Polymorphisms in the Janus kinase 2 (JAK)/signal transducer and activator of transcription (STAT) genes: putative association of the STAT gene region with familial breast cancer

Annika Vaclavicek¹, Justo Lorenzo Bermejo¹, Rita K Schmutzler², Christian Sutter³, Barbara Wappenschmidt², Alfons Meindl⁴, Marion Kiechle⁴, Norbert Arnold⁵, Bernhard H F Weber⁶, Dieter Niederacher⁷, Barbara Burwinkel¹,8, Claus R Bartram³, Kari Hemminki¹,9 and Asta Försti¹,9

¹Division of Molecular Genetic Epidemiology C050, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany
²Division of Molecular Gynaeco-Oncology, Department of Gynaecology and Obstetrics, Center of Molecular Medicine Cologne (CMMC), University Hospital of Cologne, Cologne, Germany
³Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany
⁴Department of Gynaecology and Obstetrics, Klinikum rechts der Isar at the Technical University, Munich, Germany
⁵Division of Oncology, Department of Gynaecology and Obstetrics, University Hospital Schleswig-Holstein, Kiel, Germany
⁶Institute of Human Genetics, University of Regensburg, Regensburg, Germany
⁷Division of Molecular Genetics, Department of Gynaecology and Obstetrics, Clinical Center University of Düsseldorf, Düsseldorf, Germany
⁸Helmholtz-University Group Molecular Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
⁹Center for Family and Community Medicine, Karolinska Institute, Huddinge, Sweden

(Requests for offprints should be addressed to A Vaclavicek; Email: a.vaclavicek@dkfz.de)

Abstract

The Janus kinase (JAK)/signal transducer and activator of transcription (STAT) pathway mediates the signals of a wide range of cytokines, growth factors and hormones. Thus, aberrant activation of the JAK/STAT pathway may predispose to malignancy due to deregulation of proliferation, differentiation or apoptosis. In this study, we investigated whether genetic variation in the JAK2 gene and the STAT gene region (STAT3, STAT5A and STAT5B) is associated with breast cancer (BC) risk. We carried out a case-control study using a German sample set with 441 familial, unrelated BC cases and 552 controls matched by age, ethnicity and geographical region. Haplotypes in the JAK2 gene were not associated with the risk of BC. In the STAT gene region, the rare haplotype CAGCC containing the variant alleles of each single nucleotide polymorphism (SNP) was associated with an increased risk odds ratio (OR = 5.83, 95% confidence interval (CI) 1.51–26.28). According to Akaike’s information criterion, the best model to describe the relationship between the haplotypes and BC was based on the SNPs rs6503691 (STAT5B) and rs7211777 (STAT3). Carriers of the AC haplotype, which represents the variant alleles of both SNPs, were at an increased risk (OR = 1.41, 95% CI 1.09–1.82). A decreased risk was observed for carriers of the AT haplotype (OR = 0.60, 95% CI 0.38–0.94). Furthermore, individuals with the AC/GC diplotype were at a significantly increased risk (OR = 1.88, 95% CI 1.13–3.14). The observed genetic variation may also influence the inter-individual variation in response to STAT-signalling targeted therapy.
A Vaclavicek et al.: Polymorphisms in the JAK2/STAT genes

Introduction

The Janus kinase (JAK)/signal transducer and activator of transcription (STAT) pathway transmits a wide range of cytoplasmic signals from cytokines, growth factors and hormones (e.g. prolactin) that bind to specific cell surface receptors (Levy & Darnell 2002, Calo et al. 2003, Clevenger et al. 2003, Verma et al. 2003). STAT activation requires tyrosine phosphorylation by a receptor-associated JAK2 kinase which is induced after cell surface receptor activation by a ligand (Watson 2001, Levy & Darnell 2002, Calo et al. 2003). As a consequence, the STAT proteins are phosphorylated on a single tyrosine residue, ensued by dimerisation/multimerisation and nuclear translocation. This results in the interaction of the STAT complex with its recognition site in the promoters of target genes. Moreover, unphosphorylated STAT3 has been assumed to regulate gene expression (Yang et al. 2005).

The STAT proteins comprise a family of several members that remain latent in the cytoplasm (Watson 2001, Levy & Darnell 2002, Calo et al. 2003, Yu & Jove 2004). They participate in a series of normal cellular processes such as differentiation, proliferation, cell growth, survival and apoptosis. A fundamental role especially for STAT3 and STAT5 in the normal development of the mammary gland and the pathogenesis of human breast cancer (BC) has been established (Watson 2001, Clevenger 2004). Although both STAT3 and STAT5 are activated during the development of the mammary gland, they undergo reciprocal activation patterns (Philip et al. 1996, Watson 2001, Calo et al. 2003). Whereas the prolactin-activated transcription factor STAT5 is essential for the development and differentiation of the mammary gland as well as for lactogenesis, STAT3 is involved in the involution of the post-lactating mammary gland which constitutes an apoptotic process.

A persistent activation of STAT signalling, often due to overactive tyrosine kinases, has been observed in a wide number of human cell lines and primary tumours including BC (Bromberg & Darnell 2000, Garcia et al. 2001, Watson 2001, Coletta et al. 2004, Yu & Jove 2004). Additionally, proteins whose expression is driven by overexpression of unphosphorylated STAT3 have been found in many cancers (Yang et al. 2005). Several studies have demonstrated increased levels of STAT3 in primary mammary tumours. Immunohistochemical approaches in humans have found increased levels of nuclear localised STAT3 in malignant BCs when compared with normal tissues (Clevenger 2004). Nuclear expression of STAT5A in chemically induced rat mammary gland cancers and preneoplastic lesions has been detected, when compared with cytosolic STAT5A expression in the control tissue (Shan et al. 2004). In the same study, a significant correlation between a high STAT5A nuclear expression and high tumour grade was observed. A significantly increased nuclear localisation of STAT5A has also been shown in human BCs (Cotarla et al. 2004, Shan et al. 2004).

Family history is a well-established and important risk factor for BC. However, the known BC susceptibility genes, such as BRCA1 and BRCA2, explain only a minority of familial aggregation of the disease (Antoniou & Easton 2006). A polygenic model, with low-penetrance variants in many genes contributing jointly to the BC risk, has been suggested to explain best the excess familial risk (Antoniou & Easton 2006).

In this study, we wanted to investigate whether genetic variation in the JAK2 and the STAT3, STAT5A and STAT5B gene regions is associated with the risk of BC. A total of ten single nucleotide polymorphisms (SNPs) were chosen for genotype analyses in a case-control study, where the cases had a family history of BC, but did not carry BRCA1 or BRCA2 mutations.

Materials and methods

Study population

A case-control study was carried out using a German study population consisting of 441 familial, unrelated cases of female BC (mean age 43.8 years, range 19–82). The cases were collected by the German Consortium for Hereditary Breast and Ovarian Cancer at the Institute of Human Genetics (Heidelberg, Germany), the Department of Gynaecology and Obstetrics, Center of Molecular Medicine (Cologne, Germany) and the Department of Medical Genetics (Munich, Germany). A detailed description of the case population has been published previously (Meindl 2002). The cases belonged to: (F1) families with two or more BC cases including at least two cases with onset before the age of 50 (131 cases); (F2) families with at least one BC and one ovarian cancer (68 cases); (F3) families with at least two BC cases not included in F1 or F2 (201 cases) and (F4) families with a single case of BC diagnosed before the age of 35 years (16 cases). In addition, three cases had a family history of both female and male BCs, 13 cases had bilateral BC diagnosed below the age of 50; information on family history was missing for nine cases (Antoniou & Easton 2006).
cases. The biological samples consisted of genomic DNA extracted from blood.

A second, independent sample set was used to validate the genotyping results of the SNPs rs6503691 (STAT5B) and rs7211777 (STAT3). The second sample set was provided by four Centers of the German Consortium for Hereditary Breast and Ovarian Cancer (Cologne, Kiel, Würzburg, Düsseldorf, see author affiliations; Meindl 2002). The second sample set consisted of 381 familial, unrelated BC cases of German women (mean age 46.1 years, range 23–80). The distribution of the cases according to the family histories described above was F1 = 196 cases, F2 = 67 cases, F3 = 101 cases and F4 = 8 cases. In addition, six cases had a family history of both female and male BCS; information on family history was missing from three cases. The samples of the second set consisted of whole genome amplified DNA. The whole genome amplification (WGA) of original genomic DNA was performed with the GenomiPhi DNA amplification kit (Amersham Biosciences) according to Wong et al. (2004) and the Φ29 DNA polymerase as described by Paez et al. (2004) respectively. The amplification results were controlled by genotyping two frequent SNPs using TaqMan allelic discrimination assays. The SNPs were genotyped previously using the original genomic DNA. The genotyping results of the WGA samples were concordant with the corresponding results of the genotyped genomic DNA. About 0.57% of the WGA genotypes could not be determined or did not agree with the data of the genomic DNA.

The cases of both sample sets were collected during the years 1997–2005. Family histories covered three generations with the index case having at least one first degree relative diagnosed with BC or ovarian cancer. In all samples, the entire coding regions of the BRCA1 and BRCA2 genes were screened and cases carrying deleterious BRCA1/2 mutations were excluded (Meindl 2002). The present study focused on familial cases, because it has been shown before that the use of familial cases significantly increases the power to detect rare alleles contributing to the risk for BC (Antoniou & Easton 2003, Houlston & Peto 2003). All BC patients gave informed consent to the use of their samples.

The age-matched control population included healthy and unrelated female blood donors (first sample set: 552 controls, mean age 50.7 years, range 26–68, genomic DNA from blood; second sample set: 460 controls, mean age 43.6 years, range 18–68, whole genome amplified DNA) collected by the Institute of Transfusion Medicine and Immunology (Mannheim, Germany). All controls were randomly selected during the years 2004 and 2005 and shared the ethnic and geographical background of the BC patients. According to the German guidelines for blood donation, all blood donors were examined by a standard questionnaire and consented to the use of their samples for research purposes. The study was approved by the Ethics committee of the University of Heidelberg (Heidelberg, Germany).

**SNP selection**

In the present study, we screened the promoter regions of the JAK2 (chromosome 9; Fig. 1) and the STAT3 (chromosome 17; Fig. 2a) genes for polymorphisms as well as the coding SNPs published in the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov). In order to cover the entire regions of these genes, the SNPBrowser software (version 3.5, Applied Biosystems, Foster City, CA, USA) was used to select tagging SNPs within the gene regions. SNPs with a minor allele frequency in Caucasians under 10% were excluded, as well as validated SNPs whose genotypes matched more than 80% with the selected tagging SNPs. To determine the linkage disequilibrium (LD) patterns

---

**Figure 1** Schematic structure of the JAK2 gene. The grey boxes correspond to the coding exons 3–25. The regions which were screened by sequencing are underlined. Verified SNPs are marked with an arrow and the SNPs selected for further analyses are typed in bold. Asterisks mark tagging SNPs selected with the SNPBrowser software by Applied Biosystems.

---

www.endocrinology-journals.org
and the haplotype blocks in the JAK2 and STAT3 genes and their surrounding regions, HapMap population data of Utah residents with northern and western European ancestry were used (www.hapmap.org). The LD calculation was performed by the Haploview software (Barrett et al. 2005). The haplotype blocks were defined by the methods of Gabriel et al. (2002). In our attempt to cover the whole STAT gene region, four more tagging SNPs on chromosome 17 in the STAT5A and STAT5B genes and its surrounding region were chosen (Fig. 2b). The selected SNPs span a region of 215.314 bp for the STAT gene region and 125.158 bp for the JAK2 gene region. According to the available HapMap data, the SNPs in the JAK2 gene region are relatively highly linked (Fig. 3). The SNPs in the STAT gene region form two haplotype blocks separated by the SNP rs2293152 (Fig. 4). Based on our sequencing results and the LD derived from the HapMap data, a total of five SNPs were selected for genotyping in the JAK2 gene and a total of five SNPs in the STAT gene region. They were chosen because they seemed not to be in high LD and therefore would give information about the whole gene region in the haplotype analysis.

**PCR amplification**

The PCR was performed with 5 ng DNA in a total reaction volume of 10 μl as described earlier by

---

**Figure 2** (a) Schematic structure of the STAT3 gene. The grey boxes correspond to the coding exons 2–24. The regions which were screened by sequencing are underlined. Verified SNPs are marked with an arrow. The SNPs selected for genotyping and further analyses are typed in bold. Asterisks mark tagging SNPs selected with the SNPbrowser software by Applied Biosystems. (b) Schematic illustration of the STAT gene region. The SNPs selected for genotyping are marked with an arrow. Asterisks mark tagging SNPs selected with the SNPbrowser software by Applied Biosystems.

**Figure 3** The LD block structure observed in the JAK2 gene using genotype data from Utah residents with northern and western European ancestry available at the HapMap database. Each square box indicates the pairwise magnitudes of LD. The darker the colour, the higher the LD: black-coloured squares indicate a very strong LD. The genotyped SNPs in the JAK2 gene are marked with an arrow in the following order: rs1887429, rs2230722, rs7034539, rs1410779, rs11793659. Data of the SNP rs2230722 was not available in HapMap. It is located between rs1328917 and rs2149555 as marked in the LD map.
Vaclavicek et al. (2006). Primer sequences and annealing temperatures are available from the corresponding author on request.

SNP screening by sequencing

A set of 23 samples from BC cases was investigated by sequencing to confirm the presence, frequency and LD data of the polymorphisms and to obtain standards for genotyping with TaqMan (Vaclavicek et al. 2006). The PCR product was purified with ExoSapIT (USB Amersham) at 37 °C for 40 min followed by 85 °C for 15 min. The Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) was used to carry out the sequencing reaction. The ABI PRISM 3100 Genetic analyser (Applied Biosystems) was used and the obtained sequences were aligned with DNAStar SeqManII software (DNASTAR Inc., Madison, WI, USA).

Genotyping using TaqMan assays

The SNPs were investigated using the allelic discrimination method. TaqMan assays as assay on demand were ordered from Applied Biosystems: C_12095829 (rs1887429), C_22273134 (rs7034539), C_7590450 (rs1410779), C_1417102 (rs11793659), C_29619946 (rs9912576), C_30619370 (rs6503691), C_15961941 (rs2272087), C_3140302 (rs2293152), C_1952182 (rs7211777). Primer and probe sequences for rs2230722 (assay by design, Applied Biosystems) are available from the corresponding author on request.

The reaction was performed in 5 µl using 225 nM each primer, 50 nM each probe and 2.5 µl Taq-Man Universal 2× PCR Master Mix (Applied Biosystems). PCR conditions were as described previously (Vaclavicek et al. 2006). The samples were read and analysed in an ABI Prism 7900HT sequence detection system using SDS 1.2 software (Applied Biosystems). About 10% of the samples were sequenced to confirm the results. All the sequencing results were concordant with the original results.

Statistical analysis of single SNPs

Genotype distributions in BC cases and controls were tested for the Hardy–Weinberg equilibrium (HWE) by χ² tests. The differences in genotype frequencies between the cases and the controls were assessed for statistical significance by χ² tests; when the expected number of cases was smaller than five, Fisher’s exact test was used. Odds ratios (ORs) with 95% confidence intervals (95% CIs) were calculated to evaluate the association between the genotypes and BC. A joint analysis was performed using the Mantel–Haenszel adjustment. Statistical analyses were carried out using the Hardy–Weinberg equilibrium test tool offered by the Institute of Human Genetics, Technische Universität Munich (http://ihg.gsfe.de/cgi-bin/hw/hwa1.pl) and the Epi Info 2000 software (http://www.cdc.gov/epiinfo). Power calculations were carried out in the PS software for power and sample size calculation (http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/Power

Figure 4 The LD block structure observed in the STAT gene region using genotype data from Utah residents with northern and western European ancestry available at the HapMap database. Each square box indicates the pairwise magnitudes of LD. The darker the colour the higher the LD: black-coloured squares indicate a very strong LD. The genotyped SNPs are marked with an arrow in the following order: rs6503691 (STAT5B), rs2272087 (STAT5A), rs2293152 (STAT3), rs7211777 (STAT3). For the SNP rs9912576 in the KCNH4 gene which is located 75356 bp downstream from rs6503691, no HapMap data was available.
SampleSize: Dupont & Plummer 1998). With our sample size, we had a power of 80% to detect an OR of 1.5 at a significance level of 0.05.

Statistical analysis of haplotypes

The LD was calculated with Haploview (Barrett et al. 2005). Haplotypes based on any combination of 2–5 SNPs were inferred using the Haplotype procedure of SAS/Genetics Software. After haplotype inference, the relationship between the haplotypes and susceptibility to BC was assessed by logistic regression. Since the models based on different combinations of SNPs are generally not nested, the optimal combination of SNPs to describe the relationship between the genotypes and BC was selected using Akaike’s information criterion (AIC; Akaike 1973). It tries to maximise the likelihood of the observed data under the simplest model. In our calculations, the lower the AIC value, the better the model. After the model selection, ORs with 95% CIs were estimated for haplotypes and diplotypes.

Results

The present case-control study was performed to clarify whether genetic variation in the JAK2, STAT3, STAT5A and STAT5B genes region is associated with the risk of BC. First, the JAK2 promoter and all exons with coding SNPs published by NCBI and their surrounding regions were sequenced in a small set of 23 BC samples (Fig. 1). Furthermore, four tagging SNPs in the JAK2 gene region and in the 5′ promoter region, selected with the SNPbrowser software, were sequenced. In addition to the tagging SNPs, we could verify only two promoter SNPs and the coding SNPs rs2230722 in exon 6 and rs2230724 in exon 19. The two confirmed exonic SNPs are synonymous and do not change amino acid.

In the STAT3 gene, we first screened the promoter region from −40 868 to −40 345 bp (numbering relative to ATG in exon 2) by sequencing (Fig. 2a). No SNPs were detected. We also screened SNPs published by NCBI in the exons and their surrounding regions. Except the intronic SNP rs2306581 near to exon 2, we did not verify any SNPs. Therefore, six tagging SNPs were selected with the SNPbrowser software in the STAT3, STAT5A, STAT5B and the adjacent KCNH4 gene (Fig. 2b). All of them were confirmed in our samples.

Based on the LD data derived from the HapMap data (Figs 3 and 4) and our sequencing results, a total of ten SNPs (JAK2 gene: rs1887429, rs2230722, rs7034539, rs1410779, rs11793659; STAT gene region: KCNH4 gene: rs9912576, STAT5B gene: rs6503691, STAT5A gene rs2272087, STAT3 gene: rs2293152, rs7211777) were selected for genotyping in the German sample set. The genotype distributions of all the chosen polymorphisms among the controls were consistent with HWE. The observed allele frequencies of the SNPs agreed with the published data in the Caucasians (NCBI, HapMap, Applied Biosystems).

JAK2 haplotypes are not associated with the risk of BC

The promoter SNP rs1887429 showed a protective effect for the homozygous variant TT genotype with an OR of 0.50 (95% CI 0.26–0.96; Table 1). The rare homozygous TT genotype of the SNP rs2230722 was more frequent in the cases than in the controls with an OR of 1.59 (95% CI 1.01–2.49). For the SNP rs7034539 near exon 19, a significant protective effect for carriers of the variant GG genotype was observed (OR = 0.45, 95% CI 0.20–0.99).

After the inference of haplotypes based on all five SNPs, the haplotype distribution of the cases and the controls were compared using the wild-type haplotype as reference (Table 1). As expected from the high LD within the JAK2 gene region, the four commonest haplotypes accounted for ~86% of the total number of haplotypes. Data on haplotypes with a frequency under 5% are not shown. We could not identify any significant association between the haplotypes based on the five SNPs and the risk of BC. All other SNP combinations were investigated using the AIC algorithm, but no haplotype was significantly associated with the risk of BC (data not shown).

STAT gene region shows significant association with BC risk

The SNPs in the STAT gene region were not individually associated with familial BC (Table 2). When we divided the cases into the subgroups F1 to F4 based on the family history of BC, the genotype distributions showed no significant differences. While the SNPs in the STAT3 and STAT5A genes were tightly linked, the SNPs in the STAT5B and the adjacent KCNH4 gene showed a lower LD (Fig. 5). The analysis of the inferred haplotypes revealed that the rare CAGCC haplotype (frequency ≤1%), which contained the variant alleles of each SNP, was associated with an increased BC risk (OR = 5.83, 95% CI 1.51–26.28, P = 0.002; Table 2).

The AIC values indicated that the best model to describe the relationship between the haplotypes and BC was based on the SNPs rs6503691 (STAT5B) and rs7211777 (STAT3; data not shown). We observed an
increased risk for the AC haplotype, which contains the variant alleles of the two SNPs (OR = 1.72, 95% CI 1.19–2.48, \( P = 0.002 \)), and a decreased risk for the AT haplotype (OR = 0.51, 95% CI 0.25–1.01, \( P = 0.04 \)).

The findings for these two SNPs were validated using a second, independent sample set. The genotype distribution was similar to the one in the first sample set and no association between the genotypes and the risk of BC was observed (data not shown). Nor did the haplotypes associate with the risk of BC (data not shown). However, the joint analysis of the two sample sets still detected a haplotype effect (global \( P = 0.004 \)). The haplotype analysis of the two SNPs, based on the first and second sets of samples, revealed that the AC haplotype, with a frequency of 7% in the control population, was associated with an increased risk (OR = 1.41, 95% CI 1.09–1.82, \( P = 0.007 \); Table 3). Moreover, a significantly decreased risk was observed for the AT haplotype (OR = 0.60, 95% CI 0.38–0.94, \( P = 0.02 \)).

Consideration of the two haplotypes for each individual (diplotypes) showed that the AC/GC diplotype carriers were at a significantly increased risk (OR = 1.88, 95% CI 1.13–3.14, \( P = 0.01 \); Table 3). The AC/AC diplotype was also more frequent in the cases than in the controls (OR = 3.03, 95% CI 0.78–11.80, \( P = 0.09 \)), but it was very rare (frequency \( \% \) 1%). The AT/GT diplotype which consists of the protective AT and the wild-type GT haplotype respectively showed a protective effect of borderline significance (OR = 0.57, 95% CI 0.31–1.05, \( P = 0.05 \)).

The SNPs rs6503691 and rs7211777 showed a LD of \( D^j = 0.49 \). The haplotype distribution in our study was similar to the available population data of Caucasians at NCBI and HapMap. In addition, there were no differences in the haplotype distributions between the two sample sets, nor between the samples collected by the different Centers.

**Discussion**

The present case-control study used a German study population to investigate the influence of genetic
variation in the JAK2 and STAT3, STAT5A and STAT5B genes on familial BC risk. However, the effect of the studied SNPs would be expected to be the same also in women without a family history.

Our results showed an association between the STAT gene region and susceptibility to BC. According to AIC, the best model to explain this association was based on the SNPs rs6503691 (STAT5B) and rs7211777 (STAT3). Carriers of the AC haplotype containing the variant alleles of these SNPs were at an increased risk (Table 3). The significantly increased risk was also observed for the AC/GC diplotype. The homozygous AC/AC diplotype increased the OR even more. However, due to the rarity of the diplotype, the association was not statistically significant.

The diplotype AT/GT, which consists of the protective AT and the wild-type GT haplotypes respectively, showed a protective effect of borderline significance. The three SNPs in the JAK2 gene that were individually marginally associated with BC are unlikely to be true risk factors because of lacking haplotype effects (Table 2).

The SNPs rs6503691 and rs7211777 within the STAT gene region span a region of 140 kb, and according to the available HapMap data, they surround two haplotype blocks (Fig. 4). We observed a slightly higher LD among the SNPs within this region than the HapMap suggested ($D^\prime = 0.49$ and

Table 2 Genotype, allele and haplotype distribution of the single nucleotide polymorphisms in the signal transducer and activator of transcription gene regions in the German breast cancer patients and the controls (first sample set). Variant alleles within haplotypes are typed in bold

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Cases</th>
<th>Controls</th>
<th>OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGA/C</td>
<td>245 (0.31)</td>
<td>357 (0.35)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>TGAGT</td>
<td>207 (0.26)</td>
<td>257 (0.25)</td>
<td>1.17 (0.91–1.51)</td>
<td>0.2</td>
</tr>
<tr>
<td>TGAGC</td>
<td>115 (0.14)</td>
<td>142 (0.14)</td>
<td>1.18 (0.87–1.60)</td>
<td>0.27</td>
</tr>
<tr>
<td>TGGGC</td>
<td>72 (0.09)</td>
<td>99 (0.1)</td>
<td>1.06 (0.74–1.52)</td>
<td>0.74</td>
</tr>
<tr>
<td>CAGCC</td>
<td>12 (0.01)</td>
<td>3 (0.003)</td>
<td>5.83 (1.51–26.28)</td>
<td>0.002</td>
</tr>
</tbody>
</table>

The bases are listed in the order rs1887429, rs2230722, rs7034539, rs1410779, rs11793659.
The differences may be explained by the different origin of the samples and the sample sizes. The LD data in the HapMap is based on the genotypes of 30 trios from Utah residents with northern and western European ancestry. Our LD data of the five SNPs analysed in the \textit{STAT} gene region are based on 993 German individuals (Fig. 5) and the LD data between the SNPs rs6503691 and rs7211777 on 1834 German individuals. There were no population differences in the genotype or haplotype distribution between the two sample sets used in our study, nor between the sample sets collected by the different Centers of the German Consortium for Hereditary Breast and Ovarian Cancer. However, the significant haplotype effect was based mainly on the results from the first sample set.

Chromosome 17 is well known to be one of the most frequent targets of genetic damage in human neoplasia (Osborne & Hamshere 2000). Loss of heterozygosity, as well as amplification, has been often reported in BC (Osborne & Hamshere 2000, Maguire \textit{et al.} 2005). The erythroblastic leukaemia viral oncogene homologue 2 (\textit{ERBB2}) gene, which encodes a transmembrane tyrosine kinase receptor, is located on chromosome 17 (Revillion \textit{et al.} 1998), about 2.5 Mb downstream of the \textit{STAT5B} gene. \textit{ERBB2} is expressed in a variety of normal tissues and has been reported to be over-expressed and amplified in 20–30\% of breast, ovarian and renal cell carcinomas (Revillion \textit{et al.} 1998). Also the \textit{BRCA1} gene is located on chromosome 17, about 650 Kb upstream of the \textit{STAT3} gene. Allelic loss is a key mechanism of inactivation of the wild-type allele and is known to occur in breast tumours of \textit{BRCA1} mutation carriers (Miki \textit{et al.} 1994). Studies to identify new predisposing genes on chromosome 17q in breast tumours have been carried out without success.

![Figure 5](image-url) The LD block structure observed in the \textit{STAT} gene cluster region using the genotype data from the German control samples. Each square box indicates the pairwise magnitudes of LD. The darker the colour, the higher the LD: black coloured squares indicate a very strong LD.

![Table 3](table-url) Distributions of the haplotypes and diplotypes containing the two single nucleotide polymorphisms rs6503691 signal transducer and activator of transcription B (\textit{STAT5B}) and rs7211777 (\textit{STAT3}; joint data of both sample sets). Variant alleles within haplotypes and diplotypes are typed in bold.

<table>
<thead>
<tr>
<th>Haplotype(^a)</th>
<th>Cases</th>
<th>Controls</th>
<th>OR (95% CI)</th>
<th>(P) value</th>
<th>Gobal (P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT</td>
<td>942 (0.61)</td>
<td>1200 (0.62)</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GC</td>
<td>427 (0.28)</td>
<td>533 (0.28)</td>
<td>1.02 (0.87–1.19)</td>
<td>0.79</td>
<td>0.004</td>
</tr>
<tr>
<td>AC</td>
<td>147 (0.09)</td>
<td>133 (0.07)</td>
<td>1.41 (1.09–1.81)</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>AT</td>
<td>32 (0.02)</td>
<td>68 (0.04)</td>
<td>0.60 (0.38–0.94)</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diplotype</th>
<th>Cases</th>
<th>Controls</th>
<th>OR (95% CI)</th>
<th>(P) value</th>
<th>Gobal (P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT/GT</td>
<td>290 (0.37)</td>
<td>376 (0.39)</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GC/GT</td>
<td>262 (0.34)</td>
<td>317 (0.33)</td>
<td>1.07 (0.85–1.35)</td>
<td>0.54</td>
<td>0.02</td>
</tr>
<tr>
<td>AC/GT</td>
<td>82 (0.11)</td>
<td>91 (0.09)</td>
<td>1.17 (0.82–1.66)</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>GC/GC</td>
<td>56 (0.07)</td>
<td>82 (0.08)</td>
<td>0.89 (0.60–1.31)</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>AC/GC</td>
<td>45 (0.06)</td>
<td>31 (0.03)</td>
<td>1.88 (1.13–3.14)</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>AT/GT</td>
<td>18 (0.02)</td>
<td>41 (0.04)</td>
<td>0.57 (0.31–1.05)</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>AT/GC</td>
<td>8 (0.01)</td>
<td>20 (0.02)</td>
<td>0.52 (0.21–1.26)</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>AC/AT</td>
<td>6 (0.008)</td>
<td>6 (0.006)</td>
<td>1.30 (0.37–4.58)</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>AC/AC</td>
<td>7 (0.01)</td>
<td>3 (0.003)</td>
<td>3.03 (0.78–11.80)</td>
<td>0.09</td>
<td></td>
</tr>
</tbody>
</table>

\(\text{\(D\)'=0.03\) respectively.\}}
(De Marchis et al. 2004, Maguire et al. 2005). Our results add important information to this well-characterised allelic imbalance data on chromosome 17q by indicating a role of individual’s genotype on BC susceptibility.

As STATs integrate the signals from several cytokines, growth factors and hormones, an effect on aberrant activation of transcription and the pathogenesis of cancer can be assumed through various genes interacting with each other. For the same reason, STATs and STAT-signalling could represent effective molecular targets for the development and prognosis of new anticancer treatments (Calo et al. 2003). Although our findings need further evaluation, the present results suggest a role for the genetic variation in the STAT genes in the pathogenesis of human BC. The observed genetic variation may also cause interindividual variation in the response to STAT-signalling targeted therapy.

Acknowledgements
The authors are grateful to Julia Schmutzhard, Helmholtz-University Group Molecular Epidemiology, DKFZ, for performance and evaluation of the whole genome amplification. The German controls were collected by Peter Bugert, Institute of Transfusion Medicine and Immunology, Red Cross Blood Service of Baden-Württemberg-Hessa, University of Heidelberg, Faculty of Clinical Medicine, Mannheim, Germany. The German breast cancer samples were collected within a project funded by the Deutsche Krebshilfe, supported by the Center of Molecular Medicine Cologne (CMMC) and coordinated by Prof. Rita K Schmutzler. This study was supported by the Tumorzentrum Heidelberg/Mannheim and a grant from the EU (LSHC-CT-2005-503465). The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

References
Meindl A 2002 Comprehensive analysis of 989 patients with breast or ovarian cancer provides BRCA1 and BRCA2 mutation profiles and frequencies for the German population. International Journal of Cancer 97 472–480.


