[CANCER RESEARCH 64, 7–12, January 1, 2004]

Advances in Brief

Modification of Breast Cancer Risk in Young Women by a Polymorphic Sequence in the *egfr* Gene

Burkhard Brandt,[1] Silke Hermann,[2] Kurt Straif,[3] Nicola Tidow,[1] Horst Buerger,[4] and Jenny Chang-Claude[2]

[1]Institute for Clinical Chemistry and Laboratory Medicine, University of Muenster, Muenster, Germany; [2]German Cancer Research Center, Heidelberg, Germany; [3]Carcinogen Identification and Evaluation, IARC, WHO, Lyon, France; and [4]Institute of Pathology, University of Muenster, Muenster, Germany

Abstract

The regulation of the epidermal growth factor receptor (*egfr*) gene in human cancer is not yet fully understood. Recent data on a polymorphic CA repeat located at the 5′-regulatory sequence in intron 1 of the *egfr* gene [*egfr* CA simple sequence repeat (SSR) I] point to a possible inheritance of cancer risk associated with the *egfr* gene. Furthermore, we have detected frequent allelic imbalances restricted to the *egfr* CA SSR I in breast cancer tissue and nontumorous breast tissue adjacent to invasive and *in situ* breast cancer representing amplifications. Therefore, we conducted a population-based case-control study to assess the relationship between the *egfr* polymorphism and breast cancer risk. Cases with a first primary breast cancer by age 50 years and age-matched population controls provided information on known and suspected risk factors. The allelic length of the *egfr* CA SSR was determined in 616 cases and 1072 population-sampled controls. Genotypes were categorized for analysis by allele length. Multivariate logistic regression was used to compare genotype distributions, accounting for other risk factors, and to investigate gene-environment interactions. We found a modifying effect, albeit no main effect, of the allelic length of the *egfr* polymorphism on breast cancer risk. The presence of two long alleles (≥19 CA) was associated with a significantly elevated odds ratio (OR) of 10.4 [95% confidence interval (CI), 1.85–58.70] among women with a first-degree family history of breast cancer ($P = 0.015$ for interaction). The risk increase associated with high red meat consumption (OR, 10.68; 95% CI, 1.57–72.58) and the protective effect of high vegetable intake (OR, 0.07; 95% CI, 0.004–1.07) was also most pronounced among carriers of two long alleles (≥19 CA). The length of the *egfr* CA SSR may increase the risk for familial breast cancers, and its effect could be modulated by dietary factors.

Introduction

Breast cancer is the most common incident cancer and the second most common fatal cancer among women in Western countries. Some 15–20% of cases are associated with family history of breast cancer, but only 15% of familial risk can be explained by an inherited mutation in highly penetrant genes, such as *BRCA1* and *BRCA2* (1–3). It is now widely accepted that cancer susceptibility can also be due to low-penetrance genetic variants and gene-environment interactions (4). The search for those genes has been directed mainly to functionally significant variants in genes that are known or assumed to be biologically related to breast cancer (5).

Animal models have shown that an increased expression of oncogenes coding for tyrosine kinase receptors is involved in early breast carcinogenesis (6). In particular, an increased rate of breast cancer due to overexpression of the epidermal growth factor receptor (EGFR) was observed in transgenic mice (7). However, the regulation of the *egfr* gene in human cancers is not yet fully understood. Data from studies on a polymorphic CA repeat located at the 5′-regulatory sequence in the intron 1 of the *egfr* gene [*egfr* CA simple sequence repeat (SSR) I] suggest that this polymorphic site may play a role in cancer susceptibility. Recently, we reported that the basal transcription activity of the gene was inversely related to the number of CA repeats in that CA SSR I (8). Moreover, we detected frequent allelic imbalances (AIs) restricted to the *egfr* CA SSR I in breast cancer tissue indicative of amplifications (9, 10). These amplifications almost always involved the regulatory sequence centered by the CA SSR I of intron 1 and led to EGFR overexpression (10). As also described recently by our research group, the close proximity of an inducible fragile site next to the *egfr* gene suggests that exogenous factors such as nutritional habits may play an important role in the induction of *egfr* amplifications (10). Therefore, we investigated whether the length of the inherited alleles of the CA SSR I in the *egfr* gene is associated with disease risk, particularly with regard to family history of cancer and dietary pattern, in a large case-control study of breast cancer.

Materials and Methods

Study Population. The population-based case-control study consisted of German-speaking women resident in two study regions (Freiburg and Rhein-Neckar-Odenwald) in southern Germany. Breast cancer patients under the age of 51 years at the time of diagnosis of incident *in situ* or invasive breast cancer were eligible for the study. They were either approached in the hospital or contacted by a letter from the attending physician after discharge. Two controls matched by exact age and study region were selected for each case from a random list of residents provided by the population registries. These control persons were invited to participate by letter. Overall, 706 of 1005 eligible cases (70.2%), and 1381 of 2257 eligible controls (61.2%) participated (11).

The study participants completed a self-administered risk factor questionnaire including information on demographic factors, anthropometric measures, and menstrual, reproductive, and breast-feeding history and family history of cancer. For analysis of dietary factors, the 1451 participants of the study region Rhein-Neckar-Odenwald were asked to complete a 176-item food frequency questionnaire after they had filled out a risk factor questionnaire. The participants were asked to document their nutritional habits in the year before the diagnosis of breast cancer or questionnaire completion. The food frequency questionnaire was developed at the German Cancer Research Center (Deutsches Krebsforschungszentrum) in Heidelberg, Germany and validated for food group, energy, and nutrient intake (11, 12). The food item list was defined on the basis of the dietary intake results in the German National Food Consumption Survey. The food frequency questionnaire was returned by 1288 of the 1451 participants (88.8%). Of these, 258 participants were excluded due to various reasons such as a high number of missing items or extreme over- and underreporting (13).

This analysis includes only German cases and controls, from whom questionnaire data and DNA from a blood sample were available. A total of 616 cases and 1072 controls were included in molecular analysis; genotyping was successful for 604 cases and 1063 controls. Nutritional and genotyping data were available from 1000 participants (311 cases and 689 controls).

Genotype Analysis. Genomic DNA isolation from peripheral blood was performed using the QIAamp Blood Kit (Qiagen), and tumor tissue DNA was

Received 8/21/03; revised 10/17/03; accepted 11/6/03.

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: B. Brandt and J. Chang-Claude contributed equally to this work.

Requests for reprints: Burkhard Brandt, Institute of Clinical Chemistry and Laboratory Medicine, University of Muenster, Albert-Schweitzer-Strasse 33, 48149 Muenster, Germany. Phone: 49-251-83-47226; Fax: 49-251-83-47226; E-mail: brandt@uni-muenster.de.

extracted using the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's instructions. As a control for PCR fragment length, DNA from the tumor cell line MDA-MB-468 was analyzed.

PCR amplification was performed using AmpliTaq DNA polymerase (Applera) in a 25-μl reaction volume containing 200 nM of each primer, 1× GeneAmp buffer II, 2 mM $MgCl_2$, 100 μM of each GeneAmp deoxynucleoside triphosphate (Applera), and 20 ng of sample DNA. The primer sequences are specific for microsatellite marker egfr in intron 1 of *egfr* gene*:* forward primer, 5′-GTT-TGA-AGA-ATT-TGA-GCC-AAC-C; and reverse primer, 5′-TTC-TTC-TGC-ACA-CTT-GGC-AC. Downstream primers were labeled with a fluorescent dye ([F]-amidite-6-carboxy fluorescein). Separation was done with a four-color laser-induced fluorescence capillary electrophoresis system (ABI PRISM 310 DNA Analyzer and ABI PRISM 3700 DNA Analyzer). One to two μl of the amplified PCR products were diluted in 20 μl of water (high-performance liquid chromatography grade) containing 0.5 μl of GENESCAN 500 (Tamra) or 400 HD (Rox) fluorescent size standard (Applera). Denatured PCR fragments were separated on the ABI PRISM 310 and ABI PRISM 3700 DNA Analyzer (Applera). Evaluation of the collected data was accomplished with GeneScan Analysis Software (Applera). All analyses were performed at least in duplicates of independent PCRs.

Genotypes were categorized for analysis by allele size and frequency. The length of the CA repeat is inversely associated with transcriptional activity of the gene and directly associated with the likelihood of an AI. Therefore, we first modeled the *egfr* polymorphism in three ways to reflect the activity of specific alleles: (*a*) having at least one short allele, considering the shorter of the two alleles for each subject; (*b*) having at least one long allele; and (*c*) the combined effect of the two alleles, by assessing the mean allele length for each subject. In addition, the *egfr* polymorphism was assessed by dichotomization of the sample into two groups based on the observed distribution of the repeats as well as previous analysis of transcriptional activity and likelihood for AI by allele size. We defined the cutoffs for a shorter allele with ≤16, ≤17, and ≤18 CA repeats and a longer allele with ≥17, ≥18, and ≥19 CA repeats, respectively.

Statistical Methods. The allele frequencies in cases and controls were compared, and the difference was assessed by the χ^2 test. The risk of breast cancer associated with *egfr* genotype and other risk factors was assessed using logistic regression modeling with stratification according to age (in years). Odds ratios (ORs) and their 95% confidence intervals (CIs) were calculated using the proportional hazards regression procedure of the statistical software package SAS Release 8.12 (SAS Institute, Cary, NC). Analyses were adjusted for lifetime duration of breastfeeding in months (as a continuous variable), first-degree family history of breast cancer (mother, sister, or daughter), number of full-term pregnancies (classified in three categories of 0, 1–2, and ≥3 full-term pregnancies), age at menarche (classified in three categories of <13 years, 13–14 years, and ≥15 years of age), and alcohol consumption (0 g/day, ≤18 g/day, and ≥19 g/day). The procedure of nutritional analyses has been described in detail elsewhere (13). In short, for evaluations considering nutritional factors, the calculated consumption of various food groups was analyzed as a categorical variable (in quartiles) based on item consumption among controls only. The relative risk of breast cancer was estimated for higher quartiles of food consumption in comparison with the lowest quartile of consumption. To test categorized variables for trend, the categories were scored and entered into the regression analysis as ordinal variables.

We also investigated whether the different genotypes interact with other risk factors previously shown to be related to breast cancer. Effect modification of dietary risk factors and risk associated with family history by genotype was evaluated for three models by defining the long allele to be ≥17, 18, and 19 CA repeats. Interactions between genetic and other variables were measured by using multiplicative terms and evaluated by the likelihood ratio tests.

Results

Mean ages of the 604 cases and 1063 controls were 42.4 and 42.6 years, respectively. Table 1 shows the distribution of different risk factors for cases and controls and the corresponding ORs. As reported previously, a larger number of full-term births and a longer duration of breastfeeding were associated with a significantly reduced breast cancer risk. Cases reported the occurrence of first-degree family history of breast cancer significantly more often than controls.

The *egfr* genotype distributions are presented in Table 2. The number of CA repeats among German participants (n = 1667) ranged from 8 to 23. There were 44 different diploid genotypes, and a trimodal distribution of the genotypes was observed (Table 2).

For the shorter allele, the most commonly observed allele was 16 repeats, both for the cases (63.2%) and the controls (66.6%). The most frequent longer allele was 20 repeats, both for the cases (36.4%) and the controls (34.5%). The three most frequent allele combinations among controls were 16 CA repeats/16 CA repeats (19.4%), 16 CA repeats/20 CA repeats (18.9%), and 16 CA repeats/18 CA repeats (15.8%). Among cases, these were 16 CA repeats/20 CA repeats (19.7%), 16 CA repeats/16 CA repeats (18.7%), and 16 CA repeats/18 CA repeats (12.8%). There is no apparent difference in the distribution between cases and controls (Table 2).

Table 1 *Distribution of selected risk factors in German cases and controls and the corresponding odds ratios*

	Cases (n = 604)	Percentage	Controls (n = 1063)	Percentage	OR[a] (95% CI)
Parity					
0 full-term pregnancies	131	21.69	224	21.07	1
1–2 full-term pregnancies	411	68.05	671	63.12	1.08 (0.84–1.40)
3+ full-term pregnancies	62	10.26	168	15.80	0.65 (0.44–0.94)
Total time of breastfeeding (mo)					
0	284	47.02	467	43.93	1
1–6	220	36.42	353	33.21	1.05 (0.83–1.32)
7+	100	16.56	243	22.86	0.69 (0.52–0.91)
First-degree family history					
No	528	87.42	1004	94.45	1
Yes	76	12.58	59	5.55	2.42 (1.69–3.45)
Menarche (yrs)					
≤12	224	37.09	388	36.50	1
13–14	292	48.34	504	47.41	0.98 (0.78–1.22)
15+	85	14.07	170	15.99	0.84 (0.62–1.15)
Missing	3	0.50	1	0.09	
Education					
Low	81	13.41	147	13.83	1
Medium	405	67.05	662	62.27	1.10 (0.81–1.48)
High	118	19.54	254	23.89	0.82 (0.57–1.17)
Alcohol consumption (g/day)					
None	120	19.9	173	16.3	1
≤18	393	65.1	796	74.9	0.72 (0.55–0.94)
19+	91	15.1	94	8.8	1.42 (0.97–2.06)

[a] OR, odds ratio; CI, confidence interval.

Table 2 *Distribution of allele combinations of the CA SSR in the egfr gene (by length of CA repeat) in German breast cancer patients and population controls*

		Cases		Controls	
EGFR1	EGFR2	Frequency	Percentage	Frequency	Percentage
8	18	1	0.17	0	0
13	16	0	0	1	0.09
13	18	0	0	1	0.09
14	15	0	0	1	0.09
14	16	7	1.16	12	1.13
14	17	0	0	2	0.19
14	18	0	0	4	0.38
14	19	0	0	1	0.09
14	20	1	0.17	2	0.19
14	21	2	0.33	0	0
14	22	0	0	1	0.09
15	15	0	0	1	0.09
15	16	12	1.99	14	1.32
15	17	0	0	3	0.28
15	18	4	0.66	3	0.28
15	19	2	0.33	0	0
15	20	3	0.50	8	0.75
15	21	1	0.17	4	0.38
16	16	113	18.71	206	19.38
16	17	28	4.64	46	4.33
16	18	77	12.75	168	15.8
16	19	6	0.99	17	1.60
16	20	119	19.70	201	18.91
16	21	36	5.96	66	6.21
16	22	2	0.33	3	0.28
16	134	1	0.17	1	0.09
17	17	6	0.99	3	0.28
17	18	10	1.66	21	1.98
17	19	0	0	4	0.38
17	20	16	2.65	14	1.32
17	21	11	1.82	13	1.22
18	18	21	3.48	30	2.82
18	19	7	1.16	5	0.47
18	20	46	7.62	75	7.06
18	21	11	1.82	27	2.54
18	22	2	0.33	2	0.19
18	132	0	0	1	0.09
19	19	1	0.17	1	0.09
19	20	1	0.17	9	0.85
20	20	34	5.63	58	5.46
20	21	20	3.31	25	2.35
20	22	0	0	3	0.28
21	21	3	0.50	5	0.47
21	22	0	0	1	0.09
		604	100	1063	100

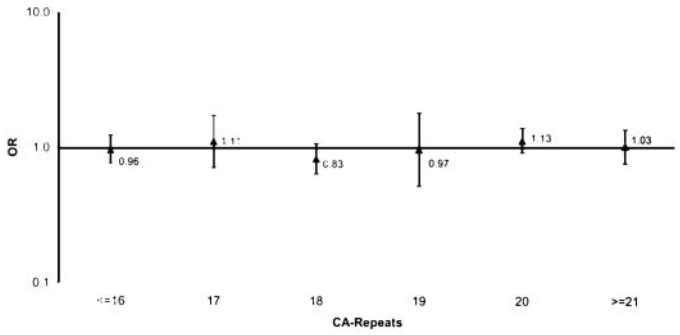

Fig. 1. Odds ratios of breast cancer associated with the longer allele. Odds ratios and the 95% confidence intervals are presented for those having at least one long allele (designated by the CA repeat lengths on the *X* axis), considering the longer of the two alleles for each subject.

Association of the Allelic Length of the *egfr* CA SSR with Breast Cancer Risk. The exploratory analysis did not yield significantly increased ORs associated with carrying a shorter or longer of the two alleles at a specific cutpoint (Figs. 1 and 2).

The *egfr* genotype was compared between cases and controls, applying 17, 18, or 19 CA repeats as the cutoff value between long and short alleles. Using women carrying two short alleles of less than 17, 18, and 19 CA repeats as reference group, respectively; breast cancer risk was not found to be increased in women carrying one or two long alleles. ORs were 1.19 for carriers of one ≥19 CA allele and 1.13 for carriers of two ≥19 CA alleles. Similar results were obtained using 17 and 18 CA repeats as cutoffs for group classification (data not shown).

Effect Modification of Long *egfr* CA SSR Alleles. We analyzed the data for a modification of the effect of the length of the *egfr* CA SSR alleles on breast cancer risk by family history. Breast cancer patients with family history of breast cancer in a first-degree relative were more frequently carriers of two alleles of ≥19 CA repeats than patients without family history (17.1% and 8.7%, respectively). Merely 3.4% of the controls with a positive family history carried two long alleles (Table 3). The presence of two long alleles *egfr* CA SSR was associated with an elevated OR of 10.4 (95% CI, 1.8–58.7) among women with a first degree family history of breast cancer, but not among women without such a family history (OR, 1.8; 95% CI, 0.7–4.6), providing strong evidence for differential effects of the *egfr* CA SSR polymorphism on disease risk by a first-degree family history ($P = 0.015$ for interaction). Case-only analysis confirmed these results, albeit with borderline statistical significance. A much weaker differential effect of allele length on risk of breast cancer by family history was observed at a cutoff of ≥18 CA repeats for the longer allele ($P = 0.0047$), but not at a cutoff of ≥17 CA repeats for the longer allele (data not shown).

Following the hypothesis that a longer CA repeat is more often associated with *egfr* mutation by environmental factors, we investigated a possible effect modification by dietary risk factors. In the evaluation of the original study group without considering genetic factors (13), breast cancer risk increased with a higher consumption of red meat (P for trend = 0.016), and women with the highest consumption level had an 85% elevated breast cancer risk when compared with the lowest quartile (OR, 1.85, CI, 1.23–2.78). On the other hand, risk was inversely associated with vegetable consumption levels (P for trend = 0.034), and the OR for the fourth quartile of vegetable intake compared with the first quartile was 0.64 (95% CI, 0.43–0.96).

For red meat intake, the analysis of dietary factors stratified by the *egfr* genotype yielded a significant trend in breast cancer risk with increasing consumption for both carriers of two longer alleles and carriers of two shorter alleles. Again, the most pronounced effect was observed at a cutoff of ≥19 CA repeats for the longer allele and ≤18 CA for the shorter allele; therefore, these results will be presented (Table 4; P for trend, 0.03 and 0.02, respectively). The significant trend for homozygous carriers of either allele ($P = 0.03$ and 0.02, respectively) was predominantly due to the high ORs associated with the highest consumption quartile, although stronger for carriers of two

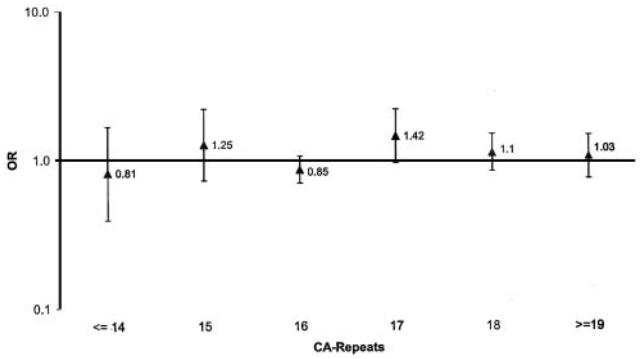

Fig. 2. Odds ratios of breast cancer associated with the shorter allele. Odds ratios and the 95% confidence intervals are presented for those having at least one short allele (designated by the CA repeat lengths on the *X* axis), considering the shorter of the two alleles for each subject.

Table 3 *egfr genotype frequencies and odds ratios for breast cancer overall and by family history of breast cancer*

			egfr genotypes[a]			
			ss	sl	ll	
		N	*N* (%)	*N* (%)	*N* (%)	*P*[b]
Overall	Controls	1063	516 (48.8)	445 (41.9)	102 (9.6)	0.64
	Cases	604	279 (46.2)	266 (44.0)	59 (9.8)	
OR[c] (95% CI)			1.0	1.19 (0.96–1.48)	1.13 (0.78–1.62)	
By family history						
No family history	Controls	1004	480 (47.8)	424 (42.2)	100 (10.0)	0.43
	Cases	528	242 (45.8)	240 (45.4)	46 (8.7)	
OR[d] (95% CI)			1.0	1.17 (0.93–1.47)	0.95 (0.65–1.40)	
Positive family history	Controls	59	36 (61.0)	21 (35.6)	2 (3.4)	0.04
	Cases	76	37 (48.7)	26 (34.2)	13 (17.1)	
OR[d] (95% CI)			1.0	1.79 (0.70–4.57)	10.43 (1.85–58.70)	
Case-only analysis	No family history	528	242 (45.8)	240 (45.5)	46 (8.7)	0.03
	Positive family history	76	37 (48.7)	26 (34.2)	13 (17.1)	
OR[e] (95% CI)			1.0	0.72 (0.42–1.24)	1.83 (0.90–3.72)	

[a] Genotypes according to a short (≤18 CA repeats) and a long (≥19 CA repeats) allele: ss (short/short), sl (short/long), ll (long/long).
[b] *P* of the χ^2 test of the genotype distribution.
[c] Adjusted for number of full-term pregnancies (0, 1–2, ≥3), age at menarche (≤12, ≥13 years), duration of breastfeeding (continuous), menopausal status, and family history, alcohol consumption (0, ≤18, or ≥19 g/day). OR, odds ratio; CI, confidence interval.
[d] Adjusted for number of full-term pregnancies (0, 1–2, ≥3), age at menarche (≤12, ≥13 years), duration of breastfeeding (continuous), menopausal status, alcohol consumption (0, ≤18, or ≥19 g/day).
[e] Case-only analysis. Odds ratio adjusted for number of full-term pregnancies (0, 1–2, ≥3), age at menarche (≤12, ≥13 years), duration of breastfeeding (continuous), menopausal status, alcohol consumption (0, ≤18, or ≥19 g/day).

≥19 CA repeat alleles (OR, 10.68; 95% CI, 1.57–72.58) than carriers of two ≤18 CA repeat alleles (OR, 1.86; 95% CI, 1.06–3.27; Table 4). An association was not found for the heterozygous carriers of one long and one short allele (*P* for trend, 0.85). Interaction between the *egfr* polymorphism and red meat consumption was tested by scoring the red meat consumption categories but was not found to be significant.

Similarly, breast cancer risk was reduced with increasing vegetable consumption for both carriers of ≥19 CA alleles and carriers of two ≤18 CA alleles (*P* for trend, 0.07 and 0.07, respectively), but not for the heterozygous carriers ($P = 0.95$). Again, the reduction in risk was stronger for carriers of two long alleles than for carriers of two short alleles and was of borderline significance for the highest consumption quartile, with ORs of 0.07 (95% CI, 0.004–1.07) and 0.56 (95% CI, 0.32–1.00), respectively. We found no statistically significant effect modification by *egfr* genotype of the breast cancer risk associated with vegetable intake when testing for interaction by scoring the vegetable consumption categories.

Discussion

In a population-based case-control study in Germany, we tested the hypothesis that the CA SSR polymorphism in intron 1 of *egfr* affects breast cancer risk. Overall, we did not find an association of allelic length of the *egfr* CA SSR polymorphism with breast cancer risk. However, we observed that the presence of two long alleles, particularly when defined as ≥19 CA repeats, of the *egfr* CA SSR was associated with a significantly elevated OR among women with a first-degree family history of breast cancer, but not among women without such a family history. The differential effects of allelic length

Table 4 *Association of dietary factors with breast cancer risk by egfr genotypes*

	egfr genotypes[a]											
	ll				sl				ss			
Quartiles	Controls (*n* = 58)	Cases (*n* = 27)	OR[b] (95% CI)	Trend[c] *P*	Controls (*n* = 290)	Cases (*n* = 132)	OR[b] (95% CI)	Trend[c] *P*	Controls (*n* = 341)	Cases (*n* = 152)	OR[b] (95% CI)	Trend[c] *P*
Vegetables (in g/day)				0.07				0.85				0.07
1–177	9 (15.5%)	5 (18.5%)	1.0		79 (27.2%)	30 (22.7%)	1.0		84 (24.6%)	47 (30.9%)	1.0	
178–254	14 (24.1%)	4 (14.8%)	0.32 (0.03–3.60)		71 (24.5%)	37 (28.0%)	1.25 (0.69–2.29)		88 (25.8%)	35 (23.0%)	0.85 (0.48–1.48)	
255–372	19 (32.8%)	7 (25.9%)	0.47 (0.05–4.18)		80 (27.6%)	39 (29.6%)	1.14 (0.63–2.08)		78 (22.9%)	39 (25.7%)	0.88 (0.50–1.56)	
≥373	16 (27.6%)	11 (40.7%)	0.07 (0.004–1.07)		60 (20.7%)	26 (19.7%)	1.09 (0.57–2.09)		91 (26.7%)	31 (20.4%)	0.56 (0.32–1.00)	
Red meat (in g/day)				0.03				0.95				0.02
Missing	1 (0.02%)				8 (2.76%)	3 (2.27%)			14 (4.11%)	3 (1.97%)		
1–21	24 (41.4%)	6 (22.2%)	1.0		88 (30.3%)	39 (29.5%)	1.0		109 (32.0)	47 (30.9%)	1.0	
22–39	9 (15.5%)	3 (11.1%)	1.20 (0.12–12.40)		73 (25.2%)	33 (25%)	1.10 (0.61–1.96)		105 (30.8)	29 (19.1%)	0.71 (0.41–1.23)	
40–64	14 (24.1%)	4 (14.8%)	1.30 (0.16–10.58)		72 (24.8%)	30 (22.7%)	0.97 (0.54–1.74)		61 (17.9%)	32 (21.1%)	1.39 (0.78–2.50)	
≥65	10 (17.3%)	14 (51.9%)	10.68 (1.57–72.58)		49 (16.9%)	27 (20.5%)	1.07 (0.57–2.05)		52 (15.2%)	41 (26.9%)	1.86 (1.06–3.27)	

[a] Genotype according to a short (≤18 CA repeats) and a long (≥19 CA repeats) allele: ss (short/short); sl (short/long); ll (long/long).
[b] Adjusted for number of full-term pregnancies (0, 1–2, ≥3), age at menarche (≤12 or ≥13 years), duration of breastfeeding (continuous), menopausal status, and family history, alcohol consumption (0, ≤18, or ≥19 g/day). OR, odds ratio; CI, confidence interval.
[c] *P* for test of trend of the odds ratios for the quartiles of consumption scored as ordinal variables 1 to 4.

of the *egfr* CA SSR on disease risk by a first-degree family history were statistically significant, even after a Bonferroni adjustment for multiple comparisons using three different definitions for the longer allele ($P = 0.015$ for interaction). This was confirmed in a case-only analysis of the patients. To our knowledge, this is the first report of a possible relevance of this common intronic polymorphism for familial breast cancer.

Furthermore, our data suggested that the risk increase associated with high red meat consumption and the protective effect of high vegetable intake may be additionally enhanced among carriers of two longer alleles, particularly for alleles of ≥19 CA repeats, in comparison with the total population (13).

We cannot exclude that selection bias may have contributed to the study results because not all women had provided blood samples. However, the distribution of all relevant epidemiological variables did not differ between those with DNA samples and those without DNA samples. Dietary information was not available from all study subjects of the one study region involved. Here again, we found little difference between the original study population and the subgroup providing dietary information with respect to relevant epidemiological variables, so that the effect of selection bias may be negligible.

Recall bias of family history of breast cancer would have to be differential for noncarriers and carriers of the long allele to result in the observed differential effect by family history, and this seems rather unlikely. Recall bias may be a source of measurement error of dietary habits. Our previously reported results of an inverse association with vegetable intake and a positive association with high red meat consumption are compatible with findings of other studies. The dietary intake levels observed in our study population are comparable with the intake data described in other large German studies (German European Prospective Investigation into Nutrition and Cancer centers) by means of the same validated assessment tool (14, 15). Therefore, we have some assurance that the collected data are valid.

As mentioned above, the presence of two long alleles of the *egfr* CA SSR (most pronounced for ≥19 CA repeats) was strongly associated with an increased risk among women with a first-degree family history of breast cancer in this study. In experimental studies, we found that the length of the inherited polymorphic CA SSR in the 5′-regulatory sequence of intron 1 of the *egfr* gene, on one hand, was inversely associated with its transcriptional activity and, on the other hand, directly associated with the likelihood of AI (9, 10). The second observation was further supported by interethnic studies (16). Japanese women presented very homogenously with predominantly longer alleles containing ≥19 CA dinucleotides, as compared with Caucasians. Moreover, Japanese breast cancers had a significantly higher prevalence of AI, indicating gene amplifications, as compared with their German counterparts. Furthermore, it has already been shown that the number of repeats itself affects the mutation rate of nucleotide repeats (17).

We can also postulate that other genetic factors and/or lifestyle factors in women with a family history of breast cancer may render a higher probability of mutation in the longer allele. This is further supported by our most recent data that a putative fragile site is located close to the *egfr* gene and that these AIs were already present in the surrounding normal, nontumorous breast tissue (10). Besides, EGFR expression underlies further complex regulation mechanisms influenced by extrinsic factors. These factors appear to predominantly concern lifestyle, especially dietary factors, which are likely to be shared in a family. This is known, for instance, for genistein derived from soy, which reduced EGFR expression in rats (6). We did not have adequate statistical power to show how dietary factors may have contributed to the increased risk associated with the presence of two long alleles of the *egfr* CA SSR in women with a positive family history. In the whole study group, however, the risk increase associated with high red meat consumption and the protective effect of high vegetable intake were most pronounced among carriers of two ≥19 CA alleles. Heterocyclic aromatic amines are found in meat cooked at high temperatures for long duration. They could induce mutations because, after metabolic activation, heterocyclic aromatic amines can react directly with DNA and efficiently induce mitotic recombination (18). On the other hand, protection from DNA damage can be mediated by antioxidants in vegetables (19). Therefore, longer alleles of the *egfr* CA SSR may contribute to a higher risk of breast cancer caused by environmental factors. We speculate that in the case of two longer alleles of the CA SSR I, a mutant clone would become dominant after an undetermined number of cell divisions due to natural selection acting to increase the frequency of advantageous alleles in breast glandular cells. However, the impact of both the increase in transcription activity with a low number of CA repeats and the increase of likelihood for an AI at the *egfr* locus leading to EGFR overexpression with a high number of CA repeats should reach an equilibrium for the genotypes with a long allele and a short allele. In fact, no significant association between food pattern and cancer risk could be determined for this subgroup in our study. In the case of the two shorter alleles, extrinsic factors may play only a secondary role because of the inherent higher transcriptional activity and lower likelihood for AI.

In conclusion, our data support the assumption that the *egfr* CA SSR polymorphism is a causal factor for some cases of familial breast cancer and that its effect may be modulated by dietary factors. Additional epidemiological studies are needed to support this association, and further experimental work is needed to define the molecular mechanisms leading to AIs at the *egfr* CA SSR I locus and to explain the intriguing network of exogenous and endogenous factors causing breast cancer.

References

1. Wooster, R., and Weber, B. L. Breast and ovarian cancer. N. Engl. J. Med., *348:* 2339–2347, 2003.
2. Narod, S. A. Modifiers of risk of hereditary breast and ovarian cancer. Nat. Rev. Cancer, *2:* 113–123, 2002.
3. Chang-Claude, J. Inherited genetic susceptibility to breast cancer. *In:* A. B. Miller, H. Bartsch, P. Boffeta, L. Dragsted and H. Vainio (eds.), Biomarkers in Cancer Prevention. IARC Scientific Publication No. 154, pp. 177–190. Lyon, France: IARC, 2001.
4. Houlston, R. S., and Tomlinson, I. P. Detecting low penetrance genes in cancer: the way ahead. J. Med. Genet., *37:* 161–167, 2000.
5. Dunning, A. M., Healey, C. S., Pharoah, P. D., Teare, M. D., Ponder, B. A., and Easton, D. F. A systematic review of genetic polymorphisms and breast cancer risk. Cancer Epidemiol. Biomark. Prev., *8:* 843–854, 1999.
6. Brown, N. M., Wang, J., Cotreno, M. S., Zhao, Y. X., and Lamartiniere, C. A. Prepubertal genistein treatment modulates TGF-α, EGF and EGF-receptor mRNAs and proteins in the rat mammary gland. Mol. Cell. Endocrinol., *144:* 149–165, 1998.
7. Brandt, R., Eisenbrandt, R., Leenders, F., Zschiesche, W., Binas, B., Juergensen, C., and Theuring, F. Mammary gland specific hEGF receptor transgene expression induces neoplasia and inhibits differentiation. Oncogene, *20:* 2129–2137, 2000.
8. Gebhardt, F., Zänker, K. S., and Brandt, B. Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. J. Biol. Chem., *274:* 13176–13180, 1999.
9. Buerger, H., Gebhardt, F., Beckmann, A., Hutmacher, K., Simon, R., Lelle, R. J., Boecker, W., and Brandt, B. Length and loss of heterozygosity of an intron 1 polymorphic sequence of *egfr* is related to cytogenetic alterations and EGFR-expression. Cancer Res., *60:* 854–857, 2000.
10. Tidow, N., Boecker, A., Schmidt, H., Agelopoulos, K., Boecker, W., Buerger, H., and Brandt, B. Distinct amplification of an untranslated regulatory sequence in the *egfr* gene contributes to early steps in breast cancer development. Cancer Res., *63:* 1172–1178, 2003.
11. Chang-Claude, J., Eby, N., Kiechke, M., Bastert, G., and Becher, H. Breastfeeding and breast cancer risk by age 50 among women in Germany. Cancer Causes Control, *11:* 687–695, 2000.

12. Liu, W., Innocenti, F., Chen, P., Das, S., Cook, E. H., Jr., and Ratain, M. J. Interethnic difference in the allelic distribution of human epidermal growth factor receptor intron 1 polymorphism. Clin. Cancer Res., *9:* 1009–1012, 2003.
13. Hermann, S., Linseisen, J., and Chang-Claude, J. Nutrition and breast cancer risk by age 50: a population-based case-control study in Germany. Nutr. Cancer, *44:* 23–34, 2002.
14. Bohlscheid-Thomas, S., Hoting, I., and Boeing, H., Wahrendorf: reproducibility and relative validity of food group intake in a food frequency questionnaire developed for the German part of EPIC Project. Int. J. Epidemiol., *26:* S59–S71, 1997.
15. Bohlscheid-Thomas, S., Hoting, I., and Boeing, H. Wahrendorf: reproducibility and relative validity of energy and macronutrient intake of a food frequency questionnaire developed for the German part of EPIC Project. Int. J. Epidemiol., *26:* S71–S81, 1997.
16. Bürger, H., Packeisen, J., Kersting, C., Isola, J., Böcker, W., Böcker, A., Tidow, N., Nakachi, K., Bielwaski, K., Yatabi, Y., and Brandt, B. Allelic length of a CA dinucleotide repeat in the *egfr* gene correlates with the frequency of amplifications of this sequence. First results of an interethnic breast cancer study. J. Pathol., in press, 2003.
17. Eckart, K. A., Yan, G., and Hile, S. E. Mutation rate and specificity analysis of tetranucleotide microsatellite DNA alleles in somatic cells. Mol. Carcinog., *34:* 140–150, 2002.
18. Paladino, G., Weibel, B., and Sengstag, C. Heterocyclic aromatic amines efficiently induce mitotic recombination in metabolically competent *Saccharomyces cerevisiae* strains. Carcinogenesis (Lond.), *20:* 2143–2152, 1999.
19. Wu, A. H., Wan, P., Hankin, J., Tseng, C. C., Yu, M. C., and Pike, M. C. Adolescent and adult soy intake and risk of breast cancer in Asian-Americans. Carcinogenesis (Lond.), *23:* 1491–1496, 2002.

Modification of Breast Cancer Risk in Young Women by a Polymorphic Sequence in the *egfr* Gene

Burkhard Brandt, Silke Hermann, Kurt Straif, et al.

Cancer Res 2004;64:7-12.