A Novel Polymorphism in the Promoter Region of *ERBB4* Is Associated with Breast and Colorectal Cancer Risk

Matjaž Rokavec,[1,2,3] Christina Justenhoven,[1,2] Werner Schroth,[1,2] Monica Adina Istrate,[1,2] Susanne Haas,[4] Hans-Peter Fischer,[4] Caren Vollmert,[5] Thomas Illig,[5] Ute Hamann,[6] Yon-Dschun Ko,[7] Damjan Glavač,[3] and Hiltrud Brauch[1,2]

Abstract

Purpose: The receptor tyrosine kinase ERBB4/HER4 plays a role in cell division, migration, differentiation, as well as apoptosis, and is frequently overexpressed in breast and colorectal tumors. To understand the role of genetic variations in the regulation of *ERBB4* expression, we identified new polymorphisms and investigated their functional implication and risk association with breast and colorectal cancer.

Experimental Design: We screened colorectal tumors from 92 patients for genetic variants at the *ERBB4* ATG -1000 bp 5′-regulatory region by denaturing high-performance liquid chromatography and sequencing. Variants were subjected to DNA-protein interaction analyses (electrophoretic mobility shift assay), reporter gene assays in breast cancer cell lines MDA134 and MDA157, and immunohistochemical analyses of breast tumors. We established genotype frequencies within a breast cancer case-control collection (1,021 cases, 1,015 population-based controls) and a colorectal cancer case-control collection (459 cases, 569 blood donors) using matrix-assisted laser desorption ionization/time of flight mass spectrometry. Adjusted odds ratios (OR) and 95% confidence intervals (CI) were assessed by multivariate logistic regression.

Results: We identified five new germ line variants -815 A>T, -782 G>T, -638 insTC, -267 C>G, and -219 del10bp. Two variants showed *in vitro* functional effects. The -782T allele showed lower protein binding affinity and lower promoter activity compared with the -782G allele, however, the -815T allele showed higher protein binding affinity and higher promoter activity. The -782T variant was identified as a risk allele for breast and colorectal cancer (OR, 1.59; 95% CI, 1.06-2.34 and OR, 2.21; 95% CI, 1.22-3.99, respectively).

Conclusion: The *ERBB4* -782 G>T polymorphism, by virtue of its *in vitro* functional implication and incidence, is a risk factor for breast and colorectal cancer.

Authors' Affiliations: [1]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany; [2]Department of Clinical Pharmacology University Tübingen, Tübingen, Germany; [3]Department of Molecular Genetics, Institute of Pathology, Medical Faculty, University of Ljubljana, Slovenia; [4]Institute of Pathology, Medical Faculty of the University of Bonn, Germany; [5]GSF-National Research Center for Environment and Health, Institute of Epidemiology, Neuherberg, Germany; [6]Deutsches Krebsforschungszentrum, Heidelberg, Germany; and [7]Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus Bonn, Germany

Received 2/21/07; revised 9/13/07; accepted 9/20/07.

Grant support: Robert Bosch Foundation of Medical Research, Stuttgart, Germany, the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0, and 01KW0114, and the Slovenian Research Agency (ARRS-P3-054). M. Rokavec was a Marie Curie Fellow of the European Commission under the Program HPMT-CT-2001-00269, "Fighting Breast Cancer" Tübingen/Stuttgart.

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.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Requests for reprints: Hiltrud Brauch, Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Auerbachstrasse 112, 70376 Stuttgart, Germany. Phone: 49-711-8101-3705; Fax: 49-711-85-9295; E-mail: hiltrud.brauch@ikp-stuttgart.de.

doi:10.1158/1078-0432.CCR-07-0457

The ERBB4 protein, also known as HER4, is a member of the epidermal growth factor (EGF/ERBB) family of receptor tyrosine kinases, which are key activators of signaling pathways involved in cell division, migration, adhesion, differentiation, and apoptosis (1). ERBB4 participates together with other members EGFR/ERBB1/HER1, ERBB2/HER2, and ERBB3/HER3 in ERBB signaling of which the mitogen-activated protein kinase and phosphatidylinositol-3 kinase – activated AKT pathways are the most prominent targets (1). Aberrant ERBB signaling is a frequent phenomenon of many human cancers with receptor tyrosine kinase overexpression correlating with the oncogenic potential (1). Particularly, EGFR and ERBB2 have well-characterized roles in carcinogenesis and prognosis (2). Both receptor tyrosine kinases are frequently overexpressed in tumor tissues (3), and cancer patients whose tumors have EGFR or ERBB2 alterations tend to have a more aggressive disease associated with variables predicting a poor outcome (2, 4). A detailed biochemical understanding and availability for extracellular manipulation made them attractive targets for pharmacologic intervention. As of today, a therapeutic monoclonal antibody for the inhibition of ERBB2, i.e., Herceptin

(trastuzumab), and a low–molecular weight molecule for the inhibition of EGFR, i.e., Iressa (gefinitib) have been introduced into clinical practice and were proven to successfully control disease with respect to metastatic breast and non–small cell lung cancer as well as head and neck squamous cell carcinoma, respectively (3). Whether ERBB4 may have a similar role in tumor biology and treatment is unclear.

ERBB4 is expressed at low levels in the majority of adult tissues including breast and colon (5). Insight into its relevance during embryonic development comes from targeted inactivation in mice, which die by embryonic day 11 due to defective trabeculation in the heart (6). Although ERBB4 seems to be dispensable for ductal branching during mammary gland development of the virgin mouse and early lobulo-alveolar development at the onset of pregnancy, it is required at the penultimate stage of lobulo-alveolar expansion and epithelial functional differentiation (7). The mechanism of *ERBB4* regulation is not fully understood, but there is increasing evidence that the estrogen receptor (ER) plays a role (8). In tumors, *ERBB4* is overexpressed in ~20% of breast cancers with mRNA levels reaching 170-fold expression compared with those observed in normal mammary glands (9). Although the frequency and range of overexpression are reminiscent of *ERBB2*, the underlying mechanism for *ERBB4* overexpression cannot be explained by gene amplification. Also, the role of *ERBB4* expression in tumors is controversial (discussed in ref. 9). Although *ERBB4* expression has been found to be associated with a well-differentiated phenotype (10) and favorable prognosis in breast cancer (5, 11, 12), other studies favor an oncogenic role (13, 14). Recently, it has been shown that the prognostic value of ERBB4 depends on the subcellular localization of the ERBB4 protein, i.e., nuclear ERBB4 staining has been associated with worse clinical outcome compared with cell membrane–associated ERBB4 staining (9). The subcellular localization depends on the expressed ERBB4 isoform in that only protease-cleavable isoforms can translocate to the nucleus (9).

On the gene structure level, *ERBB4* is subject to genetic variation; however, as of August 2007, limited data on variants in the 5′-regulatory region and their functional consequences are available. Because genetic variants in the 5′-regulatory region of *EGFR* have been shown to alter promoter activity and gene expression *in vitro* and *in vivo* (15), there is a possibility that variants of the *ERBB4* 5′-regulatory region may also exert such a role. We therefore screened the *ERBB4* ATG -1000 bp including the 5′-untranslated region with the transcriptional starting site at -98 for the presence of novel genetic variants and studied their functional consequences as well as potential roles in breast and colorectal cancer.

Materials and Methods

Mutation analyses and identification of germ line variants at the* ERBB4 *5′-regulatory region. Tumor DNA from 92 patients with primary colorectal adenocarcinoma was isolated by standard phenol chloroform extraction. PCR was used to amplify the 1 kb *ERBB4* 5′-regulatory region by three independent reactions. Fragments spanning ATG -1009 to -678 (primers, 5′-TTTCCGTAAGCGGCCCTC-3′ and 5′-GGGCGGTGGGAATAAAAC-3′; annealing temperature, 55°C) and ATG -701 to -261 (primers, 5′-GGGTTGTTTTATTCCCACCG-3′ and 5′-GGCCGGACTGTGCAGCTATT-3′; annealing temperature, 56°C) were amplified using Qiagen hotstart polymerase (Qiagen). The ATG -313 to +77 fragment (primers, 5′-AGGGAGCGAGCAAGCGAAGG-3′ and 5′-GAATCGCTGGGCTGGACGGT-3′; annealing temperature, 64°C) was amplified using Failsafe kit (buffer J; Epicentre). Reactions were carried out according to the manufacturer's instructions. PCR products were subjected to denaturing high-performance liquid chromatography analyses using the Transgenomic WAVE DNA Fragment Analysis System and Wavemaker software (Transgenomic). Chromatograms of each fragment were analyzed and amplimers with variant elution profiles were subsequently sequenced to confirm putative sequence variations. For this, PCR products were purified with Qiagen PCR purification kit (Qiagen), sequencing reactions were carried out with the ABI PRISM Big Dye terminator cycle sequencing kit 1.1 and sequence analysis was done with an ABI prism 310 automatic sequencer (Applied Biosystems). All *ERBB4* mutations identified in tumor tissue were analyzed accordingly in corresponding normal tissue.

Plasmid constructs. For the construction of a recombinant plasmid, the *ERBB4* 5′-regulatory region (-1006 to -5) was amplified from human genomic DNA with forward primer 5′-AGAACGCGTCCG-TAAGCGGCCCTCC-3′ and reverse primer 5′-AGACTCGAGGGAAGTCT-CAGATCCCG-3′. These oligonucleotides introduced a *Mlu*I and *Xho*I restriction site at the 5′- and 3′-ends of the amplified fragment, respectively. The PCR product was digested with *Xho*I and *Mlu*I (GeneCraft) and cloned into appropriately digested luciferase reporter gene vector pGL3-basic (Promega). The identity of the cloned *ERBB4* 5′-regulatory region was verified by sequencing. Variants of the 5′-regulatory region representing the genetic variants of interest were generated by introducing point mutations using Quickchange II site-directed mutagenesis kit (Stratagene) according to the manufacturer's recommendations.

Transient transfections and reporter gene assays. Reporter gene assays were carried out for the investigation of differential promoter activities of variant *ERBB4* alleles according to Burk et al. (16). Two breast cancer cell lines MDA134 (ER positive) and MDA157 (ER negative) were used for transfection. Statistical significance was determined using one-way ANOVA test with Dunett's post-test and $P < 0.01$ was considered statistically significant. The statistical tests were done by GraphPAD Prism version 3.03 software (GraphPad Software, Inc.).

Electrophoretic mobility shift assays. DNA-protein interaction was investigated by electrophoretic mobility shift assay (EMSA). The nuclear extract from the MDA134 BC cell line was prepared as described in Schreiber et al. (17). We used the transcription factor binding sites prediction software (TFSEARCH, version 1.3; ref. 18) to predict proteins that might bind to the *ERBB4* promoter regions containing the variants under investigation and identified Sp1 as a potential candidate. Therefore, nuclear extract or recombinant Sp1 protein (Promega) were used as protein sources in the EMSA experiments. The sequences of the oligonucleotide probes used are listed in Supplementary Table S1. The sense and antisense oligonucleotides of the respective probes were annealed by incubation at 99°C for 5 min and slow cooling to 25°C. Double-stranded probes were labeled with Cy5.5-dCTP (Amersham Biosciences) by Klenow polymerase (Fermentas UAB) for 2 h at 37°C and cleaned by Sephadex G50 (Amersham Biosciences) column purification. Gel retardation assays were carried out as follows: nuclear extract or recombinant Sp1 protein was preincubated with 1 μg of pdIdC in Goodwin binding buffer containing 10 mmol/L of HEPES, 60 mmol/L of KCl, 0.2% NP40, 2.5% Ficoll, 2 mmol/L of DTT for 10 min at room temperature in the absence or presence of specific competitors (unlabeled oligonucleotide probes) or 2 μg of Sp1 antibody (Sp1 mouse monoclonal antibody sc-17824X; Santa Cruz Biotechnology). The labeled probe (1 ng) was added and incubated for 20 min at room temperature. Samples were loaded on a 30-min prerun native 6% polyacrylamide gel and electrophoresis was carried out at room temperature in 0.5× Tris-borate EDTA buffer at 80 V for 80 min. Gel imaging was carried out using Odyssey Infrared Imaging System (LI-COR) at a wavelength of 700 nm.

Immunohistochemical ERBB4 analysis. Immunohistochemical stainings were done on 4-μm tissue sections of formalin-fixed paraffin-embedded breast tumors using a 1:50 dilution of monoclonal anti-ERBB4 (HFR-1) mouse monoclonal antibody (Labvision). Immunostaining and scoring was done as described elsewhere (19). Stained tissues were evaluated by an expert pathologist who was unaware of the clinopathologic data and genotype. The differences in mean expression with respect to the genotype were calculated using a nonparametric *t* test. Statistical tests were done using GraphPAD Prism 3.03 software (GraphPad Software).

ERBB4 ***promoter polymorphisms and investigation of an associated breast and colorectal cancer risk.*** A colorectal case-control collection from Slovenia and the German population–based breast cancer case-control collection GENICA were investigated. The colorectal case-control collection consisted of 459 patients with histologically confirmed colorectal carcinoma collected at the Institute of Oncology, Ljubljana between 1996 and 2004, and 569 unrelated blood donors collected at the Blood Transfusion Center of Slovenia (20). Genomic DNA from patients were isolated from snap-frozen normal colonic mucosa using NucPrep chemistry (Applied Biosystems) on ABI 6100 (Applied Biosystems) and of controls from blood mononuclear cells using ethanol/NaCl precipitation. The German GENICA breast cancer case-control collection (21, 22) included 1,021 cases with histologically confirmed breast cancer and 1,015 population-based controls matched in 5-year age classes. Genomic DNA was isolated from blood mononuclear cells using Puregene kit (Gentra Systems, Inc.). Patient and control characteristics are listed in Table 1. All study participants were of European descent. The GENICA study was approved by the Ethics Committee of the University of Bonn, Germany, and the colorectal cancer study was approved by the National Medical Ethics Committee of the Republic of Slovenia.

Germ line variants *ERBB4* -815 A>T, -782 G>T, -718 C>T, -638 insTC, -609 G>A, -267 C>G, and -219 del10bp were genotyped to establish allele and genotype frequencies in both case-control collections. Genotyping was carried out with matrix-assisted laser desorption ionization/time of flight mass spectrometry (21, 23) using SpectroCHIP microarray and Bruker Autoflex (Sequenom) as well as MTP Anchor Chip 400/384 TF and Bruker Ultraflex (Bruker Daltonics). All assays were carried out in 384-well plates including four negative controls with no DNA and duplicates of 10% of the samples as a quality control. The position of polymorphisms is reported relative to the initiator ATG (with A as +1) codon. Genotype frequencies and odds ratios (OR) with 95% confidence intervals (CI) were calculated using GraphPad InStat software, version 3.06 (GraphPad Software), and checked for Hardy-Weinberg equilibrium according to Pearson's χ^2 test. All OR and *P* values were calculated from a 2×2 table based on the frequency of each genotype versus the frequency of the major genotype between cases and controls. Multivariate logistic regression analyses were used to obtain adjusted ORs and 95% CIs. Because there was a significant difference with respect to breast cancer case and

Table 1. Characteristics of the study participants

Variables	German GENICA breast cancer case-control collection			Slovenian colorectal cancer case-control collection		
	Cases (*n* = 1,020)	Controls (*n* = 1,015)	*P**	Cases (*n* = 459)	Controls (*n* = 569)	*P**
	n (%)	*n* (%)		*n* (%)	*n* (%)	
Sex						
Male	0 (0)	0 (0)		253 (55.36)	396 (69.59)	<0.0001
Female	1,020 (100)	1,015 (100)		204 (44.64)	173 (30.41)	
Age (y)						
<50	225 (22.06)	226 (22.27)	0.910	60 (13.13)	480 (84.36)	<0.0001
≥50	795 (77.94)	789 (77.73)		397 (86.87)	89 (15.64)	
Menopausal status †						
Pre	248 (24.65)	235 (24.65)	0.607	N.A.		
Post	758 (75.35)	758 (75.35)				
BC in mother or sisters						
No	886 (86.86)	940 (92.61)	<0.0001	N.A.		
Yes	134 (13.14)	75 (7.39)				
OC use (y)						
Never	371 (36.44)	368 (36.33)	0.822	N.A.		
>0 < 5	180 (17.68)	185 (18.26)				
5 < 10	134 (13.16)	120 (11.85)				
≥10	333 (32.71)	340 (33.56)				
HRT use (y)						
Never	505 (49.70)	509 (50.25)	0.009	N.A.		
>0 < 10	245 (24.11)	290 (28.63)				
≥10	266 (26.18)	214 (21.13)				
BMI (kg/m^2)						
<20	90 (8.82)	73 (7.20)	0.446	N.A.		
20 < 25	469 (45.98)	471 (46.65)				
25 < 30	305 (29.90)	324 (31.95)				
≥30	156 (15.29)	146 (14.40)				
Smoking						
Never	585 (57.41)	555 (54.68)	0.350	N.A.		
Former	192 (18.84)	215 (21.18)				
Current	242 (23.75)	245 (24.14)				

Abbreviations: OC, oral contraceptives; HRT, hormone replacement therapy; BMI, body mass index; N.A., not available.
*The differences between cases and controls were evaluated by two-sided χ^2 test.
† Women were considered premenopausal if they reported bleedings in the year of interview. All other women were considered postmenopausal by either natural or surgically induced menopause.

control status for the variables family history of breast cancer and hormone replacement therapy usage (Table 1), we adjusted for these variables in the multivariate analysis. Likewise, there was a significant difference with respect to the colorectal cancer case and control status (Table 1) for the variables age and sex, the reason why we adjusted for these variables in multivariate analysis. Pairwise linkage disequilibrium (LD) and haplotypes were calculated and estimated using Arlequin 2.0 software.

Results

Identification of genetic variants within the* ERBB4 *5′-regulatory region. Seven genetic variants were identified within the ATG -1000 bp region by mutational analysis of 92 colorectal cancer patients. These included five novel variants, i.e., -815 A>T (2 patients), -782 G>T (8 patients), -638 insTC (1 patient), -267 C>G (5 patients), and -219 del10bp (1 patient), and two known polymorphisms -718 C>T (rs6735626, 61 patients) and -609 G>A (rs6735267, 28 patients; Fig. 1A). To investigate whether the five novel variants represented tumor-specific mutations or germ line variants, we sequenced the patients' corresponding constitutional DNA and observed that whenever a variant was present in the tumor, it was also present in the germ line.

ERBB4 ***variants -782T and -815T affect promoter activity in human breast cancer cell lines.*** We investigated the promoter activity of *ERBB4* variants -815T, -782T, -718T, -609A, and -267G by reporter gene assays in MDA134 (ER positive) and MDA157 (ER negative) breast cancer cell lines. Significant changes in promoter activity were observed for the *ERBB4* -782T and -815T variants. When compared with the major -782G allele, the promoter activity of the minor -782T allele was significantly reduced to 40% in ER positive ($P < 0.01$) and to 35% in ER negative cells ($P < 0.01$), but *ERBB4* promoter activity was not completely abolished (Fig. 1B). In contrast, the promoter activity of the -815T variant was significantly increased by 35% in ER positive ($P < 0.01$) and 55% in ER negative cells ($P < 0.01$). Other variants did not show any significant changes in promoter activity (Fig. 1B).

Effect of* ERBB4 *variants on DNA-protein interaction. We performed EMSAs using nuclear proteins from MDA134 cells (Figs. 2 and 3). The DNA-protein complex with the major *ERBB4* -782G allele probe showed a more intense band when compared with the complex obtained with the minor T allele probe (Fig. 2A, *lanes 2* and *6*). In competition assays with labeled G allele probes and an excess of unlabeled competitor oligonucleotides, we observed that the protein-DNA complex 1 required higher concentrations of T allele competitor than G allele competitor to diminish the band intensity (Fig. 2C; paired *t* test, $P = 0.0274$). These data show that the binding affinity of nuclear proteins to the minor T allele was significantly lower compared with the major G allele.

In silico analysis identified the Sp1 protein as a potential binding partner of the *ERBB4* -782 and -815–containing region, a reason why we did additional EMSA experiments with recombinant Sp1 protein. In the case of the -782G allele, we observed a Sp1-containing complex (Sp1 complex) with reduced electrophoretic mobility compared with that of complex 1, which was confirmed by a supershift with the Sp1 antibody (Fig. 2B, *lanes 5* and *6*). Incubation of the *ERBB4* -782G allele probe with nuclear extract led to the formation of complex 1 (Fig. 2B, *lane 1*), which was not supershifted by Sp1 antibody (Fig. 2B, *lane 2*). The same binding pattern was observed with the -782T probe (Supplementary Fig. S1, *lanes 1-4*). Lanes 3, 4, 7, and 8 of Fig. 2B show control reactions with a Sp1 consensus probe. The formation of a Sp1 complex and a supershifted Sp1 immunocomplex from the MDA134 nuclear extract suggests that Sp1 protein is present in this nuclear extract (Fig. 2B, *lanes 3* and *4*). The observed differences in the electrophoretic mobility of the DNA-protein complex 1 and Sp1 complex, together with the absence of a supershifted immunocomplex, however, suggest that in the MDA134 nuclear extract, a protein(s) other than Sp1 binds to the *ERBB4* -782–containing promoter region.

In the case of the *ERBB4* -815 A>T variant, we observed strong Sp1 binding (Fig. 3B). Incubation of the -815A allele probe with MDA134 nuclear extract or recombinant Sp1 protein led to the formation of a Sp1 complex (Fig. 3B, *lanes 1* and *5*), which was confirmed by supershift with Sp1 antibody in both cases (Fig. 3B, *lanes 2* and *6*). In addition, we observed a faster migrating complex (complex 2), which was also present when using the Sp1 consensus probe (Fig. 3B, *lanes 1* and *3*). However, we infer that this complex may not contain Sp1 protein due to a lack of supershift (Fig. 3B, *lanes 2* and *4*). The same binding pattern was observed with the -815T probe (Supplementary Fig. S1, *lanes 5-8*). Both protein-DNA complexes (Sp1 complex and complex 2) were more pronounced

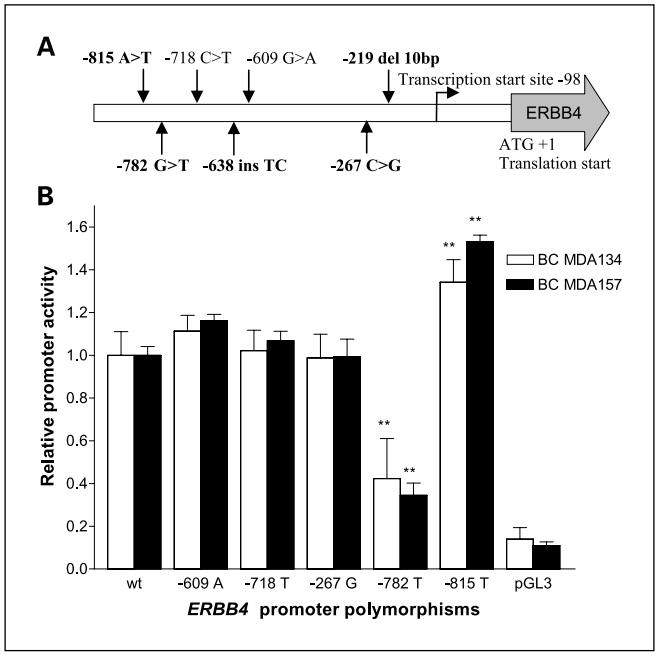

Fig. 1. *A,* 5′-regulatory region of the human *ERBB4* gene. Arrows, positions of genetic variants, transcription start site, and ATG start codon. Numbers, base positions with respect to the initiator ATG (with A as +1) codon. Novel genetic variants (*boldface*). *B,* transient reporter gene assays with constructs containing variants of the *ERBB4* ATG -1000 bp 5′-regulatory region. *ERBB4* promoter reporter constructs were transiently transfected into MDA134 (*white columns*) and MDA157 cells (*black columns*). Luciferase activity was measured 48 h after transfection and normalized by β-galactosidase. Relative promoter activity was determined by setting the activity of the wild-type (*wt*) variant to 1. Columns, mean relative promoter activities from three independent transfection experiments, each done in triplicate; bars, SD. **, $P < 0.01$, significant difference compared with the wild-type promoter activity (one-way ANOVA with Dunnet's post-test). In comparison with the -782G allele, the promoter activity of the -782T variant was significantly decreased to 40% and 35%, respectively. In comparison with the empty vector (pGL3 basic), the activity was not completely abolished. Likewise, the -815T variant was significantly increased by 35% and 55%, respectively.

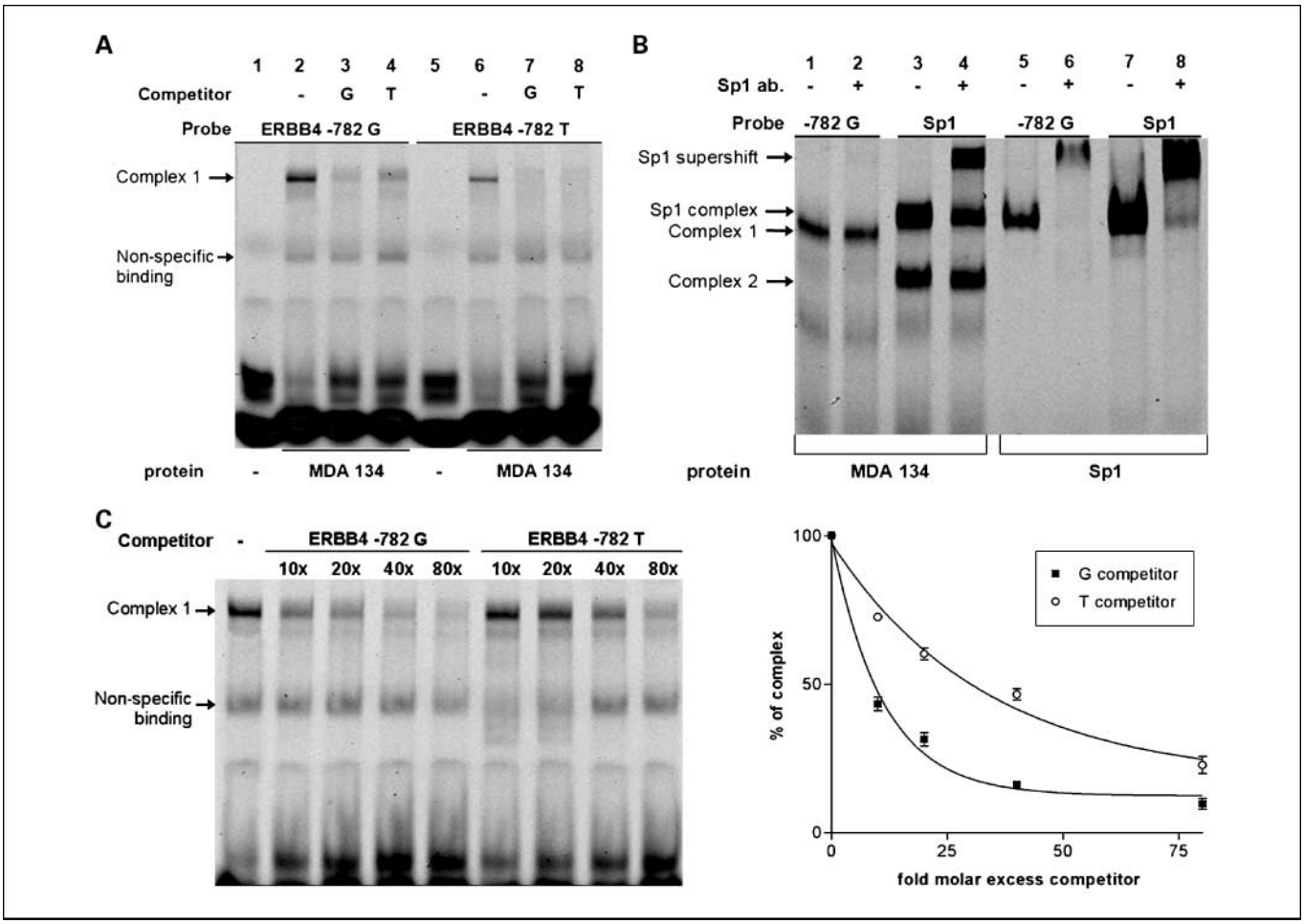

Fig. 2. EMSA analyses of the -782G>T – containing *ERBB4* 5′-regulatory region using nuclear protein extract of MDA 134 breast cancer cells or recombinant Sp1 protein. *A,* the nuclear proteins show higher affinity to interact with the G allele probe compared with the T allele probe (*complex 1, lanes 2* and *6*). In competition experiments done with 100-fold molar excess of the unlabeled G or T competitor, the DNA-protein complex 1 vanished, indicating specific binding (*lanes 3, 4, 7,* and *8*). Lanes 1 and 5, free -782G and -782T probes, respectively. *B,* recombinant Sp1 protein binds to the -782G probe (*lane 5*) as confirmed by supershift with Sp1 antibody (*lane 6*). Sp1 protein from MDA 134 nuclear extract does not visibly form an Sp1 complex with the -782G probe suggesting that complex 1 is instead formed by another nuclear protein (*lane 1*). This is supported by the absence of a supershifted immunocomplex (*lane 2*). Control reactions with Sp1 consensus oligonucleotide show specific Sp1 containing DNA-protein and immunocomplexes with nuclear extract (*lanes 3* and *4*) and recombinant Sp1 protein (*lanes 7* and *8*). *C,* competition assay with labeled G allele probe and 10, 20, 40, and 80-fold excesses of unlabeled G or T allele competitors. Retarded complexes were quantified and expressed as a percentage of the initial complex based on three competition gel shifts (■, G competitor; ○, T competitor). Bars, SE. Complex 1 was competed out at significantly lower amounts of G allele competitor compared with the same amounts of T allele competitor (paired *t* test, $P = 0.0274$).

when using the minor T allele probe in comparison with the major A allele probe (Fig. 3A, *lanes 6* and *2*, respectively). Competition assays with labeled T allele probe and excess of unlabeled competitor oligonucleotide showed that complex 2 was competed out at lower amounts of T allele competitor (Fig. 3C) compared with the same amounts of A allele competitor, a difference that was close to significance (paired *t* test, $P = 0.078$). There was no difference in competition of Sp1 complex with respect to the two competitor oligonucleotides (Supplementary Fig. S2; paired *t* test, $P = 0.14$). These data indicate that the binding affinity of some nuclear proteins to the minor T allele might be higher when compared with the major A allele. No DNA-protein interactions were observed for the other analyzed *ERBB4* promoter variants.

Expression of ERBB4 protein in breast tumors does not differ between carriers of the -782G/G and the G/T genotype. We analyzed the expression of ERBB4 protein in breast tumor tissues from 36 patients carrying the major -782G/G genotype and 21 patients carrying the heterozygous G/T genotype. There was no difference in expression levels between the two groups with respect to membranous, cytoplasmic, or nuclear staining (data not shown).

The* ERBB4 *-782T allele is associated with an increased risk for breast and colorectal cancer. To perform cancer risk association studies, we first needed to investigate whether the newly identified variants were single nucleotide polymorphisms or rare disease–associated germ line mutations. Therefore, we established allele and genotype frequencies in populations from Slovenia and Germany. The variants *ERBB4* -638 insTC and -219 del10bp were rare, with allele frequencies of <0.5%, and were not further analyzed. The minor allele frequencies of -815 A>T, -782 G>T, and -267 C>G ranged between 0.8% and 3%. These variants, together with the known polymorphisms, -718 C>T (minor allele frequency 43% and 45%, respectively) and -609 G>A (minor allele frequency 27% and 29%, respectively) met Hardy-Weinberg equilibrium in cases and

controls and were further analyzed in subsequent association studies (Table 2). Based on genotype frequencies in breast cancer and colorectal cancer case-control collections, we calculated ORs to obtain genotype-associated disease risk estimations (Table 2). The heterozygous genotype *ERBB4*-782G/T was significantly associated with an increased risk of developing breast cancer (OR, 1.59; 95% CI, 1.06-2.34) and colorectal cancer (OR, 2.21; 95% CI, 1.22-3.99). No other cancer risk associations were observed.

LD and haplotype prediction. To better understand the contribution of polymorphisms of the *ERBB4* 5′-regulatory region to cancer risk, we performed haplotype analyses. Pairwise LD across the 1 kb region is shown in Table 3A. The highest LD was observed between -718 C>T and -609 G>A ($r^2 = 0.47$), whereas other polymorphisms showed low levels of LD. Six haplotypes with frequencies >0.5% in populations of European descent, as well as corresponding frequencies in breast and colorectal cancer populations, are listed in Table 3B. Haplotype 5 (*ERBB4* -815A, -782T, -718T, -609G, and -267C) was associated with an increased risk for breast cancer (OR, 1.73; $P = 0.020$) and colorectal cancer (OR, 2.10; $P = 0.013$). It is noteworthy that this is the sole haplotype including the -782T variant, which has been identified as the at-risk allele in the preceding single locus analysis.

Discussion

According to a current hypothesis, cancer susceptibility is largely polygenic, which infers that many genetic polymorphisms may cooperate in the carcinogenic process; each conferring a small effect on the overall cancer risk (24). Based on clinical observations (5, 11, 12) and biological plausibility (3, 5, 8), we attempted a candidate gene approach for breast and colorectal cancer susceptibility by scrutinizing the *ERBB4* promoter region for its cancer risk potential. On the gene structure and functional level, we identified five novel variants and showed their effects on promoter activity. On the population level, we established allele and genotype frequencies in two

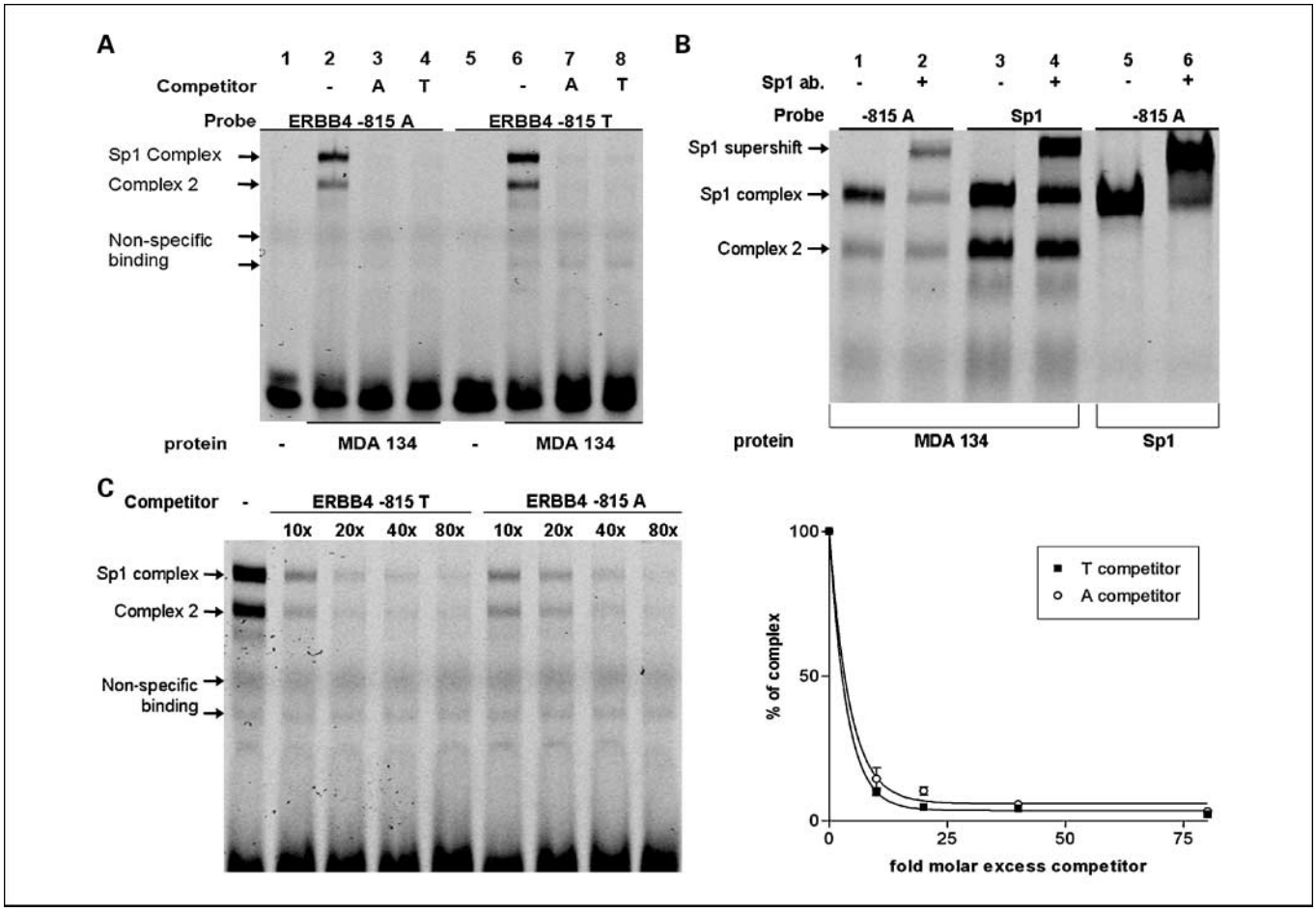

Fig. 3. EMSA analyses of the -815 A›T – containing *ERBB4* 5′-regulatory region using nuclear protein extract of MDA 134 breast cancer cells or recombinant Sp1 protein. *A,* the nuclear proteins show a higher affinity to interact with the T allele probe compared with the A allele probe (Sp1 complex and complex 2, *lanes 6* and *2*, respectively). In competition experiments done with 100-fold molar excess of the unlabeled A or T competitor, the DNA-protein complexes vanished, indicating specific binding (*lanes 3, 4, 7,* and *8*). Lanes 1 and 5, free -815A and -815T probes, respectively. *B,* recombinant Sp1 protein and Sp1 protein from the nuclear extract bind to the -815A probe (*lanes 5* and *1*) as confirmed by supershift of the immunocomplex with Sp1 antibody (*lanes 6* and *2*). Control reactions with Sp1 consensus oligonucleotide show a specific Sp1-containing DNA-protein and immunocomplexes (*lanes 3* and *4*). Complex 2 was not supershifted by the Sp1 antibody, indicating that Sp1 protein is not part of complex 2. *C,* competition assay with labeled T allele probe and 10, 20, 40, and 80-fold excesses of unlabeled T or A allele competitors. Complex 2 was quantified and expressed as a percentage of the initial complex based on three competition gel shifts (■, T competitor; ○, A competitor). Bars, SE. Complex 2 was competed out at lower amounts of T allele competitor compared with the same amounts of A allele competitor, a difference which was close to significance (paired *t* test, $P = 0.078$). There was no difference between competition of Sp1 complex with T or A allele competitor (Supplementary Fig. S2; paired *t* test, $P = 0.14$).

Table 2. Frequency of *ERBB4* genetic variants in German and Slovenian populations as well as corresponding breast and colorectal cancer populations

Single nucleotide polymorphism*	Genotype/ minor allele	GENICA controls (%)	GENICA breast cancer cases (%)	Adjusted OR (95% CI) †	Slovenian controls (%)	Slovenian colorectal cancer cases (%)	Adjusted OR (95% CI) ‡
-815 A>T	A/A	972 (98.48)	977 (98.99)	1	520 (98.48)	422 (99.06)	1
	A/T	15 (1.52)	10 (1.01)	0.68 (0.30-1.52)	8 (1.52)	4 (0.94)	0.44 (0.10-1.88)
	T/T	0	0	—	0	0	—
	T frequency	0.008			0.008		
-782 G>T	G/G	949 (96.74)	939 (94.66)	1	511 (96.60)	412 (92.79)	1
	G/T	32 (3.26)	53 (5.34)	1.59 (1.06-2.34)	18 (3.40)	32 (7.21)	2.21 (1.22-3.99)
	T/T	0	0	—	0	0	—
	T frequency	0.02			0.02		
-718 C>T (rs6735626)	C/C	325 (31.40)	336 (33.50)	1	164 (29.23)	135 (29.67)	1
	C/T	525 (50.72)	475 (47.36)	0.87 (0.71-1.07)	287 (51.16)	231 (50.77)	1.035 (0.69-1.56)
	T/T	185 (17.87)	192 (19.14)	1.02 (0.79-1.33)	110 (19.61)	89 (19.56)	0.97 (0.75-1.25)
	T frequency	0.43			0.45		
-609 G>A (rs6735267)	G/G	514 (53.54)	515 (53.09)	1	233 (51.32)	238 (57.49)	1
	G/A	375 (39.06)	375 (38.66)	1.01 (0.83-1.22)	182 (40.09)	150 (36.23)	0.75 (0.49-1.13)
	A/A	71 (7.40)	80 (8.25)	1.19 (0.84-1.69)	39 (8.59)	26 (6.28)	0.56 (0.22-1.09)
	A frequency	0.27			0.29		
-267 C>G	C/C	928 (93.55)	935 (91.76)	1	548 (97.34)	439 (96.06)	1
	C/G	52 (5.24)	64 (6.28)	1.17 (0.82-1.68)	15 (2.66)	18 (3.94)	1.43 (0.49-4.11)
	G/G	2 (0.21)	1 (0.10)	—	0	0	—
	G frequency	0.03			0.01		

*The positions of the genetic variants were counted from the ATG codon. OR and 95% CIs were calculated based on the frequency of each genotype versus the major genotype between cases and controls.
† Adjusted for family history of breast cancer and hormone replacement therapy usage.
‡ Adjusted for age and sex.

European populations and did association tests in large population-based case-control collections to understand whether any of these variants might be associated with the risk of developing breast or colorectal cancer.

In search of *ERBB4* promoter variants, we first analyzed DNA from colorectal tumors for the presence or absence of mutations in the ATG -1000 bp 5′-regulatory region. We chose these tumors because ERBB4 is similarly expressed in colorectal and breast tumors; however, the latter have not been available to us. We observed seven genetic variants which, upon analyses of the corresponding normal mucosa, were of constitutional origin. Among these, *ERBB4* -815 A>T, *ERBB4* -782 G>T, *ERBB4* -638 insTC, *ERBB4* -267 C>G, and *ERBB4* -219 del10bp are novel variants, which to the best of our knowledge, are described for the first time herein. The other variants, *ERBB4* -718 C>T and *ERBB4* -609 G>A, refer to known polymorphisms rs6735626 and rs6735267, respectively. The highest LD was between *ERBB4* -718 C>T and *ERBB4* -609 G>A (r^2 = 0.47), whereas other paired polymorphisms showed low levels of LD. The three novel variants *ERBB4* -815 A>T, *ERBB4* -782 G>T, and *ERBB4* -267 C>G had heterozygous genotype frequencies ranging from 1.5% to 5.2%, and therefore, may play a possible role in colorectal and breast cancer risk.

To provide supportive *in vitro* functional evidence for a possible role in cancer, we set out to investigate the binding capacity of major and minor alleles with respect to nuclear proteins and their influence on promoter activity. The two most upstream located polymorphisms, *ERBB4* -815 A>T and *ERBB4* -782 G>T, showed significant effects in experiments addressing these biochemical aspects. In binding experiments with nuclear extracts from MDA134 breast cancer cells, both polymorphic sites formed specific DNA-protein complexes. Because *in silico* analyses pointed to the transcription factor Sp1 as a likely binding protein due to the presence of a Sp1 binding site within the *ERBB4* -815 and -782 polymorphic region, we further tested whether this DNA-protein interaction involved Sp1. This is of particular interest because Sp1 is an important regulator of *ERBB2* (25) and *EGFR* (15) expression. Sp1 interaction was unanimously shown for the -815 polymorphic site, however, the situation with respect to the -782 site was more complex. Although the latter was able to bind recombinant Sp1, it did not visibly bind Sp1 from the nuclear extract. Altogether, our experiments suggest that in addition to Sp1, there may be at least one other nuclear protein with even higher binding affinity. It is of note that in the case of the *ERBB4* -782 polymorphic site, the major G allele showed higher protein binding affinity in competition experiments compared with the minor T allele. Accordingly, the minor T allele may be considered compromised in its capacity to bind nuclear protein, a notion that is well in line with the >2.5-fold observed *in vitro* reduction of promoter activity of this variant. Interestingly, the protein binding affinity of the minor *ERBB4* -815T allele was apparently stronger than that of the major A allele, a circumstance that might help to explain the up to 1.5-fold observed increase of promoter activity. Altogether, our results underscore that both polymorphisms show effects on promoter activity. Moreover, they suggest a more versatile role for the *ERBB4* -782 G>T polymorphism due to its putative multiple protein binding partners and capacity to obliterate transcriptional activity to a substantial degree.

These interpretations are in line with our patient-based results from breast and colorectal cancer risk association studies from Germany and Slovenia. From the five *ERBB4* promoter polymorphisms included in the analyses, the -782 G>T polymorphism was significantly associated with higher cancer risks. In detail, the frequency of the heterozygote genotype G/T was 3.3% in the GENICA population-based controls and 3.4% in the Slovenian blood donor controls. This frequency was significantly increased in the corresponding cancer groups, which translates into a 1.6-fold increased risk for breast cancer and a 2.2-fold increased risk for colorectal cancer. The -782 G>T polymorphism is less frequent than the known promoter polymorphisms rs673626 and rs6735267. Yet, our findings draw substance from the large sizes of case-control comparisons (i.e., >2,000 study subjects from the GENICA breast cancer collection and >1,000 study subjects from the Slovenian colorectal cancer collection), appropriate study design, and high genotyping call rates of >97%. Despite the lower size of the colorectal cancer collection, we had an 85% power to detect a risk allele with a frequency of 2%, which confers a relative risk of 2.2 at the 0.05 level of significance. In the colorectal cancer study, cases and controls did not match with respect to age and sex; however, genotype frequencies matched those of the larger GENICA control group and were therefore regarded relevant. Haplotype analysis did not add more information about the influential single nucleotide polymorphism's accuracy. The only haplotype associated with an increased risk for breast and colorectal cancer was the haplotype that included the -782T variant, which has been identified as the at-risk allele in the single locus analysis. At this point, we are aware of the possibility that some of our observed associations may vanish upon stringent correction for multiple testing. Due to the explorative nature of our study, we purposely did not perform multiple testing corrections in order to not eliminate potentially important results from this first investigation. To understand the relevance of our data, however, it will be important to replicate these findings in independent data sets.

Our mutational, *in vitro* functional, and population-based cancer studies suggest a role for the *ERBB4* -782 G>T polymorphism in breast and colorectal cancer risk. Whether this polymorphism will be useful to identify individuals at increased cancer risk should be further pursued in large international population-based studies. This is encouraged by recent developments in which *CASP8* (rs1045485), originally identified as a putative breast cancer susceptibility gene in two British populations (26), was successfully confirmed within a comprehensive international effort of the Breast Cancer Association Consortium on the basis of >30,000 breast cancer cases and controls (27).

In addition, the question arises whether the effects of the polymorphisms may also influence tumor biology and behavior. We observed cytoplasmic, nuclear, and membranous ERBB4 protein expression in breast tumors; however, there was no difference with respect to the underlying major or minor -782 genotypes. This may not be surprising because breast cancer risk may rather be reflected by features of the normal mammary gland, yet ERBB4 expression in normal breast is too low to be quantitatively analyzed (9). Therefore, our ERBB4 expression findings do not contradict the observed genotype-associated risk associations, but rather point to other regulatory mechanisms of ERBB4 expression in tumors. We infer that the role of ERBB4 genotypes may be limited to normal tissue.

Recently, it has been shown that ERBB4 functions as a proapoptotic protein suppressing the growth of malignant cells, and when activated, induces cell killing of breast tumor cells and enhances apoptosis (28). From this, we may infer that

Table 3. Pairwise LD of five *ERBB4* genetic variants and major haplotypes with their frequencies in German and Slovenian control as well as cancer populations

(A) Pairwise LD

	-815 A>T	-782 G>T	-718 C>T	-609 G>A	-267 C>G
-815 A>T		0.00049	0.00057	0.00052	0.00003
-782 G>T	0.03338		0.01637	0.00664	0.00052
-718 C>T	0.30979	0.83839		0.47183	0.02272
-609 G>A	0.42306	0.99997	0.98746		0.01080
-267 C>G	0.01021	0.99252	1.00000	0.9999	

(B) *ERBB4* haplotypes and their frequencies in analyzed populations

ID	Major haplotypes*	GENICA controls (%)	GENICA breast cancer cases (%)	OR†	P	Slovenian controls (%)	Slovenian colorectal cancer cases (%)	OR†	P
1	AGCGC	53.56	52.72	1		52.69	52.21	1	
2	AGTAC	26.70	27.53	1.05	0.810	28.93	23.82	0.83	0.096
3	AGTGC	15.17	13.34	0.89	0.218	14.09	17.66	1.27	0.071
4	AGCGG	2.65	3.23	1.24	0.304	1.41	1.97	1.42	0.375
5	ATTGC	1.52	2.59	1.73	0.020	1.71	3.55	2.10	0.013
6	TGCGC	0.72	0.49	0.70	0.495	0.71	0.46	0.66	0.562

NOTE: In A, D' abs (bottom left half) and r^2 (top right half).
*The order of genetic variants from left to right are: -815 A>T, -782 G>T, -718 C>T, -609 G>A and -267 C>G with the minor allele being highlighted in black.
† Odds ratios and *P* values were calculated based on the frequency of each haplotype versus haplotype 1 between cases and controls.

high levels of ERBB4 may have a protective cellular effect, thus explaining why a low constitutional expression genotype may confer breast and colorectal cancer risk. This interpretation is in line with the reported improved breast cancer survival times of patients with high *ERBB4* expression in their tumors (11, 12). Moreover, there has been mention that the nuclear localization of an ERBB4 isoform is associated with worse outcome compared with membrane localization (9). Whether the *ERBB4* -782 G>T polymorphism plays a role in the aberrant expression of this or any other deleterious ERBB4 isoform remains to be determined. Finally, it will be important to understand the role of this polymorphism on the ERBB effector pathways mitogen-activated protein and phosphatidylinositol-3 kinase–activated AKT, which may eventually stimulate progress into the development of targeted drug treatment for breast and colorectal cancer.

Acknowledgments

We thank all participants of the German GENICA study and the Slovenian colorectal cancer study and Sandra Brod for technical assistance.

References

1. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol 2001;2:127–37.
2. Ross JS, Fletcher JA. The HER-2/neu oncogene in breast cancer: prognostic factor, predictive factor, and target for therapy. Oncologist 1998;3:237–52.
3. Holbro T, Civenni G, Hynes NE. The ErbB receptors and their role in cancer progression. Exp Cell Res 2003;284:99–110.
4. Nicholson RI, Gee JM, Harper ME. EGFR and cancer prognosis. Eur J Cancer 2001;37 Suppl 4:S9–15.
5. Srinivasan R, Poulsom R, Hurst HC, Gullick WJ. Expression of the c-erbB-4/HER4 protein and mRNA in normal human fetal and adult tissues and in a survey of nine solid tumour types. J Pathol 1998;185:236–45.
6. Gassmann M, Casagranda F, Orioli D, et al. Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. Nature 1995;378:390–4.
7. Jones FE, Golding JP, Gassmann M. ErbB4 signaling during breast and neural development: novel genetic models reveal unique ErbB4 activities. Cell Cycle 2003;2:555–9.
8. Zhu Y, Sullivan LL, Nair SS, et al. Coregulation of estrogen receptor by ERBB4/HER4 establishes a growth-promoting autocrine signal in breast tumor cells. Cancer Res 2006;66:7991–8.
9. Junttila TT, Sundvall M, Lundin M, et al. Cleavable ErbB4 isoform in estrogen receptor-regulated growth of breast cancer cells. Cancer Res 2005;65:1384–93.
10. Bacus SS, Chin D, Yarden Y, Zelnick CR, Stern DF. Type 1 receptor tyrosine kinases are differentially phosphorylated in mammary carcinoma and differentially associated with steroid receptors. Am J Pathol 1996;148:549–58.
11. Witton CJ, Reeves JR, Going JJ, Cooke TG, Bartlett JM. Expression of the HER1-4 family of receptor tyrosine kinases in breast cancer. J Pathol 2003;200:290–7.
12. Tovey SM, Witton CJ, Bartlett JM, Stanton PD, Reeves JR, Cooke TG. Outcome and human epidermal growth factor receptor (HER) 1-4 status in invasive breast carcinomas with proliferation indices evaluated by bromodeoxyuridine labelling. Breast Cancer Res 2004;6:R246–51.
13. Bieche I, Onody P, Tozlu S, Driouch K, Vidaud M, Lidereau R. Prognostic value of ERBB family mRNA expression in breast carcinomas. Int J Cancer 2003;106:758–65.
14. Lodge AJ, Anderson JJ, Gullick WJ, Haugk B, Leonard RC, Angus B. Type 1 growth factor receptor expression in node positive breast cancer: adverse prognostic significance of c-erbB-4. J Clin Pathol 2003;56:300–4.
15. Liu W, Innocenti F, Wu MH, et al. A functional common polymorphism in a Sp1 recognition site of the epidermal growth factor receptor gene promoter. Cancer Res 2005;65:46–53.
16. Burk O, Tegude H, Koch I, et al. Molecular mechanisms of polymorphic CYP3A7 expression in adult human liver and intestine. J Biol Chem 2002;277:24280–8.
17. Schreiber E, Matthias P, Muller MM, Schaffner W. Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells. Nucleic Acids Res 1989;17:6419.
18. Heinemeyer T, Wingender E, Reuter I, et al. Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res 1998;26:362–7.
19. Tovey SM, Dunne B, Witton CJ, Cooke TG, Bartlett JM. HER4 in breast cancer: comparison of antibodies against intra- and extra-cellular domains of HER4. Breast Cancer Res 2006;8:R19.
20. Potocnik U, Ravnik-Glavac M, Golouh R, Glavac D. Naturally occurring mutations and functional polymorphisms in multidrug resistance 1 gene: correlation with microsatellite instability and lymphoid infiltration in colorectal cancers. J Med Genet 2002;39:340–6.
21. Justenhoven C, Hamann U, Pesch B, et al. ERCC2 genotypes and a corresponding haplotype are linked with breast cancer risk in a German population. Cancer Epidemiol Biomarkers Prev 2004;13:2059–64.
22. Pesch B, Ko Y, Brauch H, et al. Factors modifying the association between hormone-replacement therapy and breast cancer risk. Eur J Epidemiol 2005;20:699–711.
23. Vollmert C, Windl O, Xiang W, et al. Significant association of a M129V independent polymorphism in the 5′ UTR of the PRNP gene with sporadic Creutzfeldt-Jakob disease in a large German case-control study. J Med Genet 2006;43:e53.
24. Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. Nat Rev Cancer 2004;4:850–60.
25. Chen Y, Gill GN. Positive and negative regulatory elements in the human erbB-2 gene promoter. Oncogene 1994;9:2269–76.
26. MacPherson G, Healey CS, Teare MD, et al. Association of a common variant of the CASP8 gene with reduced risk of breast cancer. J Natl Cancer Inst 2004;96:1866–9.
27. Cox A, Dunning AM, Garcia-Closas M, et al. A common coding variant in CASP8 is associated with breast cancer risk. Nat Genet 2007;39:352–8.
28. Naresh A, Long W, Vidal GA, et al. The ERBB4/HER4 intracellular domain 4ICD is a BH3-only protein promoting apoptosis of breast cancer cells. Cancer Res 2006;66:6412–20.