both endogenous and exogenous sources [1]. Unrepaired DNA damage can result in genomic instability, leading to point mutations, deletions and insertions, as well as chromosomal alterations. These defects increase the probability of a hit to an oncogene or tumor suppressor and may lead to carcinogenesis. To maintain genomic integrity, there is an intricate system of damage response mechanisms [2]. Researchers have identified at least 15 different DNA polymerases in humans which are essential for DNA replication, DNA repair, and the tolerance of DNA damage [3].

DNA replicative polymerases which carry out the bulk of DNA synthesis have evolved to be precise and efficient [4]. Despite this high fidelity, a replication error may generate a one-sided double-strand break (DSB) or degrade to a full DSB if it is not repaired prior to initiation of DNA replication [5, 6]. To resume DNA replication at a stalled replication fork, two damage tolerance mechanisms have been proposed [7]. During template switching, synthesis on the undamaged template strands can continue to a limited extent [3, 7, 8]. In contrast, the specialized DNA polymerases that conduct translesion synthesis do not directly repair the damage, but rather bypass the damage to prevent replication fork stalling. Unlike replicative polymerases, bypass polymerases lack 3′ to 5′ exonuclease (proofreading) activity and are able to resume replication without an undamaged template. However, this also contributes to their low fidelity and potential misincorporation of nucleotides [9].

Previous research has shown that mutations in bypass polymerases may be associated with the risk of cancer [10]. *POLH (pol eta)* was shown to be highly efficient in the bypass of UV lesions, such as cyclobutane pyrimidine dimer (CPD) [11]. Germline mutations in *POLH* identified in xeroderma pigmentosum (XP) patients were the first evidence that bypass polymerases may be involved in human cancer [12]. XP is an autosomal recessive genetic

disorder of DNA repair in which individuals are unable to repair damage caused by UV light and thus are at high risk of developing skin cancer [13].

While genetic variation in other DNA repair pathway genes have been studied extensively in association with breast cancer, focus on DNA bypass polymerases has increased more recently. Several studies have evaluated bypass polymerases in association with breast cancer risk [14–19]. Two reports from the NHS (Nurses' Health Study) II cohort evaluated single-nucleotide polymorphisms (SNPs) from 5 bypass polymerase genes [18, 19]. Han et al. reported that 3 SNPs in *POLK* (rs3213801, rs5744533, and rs3756558) were associated with pre-menopausal breast cancer risk (239 cases, 477 controls) ($p < 0.05$) [18]. However, in the study of postmenopausal women (1,145 cases, 1,142 controls), there were no associations with any of the bypass polymerase SNPs [19]. In an in vivo study of breast cancer cells, Yang et al. reported elevated *POLI* expression when exposed to UV radiation [14]. In a gene sequencing study, Wang et al. identified several mutations in *POLB*, including an 87-bp deletion in the catalytic domain of the gene [15]. In two other reports, *POLQ* overexpression in tumors was associated with poor prognosis of breast cancer [17, 20].

We evaluated the association between germline DNA bypass polymerases variants and breast cancer risk in the Carolina Breast Cancer Study (CBCS), a large population-based case–control study with a racially diverse study population and tumor subtype data. This analysis offered a unique opportunity to evaluate both breast cancer subtype and race-specific effects of seven bypass polymerase genes.

Materials and methods

Study population

The CBCS is a population-based case–control study of breast cancer conducted in 24 counties of central and eastern North Carolina and has been described previously [21, 22]. Briefly, rapid case ascertainment was implemented to identify eligible cases from the North Carolina Central Cancer Registry (NCCCR) [23]. Eligible cases included women ages 20–74, living in the selected North Carolina counties during their primary breast cancer diagnosis. There were 2 phases of enrollment: Phase 1 (1993–1996) enrolled only invasive cancers, while Phase 2 (1996–2001) also included women with in situ cancer. Eligible controls were identified using DMV records for women under age 65 and Medicare lists for women ages 65 and older. Controls were frequency matched to cases based on race and age using randomized recruitment to oversample African American and younger women, a subgroup underrepresented in research studies of breast cancer [24]. This study was approved by the Institutional Review Board of the University of North Carolina at Chapel Hill.

Study subjects who met eligibility criteria and provide written informed consent were scheduled for an in-home visit that included a comprehensive interview and specimen collection by a trained study nurse. Permission was obtained from cases to access medical records and tumor tissue. The nurse-administered interview collected information about demographic factors and known and suspected breast cancer risk factors. A 30-mL blood

sample was collected at the end of the nurse visit. Blood samples were collected from 88 % of cases and 90 % of controls. Whites were more likely to provide blood samples than African Americans (88 vs. 83 %), but there were no significant differences in other risk factors for those who provided a blood sample and those who did not [25, 26]. A total of 2,311 cases (894 African American and 1,417 Whites) and 2,022 controls (788 African Americans and 1,234 Whites) were successfully enrolled in the study. This included 862 cases and 790 from Phase 1. The overall response rates for cases and controls were 78 and 57 %, respectively. Other study response rates have been reported previously [26].

SNP selection and genotyping

We selected 30 candidate SNPs in seven bypass polymerase genes which were included as part of a larger genotyping panel. Given cost limitations, we targeted SNPs across a large number of genes in the TLS pathway which met one of two criteria: (1) prior association analysis identification or (2) data from in vitro or in silico reports. These SNPs included non-synonymous missense, regulatory (5′ UTR and 3′ UTR), and intronic variants (including splice SNPs) with a minor allele frequency (MAF) of at least 5 % in African Americans or Whites (Supplemental Table 1)

DNA was extracted from peripheral blood lymphocytes by standard methods using an automated ABI-DNA extractor (Nuclei Acid Purification System, Applied Biosystems, Foster City, CA, USA) [22]. High-throughput genotyping of selected SNPs was conducted as part of a larger set of 1,536 SNPs by the UNC Mammalian Genotyping Core using Illumina GoldenGate assay (Illumina, Inc., San Diego, CA) [27]. Assay intensity data and genotype cluster images for all SNPs were reviewed individually. Overall, 1,373 of 1,536 (89 %) SNPs passed quality control. Out of the genotyped 30 bypass polymerase SNPs, we excluded five SNPs for which genotyping resulted in poor signal intensity or genotyping clustering, as well as, loci that were non-polymorphic overall (rs3730823, rs28382644, rs28382635) or in either race (five SNPs in Whites, two SNPs in African Americans). All SNPs tested were in Hardy–Weinberg Equilibrium (HWE) ($p < 0.05$) (Supplemental Table 2). Our final analysis included genotyped data for 22 SNPs in the bypass polymerase pathway in 1,946 of 2,311 (85 %) cases and 1,747 of 2,022 (88 %) controls. Participants that did not self-identify as White or African American were excluded in the final analysis (N = 26 cases and N = 29 controls).

In addition, 144 ancestry informative markers (AIMs) were genotyped to estimate African and European ancestry using maximum likelihood estimates (MLE). In this study, we selected SNPs with minor allele frequency (MAF) differences of at least 60 % between the HapMap CEU and YRI populations that had high values for Fisher's information criterion, which is the inverse of the maximum likelihood estimation (MLE) of the ancestral proportion [28].

IHC analysis and subtype ascertainment

Immunohistochemical (IHC) markers were used as a surrogate for gene expression based subtyping [29]. IHC staining and scoring procedures have been explained previously in detail [29–32]. Briefly, tumor tissue blocks were used to confirm diagnosis by a pathologist

and to conduct IHC subtyping. Formalin-fixed paraffin-embedded (FFPE) tumor tissue was available 80 % of cases and immunohistochemistry was completed for 62 % of cases. ER/PR status was abstracted from medical records for 80 % of cases while IHC was used for the remaining 20 % of cases. The concordance between these two methods was 81 % [33]. A total of 1,424 (77 % of available tumor blocks) were successfully subtyped and classified into one of five “intrinsic” subtype groups: Luminal A (ER+ and/or PR+, *HER2*−, Luminal B (ER+ and/or PR+, *HER2*+), *HER2*+/ER− (ER−, PR−, *HER2*+), and basal-like (ER−, PR−, *HER2*−, *HER1*+ and/or *CK 5/6*+), with those negative for all 5 markers considered ‘unclassified’ [30].

For the current study, we classified tumors as either Luminal (ER+ and/or PR+; $n = 788$), basal-like (ER−, PR−, *HER2*−, CK 5/6+ and/or EGFR+; $n = 199$) or *HER2*+/ER− ($n = 94$). We excluded ‘unclassified’ tumors from further analysis due to their uncertain status. The major distinction between the two Luminal subtypes are their proliferation signatures, measured by the expression of *CCNB1*, *MKI67*, and *MYBL2* (49). *HER2* expression only identifies about 30 % of Luminal B tumors, and information about proliferation markers was not available in the current study. Therefore Luminal A and B tumors were combined into a single ‘Luminal’ category (48, 49). Additionally, most other studies do not have subtype data available and only have estrogen receptor status data. Therefore, we conducted an additional exploratory analysis using estrogen receptor (ER) status to evaluate comparability to “intrinsic” subtype results. We found that ER positive effects were concordant with Luminal subtype results; while ER negative (ER−) effects correlated with those of basal-like and HER2+/ER− subtypes (Supplemental Table 3). There were no differences between CBCS cases with and without subtyping data in terms of age, menopausal status, or family history.

Statistical analysis

We calculated allele and genotype frequencies stratified by case status and self-reported race (African American or White). We assessed departure from HWE for each locus by comparing expected versus observed genotype frequencies among race-specific (White and African American) controls using χ^2 tests ($p < 0.05$). We calculated pairwise linkage disequilibrium (LD) r^2 using SAS Genetics (version9.1.3) (SAS Institute, Cary, NC) stratified by race.

We used unconditional logistic regression models to estimate odds ratios (ORs) and 95 % confidence intervals (CIs) for race-stratified effects of base excision repair SNPs on breast cancer, based initially on the additive model. We coded genotype as an ordinal variable (0, 1, or 2 for the number of minor alleles carried by the individual). We excluded non-polymorphic SNPs or SNPs with a minor allele frequency of less than 0.05 in either race. Due to large number of rare homozygote variants, the final analysis was conducted using a dominant genetic model. Less than 2 % of participants self-identified as another race and were not included in the final analysis. We adjusted for proportion of African ancestry, as measured with a set of 144 AIMs [34, 35]. Final models were adjusted for age at diagnosis, proportion of African ancestry, and offset term for the sampling design [24].

Subtype analyses

We coded breast cancer subtype as a categorical variable with four levels (control, Luminal, HER2+/ER−, and basal-like). We used unconditional polytomous regression models to estimate ORs and 95 % CI for each subtype compared to controls.

Correction for multiple testing

We used FDR (false discovery rate) correction for multiple testing, following the method of Benjamini and Hochberg [36]. Corrections were based on the number of SNPs tested and were performed separately for African American and Whites in the race-stratified analysis and separately for Luminal, *HER2*+/ER− and basal-like categories in the subtype analysis. Observed *p* values (trend) from the additive model were used to determine *q* values. The *q* value is defined as the minimum FDR that can be attained when calling a SNP significant (i.e., expected proportion of false positives) [37]. *Q* values were computed using the software package R. Statistical significance was set at $q < 0.10$.

Pathway-based analysis

We used SNP-set kernel association test (SKAT) to evaluate the combined effects of the genotyped SNPs in the bypass polymerase pathway [38]. A SNP-set refers to a set of related SNPs that are grouped based on prior biological knowledge. In the current study, a SNP-set was formed based on SNPs in bypass polymerase genes [38]. The formation of SNP-sets harnesses the potential correlation between SNPs to increase power [39]. We chose a linear kernel under the assumption of a log-linear model. Kernel regression methods convert genomic information for a pair of individuals to a kernel score representing either similarity or dissimilarity. When applied to all pairs of the individuals, this information formed a positive semi-definite matrix [40]. We tested the global hypothesis for SNPs in the pathway separately for White and African American participants [38].

Results

Characteristics of the CBCS population with genotyping data are described in Table 1. The distributions of age, proportion of African ancestry, and menopausal status were similar between cases and controls. African American cases were more likely to be diagnosed at a later stage and were more likely to have tumors that were ER negative. African Americans were twice as likely to be classified as having basal-like tumors compared to Whites (22 vs. 11 %).

Genotype associations by race

The race-stratified adjusted ORs for TLS SNPs are summarized in Table 2. Most SNPs did not show a meaningfully increased or decreased odds ratio. However, for both race groups, SNPs in *POLQ* were associated with an increased odds ratio under the dominant genetic model ($p < 0.05$). *POLQ* rs487848 (AG + AA vs. GG) showed a statistically significant (uncorrected) positive association with breast cancer risk in Whites (OR = 1.31; 95 % CI = 1.08, 1.68) and African Americans (OR = 1.22; 95 % CI = 1.00, 1.49). *POLQ* SNP rs532411 (CT + TT vs. CC) was also significantly associated with increased breast cancer among both races (OR = 1.31; 95 % CI = 1.02, 1.66) and (OR = 1.22; 95 % CI = 1.00, 1.48),

respectively. Finally, *POLQ* SNP rs3218634 (CG + CC vs. GG) showed an increased risk in breast cancer in Whites (OR = 1.29; 95 % CI 1.02, 1.65). After adjustment for multiple testing, none of the SNPs remained significant at the 0.10 FDR level (Table 3).

Genotype associations by subtype

In subtype-specific analyses, the 3 *POLQ* SNPs were significantly associated with Luminal breast cancer ($p < 0.05$ without FDR correction): rs487848 AG + AA vs. GG (OR = 1.34, 95 % CI 1.02–1.67); rs532411 CT + TT vs. CC, (OR = 1.33, 95 % CI 1.06–1.65); rs3218634 CG + CC vs. GG, (OR = 1.26, 95 % CI 1.01–1.57). Additionally, another *POLQ* SNP (rs1381057 CT + TT vs. CC) was significantly associated with HER+/ER− breast cancer (OR = 1.44; 95 % CI = 1.06, 1.93). The same set of *POLQ SNPs* was significantly associated with ER+ breast cancer (Supplemental Table 1). However, after FDR adjustment for multiple testing, none of these SNPs were significantly associated with ER+ breast cancer ($q = 0.10$).

Pathway-based analysis

We assessed the global *p* value for two different SNP-sets stratified by race using the SNP-set Kernel Association Test (SKAT), adjusted for AIMs, and offset term. We did not find any significant associations for SNP-sets. A Kernel machine test of no linear effects yielded a global *p* value of 0.40, and 0.54 for African Americans and Whites, respectively (Supplemental Table 4).

Discussion

Given the relatively low fidelity and high mutational potential of bypass polymerases, it was initially hypothesized that SNPs in DNA bypass polymerases may be linked to increased cancer risk. We did not find a consistent pattern of association with breast cancer risk overall or within a given subtype for most SNPs we evaluated. Subsequently, specialized bypass polymerases were shown to bypass lesions in an error-free manner [41–43]. Therefore, functional redundancy in this pathway may weaken associations between specific bypass polymerases SNPs and breast cancer. Indeed, lesion specificity and functional redundancy are both evolutionary tactics which may ensure that genomic integrity is maintained [44].

Despite the weak results for most bypass polymerases, we did observe evidence for both race- and subtype-specific associations between three *POLQ* variants and an increased breast cancer risk. To our knowledge, other studies have not investigated these associations. All of the SNPs showing an association appeared to predict increased risk of Luminal breast cancer. Our power to detect associations in the other subtypes (i.e., basal-like) was limited. Although not statistically significant using the FDR, these findings are suggestive and warrant replication in other studies. Within each race, these three *POLQ* SNPs were in LD with each other making it difficult to identify which, if any SNPs were most likely to have functional effects. Two out of three identified *POLQ* SNPs had a SIFT score 0.05 or less, indicative of being a damaging functional SNP (Supplemental Table 1). However, functional

studies and fine mapping of the region are needed to further hone in on a potential causal variant.

The *POLQ* gene, located at chromosome 3q, is a member of the A Family that encodes the protein polymerase theta. *POLQ* has also been implicated as playing a role in other DNA repair mechanisms such as base excision repair (BER) and crosslink repair [45, 46]. *POLQ* is able to efficiently bypass oxidative DNA lesions such as abasic (AP) sites and thymine glycol in vitro [41, 47–49]. Another lab study showed that *POLQ* successfully extends from mismatches and bases opposite (6–4) photoproducts [41]. On the other hand, *POLQ*-deficient mutants exhibited hypersensitivity to oxidative base damage induced by H_2O_2 [50]. The results of the current study, together with previous experimental evidence, suggest *POLQ* may play an important role in breast cancer risk.

A pair of recent studies has linked *POLQ* overexpression in tumors to breast cancer progression and poorer prognoses [16, 17, 51]. Lemee et al. examined gene expression profiles of tumors from two cohorts of European women with untreated primary breast cancer. Patients' tumor cells that overexpressed *POLQ* had a 4.3-fold increased risk of death compared those with normal expression [17]. Higgins et al. also found elevated levels of *POLQ* expression in breast cancer cells, which was linked to poor prognosis in early breast cancer patients [16]. While these findings emphasize the role of *POLQ* after disease onset, genes that influence progression also have been shown to influence early disease/etiology and therefore these findings also suggest that *POLQ* merits further investigation.

These findings should be considered in light of strengths and limitations. Compared to other genetic association studies of breast cancer, CBCS has a larger proportion of African Americans (over 40 %). In addition, CBCS has detailed subtype data on tumors from a large population-based sample of women allowing a unique investigation of the genetics of specific breast cancer subtypes as well as the ability to extend study results to the population as a whole. Stratification by subtype does reduce power for some race-specific and subtype comparisons, especially for HER2+/ER− and basal-like tumors, which may explain why only associations in Luminal tumors were found. Future research that includes large numbers of breast cancer cases with less common subtypes and focuses on oversampling African American cases should have improved power to more precisely estimate subtype associations, especially among African American women.

We had genotype and subtype data for a large proportion of CBCS participants. Tumor tissue was available for 1,845 of 2,311 cases (80 %) and subtyping using IHC was completed for 1,424 of 2,311 cases (62 %) [33]. A comparison of subtyped and non-subtyped CBCS cases showed that the subtyped cases were not significantly different from the CBCS as a whole with respect to age and menopausal status. However, cases with subtype data were more likely to be African American and to have a later stage at diagnosis, which may bias estimates for SNPs related to race or disease aggressiveness [31].

It is also noteworthy that definitions for Luminal breast cancer have evolved since original CBCS IHC subtyping methods were published [29]. As a result, Luminal breast cancers cannot be divided into finer categories (Luminal A vs. Luminal B) without information on

proliferation markers or PR percentage [52, 53]. Therefore, there is heterogeneity within the group of Luminal breast cancers defined here. Nonetheless, our subtyping methods have the advantage of excluding tumors that were negative for all markers tested. Only triple negatives that were also positive for a basal-like marker are included among basal-like cancers, reducing outcome misclassification potential in this important subgroup.

Although we did not find any significant combined effects of SNPs in the TLS pathway using SKAT, use of kernel-based machine learning to assess pathway effects in breast cancer is an important advance in studying gene–gene interactions [19, 54]. While our pathway analysis was limited by the density of SNP coverage across TLS pathway genes, it is important to understand gene–gene interactions in breast cancer pathways. Future application of SKAT to similar data should consider tag-SNP approaches, which may better capture variation in candidate pathways.

In summary, this study adds important new information on the role of bypass polymerases in breast cancer etiology using tumor tissue to evaluate subtype-specific effects and considers carefully selected regulatory and coding SNP-sets in a biologically established DNA repair pathway. We identified three novel SNPs in the *POLQ* gene not previously associated with breast cancer. Larger studies such as the CBCS Phase 3 with improved power for race- and subtype-specific analyses and collaborative consortia [55, 56] will help gain further insight into the role of genetic variation in the DNA bypass polymerases and the risk of breast cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the participants of the Carolina Breast Cancer Study, the UNC BioSpecimen Processing Facility for our DNA extractions, blood processing, storage, and sample disbursement (https://genome.unc.edu/bsp), the UNC Mammalian Genotyping Core for CBCS sample genotyping (http://mgc.unc.edu), and Jessica Tse and Katie O'Brien for their technical assistance and support. Finally, we would like to acknowledge Dr. Robert Millikan, the previous Principal Investigator of the CBCS, for his guidance and inspiration for this paper. This research was funded in part by the University Cancer Research Fund of North Carolina, the National Cancer Institute Specialized Program of Research Excellence (SPORE) in Breast Cancer (NIH/NCI P50-CA58223).

References

1. Gieseking S, Bergen K, Di Pasquale F, et al. Human DNA polymerase beta mutations allowing efficient abasic site bypass. J Biol Chem. 2011; 286:4011–4020.10.1074/jbc.M110.176826 [PubMed: 21107011]
2. Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. Nature. 2001; 411:366–374.10.1038/35077232 [PubMed: 11357144]
3. Lange SS, Takata K, Wood RD. DNA polymerases and cancer. Nat Rev Cancer. 2011; 11:96–110.10.1038/nrc2998 [PubMed: 21258395]
4. McCulloch SD, Kunkel TA. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. Cell Res. 2008; 18:148–161.10.1038/cr.2008.4 [PubMed: 18166979]
5. Cleaver JE. Mechanisms by which human cells bypass damaged bases during DNA replication after ultraviolet irradiation. Scientific World Journal. 2002; 2:1296–1305.10.1100/tsw.2002.348 [PubMed: 12805912]

6. Li X, Heyer WD. Homologous recombination in DNA repair and DNA damage tolerance. Cell Res. 2008; 18:99–113.10.1038/cr.2008.1 [PubMed: 18166982]
7. Budzowska M, Kanaar R. Mechanisms of dealing with DNA damage-induced replication problems. Cell Biochem Biophys. 2009; 53:17–31.10.1007/s12013-008-9039-y [PubMed: 19034694]
8. Wood RD, Mitchell M, Lindahl T. Human DNA repair genes, 2005. Mutat Res. 2005; 577:275–283.10.1016/j.mrfmmm.2005.03.007 [PubMed: 15922366]
9. Friedberg EC, Gerlach VL. Novel DNA polymerases offer clues to the molecular basis of mutagenesis. Cell. 1999; 98:413–416. [PubMed: 10481906]
10. Albertella MR, Lau A, O'Connor MJ. The overexpression of specialized DNA polymerases in cancer. DNA Repair (Amst). 2005; 4:583–593.10.1016/j.dnarep.2005.01.005 [PubMed: 15811630]
11. Choi JH, Pfeifer GP. The role of DNA polymerase eta in UV mutational spectra. DNA Repair (Amst). 2005; 4:211–220.10.1016/j.dnarep.2004.09.006 [PubMed: 15590329]
12. Johnson RE, Kondratick CM, Prakash S, et al. hRAD30 mutations in the variant form of xeroderma pigmentosum. Science. 1999; 285:263–265. [PubMed: 10398605]
13. Lehmann AR. Replication of UV-damaged DNA: new insights into links between DNA polymerases, mutagenesis and human disease. Gene. 2000; 253:1–12. [PubMed: 10925197]
14. Yang J, Chen Z, Liu Y, et al. Altered DNA polymerase iota expression in breast cancer cells leads to a reduction in DNA replication fidelity and a higher rate of mutagenesis. Cancer Res. 2004; 64:5597–5607.10.1158/0008-5472.CAN-04-0603 [PubMed: 15313897]
15. Wang L, Banerjee S. Mutations in DNA-polymerase-Beta occur in breast, prostate and colorectal tumors. Int J Oncol. 1995; 6:459–463. [PubMed: 21556560]
16. Higgins GS, Harris AL, Prevo R, et al. Overexpression of POLQ confers a poor prognosis in early breast cancer patients. Oncotarget. 2010; 1:175–184. [PubMed: 20700469]
17. Lemee F, Bergoglio V, Fernandez-Vidal A, et al. DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. Proc Natl Acad Sci USA. 2010; 107:13390–13395.10.1073/pnas.0910759107 [PubMed: 20624954]
18. Han J, Haiman C, Niu T, et al. Genetic variation in DNA repair pathway genes and premenopausal breast cancer risk. Breast Cancer Res Treat. 2009; 115:613–622.10.1007/s10549-008-0089-z [PubMed: 18551366]
19. Monsees GM, Kraft P, Chanock SJ, et al. Comprehensive screen of genetic variation in DNA repair pathway genes and postmenopausal breast cancer risk. Breast Cancer Res Treat. 2011; 125:207–214.10.1007/s10549-010-0947-3 [PubMed: 20496165]
20. Higgins MJ, Baselga J. Breast cancer in 2010: novel targets and therapies for a personalized approach. Nat Rev Clin Oncol. 2011; 8:65–66.10.1038/nrclinonc.2010.217 [PubMed: 21278769]
21. Newman B, Moorman PG, Millikan R, et al. The carolina breast cancer study: integrating population-based epidemiology and molecular biology. Breast Cancer Res Treat. 1995; 35:51–60. [PubMed: 7612904]
22. Millikan R, Eaton A, Worley K, et al. HER2 codon 655 polymorphism and risk of breast cancer in African Americans and Whites. Breast Cancer Res Treat. 2003; 79:355–364. [PubMed: 12846420]
23. Aldrich TE, Vann D, Moorman PG, et al. Rapid reporting of cancer incidence in a population-based study of breast cancer: one constructive use of a central cancer registry. Breast Cancer Res Treat. 1995; 35:61–64. [PubMed: 7612905]
24. Weinberg CR, Sandler DP. Randomized recruitment in case-control studies. Am J Epidemiol. 1991; 134:421–432. [PubMed: 1877602]
25. Millikan RC, Player JS, Decotret AR, et al. Polymorphisms in DNA repair genes, medical exposure to ionizing radiation, and breast cancer risk. Cancer Epidemiol Biomark Prev. 2005; 14:2326–2334.10.1158/1055-9965.EPI-05-0186
26. Nyante, SJ. Ph D dissertation. University of North Carolina; Chapel Hill: 2009. Single nucleotide polymorphisms and the etiology of basal-like and luminal A breast cancer: a pathway-based approach.
27. Shen R, Fan JB, Campbell D, et al. High-throughput SNP genotyping on universal bead arrays. Mutat Res. 2005; 573:70–82.10.1016/j.mrfmmm.2004.07.022 [PubMed: 15829238]

28. Barnholtz-Sloan JS, Chakraborty R, Sellers TA, et al. Examining population stratification via individual ancestry estimates versus self-reported race. Cancer Epidemiol Biomark Prev. 2005; 14:1545–1551.10.1158/1055-9965.EPI-04-0832
29. Nielsen TO, Hsu FD, Jensen K, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res. 2004; 10:5367–5374.10.1158/1078-0432.CCR-04-0220 [PubMed: 15328174]
30. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA. 2006; 295:2492–2502.10.1001/jama.295.21.2492 [PubMed: 16757721]
31. Millikan RC, Newman B, Tse CK, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat. 2008; 109:123–139. [PubMed: 17578664]
32. Huang WY, Newman B, Millikan RC, et al. Hormone-related factors and risk of breast cancer in relation to estrogen receptor and progesterone receptor status. Am J Epidemiol. 2000; 151:703–714. [PubMed: 10752798]
33. Nyante SJ, Gammon MD, Kaufman JS, et al. Common genetic variation in adiponectin, leptin, and leptin receptor and association with breast cancer subtypes. Breast Cancer Res Treat. 2011; 129:593–606.10.1007/s10549-011-1517-z [PubMed: 21516303]
34. Barnholtz-Sloan JS, McEvoy B, Shriver MD, et al. Ancestry estimation and correction for population stratification in molecular epidemiologic association studies. Cancer Epidemiol Biomark Prev. 2008; 17:471–477.10.1158/1055-9965.EPI-07-0491
35. Pfaff CL, Barnholtz-Sloan J, Wagner JK, et al. Information on ancestry from genetic markers. Genet Epidemiol. 2004; 26:305–315.10.1002/gepi.10319 [PubMed: 15095390]
36. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. JSTOR. 1995; 57:289.
37. Storey JD, Tibshirani R. Statistical methods for identifying differentially expressed genes in DNA microarrays. Methods Mol Biol. 2003; 224:149–157.10.1385/1-59259-364-X:149 [PubMed: 12710672]
38. Wu MC, Kraft P, Epstein MP, et al. Powerful SNP-set analysis for case-control genome-wide association studies. Am J Hum Genet. 2010; 86:929–942.10.1016/j.ajhg.2010.05.002 [PubMed: 20560208]
39. Li H. U-statistics in genetic association studies. Hum Genet. 201210.1007/s00439-012-1178-y
40. Schaid DJ. Genomic similarity and Kernel methods I: advancements by building on mathematical and statistical foundations. Hum Hered. 2010; 70:109–131.10.1159/000312641 [PubMed: 20610906]
41. Seki M, Wood RD. DNA polymerase theta (POLQ) can extend from mismatches and from bases opposite a (6–4) photoproduct. DNA Repair (Amst). 2008; 7:119–127.10.1016/j.dnarep.2007.08.005 [PubMed: 17920341]
42. Choi JY, Lim S, Kim EJ, et al. Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases alpha, delta, eta, iota, kappa, and REV1. J Mol Biol. 2010; 404:34–44.10.1016/j.jmb.2010.09.015 [PubMed: 20888339]
43. Johnson RE, Haracska L, Prakash S, et al. Role of DNA polymerase zeta in the bypass of a (6–4) TT photoproduct. Mol Cell Biol. 2001; 21:3558–3563.10.1128/MCB.21.10.3558-3563.2001 [PubMed: 11313481]
44. Wang Z. DNA damage-induced mutagenesis : a novel target for cancer prevention. Mol Interv. 2001; 1:269–281. [PubMed: 14993366]
45. Shcherbakova PV, Bebenek K, Kunkel TA. Functions of eukaryotic DNA polymerases. Sci Aging Knowl Environ. 2003:RE3.
46. Muzzini DM, Plevani P, Boulton SJ, et al. Caenorhabditis elegans POLQ-1 and HEL-308 function in two distinct DNA interstrand cross-link repair pathways. DNA Repair (Amst). 2008; 7:941–950.10.1016/j.dnarep.2008.03.021 [PubMed: 18472307]
47. Hogg M, Seki M, Wood RD, et al. Lesion bypass activity of DNA polymerase theta (POLQ) is an intrinsic property of the pol domain and depends on unique sequence inserts. J Mol Biol. 2011; 405:642–652.10.1016/j.jmb.2010.10.041 [PubMed: 21050863]

48. Yoon JH, Bhatia G, Prakash S, et al. Error-free replicative bypass of thymine glycol by the combined action of DNA polymerases kappa and zeta in human cells. Proc Natl Acad Sci USA. 2010; 107:14116–14121.10.1073/pnas.1007795107 [PubMed: 20660785]

49. Prasad R, Longley MJ, Sharief FS, et al. Human DNA polymerase theta possesses 5′-dRP lyase activity and functions in single-nucleotide base excision repair in vitro. Nucleic Acids Res. 2009; 37:1868–1877.10.1093/nar/gkp035 [PubMed: 19188258]

50. Yoshimura M, Kohzaki M, Nakamura J, et al. Vertebrate POLQ and POLbeta cooperate in base excision repair of oxidative DNA damage. Mol Cell. 2006; 24:115–125.10.1016/j.molcel.2006.07.032 [PubMed: 17018297]

51. Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. Science. 2006; 314:268–274.10.1126/science.1133427 [PubMed: 16959974]

52. Prat A, Cheang MC, Martin M, et al. Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal a breast cancer. J Clin Oncol. 2013; 31:203–209.10.1200/JCO.2012.43.4134 [PubMed: 23233704]

53. Cheang MC, Chia SK, Voduc D, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst. 2009; 101:736–750.10.1093/jnci/djp082 [PubMed: 19436038]

54. Kazma R, Babron MC, Gaborieau V, et al. Lung cancer and DNA repair genes: multilevel association analysis from the International Lung Cancer Consortium. Carcinogenesis. 2012; 33:1059–1064.10.1093/carcin/bgs116 [PubMed: 22382497]

55. McGee SA, Durham DD, Tse CK, et al. Determinants of breast cancer treatment delay differ for African American and White women. Cancer Epidemiol Biomark Prev. 2013; 22:1227–1238.10.1158/1055-9965.EPI-12-1432

56. Palmer JR, Ambrosone CB, Olshan AF. A collaborative study of the etiology of breast cancer subtypes in African American women: the AMBER consortium. Cancer Causes Control. 2014; 25:309–319.10.1007/s10552-013-0332-8 [PubMed: 24343304]

Table 1

Characteristics of CBCS participants with genotyping data

Characteristic	Cases				Cases			
	African American		White		African American		White	
	N	%	N	%	N	%	N	%
Total (N)	742	100	1,204	100	658	100	1,089	100
Average age at selection	52		52		52		53	
Average proportion of African ancestry	0.78		0.064		0.774		0.066	
Menopausal status								
Pre-menopausal	324	41.36	539	30.53	290	59.09	456	62.67
Postmenopausal	418	58.64	665	69.47	368	40.91	633	37.33
Stage								
CIS[a]	88	11.90	349	29				
1	216	29.10	393	32.6				
2	299	40.30	328	27.2				
3	76	10.20	68	5.7				
4	27	3.60	15	1.3				
Missing	36	4.90	51	4.2				
Subtype								
Luminal	269	56.31	519	75.63				
HER2+/ER−	38	7.59	56	5.82				
Basal-like	108	21.98	91	10.68				
Estrogen receptor (ER) status								
Positive	235	49.57	482	68.91				
Negative	251	50.43	242	31.09				

[a]Carcinoma in situ

Table 2

Association between selected bypass polymerase SNPs and breast cancer by race in the CBCS

Gene	SNP	Genotype	White					African Americans				
			Cases (N)	Controls (N)	OR (95 % CI)	*p* value	*q* value	Cases (N)	Controls (N)	OR (95 % CI)	*p* value	*q* value
POLE	rs9333555	AA	1,142	1,034	Referent			733	653	Referent		
		AG + GG	62	55	0.94 (0.64, 1.38)	0.7398	0.8088	9	5	1.92 (0.62, 6.01)	0.2601	0.6134
	rs6899628	CC	1,117	1,007	Referent			336	275	Referent		
		CT + TT	87	82	0.92 (0.65, 1.29)	0.2778	0.716	405	383	0.96 (0.81, 1.13)	0.5979	0.7972
POLI	rs3218786	TT	1,120	1,032	Referent			732	650	Referent		
		CT + CC	84	57	1.41 (0.98, 2.03)	0.065	0.2763	10	8	1.04 (0.39, 2.76)	0.9358	0.9943
POLL	rs8305	AA	599	543	Referent			657	573	Referent		
		AG + GG	605	546	1.06 (0.93, 1.21)	0.3791	0.7161	85	85	0.84 (0.61, 1.16)	0.2876	0.6134
	rs3730477	CC	739	659	Referent			662	565	Referent		
		CT + TT	464	430	0.93 (0.80, 1.08)	0.3286	0.7161	80	93	0.79 (0.57, 1.08)	0.1438	0.6134
	rs3730475	TT	613	548	Referent			511	430	Referent		
		CT + CC	591	541	0.98 (0.85, 1.12)	0.7612	0.8088	231	228	0.91 (0.74, 1.12)	0.3761	0.6134
POLM	rs3730463	AA	1,041	950	Referent			599	525	Referent		
		AC + CC	163	139	1.12 (0.88, 1.43)	0.3416	0.7161	143	133	1.00 (0.78, 1.28)	0.9802	0.9943
	rs28382653	GG	1,204	1,086	Referent			670	592	Referent		
		GT + TT	0	3	N/A	N/A		72	66	1.03 (0.73, 1.46)	0.8481	0.9943
POLQ	rs487848	GG	1,024	957	Referent			469	438	Referent		
		AG + AA	180	132	1.31 (1.03, 1.68)	**0.0279**	0.2131	273	220	1.22 (1.00, 1.49)	**0.0487**	0.482
POLQ	rs3218651	AA	840	771	Referent			577	521	Referent		
		AG + GG	364	318	0.96 (0.82, 1.13)	0.6463	0.8088	165	137	1.13 (0.88, 1.43)	0.3411	0.6134
	rs532411	CC	1,024	956	Referent			469	438	Referent		
		CT + TT	180	133	1.31 (1.02, 1.66)	**0.0318**	0.2131	272	220	1.22 (1.00, 1.48)	**0.0484**	0.482
	rsl381057	CC	575	534	Referent			298	276	Referent		
		CT + TT	629	555	1.02 (0.90, 1.17)	0.7355	0.8088	444	382	1.10 (0.94, 1.29)	0.2464	0.6134
	rs3218634	GG	1,024	955	Referent			30	15	Referent		
		CG + CC	180	134	1.29 (1.02, 1.65)	**0.0376**	0.2131	712	643	1.20 (0.98, 1.47)	0.0723	0.482
POLQ	rs3218637	GG	1,202	1,089	Referent			672	586	Referent		

Gene	SNP	Genotype	White					African Americans				
			Cases (N)	Controls (N)	OR (95 % CI)	*p* value	*q* value	Cases (N)	Controls (N)	OR (95 % CI)	*p* value	*q* value
		AG + AA	2	0	N/A	N/A		69	72	0.88 (0.62, 1.24)	0.4559	0.6513
	rs3218649	GG	497	457	Referent			221	201	Referent		
		CG + CC	704	632	1.01 (0.89, 1.15)	0.8453	0.8453	521	457	1.08 (0.92, 1.26)	0.3358	0.6134
REV1L	rs3087403	GG	623	543	Referent			377	348	Referent		
		AG + AA	580	546	0.95 (0.83, 1.09)	0.4451	0.7424	365	310	1.09 (0.92, 1.29)	0.3406	0.6134
	rs3087386	CC	385	329	Referent			383	329	Referent		
		CT + TT	819	760	0.96 (0.85, 1.09)	0.5317	0.7532	359	329	0.98 (0.83, 1.17)	0.8511	0.9943
	rs3087399	**AA**	915	843	Referent			411	360	Referent		
		AG + GG	306	259	1.07 (0.89, 1.29)	0.4804	0.7424	331	298	1.00 (0.84, 1.19)	0.9943	0.9943
REV3L	rs458017	TT	1,054	945	Referent			684	595	Referent		
		CT + CC	150	144	0.87 (0.68, 1.11)	0.2642	0.7161	58	63	0.83 (0.58, 1.20)	0.3222	0.6134
	rsl7539651	CC	1,201	1,085	Referent			582	508	Referent		
		CT + TT	3	4	N/A	N/A		160	150	0.90 (0.70, 1.15)	0.3987	0.6134

Bold values indicate the significance level was set at alpha less than or equal to 0.05

[a] Odds ratio and 95 % confidence interval, adjusted for age, African ancestry, and offset term

[b] *P* value is unadjusted for multiple comparisons

[c] *q* value is adjusted for multiple comparisons (false discovery rate; FDR)

[d] N/A indicates non-polymorphic SNP

Table 3

Association between selected bypass polymerase SNPs and breast cancer by subtype in the CBCS

Gene	SNP	Luminal				HER2+/ER−			Basal-like		
		Genotype	OR (95 % CI)[a]	*p* value[b]	*q* value[c]	OR (95 % CI)	*p* value	*q* value	OR (95 % CI)	*p* value	*q* value
POLH	rs6899628	**CC**	Referent			Referent			Referent		
		CT + TT	0.94 (0.74, 1.21)	0.6402	0.8730	1.00 (0.57, 1.77)	0.9991	0.9991	0.80 (0.55, 1.17)	0.2550	0.4250
POLL	rs8305	**AA**	Referent			Referent			Referent		
		AG + GG	0.99 (0.81, 1.20)	0.8863	0.8863	1.22 (0.76, 1.97)	0.4115	0.8763	0.91 (0.63, 1.30)	0.5897	0.6719
	rs3730477	**CC**	Referent			Referent			Referent		
		CT + TT	0.98 (0.80, 1.20)	0.8344	0.8863	0.72 (0.43, 1.21)	0.2144	0.6432	0.79 (0.54, 1.15)	0.2168	0.4250
	rs3730475	TT	Referent			Referent			Referent		
		CT + CC	0.97 (0.81, 1.17)	0.7731	0.8863	0.68 (0.43, 1.06)	0.8670	0.9389	0.93 (0.68, 1.26)	0.6271	0.6719
POLM	rs3730463	**AA**	Referent			Referent			Referent		
		AC + CC	0.92 (0.71, 1.18)	0.4994	0.8730	0.86 (0.47, 1.58)	0.6358	0.9166	1.34 (0.92, 1.96)	0.1236	0.4250
POLQ	rs487848	**GG**	Referent			Referent			Referent		
		AG + AA	1.34 (1.02, 1.67)	**0.0096**	0.1440	1.21 (0.72, 2.02)	0.4753	0.8763	1.23 (0.86, 1.76)	0.2492	0.4250
	rs3218651	**AA**	Referent			Referent			Referent		
		AG + GG	1.16 (0.65, 2.07)	0.6258	0.8730	1.18 (0.28, 5.01)	0.8215	0.9289	1.10 (0.38, 3.15)	0.2492	0.4250
	rs532411	**CC**	Referent			Referent			Referent		
		CT + TT	1.33 (1.06, 1.65)	**0.0118**	0.3918	1.20 (0.72, 2.02)	0.4790	0.8763	1.23 (0.86, 1.76)	0.2487	0.4250
	rsl381057	**CC**	Referent			Referent			Referent		
		CT + TT	1.11 (0.97, 1.27)	0.1306	0.3918	1.44 (1.06, 1.93)	**0.0193**	0.2895	1.10 (0.88, 1.38)	0.3890	0.5305
	rs3218634	**GG**	Referent			Referent			Referent		
		CG + CC	1.26 (1.01, 1.57)	**0.0376**	0.2820	1.18 (0.71, 1.98)	0.5258	0.8763	1.24 (0.87, 1.77)	0.2331	0.4250
	rs3218649	**GG**	Referent			Referent			Referent		
		CG + CC	1.08 (0.95, 1.23)	0.2505	0.6263	1.27 (0.94, 1.71)	0.1235	0.6308	1.04 (0.84, 1.30)	0.7263	0.7263
REV1L	rs3087403	**GG**	Referent			Referent			Referent		
		AG + AA	1.01 (0.88, 1.16)	0.8611	0.8863	0.78 (0.55, 1.11)	0.1682	0.6308	1.22 (0.97, 1.53)	0.0931	0.4250
	rs3087386	**CC**	Referent			Referent			Referent		
		CT + TT	0.95 (0.83, 1.08)	0.4133	0.8730	1.25 (0.93, 1.69)	0.1404	0.6308	0.86 (0.68, 1.07)	0.1771	0.4250
	rs3087399	**AA**	Referent			Referent			Referent		

Gene	SNP	Luninal				HER2+/ER−			Basal-like		
		Genotype	OR (95 % CI)[a]	*p* value[b]	*q* value[c]	OR (95 % CI)	*p* value	*q* value	OR (95 % CI)	*p* value	*q* value
REV3L	rs458017	AG + GG	1.05 (0.87, 1.28)	0.6069	0.8730	0.91 (0.57, 1.46)	0.7042	0.9166	0.84 (0.60, 1.17)	0.3085	0.4626
		TT	Referent			Referent			Referent		
		CT + CC	0.76 (0.57, 1.02)	0.0651	0.3255	1.11 (0.60, 2.05)	0.7333	0.9166	0.84 (0.52, 1.38)	0.4933	0.6166

Bold values indicate the significance level was set at alpha less than or equal to 0.05

[a] Odds ratio and 95 % confidence interval, adjusted for age, African ancestry, and offset term

[b] *P* value is unadjusted for multiple comparisons

[c] *q* value is adjusted for multiple comparisons (false discovery rate; FDR)