Cancer Causes Control (2015) 26:121–131
DOI 10.1007/s10552-014-0491-2

ORIGINAL PAPER

Genetic variation in estrogen and progesterone pathway genes and breast cancer risk: an exploration of tumor subtype-specific effects

Sarah J. Nyante · Marilie D. Gammon · Jay S. Kaufman · Jeannette T. Bensen · Dan Yu Lin · Jill S. Barnholtz-Sloan · Yijuan Hu · Qianchuan He · Jingchun Luo · Robert C. Millikan

Received: 31 December 2013 / Accepted: 4 November 2014 / Published online: 25 November 2014

Abstract
Purpose To determine whether associations between estrogen pathway-related single nucleotide polymorphisms (SNPs) and breast cancer risk differ by molecular subtype, we evaluated associations between SNPs in cytochrome P450 family 19 subfamily A polypeptide 1 (*CYP19A1*), estrogen receptor (*ESR1*), 3-beta hydroxysteroid dehydrogenase type I (*HSD3B1*), 17-beta hydroxysteroid dehydrogenase type II (*HSD17B2*), progesterone receptor (*PGR*), and sex hormone-binding globulin (*SHBG*) and breast cancer risk in a case–control study in North Carolina.
Methods Cases ($n = 1{,}972$) were women 20–74 years old and diagnosed with breast cancer between 1993 and 2001. Population-based controls ($n = 1{,}776$) were frequency matched to cases by age and race. A total of 195 SNPs were genotyped, and linkage disequilibrium was evaluated using the r^2 statistic. Odds ratios (ORs) and 95 % confidence intervals (CIs) for associations with breast cancer overall and by molecular subtype were estimated using logistic regression. Monte Carlo methods were used to control for multiple comparisons; two-sided p values $<3.3 \times 10^{-4}$ were statistically significant. Heterogeneity tests comparing the two most common subtypes, luminal A ($n = 679$) and basal-like ($n = 200$), were based on the Wald statistic.
Results *ESR1* rs6914211 (AA vs. AT+TT, OR 2.24, 95 % CI 1.51–3.33), *ESR1* rs985191 (CC vs. AA, OR 2.11, 95 % CI 1.43–3.13), and *PGR* rs1824128 (TT+GT vs. GG, OR 1.33, 95 % CI 1.14–1.55) were associated with risk after accounting for multiple comparisons. Rs6914211 and rs985191 were in strong linkage disequilibrium among controls (African-Americans $r^2 = 0.70$; whites $r^2 = 0.95$). There was no evidence of heterogeneity between luminal A and basal-like subtypes, and the three SNPs were also

Robert C. Millikan: Deceased.

Electronic supplementary material The online version of this article (doi:10.1007/s10552-014-0491-2) contains supplementary material, which is available to authorized users.

S. J. Nyante (✉)
Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
e-mail: nyantesj@mail.nih.gov

M. D. Gammon · J. T. Bensen · R. C. Millikan
Epidemiology Department, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

M. D. Gammon · J. T. Bensen · D. Y. Lin · J. Luo · R. C. Millikan
Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

J. S. Kaufman
Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada

D. Y. Lin · Y. Hu
Biostatistics Department, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

J. S. Barnholtz-Sloan
Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH, USA

Q. He
Fred Hutchinson Cancer Research Center, Seattle, WA, USA

associated with elevated risk of the less common luminal B, HER2+/ER−, and unclassified subtypes.
Conclusions *ESR1* and *PGR* SNPs were associated with risk, but lack of heterogeneity between subtypes suggests variants in hormone-related genes may play similar roles in the etiology of breast cancer molecular subtypes.

Keywords Breast cancer · Single nucleotide polymorphisms · Estrogen receptor · Progesterone receptor · Cytochrome P450 family 19 subfamily A polypeptide 1 · 17-Beta hydroxysteroid dehydrogenase type II · Sex hormone-binding globulin · 3-Beta hydroxysteroid dehydrogenase type I

Introduction

Breast cancer can be classified into molecular subtypes that differ with respect to biology, survival, and some risk factors [1–6]. Although more than 70 single nucleotide polymorphisms (SNPs) have been associated with breast cancer risk [7, 8], there have been few investigations of genetic variation and risk of specific breast tumor molecular subtypes. A pooled analysis examined 10 risk loci and demonstrated heterogeneity of risk associations by subtype [9]. More recently, the Carolina Breast Cancer Study (CBCS) demonstrated that subtype associations for 78 SNPs in several susceptibility pathways varied by tumor subtype [10]. Beyond these two studies, information regarding the heterogeneity of association between genetic variants and risk of breast cancer subtypes is limited. Greater understanding of the relationship between genetic polymorphisms and specific tumor types may provide clues regarding which biological mechanisms are involved in the tumorigenesis of different subtypes.

Estrogen and progesterone-related genes are likely candidates for harboring breast cancer risk variants due to the central role that estrogen and progesterone play in normal breast development and breast cancer [11–13] and may provide clues as to how the estrogen pathway functions in tumors. Estrogen bound to the estrogen receptor (ER), encoded by *ESR1*, can act as a transcription factor for genes associated with cell proliferation and survival, including growth factors, tumor suppressors, and pro-apoptotic genes [14]. Progesterone signaling via the progesterone receptor (PR), encoded by *PGR*, also influences transcription and can act through multiple pathways to stimulate proliferation [15, 16].

Previous analyses of the relationship between non-genetic risk factors and breast cancer subtypes in the CBCS indicate that there may be etiologic heterogeneity between breast cancer subtypes [4]. Millikan et al. [4] previously reported that associations for some reproductive risk factors were qualitatively different between luminal A breast cancers, which are defined in part by ER and/or PR expression, and basal-like breast cancers, which are defined in part by lack of ER and PR expression. Specifically, parity (vs. nulliparity) was inversely associated with luminal A breast cancer regardless of breastfeeding status, whereas parity in women who breastfed was unassociated with the basal-like subtype, and parity in women who did not breastfeed was positively associated with basal-like breast cancer [4]. Additionally, earlier age at first birth was associated with a lower risk of luminal A, but increased risk of basal-like breast cancer [4]. The fact that these differences were observed among reproductive risk factors suggests that this heterogeneity may have a hormonal link.

We hypothesized that hormonal factors related to parity and lactation might influence pathways leading to luminal A and basal-like breast cancer in different ways and that some of these differences may be rooted in genetic variation. Therefore, we examined associations between SNPs in six genes related to the production, bioavailability, and signaling of estrogen and progesterone and breast cancer risk. In addition to *ESR1* and *PGR*, we examined genes for cytochrome P450 family 19 subfamily A polypeptide 1 (*CYP19A1*), which converts androgens to estrogens [17]; 17-beta hydroxysteroid dehydrogenase type II (*HSD17B2*), which oxidizes active sex steroids into inactive precursors, including conversion of estradiol into estrone [18]; sex hormone-binding globulin (*SHBG*), which binds estrogen in the blood controlling its availability to bind to ER [19, 20]; and 3-beta hydroxysteroid dehydrogenase type I (*HSD3B1*), which converts pregnolone to progesterone [17, 21]. Associations were estimated for breast cancer risk overall and for breast cancer molecular subtypes. This investigation was informed by previous findings of *non-genetic* risk factor heterogeneity between luminal A and basal-like subtypes [4], therefore we focused subtype-specific *genetic* risk factor comparisons primarily on luminal A and basal-like tumors.

Materials and methods

Study population

The CBCS is a population-based case–control study of breast cancer in North Carolina and has been described previously [22, 23]. Briefly, eligible cases included women 20–74 years old who were diagnosed with primary invasive breast cancer from 1993 to 2001 and lived within the 24-county study area at diagnosis. Women diagnosed with breast carcinoma in situ (CIS) were also enrolled from 1996 to 2001. Cases were identified through the North

Carolina Central Cancer Registry using rapid case ascertainment. Randomized recruitment was used to oversample invasive cases that were African-American or <50 years old [24]. Eligible controls included women 20–74 years old who lived within the study area and had no personal history of breast cancer. Controls were identified through Department of Motor Vehicles (<65 years old) or Health Care Financing Administration (≥65 years old) records and were frequency matched to cases by race and 5-year age groups. Case and control response rates were 76 and 55 % with respect to invasive case recruitment, and 83 and 65 % for CIS recruitment.

A total of 2,311 cases and 2,022 controls provided informed consent and were interviewed about breast cancer risk factors, including reproductive and menstrual history, exogenous hormone use, and family history of cancer. A total of 2,045 (88 %) cases and 1,818 (90 %) controls also provided a blood sample. DNA was extracted from blood and stored at −80 °C. This study was approved by the University of North Carolina at Chapel Hill Institutional Review Board.

Molecular subtype

Breast tumor molecular subtype was determined by immunohistochemical analysis of ER, PR, epidermal growth factor receptor (EGFR), c-erb B2/neu (HER2), and cytokeratin 5/6 (CK 5/6), using methods described previously [4, 25–28]. Tumor tissue was available for 1,845 of 2,311 cases, and immunohistochemistry was completed for 1,424. Cases where molecular subtype was determined were more likely to be African-American and have later stage at diagnosis than cases where subtype data were incomplete, but did not differ by other characteristics [4]. Tumors were classified as luminal A (ER+ and/or PR+, HER2−); basal-like (ER−, PR−, HER2−, CK5/6+ and/or EGFR+); luminal B (ER+ and/or PR+, HER2+); HER2+/ER− (ER−, PR−, HER2+); and unclassified (ER−, PR−, HER2−, CK5/6−, EGFR−).

SNP selection

Tag SNPs for *ESR1*, *HSD3B1*, *HSD17B2*, *PGR,* and *SHBG* were selected from International HapMap Project CEU (European) and YRI (West African) population data [29] using Haploview Tagger software [30–32]; data from the HapMap African-American population were unavailable at the time of SNP selection. A minimum pairwise r^2 of 0.80 was used to define tags within the genomic regions specified in Table 1. Selection was restricted to SNPs with a minimum minor allele frequency of 0.10. SNPs selected from each population were combined into a single list. *CYP19A1* SNPs were selected from a published list of linkage disequilibrium (LD) block tag SNPs in African-American and white women [33, 34]. We also included 14 SNPs, identified through a literature search, which had been investigated previously for an association with breast cancer risk or a functional effect (Table 1). In total, 207 SNPs were selected for study.

Genotyping

SNPs were genotyped at the University of North Carolina Mammalian Genotyping Core using the Illumina GoldenGate assay (Illumina, Inc., San Diego, CA, USA) as part of a 1,536-SNP panel. Panel results and quality control measures were reported in detail previously [35]. Exact tests for deviation from Hardy–Weinberg equilibrium were conducted among controls, stratified by self-identified race [36]. Genotyping cluster images were re-reviewed for SNPs with Hardy–Weinberg test p values <0.01; all SNPs reviewed during this process had good signal intensity and none were excluded. A total of 195 of the 207 (94 %) selected SNPs had acceptable data and were included in this analysis (Table 1). Additionally, ancestry informative markers (AIMs) specifically selected to maximize the differences between African and European ancestral populations based on differences in allele frequencies and Fisher's information content were genotyped for each participant [35, 37]. A total of 144 of 158 AIMs passed quality control and were used in the analysis.

A total of 2,039 cases and 1,818 controls had DNA available for genotyping. Subjects were excluded from analysis if there were genotype calls for <95 % of SNPs ($n = 103$), genotype-determined sex was male ($n = 5$), or there was suspected contamination ($n = 1$), resulting in data for 1,972 cases and 1,776 controls (Online Resource 1). Subjects without genotype data were more likely to be cases, recruited after 1996, or African-American, but did not differ by other risk factors or molecular subtype (among cases). A total of 1,220 cases had both genotype and molecular subtype data (luminal A, $n = 679$; basal-like, $n = 200$, luminal B, $n = 116$; HER2+/ER−, $n = 94$; unclassified, $n = 131$).

Covariates

Age was defined as age in years at breast cancer diagnosis (cases) or recruitment (controls). Self-identified race was reported during the study interview. Less than 2 % of participants were Native American/Eskimo, Asian/Pacific Islander, Hispanic, or mixed race and were grouped with white women as non-African-American. Self-identified race was missing for two participants, who were excluded from analysis. Proportion of African or European ancestry was estimated from 144 AIMs using maximum likelihood

Table 1 Estrogen and progesterone pathway genes evaluated for associations with breast cancer risk in the Carolina Breast Cancer Study

Gene	Chromosomal location	Base pair range used to select tag SNPs[a]	No. of tag SNPs in analysis	Potentially functional SNPs
Estrogen receptor (*ESR1*)	6q25.1	152,168,507–152,466,237	97	rs2077647 (S10S)
				rs2234693 (+397 C/T)
				rs3798577 (3′ UTR)
				rs851982 (−104062 C/T)
				rs9341070 (3′ UTR)
Progesterone receptor (*PGR*)	11q22	100,412,565–100,506,910	26	rs10895068 (+331 G/A)
Cytochrome P450 family 19 subfamily A polypeptide 1 (*CYP19A1*)	15q21.1	N/A[b]	24	rs10046 (3′ UTR)
				rs1008805 (Intron 2 A/G)
				rs2236722 (W39R)
				rs700518 (V80V)
				rs700519 (R264C)
17-beta hydroxysteroid dehydrogenase type 2 (*HSD17B2*)	16q24.1–q24.2	80,624,461–80,689,799	40	rs4445895 (5′ UTR)
				rs8191136 (5′ UTR)
3-beta hydroxysteroid dehydrogenase type 1 (*HSD3B1*)	1p13.1	119,849,985–119,859,622	7	
Sex hormone-binding globulin (*SHBG*)	17p13-p12	7,472,680–7,477,833	1	rs1799941 (−67 G/A)

[a] Data based on NCBI B36 assembly, dbSNP b126

[b] SNPs selected from a published analysis of SNPs that tag LD blocks in African-Americans and whites (33, 34)

estimation, based on the observed allele frequencies in the study population, the contributions from the ancestral populations, and the differences in allele frequency between the ancestral populations [37–39]. Methods assumed two ancestral populations (European and African), such that the proportion of African ancestry was equal to 1 minus the proportion of European ancestry. Estimates of African ancestry were included in regression models as a proportion ranging from 0 to 0.96 (the maximum observed value).

Statistical analysis

Genotype frequencies for each SNP were calculated stratified by self-identified race and adjusted for the sampling probabilities used to select participants. LD between SNPs was estimated by calculating pairwise r^2 in Haploview according to self-identified race and case status [30]. Odds ratios (ORs) and 95 % confidence intervals (CIs) for the association between genotypes and overall breast cancer risk were estimated using unconditional binary logistic regression in SAS v9.1 (SAS, Cary, NC, USA). Subtype-specific associations were estimated using unconditional polytomous logistic regression. Genotype associations were modeled using the general statistical model (2 degrees of freedom). If the number of participants homozygous for the rare allele was <5 cases, <5 controls, or <10 cases and 10 controls, the rare homozygote and heterozygote groups were combined. Furthermore, if the results indicated that the underlying genetic model was recessive, dominant, or additive, additional analyses specific to that model were conducted. The Wald test was used to evaluate heterogeneity of basal-like and luminal A regression coefficients from the polytomous logistic regression model. The null hypothesis was that the coefficients were equal for the two subtypes. Additionally, we conducted exploratory analyses examining associations with luminal B, HER2+/ER−, and unclassified subtypes and estimating race-stratified subtype associations. In order to evaluate the effect of including CIS cases in the analysis, we re-evaluated associations excluding CIS study participants.

All statistical tests were two sided. Monte Carlo methods were used to approximate the joint distribution of test statistics and evaluate the family wise error rate to control for multiple comparisons, insuring that correlation among SNPs did not result in an overly conservative correction [40]. The nominal alpha level indicating statistical significance for genotype associations was 0.05; the Monte Carlo-adjusted level was 3.3×10^{-4}.

All models were adjusted for the frequency-matching factors (age and self-identified race) and the proportion of African ancestry in order to control for residual population stratification. Models also included an offset term in order to account for the randomized recruitment case-sampling

strategy and to produce population-based effect estimates [24, 41].

Results

Of 195 SNPs genotyped (Online Resource 2, Online Resource 3), *ESR1* rs6914211 (AA vs. AT+TT, OR 2.24, 95 % CI 1.51–3.33), *ESR1* rs985191 (CC vs. AA, OR 2.11, 95 % CI 1.43–3.13), and *PGR* rs1824128 (GT+TT vs. GG, OR 1.33, 95 % CI 1.14–1.55) were positively associated with overall breast cancer risk after accounting for multiple comparisons (Table 2). Associations between rs6914211 and rs985191 and the luminal A and basal-like subtypes were similar to associations with breast cancer overall and were not statistically different from each other (heterogeneity test: rs6914211, $p = 0.58$; rs985191, $p = 0.57$), though associations with basal-like breast cancer were not statistically significant. SNPs rs6914211 and rs985191 are approximately 15 Kb apart on chromosome 6 and were in strong LD in the CBCS (African-American: controls $r^2 = 0.70$, cases $r^2 = 0.76$; white: controls $r^2 = 0.95$, cases $r^2 = 0.96$). Rs6914211 and rs985191 were also moderately correlated with a third *ESR1* SNP, rs9397463, among whites but not African-Americans (Online Resource 4). Rs9397463 was positively associated with risk of all subtypes of breast cancer, but the association was statistically significant only for risk of luminal B tumors (Online Resource 3).

Associations between rs1824128 and the luminal A and basal-like subtypes were weaker than the association with breast cancer overall, and there was no evidence of heterogeneity between the subtypes ($p = 0.80$; Table 2). Rs1824128 was in strong LD with *PGR* rs11224575 among all subjects, and in moderate LD with *PGR* rs11224579, among white cases and controls (Online Resource 4); rs11224575 and rs11224579 were positively, but not significantly, associated with breast cancer risk (Online Resource 3). All associations remained similar after the exclusion of CIS study participants (data not shown).

To determine whether the rs6914211, rs985191, and rs1824128 associations were consistent across other tumor types, we examined their associations with HER2+/ER−, luminal B, and unclassified subtypes. The direction of association was consistent with what was seen for breast cancer overall, with the exception that rs1824128 was not associated with the luminal B subtype (Online Resource 5). Associations between rs6914211 and rs985191 and luminal B breast cancer were particularly strong. For each, two copies of the risk allele were associated with an approximately fivefold increase in luminal B risk (rs6914211 AA vs. AT+TT, OR 4.96, 95 % CI 2.37–10.38; rs985191 CC vs. AA, OR 5.12, 95 % CI 2.41–10.86). With the exception of the previously mentioned luminal B associations with rs6914211, rs985191, and rs9397463, no other SNPs were significantly associated with luminal B, HER2+/ER−, or unclassified tumors after accounting for multiple comparisons. When we explored whether subtype-specific associations for these SNPs differed by self-identified race, we found that the association between *PGR* rs1824128 and risk was stronger among African-Americans than non-African-Americans for all subtypes except the unclassified tumors. However, these associations were imprecise due to the small number of cases in some subgroups (Online resource 6).

No suspected functional SNPs met our Monte Carlo-adjusted criteria for an association with risk; however, *CYP19A1* rs700519 was associated with a reduced risk of breast cancer overall at the nominal significance level of 0.05 (TT vs. CC, OR 0.44, 95 % CI 0.21–0.89; Table 3). An inverse association was also noted for luminal A and basal-like subtypes, although the luminal A association was slightly weaker than the overall association.

Discussion

We investigated the association between SNPs in *CYP19A1, ESR1, HSD3B1, HSD17B2, PGR,* and *SHBG* and breast cancer risk and identified two SNPs in *ESR1* and one in *PGR* that were associated with breast cancer risk. There was no association between *HSD3B1*, *HSD17B2,* or *SHBG* variants and risk. Statistical significance criteria were adjusted for the effective number of comparisons made, but it is still possible that some of the significant associations we observed were due to chance.

An increasing number of studies have evaluated heterogeneity of SNP associations by tumor characteristics, including ER, PR, and HER2 and other clinical characteristics [42–49], but few have considered intrinsic molecular subtype, which defines basal-like breast cancer as a specific type of ER-negative breast cancer. Studies that have evaluated molecular subtypes have shown that genetic associations [9, 10] and gene expression levels [50] differ by molecular subtype for some risk alleles. Although variants in estrogen and progesterone signaling pathway-related genes have been investigated previously for associations with overall breast cancer risk, we examined associations between these variants and luminal A and basal-like breast cancer subtypes because subtype differences in hormonally related non-genetic risk factors and differences in ER and PR expression suggested that genetic risks related to hormonal pathways may differ between these subtypes as well. However, in this study, associations with genetic variants in estrogen and progesterone pathway genes were largely similar for the luminal A and basal-like

Table 2 Estrogen and progesterone pathway SNPs associated with breast cancer risk in the Carolina Breast Cancer Study

	Controls	All cases			Breast cancer subtype						
					Luminal A			Basal-like			
	n	*n*	OR (95 % CI)[a]	*p* value	*n*	OR (95 % CI)[a]	*p* value	*n*	OR (95 % CI)[a]	*p* value	$p^{b}_{\text{heterogeneity}}$
ESR1											
rs6914211											
AT+TT	1,736	1,887	Referent		646	Referent		190	Referent		
AA	40	85	2.24 (1.51, 3.33)	1.0×10^{-4}	33	2.70 (1.66, 4.39)	1.0×10^{-4}	10	2.19 (1.06, 4.53)	0.03	0.58
rs985191											
AA	1,202	1,312	Referent		444	Referent		117	Referent		
AC	532	574	0.99 (0.86, 1.16)	0.94	203	1.08 (0.88, 1.33)	0.45	74	1.34 (0.97, 1.84)	0.07	0.23
CC	42	86	2.11 (1.43, 3.13)	2.0×10^{-4}	32	2.47 (1.51, 4.04)	3.0×10^{-4}	9	1.98 (0.92, 4.24)	0.08	0.57
PGR											
rs1824128											
GG	1,348	1,394	Referent		489	Referent		148	Referent		
GT+TT	428	577	1.33 (1.14, 1.55)	3.0×10^{-4}	190	1.22 (0.99, 1.50)	0.06	52	1.16 (0.83, 1.64)	0.39	0.80

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

[b] *p* value for heterogeneity of luminal A and basal-like subtype associations

Table 3 Association between estrogen and progesterone pathway functional SNPs[a] and breast cancer risk in the Carolina Breast Cancer Study

SNP	Genotype	Controls	All cases		Breast cancer subtype			
					Luminal A		Basal-like	
		n	*n*	OR (95 % CI)[b]	*n*	OR (95 % CI)[b]	*n*	OR (95 % CI)[b]
CYP19A1								
rs700518	AA	690	774	Referent	252	Referent	94	Referent
V80V	AG	768	884	1.07 (0.92, 1.25)	306	1.08 (0.87, 1.34)	75	0.93 (0.66, 1.32)
	GG	315	312	0.87 (0.70, 1.08)	120	0.93 (0.70, 1.25)	31	1.05 (0.64, 1.70)
rs700519	CC	1,514	1,694	Referent	578	Referent	162	Referent
R264C	CT	239	264	0.99 (0.81, 1.21)	95	1.17 (0.89, 1.56)	36	1.06 (0.70, 1.60)
	TT	23	14	0.44 (0.21, 0.89)	6	0.67 (0.26, 1.71)	<5	0.48 (0.11, 2.13)
rs1008805	CC	236	256	Referent	88	Referent	20	Referent
Intron 2 A/G	CT	722	836	1.10 (0.95, 1.28)	289	1.07 (0.87, 1.31)	77	1.02 (0.73, 1.41)
	TT	816	880	0.98 (0.78, 1.22)	302	0.91 (0.68, 1.23)	103	0.86 (0.50, 1.46)
rs10046	CC	628	695	Referent	225	Referent	84	Referent
3′ UTR	CT	791	918	1.11 (0.95, 1.30)	324	1.16 (0.94, 1.44)	82	1.00 (0.71, 1.41)
	TT	357	359	0.90 (0.73, 1.10)	130	0.90 (0.68, 1.20)	34	0.98 (0.61, 1.56)
rs2234693	CC	398	470	Referent	158	Referent	45	Referent
+397 C/T	CT	908	984	0.92 (0.77, 1.09)	358	0.98 (0.78, 1.23)	108	1.09 (0.75, 1.59)
	TT	469	518	0.90 (0.74, 1.09)	163	0.79 (0.61, 1.04)	47	0.93 (0.60, 1.45)
rs2077647	TT	461	534	Referent	171	Referent	50	Referent
+29 T/C	CT	913	986	0.96 (0.82, 1.13)	354	1.10 (0.88, 1.38)	103	1.08 (0.75, 1.55)
S10S	CC	399	452	0.99 (0.81, 1.20)	154	1.12 (0.85, 1.46)	47	1.10 (0.71, 1.69)
rs3798577	TT	527	577	Referent	156	Referent	38	Referent
3′ UTR	CT	851	977	1.08 (0.92, 1.26)	328	1.08 (0.87, 1.35)	105	1.27 (0.90, 1.80)
	CC	398	417	0.94 (0.78, 1.14)	195	1.06 (0.82, 1.37)	57	0.97 (0.62, 1.50)
rs851982	CC	229	265	Referent	84	Referent	24	Referent
−104062 C/T	CT	730	846	1.03 (0.83, 1.28)	294	1.14 (0.85, 1.54)	89	1.08 (0.66, 1.77)
	TT	816	859	0.93 (0.75, 1.15)	301	1.08 (0.80, 1.46)	87	0.80 (0.48, 1.32)
rs4445895	CC	733	789	Referent	256	Referent	83	Referent
5′ UTR	CT	807	947	1.09 (0.94, 1.26)	344	1.21 (1.00, 1.48)	88	0.98 (0.71, 1.35)
	TT	236	236	0.92 (0.74, 1.14)	79	0.94 (0.69, 1.27)	29	1.14 (0.72, 1.80)
rs10895068	GG	1,637	1,803	Referent	625	Referent	187	Referent
+331G/A	AG	135	164	1.10 (0.85, 1.41)	51	0.90 (0.63, 1.29)	13	1.00 (0.54, 1.84)
	AA	<5	5	–[c]	<5	–[c]	0	
rs1799941	GG	1,188	1,356	Referent	461	Referent	141	Referent
−67 G/A	AG	516	527	0.93 (0.80, 1.09)	182	0.92 (0.74, 1.14)	50	1.07 (0.74, 1.54)
	AA	71	89	1.31 (0.93, 1.85)	36	1.52 (0.98, 2.37)	9	1.92 (0.90, 4.07)

[a] *CYP19A1* rs2236722, *ESR1* rs9341070, and *HSD17B2* rs8191136 not included in table due to monomorphism or too few subjects with variant alleles

[b] Odds ratio and 95 % confidence interval, adjusted for age, self-identified race, African ancestry, and offset term

[c] Not estimated due to small sample size

subtypes. Although case numbers were small, positive associations for rs6914211, rs985191, and rs1824128 were generally consistent for luminal B, HER2+/ER−, and unclassified tumors as well.

ESR1 rs6914211 and rs985191 were associated with breast cancer risk in this study. Previously, in a large pooled analysis, rs985191 was found to be positively, but not significantly associated with breast cancer risk [51]; to our knowledge, no previous studies have identified rs6914211 as having an association with risk. Rs6914211 and rs985191 are located close to other risk-associated regions on chromosome 6. *ESR1* rs3020314 and *ESR1* rs3020401 were identified as potentially associated with risk in a two-stage candidate SNP analysis [52], and

rs3020314 was associated with ER-positive breast cancer in a three-stage study of *ESR1* tag SNPs [51]. Rs3020314 and rs3020401 were in low to moderate LD with rs6914211 and rs985191 among white CBCS participants (r^2 ranging from 0.28 to 0.32). In this study, rs3020401 was positively associated with risk of all subtypes except for HER2+/ER−, but at levels below our criteria accounting for multiple comparisons, whereas rs3020314 was not associated with overall or subtype-specific risk. It is possible that rs3020314/rs3020401 and rs6914211/rs985191 are reflecting associations related to the same causal locus and inconsistencies between studies may be a result of regional variation in LD between study populations. Fine-mapping of this region may help illuminate the nature of the causal locus (or loci) that may be near these SNPs.

In addition to these tag SNP associations, genome wide association studies (GWAS) have also identified breast cancer risk loci on 6q25.1: one within *ESR1* [46] and two >60 Kb upstream of *ESR1* [53, 54]. The upstream SNPs are in close proximity to *C6ORF97* and *C6ORF211*, genes of unknown function whose expression is highly correlated with *ESR1* expression [55]. These GWAS risk variants were not genotyped in this analysis; however, future investigations of 6q25.1 should consider joint analyses of regions within and upstream of *ESR1*, as the co-expression profiles suggest a biological relationship.

PGR intronic SNP rs1824128 was also associated with breast cancer risk. To our knowledge, previous studies have not identified rs1824128 as associated with risk nor has 11q22.1 been identified as a harboring risk loci in breast cancer GWAS. Much of the investigation into *PGR* variants and breast cancer risk has focused on the functional polymorphism +331 G/A (rs10895068), which was not associated with risk in this study or a recent meta-analysis [56]. Thus, replication of the rs1824128 association is necessary before the meaning of this association can be assessed.

Our analysis of potentially functional SNPs suggested an inverse association between the *CYP19A1* rs700519 (R264C) TT genotype and risk of breast cancer overall. Although this association did not meet the multiple comparison-adjusted criteria for statistical significance, there is prior evidence that the R264C change may be associated with risk. Some have reported that 264C results in reduced hydrophobicity, reduced *CYP19A1* expression, and lower aromatase activity [57, 58]; lower aromatase activity would be consistent with the association between the TT genotype and lower breast cancer risk in the CBCS. However, another study reported no difference in enzymatic activity between 264R and 264C [59], and rs700519 was not associated with risk in other epidemiologic studies [33, 60, 61]. The rs700519 TT genotype is uncommon and was rare in CBCS non-African-American participants. A lack of association with breast cancer risk in previous studies may be explained by the low prevalence of the at-risk genotype in populations of non-African descent.

Molecular subtype data were not available for all enrolled cases; however, genotype distributions only differed between cases with and without molecular subtype data for four SNPs (*HSD17B2* rs3111351 and rs8191136 in African-Americans, and *ESR1* rs6557177 and rs985695 in non-African-Americans). This is within the number expected to differ due to chance alone, suggesting that results were likely not biased due to the inability to include all cases in subtype-specific analyses. Other limitations of this analysis included low SNP density in some genes and limited statistical power for identifying subtype heterogeneity. The SNP selection method we employed selected SNPs based on LD with other SNPs (i.e., tag SNPs), such that a single SNP may be a marker for other correlated SNPs. This tagging method has been shown to provide efficient coverage of gene regions [32]; however, failure of tag SNPs during genotyping may have led to low coverage in some areas. The number of basal-like cases was relatively low in comparison with the number of luminal A cases, and heterogeneity tests may have lacked power. However, visual comparison of ORs for the three SNPs most strongly associated with overall risk shows similar associations luminal A and basal-like subtypes with the same direction of effect; thus it is unlikely that any meaningful heterogeneity was overlooked for these particular SNPs.

The major strength of this analysis was the ability to examine molecular subtypes in a large, population-based study. These IHC-determined subtypes have unique risk factor profiles and are related to breast cancer prognosis [4, 5, 25], but the contribution of genetic susceptibility to the risk of each subtype has not yet been determined. Our ability to detect statistical differences between luminal A and basal-like associations was limited, but our results show clear positive associations between the three risk SNPs and basal-like and luminal A breast cancers, as well as luminal B, HER2+/ER−, and unclassified subtypes. These data suggest that genetic variants in pathways primarily related to hormonal mechanisms may be important factors for ER-negative as well as ER-positive subtypes. We can only speculate as to why differences in association by molecular subtype were not observed. One possible explanation is that heterogeneity may be related to polymorphisms in hormonally related genes not evaluated in this study. We evaluated six genes, but there are several other hormonally related candidate genes that may be differentially related to risk of breast cancer subtypes. Another possibility is that the non-genetic risk factor heterogeneity observed may be related to aspects of parity and lactation that are not directly related to sex hormones.

Further investigation of the relationship between hormonal and non-hormonal biomarkers and risk of breast cancer subtypes is needed to understand the relative contributions of each biological pathway.

Another strength was that African-American participants were oversampled, providing an opportunity for exploratory analyses of molecular subtype risk stratified by self-identified race. Despite recent efforts to conduct large-scale studies of genetic variation and breast cancer risk in African-Americans, most early genetic susceptibility studies have consisted primarily of women of European descent. We explored associations stratified by self-identified race with the hypothesis that there may be risk-associated variants that are more prevalent in African or African-American populations. This may include variants associated with risk of basal-like breast cancer, a subtype that is more prevalent among African-Americans [4]. Furthermore, because LD range can differ between populations of African as compared with European descent [62, 63], it is possible that a tag SNP may have different levels of correlation with causal loci in different populations. Thus, a SNP may be associated with risk among subjects with one genetic background and weakly or not associated among subjects with another. In this population, *CYP19A1* R264C was more prevalent among African-Americans and associations with *PGR* rs1824128 were stronger among African-Americans, an observation that was largely consistent across subtypes but was based on small numbers and must be interpreted with caution.

Genotype frequencies for rs1824128 were similar comparing African-American and non-African-American participants and so differing prevalence of the risk allele is not a fitting explanation for why the rs1824128 association might be limited to African-Americans. As mentioned previously, there is no data pointing to any functional effect associated with rs1824128 and the association requires replication in order to be confirmed. It is possible that the observed differences by race were due to chance. In an analysis of GWAS-identified and candidate gene SNPs, O'Brien et al. [10] found that molecular subtype-specific associations differed between African-Americans and non-African-Americans for a minority of SNPs they examined; however, that study was also carried out within the CBCS and was subject to the same power limitations that we experienced in this study. Future analyses intended to focus primarily on racial differences in risk of breast cancer subtypes will require even larger case numbers and would be most efficiently carried out through pooled analyses. Finally, all statistical models were adjusted for the proportion of African ancestry, controlling for potential confounding due to population stratification.

In conclusion, this analysis identified genetic variants in *ESR1*, *PGR,* and *CYP19A1* that were associated with breast cancer risk, and there was little evidence of heterogeneity between luminal A and basal-like subtypes. Associations in *ESR1* and *CYP19A1* were consistent with data from other studies. Further characterization of the effects of these SNPs or related loci on gene expression could increase our understanding of the biological pathways active in breast carcinogenesis.

Acknowledgments This work was supported by funding from the National Institutes of Health [P50-CA58223, P30-CA16086, R25-CA57726 to S.J.N., P30ES10126 to M. D. G. and R. C. M.]. The authors wish to thank Patricia Basta of the University of North Carolina at Chapel Hill Epidemiology Department, and Michael Andre and Amanda Beaty of the University of North Carolina Lineberger Comprehensive Cancer Center for their work preparing the DNA samples and conducting the genotyping. The authors also thank Jessica Tse of the University of North Carolina at Chapel Hill Epidemiology Department for advice and statistical support.

Conflict of interest Sarah J. Nyante is currently a postdoctoral fellow at the National Institutes of Health, which funded the research undertaken at the University of North Carolina at Chapel Hill. There are no other conflicts of interest to declare.

References

1. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D (2000) Molecular portraits of human breast tumours. Nature 406:747–752
2. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P, Borresen-Dale AL (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 98:10869–10874
3. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 100:8418–8423
4. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, Smith LV, Labbok MH, Geradts J, Bensen JT, Jackson S, Nyante S, Livasy C, Carey L, Earp HS, Perou CM (2008) Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 109:123–139
5. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, Hewitt SM, Anderson WF, Szeszenia-Dabrowska N, Bardin-Mikolajczak A, Zatonski W, Cartun R, Mandich D, Rymkiewicz G, Ligaj M, Lukaszek S, Kordek R, 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
6. O'Brien KM, Cole SR, Tse CK, Perou CM, Carey LA, Foulkes WD, Dressler LG, Geradts J, Millikan RC (2010) Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. Clin Cancer Res 16:6100–6110. doi:10.1158/1078-0432.CCR-10-1533
7. Easton D (2012) Germline polymorphisms and susceptibility to breast cancer. San Antonio Breast Cancer Symposium, San Antonio, TX

8. Peng S, Lu B, Ruan W, Zhu Y, Sheng H, Lai M (2011) Genetic polymorphisms and breast cancer risk: evidence from meta-analyses, pooled analyses, and genome-wide association studies. Breast Cancer Res Treat 127:309–324. doi:10.1007/s10549-011-1459-5
9. Broeks A, Schmidt MK, Sherman ME et al (2011) Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. Hum Mol Genet 20:3289–3303. doi:10.1093/hmg/ddr228
10. O'Brien KM, Cole SR, Engel LS, Bensen JT, Poole CL, Herring AH, Millikan RC (2013) Breast cancer subtypes and previously established genetic risk factors: A Bayesian approach. Cancer Epidemiol Biomarkers Prev. doi:10.1158/1055-9965.epi-13-0463
11. Barbieri RL (2004) The Breast. In: Strauss JF, Barbieri RL (eds) Yen and Jaffe's reproductive endocrinology, 5th edn. Elsevier, Philadelphia, pp 307–326
12. Conneely OM, Mulac-Jericevic B, Lydon JP (2003) Progesterone-dependent regulation of female reproductive activity by two distinct progesterone receptor isoforms. Steroids 68:771–778
13. Gruber CJ, Tschugguel W, Schneeberger C, Huber JC (2002) Production and actions of estrogens. N Engl J Med 346:340–352
14. Frasor J, Danes JM, Komm B, Chang KC, Lyttle CR, Katzenellenbogen BS (2003) Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. Endocrinology 144:4562–4574
15. Ballare C, Uhrig M, Bechtold T, Sancho E, Di Domenico M, Migliaccio A, Auricchio F, Beato M (2003) Two domains of the progesterone receptor interact with the estrogen receptor and are required for progesterone activation of the c-Src/Erk pathway in mammalian cells. Mol Cell Biol 23:1994–2008
16. Ballare C, Vallejo G, Vicent GP, Saragueta P, Beato M (2006) Progesterone signaling in breast and endometrium. J Steroid Biochem Mol Biol 102:2–10
17. Payne AH, Hales DB (2004) Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. Endocr Rev 25:947–970
18. Luu-The V (2001) Analysis and characteristics of multiple types of human 17beta-hydroxysteroid dehydrogenase. J Steroid Biochem Mol Biol 76:143–151
19. Catalano MG, Frairia R, Boccuzzi G, Fortunati N (2005) Sex hormone-binding globulin antagonizes the anti-apoptotic effect of estradiol in breast cancer cells. Mol Cell Endocrinol 230:31–37
20. Fortunati N, Fissore F, Fazzari A, Becchis M, Comba A, Catalano MG, Berta L, Frairia R (1996) Sex steroid binding protein exerts a negative control on estradiol action in MCF-7 cells (human breast cancer) through cyclic adenosine 3′,5′-monophosphate and protein kinase A. Endocrinology 137:686–692
21. Duax WL, Thomas J, Pletnev V, Addlagatta A, Huether R, Habegger L, Weeks CM (2005) Determining structure and function of steroid dehydrogenase enzymes by sequence analysis, homology modeling, and rational mutational analysis. Ann N Y Acad Sci 1061:135–148
22. Millikan R, Eaton A, Worley K, Biscocho L, Hodgson E, Huang WY, Geradts J, Iacocca M, Cowan D, Conway K, Dressler L (2003) HER2 codon 655 polymorphism and risk of breast cancer in African Americans and whites. Breast Cancer Res Treat 79:355–364
23. Newman B, Moorman PG, Millikan R, Qaqish BF, Geradts J, Aldrich TE, Liu ET (1995) The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. Breast Cancer Res Treat 35:51–60
24. Weinberg CR, Sandler DP (1991) Randomized recruitment in case–control studies. Am J Epidemiol 134:421–432
25. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Millikan RC (2006) Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 295:2492–2502
26. Huang WY, Newman B, Millikan RC, Schell MJ, Hulka BS, Moorman PG (2000) Hormone-related factors and risk of breast cancer in relation to estrogen receptor and progesterone receptor status. Am J Epidemiol 151:703–714
27. Livasy CA, Perou CM, Karaca G, Cowan DW, Maia D, Jackson S, Tse CK, Nyante S, Millikan RC (2007) Identification of a basal-like subtype of breast ductal carcinoma in situ. Hum Pathol 38:197–204
28. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L, Akslen LA, Ragaz J, Gown AM, Gilks CB, van de Rijn M, Perou CM (2004) Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 10:5367–5374
29. Frazer KA, Ballinger DG, Cox DR et al (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861
30. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265
31. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 74:106–120
32. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D (2005) Efficiency and power in genetic association studies. Nat Genet 37:1217–1223
33. Haiman CA, Dossus L, Setiawan VW et al (2007) Genetic variation at the CYP19A1 locus predicts circulating estrogen levels but not breast cancer risk in postmenopausal women. Cancer Res 67:1893–1897
34. Haiman CA, Stram DO, Pike MC, Kolonel LN, Burtt NP, Altshuler D, Hirschhorn J, Henderson BE (2003) A comprehensive haplotype analysis of CYP19 and breast cancer risk: the multiethnic cohort. Hum Mol Genet 12:2679–2692
35. Nyante SJ, Gammon MD, Kaufman JS, Bensen JT, Lin DY, Barnholtz-Sloan JS, Hu Y, He Q, Luo J, Millikan RC (2011) Common genetic variation in adiponectin, leptin, and leptin receptor and association with breast cancer subtypes. Breast Cancer Res Treat 129:593–606. doi:10.1007/s10549-011-1517-z
36. Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy–Weinberg equilibrium. Am J Hum Genet 76:887–893
37. Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF (2006) A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. Am J Hum Genet 79:640–649
38. Barnholtz-Sloan JS, Chakraborty R, Sellers TA, Schwartz AG (2005) Examining population stratification via individual ancestry estimates versus self-reported race. Cancer Epidemiol Biomark Prev 14:1545–1551
39. Barnholtz-Sloan JS, McEvoy B, Shriver MD, Rebbeck TR (2008) Ancestry estimation and correction for population stratification in molecular epidemiologic association studies. Cancer Epidemiol Biomark Prev 17:471–477
40. Lin DY (2005) An efficient Monte Carlo approach to assessing statistical significance in genomic studies. Bioinformatics 21:781–787. doi:10.1093/bioinformatics/bti053
41. Weinberg CR, Wacholder S (1990) The design and analysis of case–control studies with biased sampling. Biometrics 46:963–975

42. Stacey SN, Manolescu A, Sulem P et al (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 39:865–869
43. Garcia-Closas M, Hall P, Nevanlinna H et al (2008) Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet 4:e1000054
44. Palmer JR, Ruiz-Narvaez EA, Rotimi CN, Cupples LA, Cozier YC, Adams-Campbell LL, Rosenberg L (2013) Genetic susceptibility loci for subtypes of breast cancer in an African American population. Cancer Epidemiol Biomarkers Prev 22:127–134. doi:10.1158/1055-9965.epi-12-0769
45. Stacey SN, Manolescu A, Sulem P et al (2008) Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 40:703–706
46. Siddiq A, Couch FJ, Chen GK et al (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:10.1093/hmg/dds381
47. Figueroa JD, Garcia-Closas M, Humphreys M et al (2011) Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. Hum Mol Genet 20:4693–4706. doi:10.1093/hmg/ddr368
48. Stevens KN, Fredericksen Z, Vachon CM et al (2012) 19p13.1 is a triple-negative-specific breast cancer susceptibility locus. Cancer Res 72:1795–1803. doi:10.1158/0008-5472.can-11-3364
49. Stevens KN, Vachon CM, Lee AM et al (2011) Common breast cancer susceptibility loci are associated with triple-negative breast cancer. Cancer Res 71:6240–6249. doi:10.1158/0008-5472.can-11-1266
50. Nordgard SH, Johansen FE, Alnaes GI, Naume B, Borresen-Dale AL, Kristensen VN (2007) Genes harbouring susceptibility SNPs are differentially expressed in the breast cancer subtypes. Breast Cancer Res 9:113
51. Dunning AM, Healey CS, Baynes C et al (2009) Association of ESR1 gene tagging SNPs with breast cancer risk. Hum Mol Genet 18:1131–1139
52. Mavaddat N, Dunning AM, Ponder BA, Easton DF, Pharoah PD (2009) Common genetic variation in candidate genes and susceptibility to subtypes of breast cancer. Cancer Epidemiol Biomark Prev 18:255–259
53. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, Wen W, Levy S, Deming SL, Haines JL, Gu K, Fair AM, Cai Q, Lu W, Shu XO (2009) Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet 41:324–328
54. Turnbull C, Ahmed S, Morrison J et al (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet 42:504–507. doi:10.1038/ng.586
55. Dunbier AK, Anderson H, Ghazoui Z, Lopez-Knowles E, Pancholi S, Ribas R, Drury S, Sidhu K, Leary A, Martin LA, Dowsett M (2011) ESR1 is co-expressed with closely adjacent uncharacterised genes spanning a breast cancer susceptibility locus at 6q25.1. PLoS Genet 7:e1001382. doi:10.1371/journal.pgen.1001382
56. Yu KD, Chen AX, Shao ZM (2010) No association between a progesterone receptor gene promoter polymorphism (+331G>A) and breast cancer risk in Caucasian women: evidence from a literature-based meta-analysis. Breast Cancer Res Treat 122:853–858. doi:10.1007/s10549-010-0738-x
57. Kagawa N, Hori H, Waterman MR, Yoshioka S (2004) Characterization of stable human aromatase expressed in *E. coli*. Steroids 69:235–243
58. Ma CX, Adjei AA, Salavaggione OE, Coronel J, Pelleymounter L, Wang L, Eckloff BW, Schaid D, Wieben ED, Adjei AA, Weinshilboum RM (2005) Human aromatase: gene resequencing and functional genomics. Cancer Res 65:11071–11082
59. Watanabe J, Harada N, Suemasu K, Higashi Y, Gotoh O, Kawajiri K (1997) Arginine–cysteine polymorphism at codon 264 of the human CYP19 gene does not affect aromatase activity. Pharmacogenetics 7:419–424
60. Lee KM, Abel J, Ko Y, Harth V, Park WY, Seo JS, Yoo KY, Choi JY, Shin A, Ahn SH, Noh DY, Hirvonen A, Kang D (2003) Genetic polymorphisms of cytochrome P450 19 and 1B1, alcohol use, and breast cancer risk in Korean women. Br J Cancer 88:675–678
61. Miyoshi Y, Iwao K, Ikeda N, Egawa C, Noguchi S (2000) Breast cancer risk associated with polymorphism in CYP19 in Japanese women. Int J Cancer 89:325–328
62. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229
63. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. Nature 411:199–204