Large-Scale Association Study Identifies ICAM Gene Region as Breast and Prostate Cancer Susceptibility Locus

Stefan Kammerer,1 Richard B. Roth,1 Richard Reneland,1 George Marnellos,1 Carolyn R. Hoyal,1 Nathan J. Markward,1 Florian Ebner,2 Marion Kiechle,3 Ulrike Schwarz-Boeger,3 Lyn R. Griffiths,4 Christian Ulbrich,5 Korbinian Chrobok,6 Gerhard Forster,5 Georg M. Praetorius,5 Peter Meyer,7 Joachim Rehbock,2 Charles R. Cantor,1 Matthew R. Nelson,1 and Andreas Braun1

ABSTRACT
We conducted a large-scale association study to identify genes that influence nonfamilial breast cancer risk using a collection of German cases and matched controls and >25,000 single nucleotide polymorphisms located within 16,000 genes. One of the candidate loci identified was located on chromosome 19p13.2 [odds ratio (OR) = 1.5, P = 0.001]. The effect was substantially stronger in the subset of cases with reported family history of breast cancer (OR = 3.4, P = 0.001). The finding was subsequently replicated in two independent collections (combined OR = 1.4, P < 0.001) and was also associated with predisposition to prostate cancer in an independent sample set of prostate cancer cases and matched controls (OR = 1.4, P = 0.002). High-density single nucleotide polymorphism mapping showed that the extent of association spans 20 kb and includes the intercellular adhesion molecule genes ICAM1, ICAM4, and ICAM5. Although genetic variants in ICAM5 showed the strongest association with disease status, ICAM1 is expressed at highest levels in normal and tumor breast tissue. A variant in ICAM5 was also associated with disease progression and prognosis. Because ICAMs are suitable targets for antibodies and small molecules, these findings may not only provide diagnostic and prognostic markers but also new therapeutic opportunities in breast and prostate cancer.

INTRODUCTION
Breast cancer is the most common form of cancer in women and accounts for over 400,000 deaths each year worldwide. Among the risk factors identified are family history, diet, reproductive history, exogenous estrogen exposure, smoking, and alcohol consumption. The heritability of breast cancer is estimated at about 25% (1), but germine mutations in so-called high-penetrance susceptibility genes, such as BRCA1 and BRCA2 (2), explain <10% of all cases. Therefore, it is expected that more common genetic variations with low penetrance account for a high proportion of breast cancer cases.

Family based linkage studies have proven very successful for identifying genes with highly penetrant alleles that segregate in a simple Mendelian fashion within families (3). However, successes in which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare diseases and to selected “familial” subsets of common disorders, which linkage methods are used have largely been restricted to rare disease suscep-

With different families, the contributions of the different genes will likely vary, thus complicating the ability to detect and replicate linkage.

For common disease susceptibility genes, direct association approaches that compare the distribution of genetic marker frequencies between groups of unrelated individuals are expected to have greater power than traditional linkage studies (5). Recently, there has been increasing interest in the use of whole-genome association methods to identify genes that are involved in complex trait variation (6–8). To date, however, few large-scale studies have been reported (9–11), and the search for common polymorphisms that influence breast cancer risk has yet to produce useful markers. In an effort to identify genes and variants that influence breast cancer susceptibility, we conducted a large-scale case-control study using >25,000 single nucleotide polymorphisms (SNPs) located within approximately 16,000 genes. We report the identification of intercellular adhesion molecule variants on chromosome 19p13.2 that confer increased breast and prostate cancer risk and might predict disease outcome.

MATERIALS AND METHODS

Study Subjects and Preparation. The discovery sample comprised 254 breast cancer patients attending the Frauenklinik Innenstadt, Munich, Germany. Lymph node status was positive at time of assessment in 94 cases (37%), and 18 cases (7%) had known distant metastases. Twenty-four cases (11%) reported a positive family history of breast cancer. The median age at diagnosis was 56 years (range = 23–87 years). Two hundred and sixty-eight controls with a median age of 57 years (range = 17–88 years) were recruited from patients with benign disease attending the clinic during the same period. Controls with a family history of breast or ovarian cancer were excluded. Both parents of each study participant were reported to be of German descent.

The German replication sample consisted of 188 cases and 150 controls recruited at the Department of Obstetrics and Gynecology, Technical University of Munich. The majority of breast cancer cases were recruited at preoperative visits, and female controls were recruited from healthy individuals or patients with nonmalignant diagnoses. Median age of diagnosis for cases was 59 years (range = 22–87 years), and median age of controls was 50 years (range = 19–91 years). All but two participants reported both parents were of German descent. The two exceptions each reported one parent of non-German, Eastern European origin.

The Australian replication sample comprised 180 breast cancer cases recruited by the Pathology Department of Gold Coast Hospital or by the Genomics Research Center, Southport. Median age of diagnosis was 50 years (range = 24–74 years). Controls consisted of 180 healthy volunteers recruited through the Genomics Research Center. Only controls with no family history of cancer or precancerous conditions were included. Controls were individually age matched to cases (±5 years). Median age for controls was 60 years (range = 28–94 years).

The prostate cancer sample consisted of 368 German patients with a median age of diagnosis of 65 years (range = 43–90 years) recruited at the Urology Clinic Munich-Planegg. and 368 controls without symptomatic prostate disease with a median age of 68 years (range = 25–92 years) collected at the
University Hospital of Tuebingen. All subjects involved reported both parents to be of German ancestry.

All subjects involved in our studies signed a written informed consent and the institutional ethics committees of participating institutions approved the experimental protocols. For discovery samples, we extracted DNA from 5 mL of blood from each subject using a desalting method and quantified fluorometrically (Fluoroskan Ascent CF, Labsystems, Helsinki, Finland) using Pico green. DNA pools were generated by combining equimolar amounts of each sample as described elsewhere (12, 13).

**SNP Markers and Genotyping.** A set of 25,494 SNP markers was selected from a collection of 125,799 experimentally validated polymorphic variations (14). This set was limited to SNPs located within gene coding regions, minor allele frequencies >0.02 (95% have frequencies >0.1), and a target inter-marker spacing of 40 kb. SNP annotation is based on National Center for Biotechnology Information dbSNP database, refSNP, build 118.8 Genomic annotation is based on National Center for Biotechnology Information genome build 34. Gene annotation is based on LocusLink genes for which National Center for Biotechnology Information was providing positions on the Mapview FTP site.9

For pooled DNA assays, 25 ng of case and control DNA pools were used for amplification at each site. We conducted all PCR and MassEXTEND (Sequenom, Inc., San Diego, CA) reactions using standard conditions (13). Relative allele frequency estimates were derived from area under the peak calculations of mass spectrometry measurements from four analyte aliquots as described elsewhere (13). For individual genotyping, the same procedure was applied except only 2.5 ng of DNA was used, and only one mass spectrometry measurement was taken.

**Statistical Analysis.** We carried out tests of association between disease status and each SNP using pooled DNA in a similar fashion as explained elsewhere (15). Sources of measurement variation included pool formation, PCR/mass extension, and chip measurement. When three or more replicate measurements of a SNP were available within a model level, the corresponding variance component was estimated from the data. Otherwise, the following historical laboratory averages were used: pool formation = 5.0 × 10^{-5}, PCR/mass extension = 1.7 × 10^{-4}, and chip measurement = 1.0 × 10^{-4}. We carried out tests of association using individual genotypes that used a χ² test of heterogeneity based on allele and genotype frequencies.

The DerSimonian-Laird random effects meta-analysis method (16) was used for the analysis of heterogeneity based on allele and genotype frequencies. The DerSimonian-Laird random effects meta-analysis method (16) was used for the analysis of association of SNPs in ICAM1 and ICAM5 with breast cancer susceptibility.

**RESULTS**

We carried out a genome-wide association study using 25,494 SNPs located within 10 kb of 15,995 LocusLink annotated genes and a sample consisting of 254 breast cancer cases and 268 age-matched controls of German descent. We used a high-throughput approach using DNA pools, chip-based mass spectrometry (11–13, 19), and a three-step SNP selection strategy. In the first step, we did a single PCR and primer extension reaction for each SNP on one DNA pool consisting of cases and on one consisting of controls. Relative allele frequencies obtained from four mass spectrometry measurements of the extension products were compared between pools. In the second step, the 1,619 SNPs (~5%) with the most statistically significant associations were remeasured in triplicate on each DNA pool. In the third step, the 74 most significant SNPs (~5%) from step two were individually genotyped in each sample. Fifty-two SNPs were confirmed to have statistically significant differences between cases and controls (P < 0.05).

**Table 1** Analysis of association of SNPs in ICAM1 and ICAM5 with breast cancer susceptibility.

| Sample        | N * | MAF † | OR ‡ | P value § | Genotype frequencies ¶ | P value ||
|---------------|-----|-------|------|-----------|-----------------------|--------|
| rs1056538     |     |       |      |           |                       |        |
| ICAM5 Exon 5  |     |       |      |           |                       |        |
| (V301I)       | German Controls | 266  | 0.44 | 1.54 0.001 | 0.21 0.47 0.32 0.006 |
|               | (Discovery) Cases | 231  | 0.34 | 1.10 0.28 | 0.13 0.42 0.45 |
|               | German Controls | 143  | 0.43 | 1.35 0.07 | 0.18 0.50 0.32 0.15 |
|               | (Replication) Cases | 181  | 0.36 | 1.10 0.27 | 0.14 0.43 0.43 |
|               | Australian Controls | 175  | 0.41 | 1.27 0.03 | 0.19 0.43 0.38 0.21 |
|               | (Replication) Cases | 175  | 0.35 | 1.30 0.01 | 0.17 0.36 0.47 |
|               | Total |     |      |           | 1.39 0.0001 |        |
| rs5030382     |     |       |      |           |                       |        |
| ICAM5 Exon 6  |     |       |      |           |                       |        |
| (K469E)       | German Controls | 265  | 0.46 | 1.39 0.01 | 0.22 0.49 0.29 0.04 |
|               | (Discovery) Cases | 242  | 0.38 | 1.19 0.27 | 0.16 0.45 0.39 |
|               | German Controls | 142  | 0.48 | 1.27 0.07 | 0.25 0.47 0.28 0.45 |
|               | (Replication) Cases | 178  | 0.42 | 1.14 0.20 | 0.19 0.48 0.34 |
|               | Australian Controls | 170  | 0.41 | 1.19 0.20 | 0.19 0.44 0.36 0.73 |
|               | (Replication) Cases | 167  | 0.38 | 1.20 0.05 | 0.17 0.43 0.40 |
|               | Total |     |      |           | 1.28 0.0004 |        |
| rs2814394     |     |       |      |           |                       |        |
| ICAM5 542 bp  |     |       |      |           |                       |        |
| upstream      | German Controls | 257  | 0.23 | 1.09 0.57 | 0.04 0.37 0.59 0.74 |
|               | (Discovery) Cases | 242  | 0.24 | 1.19 0.57 | 0.06 0.35 0.70 |
|               | German Controls | 143  | 0.19 | 1.19 0.57 | 0.06 0.26 0.69 0.03 |
|               | (Replication) Cases | 179  | 0.26 | 1.19 0.57 | 0.06 0.35 0.63 0.25 |
|               | Australian Controls | 174  | 0.21 | 1.33 0.05 | 0.05 0.33 0.63 0.25 |
|               | (Replication) Cases | 176  | 0.26 | 1.41 0.004 | 0.09 0.35 0.56 |
|               | Total |     |      |           | 1.27 0.02          |        |

* Number of subjects with genotypes.
† Minor relative allele frequency.
‡ Odds ratio with reference to allele that increases frequency in cases.
§ P value for test comparing allele frequencies between cases and controls.
¶ Relative genotype frequencies.
|| P value for tests comparing genotype frequencies between cases and controls.

original marker rs1056538 (V301I) and rs2228615 (A348T), which were in near complete linkage disequilibrium (Fig. 2E). A nonsynonymous SNP in ICAM1, rs5030382 (K469E), was significantly associated in the discovery sample but not in the replication samples. Analyses of haplotypes consisting of subsets of the 15 genotyped SNPs did not reveal any haplotype with stronger association than individual SNPs (data not shown).

Because there are several common features between breast and

Fig. 1. Association fine mapping of breast cancer susceptibility region on chromosome 19p13.2. Sixty public domain SNPs in a 100-kb window around the incident SNP (bold “+” in center) were compared between pools of cases and controls. The Y-axis corresponds to their chromosomal position and the Y-axis to the test P values (shown on the −log10 scale). The continuous dark line presents the results of a goodness-of-fit test for an excess of significance (compared with 0.05) in a 10-kb sliding window assessed at 1-kb increments. The continuous light gray line is the result of a nonlinear smoothing function showing a weighted average of the P values across the region. The darkness of each point corresponds to the minor allele frequency of each SNP in the control sample (see legend below graph). The LocusLink gene annotations for National Center for Biotechnology Information genome build 34 are included.

Fig. 2. Genotype analysis of 15 SNPs in a 20-kb window around the ICAM region. A, P values for each SNP in the breast cancer discovery sample. B, meta-analysis P values of the breast cancer replication samples (gray) and all three samples (black). C, P values of the prostate cancer samples. D, Gene map and chromosomal positions as in Fig. 1. SNP locations are indicated as tick marks. E, estimates of linkage disequilibrium in the combined breast cancer sample. Grey-scale ranges from white (linkage disequilibrium = 0) to black (linkage disequilibrium = 1) with increments of 0.1.
prostate cancer (20), we also tested this locus for association with prostate cancer susceptibility in an independent collection of German cases and controls. Thirteen of the 15 SNPs genotyped in the breast cancer sample were genotyped in the prostate cancer sample. The resulting association pattern was similar to that observed for breast cancer. The nonsynonymous SNPs, rs1056538 and rs2228615, in *ICAM5* were most strongly associated (OR = 1.4, P = 0.002; Fig. 2C).

The discovery collection of German breast cancer patients included information related to family history, age of onset, and severity of disease. The genotyped SNPs were tested for association with these variables. The SNP rs1056538 and rs2228615 were most strongly associated with a positive family history of breast cancer (P = 0.0065, Fisher’s exact test). Fifteen percent of those homozygous for the susceptible allele (C) had a positive family history, compared with 9% of the heterozygotes and none of those homozygous for the protective allele. Comparing allele frequencies between cases with a positive family history and controls results in a substantially larger estimated effect size than in all cases (OR = 3.4, P = 0.001). There was no association between rs1056538 and age of diagnosis. Also noteworthy, the SNP most strongly associated with breast cancer status in the replication samples, rs281439 located 542 bp 5' of *ICAM5* (Table 1), was also associated with indicators of cancer severity (Table 2). Patients carrying the allele associated with breast cancer susceptibility (G) had a significantly shorter time span between diagnosis and recruitment, suggesting shorter survival time, and higher rates of metastases to other organs. These results suggest that one or more variants in this region are risk factors for breast and prostate cancer and may influence disease progression and prognosis.

Using semiquantitative RT-PCR, we found expression of *ICAM1* in a variety of normal tissues, including breast and prostate and several breast tumor samples tested (data not shown). As described previously, *ICAM5* showed strongest expression in brain and, along with *ICAM4*, was expressed at very low levels in various other tissues, including breast and prostate, and a number of breast tumor tissues and cell lines.

**DISCUSSION**

In an association study using SNPs in nearly 16,000 genes, we obtained evidence that a region on chromosome 19p13.2 containing the genes *ICAM1, ICAM4*, and *ICAM5* influences breast and prostate cancer risk. The identification of a susceptibility region influencing both breast and prostate cancer is particularly interesting because breast and prostate cancer have many common features, such as hormone-sensitivity, parallel incidence rates in various countries, and common genetic alterations (20). The role of intercellular adhesion molecules in breast cancer has been described (21, 22). *ICAM1* is constitutively expressed in cells involved in immune response and induced on other cells, including endothelial and epithelial cells. It has a well-known role in inflammation-related processes and immune surveillance, and derangements of its expression have been implicated in the development of a variety of inflammatory diseases (23–25) as well as in tumor progression of several cancers, including breast cancer (26–28). *ICAM1* is also involved in transmembrane signal transduction after binding to β2 integrin ligands and multimer formation (29), activating the mitogen-activated protein kinase pathway (30) and eventually transcription factors like AP-1 that regulate cell proliferation events (31). The inhibition of AP-1 has been shown to inhibit breast cancer cell growth (32). Relatively little is known about the roles of *ICAM4* and *ICAM5* in cell signaling events and tumor surveillance, but their involvement in similar pathways is likely. *ICAM4* has been reported to be exclusively expressed in erythrocytes and has a suggested role in cell interaction events, including hemostasis and thrombosis (33). *ICAM5* is mainly expressed in specific areas of the brain (34) and has been implicated in dendritic outgrowth (35) and rapid cell spreading of microglia (36). Although the reported expression patterns and described functions of *ICAM4* and *ICAM5* are not indicative of a role in breast and prostate cancer susceptibility, their roles in cell adhesion and cell signaling together with their low level expression in cancer-relevant tissues leave the possibility that their dysregulation or dysfunction may increase cancer risk.

Our findings are in agreement with previous reports on the involvement of *ICAM1* in tumor progression, and the K469E variant is a potential candidate to influence this process. However, the data also suggests an influence of other variants in this region on breast and prostate cancer susceptibility. Although the genetic evidence favors *ICAM5*, higher relative expression levels of *ICAM1* make it a more favorable candidate for predisposition and potentially tumor progression. However, current data cannot exclude any of the three *ICAM* genes as biologically responsible for the observed association.

The route by which these genetic associations were arrived at and the potential for spurious association must certainly be considered. Recent published work has brought much needed light to the need for proper validation to verify genetic findings for complex traits (37–39). In the current study, the initial association found between the *ICAM5* marker and breast cancer status was one result from over 25,000 hypothesis tests. A conservative Bonferroni adjustment to yield an experiment-wide type I error rate of 0.05 would demand a test-wise *P* value on the order of 10^-6_. Given the modest sample size, only common variants with relatively large effects (OR > 2) would reach such significance levels. Instead, we chose to be more mindful of the role of type II error rates and apply a more liberal set of criteria in the initial phases of the study and verify true genetic effects by independent replication. The analysis of 52 selected markers in the German and Australian replication samples resulted in multiple associations of continuing interest, with the *ICAM5* providing the most consistent and statistically significant association (P = 0.01, OR = 1.3). This would not be considered significant on an experiment-wide level after Bonferroni adjustment, which would require a test-wise *P* value on the order of 0.001. Indeed, if the true size of the genetic effect is 1.4, then the total replication sample size of 368 cases and 330 controls has <50% power to reject the false null hypothesis at this level.

It should be noted that we observed two additional lines of evidence supporting the validity of the association of this region to breast cancer susceptibility. First, the risk allele for coding SNP rs1056538 is significantly more common in cases with a family history of breast cancer. Second, the independent identification of the same variation with remarkably similar effects observed in a larger prostate cancer study gives us added reason to believe that one or more variations in these *ICAM* genes are influencing breast and prostate cancer susceptibility in populations of European origin. These are characteristics that have also been observed for the germ line variations in *CHEK2* (40, 41).

The successful identification of a small gene region associated with breast and prostate cancer susceptibility through large-scale associa-

Table 2 Association of rs281439 with clinical indicators of breast cancer severity

<table>
<thead>
<tr>
<th>Genotype</th>
<th>N (N = 14)</th>
<th>GC (N = 90)</th>
<th>CC (N = 139)</th>
<th><em>P</em> value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of diagnosis (years)</td>
<td>235</td>
<td>49/54/61 *</td>
<td>49/56/63</td>
<td>0.67</td>
</tr>
<tr>
<td>Years since diagnosis</td>
<td>233</td>
<td>0.6/4.4/4.6</td>
<td>0.3/5.5/2</td>
<td>1.1/3.0/6.6</td>
</tr>
<tr>
<td>Lymph node metastases (yes)</td>
<td>254</td>
<td>43% (6)</td>
<td>29% (26)</td>
<td>29% (41)</td>
</tr>
<tr>
<td>Organ metastases (yes)</td>
<td>254</td>
<td>29% (4)</td>
<td>8% (7)</td>
<td>4% (6)</td>
</tr>
</tbody>
</table>

* Quantitative variables are expressed as 1st quartile/median/3rd quartile. Categorical variables are expressed as column percentage (count).
tion supports the promise of more comprehensive genome-wide studies. The approach presented here using DNA pools and genome-wide gene-based SNPs is very useful to quickly discover susceptible genes for various cancer types and other complex disorders. However, given the complexity of the genetic architecture underlying complex trait variation, genetic analyses, alone, are not likely to unambiguously identify the genes or genetic variations responsible. Further validation in independent samples must be carried out to support genetic evidence, and functional experiments must be conducted to identify which variations in which genetic and environmental contexts are associated with disease susceptibility and prognosis. The observed association of genetic variants to disease severity and progression has promising implications for patient management. Because ICAMs are cell surface molecules and therefore obvious targets for antibodies and small molecules, these findings also offer new opportunities for therapeutical intervention.

ACKNOWLEDGMENTS

We would like to thank Richard Kolodner for critically reviewing the report and making helpful comments. We would also like to thank all patients participating in this study and the members of Sequenom’s high throughput center team for generating the data.

REFERENCES