Estrogen Receptor Genotypes and Haplotypes Associated with Breast Cancer Risk

Bert Gold,1 Francis Kalush,2 Julie Bergeron,3 Kevin Scott,3 Nandita Mitra,4 Kelly Wilson,2 Nathan Ellis,4 Helen Huang,4 Michael Chen,3 Ross Lippert,5,6 Bjarni V. Halldorsson,5 Beth Woodworth,1 Thomas White,2 Andrew G. Clark,2 Fritz F. Parl,7 Samuel Broder,2 Michael Dean,1 and Kenneth Offit4

1Human Genetics Section, Laboratory of Genomic Diversity, National Cancer Institute at Frederick, Frederick, Maryland; 2Celera Diagnostics or Celera, Inc, Rockville, Maryland; 3Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts; and 4Department of Pathology, Vanderbilt University School of Medicine, Nashville, Tennessee

ABSTRACT

Nearly one in eight US women will develop breast cancer in their lifetime. Most breast cancer is not associated with a hereditary syndrome, occurs in postmenopausal women, and is estrogen and progesterone receptor-positive. Estrogen exposure is an epidemiologic risk factor for breast cancer and estrogen is a potent mammary mitogen. We studied single nucleotide polymorphisms (SNPs) in estrogen receptors in 615 healthy subjects and 1011 individuals with histologically confirmed breast cancer, all from New York City. We analyzed 13 SNPs in the progesterone receptor gene (PGR), 17 SNPs in estrogen receptor 1 gene (ESR1), and 8 SNPs in the estrogen receptor 2 gene (ESR2). We observed three common haplotypes in ESR1 that were associated with a decreased risk for breast cancer [odds ratio (OR), ~ 0.4; 95% confidence interval (CI), 0.2–0.8; P < 0.01]. Another haplotype was associated with an increased risk of breast cancer (OR, 2.1; 95% CI, 1.2–3.8; P < 0.05). A unique risk haplotype was present in ~7% of older Ashkenazi Jewish study subjects (OR, 1.7; 95% CI, 1.2–2.4; P < 0.003). We narrowed the ESR1 risk haplotypes to the promoter region and first exon. We define several other haplotypes in Ashkenazi Jews in both ESR1 and ESR2 that may elevate susceptibility to breast cancer. In contrast, we found no association between any PGR variant or haplotype and breast cancer. Genetic epidemiology study replication and functional assays of the haplotypes should permit a better understanding of the role of steroid receptor genetic variants and breast cancer risk.

INTRODUCTION

Only a small fraction (~<5%) of women diagnosed with breast cancer have a clear hereditary predisposition (1–3), and of these, about one half have predisposing mutations in BRCA1, BRCA2, PTEN, TP53, or other known cancer predisposing genes. However, twin studies indicate that the heritability of breast cancer is ~30% (4), suggesting that genes other than the well-mapped regions act as modifiers of breast cancer risk. Although it is likely low penetrance as well as high penetrance genes may be involved in the etiology, it remains unclear which genomic regions and which biochemical functions or signal transduction pathways account for additional, heritable breast cancer incidence or progression.

Abundant epidemiologic evidence suggests that estrogen plays a crucial role in most breast cancers. Nulliparous women are at significantly elevated risk, as are women who have had children late in their lives, women who have early menarche or women who have late menopause. Obesity is also associated with breast cancer risk; estrogen synthesis in adipose tissue is proposed to account for this increase in risk. Whereas estrogen receptor (ER)-positive and progesterone receptor (PgR)-positive breast cancers have better short-term prognosis than those that have become hormone independent (5), receptor status varies as a function of age and menopausal status. Younger patients are more likely receptor negative and hormonally unresponsive; older patients are more often receptor positive and hormonally responsive. Recent and complete reviews of the genetics of breast cancer and its relation to the estrogen and progesterone receptor are available (6–12).

In this report, we seek to identify candidate steroid hormone receptor gene variants in ESR1, ESR2, and PGR that might be associated with risk of breast cancer, perhaps leading to accelerated or slower rates of neoplastic transformation.8

MATERIALS AND METHODS

5′-Nucleotidase Assay Designs. Single nucleotide polymorphisms (SNPs) discovered through data mining of the Celera Discovery System, a Celera Proprietary Database, or deposits into dbSNP were chosen for assay design. Limited resequencing permitted discovery of the rare variant in ESR2 exon 4 reported here. Sequences chosen for scoring on the entire cohort were either purchased from Applied Biosystems as Assays-on-Demand (AOD) or submitted to an Applied Biosystems Assays-by-Design pipeline. One difficult assay (G393G in PGR) was designed by Raymond Stephens of Celadon Labs., College Park, MD. Propynyl T oligonucleotide probes for G393G were manufactured under special license agreement by Trilink Biotech, San Diego, CA.

5′-Nucleotidase Assay Method. Five nanograms of genomic patient and control DNAs were aliquoted with a Hydra liquid handler (HYDRA Robbins Molecular BioProducts, San Diego, CA) into 384-well bar-coded optical thermostable plates compatible with the ABI PRISM 7900HT sequence detection system (ABI Prism 7900 HT, Applied Biosystems, Inc., Foster City, CA). Before assay, these were rehydrated with 2.4 µl of deionized water with a QiFill2 automated pipetter (QFILL2 Genetix Ltd., Queensway, New Milton Hampshire, United Kingdom). For Assays-on-Demand products (denoted by a catalog number and paucity of sequence information in Table 1), 2.5 µL of TaqMan Universal PCR master mix was added with 125 nL of Assay-on-Demand mix per well. For Assays-by-Design products or assays devised in house, 2.5 µL of PCR master mix was combined with 100 µmol of each primer (45 nL each) and 100 µmol of each probe (10 nL each). Plates were sealed and cycled at 95°C for 10 minutes, followed by 50 cycles of 95°C for 15 seconds, 58°C for 1 minute in an ABI GeneAmp PCR System 9700 thermocycler set for 9600 emulation. At the end of cycling, plates were held at 25°C until reading in a 7900 HT sequence detection system. Each plate contained controls of each genotype and no template controls. Data from plates failing any control were discarded. Manual genotype calls were done conservatively, consistent with the standards discussed in Clark et al. (13) and missing values were excluded from the analysis as detailed in Results. Aggregate indeterminate genotypes

Received 4/9/04; revised 10/1/04; accepted 10/6/04.

Grant support: Funded in whole or in part with federal funds from the Center for Cancer Research of the National Cancer Institute and the NIH under contract NO1-CO-12400, the Barbara Goldsmith Foundation, The Lymphoma Foundation, The Frankel Foundation, and the Academic Medicine Development Company. The New York Cancer Project is administered and funded by AMDeC Foundation, Inc.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Bert Gold, Human Genetics Section, Laboratory of Genomic Diversity, Center for Cancer Research, National Cancer Institute at Frederick, Building 560, Room 21-21, Frederick, MD 21702. Phone: (301) 846-5098; Fax: (301) 846-1909; E-mail: goldb@ncifcrf.gov.

©2004 American Association for Cancer Research.

8 Genetic loci are identified in this publication as ESR1, ESR2 and PGR consistent with Human Genome Organization (HUGO) guidelines. ERα, ERβ and PgR refer to the respective peptide products.
we used SNPhap,11 PHASE12 (14), SNPEM13 (15), MLOCUS14 (16), and SPSS (SPSS Inc., Chicago, IL). For haplotype estimation was carried out with SAS (SAS Institute, Inc., Cary, NC), SAS/Genetics (SAS extension of the Clark (17) algorithm with expectation maximization (EM; ref. 18); in the rs910416 rs1884051, rs6905370, rs926778, rs3020366, rs75066, rs2228480, rs7398677, rs3020366, rs910416

* AOD, Assays-on-Demand (number beginning with “C.” is a catalog number).

averaged 1% of total ESR1 SNPs sampled (range, 0.4–2%). Genotype tabulations and missing value details are provided as supplementary data.8

The following SNPs were assayed in the ESR1 gene: rs851984, rs2881766, ESR1002, rs2077647, rs82471, rs932231, rs172221, bCV1576295, rs1801132, rs1884051, rs6905370, rs926778, rs6905370, rs851984, rs3020366, rs75066, rs2228480, rs7398677, rs3020366, rs910416

Institute, Inc.), and SPSS (SPSS Inc., Chicago, IL). For haplotype estimation was carried out with SAS (SAS Institute, Inc., Cary, NC), SAS/Genetics (SAS extension of the Clark (17) algorithm with expectation maximization (EM; ref. 18), and haplo.score19 (19). Self-assigned demography was checked with STRUCTURE.16 Each analysis was conducted for the case–control population as a whole; stratifying by one of six ethnicities (Asian American, African American, Hispanic, Ashkenazi Jewish, Unknown, or European American) and by age (age ≤50 and age >50). To address the effect of age of menopause, a surrogate marker (age < or ≥50) was used. Five males with breast cancer and two male controls were included in the study, but excluded for the statistical analysis presented here. No result presented here was significantly impacted when we included or excluded these males. For the purpose of haplotype estimation, the most common alleles in the study as a whole were represented by “A” for ESR1 loci; and “A” for ESR2 loci; in addition, efforts are made to describe each haplotype by the base letter of the variants that compose it. Map coordinates provided are those from NCBI Build 33 (April 2003).17

SNP name designations are consistent with those defined in dbSNP when possible.

A map with SNP location details is provided as Fig. 1.

### Statistical Methods.

SNP typing in control samples was checked for compliance with Hardy–Weinberg equilibrium with Tools For Population Genetic Analyses (TFPGA).10 Contingency table analysis for individual SNPs was carried out with SAS (SAS Institute, Inc., Cary, NC), SAS/Genetics (SAS Institute, Inc.), and SPSS (SPSS Inc., Chicago, IL). For haplotype estimation we used SNPship.11 PHASE112 (14), SNPEM13 (15), MLOCUS14 (16), an extension of the Clark (17) algorithm with expectation maximization (EM; ref. 18), and haplo.score119 (19). Self-assigned demography was checked with STRUCTURE.16 Each analysis was conducted for the case–control population as a whole; stratifying by one of six ethnicities (Asian American, African American, Hispanic, Ashkenazi Jewish, Unknown, or European American) and by age (age ≤50 and age >50). To address the effect of age of menopause, a surrogate marker (age < or ≥50) was used. Five males with breast cancer and two male controls were included in the study, but excluded for the statistical analysis presented here. No result presented here was significantly impacted when we included or excluded these males. For the purpose of haplotype estimation, the most common alleles in the study as a whole were represented by “A” for ESR1 loci; and “A” for ESR2 loci; in addition, efforts are made to describe each haplotype by the base letter of the variants that compose it. Map coordinates provided are those from NCBI Build 33 (April 2003).17

SNP name designations are consistent with those defined in dbSNP when possible.

A map with SNP location details is provided as Fig. 1.

### Statistical Methods.

SNP typing in control samples was checked for compliance with Hardy–Weinberg equilibrium with Tools For Population Genetic Analyses (TFPGA).10 Contingency table analysis for individual SNPs was carried out with SAS (SAS Institute, Inc., Cary, NC), SAS/Genetics (SAS Institute, Inc.), and SPSS (SPSS Inc., Chicago, IL). For haplotype estimation we used SNPship.11 PHASE112 (14), SNPEM13 (15), MLOCUS14 (16), an extension of the Clark (17) algorithm with expectation maximization (EM; ref. 18), and haplo.score119 (19). Self-assigned demography was checked with STRUCTURE.16 Each analysis was conducted for the case–control population as a whole; stratifying by one of six ethnicities (Asian American, African American, Hispanic, Ashkenazi Jewish, Unknown, or European American) and by age (age ≤50 and age >50). To address the effect of age of menopause, a surrogate marker (age < or ≥50) was used. Five males with breast cancer and two male controls were included in the study, but excluded for the statistical analysis presented here. No result presented here was significantly impacted when we included or excluded these males. For the purpose of haplotype estimation, the most common alleles in the study as a whole were represented by “A” for ESR1 loci; and “A” for ESR2 loci; in addition, efforts are made to describe each haplotype by the base letter of the variants that compose it. Map coordinates provided are those from NCBI Build 33 (April 2003).17

SNP name designations are consistent with those defined in dbSNP when possible.

A map with SNP location details is provided as Fig. 1.

### Statistical Methods.

SNP typing in control samples was checked for compliance with Hardy–Weinberg equilibrium with Tools For Population Genetic Analyses (TFPGA).10 Contingency table analysis for individual SNPs was carried out with SAS (SAS Institute, Inc., Cary, NC), SAS/Genetics (SAS Institute, Inc.), and SPSS (SPSS Inc., Chicago, IL). For haplotype estimation we used SNPship.11 PHASE112 (14), SNPEM13 (15), MLOCUS14 (16), an extension of the Clark (17) algorithm with expectation maximization (EM; ref. 18), and haplo.score119 (19). Self-assigned demography was checked with STRUCTURE.16 Each analysis was conducted for the case–control population as a whole; stratifying by one of six ethnicities (Asian American, African American, Hispanic, Ashkenazi Jewish, Unknown, or European American) and by age (age ≤50 and age >50). To address the effect of age of menopause, a surrogate marker (age < or ≥50) was used. Five males with breast cancer and two male controls were included in the study, but excluded for the statistical analysis presented here. No result presented here was significantly impacted when we included or excluded these males. For the purpose of haplotype estimation, the most common alleles in the study as a whole were represented by “A” for ESR1 loci; and “A” for ESR2 loci; in addition, efforts are made to describe each haplotype by the base letter of the variants that compose it. Map coordinates provided are those from NCBI Build 33 (April 2003).17

SNP name designations are consistent with those defined in dbSNP when possible.

A map with SNP location details is provided as Fig. 1.

### Statistical Methods.

SNP typing in control samples was checked for compliance with Hardy–Weinberg equilibrium with Tools For Population Genetic Analyses (TFPGA).10 Contingency table analysis for individual SNPs was carried out with SAS (SAS Institute, Inc., Cary, NC), SAS/Genetics (SAS Institute, Inc.), and SPSS (SPSS Inc., Chicago, IL). For haplotype estimation we used SNPship.11 PHASE112 (14), SNPEM13 (15), MLOCUS14 (16), an extension of the Clark (17) algorithm with expectation maximization (EM; ref. 18), and haplo.score119 (19). Self-assigned demography was checked with STRUCTURE.16 Each analysis was conducted for the case–control population as a whole; stratifying by one of six ethnicities (Asian American, African American, Hispanic, Ashkenazi Jewish, Unknown, or European American) and by age (age ≤50 and age >50). To address the effect of age of menopause, a surrogate marker (age < or ≥50) was used. Five males with breast cancer and two male controls were included in the study, but excluded for the statistical analysis presented here. No result presented here was significantly impacted when we included or excluded these males. For the purpose of haplotype estimation, the most common alleles in the study as a whole were represented by “A” for ESR1 loci; and “A” for ESR2 loci; in addition, efforts are made to describe each haplotype by the base letter of the variants that compose it. Map coordinates provided are those from NCBI Build 33 (April 2003).17

SNP name designations are consistent with those defined in dbSNP when possible.

A map with SNP location details is provided as Fig. 1.

### Statistical Methods.

SNP typing in control samples was checked for compliance with Hardy–Weinberg equilibrium with Tools For Population Genetic Analyses (TFPGA).10 Contingency table analysis for individual SNPs was carried out with SAS (SAS Institute, Inc., Cary, NC), SAS/Genetics (SAS Institute, Inc.), and SPSS (SPSS Inc., Chicago, IL). For haplotype estimation we used SNPship.11 PHASE112 (14), SNPEM13 (15), MLOCUS14 (16), an extension of the Clark (17) algorithm with expectation maximization (EM; ref. 18), and haplo.score119 (19). Self-assigned demography was checked with STRUCTURE.16 Each analysis was conducted for the case–control population as a whole; stratifying by one of six ethnicities (Asian American, African American, Hispanic, Ashkenazi Jewish, Unknown, or European American) and by age (age ≤50 and age >50). To address the effect of age of menopause, a surrogate marker (age < or ≥50) was used. Five males with breast cancer and two male controls were included in the study, but excluded for the statistical analysis presented here. No result presented here was significantly impacted when we included or excluded these males. For the purpose of haplotype estimation, the most common alleles in the study as a whole were represented by “A” for ESR1 loci; and “A” for ESR2 loci; in addition, efforts are made to describe each haplotype by the base letter of the variants that compose it. Map coordinates provided are those from NCBI Build 33 (April 2003).17

SNP name designations are consistent with those defined in dbSNP when possible.
chosen through use of the software program PHASEopybest.py.\(^{16}\) Rare haplotypes representing less than 1% of the total were deleted for the purpose of determining hSNPs. HT SNP Tester (at the same web site) was used to determine the final set of hSNPs used in the analysis, which included rare haplotypes. We used a permutation program [SNPem (22) or PHASE 2.02 (18)] to evaluate the statistical significance of any association observed between haplotype and disease state. Schaid’s program, haplo.score [Schaid et al. (19)] which uses an E-M algorithm to estimate haplotypes and then tests disease association through a general linear model, was used to verify significant associations discovered with SNPem. For sparse data, haplo.score computes simulation \(P\)-values for all score tests of association. In addition to our haplotype work, we performed two separate haplotype association analyses on blocks of strong linkage disequilibrium in ESR1.

Case–Control Sample Description. Breast cancer cases consisted of 1,006 female patients with histologically confirmed breast cancer who presented for treatment or consultation at Memorial Sloan Kettering Cancer Center from January 2000 through December 2001. DNA was obtained from peripheral blood samples. Information was obtained on age at time of diagnosis of breast cancer, age at donation of blood sample, histologic confirmation of breast cancer diagnosis, sex, and ethnicity. All DNA samples were permanently anonymized according to an Institutional Review Board (IRB)-approved protocol. Samples were unselected for family history or any other demographic characteristic. Control subjects were matched to cases on age and ethnicity, 39 Asians, and 35 subjects for whom no ethnicity or race was provided. 388 Ashkenazi Jews, 149 African Americans, 81 research subjects of Hispanic ethnicity, and 388 Ashkenazi Jews, 149 African Americans, 81 research subjects of Hispanic ethnicity, 39 Asians, and 35 subjects for whom no ethnicity or race was provided.

Verification Data Set. On completion of genotyping and analysis from the New York Academic Medical Development Corporation (NY AMDeC) study, we were provided a set of incomplete genotyping data from 298 breast cancer cases and 94 controls from a study conducted at Vanderbilt University School of Medicine. Investigators there (24) had typed five SNPs (intron 1 and early 1990s, but have recently added to their data set. Although these investigators have now typed seven SNPs (\(rs2228480\), and \(rs1801132\)) provide sufficient information to assemble 564 useful haplotypes from 282 individuals (190 cases and 94 controls).

\(^{16}\) Written by Ross Lazarus at http://www.innateimmunity.net.
Fig. 1. Maps of typed steroid receptor gene SNPs. A, typed ESR1 polymorphisms. Top bar, maps all 17 typed polymorphisms with dbSNP names and including the three synonymous polymorphisms typed in the coding region (rs2077647 is ESR1:390, which is also known as S10S in exon 1; rs1801132 is ESR1:1071 which is P325P in exon 4; and rs2228480 is ESR1:1034 which is T594T in exon 8). SYNE1, Nuclear Envelope Spectrin Repeat Protein 1, also known as Synaptic Nuclear Envelope Protein 1. Bottom bar, only those SNPs mapping to the transcribed (RefSeq) region of the gene. B, typed ESR2 polymorphisms. SYNE2, Nuclear Envelope Spectrin Repeat Protein 2, also known as Synaptic Nuclear Envelope Protein 2. C, typed PGR polymorphisms.

RESULTS

Linkage Disequilibrium. Both PGR and ESR2 SNPs have a large amount of linkage disequilibrium (Fig. 2B and C). Thus, we see a small number of haplotypes when we do frequency estimation by EM. ESR1 haplotypes indicated only moderate linkage disequilibrium in the region as represented by the D’ statistic, and more modest linkage disequilibrium as measured by a correlation coefficient (R²; Fig. 2A). An estimation of the total frequencies was done by EM and PHASE as described in Materials and Methods. These showed 585 haplotypes for 1,626 individuals with 17 loci. Codominant segregation of five alleles of ESR1 (rs851984, ESR1:1002, rs2077647, rs1801132, and rs2228480) was confirmed by screening several Centre d’Etudes du Polymorphisme Humain (CEPH) families. European-American allele frequencies in the NY AMDeC study were compared with those of CEPH founders. Each was used separately for linkage disequilibrium estimation (data not shown). Our linkage disequilibrium measurements in ESR1 were consistent with those provided by Zuppan et al. (25).

Analysis of Individual Single Nucleotide Polymorphisms in ESR1. Seventeen SNPs were typed in ESR1 with an average distance of 25,606 bp. Fifteen of these 17 variants are in the public dbSNP, whereas two are unique to Celera (hCV1576295, ESR1:1002). The latter is a T/G SNP identified by Celera as located in the promoter region upstream of untranslated exon 1C. It is ~1831 bp from the transcription start site specified by NM_000125 as elaborated in the April 2003 (Build 33, hg15) version of the human genome map. After checking Hardy–Weinberg equilibrium in each of the six sample groups (Asians, African Americans, Hispanics, Ashkenazi Jews, Unknown ethnicity, and European Americans), was carried out a two-by-two contingency table analysis on each SNP in every gene with affection status. Hardy–Weinberg compliance was obtained in each population in every control group, so long as cell numbers did not reach single digits. When cell numbers were very small, such as the Asian population, Fisher’s Exact Test evidenced no violations of Hardy–Weinberg in the control population. Each population was also stratified into two age categories: age >50 and age ≤50. A comparison of the age match between cases and controls is available as supplementary data at ftp://ftp.ncifcrf.gov/pub/users/goldb/in a folder labeled CANCERRESEARCH. Differences in the genotype distributions between cases and controls were tested with the χ² test, Fisher’s Exact Test, and a Monte Carlo test. Although no SNPs were differentially distributed among aggregate cases and controls in a statistically significant way, when the population was age-stratified into an age 50-and-under and an over-50 age group, three ESR1 SNPs were statistically significantly associated (with P-values ~ 0.01, 0.001, and 0.003 by Fisher’s Exact Test) with disease in the Jewish population; these are ESR1:1002, rs2077647, and rs827421 respectively (see Fig. 3). Whereas ESR1:1002 is located at a putative promoter site, rs2077647 is a synonymous SNP in exon 1 often described as S10S, and rs827421 is located in intron 1. Among the Ashkenazi Jewish control subjects over age 50, 82, 43, and 57% were distributed with the more common allele, at the three SNPs, respectively; whereas 89, 56, and 45 of the cases, possess that allele. Both the 3 × 2 genotype distributions and the 2 × 2 allele distributions were statistically significant departures from that expected for two SNPs. These SNPs implicate the region, denoted as the A/B region of the steroid hormone receptor, that encodes the ligand-dependent transactivation domain. As expected, a haplotype resulting from SNPs ESR1:1002 and rs2077647, which is the most common haplotype among the older Ashkenazi Jews, accounting for 50% of all haplo-
types among them, 11 (T-C), is associated with an increased risk for breast cancer among Ashkenazi Jews over 50 years of age or older [odds ratio (OR), 1.706; 95% confidence interval (CI), 1.213–2.399; \( P < 0.05 \) from SNPEM, for the dominant model and OR = 2.916, 95% CI 1.598–5.320 for the recessive model]. The details of the contingency tables are available as supplementary data at ftp://ftp.ncbi.nlm.gov/pub/users/goldb/in a folder labeled CANCERRESEARCH.9

Typing of eight htSNP (haplotype tagging) sites, composed of rs851984, rs2881766, rs20776747, hCV1576295, rs1801132, rs6905370, rs2228480, and rs3798577 in the ESR1 locus, permitted identification of individual haplotypes associated with either an increased or a decreased risk of breast cancer (see Table 2).

**Protective Haplotypes of ESR1.** Three protective haplotypes, H4, H6 and H13, evidenced a statistically significant level of protection among overall female research subjects. When only female European-
American study subjects were evaluated, only H6 conferred statistically significant protection. We sought to further define the location of the statistically protective signal through two methods: (a) redefining the haplotype with a rapid redefinition feature of SNPEM, and (b) stratifying the population by age. Research subjects over 50, bearing H6, evidenced significant protection that could be localized to the first six SNPs of the eight-SNP haplotype as H6b (see Table 2). H6b was only marginally significant with the Dirichlet method in Northern Europeans. Peak significance of the protective haplotype was obtained when only the first six htSNPs were used to estimate haplotypes in the older population sample ($P = 0.004$). With these parameters, 1.8% of the older research subjects with a six-SNP haplotype, 112221 (G-T-C-C-G-G), provided an OR of 0.330 (95% CI, 0.136–0.799). The use of PHASE 2.02, a Bayesian haplotype estimation method, in case–control mode, affirmed the statistical significance of the association (data not shown).

### Table 2. Selected ESR1 haplotypes

<table>
<thead>
<tr>
<th>Groups</th>
<th>n</th>
<th>Cases and controls</th>
<th>No. of 8 SNP haplotypes</th>
<th>Haplotype name</th>
<th>Significant haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1619</td>
<td>1006 cases; 613 controls</td>
<td>131</td>
<td>H1 11111111 (G-T-C-T-G-G-G-T)</td>
<td></td>
</tr>
<tr>
<td>European Americans</td>
<td>927</td>
<td>582 cases; 345 controls</td>
<td>112</td>
<td>H1 11111111 (G-T-C-T-G-G-G-T)</td>
<td></td>
</tr>
<tr>
<td>African Americans</td>
<td>149</td>
<td>92 cases; 57 controls</td>
<td>64</td>
<td>H1 11111111 (G-T-C-T-G-G-G-T)</td>
<td></td>
</tr>
<tr>
<td>Ashkenazi Jews</td>
<td>388</td>
<td>238 cases; 150 controls</td>
<td>76</td>
<td>H1 11111111 (G-T-C-T-G-G-G-T)</td>
<td></td>
</tr>
<tr>
<td>Asians</td>
<td>39</td>
<td>26 cases; 13 controls</td>
<td>28</td>
<td>H1 11111111 (G-T-C-T-G-G-G-T)</td>
<td></td>
</tr>
<tr>
<td>Hispanics</td>
<td>81</td>
<td>49 cases; 32 controls</td>
<td>47</td>
<td>H1 11111111 (G-T-C-T-G-G-G-T)</td>
<td></td>
</tr>
<tr>
<td>Other ethnicities</td>
<td>35</td>
<td>19 cases; 16 controls</td>
<td>36</td>
<td>H1 11111111 (G-T-C-T-G-G-G-T)</td>
<td></td>
</tr>
<tr>
<td>Age 50 and under</td>
<td>606</td>
<td>447 cases; 159 controls</td>
<td>103</td>
<td>H1 11111111 (G-T-C-T-G-G-G-T)</td>
<td></td>
</tr>
<tr>
<td>Age over 50</td>
<td>1008</td>
<td>559 cases; 454 controls</td>
<td>115</td>
<td>H1 11111111 (G-T-C-T-G-G-G-T)</td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** H, haplotype; NS, not significant.

† Five individuals did not provide age.
‡ No controls of this ethnicity with this haplotype.
Susceptible Haplotypes of ESR1. Among all female study participants, H3 and H4 evidenced statistically significant susceptibility to breast cancer. H3 evidenced susceptibility when only older study subjects were considered. This haplotype (H3) was very rare among African Americans, and did not evidence significant susceptibility in study participants ages 50 or under. Haplotype H8: 12111111 (G-G-C-T-G-G-G-T) was not associated with an increased risk of breast cancer when the data set as a whole was considered, but was significant by two measures (P-value from SNPEM = 0.014 and P-value from haplo.score = 0.02) when only self-identified Ashkenazi-Jewish research subjects were analyzed, although the 95% CIs encompass unity when males are excluded from the analysis [OR, 3.292; 95% CI, 0.945–11.471 (OR, 3.706; 95% CI, 1.076–12.757; P < 0.01 with males included in the analysis)]. A separate susceptible haplotype localized to five (of the eight htSNPs) distal to the protective haplotype. In this analysis, rs2077647, hCV1576295, rs1801132, rs6905370, and rs2228480 formed haplotype H2a 11221 (C-T-C-A-G) that accounted for 2.6% of the study subjects with a P-value for risk, an OR of 1.776 (95% CI, 1.001–3.152). This is consistent with a susceptible haplotype mapping to exon 4 and intron 4, which might implicate the ligand-binding domain of ERα.
We reanalyzed the ESR1 data with two blocks of SNPs in clear linkage disequilibrium for association with breast cancer, i.e., SNPs 2, 3, 4, 5, 6, and 7 as one block and SNPs 10, 11, 12, and 13 as another block for association analysis. We found some susceptibility in older European-American study participants (P < 0.05) in a haplotype of 221 or G-A-C for rs1884051-rs6905370-rs926778, but the ORs varied only between 1.4 and 1.8, with each 95% CI overlapping unity, yet with significance with the SNPEM permutation algorithm (22) preserved among older European Americans (P < 0.05). This haplotype was very rare in Jews, observed only twice, both times among cases.

**Results from the Vanderbilt Validation Data Set.** One haplotype, consisting of four SNPs, from a genotype complete subset of the Vanderbilt case–control study (see Materials and Methods), provided a statistically significant association with breast cancer when analyzed with SNPEM and when ORs were computed with SNPhap and SPSS. This haplotype 1211 (C-C-A-G) provided a P-value <0.05 through comparison with a permuted distribution in SNPEM, and a P < 0.004 with Fisher’s Exact Test (OR, 4.619, 95% CI, 1.378–15.481). The haplotype was present in 29 (~5%) of the estimated 568 haplotypes. Among cases, 26 among 345 weighted haplotypes were 1211 (C-C-A-G); among 173 weighted control haplotypes only 3 were 1211 (C-C-A-G). This data set and haplotype also evidenced a greater risk for breast cancer for bearers over age 50 (P < 0.01 by Fisher’s Exact Test); limiting the analysis to that group demonstrated an OR of 7.9 (95% CI, 1.025–60.868). With weighted probabilities and SNPhap on subjects over age 50 to determine the haplotypes, there were 14 cases and only 1 control with this haplotype. Whether 1211 (C-C-A-G) in the Vanderbilt data set overlaps with, or is identical to, one of the haplotypes in the NY AMDc data set cannot be determined at this time, because there is only a single SNP typed in common in the two case–control data sets (rs2077547, or S105).

**Susceptible Haplotype in ESR2 among Ashkenazi Jews.** The 388 self-identified Ashkenazi Jewish females in the study were typed for the eight markers in ESR2 at 14q23.2 described in Fig. 1. One haplotype, E2H1, formed from the last seven of these SNPs, AAAAAAA (T-C-G-G-T-A-C), beginning with rs1256030, manifested a statistically significant (P = 0.037) susceptibility for breast cancer (OR, 2.317; 95% CI, 1.042–5.155; see Table 3). This was the third most common ES2 haplotype among the Ashkenazi Jews tested, making up 11.8% of the predicted haplotypes. The 382 Ashkenazim in the study for whom complete genotypes were provided by TaqMan were further investigated with SNPEM. On reduction of the haplotype under investigation to the final five 3’ SNPs in ESR2, a new but related haplotype, E2H5, was statistically significant among the Ashkenazim, AAAAA (C-G-G-T-A-C; P = 0.001 OR, 1.82; 95% CI, 1.213–2.737). Localization of the SNPs providing the haplotype association has been possible through the use of the SNPEM permutation algorithm (Table 3). This implicated the four distal (3’) SNPs of ESR2 as giving rise to the significant haplotype because the exon 4 C or T polymorphism identified in this study is invariant among the Ashkenazi Jewish subjects.

**DISCUSSION**

These results suggest that a portion of hereditary predisposition to breast cancer can be accounted for by allele polymorphism in genes in the steroid hormone pathway. Individual differences in hormonal regulation may result from haplotypes that confer an increased risk or protection from risk of breast cancer in a subset of the population.

Although a linkage between late-onset breast cancer and ESR1 was published in 1991 (25) and several published studies find statistically significant associations between ESR1 polymorphisms and breast cancer (24, 26–30), other studies have not shown linkage or association (31–33). Each of these studies either had a relative paucity of samples or chose only a few SNPs, with no efforts to generate haplotypes for haplotype estimation and association analysis. Recently, a breast cancer protective association was reported for a GT dinucleotide repeat polymorphism 6627 bp upstream from the transcription start site of ESR1 exon 1 in a large, ethnically homogeneous Han Chinese cohort (GT18 allele; OR, 0.58; 95% CI, 0.36–0.94; ref. 29). Previously, that same group had reported a PvUII polymorphism in intron 1 that was associated with increased breast cancer risk (genotype pp; OR, 1.4; 95% CI, 1.1–1.8; ref. 30). Data suggest that the most frequent allele of this common PvUII polymorphism eliminates a functional binding site for transcription factor B-myb, which may, therefore, down-regulate ESR1 transcription (34, 35). Although we have not typed the variants described by Cai et al. (29, 30), the magnitude and direction of the protective effects seen by them are consistent with the haplotype data reported here.

Since the discovery of ESR2 in 1996 (36, 37), several groups have characterized its unique expression profile (38–44), but few have searched for polymorphisms associated with breast cancer risk (45–47). Five ESR2 polymorphisms have been identified in the German population (48), among which, one, rs1256049, provided evidence for an association with anorexia nervosa. This same ESR2 polymorphism had a highly statistically significant association with ovulatory dysfunction in a Chinese population (49). In addition, an intragenic CA repeat polymorphism in ESR2 has been associated with bone mineral density in a Japanese research subject population (50). More recently, the Shanghai breast study group reported an ESR2 exon 7 synonymous SNP (rs1256054, L392L) as conferring increased risk of breast cancer (OR, 2.37; 95% CI, 1.18–4.77) in a robust study (47). They hypothesized that this SNP may act as an exonic splicing enhancer. We are currently in the process of typing this SNP; although preliminary results indicate that it is quite rare and will likely be uninformative.

---

**Table 3 Significant ESR2 haplotypes in Ashkenazi Jewish NY AMDc research subjects**

<table>
<thead>
<tr>
<th>Haplotype name</th>
<th>Haplotype</th>
<th>Haplotype % among Ashkenazi</th>
<th>P-value using Dirichlet Distribution from SNPEM</th>
<th>P-value from haplo. score</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2H1</td>
<td>XAAAAAAA (T or G)-T-C-G-G-T-A-C</td>
<td>11.8%</td>
<td>0.037</td>
<td>0.03766</td>
<td>2.317 (1.042–5.155)</td>
</tr>
<tr>
<td>E2H2</td>
<td>XAAAAAAA (T or G)-T-C-G-G-T-A</td>
<td>11.8%</td>
<td>0.002</td>
<td>0.00257</td>
<td>2.043 (1.243–3.358)</td>
</tr>
<tr>
<td>E2H3</td>
<td>XAAAAAA (T or G)-T-C-G-G-T</td>
<td>11.8%</td>
<td>0.002</td>
<td>0.00344</td>
<td>1.951 (1.195–3.184)</td>
</tr>
<tr>
<td>E2H4</td>
<td>XAAAAAA (T or G)-T/C-C-G-G-T-A</td>
<td>11.8%</td>
<td>0.009</td>
<td>0.01391</td>
<td>2.339 (1.140–4.797)</td>
</tr>
<tr>
<td>E2H5</td>
<td>XAAAAAA (T or G)-T/C-C-G-G-T-A</td>
<td>17.6%</td>
<td>0.001</td>
<td>0.00267</td>
<td>1.822 (1.213–2.737)</td>
</tr>
</tbody>
</table>

Note: Places in the haplotype table above represent variants in the ESR2 gene typed in Ashkenazi Jewish research subjects in this study. The most common alleles were typed as “A” for SNPEM input. Most common alleles were “T” in rs1271572, “T” in rs1256030, “C” in EX4CorT, “G” in rs1256049, “T” in rs4986938, “A” in rs928554, and “G” in rs1255998. “C” in EX4CorT was invariant among Ashkenazi Jews tested.
We typed six intronic and 3’ SNPs in addition to six of those characterized by De Vivo et al. (51: +44 CT, +331 GA, S344T, G393G, V660L, and H770H). As can be seen through inspection of Fig. 2C, we found significant linkage disequilibrium throughout the PGR gene as gauged by either the D’ or R² statistic computed for the 12 SNPs that we typed. However, we found no single SNP, nor any haplotype, that was significantly associated with breast cancer when we stratified by age, ethnicity, or both. We sought to determine why we were not able to replicate the association of the functional +331 G/A polymorphism with breast cancer risk reported by De Vivo et al. (51). Whereas they found 87% GG at this site among their cases and 90% GG among their controls, we found 93.4% and 93% among our cases and controls, respectively; and whereas they found 87% of this same genotype among their postmenopausal cases with 90% among their postmenopausal controls, we found 93% GG and 93.2% GG among our older cases and controls, respectively. We sought to determine whether this disparity could be explained by demographic differences between the Nurses Health Study participants, polled by De Vivo, and our research subjects by stratifying our results by both ethnicity and age. This analysis of +331 G/A genotypes in cases and controls also yielded no association.

This study has several limitations inherent in its study design. To gain IRB approval, we permanently anonymized DNA samples after collection of a minimum clinical annotation. Thus, it is not now possible to retrospectively examine clinical or demographic records to examine additional potential confounding variables, such as endogenous or exogenous estrogen exposure and other environmental variables. To address the effect of age of menopause, we used a surrogate marker (age < or >50).

Second, our study, although among the largest to date, still lacks statistical power to come to firm conclusions concerning the relationship of ESR1, ESR2, or PGR SNPs or haplotypes with respect to the specific populations tested. Our genotyping adhered to reproducibility and control standards published elsewhere (13), and we meet recently published genotyping standards (52). In addition, we demonstrated haplotype segregation in the CEPH families on the five initial SNPs published by Celera and tested and obtained Hardy–Weinberg equilibrium in each of our control populations, even although some were quite small.

Although we made an effort to verify our findings in an independent data set, this additional analysis was also underpowered. Nonetheless, the Vanderbilt verification data set confirmed the existence of an ESR1 risk haplotype, which includes the aforementioned P40II SNP, although it may not be precisely the same haplotype identified in the NY AMDeC study or the Shanghai breast cancer study.

With regard to the statistical limitations inherent in our study design, we used a variety of methods of imputing haplotypes to provide evidence for statistically significant disease associations. Whereas Nyholt (53) has emphasized a critical need for multiple test corrections in disease association studies to avoid Type I error, Krawczak et al., (54) and others (55, 56) have questioned the overzealous application of the Bonferroni (57) correction. In an attempt to avoid potential errors in inference about associations stemming from the method of determination of haplotypes, we applied several different haplotype inference procedures. We observed that the methods found the common haplotypes accurately, but there was some disagreement in the rare haplotypes. With a variety of methods, some assuming underlying normal distributions, others being permutation based and still others being Bayesian, we have found similar haplotype frequencies in the genes under investigation. Such methods include those of Fallin et al.(15) and Fallin and Schork (22, 58), whose haplotype assignment, affection status permutation algorithm has been adopted by Schaid et al. (19) and modified by Stephens and Donnelly (59). These methods (22) use EM haplotype estimations, which assume compliance with Hardy–Weinberg equilibrium, whereas Stephens’ Gibbs sampling algorithm bases haplotype estimations on existent allele frequencies. In all cases of significant association between haplotypes and breast cancer risk, these methods provided consistent haplotype inferences.

Using a variety of statistical and laboratory methods, we have discovered and validated the presence of common polymorphisms in three sex steroid hormone receptor genes. These candidate genes for disease association analysis have functional significance for the etiology under study and, therefore, cannot be thought of as impartially selected for statistical association testing. With cases and controls from the same geographical area and matched for age, we analyzed patterns of linkage disequilibrium and affection trait association with these genetic variants. We were able to define ESR1 haplotypes that conferred significant association with breast cancer risk in a North American population. Whereas the “protective” alleles that have been identified are quite rare in the populations studied, and their overall contribution to disease may be quite small, this analysis of SNP genotypes provides a means to associate variants in steroid hormone receptor genes and the breast cancer phenotype. Continued study of haplotypes of candidate genes in the steroid hormone receptor signal transduction pathway will provide additional insight into the biology of breast neoplasia.

ACKNOWLEDGMENTS

The authors thank Drs. Vanessa Clark and Daniele Fallin for helpful comments on the statistical analysis. Drs. Robert Stephens and Matthew Stephens generously provided recompiled software. The authors would like to acknowledge the New York Cancer Project, which, in connection with the publication of this study, made available biological samples from and information on control individuals. The New York Cancer Project is administered and funded by AMDeC Foundation, Inc. The content of this publication does not necessarily reflect the views of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the United States Government.

REFERENCES

Announcements

MEETING OF THE RADIATION RESEARCH SOCIETY

The annual meeting of the Radiation Research Society will be held at the State University of Iowa, Iowa City, on June 22-24, 1953. The Society will be the guest of the University, and all meetings will be held on the campus. The program will consist of: (1) Two symposia, one on "The Effects of Radiation on Aqueous Solutions," which includes the following speakers: E. S. G. Barron, Edwin J. Hart, Warren Garrison, J. L. Magee, and A. O. Allen. The second is "Physical Measurements for Radiobiology" and companion talks by Ugo Fano, Burton J. Moyer, G. Failla, L. D. Marinelli, and Payne S. Harris. (2) On Monday night, June 22, a lecture by Dr. L. W. Alvarez on meson physics has been tentatively scheduled. On Tuesday night, June 23, Dr. L. H. Gray of the Hammersmith Hospital, London, will speak on a topic to be announced. Dr. Gray's lecture is sponsored by the Iowa Branch of the American Cancer Society. Those desiring to report original research in radiation effects, or interested in attending or desiring additional information, please contact the Secretary of the Society, Dr. A. Edelmann, Biology Department, Brookhaven National Laboratory, Upton, L.I., New York.

ERRATUM

The following correction should be made in the article by Beck and Valentine, "The Aerobic Carbohydrate Metabolism of Leukocytes in Health and Leukemia. I. Glycolysis and Respiration," November, 1952, page 891; substitute for the last paragraph:

The data in Table 3 permit several interesting calculations. If one compares the amount of glucose actually disappearing with the sum of the amount equivalent to lactic acid produced plus that equivalent to O₂ consumption, it is seen that the amount of glucose "cleavage products" exceeds the amount of glucose utilized by 12 per cent in N and 27 per cent in CML and is exceeded by the glucose utilized by 16 per cent in CLL. If the assumption is made that, in this respect, the myeloid and lymphoid cells of leukemia are similar to those of normal blood, it may be that the computed normal figure represents a summation of the myeloid (M) and lymphoid (L) cells that make up the normal leukocyte population. Thus, if M = +0.27 and L = -0.16 and the normal differential is 65 per cent M and 35 per cent L, then

0.65 (+0.27) + 0.35 (-0.16) = +0.12,

a figure identical to the observed +0.12 for normal leukocytes.
Estrogen Receptor Genotypes and Haplotypes Associated with Breast Cancer Risk

Bert Gold, Francis Kalush, Julie Bergeron, et al.

Cancer Res 2004;64:8891-8900.

Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/64/24/8891

Cited articles
This article cites 53 articles, 14 of which you can access for free at:
http://cancerres.aacrjournals.org/content/64/24/8891.full.html#ref-list-1

Citing articles
This article has been cited by 15 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/64/24/8891.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.