available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Single nucleotide polymorphisms in chromosomal instability genes and risk and clinical outcome of breast cancer: A Swedish prospective case-control study

Annika Brendle[a], *Andreas Brandt*[a], *Robert Johansson*[b], *Kerstin Enquist*[c], *Göran Hallmans*[c], *Kari Hemminki*[a,d], *Per Lenner*[b], *Asta Försti*[a,d,*]

[a]*Division of Molecular Genetic Epidemiology C050, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany*

[b]*Department of Oncology, Norrlands University Hospital, Umeå, Sweden*

[c]*Department of Public Health and Clinical Medicine/Nutritional Research, Umeå University, Sweden*

[d]*Center for Family and Community Medicine, Karolinska Institute, Huddinge, Sweden*

ARTICLE INFO

Article history:
Received 9 July 2008
Received in revised form 26 September 2008
Accepted 2 October 2008
Available online 12 November 2008

Keywords:
Breast cancer
Chromosomal instability
Prognosis
Single nucleotide polymorphism
Susceptibility

ABSTRACT

Chromosomal instability (CIN) is a major characteristic of many cancers. We investigated whether putatively functional single nucleotide polymorphisms (SNPs) in genes related to CIN (*CENPF*, *ESPL1*, *NEK2*, *PTTG1*, *ZWILCH*, *ZWINT*) affect breast cancer (BC) risk and clinical outcome in a Swedish cohort of 749 incident BC cases with detailed clinical data and up to 15 years of follow-up and 1493 matched controls. As a main observation, carriers of the A allele of the *CENPF* SNP rs438034 had a worse BC-specific survival compared to the wild type genotype GG carriers (hazard ratio (HR) 2.65, 95% confidence interval (CI) 1.19–5.90), although they were less likely to have regional lymph node metastases (odds ratio (OR) 0.71, 95% CI 0.51–1.01) and tumours of stage II–IV (OR 0.73, 95% CI 0.54–0.99). As there is increasing evidence that CENPF is associated with poor prognosis in patients with primary BC, further independent studies are needed to clarify the importance of genetic variation in the *CENPF* gene in the clinic.

1. Introduction

Chromosomal instability (CIN) leading to an aberrant chromosome number has been recognised as a hallmark of human cancer.[1–3] However, the molecular mechanisms underlying CIN are poorly understood.[4–6] They include defects in chromosomal segregation, telomere stability, DNA damage response as well as defects in checkpoints preventing the reproduction of abnormal cells. CIN is supposed to contribute to tumourigenesis through accelerating the rate of loss of heterozygosity of tumour suppressor genes or the amplification of oncogenes.[2,3,7,8] Additionally, alterations in the gene expression caused by CIN can increase cell proliferation or decrease cell death, which are essential processes for tumour development and progression.[6,7]

Carter and colleagues have described a gene expression signature of aneuploidy as a significant predictor of poor clinical outcome in tumours of the breast, lung and brain.[9] A ranking of genes correlating with the high levels of CIN contains several key regulators of replication and segregation of chromosomes, including the genes *ESPL1* (extra spindle poles-like 1, separase), *NEK2* (never in mitosis gene A-related

* *Corresponding author:* Address: Division of Molecular Genetic Epidemiology C050, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany. Tel.: +49 6221 421803; fax: +49 6221 421810.
E-mail address: a.foersti@dkfz.de (A. Försti).

0959-8049/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2008.10.001

kinase 2), *PTTG1* (pituitary tumour-transforming gene 1, securin) and *ZWINT* (ZW10 interactor). Moreover, Perez de Castro and colleagues have described genes known as mitotic regulators which are mutated and over-expressed in various CIN cancers.[10] Among them are the *CENPF* (centromere protein F), *ESPL1*, *NEK2*, *PTTG1*, *ZWILCH* and *ZWINT* genes, respectively. Importantly, in those tumour types where CIN is present, there is a significant correlation between the CIN phenotype and poor prognosis, suggesting that chromosome imbalances specifically contribute to aggressive or metastatic cancer.[9,10] In another study, the association between gene expression and breast tumour copy number aberrations in genes belonging to one or more of the categories DNA replication, DNA damage/repair, cell cycle, mitosis, centrosome and telomere has been assessed.[11] In particular, Fridlyand and colleagues observed that many genes which maintain genome integrity are over-expressed and direct targets of E2F.[11] This fact supports the hypothesis that the deregulation of the Rb/E2F pathway is a major contributor to CIN in breast tumours.

Traditional prognostic factors, such as tumour size, grade and lymph node metastasis status, are still the most important prognostic factors for long-term survival in breast cancer (BC), although recent gene expression profiling studies have provided a new tool for the prediction of BC prognosis.[12] Also inherited genetic variation can affect both individual's risk of BC[13] and survival in BC.[14,15] As CIN can contribute to both tumourigenesis[2] and cancer prognosis,[9,10] genetic variation in CIN genes may affect BC susceptibility and prognosis. In our previous study we investigated the effect of nine putatively functional single nucleotide polymorphisms (SNPs) in six major spindle checkpoint genes (BUB1B, *BUB3*, *CENPE*, *MAD2L1*, *MAD2L2*, *TTK*) on familial BC risk in a German population.[16] Our results of no association between the studied SNPs and the risk of BC suggest that genetic variation in the mitotic checkpoint genes is not functionally relevant in the assessment of BC and rather may have an effect on cancer progression and prognosis.

For the present study, we selected six genes involved in different steps of mitosis that have been shown to correlate with CIN: *CENPF*, *ESPL1*, *NEK2*, *PTTG1*, *ZWILCH*, *ZWINT*.[9–11] We screened the regulatory and the coding regions for putatively functional SNPs and selected six SNPs for further analysis by genotyping. We used a large Swedish study population with detailed clinical data and a follow-up time of up to 15 years.

2. Materials and methods

2.1. Study population

The analyses were performed on DNA from 749 Swedish BC cases together with 1493 controls. The cases with the age and gender matched controls were drawn from the population-based Västerbotten intervention project and the mammary screening project, which contain blood samples collected between January 1990 and January 2001 from an ethnically homogenous population living in a geographically defined region in North Sweden.[17] Prospective cases were identified from the cohorts by record linkage to the regional cancer registry. The controls were selected from the same cohort as the corresponding case. They were matched with the case by age at baseline (+/– 6 months) and the time of sampling (+/– 2 months). The controls had to be alive at the time of diagnosis of the corresponding case and without any previous cancer diagnosis, except carcinoma in situ of cervix uteri. All participants gave an informed consent to the use of their samples for research purpose. The blood samples were stored at –80 °C until the time of sample selection and DNA isolation for genotyping analyses. After dividing the samples randomly on the 96-well plates, whole genome amplification (WGA) was performed with the GenomiPhi DNA amplification kit (Amersham Biosciences, Piscataway, NJ) according to the method described by Wong and colleagues.[18] using Φ29 DNA polymerase as described by Paez and colleagues.[19] The amplification

Table 1 – Characteristics of the Swedish breast cancer samples at the time of diagnosis.

Characteristics	Breast cancer patients, *n* (%)
Age at diagnosis, mean (range, SD)	58.1 (30.9–76.1, 8.7)
Oestrogen receptor (ER)	
Positive	200 (26.7)
Negative	80 (10.7)
Missing data	469 (62.6)
Progesterone receptor (PR)	
Positive	177 (23.6)
Negative	97 (13.0)
Missing data	475 (63.4)
Hormone receptor combination	
ER+/PR+	164 (21.9)
ER+/PR–	31 (4.1)
ER–/PR+	13 (1.7)
ER–/PR–	66 (8.8)
Missing data	475 (63.5)
Tumour size in cm	
⩽2 cm	486 (64.9)
>2 cm	214 (28.6)
Missing data	49 (6.5)
Histologic grade	
1	155 (20.7)
2	342 (45.7)
3	215 (28.7)
Missing data	37 (4.9)
Regional lymph node metastasis	
Negative	450 (60.1)
Positive	209 (27.9)
Missing data	90 (12.0)
Stage at diagnosis	
0	2 (0.3)
I	392 (52.3)
II	310 (41.4)
III	25 (3.3)
IV	14 (1.9)
Missing data	6 (0.8)
Distant metastasis	
Negative	728 (97.2)
Positive	13 (1.7)
Missing data	8 (1.1)

results were controlled by genotyping two frequent SNPs using TaqMan allelic discrimination assays. The SNPs had been genotyped previously using the original genomic DNA. Less than 0.10% of the WGA genotypes could not be determined or did not agree with the data of the genomic DNA. Genotyping was performed blinded by the case-control status of each sample using WGA DNA.

Clinical data for the BC cases were retrieved from the registry managed by the Northern Sweden Breast Cancer Group (Table 1). Follow-up was performed until 26th April 2007. Information about the date of death was collected from the Swedish population register with a BC-specific follow-up until 31st December 2004. The median follow-up time for BC-specific survival was 4.7 years. The study was approved by the ethical committee of Karolinska Institute Syd and Umeå University.

2.2. SNP screening by sequencing

Genomic DNA from blood from a randomly chosen set of 32 Swedish BC samples was used to screen by sequencing the promoters, up to –1000 bp from transcription start site, reported non-synonymous coding SNPs (NCBI dbSNP) and the 3′UTRs in six genes related to CIN (*CENPF*, *ESPL1*, *NEK2*, *PTTG1*, *ZWILCH* and *ZWINT*). Primer sequences are available from the corresponding author on request.

To identify putative transcription factor binding sites we used the TESS – Transcription Element Search System database (http://www.cbil.upenn.edu/tess/). The PolyPhen database (http://genetics.bwh.harvard.edu/pph/) was used to predict a possible impact of an amino acid substitution on the structure and function of human proteins. Putative miRNA binding sites in the 3′UTRs were determined using the online available tools microInspector (http://mirna.imbb.forth.gr/microinspector/) and TargetScan (http://www.targetscan.org/).

2.3. TaqMan allelic discrimination

The TaqMan allelic discrimination method was used to genotype the six selected SNPs as described previously.[20] The TaqMan assays (Assay-on-demand) were ordered from Applied Biosystems (Foster City, CA, USA). Call rates for genotyping ranged between 96.9% and 99.1%. The large sample set for genotyping contained 104 duplicate samples (four duplicates on each 96-well plate) which represent approximately 5% of all samples. The genotypes of the duplicates showed 100% concordance.

2.4. Statistical analysis

The observed genotype frequencies in the controls were tested for Hardy–Weinberg equilibrium (HWE) using the χ^2 test. Statistical significance for the differences in the genotype frequencies between the BC cases and the controls was determined by the Wald χ^2 test of heterogeneity with two degrees of freedom. The linkage disequilibrium (LD) between the SNPs was evaluated with the Haploview software.[21] Odds ratios (ORs) and 95% confidence intervals (95% CIs) for association between genotypes and BC risk and tumour characteristics were calculated by logistic regression (PROC LOGISTIC, SAS Version 9.1; SAS Institute, Cary, NC). For a polymorphism

Table 2 – Verified SNPs in the promoter, the coding region and the 3′UTR of genes related to CIN. Verification was done by sequencing 32 randomly selected BC samples. SNPs included in further analyses are typed in bold.

Gene	SNP reference ID	Variation	wt/wt	wt/var	var/var	Region	Putative effect
CENPF	rs3748691	[C/G]	26	5	0	5′UTR	Putative transcription factor binding sites
	rs2070065	[C/G]	26	5	0	Exon 11	Benign amino acid change [His/Gln]
	rs2666839	[G/A]	22	5	0	Exon 12	Benign amino acid change [Ala/Thr]
	rs335524	[G/A]	17	2	10	Exon 16	Unknown amino acid change [Arg/Gln]
	rs438034	**[G/A]**	**10**	**16**	**5**	**Exon 18**	Unknown amino acid change [Arg/Gln]
ESPL1	rs1976938	[T/C]	6	5	1	Promoter	Putative transcription factor binding sites
	rs6580941	**[C/T]**	**11**	**9**	**1**	**Promoter**	Putative transcription factor binding sites
	rs6580942	[A/C]	13	12	3	Exon 1	Damaging amino acid change [Ala/Asp]
	rs1318648	[A/C]	12	15	4	Exon 7	Damaging amino acid change [Ser/Arg]
	rs1056692	[A/G]	13	13	3	3′UTR	No predicted effect
NEK2	rs701927	[G/T]	12	7	1	Promoter	No predicted effect
	rs701928	**[T/A]**	**15**	**11**	**2**	**Promoter**	Putative transcription factor binding sites
	rs10429965	[A/G]	21	7	1	Exon 7	Benign amino acid change [Asn/Ser]
	rs894852	[C/T]	17	11	1	3′UTR	No predicted effect
PTTG1	**rs1862392**	**[T/A]**	**15**	**13**	**3**	**Promoter**	Putative transcription factor binding sites
	rs1862391	[A/C]	14	13	3	Promoter	No predicted effect, in LD with rs1862392
	rs2961952	[G/A]	20	8	2	3′UTR	No predicted effect
ZWILCH	**rs3087660**	**[A/G]**	**11**	**13**	**4**	**5′UTR**	No predicted effect
ZWINT	new	[C/A]	13	14	3	5′UTR: –819 bp from ATG	Putative transcription factor binding sites
	rs2241666	**[G/A]**	**10**	**14**	**7**	**Exon 6**	Benign amino acid change [Arg/Gly]

with a variant allele frequency between 25% and 45%, the study had greater than 90% power to detect an OR of 1.40 at a significance level of 0.05 (PS – software for power and sample size calculation, http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize). The survival curves for BC-specific survival were derived by the Kaplan–Meier method (PROC LIFETEST, SAS Version 9.1, SAS Institute). The relative risk of death by BC was estimated as hazard ratio (HR) using Cox regression (PROC PHREG, SAS Version 9.1, SAS Institute). Censoring events were death by another cause than BC, moving out of the study and 31st December 2004. The HRs were also calculated within subgroups of cases with a similar manifestation of a clinical factor. Furthermore, the hazard ratio for the overall risk was adjusted for the clinical factors (oestrogen receptor status, progesterone receptor status, tumour size, lymph node metastasis, histologic grade) to determine the value of the genotypes as an independent prognostic marker.

3. Results

3.1. *Selection of SNPs*

A randomly chosen set of 32 Swedish BC samples was screened to verify SNPs in the chosen genes *CENPF*, *ESPL1*, *NEK2*, *PTTG1*, *ZWILCH* and *ZWINT*. The promoter regions and the 3′UTRs of all genes were sequenced, as well as the reported non-synonymous coding SNPs (NCBI dbSNP) in the *CENPF*, *ESPL1*, *NEK2* and *ZWINT* genes, respectively. Table 2 shows the SNPs verified in the investigated gene regions in our small sample set. In the procedure of selecting SNPs for genotyping also the LD data derived from our sequenced SNPs and from the public HapMap database were taken into account. A total of six SNPs were selected for genotyping in the Swedish study population (*CENPF*: rs438034, *ESPL1*: rs6580941, *NEK2*: rs701928, *PTTG1*: rs1862392, *ZWILCH*: rs3087660, *ZWINT*: rs2241666).

According to the HapMap data, the SNPs in the *CENPF, ESPL1, NEK2* and *PTTG1* gene, respectively, are in high LD forming only one big haploblock (Supplementary Fig. 1A–D). Since this was in compliance with our sequencing results, we selected one SNP in each gene for further investigation. The *ZWILCH* gene is divided into two big haploblocks which are separated by a region of low LD (Supplementary Fig. 1E). Within the sequenced regions we could confirm only one SNP in the 5′UTR located in the first haploblock and selected it for genotyping. In the 5′UTR region of the *ZWINT* gene we detected a new frequent SNP 819 bp upstream of the translation start. However, no TaqMan genotyping assay could be designed for this SNP. Since no other promoter SNPs could be confirmed, and the LD within the *ZWINT* gene was relatively high we selected the non-synonymous coding SNP in exon 6 which was according to our sequencing results in LD with the newly identified SNP ($D' = 1.0$, $r^2 = 0.44$).

Table 3 – Associations of the genotypes of the cases of the CENPF SNP rs438034 with tumour characteristics in the Swedish population.

Tumour characteristics	Genotypes			
	GG	GA	AA	GA + AA
Oestrogen receptors				
Positive	60 (30.3)	103 (52.0)	35 (17.7)	138 (69.7)
Negative	26 (32.9)	43 (43.0)	19 (24.1)	53 (67.1)
OR (95% CI)	1.00	0.76 (0.42–1.39)	1.25 (0.61–2.58)	0.89 (0.51–1.55)
Progesterone receptors				
Positive	51 (29.0)	96 (54.6)	29 (16.4)	125 (71.0)
Negative	33 (34.4)	39 (40.6)	24 (25.0)	63 (65.6)
OR (95% CI)	1.00	0.63 (0.36–1.12)	1.28 (0.64–2.57)	1.09 (0.33–3.58)
Tumour size				
⩽20 mm	157 (32.6)	247 (51.2)	78 (16.2)	325 (67.4)
>20 mm	74 (35.2)	94 (44.8)	42 (20.0)	136 (64.8)
OR (95% CI)	1.00	0.81 (0.56–1.16)	1.14 (0.72–1.82)	0.89 (0.63–1.25)
Histologic grade				
1 + 2	156 (31.8)	256 (52.1)	79 (16.1)	335 (68.2)
3	73 (34.4)	93 (43.7)	47 (22.1)	140 (65.7)
OR (95% CI)	1.00	0.78 (0.54–1.12)	1.27 (0.81–2.01)	0.89 (0.64–1.26)
Regional lymph node metastasis				
Negative	136 (30.6)	231 (51.9)	78 (17.5)	309 (69.4)
Positive	79 (38.2)	90 (43.5)	38 (18.3)	128 (61.8)
OR (95% CI)	1.00	**0.67 (0.46–0.97)**	0.84 (0.52–1.35)	0.71 (0.51–1.01)
Stage				
0 + I	115 (29.4)	213 (54.5)	63 (16.1)	276 (70.6)
II–IV	125 (36.3)	151 (41.5)	68 (19.8)	219 (63.7)
OR (95% CI)	1.00	**0.65 (0.47–0.91)**	0.99 (0.65–1.52)	**0.73 (0.54–0.99)**

Because of missing clinical data, the numbers of cases vary and do not add to 100% of subjects. OR, odds ratio; 95% CI, 95% confidence interval.

3.2. No association of SNPs with BC susceptibility

No differences in the allele or genotype frequencies between the BC cases and the controls were detected (Supplementary Table 1). For the SNP rs3087660 (*ZWILCH*) the controls deviated slightly from the HWE ($P = 0.001$). However, as for the other SNPs, the allele and genotype frequencies were approximately concordant with the ones published for Caucasians in the HapMap database.

3.3. Association of SNPs with breast tumour characteristics at the time of diagnosis

When we stratified the BC cases by the tumour characteristics listed in Table 1, few associations were observed. For the SNP rs438034 in the *CENPF* gene, the carriers of the A allele were less likely to have regional lymph node metastases (OR 0.71, 95% CI 0.51–1.01) and tumours of stage II–IV (OR 0.73, 95% CI 0.54–0.99) than the carriers of the wild type (WT) genotype (Table 3).

For the SNP rs1862392 in the *PTTG1* gene, the carriers of the TA genotype had more often regional lymph node metastasis positive tumours than the carriers of the WT genotype (OR 1.50, 95% CI 1.05–2.14) (Supplementary Table 2). Also the carriers of the AG genotype of the SNP rs3087660 (*ZWILCH*) had more often oestrogen receptor positive tumours than the carriers of the WT genotype (OR 0.52, 95% CI 0.28–0.95) (Supplementary Table 2). However, for the *PTTG1* and the *ZWILCH* SNPs the rare homozygous genotype frequencies did not differ between the cases and the controls.

3.4. Association with BC-specific survival

The survival analysis stratified by the genotypes of the studied SNPs in all cases showed that the carriers of the AA genotype of the SNP rs438034 in the *CENPF* gene had worse BC-specific survival compared to the WT genotype carriers (HR 2.65, 95% CI 1.19–5.90) (Table 4, Fig. 1A). The other investigated SNPs were not associated with survival. For the SNP rs438034 (*CENPF*) we further examined the BC-specific survival in relation to the tumour characteristics (Supplementary Table 3). The survival was worst in women with the GA and AA genotypes, when they had tumours larger than 2 cm (HR 5.11, 95% CI 1.46–17.85; HR 4.73, 95% CI 1.28–17.51) (Fig. 1B). This effect was also observed in the higher stage tumours (HR 2.25, 95% CI 0.99–5.11; HR 2.97, 95% CI 1.24–7.08) (Fig. 1C). In the multivariate analysis, adjusting for all clinical markers of the tumours, the A allele carrier status was significantly associated with survival both when we included (HR 3.41, 95% CI 1.29–9.00) and excluded hormone receptor status (HR 2.48 95% CI 1.19–5.17) from the analysis (Table 5), indicating the value of the *CENPF* SNP rs438034 as an independent prognostic marker.

Table 4 – Breast cancer-specific survival of the selected SNPs in all cases.

Survival analysis				
Gene	Genotype	Breast cancer-specific survival		
		N_{all}	N_{died} (%)	HR (95% CI)
CENPF				
rs438034	GG	234	10 (4.3)	1.00
	GA	351	26 (7.4)	1.64 (0.79–3.40)
	AA	126	15 (11.9)	**2.65 (1.19–5.90)**
ESPL1				
rs6580941	CC	323	18 (5.6)	1.00
	CT	298	22 (7.4)	1.30 (0.70–2.42)
	TT	85	9 (10.6)	1.96 (0.88–4.36)
NEK2				
rs701928	TT	379	24 (6.3)	1.00
	TA	272	23 (8.5)	1.39 (0.78–2.46)
	AA	53	4 (7.8)	1.20 (0.42–3.47)
PTTG1				
rs1862392	TT	315	23 (7.3)	1.00
	TA	313	18 (5.7)	0.83 (0.45–1.53)
	AA	77	8 (11.0)	1.51 (0.68–3.38)
ZWILCH				
rs3087660	AA	238	16 (6.7)	1.00
	AG	330	27 (8.2)	1.18 (0.64–2.20)
	GG	128	4 (3.1)	0.45 (0.15–1.35)
ZWINT				
rs2241666	GG	277	22 (7.9)	1.00
	GA	323	21 (6.5)	0.79 (0.44–1.44)
	AA	112	8 (7.1)	1.02 (0.46–2.30)

HR, hazard ratio; 95% CI, 95% confidence interval.

4. Discussion

Functional polymorphisms, which have an effect on the regulation of gene expression or on the coded protein, can contribute to the differences between individuals in susceptibility to and severity of a disease. Many cancers show CIN, i.e. they frequently lose and gain whole chromosomes during cell divisions. Chromosomal unstable cancers are likely to have a poorer prognosis than diploid cancers and the degree of aneuploidy correlates with the severity of the disease.[3] Although unscheduled cell proliferation is usually promoted by alterations of direct regulators of mitotis, recent studies have identified several mitotic proteins that are differentially expressed in various types of cancers.[9–11] Based on these findings, we selected six genes related to CIN, to evaluate whether genetic variation in these genes affects BC risk or prognosis.

We detected several common SNPs in the promoter, the coding region and the 3′UTR of the investigated genes, many of them having a potential to affect gene expression or protein function. In the *ESPL1* gene, the sequenced promoter SNPs were located within several putative transcription factor binding sites (TESS) and the coding SNPs were predicted to have a damaging effect on the protein structure (Polyphen), thus giving a high biological plausibility for this gene to be involved in BC. According to our screening results and the data available by HapMap 10 of 14 SNPs were in high LD ($D' = 1.0$, $r^2 \geqslant 0.8$). As we observed no effect of the investigated SNP rs6580941 on BC risk and clinical outcome, we can exclude with a high probability any effect of genetic variation in *ESPL1* on BC susceptibility and prognosis.

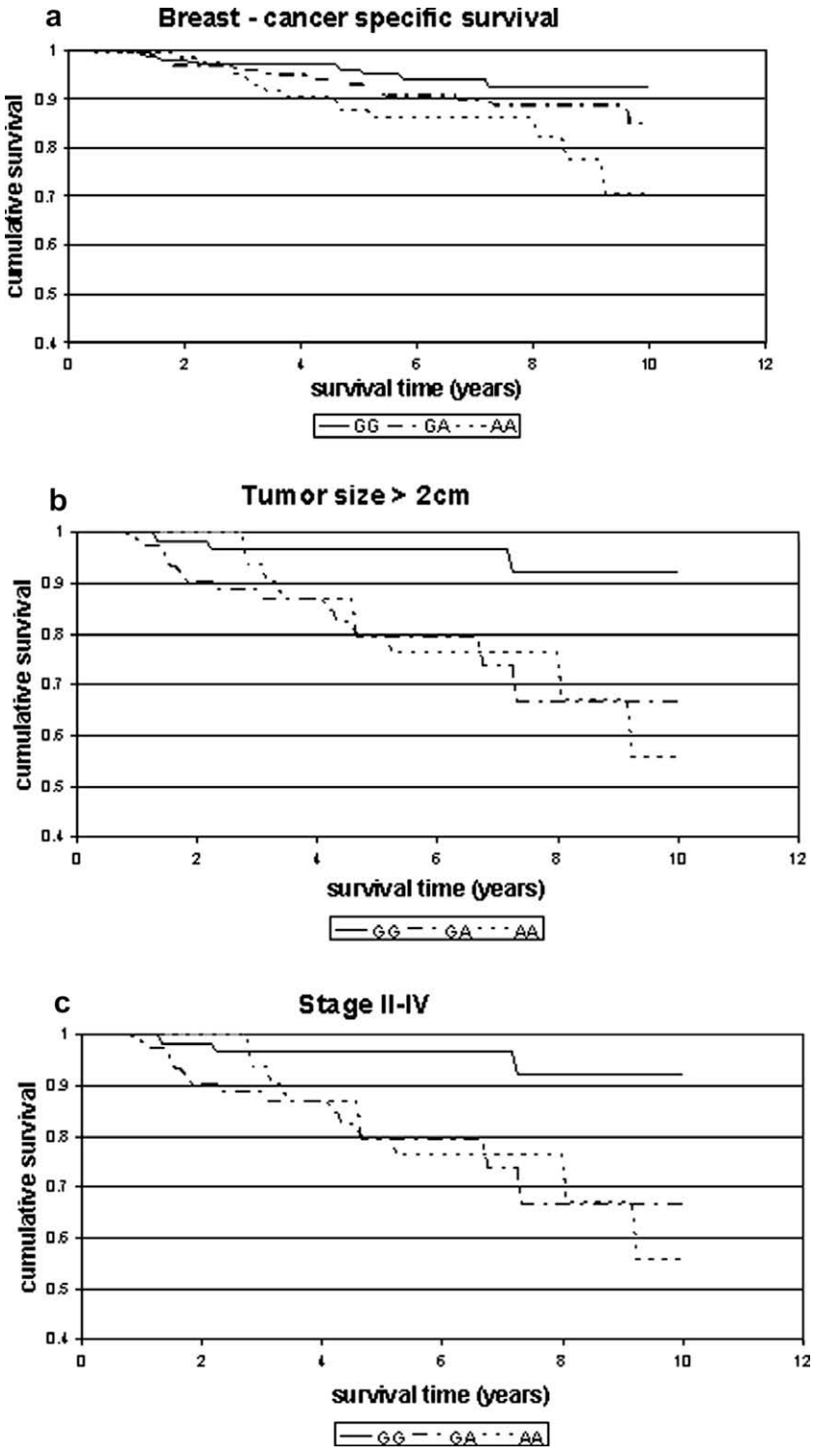

Fig. 1 – Survival of the breast cancer patients according to their genotypes of the SNP rs438034 (*CENPF*) after diagnosis of cancer. (a) Breast cancer-specific survival and (b–c) survival in relation to tumour size and stage II–IV, respectively.

So far, hardly any studies on polymorphisms in the investigated genes in relation to cancer risk and prognosis have been published. Only one multigenic study on BC susceptibility has investigated 3 SNPs in the *PTTG1* gene in a Taiwanese population.[22] It found an increased risk of borderline significance for the variant genotype carriers of rs2910203 in intron 3. According to HapMap this SNP does not exist in the CEPH population. Additionally, the HapMap data show hardly any LD between the SNPs in the *PTTG1* gene in the Asian population. In the CEPH population, both the HapMap data and our own sequencing results indicated high LD between the SNPs. We investigated only one SNP rs1862392 in the promoter region of the *PTTG1* gene and found no association with the risk and clinical outcome of BC. This result suggests that genetic

Table 5 – Multivariate analysis of hazard ratio for the SNP rs438034 (CENPF). The analyses were adjusted for all tumour characteristics at the time of diagnosis, including and excluding hormone receptor status, respectively.

Genotype	N_{all}	N_{died} (%)	HR (95% CI)	*P* value
Adjustment including hormone receptor status[a]				
GG	82	5 (6.1)	1.00	
GA	116	17 (14.6)	3.68 (1.33–10.17)	0.01
AA	49	9 (18.4)	3.01 (0.99–9.12)	0.05
GA+AA	165	26 (15.8)	3.41 (1.29–9.00)	0.01
Adjustment excluding hormone receptor status[b]				
GG	202	9 (4.4)	1.00	
GA	291	24 (8.3)	2.55 (1.17–5.54)	0.02
AA	104	12 (11.5)	2.35 (0.99–5.60)	0.05
GA+AA	395	36 (9.1)	2.48 (1.19–5.17)	0.02

HR, hazard ratio; 95% CI, 95% confidence interval.
a The median follow-up time was 6.5 years.
b The median follow-up time was 4.4 years.

variation in the *PTTG1* gene does not affect BC susceptibility or prognosis in the Caucasian population.

The main observations of our study were apparently contradictory: although the carriers of the A allele of the *CENPF* SNP rs438034 were less likely to have aggressive tumours of stage II–IV with lymph node metastases, they seemed to have worse BC-specific survival than the non-carriers. Additionally, the strongest association between the A allele carrier status and bad survival was observed in patients with aggressive tumours (larger than 2cm and high stage tumours). The multivariate analysis suggested that the *CENPF* SNP rs438034 is an independent prognostic marker of BC. This result should be taken with caution, because complete clinical data were available only for 247 patients, mainly due to the missing data of hormone receptor status. However, even when hormone receptor status was excluded from the analysis, this SNP showed up as an independent prognostic marker. Additionally, multivariate analysis is complicated by correlations between variables whereby causality cannot be inferred.

The role of the kinetochore protein CENPF in the mitotic checkpoint has created controversial discussions in the past.[23] Recent studies have shown that CENPF is not an essential component of the spindle assembly checkpoint, but rather contributes to a prolonged mitotic delay in response to unattached kinetochores[23,24] by modulating recruitment of other checkpoint components to the kinetochores.[23]

The investigated SNP rs438034 in the *CENPF* gene is located in exon 18 and causes an amino acid change from arginin, a charged polar basic amino acid, to glycin, a nonpolar amino acid. However, no functional effect on the coded protein could be predicted by the Polyphen prediction tool. The *CENPF* gene shows a high genetic variability with 22 non-synonymous coding SNPs. Among them, there are 15 rare SNPs with frequencies <2%, 3 SNPs with a frequency <6% and 4 SNPs with frequencies ⩾35%. According to HapMap the common SNPs in the *CENPF* gene are in relatively high LD forming one big haploblock. The HapMap tag SNP picker tool suggests 3 tagging SNPs (rs2807663, rs335524, rs376776) with a minor allele frequency ⩾10%. Our selected SNP is in high LD with the HapMap tagging SNPs ($D' = 1.00$, $r^2 = 0.75–1.00$). Therefore, the observed effect may derive from other functional variants which are in strong LD with rs438034. The SNPs with allele frequencies <6% are not in LD with the common SNPs and thus cannot explain the effect of rs438034.

CENPF contains multiple motifs and functional domains.[25,26] Both the investigated SNP rs438034 and the other common, non-synonymous SNPs, which are in high LD with rs438034, are located within important functional domains of CENPF. rs438034 is located within a nuclear localisation domain.[27] Moreover, CENPF contains an Rb-binding domain which is located at the C-terminal region, 18 amino acids upstream of the SNP investigated in this study. The SNP rs3748697 (Asn2396Asp) lies within a region of internal repeats and might affect CENPF–kinetochore interactions.[28] Rs335524 (Arg2729Gln) and rs7289 (Asn3106Lys) are located in regions that have been shown to be important for spindle pole localisation, CENPF homo- and heterodimerisation and to contain putative chromatin binding sites, respectively.[27,28]

The power of our study was high (>90%) to detect moderate effects (OR > 1.4), which could be expected for SNPs selected based on their predicted functionality. Additionally, many of the SNPs were in LD with other putatively functional SNPs. As in our previous study on polymorphisms in the mitotic checkpoint genes,[16] no association with BC risk was observed and only one SNP associated with some clinical tumour characteristics and survival. We concentrated on regions, known to contain regulatory elements and non-synonymous SNPs, however, there may be rare SNPs in the screened regions or functional SNPs in other, still unknown, regulatory regions, which affect BC risk and clinical outcome.

In our study, the *CENPF* SNP was associated with worse BC-specific survival. The CENPF protein expression has been shown to correlate with reduced BC-specific survival, underlining its usefulness as a BC-specific marker of poor outcome.[29] Additionally, CENPF expression has been correlated significantly with high telomerase activity.[29] Moreover, a significant proportion of tumours over-expressing CENPF have been shown to be aneuploid, strengthening the relation between CENPF and markers of CIN.[29] Our results suggest a possible role for the investigated SNP rs438034 or other SNPs in high LD with this SNP in BC due to a differential expression or function of the CENPF protein. Additional studies are needed to clarify further the role of polymorphisms in the *CENPF* gene in BC prognosis.

Conflict of interest statement

None declared.

Acknowledgements

We thank Åsa Ågren (Department of Public Health and Clinical Medicine/Nutritional Research, Umeå University, Sweden) for her efficiency and skill in keeping track of samples and data. The Northern Sweden Breast Cancer Group is appreciated for providing the clinical data. This study was supported by a grant from the European Union (LSHC-CT-2004-503465).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejca.2008.10.001.

REFERENCES

1. Jallepalli PV, Lengauer C. Chromosome segregation and cancer: cutting through the mystery. *Nat Rev Cancer* 2001;**1**(2):109–17.
2. Kops GJ, Weaver BA, Cleveland DW. On the road to cancer: aneuploidy and the mitotic checkpoint. *Nat Rev Cancer* 2005;**5**(10):773–85.
3. Rajagopalan H, Lengauer C. Aneuploidy and cancer. *Nature* 2004;**432**(7015):338–41.
4. Draviam VM, Xie S, Sorger PK. Chromosome segregation and genomic stability. *Curr Opin Genet Dev* 2004;**14**(2):120–5.
5. Gollin SM. Mechanisms leading to chromosomal instability. *Semin Cancer Biol* 2005;**15**(1):33–42.
6. Jefford CE, Irminger-Finger I. Mechanisms of chromosome instability in cancers. *Crit Rev Oncol Hematol* 2006;**59**(1):1–14.
7. Tomonaga T, Nomura F. Chromosome instability and kinetochore dysfunction. *Histol Histopathol* 2007;**22**(2):191–7.
8. Weaver BA, Cleveland DW. Does aneuploidy cause cancer? *Curr Opin Cell Biol* 2006;**18**(6):658–67.
9. Carter SL, Eklund AC, Kohane IS, et al. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 2006;**38**(9):1043–8.
10. Perez de Castro I, de Carcer G, Malumbres M. A census of mitotic cancer genes: new insights into tumour cell biology and cancer therapy. *Carcinogenesis* 2007;**28**(5):899–912.
11. Fridlyand J, Snijders AM, Ylstra B, et al. Breast tumour copy number aberration phenotypes and genomic instability. *BMC Cancer* 2006;**6**:96.
12. Soerjomataram I, Louwman MW, Ribot JG, et al. An overview of prognostic factors for long-term survivors of breast cancer. *Breast Cancer Res Treat* 2008;**107**(3):309–30.
13. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;**447**(7148):1087–93.
14. Hartman M, Lindstrom L, Dickman PW, et al. Is breast cancer prognosis inherited? *Breast Cancer Res* 2007;**9**(3):R39.
15. Hemminki K, Ji J, Forsti A, et al. Survival in breast cancer is familial. *Breast Cancer Res Treat* 2008;**110**(1):177–82.
16. Vaclavicek A, Bermejo JL, Wappenschmidt B, et al. Genetic variation in the major mitotic checkpoint genes does not affect familial breast cancer risk. *Breast Cancer Res Treat* 2007;**106**(2):205–13.
17. Kaaks R, Lundin E, Rinaldi S, et al. Prospective study of IGF-I, IGF-binding proteins, and breast cancer risk, in northern and southern Sweden. *Cancer Cause Control* 2002;**13**(4):307–16.
18. Wong KK, Tsang YT, Shen J, et al. Allelic imbalance analysis by high-density single-nucleotide polymorphic allele (SNP) array with whole genome amplified DNA. *Nucleic Acids Res* 2004;**32**(9):e69.
19. Paez JG, Lin M, Beroukhim R, et al. Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res* 2004;**32**(9):e71.
20. Vaclavicek A, Hemminki K, Bartram CR, et al. Association of prolactin and its receptor gene regions with familial breast cancer. *J Clin Endocrinol Metab* 2006;**91**(4):1513–9.
21. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;**21**(2):263–5.
22. Lo YL, Yu JC, Chen ST, et al. Breast cancer risk associated with genotypic polymorphism of the mitotic checkpoint genes: a multigenic study on cancer susceptibility. *Carcinogenesis* 2007;**28**(5):1079–86.
23. Feng J, Huang H, Yen TJ. CENP-F is a novel microtubule-binding protein that is essential for kinetochore attachments and affects the duration of the mitotic checkpoint delay. *Chromosoma* 2006;**115**(4):320–9.
24. Varis A, Salmela AL, Kallio MJ. Cenp-F (mitosin) is more than a mitotic marker. *Chromosoma* 2006;**115**(4):288–95.
25. Ma L, Zhao X, Zhu X. Mitosin/CENP-F in mitosis, transcriptional control, and differentiation. *J Biomed Sci* 2006;**13**(2):205–13.
26. Zhu X, Mancini MA, Chang KH, et al. Characterization of a novel 350-kilodalton nuclear phosphoprotein that is specifically involved in mitotic-phase progression. *Mol Cell Biol* 1995;**15**(9):5017–29.
27. Zhu X, Chang KH, He D, et al. The C terminus of mitosin is essential for its nuclear localization, centromere/kinetochore targeting, and dimerization. *J Biol Chem* 1995;**270**(33):19545–50.
28. Zhu X. Structural requirements and dynamics of mitosin-kinetochore interaction in M phase. *Mol Cell Biol* 1999;**19**(2):1016–24.
29. O'Brien SL, Fagan A, Fox EJ, et al. CENP-F expression is associated with poor prognosis and chromosomal instability in patients with primary breast cancer. *Int J Cancer* 2007;**120**(7):1434–43.