



HAL
open science

Polymorphisms in the and genes of the SUMO-conjugating system and breast cancer risk

Thomas Dünnebier, Justo Lorenzo Bermejo, Susanne Haas, Hans-Peter Fischer, Christiane B. Pierl, Christina Justenhoven, Hiltrud Brauch, Christian Baisch, Michael Gilbert, Volker Harth, et al.

► **To cite this version:**

Thomas Dünnebier, Justo Lorenzo Bermejo, Susanne Haas, Hans-Peter Fischer, Christiane B. Pierl, et al.. Polymorphisms in the and genes of the SUMO-conjugating system and breast cancer risk. *Breast Cancer Research and Treatment*, 2009, 121 (1), pp.185-194. <10.1007/s10549-009-0530-y>. <hal-00485068>

HAL Id: hal-00485068

<https://hal.science/hal-00485068v1>

Submitted on 20 May 2010

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.



HAL Authorization

Polymorphisms in the *UBC9* and *PIAS3* genes of the SUMO-conjugating system and breast cancer risk

Thomas Dünnebier¹, Justo Lorenzo Bermejo², Susanne Haas³, Hans-Peter Fischer³, Christiane B. Pierl⁴, Christina Justenhoven⁵, Hiltrud Brauch⁵, Christian Baisch⁶, Michael Gilbert¹, Volker Harth^{4,6}, Anne Spickenheuer⁴, Sylvia Rabstein⁴, Beate Pesch⁴, Thomas Brüning⁴, Yon-Dschun Ko⁶ and Ute Hamann^{1*}

¹*Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum, Heidelberg, Germany*

²*Institute of Medical Biometry and Informatics, University Hospital Heidelberg, Heidelberg, Germany*

³*Institute of Pathology, Medical Faculty of the University of Bonn, Bonn, Germany*

⁴*BGFA-Forschungsinstitut für Arbeitsmedizin der Deutschen Gesetzlichen Unfall-versicherung, Ruhr University Bochum, Bochum, Germany*

⁵*Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany*

⁶*Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany*

***Address for correspondence and contact**

Ute Hamann, PhD, Professor
Deutsches Krebsforschungszentrum, Molecular Genetics of Breast Cancer B055
Im Neuenheimer Feld 580, 69120 Heidelberg, Germany
Tel.: 0049/6221/42-2344
Fax: 0049/6221/42-4721
Email: u.hamann@dkfz-heidelberg.de

Abstract

SUMOylation consists in the covalent conjugation of small ubiquitin-related modifiers to target proteins. SUMOylation participates in processes that are tightly linked to tumorigenesis and genetic variability in the SUMO-conjugating system may influence the development of breast cancer. We recently reported that variation in the *UBC9* gene encoding the SUMO-conjugating enzyme may affect the grade of breast tumors. [Following comprehensive *in silico* analyses for detection of putative functional polymorphisms in 14 genes of the SUMO system we selected one coding SNP in *PIAS3* and seven tag SNPs in *UBC9* for association analyses.](#) Results were based on 1,021 cases and 1,015 matched controls from the population-based GENICA study. Odds ratios (OR) and 95% confidence intervals (CI) were estimated by conditional logistic regression. To explore the association with polymorphisms closely linked to the genotyped variants, multiple imputation based on HapMap data was applied. The study revealed associations of four *UBC9* polymorphisms with risk of grade 1 tumors. Comparison of genotype and haplotype models indicated that the best representation of risk solely relied on rs7187167 under dominant penetrance. Women carrying the rare allele showed an increased risk of grade 1 tumors compared with common homozygotes (OR 1.87, 95% CI 1.18-2.95). This effect appeared to be stronger in women with a family history of breast or ovarian cancer. Imputation of polymorphisms in a 300 kb region around the genotyped polymorphisms identified no variants with stronger associations. Our findings suggest that genetic variation in *UBC9* may affect the risk of grade 1 breast tumors.

Key words: *UBC9* and *PIAS3* polymorphisms, SUMOylation, breast cancer risk, tumor grade, multiple imputation

Introduction

SUMOylation is an essential cellular process, which consists in the covalent conjugation of small ubiquitin-related modifiers (SUMO-1, 2 and 3) to target proteins [1]. This kind of posttranslational modification can change and regulate the function of a protein by governing protein-protein and protein-DNA interactions. Thereby, SUMO can change the localization, activity or stability of its substrates. SUMOylation is a multistep process involving three classes of enzymes called E1, E2 and E3 [2]. First, mature SUMO is activated by a SUMO-activating enzyme complex composed of the E1 proteins AOS1 and UBA2. Next, SUMO is transferred from the E1 heterodimer to the SUMO-conjugating enzyme UBC9 (UBE2I), which represents the only E2 component of the SUMO system. Finally, UBC9 transfers SUMO to a lysine residue of the substrate resulting in isopeptide bond formation. This last step is facilitated by E3 SUMO ligases such as the protein inhibitors of activated STAT (PIAS) family proteins that confer substrate specificity [1]. Due to the existence of SUMO-deconjugating enzymes, called sentrin-specific proteases (SENPs), SUMOylation is a reversible and highly dynamic process. This class of enzymes also catalyses the proteolytic maturation of the SUMO precursor.

SUMO controls multiple events including transcription, nucleo-cytoplasmic trafficking and mitotic chromosome segregation [3, 4]. Further, it plays a pivotal role in the maintenance of genome integrity by regulating DNA replication, repair and recombination [5].

There is increasing evidence that SUMO is involved in tumorigenesis. Many tumor suppressors and oncoproteins, such as PML, WRN, BLM, c-JUN, c-FOS, TP53 and MDM2, are targets of SUMO [6-13]. In a mouse xenograft model using breast cancer cells, it was shown that increased activity of UBC9 promotes tumor growth and low activity decelerates it [14]. Moreover, several studies have described an up-regulation of enzymes involved in SUMO conjugation or deconjugation in a variety of malignancies [14-18]. In breast cancer specimens a higher expression of *UBC9* [18] and *PIAS3* [16] has been reported. In addition, SUMO regulates

the activity of several nuclear hormone receptors including estrogen receptor alpha (ER), progesterone receptor (PR) and androgen receptor [19], which play a central role in the development of hormone-driven breast tumors [20]. SUMOylation of ER is stimulated by PIAS1 and PIAS3 in a ligand-dependent manner and increases the transcriptional activity of the receptor [21]. SUMO is also conjugated to co-regulators of ER , thereby modulating their ability to interact with the nuclear receptor and to activate transcription [22-24].

Recently, we showed that genetic variation in *UBC9* is associated with the histological grade of breast tumors and is a useful marker for breast cancer prognosis [25]. Due to the close connection of SUMO with tumorigenesis, it is conceivable that genetic [variability in genes](#) of the SUMO-conjugating system also affects breast cancer susceptibility. [The present study addresses this hypothesis. Based on comprehensive *in silico* analyses for detection of putative functional polymorphisms in 14 genes of the SUMO-conjugating system \(*AOS1*, *UBA2*, *UBC9*, *PIAS1*, *PIAS2*, *PIAS3*, *PIAS4*, *SENP1*, *SENP2*, *SENP3*, *SENP5*, *SUMO1*, *SUMO2* and *SUMO3*\) we selected a coding polymorphism in *PIAS3* and seven tag SNPs in *UBC9* for association analyses. Associations with overall breast cancer risk and risks by tumor subtypes were assessed in 1,021 breast cancer cases and 1,015 population-based controls from the German GENICA study.](#)

Materials and methods

Study population

The GENICA study participants of the population-based breast cancer case-control study from the Greater Bonn Region, Germany, were recruited between 08/2000 and 9/2004 as previously described [26-28]. In brief, 1,143 incident breast cancer cases and 1,155 population controls, matched in 5-year classes, participated in the study. The GENICA study was approved by the Ethic's Committee of the University of Bonn and all study participants gave written informed consent. Cases and controls were eligible if they were of Caucasian ethnicity, current residents of

the study region and below 80 years of age. Among the recruited individuals, DNA samples were available for 1,021 (89%) breast cancer cases and 1,015 (88%) controls. Information on known and potential risk factors was collected for all participants via in-person interviews. The response rate was 88% for cases and 67% for controls. Characteristics of the study population regarding potential breast cancer risk factors included age at diagnosis, menopausal status (premenopausal, postmenopausal), family history of at least one first degree relative with breast or ovarian cancer (yes, no), use of oral contraceptives (OC) (never, >0 to <5, 5 to <10, ≥10 years), use of hormone therapy (HT) (never, >0 to <10, ≥10 years), body mass index (BMI) (<20, 20 to <25, 25 to <30, ≥30 kg/m²) and smoking status (never, former, current) (Table 1). Information on clinical and histopathological tumor characteristics was collected and included histology (ductal, lobular, ductolobular), histological grade (G1, G2, G3), tumor size (T1, T2, T3, T4), lymph node status (N0, N≥1), ER status (positive, negative), PR status (positive, negative) and HER2 status (positive, negative) (Table 2). Tumor grade was determined according to the Nottingham Criteria, which comprises formation of tubuli, nuclear pleomorphism and mitotic rate.

Search for putative functional single nucleotide polymorphisms (SNPs) by *in silico* analyses

To identify potential promoter regions, the upstream sequences of the *AOS1*, *UBA2*, *UBC9*, *PIAS1*, *PIAS2*, *PIAS3*, *PIAS4*, *SENP1*, *SENP2*, *SENP3*, *SENP5*, *SUMO1*, *SUMO2* and *SUMO3* genes were analysed by the web-tool PromoterSweep [29]. The analysed sequences encompassed the 3 kb upstream region of the transcription start site (TSS), the region between the TSS and the translation start and the 200 bp downstream region of the translation start. The latter DNA segment was included in the analyses to facilitate the automatic identification of orthologs. In case of more than one possible TSS, the most upstream TSS was used for defining the 3 kb upstream region. All sequence and TSS data were obtained from the database Ensembl (version 54). PromoterSweep employs a combination of algorithms including promoter database

searches, profile matrix searches based on known transcription factor binding sites (TFBs) and *de novo* motif discovery by comparing orthologous promoter regions. The integration of different methods leads to an improved prediction accuracy compared to single method tools. Identified promoter motifs fulfilled three criteria: they (1) showed homology to a promoter region annotated in a promoter database, (2) fitted to a profile matrix of a transcription factor and (3) appeared to be conserved among orthologous promoter regions. Predicted TFBs were screened for polymorphisms with a minor allele frequency (MAF) of more than 1% in the CEU HapMap population of European ancestry [30]. Also non-synonymous coding SNPs in all 14 genes were analyzed for their potential effect on protein function using the web-tools Polyphen [31] and Panther [32, 33].

DNA isolation and genotyping

Genomic DNA was extracted from heparinized blood samples (PuregeneTM, Gentra Systems, Inc., Minneapolis, USA) [27].

Genotyping of *UBC9* polymorphisms (rs7187167, rs11248866, rs909915, rs8052688, rs761059, rs760160 and rs8063) was performed by TaqMan[®] allelic discrimination as previously described [25]. Genotyping of the *PIAS3* polymorphism (rs17354559) was performed by PCR-based restriction fragment length (RFLP) analysis using the forward primer 5'-TGC ACC CAG CCT CAG ATT G-3' and the reverse primer 5'-GGA TCT CAT CAC AAT CTG ATC AG-3' (mismatch is underlined), and the restriction enzyme Hpy188 III. Amplified DNA fragments were digested with 2.5 U Hpy188 III (New England Biolabs, Frankfurt, Germany), separated on a 3% agarose gel containing ethidium bromide (Sigma-Aldrich, Steinheim, Germany) and scored by UV visualization. Fragment sizes are 255 bp and 23 bp for the C allele and 278 bp for the G allele.

Overall call rates ranged from 96.4% to 99.7%. Concordance rates of 135 duplicate samples (6.6%) were equal or greater than 99%. The distribution of genotypes in the control group was consistent with Hardy-Weinberg equilibrium evaluated by Fisher's exact test.

Single SNP analyses

Our study had a 90% power to detect an OR of 1.34 ($\alpha = 0.05$, two-sided test, dominant model, the lowest minor allele frequency of the seven investigated *UBC9* single nucleotide polymorphisms (SNPs) was 11%). For *PIAS3*, the power was 80% to detect an OR of 1.75 given a minor allele frequency of 2% ($\alpha = 0.05$, two-sided test, dominant model).

Associations between each SNP and breast cancer risk were analysed by logistic regression conditional on age in 5-year groups using the software package R (version 2.0.1). Risks were adjusted for potential breast cancer risk factors, which included menopausal status, family history of breast or ovarian cancer, OC use, HT use, BMI and smoking status. P values of Wald tests for individual SNPs ≤ 0.05 were considered statistically significant. Genotype specific risks were calculated as odds ratios (OR) and associated 95% confidence intervals (CI).

Associations between genotypes and risk of breast cancer subtypes were explored upon stratification of cases by histological grade, tumor histology, tumor size, lymph node status, ER, PR and HER2 status. Subgroups of cases were compared to all controls. Only strata with a size greater than five were analysed. Due to our previous results [25], we primarily aimed to investigate possible associations with the risk of tumors stratified by grade. Other stratified analyses were performed in the sense of an exploratory testing. Global probability values for associations between genotypes and grade 1 tumor risk were corrected for multiple comparisons by permutation of the case-control status of each individual (10,000 iterations) [34] considering thirty-two associations (any, 1, 2 and 3 grades x 8 SNPs).

To identify subgroups of women at risk, we further investigated possible interaction effects of each SNP and the epidemiological variables menopausal status, family history of breast or ovarian cancer, use of OC, use of HT, BMI and smoking.

Model selection for the association of risk of grade 1 breast tumors and combinations of SNPs in *UBC9*

The relationship between genotype and grade 1 tumor risk was investigated based on four different penetrance models for rs7187167 and the genetically linked SNPs rs11248866 and rs8052688 (recessive, dominant, additive and three-genotype model). In all models, information on epidemiological risk factors was included as covariates (menopausal status, family history of breast or ovarian cancer, use of OC, use of HT, BMI and smoking). The best model was selected based on likelihood ratio tests using a covariates-only model as reference. The linked SNP rs909915 was not considered in this analysis because the homozygous genotype of the rare allele was not represented in grade 1 tumor patients.

To analyze the possible effect of combinations of SNPs, we inferred haplotypes and haplogenotypes (diplotypes) using the *haplo.em* function in the R package *haplo.stats* [35]. Since the investigated polymorphisms were in linkage disequilibrium, the inference of haplotypes should increase the statistical power in comparison with an evaluation of possible additive or multiplicative interactions. Single SNP models were compared with diplotype models based on likelihood ratio tests, trying to optimize the goodness of fit of the model (large likelihood) and keeping the model as simple as possible (few degrees of freedom). Epidemiological risk factors were included in any model. Model comparisons were limited to those haplogenotypes represented in both grade 1 tumor cases and controls.

Imputation of genotypes

There is evidence that genetic association studies may benefit from combining information across SNP markers and by exploiting existing catalogues of variation [36]. We imputed genotypes based on HapMap data to investigate genetic associations at a finer grid of locations across the genome (detect possible associations with genetic variants that were not genotyped in our study). Imputation relied on inference of haplotypes by means of the expectation-maximisation (EM) algorithm in the presence of partially missing data. In brief, missing alleles were excluded from the calculation of allele frequencies. In the E-step, frequencies of partially missing genotypes were updated looping through all possible genotypes. In the M-step, all existing haplotypes that have alleles identical to the non-missing alleles of this haplotype were updated. The certainty of imputation of genotypes using HapMap data was evaluated by cross-validation and it was represented by minus the logarithm of the probability value (Pval) for Cohen's Kappa between the true and the imputed genotypes. Selection of variants for subsequent analysis of association relied on the visual inspection of recombination rates and imputation accuracies in the ± 500 kb region around the genotyped SNPs. Uncertainty in the imputed genotypes was taken into account in the conditional ordinal logistic regression by bootstrapping from the multinomial distribution of the expected genotypes given the observed, directly genotyped variants (1,000 replicates). The p values referred to a three-genotype model.

Results

Selection of SNPs in genes of the SUMO-conjugating system for association studies

To identify variants in 14 genes of the SUMO conjugating system that may influence the expression or function of the encoded proteins, we performed comprehensive *in silico* analyses using HapMap data [30]. The genes searched for functional SNPs encode the E1 components AOS1 and UBA2, the E2 protein UBC9, the E3 PIAS family members PIAS1, the splice variants PIASx and PIASx , PIAS3 and PIASy, the SUMO-specific proteases SENP1, SENP2, SENP3

and SENP5 and the modifiers SUMO-1, SUMO-2 and SUMO-3. The upstream sequences of these genes were analyzed by a novel *in silico* tool, which employs a combination of different methods to predict potential promoter regions and transcription factors binding to them. None of the currently known polymorphisms with a MAF of more than 1% in the European HapMap population was located in predicted transcription factor binding sites. Screening of these genes for non-synonymous coding polymorphisms revealed one SNP in *PIAS3* (rs17354559, MAF=0.02) resulting in an amino acid substitution of serine to cysteine at position 390 (Ensembl Protein ID ENSP00000376765) that was predicted to be of functional relevance.

For association analyses, we selected the putative functional SNP in *PIAS3* and seven tag SNPs (rs7187167, rs11248866, rs909915, rs8052688, rs761059, rs760160 and rs8063) in the *UBC9* gene and the region 10,000 base pairs upstream of its transcription start as previously described [25].

Associations of SNPs in *UBC9* and *PIAS3* with overall breast cancer risk and risk by tumor subtypes

We analyzed the eight polymorphisms in the *UBC9* and *PIAS3* genes within the GENICA study population. No association with overall breast cancer risk was observed (Table 3, any-grade). Frequencies of *UBC9* haplotypes encompassing SNPs in strong linkage disequilibrium (rs7187167, rs11248866, rs909915 and rs8052688) [25] did not differ significantly between cases and controls (data not shown).

Stratification of cases by histological grade, tumor histology, tumor size, lymph node status, ER, PR and HER2 status revealed associations with tumor grade. Four SNPs in *UBC9* were significantly associated with the risk of grade 1 breast tumors (rs7187167, global $p=0.01$; rs11248866, global $p=0.04$; rs8052688, global $p=0.05$; rs8063, global $p=0.05$) (Table 3).

Model selection for single SNPs based on likelihood ratio tests revealed that a dominant model for rs7187167 was the best representation of the association between *UBC9* genotype and risk of grade 1 breast cancer (Supplementary Table 1). Since the four *UBC9* SNPs rs7187167, rs11248866, rs909915 and rs8052688 are in linkage disequilibrium, we investigated whether the risk of grade 1 tumors was affected by combinations of SNPs rather than by single polymorphisms. The dominant model based on rs7187167 was augmented by additional polymorphisms and likelihood ratio tests showed a non significant improvement of the goodness of fit. Model comparisons were limited to haplogenotypes represented in both grade 1 tumor cases and controls.

Based on the dominant model and conditional logistic regression, women with the rare allele of *UBC9* rs7187167 had a higher risk of grade 1 breast tumors than women homozygous for the common allele (OR 1.87, 95% CI 1.18-2.95). The effect was more pronounced when cases had a family history of breast or ovarian cancer (14 women affected by grade 1 tumors with a family history, OR 2.18, 95% CI 0.70-6.77), however was statistically not significant.

To correct for multiple testing (any, 1, 2 and 3 grades x 8 SNPs) permutation analysis was performed. After correction, associations between *UBC9* polymorphisms and grade 1 breast cancer risk did not remain statistically significant.

Imputation of untyped SNPs flanking the genotyped SNPs in *UBC9*

Multiple imputation based on HapMap data was applied to investigate if rs7187167 or untyped SNPs in its proximity were responsible for the association with grade 1 tumor risk. Visual inspection of recombination rates in a 500 kb region centered on the genotyped SNPs, together with the results on certainty of imputation, suggested the selection of a 300 kb region comprising 210 SNPs, which showed heterozygosity in HapMap (Supplementary Figure 1). Conditional logistic regression based on genotyped and imputed SNPs indicated that rs7187167 shows the

strongest association with risk of grade 1 breast cancer (Figure 1A). rs7187167 showed no significant association with risk of grade 2, grade 3 and any-grade breast tumors (Figures 1B, 1C and 1D).

Association between SNPs and breast cancer risk by epidemiological parameters

To identify subgroups of women at risk, we further investigated possible interaction effects of the genotyped polymorphisms and menopausal status, family history of breast or ovarian cancer, OC use, HT use, BMI and smoking status. No statistically significant interactions were found.

Discussion

There is increasing evidence of a crucial role of SUMO modification in tumorigenesis. Therefore, we hypothesized that variation in genes of the SUMO-conjugating system may affect the risk of breast cancer. In an initial step we screened 14 genes involved in SUMO modification for putative functional polymorphisms in the promoter and coding regions by performing comprehensive *in silico* analyses. We obtained evidence for one functional polymorphism in the coding region of the SUMO ligase gene *PIAS3*. In addition, we selected seven tag SNPs covering the complete promoter and gene region of *UBC9*, which encodes a key enzyme of SUMOylation. All SNPs were analyzed for associations with breast cancer risk in the population-based case-control study GENICA.

None of the investigated polymorphisms seemed to alter the overall risk of breast cancer. However, stratification of patients by tumor grade revealed that four SNPs in *UBC9* (rs7187167, rs11248866, rs8052688 and rs8063) were associated with the risk of grade 1 tumors. The best association model included rs7187167 under dominant penetrance, and this model was not improved by inclusion of additional SNPs in linkage disequilibrium. Women carrying the rare allele of rs7187167 had a risk of grade 1 tumors 1.9 times higher than women homozygous for

the common allele. The identification of three additional association signals in the region adds consistency to this finding. Individual variants are expected to show stronger effects in familial, genetically enriched cases [37]. Although not significant, we found that carriers with a family history of breast or ovarian cancer had a 2.2 times increased risk of grade 1 tumors, thus further supporting the biological relevance of rs7187167 variation. Results from multiple imputation confirmed that rs7187167 showed the strongest association with the risk of grade 1 tumors. Thus, the identification of multiple associations in the region, the stronger effect for familial cases and the imputation results suggest a contribution of rs7187167 to the risk of grade 1 breast tumors.

Multiple imputation has been shown to facilitate the detection of causal variants that have not been directly genotyped in the study [36]. We applied imputation techniques followed by conditional logistic regression to assess if rs7187167, or rather a linked SNP in its proximity, was responsible for the observed association. Imputation of genotypes relied on the assumption of similar patterns of genetic linkage in the GENICA population and in the public repository HapMap. In this study, none of the imputed SNPs in the investigated gene region showed a signal of association stronger than that of rs7187167. Although the causal variant is yet unidentified, the result suggests that rs7187167 or a closely linked SNP may be responsible for the observed association with risk of grade 1 breast tumors. The causal variant could be a rarer SNP close to rs7187167, which was not genotyped or imputed in our study.

As yet, functional studies on rs7187167 are lacking. The polymorphism is located 14.8 kb upstream of the *UBC9* translation start and disrupts a potential binding site for the transcription factor SP1 [25]. In consequence, the rare allele may result in a lower *UBC9* expression. To date, however, data on *UBC9* expression levels supporting this hypothesis are not available. One previous study using an *in silico* approach to predict putative genes on the short arm of chromosome 16 suggested an open reading frame encompassing rs7187167 [38]. Thus, we

cannot exclude the possibility that this polymorphism influences the function of a yet unidentified protein.

Many tumor suppressors and oncoproteins have been identified as SUMO substrates emphasizing SUMO's crucial role in processes that are tightly linked to cancerogenesis. Although SUMO modification seems to be highly regulated, both spatially and temporally, it is conceivable that expression levels of *UBC9* influence the balance between SUMO conjugation and deconjugation and thus the function of target proteins. Several SUMO targets are involved in cell differentiation, proliferation and cell cycle control [39], such as c-JUN, c-FOS [40], p53 [41, 42] and ER [43], providing a possible link between SUMOylation and tumor grade. Two recent studies have reported a link between decreased SUMO conjugation and reduced growth of tumor cells. The first study used a mouse xenograft model and breast cancer cells to show that decreased *UBC9* activity slows down tumor growth *in vivo* [14]. Yang and Paschen reported that blocking SUMO-2 and 3 conjugation altered the expression of genes involved in cell proliferation and differentiation and resulted in reduced growth of neuroblastoma cells *in vitro* [44]. Thus, the assumption of rs7187167 or a closely linked SNP attenuating the expression of *UBC9* would be in line with its effect of favouring the development of slowly growing grade 1 tumors. Moreover, the increased risk of grade 1 tumors associated with the rare allele of rs7187167 is in agreement with our previous results of a higher frequency of this allele in patients with low grade tumors [25]. It is important to underline that our previous and the present study differ with respect to study design, case-only and case-control study, respectively and analyzed different events. For example, in the former study, we showed a higher frequency of the rare rs7187167 TT genotype (14.3%) in grade 1 cases compared to grade 2 (8.5%) and grade 3 (4.1%) cases. In the present study we additionally reported genotype frequencies in healthy controls. The present findings indicate that women affected by grade 1 breast tumors showed a higher TT genotype frequency (14.9%) than grade 2 (8.6%) and grade 3 (4.2%) cases as well as

unaffected controls (6.8%), which refers to an increased risk of grade 1 breast cancer among TT carriers.

In summary, in this study on genetic variation in the SUMO-conjugating system, we found associations between four *UBC9* SNPs (rs7187167, rs11248866, rs8052688 and rs8063) and the risk of grade 1 breast tumors. The strongest association was observed for rs7187167 and the effect appeared to be stronger in familial cases. A model selection strategy and multiple imputation indicated that, among genotyped and closely linked variants analyzed in the International HapMap Project, rs7187167 showed the most significant contribution to risk. Our data suggest that *UBC9* variation influences the risk of grade 1 breast tumors. Replication studies in independent populations should be carried out to confirm this result.

Acknowledgements

We are indebted to all women participating in the GENICA study. We gratefully acknowledge support by interviewers as well as physicians and pathologists of the study region. We thank Axel Benner for his contribution to the statistical analysis and Agnes Hotz-Wagenblatt as well as Karl-Heinz Glatting for their support in using the software PromoterSweep. Further gratitude goes to Antje Seidel-Renkert for expert technical assistance.

This work was supported by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9976/8, 01KW9975/5, 01KW9977/0 and 01KW0114, the Deutsches Krebsforschungszentrum, Heidelberg, the Robert Bosch Foundation of Medical Research, Stuttgart, BGFA-Forschungsinstitut für Arbeitsmedizin der Deutschen Gesetzlichen Unfallversicherung, Bochum, and the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany.

Competing interests: The authors declare that they have no competing interests.

References

1. Geiss-Friedlander R, Melchior F (2007) Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol* 8:947-956
2. Muller S, Hoege C, Pyrowolakis G, Jentsch S (2001) SUMO, ubiquitin's mysterious cousin. *Nat Rev Mol Cell Biol* 2:202-210
3. Dasso M (2008) Emerging roles of the SUMO pathway in mitosis. *Cell Div* 3:5
4. Hay RT (2005) SUMO: a history of modification. *Mol Cell* 18:1-12
5. Seeler JS, Bischof O, Nacerddine K, Dejean A (2007) SUMO, the three Rs and cancer. *Curr Top Microbiol Immunol* 313:49-71
6. Gostissa M, Hengstermann A, Fogal V, Sandy P, Schwarz SE, Scheffner M, Del Sal G (1999) Activation of p53 by conjugation to the ubiquitin-like protein SUMO-1. *EMBO J* 18:6462-6471
7. Muller S, Berger M, Lehembre F, Seeler JS, Haupt Y, Dejean A (2000) c-Jun and p53 activity is modulated by SUMO-1 modification. *J Biol Chem* 275:13321-13329
8. Muller S, Matunis MJ, Dejean A (1998) Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. *EMBO J* 17:61-70
9. Rodriguez MS, Desterro JM, Lain S, Midgley CA, Lane DP, Hay RT (1999) SUMO-1 modification activates the transcriptional response of p53. *EMBO J* 18:6455-6461
10. Eladad S, Ye TZ, Hu P, Leversha M, Beresten S, Matunis MJ, Ellis NA (2005) Intra-nuclear trafficking of the BLM helicase to DNA damage-induced foci is regulated by SUMO modification. *Hum Mol Genet* 14:1351-1365
11. Kawabe Y, Seki M, Seki T, Wang WS, Imamura O, Furuichi Y, Saitoh H, Enomoto T (2000) Covalent modification of the Werner's syndrome gene product with the ubiquitin-related protein, SUMO-1. *J Biol Chem* 275:20963-20966
12. Bossis G, Malnou CE, Farras R, Andermarcher E, Hipskind R, Rodriguez M, Schmidt D, Muller S, Jariel-Encontre I, Piechaczyk M (2005) Down-regulation of c-Fos/c-Jun AP-1 dimer activity by sumoylation. *Mol Cell Biol* 25:6964-6979
13. Xirodimas DP, Chisholm J, Desterro JM, Lane DP, Hay RT (2002) P14ARF promotes accumulation of SUMO-1 conjugated (H)Mdm2. *FEBS Lett* 528:207-211
14. Mo YY, Yu Y, Theodosiou E, Rachel Ee PL, Beck WT (2005) A role for Ubc9 in tumorigenesis. *Oncogene* 24:2677-2683
15. Cheng J, Bawa T, Lee P, Gong L, Yeh ET (2006) Role of desumoylation in the development of prostate cancer. *Neoplasia* 8:667-676
16. Wang L, Banerjee S (2004) Differential PIAS3 expression in human malignancy. *Oncol Rep* 11:1319-1324
17. Moschos SJ, Smith AP, Mandic M, Athanassiou C, Watson-Hurst K, Jukic DM, Edington HD, Kirkwood JM, Becker D (2007) SAGE and antibody array analysis of melanoma-infiltrated lymph nodes: identification of Ubc9 as an important molecule in advanced-stage melanomas. *Oncogene* 26:4216-4225
18. Wu F, Zhu S, Ding Y, Beck WT, Mo YY (2009) MicroRNA-mediated Regulation of Ubc9 Expression in Cancer Cells. *Clin Cancer Res* 15:1550-1557
19. Karamouzis MV, Konstantinopoulos PA, Badra FA, Papavassiliou AG (2008) SUMO and estrogen receptors in breast cancer. *Breast Cancer Res Treat* 107:195-210
20. Ali S, Coombes RC (2000) Estrogen receptor alpha in human breast cancer: occurrence and significance. *J Mammary Gland Biol Neoplasia* 5:271-281
21. Sentis S, Le Romancer M, Bianchin C, Rostan MC, Corbo L (2005) Sumoylation of the estrogen receptor alpha hinge region regulates its transcriptional activity. *Mol Endocrinol* 19:2671-2684

22. Chauchereau A, Amazit L, Quesne M, Guiochon-Mantel A, Milgrom E (2003) Sumoylation of the progesterone receptor and of the steroid receptor coactivator SRC-1. *J Biol Chem* 278:12335-12343
23. Kotaja N, Karvonen U, Janne OA, Palvimo JJ (2002) The nuclear receptor interaction domain of GRIP1 is modulated by covalent attachment of SUMO-1. *J Biol Chem* 277:30283-30288
24. Wu H, Sun L, Zhang Y, Chen Y, Shi B, Li R, Wang Y, Liang J, Fan D, Wu G et al (2006) Coordinated regulation of AIB1 transcriptional activity by sumoylation and phosphorylation. *J Biol Chem* 281:21848-21856
25. Dunnebier T, Bermejo JL, Haas S, Fischer HP, Pierl CB, Justenhoven C, Brauch H, Baisch C, Gilbert M, Harth V et al (2009) Common variants in the UBC9 gene encoding the SUMO-conjugating enzyme are associated with breast tumor grade. *Int J Cancer* 125:596-602
26. Pesch B, Ko Y, Brauch H, Hamann U, Harth V, Rabstein S, Pierl C, Fischer HP, Baisch C, Justenhoven C et al (2005) Factors modifying the association between hormone-replacement therapy and breast cancer risk. *Eur J Epidemiol* 20:699-711
27. Justenhoven C, Hamann U, Pesch B, Harth V, Rabstein S, Baisch C, Vollmert C, Illig T, Ko YD, Bruning T et al (2004) ERCC2 genotypes and a corresponding haplotype are linked with breast cancer risk in a German population. *Cancer Epidemiol Biomarkers Prev* 13:2059-2064
28. Justenhoven C, Pierl CB, Haas S, Fischer HP, Baisch C, Hamann U, Harth V, Pesch B, Bruning T, Vollmert C et al (2008) The CYP1B1_1358_GG genotype is associated with estrogen receptor-negative breast cancer. *Breast Cancer Res Treat* 111:171-177
29. del Val C, Pelz, O., Glatting, K-H, Barta, E., Hotz-Wagenblatt, A. (2009) PromoterSweep: A Tool for Identification of Transcription Factor Binding Sites. *Theor Chem Acc*
30. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM et al (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851-861
31. Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 30:3894-3900
32. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* 13:2129-2141
33. Thomas PD, Kejariwal A, Guo N, Mi H, Campbell MJ, Muruganujan A, Lazareva-Ulitsky B (2006) Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res* 34(Web Server issue):W645-650
34. Westfall PH, Young SS (1993) Resampling-Based Multiple Testing. John Wiley & Sons, New York
35. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 70:425-434
36. Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 39:906-913
37. Antoniou AC, Easton DF (2003) Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet Epidemiol* 25:190-202
38. Daniels RJ, Peden JF, Lloyd C, Horsley SW, Clark K, Tufarelli C, Kearney L, Buckle VJ, Doggett NA, Flint J et al (2001) Sequence, structure and pathology of the fully annotated terminal 2 Mb of the short arm of human chromosome 16. *Hum Mol Genet* 10:339-352
39. Zhao J (2007) Sumoylation regulates diverse biological processes. *Cell Mol Life Sci* 64:3017-3033

40. Shen Q, Uray IP, Li Y, Krisko TI, Strecker TE, Kim HT, Brown PH (2008) The AP-1 transcription factor regulates breast cancer cell growth via cyclins and E2F factors. *Oncogene* 27:366-377
41. Vousden KH, Lane DP (2007) p53 in health and disease. *Nat Rev Mol Cell Biol* 8:275-283
42. Vogelstein B, Lane D, Levine AJ (2000) Surfing the p53 network. *Nature* 408:307-310
43. Chen GG, Zeng Q, Tse GM (2008) Estrogen and its receptors in cancer. *Med Res Rev* 28:954-974
44. Yang W, Paschen W (2009) Gene expression and cell growth are modified by silencing SUMO2 and SUMO3 expression. *Biochem Biophys Res Commun*

Figure legends

Figure 1. Association between risk of breast cancer and genotyped (black circles) or imputed (grey circles) *UBC9* polymorphisms. The plot shows the $-\log_{10} p$ values for a three-genotype model versus a model, which does not include individual genotypes. Risk of grade 1 (A), grade 2 (B), grade 3 (C) and any-grade (D) breast tumors were evaluated. Multiple imputation was based on HapMap data and the called SNPs rs7187167, rs11248866, rs909915 and rs8052688 (black circles). Association of rs909915 with grade 1 tumor risk could not be assessed because the homozygous genotype of the rare allele was not represented in grade 1 tumor patients.

Table 1 Characteristics of the GENICA study population

Characteristic		Cases n (%)	Controls n (%)
Age (years)	20-44	126 (12.3)	128 (12.6)
	45-49	99 (9.7)	98 (9.7)
	50-54	156 (15.3)	151 (14.9)
	55-59	142 (13.9)	138 (13.6)
	60-64	197 (19.3)	197 (19.4)
	65-69	141 (13.8)	138 (13.6)
	≥70	160 (15.7)	165 (16.3)
Menopausal status	Premenopausal	249 (24.7)	235 (23.5)
	Postmenopausal	758 (75.3)	764 (76.5)
Breast or ovarian cancer in first-degree relatives	No	860 (84.2)	927 (91.3)
	Yes	161 (15.8)	88 (8.7)
OC use (years)	Never	372 (36.5)	368 (36.3)
	>0 to <5	180 (17.7)	185 (18.3)
	5 to <10	134 (13.2)	120 (11.8)
	≥10	333 (32.7)	340 (33.6)
HT use (years)	Never	506 (49.8)	509 (50.2)
	>0 to <10	245 (24.1)	290 (28.6)
	≥10	266 (26.2)	214 (21.1)
BMI (kg/m ²)	<20	90 (8.8)	73 (7.2)
	20 to <25	469 (45.9)	471 (46.4)
	25 to <30	306 (30.0)	324 (32.0)
	≥30	156 (15.3)	146 (14.4)
Smoking	Never	586 (57.5)	555 (54.7)
	Former	192 (18.8)	215 (21.2)
	Current	242 (23.7)	245 (24.1)

OC: oral contraceptive, HT: hormone therapy, BMI: body mass index

Table 2 Histopathological parameters of the incident breast tumors of the GENICA cases

Tumor parameter		Cases n (%)
Histology	Ductal	634 (69.5)
	Lobular	177 (19.4)
	Ductolobular	101 (11.1)
Histological grade	G1	77 (8.2)
	G2	567 (60.4)
	G3	295 (31.4)
Tumor size	T1	582 (61.9)
	T2	289 (30.7)
	T3	30 (3.2)
	T4	39 (4.1)
Nodal status	N0	602 (63.8)
	≥N1	342 (36.2)
ER status	Positive	755 (77.8)
	Negative	216 (22.2)
PR status	Positive	678 (70.0)
	Negative	291 (30.0)
HER2 status	Positive	189 (27.7)
	Negative	493 (72.3)

ER: estrogen receptor, PR: progesterone receptor

Table 3 Odds ratios for breast cancer risk by tumor grade

Gene/SNP	Genotype	Controls n (%)	Any-grade			Grade 1			Grade 2			Grade 3		
			Cases n (%)	OR _{adj} ^a (95% CI)	Globbal <i>p</i> value ^b	Cases n (%)	OR _{adj} ^a (95% CI)	Globbal <i>p</i> value ^b	Cases n (%)	OR _{adj} ^a (95% CI)	Globbal <i>p</i> value ^b	Cases n (%)	OR _{adj} ^a (95% CI)	Globbal <i>p</i> value ^b
<i>UBC9</i>														
rs7187167	CC	526 (52.9)	519 (52.1)	1.00 (reference)		29 (2.9)	1.00 (reference)		291 (51.9)	1.00 (reference)		158 (55.6)	1.00 (reference)	
C>T	CT	400 (40.2)	400 (40.1)	1.01 (0.92-1.11)	0.55	34 (4.5)	1.71 (1.04-2.79)	0.01	222 (39.6)	1.01 (0.88-1.17)	0.26	114 (40.1)	0.95 (0.77-1.17)	0.34
	TT	68 (6.8)	78 (7.8)	1.09 (0.93-1.29)		11 (1.4)	2.60 (1.38-4.89)		48 (8.6)	1.22 (0.96-1.55)		12 (4.2)	0.64 (0.38-1.08)	
rs11248866	AA	385 (38.8)	383 (38.5)	1.00 (reference)		19 (2.5)	1.00 (reference)		213 (38.1)	1.00 (reference)		119 (41.9)	1.00 (reference)	
A>G	AG	469 (47.2)	462 (46.4)	0.99 (0.90-1.09)	0.49	39 (5.2)	1.72 (1.01-2.92)	0.04	262 (46.9)	1.01 (0.87-1.16)	0.49	131 (6.1)	0.91 (0.73-1.13)	0.62
	GG	139 (14.0)	150 (15.1)	1.05 (0.92-1.20)		16 (2.1)	2.19 (1.17-4.11)		84 (15.0)	1.10 (0.90-1.34)		34 (12.0)	0.81 (0.58-1.14)	
rs909915	CC	784 (79.0)	789 (79.2)	1.00 (reference)		59 (7.9)	1.00 (reference)		441 (78.8)	1.00 (reference)		226 (79.6)	1.00 (reference)	
C>T	CT	195 (19.6)	187 (18.8)	0.97 (0.86-1.09)	0.46	15 (2.0)	1.56 (0.98-2.47)	0.93	110 (19.6)	0.99 (0.83-1.17)	0.82	50 (17.6)	0.90 (0.68-1.18)	0.22
	TT	14 (1.4)	20 (2.0)	1.18 (0.89-1.57)		0 (0.0)	°		9 (1.6)	1.12 (0.67-1.88)		8 (2.8)	1.56 (0.90-2.71)	
	CT+	209 (21.0)	207 (20.8)	0.99 (0.88-1.10)		15 (2.0)	0.87 (0.51-1.52)							
rs8052688	GG	691 (69.5)	701 (70.8)	1.00 (reference)		43 (5.8)	1.00 (reference)		395 (71.2)	1.00 (reference)		210 (73.9)	1.00 (reference)	
G>C	GC	279 (28.1)	261 (26.4)	0.96 (0.87-1.07)	0.26	25 (3.3)	1.56 (0.98-2.47)	0.05	147 (26.5)	0.97 (0.84-1.13)	0.50	68 (23.9)	0.83 (0.65-1.05)	0.27
	CC	24 (2.4)	28 (2.8)	1.07 (0.83-1.39)		6 (8.1)	2.49 (1.16-5.33)		13 (2.3)	0.97 (0.61-1.52)		6 (2.1)	0.81 (0.40-1.65)	
rs761059	GG	338 (34.6)	323 (34.0)	1.00 (reference)		29 (4.2)	1.00 (reference)		164 (30.7)	1.00 (reference)		102 (37.2)	1.00 (reference)	
G>A	GA	477 (48.9)	460 (48.4)	1.01 (0.91-1.12)	0.15	29 (4.2)	0.79 (0.48-1.28)	0.44	279 (52.2)	1.14 (0.97-1.33)	0.12	119 (43.4)	0.88 (0.70-1.12)	0.08
	AA	161 (16.5)	167 (17.6)	1.04 (0.91-1.18)		10 (1.4)	0.72 (0.35-1.46)		91 (17.0)	1.08 (0.88-1.33)		53 (19.3)	1.11 (0.84-1.48)	
rs761060	GG	688 (69.6)	655 (66.8)	1.00 (reference)		45 (6.1)	1.00 (reference)		368 (67.0)	1.00 (reference)		189 (67.3)	1.00 (reference)	
G>A	GA	264 (26.7)	296 (30.2)	1.08 (0.98-1.18)	0.29	25 (3.4)	°	0.46	166 (30.2)	1.11 (0.96-1.28)	0.33	83 (29.5)	1.07 (0.86-1.33)	0.71
	AA	37	30	0.92		3	°		15	0.85		9	0.92	

		(3.7)	(3.1)	(0.71-1.21)		(4.1)		(2.7)	(0.55-1.31)		(3.2)	(0.50-1.70)
	GA+	301	326	1.06		28	1.35					
	AA	(30.4)	(33.2)	(0.97-1.16)		(38.4)	(0.87-2.09)					
rs8063	GG	495	508	1.00		27	1.00		289	1.00	152	1.00
		(50.3)	(51.7)	(reference)		(38.0)	(reference)		(52.5)	(reference)	(53.7)	(reference)
	G>A	409	388	0.96	0.36	35	1.53	0.05	217	0.94	108	0.87
		(41.6)	(39.5)	(0.87-1.05)		(49.3)	(0.94-2.49)		(39.4)	(0.82-1.08)	(38.2)	(0.70-1.08)
	AA	80	86	1.03		9	1.79		45	1.00	23	0.90
		(8.1)	(8.8)	(0.88-1.20)		(12.7)	(0.89-3.61)		(8.2)	(0.78-1.29)	(8.1)	(0.61-1.32)
<hr/>												
<i>PIAS3</i>												
rs1735	CC	950	949	1.00		69	1.00		529	1.00	276	1.00
4559		(95.7)	(95.2)	(reference)		(94.5)	(reference)		(94.3)	(reference)	(96.8)	(reference)
	C>G	42	48	^c	0.78	4	^c	0.82	32	^c	9	^c
		(4.2)	(4.8)			(5.5)			(5.7)		(3.2)	
	GG	1	0	^c		0	^c		0	^c	0	^c
		(0.1)	(0.0)			(0.0)			(0.0)		(0.0)	
	CG+	43	48	1.03		4	1.37		32	1.15	9	0.71
	GG	(4.3)	(4.8)	(0.85-1.26)		(5.5)	(0.53-3.56)		(5.7)	(0.88-1.50)	(3.2)	(0.39-1.30)

^a Odds ratio conditional on age in 5-year groups adjusted for menopausal status, family history of breast or ovarian cancer, use of oral contraceptives, use of hormone therapy, body mass index and smoking.

^b P value of testing the null hypothesis of no association between SNP and risk of any-grade, grade 1, grade 2 and grade 3 breast tumors based on conditional logistic regression, not corrected for multiple comparisons. Statistically significant results are given in bold.

^c Not analyzed due to small numbers of cases and/or controls.