TGF-β Signaling Pathway and Breast Cancer Susceptibility

Serena Scollen, Craig Luccharini, Caroline Baynes, Kristy Driver, Manjeet K. Humphreys, Montserrat Garcia-Closas, Jonine Figueroa, Jolanta Lissowska, Paul D. Pharoah, Douglas F. Easton, Robin Hesketh, James C. Metcalfe, and Alison M. Dunning

Abstract

Background: TGF-β acts as a suppressor of primary tumor initiation but has been implicated as a promoter of the later malignant stages. Here associations with risk of invasive breast cancer are assessed for single-nucleotide polymorphisms (SNP) tagging 17 genes in the canonical TGF-β ALK5/SMADs 1&5 and ALK1/SMADs 1&5 signaling pathways: LTBP1, LTBP2, LTBP4, TGFBR1, TGFBR2, TGFBR3, TGFBR1(ALK5), ALK1, TGFBR2, Endoglin, SMAD1, SMAD2, SMAD3, SMAD4, SMAD5, SMAD6, and SMAD7. Significant associations were meta-analyzed with data from the NCI Polish Breast Cancer Study (PBCS; 1,966 cases and 2,347 controls) and published data from the Breast Cancer Association Consortium (BCAC).

Methods: Three-hundred-fifty-four tag SNPs (minor allele frequency > 0.05) were selected for genotyping in a staged study design using 6,703 cases and 6,840 controls from the Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH) study. Significant associations were meta-analyzed with data from the PBCS and BCAC. Tumor subtype analyses indicated that the TGFBR1 rs1982073 association may be confined to increased risk of developing progesterone receptor negative (PR−) tumors (1.18 (95% CI: 1.09–1.28), 4.1 x 10−5 (P value for heterogeneity of ORs by PR status = 2.3 x 10−4)). There was no evidence for breast cancer risk associations with SNPs in the endothelial-specific pathway utilizing ALK1/SMADs 1&5 that promotes angiogenesis.

Conclusion: Common variation in the TGF-β ALK5/SMADs 2&3 signaling pathway, which initiates signaling at the cell surface to inhibit cell proliferation, might be related to risk of specific tumor subtypes.

Impact: The subtype specific associations require very large studies to be confirmed.

Introduction

Variants that increase risk of breast cancer fall into 3 categories: rare, high-penetrance alleles; moderate-risk alleles; and common, low-risk alleles. Rare high/moderate-risk inherited mutations (BRCA1, BRCA2, TP53, CHEK2, PALB2, and BRIP1) account for approximately 27% of excess familial risk of breast cancer (1). These risk alleles were identified by linkage analysis/candidate gene resequencing approaches, and extensive efforts to find more susceptibility alleles have been unsuccessful. The remaining risk is likely to be caused by the cumulative effect of multiple moderate/lower-risk inherited variants (2). Confirming this, genome-wide association studies (GWAS) have already identified 17 common, low-penetrance loci; for example, FGFR2, TNRC9/LOC643714, and NEK10/SLC4A7 (3–7). Few common single-nucleotide polymorphism (SNP) associations, CASP8 D302H and TGFBR1 L10P (rs1982073; ref. 8), have been identified via the candidate gene approach, which has been hampered by both a lack of understanding of the biology of breast cancer and small, underpowered studies. The TGFBR1 Pro allele, which has been shown to increase TGFBR1 secretion in vitro (9), has been reported to be associated with an increased risk of breast cancer relative to the Leu allele: OR 1.08, 95% CI: 1.04–1.11, P = 2.8 x 10−5 (8). The
functionally relevant repeat length polymorphism in exon I of the TGFBR1 (TGFBR1*6A) has also been examined as a candidate; a recent meta-analysis of 15 breast cancer studies including 10,826 cases and 12,964 controls reported a significant association for allelic effect (6A vs. 9A): OR 1.16, 95% CI: 1.01–1.34, \( P = 0.04 \) (10).

Other variants in TGF-β signaling pathway genes have also been studied in relation to association with breast cancer risk, with the main focus previously being an SNP in the promoter region of TGFB1 (c-509t; 3987 cases, 3867 controls: OR 1.25 95% CI: 1.06–1.48, \( P = 0.009 \); ref. 9). The c-509t variant is in linkage disequilibrium (LD) with the L10P polymorphism previously discussed (\( r^2 = 0.69 \) in Stage 1 data presented here) and probably marks the same causal variant.

SNP associations in TGF-β signaling pathway genes have not been detected by recent breast cancer GWAS (3–7).

There is accumulating evidence that the TGF-β signaling pathway (Fig. 1) has a dual role, acting both in initial tumor development and in later tumor progression. A broad range of evidence for TGF-β acting as a tumor suppressor gene comes from humans, in particular from studies of epithelial colorectal tumors, and from transgenic mouse studies. However, when a primary tumor has been established, TGF-β may act to enhance subsequent tumor progression, based on studies of human cells in vitro and of transgenic mouse models in which TGF-β signaling is modulated (11, 12). Studies in endothelial cells in vitro have indicated that TGF-β may regulate angiogenesis via the balance of signaling through 2 pathways, activated by the antiangiogenic receptor ALK5 (TGFBR1) and the proangiogenic receptor ALK1 (Fig. 1; Refs. 13–15). It has been proposed that shifting the balance to a proangiogenic role for TGF-β may be a significant mechanism by which it enhances tumor progression. The expression of TGF-β signaling factor phospho-SMAD2 has recently been reported to be greater in human breast cancers associated with lymph node metastasis, consistent with a proprogression role for TGF-β (16). TGF-β signaling patterns were found to vary with age and pathologic features of prognostic significance, further supporting the proposal that the pathway may be associated with disease progression in humans.

Functionally relevant gene variants have been previously identified in 2 TGF-β signaling pathway genes, supporting the established rationale for additional studies of candidate genes that belong to cancer-related pathways (17). TGF-β signaling pathway genes are attractive candidates.

In this association, study using invasive breast cancer cases and controls, we comprehensively tagged the common variants in 17 genes comprising both arms of the TGF-β signaling pathways (Fig. 1).

![Figure 1. TGF-β signaling pathway](link)

Once cleaved from latent TGF-β binding protein (4 isoforms, LTBP1–4), the mature and active dimeric TGF-β (3 isoforms, TGFB1–3) can bind to heteromeric receptor complexes formed by type I (TGFBR1; ALK5 and ALK1 (activin A receptor type II-like 1)) receptors and type II (TGFBR2) receptors, initiating signal transduction. The binding of TGF-β to the receptor complex is facilitated by type III coreceptors, such as Endoglin. Endoglin is expressed at particularly high levels in endothelial cells (32). Activation of receptor kinase domains initiates downstream signaling by phosphorylation of transcription factors (e.g., SMAD1–5; SMAD family members). The SMAD complex migrates to the nucleus and acts as a transcription factor to regulate cell and phenotype-specific patterns of gene expression. SMAD6 and SMAD7 block phosphorylation of SMAD2 or SMAD3 and act as inhibitors of the TGF-β signaling pathway (15, 32). SMADs are widely expressed in most adult tissue and cell types, indicating that this TGF-β signaling mechanism is ubiquitous.
Materials and Methods

Study populations

As patients were recruited into the Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH) study, samples were collected into 3 stages, each totaling approximately 2,300 cases and 2,300 controls. Different controls were used for each stage. The geographic and ethnic background of both cases and controls is similar, with 99.7% being Caucasian. Given that only 0.31% of cases and controls are non-Caucasian (Europeans), the chance of false-positives or -negatives as a result of these differences is minimal.

Epithelial breast cancer cases were drawn from SEARCH (18), an ongoing population-based collection of breast cancer cases ascertained through the Eastern Cancer Registration and Information Centre (19). Ethical approval was obtained from the Anglia and Oxford Multicentre Research Committee, and informed consent was obtained from each patient.

All women diagnosed with invasive, epithelial breast cancer under the age of 55 years between January 1, 1991, and June 30, 1996, and who were alive at the start of the study (prevalent cases) as well as women under the age of 70 years who were diagnosed from 1996 onward (incident cases) were eligible for inclusion. Breast cancer cases were ascertained by both medical records and pathologic reports. Cases were randomly selected for Stage 1 from the first 3,500 recruited with sufficient quantity of DNA extracted from the blood collected. Stage 2 comprises the remainder of these plus the next 900 incident cases recruited, with Stage 3 comprising the next incident cases recruited; 64% eligible cases and 41% invited controls provided blood samples (Supplementary Table S1). The proportion of prevalent cases was higher in the first 2 stages; Stage 1 (33%), Stage 2 (20%), and Stage 3 (0.5%).

Median age at diagnosis was similar in Stage 1 and Stage 2 (51 and 52 years old, respectively) but higher in Stage 3 (56 years old for the incident cases and 68 years old for the prevalent cases) as the selection criteria for Stage 3 differed in age from previous stages. For Stages 1 and 2 combined, 27% were prevalent cases with a median age of 48 years and 73% were incident cases with a median age of 54 years. For Stages 1, 2, and 3 combined, the median age for incident cases was 54 years old and for the prevalent cases was 49 years old. There were no substantive differences in the morphology, histopathologic grade, or clinical stage of the cases by stage or by prevalent/incident status. Despite the different proportion of prevalent cases by stage, survival bias has not been observed for TGFB1 SNPs in the samples comprising Stage 1 of this study (20), and we have no evidence that allele frequencies of the SNPs studied here will be affected by the inclusion of prevalent and incident cases in SEARCH.

For Stages 1 and 2, volunteer female controls were selected in approximate order of recruitment from the Norfolk component of the European Prospective Investigation of Cancer (EPIC) (21). EPIC is a prospective study of diet and cancer being carried out in 9 European countries. The EPIC–Norfolk cohort comprises 25,000 individuals resident in Norfolk, East Anglia. Controls were aged between 45 and 74 years and were recruited between 1992 and 1994 from the same geographic region as cases. The female controls for Stage 3 were recruited from General Practitioners as part of SEARCH (Supplementary Table S1). Subjects diagnosed with breast cancer were excluded.

Hidden population structure was minimized by selecting cases and controls from the same region of the United Kingdom (3, 6–8, 22). Other sources of false-positive findings were limited by the use of the same DNA extraction method for all samples, uniform handling and storage of DNA, interspersed arraying of cases and controls on the same plate, and manual inspection of all genotype clusters.

SNPs were genotyped in a 3-staged SEARCH study design, each stage comprising 2,200 (Stages 1 and 2) or 2,303 (Stage 3) breast cancer cases and 2,280 controls (different for each stage). If an SNP showed marginally significant associations, it was taken through to the next stage (see selection criteria). Data from each stage were combined for analysis, as the power was greater than using each stage as an independent replication. A maximum of 6,703 cases and 6,840 controls were available for genotyping.

Additional genotyping data were obtained from the National Cancer Institute (NCI) Polish Breast Cancer Study (PBCS) to increase the power to confirm or negate any putative associations. This is a population-based study that included 1,966 incident cases diagnosed from 2000–2003 in Warsaw and Łódź, and 2,347 randomly selected controls obtained from population lists of all residents of Poland, and matched to the cases on age and city (23).

Selection of tagging single-nucleotide polymorphisms (tSNPs)

The aim of selecting tSNPs was to tag efficiently all of the common variations in a gene by genotyping a relatively small subset of the common variants to gain information on the entire region. Primary genotype data for each gene were obtained predominantly from the International HapMap project (24) on 30 Caucasian parent–offspring trios from Utah, United States (Centre d’Etude du Polymorphisme Humain data). SNP genotype data on 90 individuals representative of the U.S. population from the National Institute of Environmental Health Sciences (NIEHS) project (25) were used if the data from HapMap were not sufficient. Any unidentified SNPs are likely to be tagged by a tSNP selected for this region. The program Tagger, accessed via Haploview, was used to select tSNPs using the pairwise function (26), with the aim of selecting a minimal set of markers (tSNPs) such that all common alleles (defined as an MAF > 0.05) captured are correlated at an $r^2 > 0.8$ with a selected tSNP. If a tSNP failed to
manufacture or genotype, an alternative tSNP was selected, with 354 tSNPs genotyped in total.

Genotyping

The selected tSNPs were genotyped in Stage 1 using either the TaqMan ABI PRISM 7900 sequence detection system (Applied Biosystems; using 10 ng DNA) or the GenomeLab SNPstream Genotyping system (Beckman Coulter; using 8 ng genomic DNA) according to the manufacturer’s instructions. If a tSNP was selected for Stage 2, a TaqMan assay was designed and used for genotyping on Stage 2 and Stage 3. For accuracy and consistency, Stage 1 genotyping carried out by SNPstream was regenerated using a TaqMan assay. TaqMan genotyping resulted in a slightly higher call rate with no discordant replicate samples and it was preferable to combine data that had been generated by the same technology. Sequences for SNPs were retrieved and annotated using Seq4SNP software (27) or manually using databases available from NCBI (28) or ENSEMBL (29).

Genotyping accuracy for TaqMan and SNP stream

Each 384-well plate included 2 controls with no DNA. Genotyping 12 replicate samples from each plate on a separate 384-well plate ensured genotyping accuracy. For genotypes generated using the TaqMan platform, there was 100% concordance between the replicates (excluding failed replicate samples). For genotypes generated using the SNPstream platform, there was some discordance (maximum of 3.6% discordance for any SNP assay, excluding failed replicate samples). Failed genotypes were not repeated. The rate for failed genotypes did not exceed 3% for any of the SNPs under study by TaqMan. There is no evidence for greater deviation from Hardy–Weinberg equilibrium (HWE) than expected for the number of assays, indicating that genotyping quality was high. Of the 354 SNPs genotyped in Stage 1, 25 (7%) exhibited significant \( P < 0.05 \) deviation from HWE—see Supplementary Table S2.

Statistical methods

For each polymorphism, deviation of the genotype frequencies from those expected under HWE was assessed in the controls by a \( \chi^2 \) test. Genotype frequencies in cases and controls were compared using a \( \chi^2 \) test with 2 \( df \) \( (P_{\text{heterogeneity}}) \), and the Armitage trend test \( (\chi^2 \) with 1 \( df \)) for the trend in breast cancer risk with number of rare alleles \( (P_{\text{trend}}) \). The relative risks of breast cancer for heterozygotes and for rare homozygotes, relative to common homozygotes, were estimated as ORs with associated 95% CIs. Meta-analysis was carried out by STATA version 9.0 (Stata Corporation), weighting the studies by size.

Selection criteria

Selection criteria for genotyping in further stages were \( \chi^2 \) test \( P_{\text{trend}} < 0.1 \) or \( \chi^2 \) test \( P_{\text{heterogeneity}} < 0.1 \) revised to \( \chi^2 \) test \( P_{\text{trend}} < 0.05 \) for TGFBR1, TGFBR2, TGFBR1 (ALK5), ALK1, TGFBR2, and Endoglin SNPs. The selection criteria were revised as it became clear that potential associations at the less stringent cutoff were not replicated in further stages. The number of tSNPs selected for Stage 2 from genes studied was further reduced by analyzing correlations between the significant tSNPs, thus avoiding selecting several tSNPs that detected the same association (data not shown). A tSNP in TGFBR1 and rs1982073 was also included for genotyping in Stage 2, as it has previously been shown to be associated with breast cancer risk.

Analysis by receptor status

Relative risk estimates for specific tumor subtypes defined by estrogen receptor (ER) and progesterone

| SNP (Gene) | Genotype | Controls | OR | 95% CI | \( P \) | ER | OR | 95% CI | \( P \) | \( P_{\text{heterogeneity}} \)
|------------|----------|----------|----|--------|------|----|----|--------|------|-----------------
| rs10512263 | AA       | 7,782    | 3,994 | 1.00   | 1.282 | 1.00 | 1.00 | 0.72–1.01 | 0.060 |                
| TGFBR1     | AG       | 1,314    | 573  | 0.85   | 0.76–0.94 | 0.002 | 175 | 0.85 | 0.72–1.01 | 0.060 |
|           | GG       | 58       | 20   | 0.67   | 0.40–1.12 | 0.126 | 9   | 0.99 | 0.49–2.01 | 0.982 |
|           | Per allele | 9,154 | 4,587 | 0.85 | 0.77–0.93 | 0.001 | 1,466 | 0.83 | 0.71–0.97 | 0.021 |
|           |          |          |      |        |       |      |      |        | 0.851 |                
| rs4522809  | AA       | 2,650    | 1,369| 1.00   | 441   | 1.00 | 1.00 |                
| TGFBR2     | AG       | 4,486    | 2,275| 0.98   | 0.90–1.07 | 0.659 | 690 | 0.94 | 0.82–1.07 | 0.339 |
|           | GG       | 2,024    | 931  | 0.89   | 0.80–0.99 | 0.025 | 325 | 1.00 | 0.86–1.17 | 0.991 |
|           | Per allele | 9,160 | 4,575 | 0.95 | 0.90–1.00 | 0.033 | 1,456 | 0.98 | 0.90–1.06 | 0.572 |
|           |          |          |      |        |       |      |      |        | 0.450 |                
| rs1982073  | AA       | 3,382    | 1,625| 1.00   | 476   | 1.00 | 1.00 |                
| TGFBR1     | AG       | 4,256    | 2,087| 1.02   | 0.94–1.11 | 0.583 | 663 | 1.09 | 0.96–1.23 | 0.192 |
|           | GG       | 1,295    | 688  | 1.11   | 0.99–1.24 | 0.063 | 231 | 1.22 | 1.03–1.45 | 0.020 |
|           | Per allele | 8,933 | 4,400 | 1.04 | 0.99–1.10 | 0.103 | 1,370 | 1.12 | 1.03–1.22 | 0.006 |

\( a \) \( P \) value for heterogeneity of ORs by ER status of the tumors. Analyses are adjusted by study.
Table 2. Association between SNPs in the TGF-β pathway and breast cancer risk by PR status of the tumors

<table>
<thead>
<tr>
<th>SNP</th>
<th>Genotype</th>
<th>Controls</th>
<th>PR</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>PR</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>0.045 2 heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10512263 (TGFBR1)</td>
<td>AA</td>
<td>7,782</td>
<td>2,118</td>
<td>1.00</td>
<td></td>
<td>1.360</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>1,314</td>
<td>306</td>
<td>0.90</td>
<td>0.78–1.02</td>
<td>0.107</td>
<td>0.93</td>
<td>0.79–0.95</td>
<td>0.374</td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>58</td>
<td>9</td>
<td>0.60</td>
<td>0.30–1.22</td>
<td>0.162</td>
<td>0.11</td>
<td>0.64–2.38</td>
<td>0.537</td>
</tr>
<tr>
<td>Per allele</td>
<td>9,154</td>
<td>2,433</td>
<td>0.88</td>
<td>0.78–1.00</td>
<td>0.045</td>
<td>1.564</td>
<td>0.96</td>
<td>0.82–1.11</td>
<td>0.571</td>
</tr>
<tr>
<td>rs4522809 (TGFBR2)</td>
<td>AA</td>
<td>2,650</td>
<td>734</td>
<td>1.00</td>
<td></td>
<td>461</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>4,486</td>
<td>1,237</td>
<td>0.99</td>
<td>0.89–1.09</td>
<td>0.788</td>
<td>0.87–1.14</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>2,024</td>
<td>488</td>
<td>0.90</td>
<td>0.79–1.02</td>
<td>0.094</td>
<td>0.89–1.21</td>
<td>0.652</td>
<td></td>
</tr>
<tr>
<td>Per allele</td>
<td>9,160</td>
<td>2,427</td>
<td>0.95</td>
<td>0.89–1.01</td>
<td>0.114</td>
<td>1.557</td>
<td>0.94–1.10</td>
<td>0.677</td>
<td>0.286</td>
</tr>
<tr>
<td>rs1982073 (TGFBI)</td>
<td>AA</td>
<td>3,382</td>
<td>868</td>
<td>1.00</td>
<td></td>
<td>476</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>4,256</td>
<td>1,072</td>
<td>0.97</td>
<td>0.87–1.07</td>
<td>0.510</td>
<td>0.89–1.22</td>
<td>0.092</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>1,295</td>
<td>353</td>
<td>1.04</td>
<td>0.90–1.19</td>
<td>0.613</td>
<td>1.12</td>
<td>0.98–1.27</td>
<td>0.092</td>
</tr>
<tr>
<td>Per allele</td>
<td>8,963</td>
<td>2,293</td>
<td>1.01</td>
<td>0.94–1.08</td>
<td>0.853</td>
<td>1.445</td>
<td>1.18</td>
<td>1.09–1.28</td>
<td>4.1E-05</td>
</tr>
</tbody>
</table>

*P* value for heterogeneity of ORs by PR status of the tumors. Analyses are adjusted by study.

Results and Discussion

Assays for a total of 354 SNPs were successfully generated, together tagging 92% of the 1,254 common variants with $r^2 > 0.8$ identified in the 17 genes (Supplementary Table S3). These assays were used to genotype the 2,200 cases and 2,280 controls from the SEARCH study in Stage 1. Twenty-one of the 354 SNPs were genotyped in incident cases and 2,280 controls from the SEARCH study and 6 of these subsequently progressed to Stage 3 (a further 2,303 cases and 2,280 controls from the SEARCH study). Of these, 5 SNPs were genotyped in incident cases and controls from PBCS. Data are shown for all 3 stages in Supplementary Table S2 and weighted meta-analyses of all 3 stages and the PBCS data are shown in Figure 2.

In addition to the previously reported association with TGFBI Leu10Pro (rs1982073), 2 SNPs showed marginally significant associations with risk of invasive breast cancer (meta-analysis of the SEARCH and PBCS studies). In TGFBR1, the minor G allele of SNP rs10512263 showed a protective effect [OR (G vs. A) 0.87, 95% CI: 0.81–0.95, $P = 0.001$]. In TGFBR2, the G allele of rs4522809 also had a marginally protective effect [OR (G vs. A) 0.95, 95% CI: 0.91–0.99, $P = 0.02$]. The magnitudes of these effects are not markedly altered by removal of the hypothesis generating data [SEARCH Stage 1 (OR 0.91, 95% CI: 0.83–1.00, $P = 0.050$; OR 0.96, 95% CI: 0.91–1.01, $P = 0.094$)] for TGFBR1 and TGFBR2, respectively.

For rs1982073, a meta-analysis of all previously published data from the BCAC consortium (8) together with new data generated by this study is presented in Figure 2c. There is a dose-dependent association of the proline-encoding allele with increased risk of invasive breast cancer [OR (Pro vs. Leu) 1.05, 95% CI: 1.02–1.09, $P = 0.002$].

It has been suggested that this Pro allele association may be confined to an increased risk of developing PR tumours (8). The SEARCH and PBCS data presented here, including SEARCH Stage 3 and the entire cohort of PBCS in addition to the data previously published; Table 1 support that suggestion [PR OR (Pro vs. Leu) 1.18, 95% CI: 1.09–1.28, $P = 4.1 \times 10^{-5}$; PR OR (Pro vs. Leu) 1.01, 95% CI: 0.94–1.08, $P = 0.85$]. In this context it is of interest that TGFBI inhibits the PR expression in human endometrial stromal cells in vitro via SMAD signaling (30). If, as is likely, the same effect occurs in breast epithelial cells in vivo, the observation may account, at least in part, for the association of the Pro allele (rs1982073) in TGFBI with PR tumours, as the Pro allele causes substantial increases in TGFBI secretion in vitro compared with the Leu allele.

In light of this finding, tumor subtype associations of the TGFBR1 and TGFBR2 SNPs have also been investigated (Tables 1 and 2). For rs10512263, the protective effect is more pronounced in PR [OR (G vs. A) 0.88, 95% CI: 0.78–1.00, $P = 0.05$] than in PR tumours [OR (G vs. A) 0.96, 95% CI: 0.82–1.11, $P = 0.57$]. For rs4522809, the marginal protective association of the G allele is confined to ER tumours [OR (G vs. A) 0.95, 95% CI: 0.90–1.00, $P = 0.03$]. Confirmation of these early indications of possible tumor subtype differences would require much larger studies.

The effects of these SNPs on the strength of TGFBR1 signaling mediated by TGFBR1 (ALK5) and TGFBR2 on...
Figure 2. Relative risks of breast cancer estimated as OR and 95% CI by study/stage and meta-analysis for overall OR and 95% CI. A, TGFBR1 rs10512263 [G vs. A] and B, TGFBR2 rs4522809 [G vs. A] include data from SEARCH Stages 1, 2, and 3 and PBCS. C, TGFB1 rs1982073 [G vs. A] Pro vs. Leu) includes data from SEARCH Stages 1, 2, and 3 and BCAC (including PBCS). Studies have been weighted by size.
PR and ER expression in breast epithelial cells may provide a biological consistency test for the SNP associations with tumor subtype. It would be predicted, for example, that the protective G allele for the TGFBR1 SNP would be associated with reduced TGF-β signaling and higher PR expression. However, the associations for TGFBR1 (ALK5) and TGFBR2 need to be interpreted with caution, as they would not remain significant after Bonferroni adjustment for multiple testing (354 tSNPs, \( P \approx 1.4 \times 10^{-4} \)). Potentially contradicting our findings, TGFBR2 expression has recently been associated with prognostically favorable small tumors among ER− tumors (16).

If these SNP associations are true, the questions of how they exert their effects remain. SNP rs1982073 may be a causal variant. It is exonic and affects the signal peptide, responsible for secretion of TGF-β from the cell. In previous studies, we have shown that rs1982073 causes a 2.8-fold increase in the amount of protein secreted in vitro (9). SNP rs10512263 is located in intron 1 of TGFBR1, in a region that is conserved between species. It exhibits no strong LD with any other known variants. It is possible that this SNP is a causal variant with an unknown function or it may be marking the causal SNP. The functionally relevant TGFBR1*6A SNP that has previously been associated with breast cancer risk (10) has not been genotyped as part of the HapMap project or within our own cohorts, and it is currently beyond the means of this study to determine the LD between this SNP and SNP rs10512263. SNP rs4522809 is situated in intron 2 of the TGFBR2 gene and is most likely marking a putative causal variant. Identifying disease-associated functionally-relevant alleles in candidate genes can be more readily interpreted (31).

No tSNPs in the genes of the endothelial-specific ALK1/SMADs 1&5 pathway, which promotes angiogenesis (14, 15), have been identified as associated with breast cancer risk. The balance of signaling through the 2 pathways may regulate TGF-β angiogenic activity, but there is no detectable effect of SNPs on genes that regulate the proangiogenic pathway specifically. Any proangiogenic effects of SNPs are therefore likely to be restricted to effects on genes in the TGFBR1 (ALK5) pathway that weaken the antiproliferative effect of TGF-β and indirectly rebalance signaling toward the proangiogenic pathway.

From this tag SNP study, it has been possible to exclude large associations with invasive breast cancer of the common variants in the LTBP1, LTBP2, LTBP4, TGFBR2, TGFBR3, ALK1, Endoglin, SMAD1, SMAD2, SMAD3, SMAD4, SMAD5, SMAD6, and SMAD7 genes. It remains possible that associations of moderate effect size could have been missed (as false-negatives). This is illustrated by the example of rs1982073, which would not have been selected for Stages 2 and 3 without prior data. However, in the case of TGFBR1, the association of a correlated (\( r^2 = 0.61 \) in Stage 1) SNP rs4803455 was identified (see Supplementary Table S2), showing the effectiveness of the study to capture SNP associations.

The study provides some evidence that SNPs in the TGFBR1, TGFBR1 (ALK5), and TGFBR2 genes may play a role in breast cancer development, although further subgroup analyses are needed to confirm these associations and any associations with hormone receptor type-specific disease. However, this study and numerous breast cancer GWAS indicate that it is highly unlikely that other common variants in the TGF-β signaling pathway contribute significantly to breast cancer risk.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgment

Don Conroy for providing technical support, Helen Field for providing bioinformatic support, and the members of the SEARCH study team, particularly Melanie Maranian.

Grant Support

Financial support provided by Breast Cancer Campaign grant number 2003/474, Cancer, Research-UK grant numbers C8197/A10123, C490/A11021, C1287/A10118, C8197/A10865, and EU FP7 Health-F2-2009-223175-COGS.

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.

Received January 19, 2011; revised April 2, 2011; accepted April 12, 2011; published OnlineFirst May 17, 2011.

References
