



HAL
open science

Genetic variation in genes encoding for polymerase ζ subunits associates with breast cancer risk, tumour characteristics and survival

Verena Varadi, Melanie Bevier, Ewa Grzybowska, Robert Johansson, Kerstin Enquist, Roger Henriksson, Dorota Butkiewicz, Jolanta Pamula-Pilat, Karolina Tecza, Kari Hemminki, et al.

► To cite this version:

Verena Varadi, Melanie Bevier, Ewa Grzybowska, Robert Johansson, Kerstin Enquist, et al.. Genetic variation in genes encoding for polymerase ζ subunits associates with breast cancer risk, tumour characteristics and survival. *Breast Cancer Research and Treatment*, 2011, 129 (1), pp.235-245. 10.1007/s10549-011-1460-z . hal-00628285

HAL Id: hal-00628285

<https://hal.science/hal-00628285>

Submitted on 1 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Genetic variation in genes encoding for polymerase ζ subunits associates with breast cancer risk, tumour characteristics and survival.

Verena Varadi¹, Melanie Bevier¹, Ewa Grzybowska², Robert Johansson³, Kerstin Enquist⁴, Roger Henriksson^{3,5}, Dorota Butkiewicz², Jolanta Pamula-Pilat², Karolina Tecza², Kari Hemminki^{1,6}, Per Lenner³ and Asta Försti^{1,6}

¹Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, ² Center for Translational Research and Molecular Biology of Cancer, M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice, Poland, ³Department of Oncology, Norrlands University Hospital, Umeå, Sweden, ⁴Department of Public Health and Clinical Medicine/Nutritional Research, Umeå University, Sweden, ⁵Department of Oncology, Karolinska University Hospital, Stockholm, Sweden, ⁶Center for Primary Health Care Research, Clinical Research Center, Lund University, Malmö, Sweden

*Corresponding author: Verena Varadi; German Cancer Research Center; Department of Molecular Genetic Epidemiology; Im Neuenheimer Feld 580; 69120 Heidelberg; phone: +49 6221 42 1811
Email: v.varadi@dkfz-heidelberg.de

Keywords: Breast Cancer, SNPs, chromosomal instability, polymerases, prognostic marker

Abstract

Chromosomal instability is a known hallmark of many cancers. DNA polymerases represent a group of enzymes that are involved in the mechanism of chromosomal instability as they have a central function in DNA metabolism. We hypothesized that genetic variation in the polymerase genes may affect gene expression or protein configuration and by that cancer risk and clinical outcome. We selected four genes encoding for the catalytic subunits of the polymerases β , δ , θ and ζ (*POLB*, *POLD1*, *POLQ* and *REV3L*, respectively) and two associated proteins (*MAD2L2* and *REVI*) because of their previously reported association with chromosomal instability and/or tumorigenesis. We selected potentially functional and most informative tagging single nucleotide polymorphisms (SNPs) for genotyping in a population-based series of 783 Swedish breast cancer (BC) cases and 1562 controls. SNPs that showed a significant association in the Swedish population were additionally genotyped in a Polish population consisting of 506 familial/early onset BC cases and 568 controls. SNPs in all three polymerase ζ subunit genes associated either with BC risk or prognosis. Two SNPs in *REV3L* and one SNP in *MAD2L2* associated with BC risk: rs462779 (multiplicative model: OR 0.79, 95%CI 0.68-0.92), rs3204953 (dominant model: OR 1.28, 95%CI 1.05-1.56) and rs2233004 (recessive model: OR 0.49, 95%CI 0.28-0.86). Homozygous carriers of the minor allele C of the third SNP in *REV3L*, rs11153292, had significantly worse survival compared to the TT genotype carriers (HR 2.93, 95%CI 1.34-6.44). Minor allele carriers of two *REVI* SNPs (rs6761391 and rs3792142) had significantly more often large tumours and tumours with high histologic grade and stage. No association was observed for SNPs in *POLB*, *POLQ* and *POLD1*. Altogether, our data suggest a significant role of genetic variation in the polymerase ζ subunit genes regarding the development and progression of BC.

Introduction

DNA polymerases have a central function in DNA metabolism as they are involved in DNA replication, repair, recombination and translesion synthesis and with that ensure faithful transmission of genetic information from one generation to the next [1-4]. As a dysfunction within these genes might have severe influences, DNA polymerase genes represent potential candidates for cancer genes. A total number of 15 enzymes with a known DNA synthesis activity exist. Some of these polymerases, the so called classic DNA polymerases [3], are involved in rapid, high-fidelity replication of most of the eukaryotic genome, but are not capable of working on sites of DNA damage [5]. That often leads to a blockage of DNA synthesis with a stalled replication fork [6]. The arrest of DNA replication might even influence the ability of the cells to complete the cell cycle, and as a consequence it induces multiple mechanisms collectively referred to as DNA-damage-tolerance mechanism [2]. Several specialized DNA polymerases, the so called translesion synthesis (TLS) DNA polymerases [3], are involved in this process, and they allow strand extension across template lesions [2]. Despite their special function across template lesions, when operating on undamaged DNA or non-cognate lesions, the TLS polymerases show a considerably reduced fidelity. Thus, both over- and underexpression of these polymerases can lead to an increased mutation rate, suggesting that they could potentially act both as oncogenes and tumour suppressor genes, respectively [7,8]. Loss of specific DNA polymerases has been reported to cause chromosomal instability in mammalian cells [6] and chromosomal unstable cancers are likely to have a poorer prognosis than diploid cancers [9,10] with the degree of aneuploidy being in accordance with the severity of the disease [11].

We hypothesized, that common genetic variation in the form of single nucleotide polymorphisms (SNPs) in genes encoding classic or TLS DNA polymerases could

affect DNA replication by influencing gene expression or protein configuration. This might increase the level of mutations and influence cellular transformation and in the end affect cancer risk and clinical outcome. We focused on four genes encoding the catalytic subunits of two classical polymerases, polymerase δ (*POLD1*) and polymerase β (*POLB*), and two TLS-polymerases polymerase ζ (*REV3L*) and polymerase θ (*POLQ*). These polymerases have already been reported to be involved in different types of cancer or causing chromosomal instability [7,8,12,13,4,6]. In a second selection round, we selected SNPs in the genes *MAD2L2* and *REV1*. The encoded proteins act as a subunit (*MAD2L2*) and a scaffold protein (*REV1*) for the polymerase ζ enzyme [13,14]. We selected putatively functional SNPs and most informative tagging SNPs (tagSNPs) from these six genes and genotyped them in a population-based series of 783 Swedish BC cases with detailed clinical data and up to 15 years of follow-up together with 1562 age and gender matched controls. SNPs that showed a significant association in the Swedish population were also genotyped in a Polish population consisting of 506 familial/early onset BC cases with corresponding clinical data and 568 controls.

Material and Methods

Study populations

All selected SNPs were genotyped in a Swedish study population consisting of 783 BC cases and 1562 controls. Seven hundred forty nine cases with 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 [15]. 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 (\pm 6 months) and the time of sampling (\pm 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. Thirty three samples were collected consecutively during the same time period from untreated patients referred to the Department of Oncology for newly diagnosed BC. Their controls, also matched for age and gender, were selected from the Västerbotten intervention cohort. The genotyping for this study was performed blinded by the case-control status of each sample using whole genome amplified DNA as described by Brendle et al 2009 [16]. Clinical data for the unselected BC cases was retrieved from the registry managed by the Northern Sweden Breast Cancer Group (Table 1). Information about the date of death was collected from the Swedish population register with a BC-specific follow-up until 31st of December 2004. The median follow-up time was 4.7 years.

In order to investigate whether the SNPs that showed a significant association in the Swedish population also show an association in a population with familial/early onset BC, we genotyped a second population consisting of 506 familial/early onset BC cases and 568 regionally and ethnically matched controls [17]. The cases were collected through the Chemotherapy Clinics, the Genetic Counselling Service and the Surgery Clinics (Gliwice, Poland) between the years 1997-2009. Only cases without BRCA1/2 mutations were included in the study. Clinical data was available for 475 cases (Table 1). Follow-up was performed until 31st December 2009. The median follow-up time after surgery was 4.1 years. We used event-free survival (EFS) as the clinical endpoint. EFS was defined as the time from surgery to the occurrence of a breast event (i.e. local or distant recurrence, contralateral BC or death due to any cause).

All participants gave informed consent to the use of their samples for research purpose. The study was approved by the ethical committees of the participating institutes.

Selection of SNPs

We focused on four genes encoding for the catalytic subunits of the polymerases β , δ , θ and ζ (*POLB*, *POLD1*, *POLQ* and *REV3L*, respectively), and two associated genes, *MAD2L2* and *REVI*. SNPs with a minor allele frequency (MAF) of ≥ 0.1 located in promoters (up to 1kb upstream of transcription start), 3'UTRs, reported nonsynonymous (ns) SNPs and most informative tagSNPs, were selected in these genes using data for the CEU population (Utah residents with Northern and Western European ancestry from the CEPH - Centre d'Etude du Polymorphisme Human population) from the International HapMap Project (<http://www.hapmap.org/>) and the NCBI dbSNP database (<http://www.ncbi.nlm.nih.gov/snp>). TagSNPs were selected using the plug-in tagSNP picker incorporated in the HapMap function <Bulk data download> (<http://hapmap.ncbi.nlm.nih.gov/downloads/index.html.en>). They were sequenced for verification in a set of 32 BC samples as described earlier by Vaclavicek et al 2007 [18]. Promoter polymorphisms within or close to transcription factor binding sites may alter gene expression and contribute to tumorigenesis. To identify putative transcription factor binding sites we used the TESS-Transcription Element Search System tool (<http://www.cbil.upenn.edu/tess/>). SNPs located in the 3'UTR of the genes may have an effect on microRNA binding sites. MicroRNAs can down-regulate protein translation or even degrade mRNA. To identify putative microRNA binding sites we used the microRNA.org prediction tool (<http://www.microrna.org/microrna/getGeneForm.do>). nsSNPs were included in our analyses due to their potential effect on protein function. The SIFT (Sorting Intolerant from Tolerant) prediction tool (<http://sift.jcvi.org/>) was

used to predict a possible impact of an amino acid substitution on the structure and function of human proteins [19,20]. Additionally, an estimation of the impact of the different SNPs according to their linkage disequilibrium (LD) with other SNPs was done using the Haploview software [21]. Verified SNPs and the SNPs selected for genotyping are documented in Table 2.

Genotyping

Genotyping was performed using an allele specific PCR based KASPar SNP Genotyping System purchased from KBiosciences (Hoddesdon, Great Britain). Thermocycling was performed according to the KBiosciences` PCR conditions. Detection was performed using an ABI PRISM 7900 HT Sequence Detection System with SDS 2.2 software (Applied Biosystems, Weiterstadt, Germany). The Swedish sample set contained 104 duplicate samples as quality controls for the KASPar assays. For less than 0.1% the genotypes of the duplicate samples did not agree with each other.

Statistical analysis

The observed genotype frequencies in the controls were tested for Hardy-Weinberg equilibrium (HWE) using the χ^2 test. Statistical significance for a different genotype distribution in cases and controls was determined by global χ^2 test. When the overall genotype effect of a SNP on the risk of BC or a clinical factor was statistically significant at the level of 0.05, the best model to represent the relationship between the genotypes and the risk of BC or a clinical factor, respectively, was selected based on likelihood ratio tests (LRTs). Odds ratios (ORs) and 95% confidence intervals (95% CIs) for associations between genotypes, BC risk and tumour characteristics were calculated by logistic regression (PROC LOGISTIC, SAS Version 9.2; SAS Institute,

Cary, NC). Haplotypes and diplotypes were estimated and compared using the Haplotypes procedure of SAS/Genetics (SAS Version 9.1). The frequency of a specific haplotype or diplotype (haplotype pair) was estimated by summing the carrier probability for all women. The relationship between haplotypes/diplotypes and BC was assessed by logistic regression with the most common haplotype/diplotype as the reference. Gene-gene interaction was analysed by pairwise interaction using logistic regression.

The survival curves for the BC-specific survival were derived by Kaplan-Meier method (PROC LIFETEST, SAS Version 9.2, SAS Institute). The joint analysis was carried out using Mantel–Haenszel adjustment. The relative risk of death by BC was estimated as hazard ratio (HR) using Cox regression (PROC PHREG, SAS Version 9.2, SAS Institute). In the Swedish population, censoring events were death by another cause than BC, moving out of the study and 31st of December 2004. In the Polish study, cases without a breast event on 31st December 2009 were censored. The HRs were also calculated within subgroups of cases with a similar manifestation of a clinical factor. Furthermore, the hazard ratios were adjusted for the clinical factors (estrogen and progesterone receptor status, tumour size, lymph node metastasis, histologic grade) to determine the value of the genotypes as an independent prognostic marker.

Results

SNPs in REV3L and MAD2L2 associate with BC risk in the Swedish study population

The global χ^2 test indicated a significant association for two SNPs in *REV3L* and one SNP in *MAD2L2* with BC susceptibility (Table 3). According to LRT, the best model to represent the relationship between the genotypes of the *REV3L* SNPs and the risk of BC was multiplicative for rs462779 (OR 0.79, 95%CI 0.68-0.92) and dominant for rs3204953 (OR 1.28, 95%CI 1.05-1.56). The haplotype and diplotype analysis of the three SNPs genotyped in *REV3L* confirmed the association by showing an inverse correlation of the haplotype CCG (rs11153292 T/C - rs462779 T/C - rs3204953 G/A: OR 0.81, 95% CI 0.70-0.95) and the diplotype CCG/CCG (OR 0.62, 95% CI 0.40-0.97) with BC (data not shown).

For the *MAD2L2* SNP rs2233004, carriers of the minor allele had a significantly decreased BC risk (recessive model: OR 0.49, 95%CI 0.28-0.86). This result was confirmed by the haplotype and diplotype analysis in which the only haplotype containing the variant G allele of rs2233004, GAT, was associated with BC risk with an OR of 0.77 (95%CI 0.60-0.99) and the diplotype GAT/GAT with an OR of 0.50 (95%CI 0.28-0.90) (data not shown).

Gene-gene interaction studies between *REV3L* and *MAD2L2* with risk associated alleles showed no evidence for epistasis (data not shown). SNPs in *POLB*, *POLQ* and *POLD1* were not associated with BC susceptibility (Table 3).

None of the associations were replicated in the Polish population, which contained familial/early onset BC cases, at a statistically significant level. However, in the joint analysis the ORs remained statistically significant (Table 3).

Association of REV3L SNPs rs11153292 and rs462779 with survival

For the tagSNP rs11153292, located in intron 6 of *REV3L*, an association with BC-specific survival was observed. Homozygous carriers of the minor allele C had a significantly worse survival compared to the TT genotype carriers (HR 2.93, 95%CI 1.34-6.44) (Table 4, Fig 1A). After adjustment for tumour size, lymph node metastasis, histologic grade, estrogen and progesterone receptor status, the HR was 3.95 (95% CI 1.18-13.22) (Table 4), indicating the value of rs11153292 as an independent prognostic marker. The survival was poor especially for women with the CC genotype and aggressive tumours (positive lymph node metastasis: HR 4.20, 95%CI 1.32-13.44, histologic grade 3: HR 4.70, 95%CI 1.52-14.57, stage II-IV: HR 3.20, 95%CI 1.23-8.32) (Fig.1 B, C, D).

The *REV3L* nsSNP rs462779 was associated with survival in patients with aggressive tumours. The CC genotype carriers had worse survival than the common TT genotype carriers (positive lymph node metastasis: HR 4.15 95%CI 1.18-14.62; histologic grade 3: HR 3.37 95%CI 1.12-10.08; stage II-IV: HR 3.51 95%CI 1.20-10.23), although the association with disease-free survival was not statistically significant among all BC patients (HR 2.23 95%CI 0.78-6.38). None of the other SNPs associated with survival. No association with EFS in the Polish population was observed.

Association of REV1 SNPs rs6761390 and rs3792142 with tumour characteristics

When we examined the genotype distribution by tumour characteristics at the time of diagnosis, significant associations were observed for the *REV1* SNPs rs6761390, rs3792136, rs3792142 and rs4341989 (supplementary table1+2). Minor allele carriers of the SNPs rs6761390 and rs3792142 had more often large tumours and tumours with high histologic grade and stage than the common homozygotes. LD analysis indicated that these two SNPs were in complete LD. This was previously not known as only

genotype information from two different European populations, CEU for rs3792142 and CEPH for rs6761390, was available. For rs3792136, carriers of the minor allele were more likely to have small tumours and stage 0+I tumours than the CC carriers. Minor allele carriers of rs4341989 had more often hormone receptor positive tumours than the common homozygotes. These SNPs were not associated with tumour characteristics in the Polish population. No other associations with tumour characteristics were observed.

Discussion

Chromosomal instability is a characteristic of almost all human cancers [22]. The degree of chromosomal instability has been shown to predict clinical outcome and discern prognostically relevant BC subtypes [10]. The strength of our study lies in the selection of six polymerase subunit genes according to their previously reported association with chromosomal instability and/or tumorigenesis. We focused on SNPs with the highest probability of an influence on gene transcription or protein product, and on the most informative tagSNPs in order to cover a multitude of genetic variation within a gene. With the present sample size of the Swedish study population we had a power of >90% to detect an OR of 1.5 for the risk of BC in SNPs with a MAF \geq 0.1.

We showed that genetic variation in genes encoding for the polymerase ζ subunits REV3L, MAD2L2 and REV1 had a statistically significant influence on BC risk and the outcome of the disease. Two SNPs in *REV3L* associated with BC risk, one of these was additionally associated with survival for patients with aggressive tumours. A third SNP might even represent an independent prognostic marker. One SNP in *MAD2L2* associated with BC risk. Several SNPs in *REV1* associated with tumour characteristics. *REV3L* encodes for the catalytic subunit of polymerase ζ . Wittschieben et. al. reported that loss of *REV3L* expression could contribute to genomic instability during neoplastic

transformation and progression [6] and Brondello et. al. reported that *REV3L* depletion leads to genomic instability through double strand break accumulation [8]. *REV3L* has also been reported to have a lower expression in tumour versus normal tissue from early to late stages of colon carcinoma [8]. These findings provide evidence for a role of *REV3L* as a tumour suppressor gene.

The two SNPs in *REV3L*, which associated with BC risk, are nsSNPs: rs462279 leads to an amino acid change from threonine to isoleucine, rs3204953 from valine to isoleucine. However, no functional effect on the coded protein was predicted for any SNP by the SIFT prediction tool, nor did the three 3`UTR SNPs that were captured by the two nsSNPs change the binding sites for any microRNA. The third SNP, rs11153292, in *REV3L* associated with survival with homozygous carriers of the minor allele C having a worse survival compared with the common homozygotes. The strongest association was observed in patients with aggressive tumours (lymph node metastasis and high stage tumours). rs11153292 is a highly informative intronic tagSNP capturing 41 other intronic SNPs, with unknown function.

In *MAD2L2*, an association of the SNP rs2233004 with BC risk was observed. This promoter SNP is not in LD with any other SNP. The G-allele of the SNP creates an additional binding site for the transcription factor CCAAT enhancer binding protein alpha (*CEBPA*) (TFSEARCH). However, the homozygous minor allele genotype that was associated with risk was very rare with a frequency of 2% and 4% in cases and controls, respectively, minimizing the power of this result. The two other promoter SNPs genotyped in *MAD2L2* were not associated with BC risk. In our previous case-control study, the most common *MAD2L2* promoter SNP rs2233006 was not associated with familial BC [18]. Thus, genetic variation in the promoter of *MAD2L2* does not seem to affect strongly on BC risk.

In *REVI*, rs6761390 and rs3792142, which were in complete LD, associated with tumour size, histologic grade and stage. rs6761390 is located in the promoter of the *REVI* gene, rs3792142 in intron 5. Additionally, they capture three intron 1 SNPs, rs4143760, rs1011633 and rs896249. Only for rs4143760 a change of a transcription factor binding site was predicted by TFSEARCH, with the A-allele of the SNP creating a sequence for binding of GATA transcription factors.

None of the associations between the genotyped SNPs and BC risk, tumour characteristics and survival in the Swedish study population were confirmed in the Polish study population at a statistically significant level. The main difference in the study populations lies in the selection criteria for the cases. The Swedish study is a population-based study, whereas the Polish study comprises of cases with a family history of BC or an early onset of the disease. The sample sets show a significant difference in the median age of the cases (Swedish 58 years; Polish 43 years). Additionally, the survival probabilities for the Polish cases were calculated using any breast event as the clinical endpoint compared with death due to BC for the Swedish population. Thus, the lack of association in the Polish population may not necessarily indicate a lack of association with death due to BC, but a lack of association with BC progression. We selected the genes for the present study based on their known association with chromosomal instability. Chromosomal instability is common in both sporadic and hereditary cancer. In hereditary cancers, the presence of chromosomal instability has been linked to mutations in DNA repair genes [22]. Germline mutations in *BRCA1*, *BRCA2*, *PALB2*, *NBS1*, *WRN*, *BLM* and *RECQL4*, all of which have been linked to the repair of DNA double-strand breaks or DNA interstrand cross links, have been reported to predispose to the development of various chromosomal instable hereditary cancers [23-25]. Unlike hereditary cancers, the molecular basis of

chromosomal instability in sporadic cancers remains unclear [22]. These data indicate that different mechanisms influence chromosomal instability in hereditary and sporadic cancers, giving a possible explanation for the difference in our results for the Polish and Swedish study population.

For the genes *POLB*, *POLD1* and *POLQ*, there was no association of the selected SNPs with BC susceptibility, tumour characteristics or survival in the Swedish sample set. Similar results were obtained by Han et al. [26]. Taking our results on mainly postmenopausal women and the results of Han et al in premenopausal women into account, we can exclude, with a high probability any effect of common genetic variation within these genes on BC susceptibility and prognosis.

Altogether, all three genes encoding polymerase ζ subunits showed an association either with BC susceptibility, progression or survival. Further analysis to explain the functional consequences of the SNPs on gene expression and on BC progression are warranted.

Acknowledgements

We thank Asa Agren (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.

References

1. Caldecott KW (2008) Single-strand break repair and genetic disease. *Nat Rev Genet* 9 (8):619-631
2. Friedberg EC (2005) Suffering in silence: the tolerance of DNA damage. *Nat Rev Mol Cell Biol* 6 (12):943-953
3. Loeb LA, Monnat RJ, Jr. (2008) DNA polymerases and human disease. *Nat Rev Genet* 9 (8):594-604
4. Sweasy JB, Lauper JM, Eckert KA (2006) DNA polymerases and human diseases. *Radiat Res* 166 (5):693-714
5. Hubscher U, Maga G, Spadari S (2002) Eukaryotic DNA polymerases. *Annu Rev Biochem* 71:133-163
6. Wittschieben JP, Reshmi SC, Gollin SM, Wood RD (2006) Loss of DNA polymerase zeta causes chromosomal instability in mammalian cells. *Cancer Res* 66 (1):134-142
7. Albertella MR, Lau A, O'Connor MJ (2005) The overexpression of specialized DNA polymerases in cancer. *DNA Repair (Amst)* 4 (5):583-593
8. Brondello JM, Pillaire MJ, Rodriguez C, Gourraud PA, Selves J, Cazaux C, Piette J (2008) Novel evidences for a tumor suppressor role of Rev3, the catalytic subunit of Pol zeta. *Oncogene* 27 (47):6093-6101
9. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z (2006) A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 38 (9):1043-1048
10. Habermann JK, Doering J, Hautaniemi S, Roblick UJ, Bundgen NK, Nicorici D, Kronenwett U, Rathnagiriswaran S, Mettu RK, Ma Y, Kruger S, Bruch HP, Auer G, Guo NL, Ried T (2009) The gene expression signature of genomic instability in breast cancer is an independent predictor of clinical outcome. *Int J Cancer* 124 (7):1552-1564
11. Rajagopalan H, Lengauer C (2004) Aneuploidy and cancer. *Nature* 432 (7015):338-341
12. Kawamura K, Bahar R, Seimiya M, Chiyo M, Wada A, Okada S, Hatano M, Tokuhisa T, Kimura H, Watanabe S, Honda I, Sakiyama S, Tagawa M, J OW (2004) DNA polymerase theta is preferentially expressed in lymphoid tissues and upregulated in human cancers. *Int J Cancer* 109 (1):9-16
13. Srivastava DK, Husain I, Arteaga CL, Wilson SH (1999) DNA polymerase beta expression differences in selected human tumors and cell lines. *Carcinogenesis* 20 (6):1049-1054
14. Szuts D, Marcus AP, Himoto M, Iwai S, Sale JE (2008) REV1 restrains DNA polymerase zeta to ensure frame fidelity during translesion synthesis of UV photoproducts in vivo. *Nucleic Acids Res* 36 (21):6767-6780
15. Kaaks R, Lundin E, Rinaldi S, Manjer J, Biessy C, Soderberg S, Lenner P, Janzon L, Riboli E, Berglund G, Hallmans G (2002) Prospective study of IGF-I, IGF-binding proteins, and breast cancer risk, in northern and southern Sweden. *Cancer Causes Control* 13 (4):307-316
16. Brendle A, Brandt A, Johansson R, Enquist K, Hallmans G, Hemminki K, Lenner P, Forsti A (2009) Single nucleotide polymorphisms in chromosomal instability genes and risk and clinical outcome of breast cancer: a Swedish prospective case-control study. *Eur J Cancer* 45 (3):435-442
17. Varadi V, Brendle A, Brandt A, Johansson R, Enquist K, Henriksson R, Svenson U, Tavelin B, Roos G, Hemminki K, Lenner P, Forsti A (2009) Polymorphisms in

- telomere-associated genes, breast cancer susceptibility and prognosis. *Eur J Cancer* 45 (17):3008-3016
18. Vaclavicek A, Bermejo JL, Wappenschmidt B, Meindl A, Sutter C, Schmutzler RK, Kiechle M, Bugert P, Burwinkel B, Bartram CR, Hemminki K, Forsti A (2007) Genetic variation in the major mitotic checkpoint genes does not affect familial breast cancer risk. *Breast Cancer Res Treat* 106 (2):205-213
 19. Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31 (13):3812-3814
 20. Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. *Genome Res* 11 (5):863-874
 21. Dawn Teare M, Barrett JH (2005) Genetic linkage studies. *Lancet* 366 (9490):1036-1044
 22. Negrini S, Gorgoulis VG, Halazonetis TD Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* 11 (3):220-228
 23. Bachrati CZ, Hickson ID (2003) RecQ helicases: suppressors of tumorigenesis and premature aging. *Biochem J* 374 (Pt 3):577-606
 24. Kennedy RD, D'Andrea AD (2006) DNA repair pathways in clinical practice: lessons from pediatric cancer susceptibility syndromes. *J Clin Oncol* 24 (23):3799-3808
 25. Ripperger T, Gadzicki D, Meindl A, Schlegelberger B (2009) Breast cancer susceptibility: current knowledge and implications for genetic counselling. *Eur J Hum Genet* 17 (6):722-731
 26. Han J, Haiman C, Niu T, Guo Q, Cox DG, Willett WC, Hankinson SE, Hunter DJ (2009) Genetic variation in DNA repair pathway genes and premenopausal breast cancer risk. *Breast Cancer Res Treat* 115 (3):613-622

Table 1: Tumour characteristics of the Swedish and the Polish breast cancer samples at the time of diagnosis.

Characteristics	Swedish samples N (%)	Polish samples N (%)
Age at diagnosis, mean (range)	58.0 (30.6-86.2)	43.1 (23.0-56.0)
Estrogen receptor		
Positive	404 (51.50)	257 (50.79)
Negative	133 (16.99)	189 (37.35)
Missing	246 (31.42)	60 (11.86)
Progesterone receptor		
Positive	283 (36.14)	262 (51.78)
Negative	214 (27.33)	182 (35.97)
Missing	286 (36.53)	62 (12.25)
Estrogen/progesterone receptor		
ER+/PR+	229 (29.25)	224 (44.27)
ER+/PR-	71 (9.07)	31 (6.13)
ER-/PR+	12 (1.53)	38 (7.51)
ER-/PR-	114 (14.56)	150 (29.64)
Missing	357 (45.59)	63 (12.45)
Tumour size in mm		
≤ 20 mm	501 (63.98)	190 (37.55)
> 20 mm	228 (29.12)	289 (57.11)
Missing	54 (6.90)	27 (5.34)
Histologic grade		
Grade 1+2	516 (65.90)	NA
Grade3	225 (28.74)	NA
Missing	42 (5.36)	NA
Regional lymph node metastasis		
Negative	469 (59.90)	265 (52.37)
Positive	220 (28.10)	217 (42.89)
Missing	94 (12.01)	24 (4.74)
Stage at diagnosis		
0-I	407 (51.98)	NA
II-IV	367 (46.87)	NA
Missing	9 (1.15)	NA

N= number of samples

NA=not available

Table 2: Summarized information with possible functional effect and linkage disequilibrium (LD) for SNPs in the selected polymerase and polymerase associated genes. All SNPs complying with our

selection criteria that were confirmed by sequencing and/or database information with a minor allele frequency ≥ 0.1 are listed. The SNPs selected for genotyping are marked in bold.

Gene	SNP	region	possible functional effect	LD
<i>POLB</i>	rs2307158	promoter	putative TFBS*	rs3136717 ($D'=1, r^2=0.55$)
	rs3136717	intron 1	tag SNP	tags 4 out of 5 listed HapMap SNPs
<i>POLD1</i>	rs3219281	promoter	putative TFBS*	rs1726801 ($D'=1, r^2=1$)
	rs3219282	promoter	putative TFBS*	rs1726801 ($D'=1, r^2=1$)
	rs1726801	exon 4	R119H tolerated	
<i>POLQ</i>	rs587553	promoter	putative TFBS*	rs11713643 ($D'=1, r^2=0.52$)
	rs3806614	promoter	putative TFBS*	rs11713643 ($D'=1, r^2=1$)
	rs13065220	promoter	putative TFBS*	rs11713643 ($D'=1, r^2=0.79$)
	- 798 ATG	promoter	putative TFBS*	rs11713643 ($D'=1, r^2=0.50$)
	rs11713643	intron 1	tag SNP	tags 23 out of 53 listed HapMap SNPs
	rs487848	exon 11	A581V tolerated	
	rs3218649	exon 16	T982R tolerated	rs11713643 ($D'=1, r^2=1$)
	rs3218651	exon 16	H1201R tolerated	
	rs532411	exon 24	A2304V tolerated	rs487848 ($D'=1, r^2=1$)
	rs1381057	exon 28	Q2513R tolerated	rs11713643 ($D'=1, r^2=0.79$)
rs3218634	exon 29	L2538V possibly damaging	rs487848 ($D'=1, r^2=0.81$)	
<i>REV3L</i>	rs11153292	intron 6	tag SNP	tags 41 out of 108 listed HapMap SNPs
	rs462779	exon 15	T1224I tolerated	
	rs3204953	exon 33	V3064I tolerated	
	rs465646	3'UTR	putative microRNA binding site	rs462779 ($D'=1, r^2=0.70$)
	rs1061388	3'UTR	putative microRNA binding site	rs3204953 ($D'=1, r^2=1$)
	rs181294	3'UTR	putative microRNA binding site	rs462779 ($D'=1, r^2=0.76$)
<i>MAD2L2</i>	rs2233004	promoter	putative TFBS*	
	rs2233006	promoter	putative TFBS*	
	rs2233011	promoter	putative TFBS*	
	rs2272978	promoter	putative TFBS*	rs2233011 ($D'=1, r^2=1$)
<i>REVI</i>	rs28369925	promoter	putative TFBS*	rs3792136 ($D'=1, r^2=1$)
	rs6761390	promoter	putative TFBS*	
	rs28369929	intron 1**	putative TFBS*	rs3792136 ($D'=1, r^2=1$)
	rs6714244	intron 1**	putative TFBS*	rs3792152 ($D'=1, r^2=1$)
	rs1017697	intron 1**	putative TFBS*	rs13409359 ($D'=1, r^2=1$)
	rs4143760	intron 1**	putative TFBS*	rs3792142 ($D'=1, r^2=1$)
	rs10173883	intron 1**	putative TFBS*	rs13409359 ($D'=1, r^2=1$)
	rs6542882	intron 1**	putative TFBS*	rs3792152 ($D'=1, r^2=1$)
	rs13430962	intron 1**	putative TFBS*	rs13409359 ($D'=1, r^2=1$)
	rs1011633	intron 1**	putative TFBS*	rs3792142 ($D'=1, r^2=1$)
	rs8896249	intron 1**	putative TFBS*	rs3792142 ($D'=1, r^2=1$)
	rs737094	intron 1**	putative TFBS*	
	Gene	SNP	region	possible functional effect
<i>REVI</i>	rs1973011	intron 1**	putative TFBS*	

rs9308822	intron 1**	putative TFBS*	rs13409359 ($D'=1, r^2=1$)
rs13415713	intron 1**	putative TFBS*	rs13409359 ($D'=1, r^2=1$)
rs3792136	intron 4	tag SNP	tags 7 out of 77 listed HapMap SNPs
rs3087403	exon 5	V138M tolerated	rs4341989 ($D'=1, r^2=0.96$)
rs3792142	intron 5	tag SNP	tags 17 out of 77 listed HapMap SNPs
rs3087386	exon 6	F257S tolerated	rs3792152 ($D'=1, r^2=1$)
rs3087399	exon 6	N373S tolerated	rs13409359 ($D'=1, r^2=1$)
rs13409359	intron 8	tag SNP	tags 28 out of 77 listed HapMap SNPs
rs4341989	intron 8	tag SNP	tags 6 out of 77 listed HapMap SNPs
rs3792152	intron 12	tag SNP	tags 25 out of 77 listed HapMap SNPs

*TFBS= transcription factor binding site

** upstream of ATG

Table 3: Association of the SNPs selected in polymerase and polymerase associated genes with BC susceptibility in the Swedish study population. SNPs that showed a significant association in the Swedish population were also genotyped in the Polish population. For the SNPs with a significant association, the best model was calculated. For the SNPs with genotyping information for both populations, a joint analysis was performed.

Gene	SNP ID	Genotype	Swedish study population				Polish study population				Joint analysis			
			Cases N (%)	Controls N (%)	OR 95%CI	p-value	global p-value	Cases N (%)	Controls N (%)	OR 95%CI	p-value	global p-value	OR 95%CI	p-value
<i>POLB</i>	rs3136717	TT	589 (76.59)	1187 (77.18)										
		TC	165 (21.46)	327 (21.26)	1.02 (0.83-1.27)	0.82								
		CC	15 (1.95)	24 (1.56)	1.23 (0.65-2.36)	0.53	0.80							
<i>POLD</i>	rs1726801	GG	678 (89.56)	1335 (88.53)										
		GA	79 (10.44)	173 (11.47)	0.89 (0.67-1.18)	0.41								
		AA	0 (0)	0 (0)			0.41							
<i>POLQ</i>	rs487848	CC	671 (87.26)	1329 (86.81)										
		CT	94 (12.22)	194 (12.67)	0.96 (0.74-1.26)	0.79								
		TT	4 (0.52)	8 (0.52)	0.88 (0.26-2.95)	0.84	0.94							
	rs3218651	AA	533 (69.13)	1069 (69.60)										
		AG	217 (28.15)	424 (27.60)	1.03 (0.85-1.26)	0.77								
		GG	21 (2.72)	43 (2.80)	1.00 (0.58-1.71)	0.99	0.96							
	rs11713643	CC	320 (41.67)	618 (40.39)										
		CT	338 (44.01)	704 (46.01)	0.94 (0.78-1.13)	0.49								
		TT	110 (14.32)	208 (13.59)	1.02 (0.78-1.34)	0.88	0.72							
<i>REV3L</i>	rs11153292	TT	218 (28.42)	412 (26.82)										
		TC	393 (51.24)	797 (51.89)	0.93 (0.75-1.14)	0.47								
		CC	156 (20.34)	327 (21.29)	0.90 (0.70-1.16)	0.42	0.68							
	rs462779	TT	492 (63.81)	886 (57.72)				307 (68.07)	315 (68.48)					
		TC	250 (32.43)	562 (36.61)	0.80 (0.67-0.96)	0.02		129 (28.60)	129 (28.04)	1.03 (0.77-1.37)	0.86		0.86 (0.74-1.01)	0.06
		CC	29 (3.76)	87 (5.67)	0.59 (0.38-0.92)	0.02	0.01	15 (3.33)	16 (3.48)	0.96 (0.47-1.98)	0.92	0.98	0.68 (0.47-0.99)	0.04
	rs3204953	Best model multiplicative (number of C's)				0.79 (0.68-0.92)	0.002			1.01 (0.79-1.28)	0.94		0.85 (0.75-0.96)	0.01
		GG	552 (71.60)	1170 (76.42)				326 (71.96)	317 (72.54)					
		GA	199 (25.81)	330 (21.55)	1.27 (1.03-1.57)	0.02		115 (25.39)	110 (25.17)	1.02 (0.75-1.38)	0.92		1.19 (1.01-1.41)	0.04
		AA	20 (2.59)	31 (2.02)	1.34 (0.75-2.40)	0.33	0.05	12 (2.65)	10 (2.29)	1.17 (0.50-2.74)	0.72	0.94	1.30 (0.81-2.10)	0.28
Best model dominant ((GA+AA) vs GG)				1.28 (1.05-1.56)	0.02				1.03 (0.78-1.38)	0.85		1.20 (1.02-1.41)	0.03	
<i>MAD2L2</i>	rs2233004	AA	511 (68.22)	983 (65.36)				274 (61.30)	287 (60.42)					
		AG	222 (29.64)	455 (30.25)	0.92 (0.75-1.11)	0.37		146 (32.66)	153 (32.21)	1.00 (0.76-1.32)	1.00		0.96 (0.82-1.12)	0.59
		GG	16 (2.14)	66 (4.39)	0.48 (0.27-0.84)	0.01	0.03	27 (6.04)	35 (7.37)	0.81 (0.48-1.37)	0.43	0.72	0.61 (0.42-0.89)	0.01
	Best model recessive (GG vs (AA+AG))				0.49 (0.28-0.86)	0.01				0.81 (0.48-1.36)	0.42		0.62 (0.43-0.90)	0.01
	rs2233006	AA	322 (42.93)	644 (43.16)										
		AT	339 (45.20)	684 (45.84)	1.00 (0.82-1.21)	0.98								
		TT	89 (11.87)	164 (10.99)	1.13 (0.83-1.53)	0.43	0.69							
	rs2233011	TT	453 (61.72)	943 (64.19)										
		TG	246 (33.51)	469 (31.93)	1.08 (0.89-1.31)	0.43								
		GG	35 (4.77)	57 (3.88)	1.26 (0.80-1.97)	0.32	0.5							
<i>REVI</i>	rs6761390	AA	181 (24.93)	388 (26.58)										
		AC	362 (49.86)	703 (48.15)	1.12 (0.90-1.39)	0.32								
		CC	183 (25.21)	369 (25.27)	1.07 (0.83-1.38)	0.61	0.61							
	rs3792136	CC	502 (68.96)	985 (67.93)										
		CT	205 (28.16)	420 (28.97)	0.98 (0.80-1.21)	0.85								
		TT	21 (2.88)	45 (3.10)	0.96 (0.57-1.63)	0.88	0.97							
	rs3792142	TT	182 (24.56)	391 (26.33)										
		TG	371 (50.07)	724 (48.75)	1.11 (0.89-1.38)	0.36								
		GG	188 (25.37)	370 (24.92)	1.10 (0.85-1.42)	0.46	0.63							
	rs13409359	AA	555 (74.10)	1106 (74.23)										
		AC	174 (23.23)	342 (22.95)	1.01 (0.82-1.24)	0.96								
		CC	20 (2.67)	42 (2.82)	0.95 (0.53-1.69)	0.85	0.98							
	rs4341989	CC	340 (46.45)	675 (45.89)										
		CT	319 (43.58)	634 (43.10)	1.00 (0.82-1.21)	0.99								
		TT	73 (9.97)	162 (11.01)	0.91 (0.66-1.24)	0.55	0.82							
	rs3792152	CC	286 (38.65)	617 (41.47)										
		CT	357 (48.24)	682 (45.83)	1.15 (0.95-1.39)	0.16								
		TT	97 (13.11)	189 (12.70)	1.14 (0.85-1.52)	0.38	0.35							

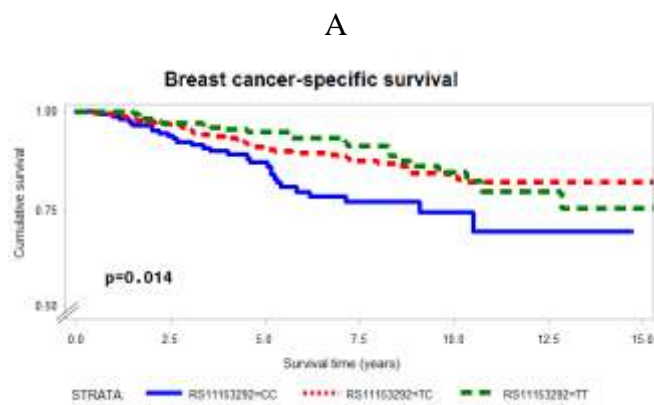
Table 4: Association of *REV3L* SNPs rs11153292 and rs462779 with BC-specific survival in the Swedish study population and with event-free survival (EFS) in the Polish study population.

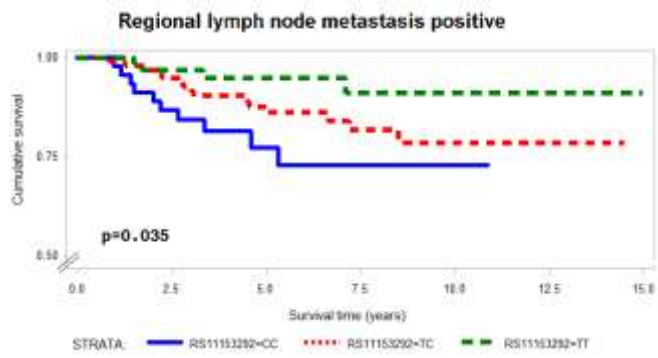
SNP	Genotype	No	No _{event} (%)	HR (95%CI)	p value	HR (95%CI)		p value
						adjusted for size, grade, regional lymph node		
Swedish study population								

		<i>metastasis</i>						
rs11153292	TT	211	10 (4.74)					
	TC	375	27 (7.20)	1.62 (0.78-3.36)	0.19	1.83 (0.81-4.14)	0.14	
	CC	148	17 (11.49)	2.93 (1.34-6.44)	0.01	3.67 (1.56-8.62)	0.003	
	global				0.02		0.01	
	Polish study population *							
	TT	123	18 (14.63)					
	TC	218	41 (18.81)	1.31 (0.75-2.28)	0.34	1.25 (0.71-2.23)	0.44	
	CC	81	15 (18.52)	1.31 (0.66-2.60)	0.44	1.29 (0.64-2.62)	0.48	
	global				0.61		0.70	
rs462779	Swedish study population							
	TT	475	28 (5.89)					
	TC	235	21 (8.94)	1.50 (0.85-2.63)	0.13	2.08 (1.12-3.86)	0.02	
	CC	28	4 (14.29)	2.23 (0.78-6.38)	0.16	3.26 (1.11-9.55)	0.03	
	global				0.18		0.02	
	Polish study population *							
	TT	287	52 (18.12)	1.00				
	TC	116	19 (16.38)	0.96 (0.56-1.62)	0.86	0.92 (0.53-1.60)	0.77	
	CC	13	1 (7.69)	0.40 (0.06-2.91)	0.37	0.56 (0.06-3.31)	0.44	
	global				0.66		0.72	

* For the Swedish population event = death due to breast cancer; for the Polish study population event = local or distant recurrence, contralateral breast cancer or death due to any cause

Figure 1: Kaplan-Meier estimates of breast cancer-specific survival according to rs11153292 genotypes in the Swedish study population. A) all breast cancer patients, B) patients with lymph node metastasis, C) patients with histologic grade 3 tumours, D) patients with stage II – IV tumours.





C



D

